Eindhoven University of Technology

BIM Based Cost Estimation Knowledge Management

The application of a Knowledge Management System by combining BIM models and the data mining process to estimate the costs of residential buildings

Y. Rekveld

Master Construction Management and Engineering

May 2017

Graduation committee:

Prof. Dr. Ir. B. (Bauke) de Vries	University supervisor (chairman graduation committee)
Dr. Dipl.-Ing. J. (Jakob) Beetz	University supervisor
Prof. Dr. C.C.P. (Chris) Snijders	University supervisor

This page is intentionally left blank

Colofon

Title: BIM Based Cost Estimation Knowledge Management

Subtitle: The application of a Knowledge Management System by combining BIM models and the data mining process to estimate the costs of residential buildings

Author:	Y. (Yonne) Rekveld
Student number:	0916003
E-mail:	yrekveld@gmail.com
University:	Eindhoven University of Technology
Graduation program:	Master Construction Management and Engineering

Graduation date: 11th of May 2017

Graduation committee:
Prof. Dr. Ir. B. (Bauke) de Vries
University supervisor (chairman graduation committee)
b.d.vries@tue.nl

Dr. Dipl.-Ing. J. (Jakob) Beetz
University supervisor
j.beetz@tue.nl

Prof. Dr. C.C.P. (Chris) Snijders
University supervisor
c.c.p.snijders@tue.nl

Company:
IBIS BV
Groot Handelsgebouw
Stationsplein 45, unit A6.002
3013 AK Rotterdam

J. (Jeroen) van de Laar
External graduation supervisor
j.van.de.laar@ibis.nl

T. (Toine) Bullens
External graduation supervisor
t.bullens@brink.nl

This page is intentionally left blank

Summary

Due to a change in what leading clients demand from the architecture, engineering and construction (AEC) industry, the AEC industry has to work more efficient. A possible solution is to apply knowledge management (KM) within the organisation. However, the problem with traditional KM techniques used in the AEC industry is that they demand too much maintenance and they only store but do not share information within the organisation. With the development of Building Information Modelling (BIM), a new possibility to create a knowledge management system has emerged. BIM provides a unique source of information as it generates and manages the data created during the building life cycle. This unique source of information can be converted into knowledge by means of data mining.

The problem this research has focused on, is that the current design process of the AEC industry is inefficient. One of the reasons is that the knowledge management systems (KMS) within the AEC industry do not meet the requirements necessary to be effective. The design process is very broad and has multiple tasks that can be made more efficient with KM. As this research is based on the limitations of the research done by Hoogeveen (2015), the KMS developed in this research has the same tasks, namely cost prediction in the early stage of the design process.

Hoogeveen (2015) proved part of this concept by using a data mining approach to predict the costs of datacentres, however he did not use the actual BIM models in his data mining process. In this research BIM models are used for automated data extraction for the data mining process. This data mining process is used to create a knowledge management system for conceptual cost estimation for residential buildings.

Based on this problem definition the following main research question has been drafted: “Can the use of a knowledge management system using BIM data and the data mining process create a more efficient process of cost estimation for residential buildings?” To answer these question a knowledge management system was created based on a literature review, which was tested with collected data.

In the literature review it was found that knowledge management can be defined as a tool set for the automation of deductive or inherent relationships between information objects, corporate users and business processes. To understand knowledge management, at first it had to be understood what knowledge is. Knowledge is defined as refined information, which means that it has meaning. Knowledge can be divided in two forms, tacit or explicit knowledge. The main difference between tacit and explicit knowledge is that explicit knowledge is documented, where tacit knowledge is implied. However, for knowledge to be useful for KM it needed to be made explicit. Especially because the goal of KM is to create new knowledge. This can only be done by combining multiple forms of explicit knowledge.

The suggested source for KM in this research is BIM. BIM is an intelligent simulation of architecture that exhibits the following six key characteristics, digital, spatial (3D), measurable, comprehensive, accessible and durable (Campbell 2006). The main product of BIM is a BIM model which is a virtual prototype of the designed building. A BIM model is characterized by building components that include data that describe how they behave and are consistent and non-redundant data (Eastman et al., 2011). There are two types of BIM models, a BIM model with a proprietary file format and a BIM model with an universal supported standard, generally the IFC file format. In this research the IFC data format was used.

To convert the data within the IFC models to knowledge the data mining process was used. Applied to the problem of converting data within the IFC model to knowledge in form of a system to estimate the costs of residential buildings in the early stage of the design process, the data mining process consists of the following steps:

1. Understanding and defining the problem
2. Collect necessary data
3. Convert non-IFC BIM models to the IFC data format
4. Convert the IFC models to a dataset
5. Add the corresponding costs
6. Clean the dataset
7. Select a useful subset
8. Engineer a useful data mining algorithm
9. Use the algorithm
10. Evaluate results with numerical prediction measures
11. Redefine data and problem
12. Use the generated model for cost prediction

The main obstacle within this research was the extraction of the information from BIM models into a dataset. In theory it should be possible to extract all physical data of the buildings from the BIM models. However during the practical implementation, the current quality of the BIM models provided problems, which have been solved by creating an extraction script that was made specific to overcome these problems.

However after applying the data mining algorithms on the dataset, no usable relations between the data could be found. It was proposed that the dataset was too small and that there was too much variation in the design of the buildings as both apartments and houses were used in the dataset. To test this assumption, addition data were collected from project analyses available in databases. These additional data have been used to create two new datasets. One dataset, where all the data of both the original data and the new data were combined and another dataset with only one type of buildings in the dataset, namely apartments.

These new datasets were tested by applying the same data mining process. From the results it could be concluded that the made assumptions were correct. The dataset with all the data showed that more data give a more consistent result, where the dataset with only apartments shows that the accuracy of the prediction are acceptable for conceptual cost estimation if only one type of building is used in the dataset.

To test the usefulness of the results, the results were used to estimate the costs of the buildings within the dataset. Although the results were close enough to use the generated cost estimation model to calculate a conceptual estimate for building project, the cost estimation model was only tested on the dataset that was used to create the model, so it should be noted that it is yet unclear how precise the model will be on unseen data.

Answering the research question, literature review has shown that the data mining process is an effective tool to create a knowledge management system and that BIM data have a great potential to serve as the data source. However during the practical implementation it became clear that there are both qualitative and quantitative problems with the available BIM models. Additional data were added to test viability of the data mining process for cost calculation for

residential buildings. With this additional data the data mining process could predict the costs of buildings within the dataset precisely enough to be a conceptual cost estimate. Nonetheless until the qualitative and quantitative problems of BIM data are solved the knowledge management system will still need the manual input of additional data to be able to estimate the costs of residential buildings.

This page is intentionally left blank

Samenvatting

Door een veranderde vraag binnen de bouwsector worden bouwbedrijven gedwongen om efficiënter te werken. Een mogelijke oplossing is het toepassen van kennismanagement (KM). Echter de traditionele kennismanagementtechnieken die gebruikt worden binnen de bouw voldoen niet. De reden is dat ze teveel onderhoud vragen en dat de kennis alleen kan worden opgeslagen en niet gedeeld worden binnen de organisatie. Door de opkomst van Building Information Modelling (BIM) zijn er nieuwe mogelijkheden ontstaan om een KM systeem te maken. BIM is namelijk een unieke bron van informatie gezien het data genereert en beheert gedurende de hele levensloop van het gebouw.

Dit onderzoek richt zich op het huidig ontwerpproces in de bouwsector, welke inefficiënt is. Een van de redenen hiervoor is dat de kennismanagementsystemen niet voldoen aan de eisen die nodig zijn om effectief te zijn. Gezien de omvang van het ontwerpproces zijn er veel taken te onderscheiden in dit proces. Dit onderzoek zal zich focussen op het calculeren van de bouwkosten in het vroege stadium van het ontwerpproces. Hiervoor is gekozen omdat dit onderzoek voortborduurt op een eerder onderzoek van Hoogeveen (2015) en daarbij verder ingaat op de beperkingen van zijn onderzoek.

Hoogeveen (2015) bewees gedeeltelijk de werking van een kennismanagementsysteem dat gebruik maakt van BIM, door data mining toe te passen op data over de bouw van datacenters. Hij maakte daarbij echter geen gebruik van de daadwerkelijke BIM modellen. In dit onderzoek worden BIM modellen gebruikt om op geautomatiseerde wijze data uit de BIM modellen te halen en deze te gebruiken in het data mining proces. Daarbij zal dit proces gebruikt worden als kennismanagementsysteem voor de conceptuele kostencalculatie voor woningen.

Gebaseerd op de bovenstaande probleemstelling is de volgende onderzoeksvraag opgesteld: “Kan het gebruiken van een kennismanagementsysteem, dat BIM data en het data mining proces gebruikt, zorgen voor een efficiënter kostencalculatieproces voor woningen?” Om deze vraag te beantwoorden is een kennismanagementsysteem ontwikkeld, gebaseerd op de kennis uit een literatuurstudie. Deze is daarna getest met projectdata die tijdens dit onderzoek verzameld zijn.

In de literatuurstudie wordt kennismanagement gedefinieerd als een set met handvaten voor het vinden van relaties tussen informatieobjecten, personen en bedrijfsprocessen. Om kennismanagement te begrijpen moet eerst begrepen worden wat kennis is. Kennis is gedefinieerd als verfijnde informatie die een betekenis heeft. Kennis kan gescheiden worden in twee vormen, ‘impliciete’ en ‘expliciete’ kennis. Het grootste verschil tussen de twee vormen van kennis is dat ‘expliciete’ kennis gedocumenteerd is, waar bij ‘impliciete’ kennis er vanuit gegaan wordt dat deze er is. Voor kennismanagement is alleen ‘expliciete’ kennis nuttig, zeker gezien het doel is nieuwe kennis te genereren. Dit kan namelijk alleen door het combineren van ‘expliciete’ kennis.

In dit onderzoek wordt voorgesteld om BIM te gebruiken als bron voor kennismanagement. BIM is een intelligente simulatie van de architectuur dat de volgende karaktereigenschappen heeft: digitaal, ruimtelijk (3D), meetbaar, allesomvattend, toegankelijk en duurzaam (Campbell, 2006). Het gebruik van BIM resulteert in een BIM model. Een BIM model is een virtueel prototype van het ontworpen gebouw. Er zijn twee typen BIM modellen, BIM modellen met een bestandsformaat van de leverancier en BIM modellen die een open

standaard als bestandsformaat hebben, vaak het IFC bestandsformaat. In dit onderzoek wordt gebruik gemaakt van het IFC bestandsformaat.

Voor het omzetten van de data in het IFC model naar kennis om kosten te calculeren is het data mining proces toegepast. Dit proces bestaat uit de volgende stappen:

1. Probleem begrijpen en definiëren
2. Verzamelen van de benodigde data
3. Omzetten van alle BIM modellen naar IFC modellen
4. Omzetten van IFC modellen naar een dataset
5. Kosten van gebouwen toevoegen aan dataset
6. Opschonen van de dataset
7. Een bruikbare subset selecteren
8. Ontwikkelen van een bruikbare data mining algoritme
9. Gebruik maken van het algoritme
10. Evalueren van resulteren met statistische indicatoren
11. Herdefiniëren van de data en het probleem
12. Het gegenereerde model gebruiken voor kostencalculaties

Het grootste knelpunt in dit onderzoek was het geautomatiseerd omzetten van de informatie in BIM modellen naar een dataset. In de theorie wordt omschreven dat het mogelijk is om alle fysieke informatie over de gebouwen in een BIM model geautomatiseerd om te zetten naar een dataset. Echter gedurende de praktische implementatie bleek dat de huidige kwaliteit van de BIM modellen problematisch is voor het geautomatiseerd omzetten van de data in BIM modellen naar een dataset. Om dit knelpunt te op te lossen is een script geschreven dat specifiek de problemen oplost voor de modellen die beschikbaar waren tijdens dit onderzoek.

Een ander probleem kwam aan het licht tijdens het gebruik van de gegenereerde dataset in het data mining algoritme, aangezien er geen bruikbare relaties in de data gevonden konden worden. Verondersteld werd dat dit probleem ontstond doordat de dataset te klein is en dat de dataset meerdere type woningen, appartementen en huizen, bevat. Om deze veronderstelling te testen zijn extra data verzameld uit publiekelijk toegankelijke projectanalyses van bouwprojecten. Deze extra data zijn gebruikt om twee datasets te creëren. Een dataset die bestaat uit alle data van zowel de oude dataset als de nieuw verworven data en een dataset die ook bestaat uit de oude dataset en de nieuw verworven data, echter met alleen de data van appartementen.

Deze datasets zijn gebruikt in hetzelfde data mining proces. Uit de resultaten kan worden geconcludeerd dat de gemaakte aannames correct zijn. De dataset met alle data gaf zeer consistente resultaten, waar de dataset met alleen appartementen liet zien dat het gegenereerde model precies genoeg de kostencalculatie kan maken voor conceptuele kostencalculaties.

Om te testen in hoeverre het gegenereerde model bruikbaar is, is het gegenereerde model toegepast op de dataset. Hoewel het model kostencalculaties maakte die dicht genoeg bij de daadwerkelijke kosten zitten voor een conceptuele kostenraming, is het model alleen getest op data die ook gebruikt zijn om het model te genereren. Het is daarom nog onbekend hoe precies het model zal werken op nieuwe, ongebruikte data.

Om de onderzoeksvraag te beantwoorden, de literatuurstudie heeft aangetoond dat het data mining proces een effectief hulpmiddel is voor het creëren van een kennismanagement systeem en dat de BIM data veel potentie hebben als databron. Echter, tijdens de praktische implementatie werd duidelijk dat er problemen zijn met zowel de hoeveelheid beschikbare BIM modellen als met de kwaliteit van de beschikbare BIM modellen. Door extra data te verzamelen was het toch mogelijk om het data mining proces succesvol toe te passen en bruikbare resultaten te krijgen. Echter, zolang de kwantitatieve en kwalitatieve problemen met BIM modellen niet zijn opgelost kost het handmatig invoeren van de data te veel tijd om het calculatieproces efficiënter te maken.

This page is intentionally left blank

Abstract

Due to a change in what leading clients demand from the architecture, engineering and construction (AEC) industry, the AEC industry has to work more efficient. Knowledge management (KM) has a great potential to make the design process of the AEC industry more efficient. Nonetheless the traditional knowledge management systems (KMS) do not meet the requirements necessary to be effective. Therefore new types of KMS are needed. A possible solution is to combine the potential of BIM and data mining into a KMS. The main scope of this research is to develop a KMS based on BIM and data mining and test the effectiveness of this KMS.

One of the most important aspects of the data mining process is the dataset that is analysed. To create such a dataset from BIM models a tool is created to automate the extraction of data from the BIM models into a dataset useful for data mining. This has led to insights about the current state of both the quality and quantity of BIM models. After the application of the data mining algorithms on the dataset it is suggested that the amount of available BIM models has to increase for the KMS to be successful.

To test this statement, additional data has been collected from publically available data. These data have been added to the dataset and the data mining process has been applied on the newly created dataset. It could be concluded that if the quality of the BIM models meets the necessary standard, the KMS system can be successful as long as enough BIM data is available and this BIM data is about buildings of a similar type.

This page is intentionally left blank

Preface & Acknowledgment

It is with great proud that I hereby present my graduation thesis written to finish the master program Construction Management and Engineering at Eindhoven University of Technology. After finishing my bachelor in the Built Environment at the Avans University, I started this master track and have really enjoyed it. Before getting into the content of the report, I would like to express my gratitude toward the people who have contributed to making this research possible.

I want to thank Jakob Beetz and Chris Snijders for their feedback and the time they have made to support me during the period of my research. They have given me the guidance I needed to successfully finish my graduation project. I would also like to thank Ibis BV for the opportunity to do my research at the company. The support of the team has really helped me during my research. Especially I would like to thank Jeroen van de Laar and Toine Bullens for being my external supervisors. Moreover, I would also like to thank Thomas Krijnen with his help with programming the python script. Without his help I would not have been able to create the script necessary for my research.

I hope that every reader will enjoy reading and learning from this thesis report.

Yonne Rekveld
Eindhoven, May 2017

This page is intentionally left blank

Abbreviations

AEC:	Architecture, Engineering and Construction
(A)NN:	(Artificial) Neural Network
AI:	Artificial Intelligence
BIM:	Building Information Modelling
ILS:	Informatieleveringsspecificatie (Information Delivery Manual)
CAD:	Computer Aided Design
CB-NL:	Nederlands conceptenbibliotheek voor de bouw
CBR:	Case-Based Reasoning
DBFMO:	Design, Build, Finance, Maintain, Operate
GD:	Generative Design
ICT:	Information and Communication Technology
IFC:	Industry Foundation Classes
KBE:	Knowledge-Based Engineering
KBS:	Knowledge-Based System
KDD:	Knowledge Discovery in Databases
KE:	Knowledge Engineering
KM:	Knowledge Management
KMS:	Knowledge Management System
K-NN:	K-Nearest Neighbour
SVM:	Support Vector Machine
TCO:	Total Cost of Ownership

This page is intentionally left blank

Contents

Colofon 3
Summary 5
Samenvatting 9
Abstract 13
Preface & Acknowledgment 15
Abbreviations 17
Contents 19
List of figures, tables and equations 21
List of figures 21
List of tables 21
List of equations 21
1. Introduction 25
1.1. Problem definition 27
1.1.1. Problem analysis and limitations 27
1.1.2. Problem definition 27
1.1.3. Research questions 27
1.2. Research approach 28
1.3. Practical and theoretical relevance 29
1.3.1. Theoretical relevance 29
1.3.2. Practical relevance 29
1.4. Reading Guide 29
2. Introduction literature review 33
3. Knowledge management 35
3.1. Definition of knowledge management 35
3.2. From data to knowledge 35
3.3. Tacit and explicit knowledge 36
3.4. Knowledge conversion and creation 36
3.5. Goal of knowledge management 37
3.6. Current state of AEC knowledge management 38
3.7. Conclusion knowledge management 40
4. BIM data as knowledge source 41
4.1. Big data 41
4.1.1. Big Data Engineering and Analytics 42
4.1.2. Data mining 43
4.1.3. Machine learning 44
4.1.4. Evaluation of the data mining performance 46
4.2. Building Information Modelling 47
4.2.1. What is BIM? 47
4.2.2. Industry Foundation Classes 48
4.2.3. Improvements of the shared BIM Models 49
4.2.4. BIM for cost estimation 50
4.3. Conclusions BIM as a knowledge source 51

4.4. Conclusion of literature review 52
5. System Design 57
5.1. Data collection and extraction 57
5.2. Data cleaning and Exploration 65
5.3. Data engineering 67
5.3.1. Sub dataset selection 67
5.3.2. Normalization 68
5.4. Conclusions data engineering 68
6. Data analytics 71
6.1. Algorithm engineering 71
6.2. Running the algorithm 71
6.3. Evaluation of result 73
6.4. Redefining data 73
6.5. Using the new datasets 75
6.6. Conclusion data analytics 78
7. Discussion 81
8. Conclusions 83
9. Recommendations 87
9.1. Recommendations for implementation 87
9.1.1. Implementation of the IFD format 87
9.1.2. Implementation of the Information Delivery Manual(IDM) 87
9.1.3. BIM as-built models 87
9.1.4. Use of additional data 88
9.2. Recommendation for further research 88
9.2.1. Quality of BIM models 88
9.2.2. Other goal of KMS 88
References 89
Appendix 95

List of figures, tables and equations

List of figures

Figure 1: Research model 28
Figure 2: SECI model of knowledge conversion 37
Figure 3: Supervised Learning compared to Unsupervised Learning 45
Figure 4: Proposed model 53
Figure 5: From input data to CSV 59
Figure 6: Structure of extractor script 62
Figure 7: Distinction between the preferred and actual working of the script 63
Figure 8: Model without structural floor 66
Figure 9: Correlation plot 68
Figure 10: Building elements composing of different construction products 85
Figure 11: Structure of the IFC data format 96
Figure 12: Price index residential buildings 137
Figure 13: Correlation coefficient of database A 138
Figure 14: Root mean squared error of database A 138
Figure 15: Mean absolute error of database A 139
Figure 16: Root relative squared error of database A 139
Figure 17: Correlation coefficient of database B 140
Figure 18: Root mean squared error of database B 140
Figure 19: Mean absolute error of database B 141
Figure 20: Root relative squared error of database B 141

List of tables

Table 1: Data mining process according to Fayyad et al. and John 43
Table 2: Building elements to be extracted 60
Table 3: Overview of building elements with the corresponding IfcElement 61
Table 4: Overview of the extracted dataset 64
Table 6: Performance of the linear regression algorithm 72
Table 7: Performance of the k-nearest neighbour algorithm 72
Table 8: Additional data from databases 74
Table 9: Performance of the linear algorithm on the dataset A 75
Table 10: Performance of the linear algorithm on dataset B 76
Table 11: Predicted prices of dataset A 77
Table 12: Predicted prices of dataset B 77

List of equations

Equation 1: Correlation coefficient 46
Equation 2: Root mean squared error 47
Equation 3: Mean absolute error 47
Equation 4: Root relative squared error 47
Equation 5: Euclidean distance 68
Equation 6: Normalization 68

Equation 7: Price normalization ……72

Part I: Problem statement

Part II: Theoretical review

Part III: Practical implementation

Part IV: Conclusions and recommendations

This page is intentionally left blank

1. Introduction

Leading clients in the architecture engineering and construction (AEC) industry are increasingly demanding higher quality products at a low cost. They demand a high reliability in quality and delivery on the required date (Brown & Adams, 2000; Stewart & Spencer, 2006). To achieve this change in demand, the construction time of the building projects has to be reduced while the quality has to be improved. As a result of this, companies within the construction process are forced to improve their efficiency as it is the ratio between input and output.

Because of the unique nature of every construction project, the construction process is difficult to standardize. Every project needs different solutions (Van der Aalst et al., 2002). Although this might be true, only 20% of all engineering activities done by experts within the building project are tasks with no pre-existing solution (Koskela, 2000). One of the reasons why this percentage is so low is because the knowledge gained within the organization is not shared. This leads to the re-engineering of a design solution while the solution already exists within the organization. Calkins, Egging and Scholz (2000) stated that when the architecture, engineering and construction (AEC) industry would achieve to systematically managing the acquired knowledge, they could use this knowledge for 80% of all the design tasks, which would lead to a reduction of process time.

The reuse of knowledge would improve the efficiency of the construction process. But in order to be able to reuse this knowledge, this knowledge has to be captured and stored. One of the solutions to achieve this is by using an effective Knowledge Management System (KMS) (Konukcu & Koseoglu, 2012).

Many companies around the world recognize the advantage of Knowledge Management (KM). KM will not only lead to a more efficient process, it is also crucial for companies to sustain their competitive advantage (Alavi & Leidner, 1999). This is the case in the upcoming years, when a large percentage of the skilled workforce will retire from work, which will lead to a lot of lost knowledge. Researchers at the Sloan Center on Aging & Work at Boston College (Sweet & Pitt-Catsouphes, 2010) examined the effect of the aging workforce within the construction industry. In this research they concluded that one of the major concerns is how to transfer the existing knowledge to lesser experienced workforce. Given that KM provides a way to transfer this knowledge to sustain competitive advantage and leads to more efficient processes, it is hard to see why KM is rarely implemented adequately within construction firms.

The reason behind this is that traditional KM techniques used in the AEC industry have some shortcomings (Deshpande, Azhar & Amireddy, 2014). According to a research of Wong (2006) one of the most critical success factors of KM is that the chosen KMS is a very dynamic system that can easily capture and transfer knowledge and requires little maintenance. Most techniques used in the AEC industry, like standards and best practise guides and post practice reviews, require paper based documentation. This requires a lot of KM maintenance and it only achieves that knowledge is captured, not shared within the organisation (Dave & Koskela, 2009).

As stated above, KM is only effective when the system demands little maintenance effort and is easy to access, both to capture and to store knowledge. This is why data from Building Information Modelling (BIM) have a great potential for KM because “BIM is the process of generating and managing data during the building life cycle” (Jan, Ho, & Tserng, 2013) and it

is already implemented in lots of building projects (Ho, Tserng & Jan, 2013). As the result is that BIM data are a unique source for KM as they provide a way to capture and store the information created during a project.

As stated by Jan, Ho and Tserng (2013), the development in BIM provides a way to share information about projects. When BIM is applied adequately within a project, the project uses a central database where all information about the project generated by the different stakeholders within the process is stored. As this database is accessible to every project member with an internet connection, the data within the database can also be easily updated and transferred to other stakeholders.

Although BIM allows stakeholders to share and store information about projects, it stores no knowledge. However data can be transformed into information and information into knowledge according to the widely accepted theory of knowledge hierarchy (Kebede, 2010). This knowledge can be used to create new knowledge when it is made explicit and combined with a other form of explicit knowledge (Nonaka, 1994). This means that if one is able to convert the information stored in the BIM database, in the form of BIM models, to explicit knowledge it can be used to gain new knowledge, by finding relationships between the different forms of knowledge that are made explicit.

Hoogeveen (2015) recognised this potential of converting BIM data into knowledge. In his research Hoogeveen used the data mining process, to create new knowledge in the form of a model to estimate the costs of building projects. The data mining process is a process in which information, in the form of data which are located in databases, is converted into knowledge. This is also called knowledge discovery in databases (KDD) (John, 1997).

Due to limitations in his research, Hoogeveen (2015) did not actually extract the information used in the process from BIM data. He used the data in the form of PDF files and extracted these data manually. This is problematic because for a KMS, in order to be effective, the system has to require little maintenance. So for a KMS to be effective, the information located within the BIM data needs to be extracted into knowledge automatically. Another limitation is the kind of building projects used in his research. In his research Hoogeveen used information about datacentres and converted this into knowledge in the form of a model that can predict the total costs of ownership for datacentres. As datacentres have relatively simple designs and the designs do not significantly differ from each other, they are pretty uniform. This makes datacentres uniquely qualified for cost prediction based on data mining (Hoogeveen, 2015), but unlikely to deliver general knowledge that allows usage outside the domain of datacentres. Given both these limitations, the presented study will focus on the creation of a KMS that automatically extracts information from BIM models and will use this KMS for cost prediction of residential buildings.

In Summary, due to a change in what leading clients demand from the AEC industry, the AEC industry has to work more efficient. A possible solution is to apply KM within the organisation. However, the problem with traditional KM techniques used in the AEC industry is that they demand too much maintenance and they only store but do not share information within the organisation. With the development of BIM, a new possibility to create a KMS has emerged. BIM provides a unique source of information as it generates and manages the data created during the building life cycle. This unique source of information can be converted into knowledge by means of data mining. Hoogeveen (2015) proved part of this concept by using a data mining approach to predict the costs of datacentres. This study will focus on the

unproven part, to extract the information in BIM models into knowledge automatically. It will also explore whether or not this concept is also applicable to predict the costs of buildings with a less general design as datacentres.

1.1. Problem definition

1.1.1. Problem analysis and limitations

The problem this research will focus on, is that the current design process of the AEC industry is inefficient. One of the reasons is that the knowledge management systems (KMS) within the AEC industry do not meet the requirement necessary to be effective. Theoretically a combination of BIM and the data mining process can provide a KMS that is effective and make the design process more efficient. The design process is very broad and has multiple tasks that can be made more efficient with KM. As this research is based on the limitations of the research done by Hoogeveen (2015), the KMS will focus on the same tasks, namely cost prediction in the early stage of the design process.

Although a combination of BIM and the data mining process can theoretically provide a KMS, the data mining process has to be able to recognise patterns within the data. This concept has only be proven with a very standardised type of building, specifically datacentres. To test if this concept can also be used for less standardised building types, another building type should be used. For this research BIM data of residential buildings were available. Therefore this KMS will be focussed on residential buildings.

The objective of this research is:

"To realize a more efficient process of cost estimation for residential buildings".

1.1.2. Problem definition

Based on the research object the problem definition is defined:

"Can the use of a knowledge management system using BIM data and the data mining process create a more efficient process of cost estimation for residential buildings?"

1.1.3. Research questions

In order to achieve the research objective of this research and to answer its problem definition the following research questions are specified.

Research question 1:

- Which are the steps the data mining process consists of?

As the goal of this research is to use the data mining process as a KMS it is useful to know what the steps are within the data mining process.

Research question 2:

- Does the IFC data format contain all the necessary data to extract a dataset for the data mining process?

The IFC data format will be used as the BIM standard for this research. This was chosen because this makes it possible to use BIM models created with all BIM authoring tools that are generally used in the AEC industry as they all support the model to be exported to the IFC data format. Also the IFC data format is becoming the Dutch BIM standard as both

companies are starting to mutually agree on using the IFC data format is the only format to share BIM models (BIM basic information delivery manual, 2016) and the government is starting to demand the IFC data format for integrated contracts such as Design, Build, Finance, Maintain and Operate (DBFMO) contracts (Central Government Real Estate Agency, 2013). Although this makes the IFC data format the logical format to use it still needs to be determined if all required information can be extracted from the IFC data format.

Research question 3:

- Is the current BIM models modelled with a high enough quality to automatically extract a dataset for the data mining process?

The effectiveness of the data mining process depends on the quality of the dataset. This dataset is extracted from the BIM models and therefore if the quality of the BIM models is not high enough, the data mining process will not work.

Research question 4:

- Can the data mining process be used to predict the cost of residential buildings in the early stages of the design process?

1.2. Research approach

This research consist of both a theoretical and practical part. In the theoretical part a knowledge management system will be suggested to create a cost estimation model for residential buildings, based on the literature. The use of this system will be tested in the practical implementation. After both parts are carried out it will be concluded if the proposed system is effective.

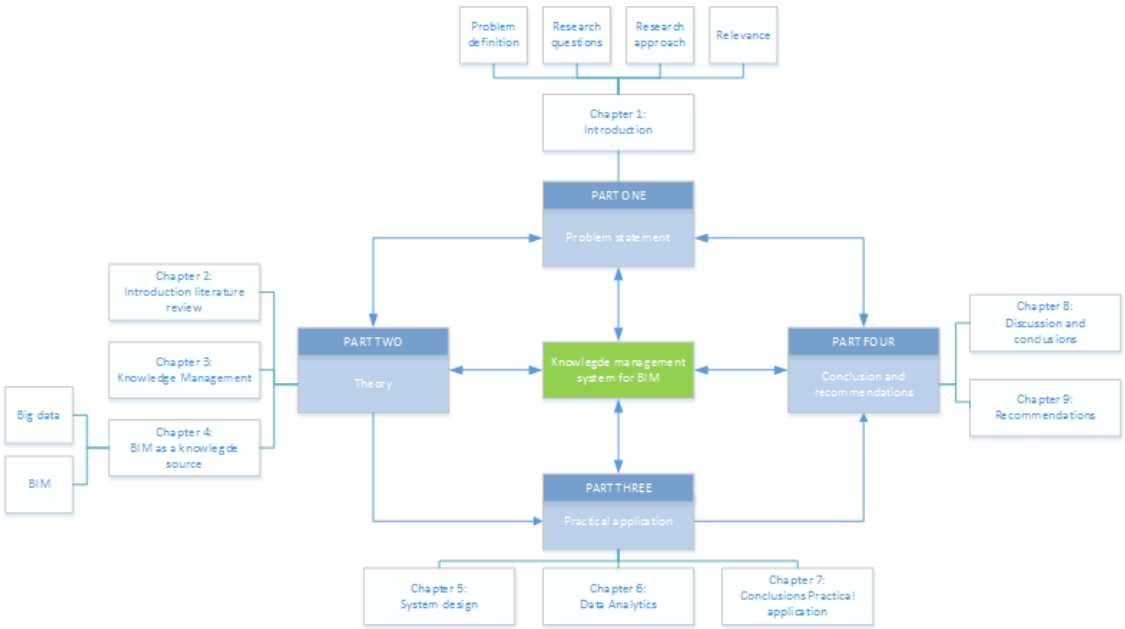

Figure 1: Research model

1.3. Practical and theoretical relevance

1.3.1. Theoretical relevance

The theoretical relevance of this research is to further explore the current possibilities and limitations of knowledge management using data mining and BIM data. This research continues the research of Hoogeveen (2015) in which BIM data and data mining were used to predict the costs of a building project. It aims to give knowledge about the possibilities outside the domain of datacentres. It also explores the potential of the current BIM models to extract data automatically by using an extraction tool for cost prediction and the potential improvement that should be made in the quality of the BIM models before it can be successful.

1.3.2. Practical relevance

The practical relevance of this research is to explore the possibility of an alternative way to create a cost estimation in the early phase of the design process. If the KMS can be automated the labour time for cost estimation will be greatly reduce. This leaves more time to do tasks that add value for the client.

For a data mining process to work effectively, the input data have to be of a high enough quality. In this research the quality of the BIM models is addressed. It will provide insight in the current quality of the BIM models and in how this quality can be improved to be useful in a data mining process.

1.4. Reading Guide

This research consists of both a theoretical approach and a practical implementation. The report starts with an introduction of the problem and the derived research questions. Thereafter a literature review is given, which is followed by a practical implementation with system development and application. The report ends with the conclusions and recommendations.

Chapter 1 elaborates the motivation of the research. The research objective, research problem and research questions are given. Also the research approach, theoretical and practical relevance and the reading guide are given.

Chapter 2 gives an introduction to the literature review. It explains in more detail how the literature research is structured.

Chapter 3 provides an overview in knowledge management. It elaborates what knowledge management is, and what the current state of knowledge management is within the AEC industry.

Chapter 4 gives more background information about the combination of big data and BIM. It explains how big data and data mining are related to each other and how the data mining process is structured. It also provides a review of BIM and explains the potential of BIM as a source for data mining. In addition it also explores the Industry Foundation Classes.

Chapter 5 describes the implementation of the data engineering part of the data mining process. It explains in more detail how data engineering should be executed. During this chapter also an overview of the practical implementation problems identified during data engineering is given.

Chapter 6 describes the implementation of the data analytic part of the data mining process. It explains in more detail how data analytics should be executed. For this chapter also an overview of the problems with the practical implementation is given.

Chapter 7 discusses a critical evaluation of the presented research.

Chapter 8 provides the conclusions of this research. In this chapter the research questions and problem statement are answered.

Chapter 9 elaborates the recommendations for implementations and the possibilities for further research in combination of data mining, BIM data and knowledge management.

Part I: Problem statement

Part II: Theoretical review

Part III: Practical implementation

Part IV: Conclusions and recommendations

This page is intentionally left blank

2. Introduction literature review

In the problem statement part of this research the research questions were defined. Some of these research questions can be answered by the literature review. In this part of the report the research question *Which are the steps the data mining process consists of?* will be answered.

As this research is a continuation on the research done by Hoogeveen (2015) most of the literature used in his research can be used for the literature review of this research. Especially on the topic of knowledge management much of the literature review is based on Hoogeveen. This literature has been verified by finding additional literature and by critically reviewing the used literature. For other topics the literature of Hoogeveen (2015) is also useful, however as this research has a different focus more research on these topics was necessary.

The literature review will start with a chapter about knowledge management. This chapter will explain what knowledge is and how it can be managed. Thereafter it will be explained how BIM data can be used as a knowledge source when using the data mining process.

This page is intentionally left blank

3. Knowledge management

The change from an industrial society to a knowledge based society has let knowledge to be the most important resource of a company. With this movement it has become more important to share knowledge within a company (Johannessen, Olaisen & Olsen, 2001). The reuse of existing organizational knowledge attained by previous experiences can reduce time spent on problem solving and increase the quality of work and therefore results in competitive advantage. This makes managing the knowledge within the organization very important. It is especially important for AEC-based organizations, because of the high amount of engineering tasks, as these are highly knowledge and experience driven (Deveraja, 2015). As knowledge is the most important asset of companies, companies must leverage their knowledge in order to maintain their sustainable competitive advantage.

3.1. Definition of knowledge management

Knowledge management (KM) is difficult to articulate and quantify because it contains elements of disciplines of both “hard” and “soft” sciences (Abaljaber et al., 1998). There is no consensus on what KM is. As the research of MIT (Abaljaber et al., 1998) shows, different articles give different solutions of KM. In the same research a small alteration of the definitions of Frappaolo and Toms (1997) is proposed: "KM is a tool set for the automation of deductive or inherent relationships between information objects, corporate users and business processes".

As knowledge management is about knowledge it is important that there is an understanding of what knowledge is and how it can be obtained. In the following section it will be made clear where knowledge comes from, how it can be defined and which types of knowledge exist.

3.2. From data to knowledge

An understanding of the concept of knowledge and knowledge taxonomies is important because theoretical developments in the knowledge management area are influenced by the distinction among the different types of knowledge (Alavi & Leidnet, 2001). The concepts of data, information and knowledge are closely related (Kock et al. 1997), and it is commonly known that knowledge has a higher level than information, and information has a higher level than data (Tuomi 1999). Data are symbols that represent the properties of objects and events without any added interpretation or analysis. They simply exist and have no significance beyond its existence (in and of itself). They can exist in any form, usable or not. They do not have meaning of their selves (Ackoff, 1989; Ackoff, 1999). Davenport and Prusak (2000) indicated that “data are a set of discrete, objective facts about events”, and “provide no judgment or interpretation and no sustainable basis of action”. Data are syntactic entities and patterns without meaning, and exist in usable or non-usable forms without any significance beyond the existence (Aamodt and Nygård 1995; Bellinger et al. 2004). Data have no meaningful relation to anything else, since they are missing a context (Uriarte, 2008).

Information is data that have been given meaning by way of relational connection. This meaning can be useful, but it does not have to be so. Aamond and Nygård (1995) explain information as structured data with meanings, which is generated from the interpretation process of data. Ackoff (1990) defined information as “data that are processed to be useful, providing answers to ‘who’, ‘what’, ‘where’, and ‘when’ questions”.

Knowledge is refined information. Knowledge is the appropriate collection of information, in such way that its intention is to be useful. Knowledge is a deterministic process. When someone "memorizes" information, then they have amassed knowledge. This knowledge has a useful meaning to them, but it does not provide for, in and of itself, an integration such as when it would infer further knowledge (Ackoff, 1990; Aamond and Nygård, 1995).

In summary, data are a carrier and storage of information and knowledge, and a media for information exchange and knowledge transfer (Kock et al. 1997). Kock et al. (1997) indicated that information is descriptive and related to the past and the present, while knowledge can be used to predict the future within a certain limit. The role of knowledge is to facilitate the processes of transforming data into information through data interpretation, deriving new information from existing through elaboration, and acquiring new knowledge through learning (Aamodt and Nygård 1995).

3.3. Tacit and explicit knowledge

Nonaka (1994) has identified two dimensions of knowledge in organizations: tacit and explicit. According to Nonaka, the tacit knowledge is rooted in action, experience, and involvement in a specific context. Tacit knowledge is comprised of both cognitive and technical elements (Nonaka, 1994). The cognitive element refers to an individual's mental model consisting of mental maps, beliefs, paradigms and viewpoints. The technical component consists of concrete know-how, crafts and skills that apply to a specific context. Tacit knowledge consists of professional expertise, individual insights and experience, and creative solutions (Pozzali & Viale, 2015).

Junnarkar and Brown (1998) suggest that "tacit knowledge is that which is implied but not actually documented". In other words knowledge can be tacit not because one is unable to articulate it, but because it has not been documented yet. This perspective is very useful because it suggests that some tacit knowledge may be more valuable when made explicit than other. The goal of knowledge management would not be to explicate all tacit knowledge but to assess first the existing tacit knowledge and determine that which has the most value before trying to make it explicit.

The explicit knowledge is articulated, codified and communicated in symbolic form and/or natural language. Most explicit knowledge exists in forms of technical or academic documents, like manuals, mathematical expressions, copyright and patents. This ''know-what'' or systematic knowledge is readily communicated and shared through printed documents, electronic methods and other formal means. Explicit knowledge is technical and requires a level of academic knowledge or understanding that is gained through formal education, or structured study. Explicit knowledge is carefully codified, stored in a hierarchy of databases and is accessed with high quality, reliable, fast information retrieval systems. Once codified, explicit knowledge assets can be reused to solve many similar types of problems or connect people with valuable, reusable knowledge (Smith, 2001). Sharing processes often require major monetary investments in the infrastructure needed to support and fund information technology (Hansen et al., 1999).

3.4. Knowledge conversion and creation

Next to the tacit-explicit distinction of knowledge another distinction between dimensions of knowledge was identified by Nonaka (1994). These dimensions, individual and collective (or

social) knowledge, in combination with the tacit-explicit dimension can be used to distinguish different kinds of knowledge conversion and creation.

Nonaka (1991) dimensioned four types of knowledge conversion on the SECI model. These four types are socialization, externalization, combination and internalization. 'Socialization' is converting individual tacit knowledge to group tacit knowledge. In the type 'externalization' tacit knowledge is made explicit. The type 'combination' is conversing separate explicit knowledge to systematic explicit knowledge and 'internalization' is the conversion from explicit knowledge to tacit knowledge.

	To tactic knowledge	to explicit knowledge
From tactic knowledge	Socialization	Externalization
From explicit knowledge	Internalization	Combination

Figure 2: SECI model of knowledge conversion (Nonaka, 1991)

As this research has as goal to convert information saved in BIM data, which is explicit, to new, well documented knowledge in the form of a tool to predict the costs of a building, this research will focus on the type 'combination'.

3.5. Goal of knowledge management

In the previous four sections, knowledge and knowledge management have been discussed. It has been made clear what knowledge and knowledge management are and why knowledge management is important. However what the actual goal of knowledge management is, has not been discussed yet. The goal of knowledge management (KM) is to connect knowledge providers and knowledge seekers to provide value creation and create sustainable competitive advantage (Abaljaber et al., 1998; Alavi & Leidner, 2001). Sustainable competitive advantage can be achieved through resources that are valuable, rare and imperfectly imitable. This resources can be property-based or knowledge based. Property based resources are legally controlled by a specific firm where knowledge based assets are protected because they are often subtle or difficult to understand or copied by outside observers.

In a study by Davenport, De Long and Beers (1997) four business objectives that fulfil the goal of KM are identified, namely:

1. To capture knowledge;
2. To improve knowledge access;
3. To enhance the knowledge environment;
4. To manage knowledge as an asset.

Capturing knowledge can be achieved by creating KM repositories. These consist of structured documents with knowledge embedded in them, stored in a way that they may be easily retrieved. Better access to knowledge can be facilitated by improving the processes of knowledge transfer between individuals and between organizations. By proactively facilitating and rewarding knowledge creation, transfer and use of an enhanced knowledge environment can be achieved. Knowledge should also be managed as an asset, and can be achieved in

multiple ways. For example some companies include their intellectual capital in the balance sheet, where other leverage their knowledge assets to generate new income or reduce costs.

3.6. Current state of AEC knowledge management

In the section below a review of the current state of knowledge management within the AEC industry is given. Much is based on the review given by Hoogeveen (2015). Also additional literature was found, both to verify the work done by Hoogeveen and to expand the review.

Knowledge is one of the most important assets of an AEC firm (Deshpande, Azhar, & Amireddy, 2014; Dave & Koskela, 2009; Woo et. Al, 2004) and since most of the firms knowledge is created during the lifecycle of a project, managing this generated knowledge is important. AEC firms mostly used to rely on individual tacit knowledge and experiences gained through previous projects. Since the last decade, more attention has been given toward KM (Al-Gahtani & Ghani, 2010; Weippert & Kajewski, 2007; Kanapeckien et al., 2010).

AEC firms have both explicit and tacit knowledge to manage (Woo, Clayton, Johnson, Flores, & Ellis, 2004). In the past, most research in the AEC industry has been focused on explicit knowledge. Explicit knowledge examples in the AEC are specifications, reports, procedure manuals, drawings and documents management systems (Al-Gahtani & Ghani, 2010).

Around 2004, a shift in attention from managing explicit knowledge to managing tacit knowledge can be found (Kamsu-Foguem & Abanda, 2015). The management of tacit knowledge requires other techniques, such as lessons learned to record personal tacit experiences, post project reviews, databases of best practices, web-based communities, decision support systems, collaborative internets, and Communities of Practice (CoP) (Deshpande, Azhar, & Amireddy, 2014; Tibaut & Jakosa, 2015; Woo, Clayton, Johnson, Flores, & Ellis, 2004; Al- Gahtani & Ghani, 2010; Kanapeckiene, Kaklauskas, Zavadskas, & Seniut, 2010).

Techniques like post project reviews were introduced to convert tacit knowledge to explicit knowledge. However these techniques are very time consuming and therefore unlikely to be effective. Even when the reviews are done properly, these documents are often stored in archives, serving no purpose and are not easily and universally accessible (Dave & Koskela, 2009). Recently, most of the scientific research has focussed on managing tactic knowledge and mostly human interaction (Al-Gahtani and Ghani, 2010). This movement in research seems to be directed towards managing tacit knowledge for processes and experiences (Woo, Clayton, Johnson, Flores, & Ellis, 2004; Dave & Koskela, 2009).

Woo et al (2004) focused on developing a web based knowledge navigator. It allows experts to search for knowledge and facilitates communication with other experts. Hence it enables socialisation. Dave and Koskela (2009) stated that the general consensus in the literature is that the AEC industry fails to retain project knowledge for future reuse. They also develop a system to enable social interaction between experts to capture tacit knowledge in the form of experiences, and make it explicit in the form of best practices with a database.

Lately, different tools of KM are integrated with BIM. Many of these tools are used to graphically represent the captured knowledge. Ho, Tserng and Jan (2013) developed a BIM-based knowledge model where experiences are graphic representations via objects and aspects of BIM models. Liu et al (2013) developed a similar model, where tacit experts'

knowledge was made explicit by using lessons learned techniques. These were linked to a BIM model, to improve the sharing of knowledge between engineers.

Lin (2014) created a 3D Construction BIM-based Knowledge Management system, called the CBIMKM system. This BIM-based KM system focused the representation of knowledge for all participants in a project. The CBIMKM system is similar to traditional expert 'yellow pages' and document management module, where stakeholders can search on a topic, via a search engine, to obtain the knowledge in the form of a BIM object with an explanation provided by the design expert.

Grover and Froese (2016) assessed the usefulness of a BIM-based social platform to manage tacit knowledge. Such application, which in generally is called SocioBIM, facilitates the sharing of collective skills, expertise, understanding and knowledge throughout the life cycle of the project. Again, the BIM application is to graphically represent the captured knowledge.

Ding, Zhong, Wu and Luo (2016) suggested a combination of BIM, ontology and semantic web technology to manage the knowledge of construction risks. Their suggested method facilitates the knowledge reuse during the construction risk analysis process. The risk database is created via documents and the BIM model serves as a graphical representation. So again the BIM part of the KM tool is to visualise the knowledge. However it uses the information within the BIM model to link automatically the existing database and the object within the building design to analyse the risks involved in the building process.

In the last couple of years, the implementation of big data in tools for KM has emerged. Bilal et al. (2016) presented a review of the current state of big data within the AEC industry. In this review a couple of (possible) KM tools were presented. Generative design (GD) is one if the possible application given. GD aims to automatically generate a design based on specified design objects. However this idea is still very conceptual. Another possible application is the generating knowledge using clash detection within BIM to provide guidelines for early design review. Again this is still very conceptual and much development within the field is necessary.

A more feasible way for implementation in short term is to predict the performance of building. As this predictions have a high dimensional dataset, big data application can provide solutions to analyse the performances (Bilal et al., 2016; Mathew, Dunn, Sohn et al., 2015). In the next chapter it will become clear why the application of big data is suitable for this application.

Another feasible implementation is to analyse the social networking with big data. As mentioned above, tools have been created with BIM to provide an application platform for experts within an AEC project. This communication can be analysed with big data for automatically classification of best practises mentioned in the social networking tool (Bidal et al., 2016).

Bidal et al. (2016) also mentioned possibilities for the future. Most of them were combining BIM model with big data to create new knowledge. Cost prediction with big data however was not mentioned in the review of Bidal et al. (2016). But in a graduation thesis Botter (2013) created a model to calculate project costs based on historical project data. Botter reviewed multiple different techniques to calculate the project costs. Although the term big data is never mentioned within the research report the used techniques to analyse the data are similar to the big data techniques. As a conclusion the usefulness of historical project data to

calculate the project costs was stated. Another conclusion was that BIM should be integrated in the process to make it less time consuming.

These findings on the trend of KM focus in the AEC can be confirmed by Rezgui et al. (2010). The authors did research regarding the evolution of KM in the AEC sector. They stated that the evolution of KM in the AEC could be divided in three generations. The first generation focused on capturing explicit knowledge via document-based knowledge (e.g. CAD drawings). The second generation was directed on the codification of knowledge in the form of BIM and associated data. In the third generation, the use of BIM was vacated and the focus was converted to capturing tacit knowledge, this resulted in systems to encourage the sharing of tacit knowledge. Walker (2016) also did a research on the evolution of KM in the AEC sector and concluded that with the emerge of big data, the new competitive advantage can be created through big data. This new possibility has led to the development of tools that use big data in KM.

Summarizing, the development of KM within the AEC industry can be defined in four generations, the three generations as defined by Rezqui et al. (2010) and the new generations defined by Walker. The first generation aimed on capturing explicit knowledge, the second generation used the emerging of BIM to capture both tacit and explicit knowledge. In the third generation the focus was on capturing tacit knowledge. With the new possibilities of big data new possibilities for KM were discovered. A new generation is emerging, which offers many possibilities to generate new knowledge. Especially the combination of BIM models and big data has a lot of potential for KM. However to date, little practical application has been created.

3.7. Conclusion knowledge management

In this chapter it has been shown that KM is very important. KM can be defined as a tool set for the automation of deductive or inherent relationships between information objects, corporate users and business processes. To understand knowledge management, at first it has to be understood what knowledge is. Knowledge is defined as refined information, which means that it has meaning. Knowledge can be divided in two forms, tacit or explicit knowledge. The main difference between tacit and explicit knowledge is that explicit knowledge is documented, where tacit knowledge is implied. However, for knowledge to be useful for KM it needs to be made explicit. Especially because the goal of KM is to create new knowledge. This can only be done by combining multiple forms of explicit knowledge. In the AEC industry great attention has been paid on KM. Rezqui et al. (2010) defined three generations of knowledge management. In this research a fourth generation is suggested, KM with big data. Already multiple researches have been conducted to explore this subject. However few of those researches resulted in practical implementation.

4. BIM data as knowledge source

In the last decade, Building Information Models (BIM) have been successfully used to enhance the performance of AEC projects. BIM is a "shared knowledge resource for information about a facility forming a reliable basis for decisions during its life-cycle; defined as existing from earliest conception to demolition" (NBIMS, nd). It has improved communication of the design between various stakeholders by enabling the identification of clashes ahead of time, by enabling the simulation of the construction sequence and improving the communication between various craft subcontractors and the general contractor (Deshpande, Azhar & Amireddy, 2014). Building information models are inherently parametric, data-rich, object based representations of the facility being designed and constructed. Alternately, building information models can be conceptualized as centralized, interconnected data stores which can contain design and some construction information about the various disciplines within a construction project. This centralized and integrated nature of the design information can potentially provide a very context rich platform for the capture, storage and dissemination of the knowledge generated during the design and construction processes. One of the requirements of an effective knowledge management system is its ability in communicating and preserving knowledge effectively across various stages of a construction project (Dave and Koskela, 2009). Because BIM models can be used over the whole span of the construction project and even evolve and are able to capture the knowledge as soon as the knowledge is created, BIM models are uniquely qualified as a knowledge source (Deshpande, Azhar & Amireddy, 2014).

Although BIM models are great knowledge sources, the knowledge within these models is not explicit. Therefore BIM models can be seen as sources of embedded knowledge. Embedded knowledge is knowledge locked in processes, products or artefacts (Argote & Ingram, 2000). Even though embedded knowledge can have an explicit form, such as BIM models, the knowledge itself is not explicit, the implications of the embedded knowledge are not immediately clear. (Gable & Blackwell, 2001).

To be able to use the embedded knowledge as a source for knowledge management, the knowledge itself has to be made explicit and usable. To do so, big data techniques will be used. What big data and the associated technique are will be explained in the sections below.

4.1. Big data

Big data is an abstract concept. At present, although the importance of big data has been generally recognized, people still have different opinions on its definition. In general, big data means the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time. There have been considerable discussions from both industry and academia on the definition of big data (Chen, Mao & Liu, 2014). One of the most cited definitions of Big Data is that of Doug Laney from the then-META (now Gartner) group. This definition consists of three V's, Volume, Velocity and Variety (Laney, 2001). Later this definition has been expanded to include a fourth V: Veracity (Ward & Barker, 2013).

- Volume: How much data?
- Velocity: How fast data are processed?
- Variety: The various types of data
- Veracity: How accurate are the data in predicting business value?

Many companies like IBM use these V's to explain big data (IBM, 2016). Although these definition is widely adopted there is also criticism. Hurwitz et al (2013) stated: "While it is convenient to simplify big data into the four V's, it can be misleading and overly simplistic. For example, you may be managing a relatively small amount of very disparate, complex data or you may be processing a huge volume of very simple data. That simple data may be all structured or all unstructured".

Ward and Barker (2013) conducted a research on the definitions used for Big Data. They concluded that there are many different definitions, which all include different factors and they created a new definition which takes all those definitions into account. "Big data is a term describing the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning" (Ward & Barker, 2013).

4.1.1. Big Data Engineering and Analytics

As the terms storage and analysis already suggest big data has two complementary aspects, Big Data Engineering (BDE) and Big Data Analytics (BDA) (Bidal et al., 2016). BDE is the domain that primarily is concerned with supporting the relevant data storage and processing activities needed for the analytics. BDE can be divided in two subdomains:

- Big Data processing
- Big Data storage

An important part of the research problem of the research is to use BIM models as input for data mining. This part of the process belongs to data processing. The literature discusses big data processing for databases in the order of terabytes and even petabytes (Dittrich & Quiané –Ruiz, 2012). These databases can be handled with data processing engines that allow parallel and distributed computation. In parallel computing many calculation or the execution of processes are carried out simultaneously. It divides the overall (large) problem in smaller ones that can be solved at the same time (Barney, 2016). In distributed computing the overall problem is also divided. However in distributed computing the calculations are carried out on multiple systems (IBM, 2017).

However the database in this research is much smaller. The problem in this research is to convert the embedded information in the BIM model to a dataset with a suitable file format that can be used in the analytics of the data. This file format has to be supported by the software package used during the data analytics. More about this in chapter 5.

Big Data Analytics relates to the tasks responsible for extracting the knowledge to drive decision-making. It mostly deals with the principles, processes and techniques to understand the Big Data (Bidal et al., 2016). In their research Bidal et al. (2016) performed a review of Big Data in de AEC industry. They divided BDA within the AEC industry in:

- *Statistics*
- *Data mining*
- *Machine learning*

Statistics is the study of collecting, analysing, and drawing conclusions from the data, with the primary focus on selecting the right tools and techniques at every data analysis stage. Right from the data collections, to efficiently analysing it, and then inferring or formulating conclusions out of it, all of these steps come under the scope of statistics (Bidal et al., 2016).

Data mining and machine learning are terms often used for the same thing, but there is an essential difference. Data mining is concerned with the process of automatic or semiautomatic exploration and analysis, of large volumes of data, to discover meaningful patterns or rules (Bidal et al., 2016). Machine learning (ML) is defined by Arthur Samuel (1957) as a "Field of study that gives computers the ability to learn without being explicitly programmed". ML uses algorithms that can learn from and make predictions on data (Kohavi & Provost, 1998). In the data mining process machine learning is one of the techniques that can be used to find the pattern or rules in data.

In this research, the goal is to extract the explicit knowledge embedded in the data to useful explicit knowledge. This is often called knowledge discovery in databases (KDD) or data mining. What data mining is and how machine learning can be implemented in this process will be explained in the following section.

4.1.2. Data mining

Data mining or knowledge discovery in databases (KDD) is the process of discovering advantageous patterns in data (John, 1997) (Fayyad, Piatetsky-Shapiro & Smyth, 1996). In the commercial field this whole process is called data mining. This in contrary with academia, where data mining is just a single step in the whole process of finding patterns (John, 1997). John proposed that academia should use the definition of the commercial field. Therefore in this research data mining is adopted as the used term.

The data mining process is “an iterative process involving several steps, beginning with the understanding and definition of a problem and ending with the analysis of result and a strategy for using the results to gain advantage (John, 1997).

Both John (1997) and Fayyad et al. (1996) describe data mining as the same process, although they split the process in different steps and use different terms. In the table below an overview of the steps according to both John and Fayyad is given.

In both researches the step that actually runs the algorithm was called data mining. However data mining is very broad. One of the sub classes of data mining is machine learning. Because this research will use machine learning as the data mining algorithm, this term is already

Table 1: Data mining process according to Fayyad et al. (1996) and John (1997)

	Fayyad et al. (1996)	**John (1997)**
Big data Engineer	**1.** Data selection **2.** Data preprocessing **3.** Data transformation	1. Data collection and extraction 2. Data cleaning and exploration 3. Data engineering
Big Data Analytics	**4.** Machine Learning	4. Algorithm engineering 5. Running the Machine Learning algorithm
	5. Interpretation/Evaluation	6. Preliminary Evaluation of Results 7. Refining the Data and Problem 8. Using the Results

implemented in the overview above. Why this research uses machine learning will become clear in the next section of this report.

4.1.3. Machine learning

The machine learning field has many techniques and algorithms. To get a better understanding about the different methods available, it is helpful to categorize them. One of the most used ways to categorize machine learning methods is supervised and unsupervised learning (Ng, 2016b).

In supervised machine learning a dataset with 'right answers' is given to the algorithm. It can be used to solve regression and classification problems. In figure 3 the supervised learning picture is a classification problem. A classification problem predicts a discrete valued output. Using the known values for the variables the algorithm can predict if the outcome of the case is likely to be for example true or false (Ng, 2016a). In unsupervised machine learning a dataset is given which is not labelled into different categories. Unsupervised learning is used to find structure in datasets. One example, which is used in figure 3, is the clustering algorithm (Ng, 2016b). The difference between supervised and unsupervised learning becomes clear with figure 3. In supervised learning the answer, a circle or a cross, is known. The algorithm will learn based on x_1 and x_2 values when the answer is a circle or a cross. When presented with a known x_1 and x_2 but an unknown answer it will be able to predict the most likely answer. In unsupervised learning the answer is not known, however based on the x_1 and x_2 it is able to create two clusters.

As the goal is to learn from datasets with known output variables, this research is focused on supervised learning. Although it is now known that a supervised learning method is used in this research, there are still a lot of methods to choose from. Therefore it is necessary to divide supervised learning even further into new categories. A logical way to divide them is in the preferred output value. Supervised learning has two main categories, regression and classification.

Regression

Regression is the method of ML that is concerned about predicting the numerical value of a target variable based on input variables. For instance, estimating the costs of the design based on design specifications (Bilal et al., 2016). Weiss and Indurkhya (1995) state that the problem of approximating the values of a continuous variable is described in the statistical literature as regression. Given samples of output (response) variable y and input (predictor) variables x, the regression task is to find a mapping y. Based on this, regression is likely to be a useful technique for this research.

The most simple form of regression is linear regression, where one variable, the explanatory variable x, is used to predict the dependable variable y. Multiple linear regressions, as the name already suggests, is a technique where more than one explanatory variable is used to predict the dependable variable and is commonly used as a regression approach. As the costs of a building are dependable on more than one variable, multiple linear regression should be considered in the system design of this research.

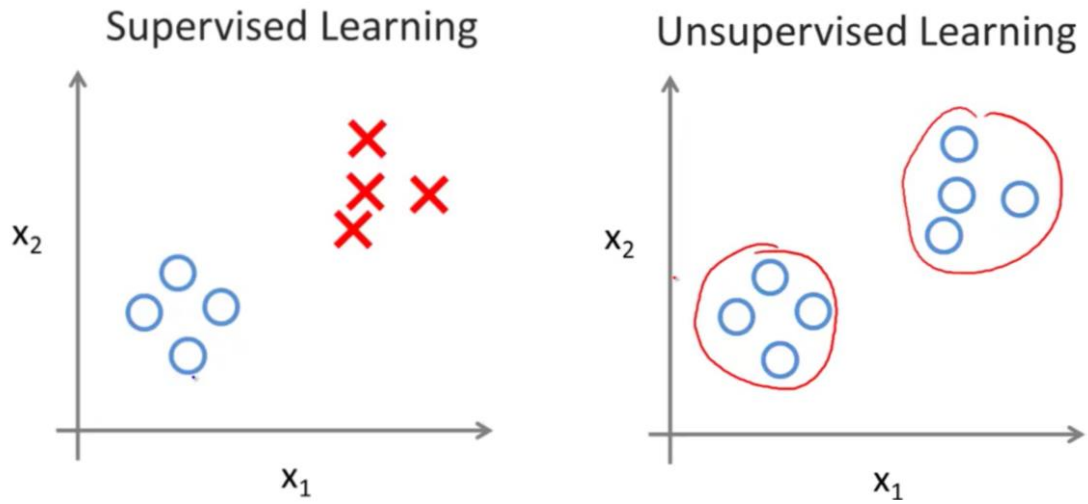

Figure 3: Supervised Learning compared to Unsupervised Learning (source: Ng, 2016b, edited)

Neural Networks are also a method that can be used as a regression technique. Although often used as a classifier, due to the relationship the logistic regression (Andew, 2011) neural network are able to solve regression problems. Berry and Linoff (2011) showed the use of neural networks to estimate the value of a house. They have been successfully used as a multivariate non-linear analytical tool, and are known to be highly effective in recognizing patterns from noisy and complex data, and estimating non-linear relationships (Khosrowshahi, 2011).

Classification

As mentioned before, besides regression there is another main category of machine learning, namely classification. Within classification there are a number of methods that can be chosen to predict costs. One of this methods is Case Based Reasoning. CBR is a data-mining technique that remembers similar situations applied to the solution of previous problems, and uses the information and knowledge from such situations to solve a new problem (Aamodt & Plaza, 1994). This technique does not require a clear model for problem-solving; rather, establishing cases is an important task in problem-solving. Since the model can be established by defining key attributes that express cases, this technique makes it easy to establish a model (Jin et al., 2012). The CBR process has four steps:

1. Retrieve: This phase inquires about and extracts the previous case most similar to the current one from a case base;
2. Reuse: This phase reuses the information and knowledge from the retrieved case for problem-solving;
3. Revise: If the retrieved case is not suitable for solving the new problem, this phase analyses the difference(s) between the new problem and the retrieved case, then revises the retrieved case accordingly;
4. Retain: This phase stores in the case base the solution proposed from the retrieved case so that it can be used in future problems

According to Vukovic et al. (2012) three techniques are usually used in the retrieval phase, namely nearest neighbour, inductive learning, and knowledge-guide. Inductive learning allows for the identification of training data or earlier knowledge patterns and similarities which are then extracted as generalized rules (Almana & Aksoy, 2014).

The most known implementation of inductive learning is the decision tree. A decision tree uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility (Quilant, 1987). Knowledge-guided approach applies existing domain and experimental knowledge to find suitable cases. Although this method is believed to be superior to the other two, knowledge-guided is difficult to apply because such knowledge often cannot be successfully captured and represented (Mi, Qian, Liu & Chang, 2008). Nearest neighbour is the most used technique, as it is easy to uses (Vukovic et al., 2012) (Sayad, 2010). It can used for classifying a new case by searching for the cases in the database that have the most similar features and assigning the class of its nearest neighbours to the new case (Vukovic et al., 2012).

K-Nearest Neighbour is a modified, extended version of Nearest Neighbour. The method gives better results because it searches for a k number of closest cases, where k is normally an odd number, and proposes the case that is most frequently found. It also allows for similarity. Similarity is the uses of weighted features. This means that the features that are the most relevant have higher weights and therefore have more influence on the predicted outcome as less relevant features (Vukovic et al., 2012).

Both classification and regression are possible for cost prediction. If classification is used the machine learning algorithm will determine the predicted cost based on the costs of the buildings most similar to the building the prediction is made for. In regression the total cost prediction is based on the sum of the input variables multiplied by their weights. More about the implementation of machine learning algorithms in the chapter 6, system analytics.

4.1.4. Evaluation of the data mining performance

As shown in table 1, one of the steps of the data mining process is to evaluate the performance of the data mining algorithm on the dataset. This performance can be tested with statistical evaluation measures. There are two types of measures. One type of measures evaluates the performance of classification and the other type evaluates the performance of numerical prediction (Witten & Frank, 2005). Although a classification algorithm is used in this research the outcome of both the regression algorithm as the classification algorithm will be a numerical prediction. Therefore the evaluation measures that can be applied on numerical prediction will be used in this research.

To evaluate the performance, the correlation coefficient, the root mean squared error (RMSE) and, mean absolute error (MAE) the root relative squared error (RRSE) are used. A correlation coefficient shows the linear dependence between the predicted value and the actual value. The result of the correlation coefficient is an index value between one and minus one, where one shows perfect positive dependence and minus one perfect negative dependence. If the value is zero there is no dependence. The most used correlation coefficient is the Pearson correlation coefficient (*r*) (Howitt & Cramer, 2005). The Pearson correlation coefficient is calculated with the equation:

$$Correlation\ coefficient = \frac{\sum(p-\bar{p})(a-\bar{a})}{\sqrt{\sum(p-\bar{p})^2}\sqrt{\sum(a-\bar{a})^2}} \quad (1)$$

$p = predicted\ value$
$\bar{p} = average\ of\ predicted\ values$
$a = actual\ value$
$\bar{a} = average\ of\ actual\ value$

The RMSE is one of the most used measures in statistics. It is used to calculate the difference between the calculated value and the actual value. An alternative to RMSE is the mean absolute error. The biggest difference between the measures is that because the RMSE is squared larger discrepancies are penalised harder. The RMSE can be calculated with the equation 2 and the MAE with equation 3.

$$Root\ mean\ squared\ error = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_n - p_n)^2} \quad (2)$$

$$Mean\ absolute\ error = \frac{1}{n}\sum_{i=1}^{n}|a_n - p_n| \quad (3)$$

The Fourth used measure to evaluate the performance is the root relative squared error (RRSE). It shows the performance of the algorithm against an more simple predictor, the average of all actual variables. If the value is below 100% the algorithm performs better than a prediction based on the average value. The closer the value comes to 0% the better the algorithm performs. The equation for the RRSE is:

$$Root\ relative\ squared\ error = \sqrt{\frac{\sum(p_n - a_n)^2}{\sum(a_n - \bar{a})^2}} \quad (4)$$

4.2. Building Information Modelling

As explained in the previous paragraphs, the BIM data will be used as a source for knowledge. Before BIM data can be used as knowledge, at first should be explained what BIM data are and where they come from. BIM data are the resulting data of using Building Information Modelling (BIM) techniques and tools in the building process.

4.2.1. What is BIM?

Building Information Modelling is a promising development within the architecture, engineering and construction (AEC) industries. With BIM technology, it is possible to create a digital and accurate virtual model of a building. When BIM is used correctly within the whole construction process the computer-generated model contains precise geometry and relevant data needed to support the construction, fabrication and procurement activities needed to realize the building (Eastman et al., 2011). But not only during the realization BIM technologies can be useful, even in the operation and maintenance phase of the building, BIM technologies can be very beneficial (Davtalab & Delgado, 2014).

A useful definition of BIM was given by Campbell (2006). BIM is an intelligent simulation of architecture that exhibits the following six key characteristics:

- Digital;
- Spatial (3D);
- Measurable (quantifiable, dimension-able, and query-able);
- Comprehensive (encapsulating and communicating design intent, building performance, constructability, and include sequential and financial aspects of means and methods);
- Accessible (to the entire AEC/ owner team through an interoperable and intuitive interface);
- Durable (usable through all phases of a facility's life).

One of the important tools of BIM Authoring Software, which is used as a tool to create a BIM model. Instead of drawing lines with a CAD program to create a 2D visualization of the building, BIM Authoring Software allows the user to create an object based 3D model. This not only allows the user to view the model in both 2D and 3D, it also allows to add other relevant data to object. For instance it is possible to add the function and material of the object (Eastman et al., 2011).

A BIM model is characterized by building components that include data that describe how they behave and are consistent and non-redundant data. Building components are objects that have digital representations and data about what they are. They can be associated with computable graphics, data attributes and parametric rules. These components also describe how they behave. This makes it possible to create analyses of the building and use it in work processes. This BIM model also contains coordinated data.

Next to the BIM model, another very important part of BIM is the interoperability between members of the building project team, to ensure that every member of the team has access to the latest project data. It is possible to allow every member to have access to all the data, these data need to be:

- Real time data exchange
- Share in a predefined format

A cloud based server is the most used way to ensure that project data is both shared in real time as it is accessible from every location.

For the predefined format there are two primary approaches, (1) use a proprietary file format and therefore stay within one software vendor's product (or use another product that is allowed by the vendor) or (2) use different vendors that can exchange data using non-proprietary file format that is an universal supported standard. The advantage of the first approach is that it allows for tighter integration among products in multiple directions. For example, a change in one model results in a change in all other linked models. The obvious disadvantage is that every member of the project team is forced to use the programs of the specific vendor. This could potentially cost a lot of money considering both licensing and training. The second approach would solve the disadvantage of the first approach. The disadvantage of this approach is that the current universal standard, IFC, is not designed to carry all relevant data.

Although the disadvantage of not being able to carry all relevant data, in this research the IFC standard will be used. The necessary data are collected via different clients of IBIS BV and therefore the acquired data will be created with software from different vendors. This makes the first approach not possible. Although this will result in some loss of data, it also has an advantage. As the tool created for the research uses a universal standard, the tool is reusable in other researches. More about the universal standard IFC in the section below.

4.2.2. Industry Foundation Classes

The Industry Foundation Classes (IFC) are the open and neutral data format for openBIM. (BuildingSMART, nd). It is developed and maintained by buildingSMART International. The IFC specification is a neutral data format to describe, exchange and share information typically

used within the building and facility management industry sector. It provides a data model structure for sharing data across various applications used in the building.

As a BIM model is characterized by building components, so is an IFC model. In an IFC model these building components are called IfcObjects. “An IfcObject is the generalization of any semantically treated thing or process” (buildingSmart, nd). These objects are divided in six subtypes, see figure 4.

Although the other five subtypes can contain a lot of information, for this research only the IfcProduct subtype is interesting because everything related to geometric or spatial context is an IfcProduct (buildingsmart, nd). An IfcProduct is also divided in subtypes (see also figure 4). As an IfcElement is a “generalization of all components that make up an AEC product” (buildingsmart, n.d.), only this subtype is relevant for this research. From the subtypes that holds an IfcElement, the IfcBuildingElement types are the types we are interested in as they are the “major functional part of a building, examples are floor, roof and wall” (buildingSmart, nd).

4.2.3. Improvements of the shared BIM Models

In the exchange of information between sender and receiver three factors are important to make sure the receiver is able to interpret the information received as the sender intents (Bell & Bjørkhaug, 2006). The first part is the format in which the information is exchanged. As already discussed, in the presented research this will be the IFC data format. The other two parts are a specification of which information to exchange and when it needs to be exchanged and a mutual understanding of what the exchanged information actually is.

The IFC data format allows various data to be shared in various ways. If the receiver of the information, in this research the extractor, wants to be sure it is able to utilize the information he receives, the sender and receiver need to be in agreement which and how the information is shared (Bell & Bjørkhaug, 2006). To ensure the agreement of how the information is shared buildingSMART created the International Framework of Dictionaries (IFD). Where the IFC format is very broad to support all the different BIM authoring tools, IFD format “specifies a language-independent information model which can be used for the development of dictionaries used to store or provide information about construction works. It enables classification systems, information models, object models and process models to be referenced from within a common framework” (International Organization for Standardization [ISO], 2007). It is a language to define ontologies, “which defines concepts and semantic relations between these concepts” (Böhms et al., 2013). The Nederlandse Conceptenbibliotheek voor de bouw (CB-NL) is the Dutch implementation of IFD. The ontologies created can be used to standardise the way information within IFC is shared. For the presented research the standardised way of sharing information is useful as the effectiveness of the extraction tool created in the practical application part of this research is highly influenced by the script ability to interpret the information within the IFC data.

Another way to ensure better quality of BIM models is to have an understanding of which information to share. The Information Delivery Manuals (IDM) can be used for the documentation of existing or new processed and describes the associated information that needs to be exchanged between parties. The main purpose of IDM is to make sure the relevant data are communicated via a structured way such as the data can be interpreted by the software on the receiving side (buildingsmart, nd.). For the Dutch AEC industry the ‘BIM basis

informatieleveringsspecificatie (BIM Basis ILS)' is defined and it has already been adopted by many companies. As the 'BIM basis ILS' has been introduced in 2016 and it started with just 14 organisations within the AEC industry it is highly unlikely that the used BIM models within the presented research are in compliance with the BIM basis ILS. In the practical implementation part of this research it will be discussed what the effect would have been if the BIM Basis ILS would have been used on all IFC.

4.2.4. BIM for cost estimation

In the previous years, a lot of research has been done in the combination of BIM and cost estimation. As a BIM model is a virtual model of the building, it is a unique source of information in the building process. If modelled correctly, a BIM model comprehends all the building elements and the relations of these building elements to each other. Most research has aimed to use BIM models as a source for quantity take off (QTO) and the use of this QTO to determine the cost estimate based on unit rates. This research also tend to focus on BIM cost estimate in the difference phases of the building process. Both these research topics could provide useful information for this research.

Part of the research on the topic of BIM for cost estimation focusses on the use of BIM tools to assist in the process of QTO. The estimator uses a virtual model of the building to estimate the QTO manually and the BIM tools only allow for a more efficient process (Shen & Issa, 2010). Examples of commercial application of this research are the software packages used in the Netherlands, like IBIS-TRAD, Vico Office and BIM ncalc (Van de Laar, personal communication, September 1, 2016; Construsoft, 2017; Knaan Bouwcomputing, 2017). All these packages are focussed on improving the process of the traditional way of creating a cost estimate, by using BIM to determine the quantities faster and with more accuracy. As this process still requires a lot of human effort, it cannot provide solutions for the presented research (Shen & Issa, 2010).

There are also researches done to automate the cost estimation. Ma, Wei and Zhang (2013) proposed a model for such automatic cost estimation. The goal of this research was to have a bill of quantities in compliances with GB50500-2008, the Chinese national mandatory specification of cost estimation for tendering of building projects. The used data format of these BIM models is IFC. In their research Ma et al. (2013) divided the IFC format in three useful parts for cost estimation. One part expresses the decomposition of building elements into construction products. Another part contains three types of information of the construction products, i.e. product type, geometric information and material information. The last part expresses accessory, opening and intersection relationships between the building elements as well as the construction products. They created a prototype QTO software application and tested the prototype's performance on more than 3000 construction products. 67% of the construction products were automatically classified and extracted to a QTO. For the other 33% it was necessary to add information manually that was lacking from the IFC model. From this research can be concluded that it is possible to extract all the necessary information for a cost estimation from the IFC data format. It can also be concluded that in some cases not all necessary information is modelled and therefore information is missing in IFC models.

Research by Cheung et al. (2012) demonstrate the use of a tool to evaluate the functional, economical properties of a building as well as the performance, during the building design. The economical properties are automatically calculated using the measurements that are

extracted from the BIM model. Four levels of cost estimation are suggested, where the level should be selected according to the level of information available. The first level is calculated based on the site boundary and massing of the building. In the second level the available information is updated with the number of floors and storey heights. On this level, default specifications of building elements are presumed to be present in the model. The third level consists of defining spaces with vertical elements. For this level the available information is updated with the floor plan of the building. In the highest level, information, like building material, is provided for the building elements.

4.3. Conclusions BIM as a knowledge source

BIM has been widely adopted in the AEC industry as it has enhanced the performance of projects. As a result BIM models are available which contain a lot of embedded knowledge. However this knowledge needs to be made explicit. In this research it is suggested that the data mining process could be used to make this knowledge explicit and even create new knowledge in the form of a cost estimation tool to estimate the building costs in the early stages of the building process.

The data mining process can be divided in a data engineering and a data analysing part. In the data engineering part the available data are converted to a usable dataset for the data analysing part. The part of data analysing consists of using algorithms to analyse the dataset and to interpret the result. One of the types of algorithms that can be used is the machine learning type. Machine learning (ML) algorithms can be used to learn from and make predictions on data. There are two categories within ML. The supervised ML category uses a dataset with known output variables to make predictions on new data. Unsupervised ML is used to find structure within the dataset. This research will implement supervised ML because our goal is to make a prediction on building costs. Within supervised ML again two categories can be found, regression and classification. As both are useful to predict costs, from both categories an algorithm will be used in this research.

The data mining process needs input data in order to be able to make predictions. As BIM has been implemented within most of the AEC industry and one of the results of BIM is a BIM model with all the information about the building, these BIM models have a great potential as a knowledge source. A universal standard for sharing BIM models is the IFC format. It provides a data model structure. For this research the physical information of the building model is interesting. All these are located within the IfcBuildingElement class.

This research closely relates to the research of cost estimation with BIM. As BIM contains all the physical information of the buildings, these information can be extracted in the form of a quantity take off. Cheung et al. (2012) have demonstrated a tool that allows for cost estimation directly from the BIM model. They suggest that during the design of the building four levels of cost estimation can be made. The first level is a cost estimation based on the site boundary and massing of the building. In the second level the model is updated with information about the amount of floors and story heights. For the third level also the spaces and vertical elements need to be defined. The most detailed level consists of a cost estimation where also specific information, like building materials, is used. This separation of levels is very useful as it allows for a method that bases cost estimation on the level of available information.

Ma Wei and Zhang (2013) showed the real potential of automatic QTO from BIM models. In their research they were able to (semi) automatic extract the QTO in compliance to the Chinese mandatory specification of cost estimation in the tender phase.

4.4. Conclusion of literature review

Based on the information obtained during the literature review, it is now known how BIM models can be used as the data source for KM. By using the data mining process one should be able to convert the knowledge embedded in BIM models to useful explicit knowledge for cost estimation. The steps necessary to convert BIM models with corresponding building costs to a cost estimation model are:

1. Understanding and defining the problem
2. Collect necessary data
3. Convert non-IFC BIM models to the IFC data format
4. Convert the IFC models to a dataset
5. Add the corresponding costs
6. Clean the dataset
7. Select a useful subset
8. Engineer a useful data mining algorithm
9. Use the algorithm
10. Evaluate results with numerical prediction measures
11. Redefine data and problem
12. Use the generated model for cost prediction

On the next page the proposed model is given, this model combines the enumeration given above with the previously mentioned SECI model on page 37.

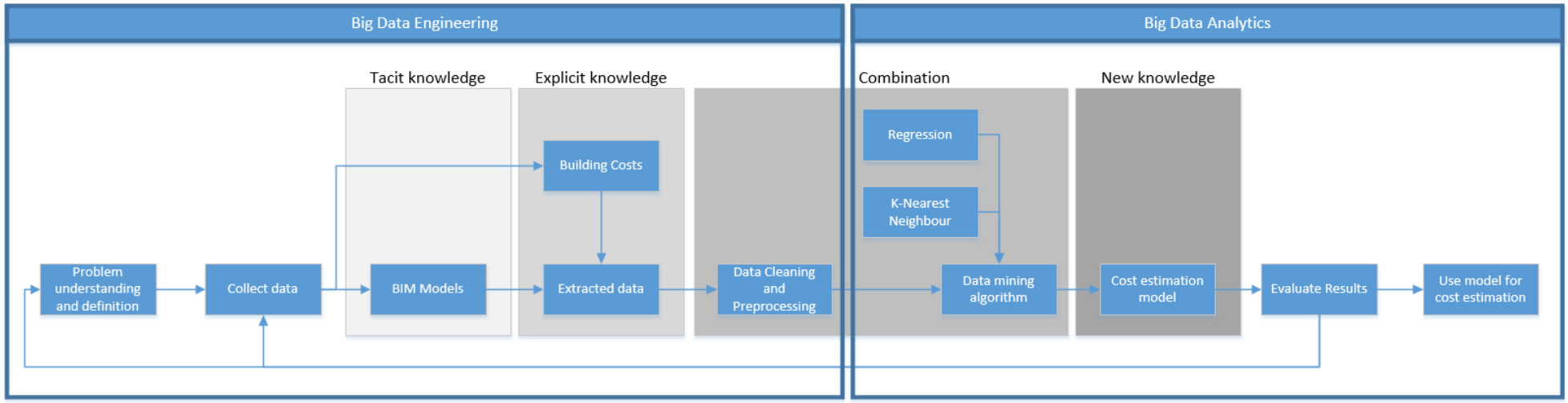

Figure 4: Proposed model

This page is intentionally left blank

Part I: Problem statement

Part II: Theoretical review

Part III: Practical implementation

Part IV: Conclusions and recommendations

This page is intentionally left blank

5. System Design

From the literature review it is known that the process of discovering advantageous patterns in data is called data mining (John, 1997). This process has been adjusted to fit the presented research and consists of the following steps.

1. Understanding and defining the problem
2. Collect necessary data
3. Convert non-IFC BIM models to the IFC data format
4. Convert the IFC models to a dataset
5. Add the corresponding costs
6. Clean the dataset
7. Select a useful subset
8. Engineer a useful data mining algorithm
9. Use the algorithm
10. Evaluate results with numerical prediction measures
11. Redefine data and problem
12. Use the generated model for cost prediction

The first step of the data mining process, understanding and defining the problem, has been dealt with in the previous part of the research. This part of the research will focus on the other steps of the process. In this chapter steps 2 to 7 are discussed. In chapter 6 the steps 8 to 12 will be explained.

5.1. Data collection and extraction

The second step of the data mining process is data collection and extraction. As discussed in the literature review, the BIM and costs data are already created for other purposes and stored in diverse databases. These data need to be collected and stored in a single database. For this research, construction companies in the Netherlands were asked to provide their BIM and matching costs data. This has resulted in 15 sets of data, each containing one building. Three of these sets were not suitable for this research because the type of buildings that were built in the projects were industrial constructions as opposed to the 12 other sets that were residential buildings.

Although 12 suitable sets of data were collected, these data are not directly usable for data mining. First the usable data need to be extracted and stored in a format that can be accessed by the data mining algorithm (John, 1997). Thus first it has to be determined in what format the original data are stored and in what format the machine learning algorithms need to analyse the data.

In the literature review is explained that BIM data could be shared in either a proprietary file format or non-proprietary file format. Also was explained that the format used in this research would be a universal standard format and therefore a non-proprietary format, because it allows all the provided BIM data to be used. As IFC is the most widely accepted non-proprietary BIM data format, IFC is used as the file format in this research.

The data collected had both types of formats as in advance there was no demand given to the construction companies in which format the BIM data should be provided. The reason for not demanding the data to be provided in the IFC format was because it would require more effort from the construction company and therefore it could reduce the chance of cooperation by

those companies. It was not worthwhile taking this risk as converting the format from the BIM authoring tool to IFC would only require access to the BIM authoring tool and this would also take time. Both were available in this research.

From the collected datasets, five of them were provided in the original format created by the BIM authoring tool and seven were provided in the IFC format. From the five provided in the original format all were provided in the RVT format, which is the original format of BIM data created in Revit, a product of Autodesk.

The costs data were provided in two file formats. One was provided in PDF, the rest was provided in XTBF, the file format used by Ibis TRAD, the building costs calculation package of Ibis BV (A. Lassche, personal communication, September 14, 2017). It was chosen to use only to total costs of the building. This choice was made based on the amount of data available. As this is a small amount, it is necessary to keep it as generalized as possible. As only the total costs data need to be determined, it was chosen to convert these data manually to the required format.

Data extraction

In the data extraction process the necessary information is extracted from the input data and written to the output data (see figure 6). As the input data are already known, a selection needs to be made on the output data and how these data are extracted. As the goal of the extraction process is to get the necessary output, first the output data will be selected. Afterwards a selection will be made of how to extract the necessary data from the IFC data file (input data).

To extract the usable data from the collected data, there needs to be a file format to store the usable data into it. This file format needs to be suitable to be used with the machine learning algorithm. It is also necessary that the programming language used to extract the usable data from the collected data can write to this specific file format, preferably as a standard option.

The machine learning algorithms will be provided by R. R is an open source software environment for statistical computing. One of the many applications of R is data mining (The R foundation, nd). The most used file format for R is a simple text format. To separate the text filed a separator is used. A comma is an easy and useful separator. A simple text file format with a comma as a separator is called a CSV, an abbreviation of Comma Separated Value. CSV is a standard file format often used to exchange data between application and a good option for importing data into R (Togaware, nd).

As it is known that we need to convert IFC to CSV, it is possible to determine how the data can be extracted. There are multiple commercial and non-commercial tools available for the interpretation of IFC models. Solibri Viewer is one of these tools that allows users to view IFC models. It is also able to determine properties of the IfcObject, like area, thickness and volume. However it does not allow for automatic data extraction to a data format that can be used with a machine learning algorithm. As mentioned before, it is important that as much as possible is automated in the knowledge management system because the more people have to do this themselves, the lower the chance of actual implementation. Because of this it was chosen to program a new extraction tool for this research.

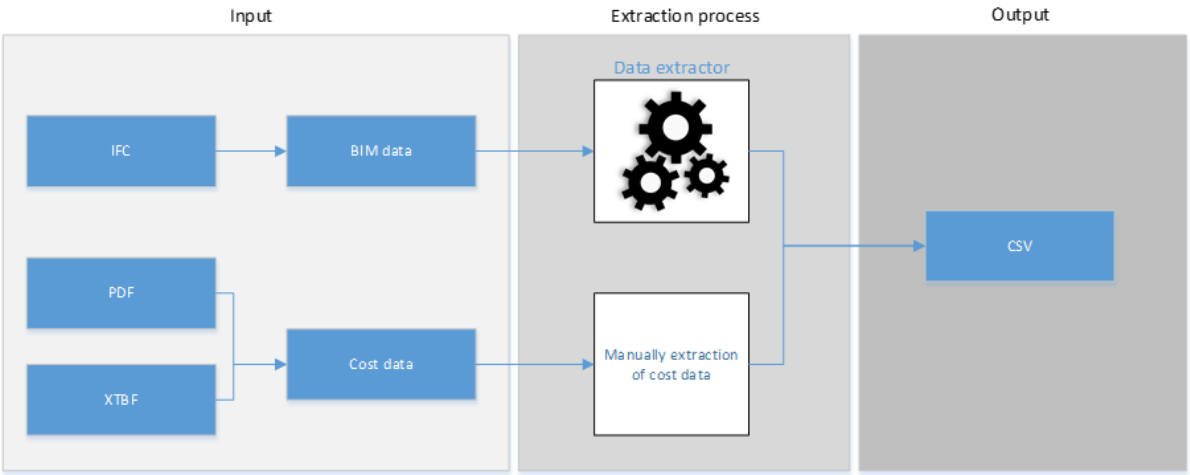

Figure 5: From input data to CSV

The programming language that is used is Python in combination with IfcOpenShell and PythonOCC, the programming language used in the research by Krijnen (2015). This was chosen because IfcOpenShell is one of the very few libraries that allow programming access to IFc model. And because Krijnen very willingly shared his knowledge, this was used in the development of the extracting script.

The Python script uses the IfcOpenShell module to interpret the object based IFC data. It sorts the data by IfcElement. This is done by creating a Python list based on the IfcElement that the extractor needs to extract using the '*by_type*' function of the module. As the extraction script needs to be able to understand what kind of elements it is extracting from the IFC file, it is necessary to program for every element how it can be classified, mostly by IfcElement and how it can calculate or interpret the necessary information to extract. Therefore it is necessary to select possible relevant IfcElements to be extracted.

The goal of the extractor is to create a CSV format file that contains the dataset to be used in the data mining algorithm. Essentially this file would contain a quantity take off (QTO) of the IFC file. Based on the literature research, the cost estimation can be determined by two different types of quantity take-off. Either a very detailed level of extraction as Ma et al. (2013) used or a very general extraction as used in the research of Cheung et al. (2012). This extracted quantity take off would be used as the dataset of the big data process. The one suggested by Ma et al. (2013) would result in a very detailed dataset, in which every element of the building is extracted. This would give a very precise outcome when estimating the costs of the building with known unit rates. However, in this research the unit rates are calculated through the machine learning algorithm. For this a statistical correlation has to be found between the different buildings. With the amount of data available it is more likely to find this correlation with general elements. Therefore the QTO suggested by Cheung et al. (2012) is chosen. Based on the research of Cheung et. al (2012) the following components need to be extracted:

Table 2: Building elements to be extracted

Group elements	Elements		Unit
Substructure	Substructure	Ground floor area	m^2
Superstructure	Frame	Area of all floors	m^2
	Upper Floors	Same as frame only without ground floor	m^2
	Roof	Area of roof as viewed from plan	m^2
	Stairs and ramps	Number of stairs within building	no
	External walls	Area of external surface of the external walls	m^2
	Windows and external doors	Area of external surface of the windows and external doors	m^2
	Internal walls and partitions	The vertical area of the inner walls	m^2
	Internal doors	Amount of inner doors	no
	Wall finishes	Vertical surface area of the walls that needs finishing, area per wall finish type	m^2
	Floor finishes	Upper surface area of the floors, area per floor finish type	m^2
	Ceiling finishes	Lower surface area of ceilings, area per ceiling finish type	m^2
Fittings and furnishings	Fitting and furnishings	Total area for fitting and furnishings	m^2
Services	Services	Total area for services	m^2

Based on the level of detail within the BIM models, it is not possible to determine the finishing of the wall, floor and ceiling, the fittings and furnishings and services of the buildings. In only four of the BIM models information about the finishing and the fittings and furnishings were available. For services it was even more problematic as for only six IFC models the spaces were defined and even if the spaces were defined only little information was available about the quality of service within the spaces. Therefore these elements are ignored in this research.

The substructure, frame, upper floors and roof could be determined by calculating the area of the IfcSlabs located on the corresponding floor(s). The stairs can normally be determined by the IfcStair element. However in some IFC models the stairs are exported to IfcBuildingElementProxy or IfcFurnishingElement. Using a filter on these IFC elements the amount of stairs could be determined. The external and internal walls can be extracted with the IfcWall elements, and the doors with the IfcDoor elements. Both needed a filter to determine if they are located inside the building or located in the external wall. The windows can be determined with IfcWindow.

Table 3: Overview of building elements with the corresponding IfcElement

Group elements	Elements		IfcElement
Substructure	Substructure	Ground floor area	IfcSlab
Superstructure	Frame	Area of all floors	IfcSlab
	Upper Floors	Same as frame only without ground floor	IfcSlab
	Roof	Area of roof as viewed from plan	IfcSlab
	Stairs	Number of stairs within building	IfcBuildingElementProxy/ IfcFurnishingElement/ IfcStair
	External walls	Area of external surface of the external walls	IfcWall
	Windows and external doors	Area of external surface of the windows and external doors	IfcWindow and IfcDoor
	Internal walls and partitions	The vertical area of the inner walls	IfcWall
	Internal doors	Amount of inner doors	IfcDoor
	Wall finishes	Total vertical surface area of the walls that need finishing	IfcWall
	Floor finishes	Total upper surface area of the floors	IfcSlab
	Ceiling finishes	Total lower surface area of ceilings	IfcSlab

This all results in a dataset with the following attributes: area of ground floor, area of upper Floors, area of roofs, amount of stairs, area of external walls, area of external openings, area internal walls, amount of internal doors.

The process of creating the extractor script contained a lot of trial and error. The biggest challenge was the data management of Python. In the original script, based on the scripts of Van Strien (2015) and Krijnen (2015) all the objects in the IFC file were imported and converted to shapes at the same time, which allow for interpreting the objects as related to each other. The problem however was the amount of objects that the Python script needed to interpret. As Van Stien (2015) and Krijnen (2015) both used a small IFC model called IfcOpenHouse to test their script and this research uses IFC models from real construction projects, the amount of objects is significantly higher in the latter. The IfcOpenHouse model contains only 47 IfcProducts in contrary to the average of 6255 IfcProducts that the models used in this

research contain. As a result the original script crashed due to the limited amount of memory available to run the script.

This problem was eventually solved by changing the order in the script. Instead of creating all shapes directly, first a list was made of elements needed to calculate a particular attribute. An overview of the element needed per attribute can be found in table 3. Because for the floor area the floor level is also important, the list of IfcSlabs was split by floor levels using the link between the IfcSlab and IfcBuildingStorey.

For some elements the location of the element matters. For example for the IfcWall it matters if it is an inside or an outside wall. In most IFC models this can be determined with the IsExternal property in the PSet_WallCommon. However as one of the IFC models did not contain the PSet_WallCommon this was not used. A filter was used based on name and material. Because all the outside walls in the IFC models are made of brickwork a filter was used to find all IfcSlabs with material type brickwork. In the theory it was already discussed that there are initiatives to ensure that the proper information is shared within the IFC model. The BIM Basis ILS prescribes the use of Pset_WallCommon to share information about the properties of the wall. If this would have been implemented by the exportation of all BIM models to the IFC data format, the Pset_WallCommon could have been used and therefore

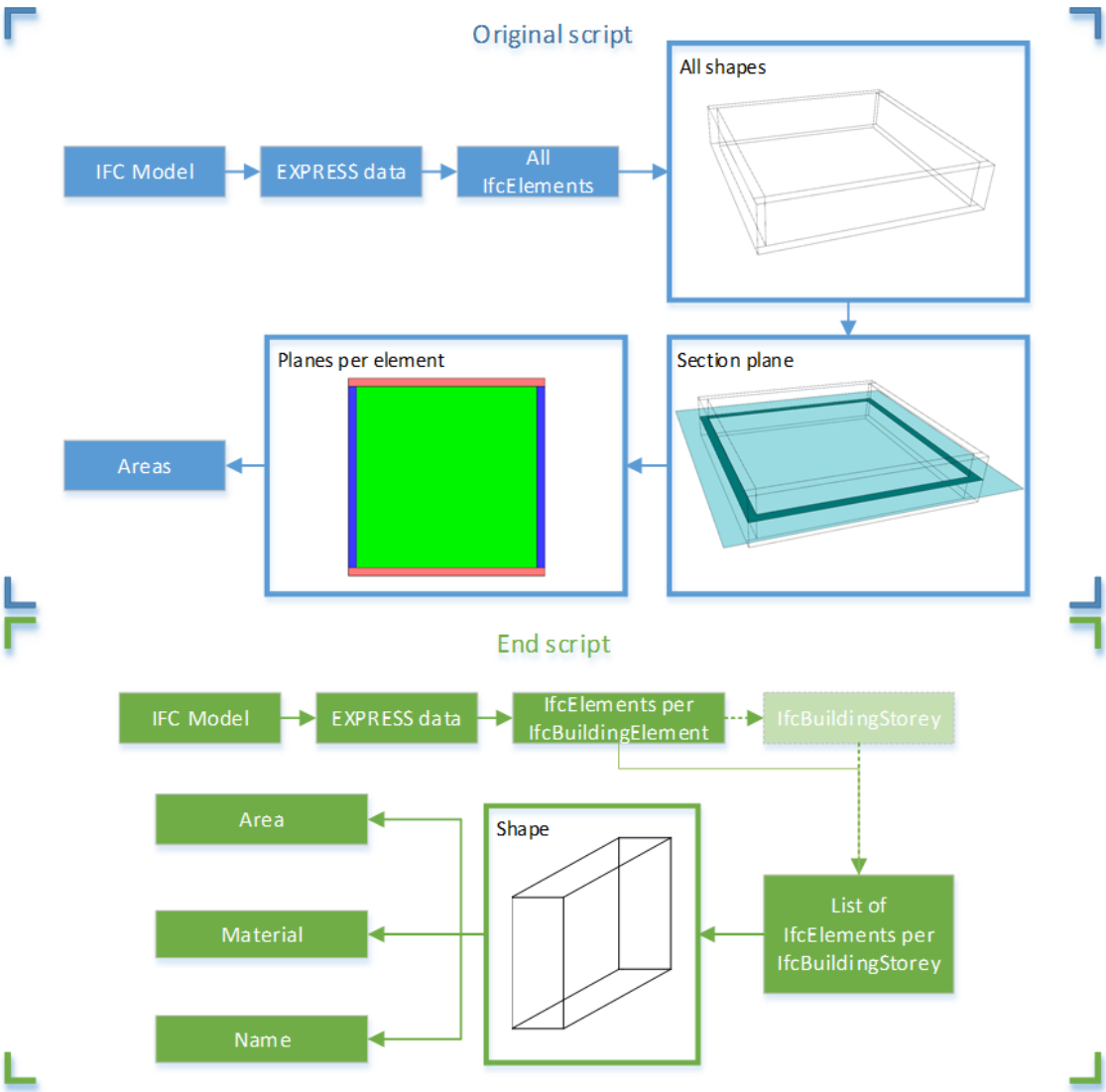

Figure 6: Structure of extractor script

the script would have worked also with external walls that are not made of brickwork.

It should be noted that although changing the order of the script made the script work, it also means a loss of information. The new script analyses every element individually. Therefore most of the information about how the elements are located to each other is lost. This made it impossible to analyse for example the area of particular rooms based on building elements. In the left part of figure 8 it is shown how the area of a room could be determined if the preferred script would have worked. On the right part of this figure the working of the actual script is shown. In the table below, the dataset as created by the extractor can be found. Notable is the 'Count Stairs' attribute as two of the return values are zero. This is because no information about stairs could be found in those IFC models. It is very common in data mining that the dataset is cleaned after extraction. More about this in the next section.

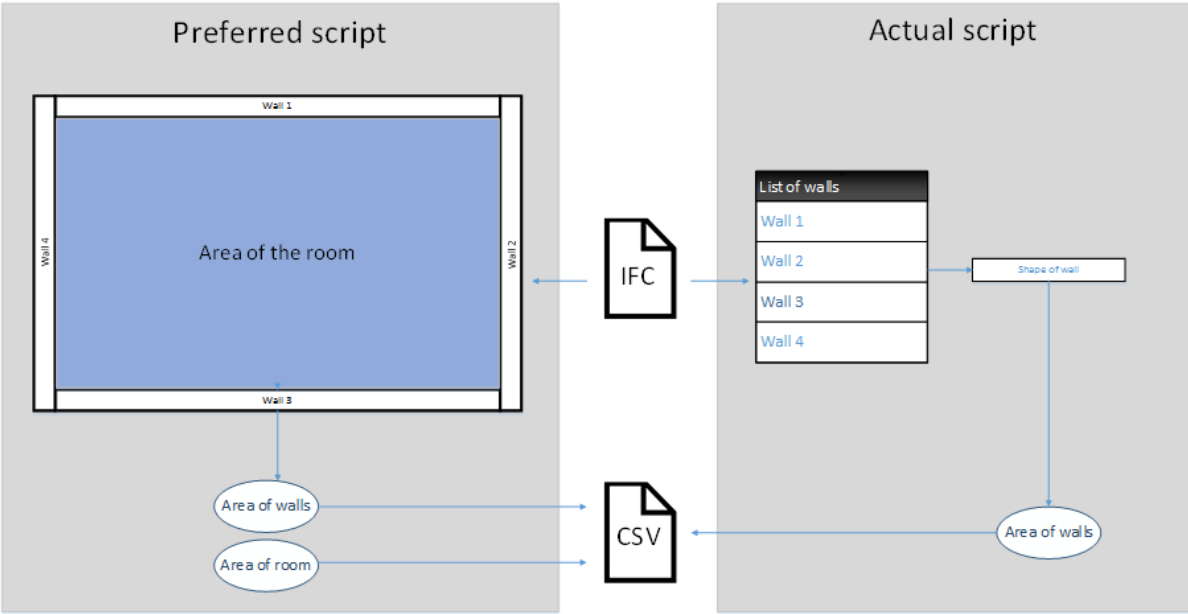

Figure 7: Distinction between the preferred and actual working of the script

Table 4: Overview of the extracted dataset

Filename	Area Ground Floor	Area Upper Floor	Area Roof	Count Stair	Area External Walls	Area External Openings	Area Internal Walls	Count Internal Doors
A.ifc	842	652	842	0	2902	726	3452	252
B.ifc	584	2101	654	9	2614	668	3352	466
C.ifc	595	1984	880	4	3607	673	2326	421
D.ifc	682	1948	850	4	3224	690	2894	426
E.ifc	567	2234	615	5	2405	647	2962	176
F.ifc	880	696	880	26	1829	500	2608	130
G.ifc	569	931	603	20	1984	315	1989	100
H.ifc	280	186	321	12	920	149	1147	54
I.ifc	233	359	252	8	960	130	724	40
J.ifc	374	1374	475	6	1891	414	1936	294
K.ifc	1496	364	1496	32	6560	2017	2846	177
L.ifc	432	575	436	0	1307	279	674	102

5.2. Data cleaning and Exploration

In a traditional data mining process, the amount of data in the database is very large. As large databases contain a lot of samples, the chance of errors in the created dataset is very likely. In order to be successful, it is required that the system is robust regarding errors in dataset (John, 1997).

For large datasets, a sophisticated method is needed to test the robustness of the datasets. This is necessary because of its size, it is not really possible to check the data manually. The dataset created by the extractor however is small. Although it is preferred to have a large database, the small size of the used database allows us to check the extracted data by hand.

In this research most of the data cleaning is actually done by the extractor. For the structure of the thesis however, it has been chosen to separate the extraction and cleaning of the data. The way the set data was checked has been done by comparing the extracted data with the data in the original costs estimate. Unfortunately the data appeared to be anything but robust. For almost every project the extracted data did not match the data from the original cost estimate. One of the most important reasons is that the IFC models were not made for this purpose. Most of them were once created by the architect and never updated in the further process. Even though the architect also created the specifications of the project, many cases were found in which the IFC model did not match the made specifications.

One of the most commonly found errors in the data is mislabelling of the IfcObjects. As explained in the literature review IfcObjects are the generalisation of the components in an IFC model. The most often mislabelling was the mislabelling of stairs. In the IFC models the stairs were either labelled as IfcStair, IfcFurnisingElement or IfcBuildingProxyElement. As explained in the previous chapter, in the different models the stairs were modelled with three different IFC elements, IfcStair, IfcBuildingElementProxy and IfcFurnishingElement.

This is a problem because as the name IfcFurnishingElement suggests, this type of IfcElement is a generalization of all furniture related objects (buildingSMART, nd). This is even more the case for IfcBuildingElementProxy as this is “a proxy definition that provides the same functionality as an IfcBuildingElement, but without having a defined meaning of the special type of building element, it represents” (buildingSMART, nd). As a result a query on the name of the IfcFurnishingElements and IfcBuildingElementProxy is necessary to find the stairs within all of the IfcFurnishingElements. This is especially hard because as there is no standardisation in the naming of IfcElements and therefore it is not necessarily the case that the word ‘stair’ or a translation of this word for example the word ‘trap’ in Dutch, is within the name of the IFC element.

In only two models the stair was modelled as an IfcStair. Five models had the stairs modelled as an IfcFurniture with the word ‘trap’ in the name of the elements. In three of the models the used element for the stair was IfcObject and there was no indication in the name that is was a stair. Therefore it was necessary to query on the used name, ‘nl0-dht_nl+1-oht_2kwart,’ to find the stair. In the dataset, twice the extractor returned zero for the attribute Count Stairs. In these IFC models the stairs were not modelled. For the dataset to be robust the actual values are updated manually.

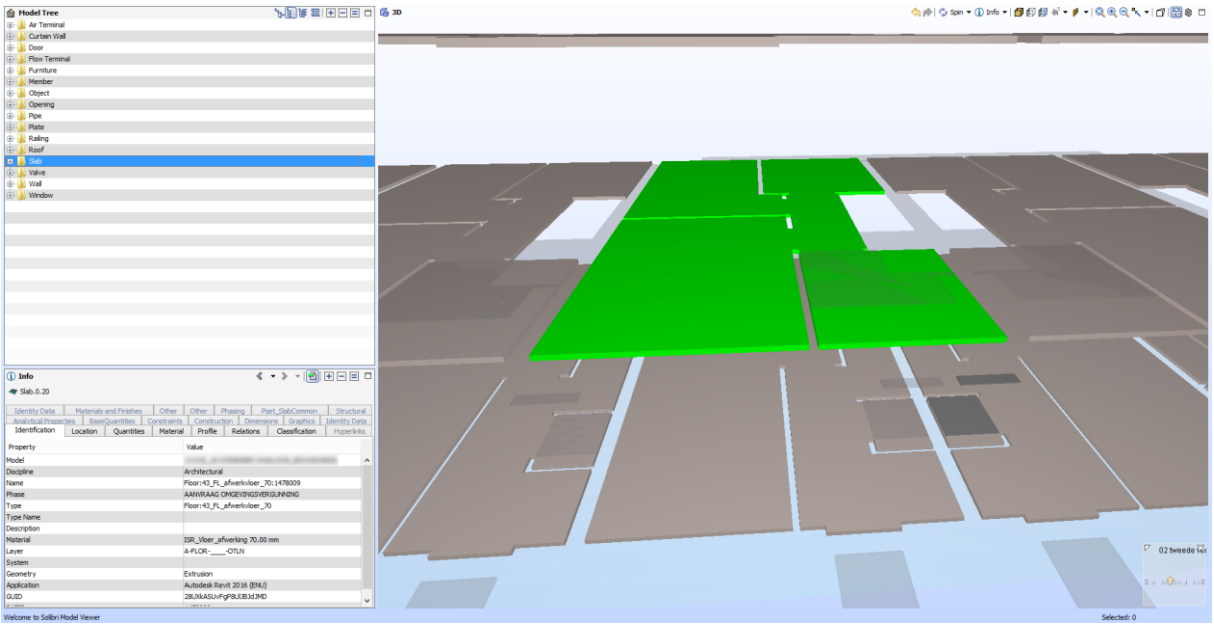

Figure 8: Model without structural floor

The general problem of mislabelling the IfcObjects could be solved with the use of either an IDM, more specific the BIM Basis ILS for the Nederlands or an IFD. The BIM Basis ILS prescribes the use of correct entities and the Nederlandse Revit Afspraken even describes how to use a Mapping Table within Revit to make sure the correct entities are exported. The Nederlandse Revit Afspraken is a document created by Pijffers (2016) in compliance with BuildingSMART Benelux which is the managing organisation of the BIM Basis ILS.

During the exploration of the IFC models two projects were found in which no structural floors were defined using IfcSlab (see figure 9). In both cases only the screed floors and roof insulation were modelled as IfcSlab elements. After some exploration, for one of the models the structural floor could be found as a IfcBuildingElementProxy. After comparing the other IFC model with the cost files, it was found that the structural floor was a wide slab floor. Because of this, the extractor needed extra code to check the presence of different types of floors and to use the appropriate type for the specific model. As already explained in the BIM Basis ILS is prescribed that the appropriate entity should be used, in this case a IfcSlab. This would have solved the problem for one of the two models. However for the other model this would not have solved it as the structural floor was never modelled.

In the dataset the floor area is important for different attributes. As it is used for different attributes calculating the total floor area is not enough. The area is needed per floor. To be able to calculate the area of a floor, the script needed to be able to interpret on which floor the IfcSlab is located, as the IfcSlabs were used to calculated the floor area. As IfcSlabs are linked to a IfcBuildingStorey and IfcBuildingStorey's are used to model different floor levels, in most IFC models a combination of IfcSlabs and IfcBuildingStorey was enough to interpret the area per floor level. However in some models, the IfcSlabs were not linked to the correct IfcBuildingStorey. For these IfcSlabs an offset was used to give them the correct height. For this problem also additional code was written (see appendix 2, lines 332-346) to get the correct output from the extractor. The code however is very model specific and therefore would not work with other IFC models that present the same kind of problem. If the model would have been made in accordance to the BIM Basis ILS this problem would have been solved as the BIM Basis ILS only allows for one main level for every building storey.

In his dissertation, John (1997) said about the data: “in the end the data stored in a database is no better than the typing skill of the data-entry clerk or the decisions made by experts that are recorded in the database” (pp. 51-52). As most IFC models are created by architects to show their design and limitedly to never updated during the rest of the design phase, few decisions made by experts are recorded in the data. In the normal building process this is not really a problem, as most contractors still use the 2D drawings and the specifications to create a cost estimate. But as the AEC industry is slowly starting to use the possibilities of BIM data, the accuracy of the BIM models will become more important. As the quality of the IFC files is problematic for automated cost estimation based on IFC files, we can already conclude that with the current state of the AEC industry regarding BIM, it is not yet possible to use the BIM files for automated knowledge management. However if the BIM Basis ILS would be implemented much of the problems that occurred would be solved.

5.3. Data engineering

5.3.1. Sub dataset selection

The data engineering part of the data mining process consists of a large number of possible activities. One of the most important activities is to select a subset of attributes that will be used in the rest of the data mining process (John, 1997).

The subset selection has already taken place in the data extraction phase. This is because every attribute that is extracted from the IFC data, had to be programmed into the data extraction script created for this research, as explained in section 7.1. Therefore already choices were made of which types of objects to extract from the data and thus which possible attributes would likely be relevant.

However in the dataset there are still 8 independent variables compared to 12 observations. This is already a high number of independent variables compared to the number of observations. This becomes even more problematic when the dataset needs to be split in a trainings dataset and a test dataset. To reduce the amount of independent variables a method is needed to select the most relevant variables. This is called feature selection. To evaluate the dataset, first a correlation plot is made of the dependent variables and all the independent variables, see figure 10. It shows that most of the attributes have a very low correlation. According to Guyon and Elisseeff a variable ranking method is the best option to use whenever the data is noisy. The result of the ranking method is the following ranking:

1. AreaInternalWalls
2. AreaExternalOpenings
3. AreaRoof
4. CountInternalDoors
5. AreaUpperFloor
6. AreaGroundFloor
7. AreaExternalWalls
8. CountStairs

Therefore it has been chosen to apply stepwise regression for the regression model. In stepwise regression every step of a variable is added to the algorithm as long as the performance of the algorithms increases.

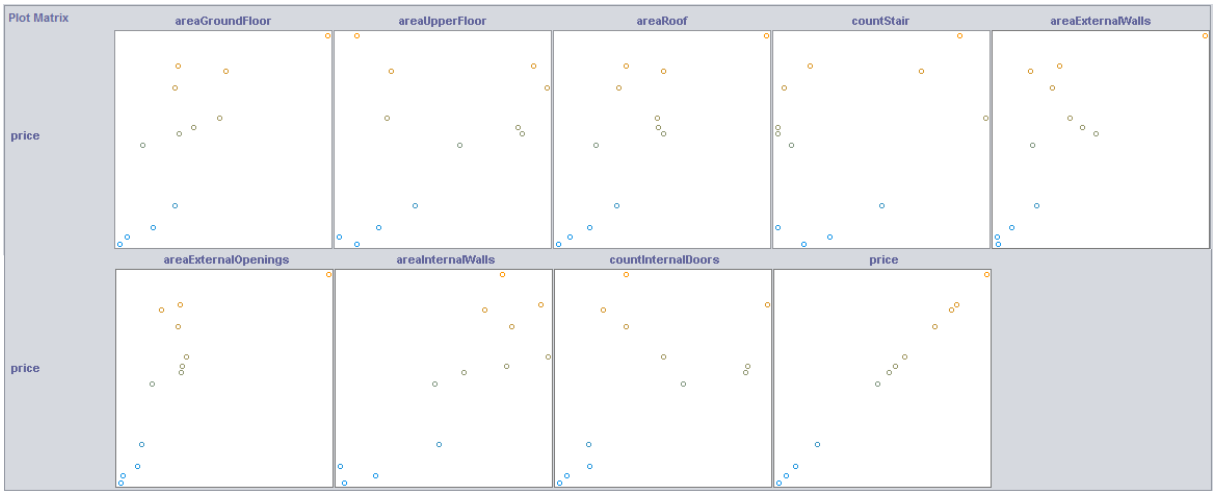

Figure 9: Correlation plot

5.3.2. Normalization

The classification algorithm k-Nearest Neighbour uses the Euclidean distance between the variables to calculate the closest neighbour. The Euclidean distance (*d)* is calculated with equation 5. It calculates the straight-line difference between two points. Therefore it is necessary to scale all variables to a similar scale. Otherwise the variables with high values would have a higher influence on the dataset as opposed to variables with low values.

$$Euclidean\ distance: \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \qquad (5)$$

The variables can be normalized using equation 6. Via this equation all variables are transformed into variables between zero and one. Via this equation the range within a variable is restricted between a zero and a one. Therefore the relative distance between a variable with a small range of values has the same influence as the relative distance between variables with a high range of values.

$$Normalization = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (6)$$

5.4. Conclusions data engineering

In this chapter the data engineering part of the data mining process is discussed. Much attention has been paid towards the extraction of the information from BIM models into a dataset. It has become clear that there is a difference between the theoretical potential of the BIM models and the actual potential of the BIM models in the current state of the AEC industry. As discussed in the literature review, theoretically it should be possible to extract all physical data of the buildings from the BIM models. However the current quality of the BIM models has presented problems.

First of all, too often the models are not a real representation of the actual building. For example in the collected BIM models part of the building were found missing. Another problem with the quality of the BIM models is the mislabelling of the building element type. Elements were modelled as walls while they were for example part of the window frame. This presents a problem in automatic QTO as the extractor script interprets the object as the wrong type and thus extracts an incorrect QTO.

As discussed in the literature both how and which information that is shared with the IFC model can be standardised with an IDM and an IFD. Using this standardisation would solve most of the problems that occurred with the extraction of the QTO.

In the literature review it was made clear what the theoretical strength of the proposed concept is. In this chapter however it was shown what the weaknesses of the concept are. The whole system depends on the quality of the BIM models and it has been shown that at this point the models available in real life situation do not meet the level of quality needed for the system to work automatically.

This page is intentionally left blank

6. Data analytics

In this chapter, it is explained how the different machine learning algorithms are applied and what the results are. Additionally the conclusions of the practical application are discussed. It should be noted that the dataset is very small and therefore not really useful for data mining. However in this chapter it is explored if some useful relationships can be found.

6.1. Algorithm engineering

In the algorithm engineering step the appropriate machine learning algorithm is chosen or if necessary created. For this research two different machine learning algorithms are chosen, as discussed in the literature review. These two algorithms are multiple linear regression and k-Nearest Neighbour. These algorithms were chosen to test the performance of both a regression algorithm as a classification algorithm, which are the two types of machine learning algorithms.

The software package R will be used for the data mining analytics. R has a large library of algorithms that can be used. Additionally a graphical user interface (GUI) for data mining called Rattle has been created by Togaware. This GUI provides an interface for R which has the advantage that it creates the code necessary to run the algorithm in R and thus prevents unnecessary coding. For this reason the Rattle package is used. Unfortunately not all algorithms that are available in software package R are available within the Rattle GUI. This is also the case for k-Nearest Neighbour. Consequently the code for running the algorithm k-Nearest Neighbour has been created during the presented research. This code can be found in appendix 3.

6.2. Running the algorithm

As discussed in the literature review the performance of the algorithm can be tested with statistical evaluation measures. To evaluate the performance, the correlation coefficient (CC), the root mean squared error (RMSE), the mean absolute error (MAE) and the root relative squared error (RRSE) are used.

For the linear regression algorithm only the input variables need to be determined before the algorithm can be run. In the previous chapter, feature selection was used to determine a ranking for the relevance of the variables. To determine the optimal amount of attributes used, stepwise regression is applied.

To test the performance of the model, the dataset is randomly split into a training set and a test set. The split percentage has been set on 75%. Three different test- and training sets have been applied as can be seen in table 5. The table shows the result of the algorithm is very dependent on the randomly selected training and test set. Therefore it can be concluded that the amount of data available is too little to have any significant meaning.

Table 5: Performance of the linear regression algorithm

Amount of variables	Measures	Result 1	Result 2	Result 3
AreaInternalWalls	CC	0.723	0.9838	-0.9976
	RMSE	560617.4449	505518.9929	627267.4708
	MAE	481468.0006	466434.9967	508912.5596
	RRSE	48.0045 %	36.2489 %	64.5285 %
AreaInternalWalls	CC	*0.7870*	0.9902	-0.9909
AreaExternalOpenings	RMSE	*433203.4989*	555776.8736	621719.9478
	MAE	*402421.8829*	435928.7036	570460.5939
	RRSE	*37.0943 %*	39.8527 %	63.9578 %
AreaInternalWalls	CC	0.6855	0.9674	-0.8468
AreaExternalOpenings	RMSE	459702.867	959054.6101	791664.2357
AreaRoof	MAE	428591.8218	825663.6968	614111.1752
	RRSE	39.3634 %	68.7702 %	81.4404 %
AreaInternalWalls	CC	0.4980	0.9142	-0.8131
AreaExternalOpenings	RMSE	544428.3547	355500.1695	785137.176
AreaRoof	MAE	513398.0693	401117.7355	1103176.0896
CountInternalDoors	RRSE	46.6183 %	40.1181 %	113.4863 %

For the k-Nearest neighbour algorithms besides the amount of variables also the amount of nearest neighbours used to predict the value are important. As the k is the amount of samples used to determine the costs based on their average, the result of the algorithm is highly influenced by the number of k's used. In this research the following k's are evaluated, k: 1, k: 3 and k: 5. For all three k's the performance was evaluated on the dataset. In table 6 the performance of the algorithm with k: 3 and only areaExternalOpenings as variable is shown.

Although the difference in outcome between the three results is relatively the same for RMSE MAE and RRSE, the difference in the correlation coefficient indicates that also the k-nearest neighbour is not useful on the data. The result of the first and third result indicate that the area of external openings is positively correlated with the costs of the building, where the second result indicates that the correlation is negative.

Table 6: Performance of the k-nearest neighbour algorithm

Measure	Result 1	Result 2	Result 3
Correlation coefficient	0.9425	-0.7737	0.9994
Root mean squared error	521440.4598	401838.3491	626107.6
Mean absolute error	460706.2222	347361.6667	473364.6667
Root relative squared error	44.6499 %	41.3381 %	44.8958 %

6.3. Evaluation of result

Based on the result in the previous chapter it can be concluded that the dataset is indeed not suitable to apply data mining. The main reason is that the dataset is too small. If a dataset is small the dataset has low statistical power and the research is therefore unlikely to result in reprehensive conclusions (Button, Ioannidis, Mokrysz, Nosek, Flint, Robinson & Munafò, 2013). As only 12 projects were used in the evaluation, this is certainly the case. However in the current situation not many firms in the AEC industry will have many more BIM models of one type of buildings. So even though Hoogeveen (2015) proved that the concept works theoretically, for the concept to work within real-world situations, more BIM models need to be generated. And as already concluded, these BIM models need to have a higher level of quality as the current standard. Another problem with the dataset is that because of the small amount of projects no difference was made between apartment buildings and houses. Although they both have the same function their designs differentiate.

6.4. Redefining data

In the last part it was concluded that the dataset was not suitable for data mining. Therefore additional data about residential buildings have been collected and added to the dataset. These data were found using three sources, namely:

- Bouwkosten: nieuwbouw en renovatie projectanalyses (Kuhlmann, 2007)
- Bouwkosten.nl (2017)
- Casadata.nl (2017)

On the next page the collected data are given. As not all of the data contained information about the internal walls or the internal doors, these variables have been ignored. In appendix 4 the used sources can be found.

To compare the prices of the projects it is important to normalize the prices to the same reference point in time. As some of the collected data were about projects in 2002, this step was important to be able to compare the prices accurately. The used formula to normalize the price is:

$$Price\ normalization = Price_{original} * \frac{PriceIndex_{2017}}{PriceIndex_{original\ date}} \quad (7)$$

In appendix 5 the price index from the last 15 years can be found.

The old dataset was not suitable for data mining because it had had two problems, it was small and it contained two types of residential buildings, namely apartments and houses. Therefore it was chosen to create two new datasets. One dataset with all available data of both the old and new data. The other dataset contains data from both the old and new data, but only from projects that were apartments. It was chosen to create a dataset for apartments and not houses as there was more data available about apartments than about houses. In the remainder of this research the dataset with all the data is called dataset A and the dataset with only apartments is called dataset B. In the next part both datasets will be used in the algorithm to evaluate which of the two datasets perform best.

Table 7: Additional data from databases

Project name	Source	Price index	Area Ground Floor	Area Upper Floor	Area Roof	Count Stair	Area External Walls	Area External Openings	Original price	Normalized price
Nieuwbouw 14 rijenwoningen Groenoord	casadata.nl	1-3-2017	645	2085	645	42	1628	652	€ 2.178.000,00	€ 2.178.000,00
Nieuwbouw 22 rijenwoningen Dronten	casadata.nl	1-3-2017	1177	1294	1177	44	1291	317	€ 2.005.310,00	€ 2.005.310,00
Nieuwbouw 31 rijenwoningen	casadata.nl	1-3-2017	1995	1762	1995	31	2632	770	€ 3.892.552,00	€ 3.892.552,00
28 appartementen De Haere	Bouwkosten.nl	1-3-2017	1165	2137	1322	4	2092	606	€ 3.380.015,00	€ 3.380.015,00
Appartementencomplex de barchaan	Bouwkosten.nl	1-3-2017	599	1050	656	3	1528	574	€ 2.641.105,00	€ 2.641.105,00
44 apartementen de parkwachters	Bouwkosten.nl	1-3-2017	1045	4976	1368	19	4358	1561	€ 5.665.615,00	€ 5.665.615,00
29 appartementen boven winkels	Bouwkosten.nl	1-3-2017	990	2521	990	12	1981	565	€ 4.623.110,00	€ 4.623.110,00
Woontoren met 40 appartementen	Bouwkosten.nl	1-3-2017	673	5007	713	8	3603	1604	€ 4.924.405,00	€ 4.924.405,00
Allardhof te buren	Bouwkosten.nl	1-3-2017	1220	1313	1256	6	3653	565	€ 2.414.120,00	€ 2.414.120,00
17 appartementen Oudenbosch	Kuhlmann (2007)	1-4-2002	487	976	554	4	1389	380	€ 1.359.660,00	€ 1.622.132,71
49 seniorenappartementen Dronten	Kuhlmann (2007)	1-4-2002	1463	4785	1231	3	3425	1222	€ 3.677.005,00	€ 4.386.824,70
15 stadswoningen Maastricht	Kuhlmann (2007)	1-10-2002	981	1412	1393	31	2096	646	€ 2.039.045,00	€ 2.396.184,79
Appartementen Winterswijk	Kuhlmann (2007)	1-5-2002	2451	3210	2451	11	2554	616	€ 3.562.415,00	€ 4.228.964,40

6.5. Using the new datasets

For the new datasets the same methods have been applied as were for the old dataset. The input variables were determined with a ranking method. And again the dataset has been randomly split into a training set and a test set, where 75% was put into the training set and 25% into the test set. Also the same methods were used to test the performance of the algorithm, which are:

- the correlation coefficient (CC),
- the root mean squared error (RMSE)
- the mean absolute error (MAE)
- root relative squared error (RRSE)

The performance of the data mining process on the new datasets can be found in table 8 and 9. In table 8 the performance is measured on the dataset A. In comparison to the old dataset this dataset has 13 more building projects in the dataset, which is an increase of 108%. Table 9 shows the performance of the linear algorithm on the dataset B. This dataset has almost the same number of projects, 14 in the new dataset and 12 in the old dataset.

Both tables show an improvement compared to the performance of the old dataset. Especially in table 8 where the consistency in performance, based on different training- and test sets, has greatly increased in all performance measured.

Table 8: Performance of the linear algorithm on the dataset A

Amount of variables	Measures	Result 1	Result 2	Result 3
AreaUpperFloor	CC	0.6671	0.9467	0.5195
	RMSE	1131852.4198	666785.1939	1122752.2352
	MAE	1003933.5539	571196.4198	945815.3132
	RRSE	76.2631 %	35.8792 %	76.5787 %
AreaUpperFloor	CC	0.7798	0.9081	0.7441
AreaExternalOpenings	RMSE	859523.2126	742378.2462	875311.6464
	MAE	764947.4453	641098.7321	718483.882
	RRSE	57.9138 %	39.9468 %	59.7017 %
AreaUpperFloor	CC	0.7808	0.9064	0.7166
AreaExternalOpenings	RMSE	857395.0371	750482.6377	900271.0147
AreaRoof	MAE	763929.6517	637432.6853	764562.7303
	RRSE	57.7704 %	40.3829 %	61.4041 %
AreaUpperFloor	CC	0.9398	0.9854	0.9441
AreaExternalOpenings	RMSE	515213.5166	644344.289	450966.3423
AreaRoof	MAE	422329.6066	571994.0504	350937.6706
AreaGroundFloor	RRSE	34.7146 %	34.6717 %	30.7587 %
AreaUpperFloor	CC	0.9409	0.924	0.9105
AreaExternalOpenings	RMSE	626445.8633	1070879.5792	532660.4909
AreaRoof	MAE	539378.4266	788590.4438	452562.9005
AreaGroundFloor	RRSE	42.2093 %	57.6233 %	40.9216 %
AreaExternalWalls				

Table 9: Performance of the linear algorithm on dataset B

Amount of variables	Measures	Result 1	Result 2	Result 3
AreaUpperFloor	CC	0.832	0.9986	-0.4039
	RMSE	426982.2271	464236.7724	892102.2197
	MAE	354557.2523	453382.0412	801657.1538
	RRSE	31.9167 %	46.736 %	70.2887 %
AreaUpperFloor	CC	0.7304	0.9996	-0.3797
AreaExternalOpenings	RMSE	475461.7763	458896.0863	925018.2828
	MAE	411819.3968	418881.0927	843154.0895
	RRSE	35.5405 %	46.1984 %	72.8822 %
AreaUpperFloor	CC	0.8158	0.9324	0.7974
AreaExternalOpenings	RMSE	793180.8622	624127.2974	804384.9282
CountStairs	MAE	721734.6241	491184.1518	678728.1529
	RRSE	53.4438 %	33.5838 %	54.8641 %
AreaUpperFloor	CC	0.7015	0.9967	-0.4164
AreaExternalOpenings	RMSE	512596.1218	758006.4997	619014.072
CountStairs	MAE	451168.1877	700553.4729	493597.3885
AeraExternalWalls	RRSE	38.3162 %	76.3107 %	48.7721 %
AreaUpperFloor	CC	0.8081	0.979	-0.9428
AreaExternalOpenings	RMSE	523895.9156	227580.5032	835725.0506
CountStairs	MAE	418128.2352	220203.8128	827008.3429
AreaExternalWalls	RRSE	39.1609 %	22.9112 %	65.8468 %
AreaGroundFloor				

Based on the performance shown in the tables 8 and 9 the best performing linear algorithm was chosen. For dataset A, the best performing algorithm is the one with four independent variables. This is because the performance indicator is generally the best when four variables are used. This has been made more clear in a graphical representation, which can be found in appendix 6. This results in the following formula:

$$Predicted\ price = 244587{,}58 + 598.53 * areaUpperFloor + 691{,}24 * areaExternalOpening + 562{,}00 * areaRoof + 637{,}53 * areaGroundFloor \quad (8)$$

For dataset B, the performance is best when three variables are used. Although only 2 of the four indicators used perform best with three algorithms, as the RMSE and MAE perform better with four variables, the consistency in performance between the different training- and test sets makes three variables the best option. See also appendix 7 for the graphical representation. This results in the following formula:

$$Predicted\ price = 1306976.57 + 763{,}49 * areaUpperFloor + 94219{,}90 * countStair - 629{,}50 * areaExternalOpenings \quad (9)$$

The formula 8 and 9 have been used to predict the prices of the projects within the dataset (see table 10 and 11). The predicted prices are compared to the actual prices to determine the accuracy of the algorithm. According to Phaobunjong (2002) the expected accuracy for a definitive estimate of a building project is within a range of +15% to – 5% and the expected range of a conceptual estimate should be within +50% to -30%.

The margin of error of the predicted prices in table 10 and 11 are too high for a definitive cost estimate. However for a conceptual estimate the algorithms perform acceptable. Only four of the predicted prices are not within the acceptable range for a conceptual estimate. All of these are predicted based on dataset A and are houses.

Where the performance indicators of dataset A have shown that the consistency increases as the dataset becomes larger, the predicted prices of dataset B shows the increased accuracy when only the same type of residential buildings are used. Calculated based on the absolute error in percentage of the predicted price, the average error is 12.4%. As the dataset A has an average error of 25.6%, the dataset B is clearly more accurate.

Table 10: Predicted prices of dataset A

Name	Actual price	Predicted price	Error in %
Nieuwbouw 14 rijenwoning Groenoord	€ 2.178.000,00	€ 2.716.914,40	25%
Nieuwbouw 22 rijenwoning Dronten	€ 2.005.310,00	€ 2.650.061,01	32%
Nieuwbouw 31 rijenwoning	€ 3.892.552,00	€ 4.224.521,49	9%
28 appartementen De Haere	€ 3.380.015,00	€ 3.428.232,40	1%
Appartementencomplex de barchaan	€ 2.641.105,00	€ 2.020.371,10	-24%
44 apartementen de parkwachters	€ 5.665.615,00	€ 5.736.949,05	1%
29 appartementen boven winkels	€ 4.623.110,00	€ 3.331.576,18	-28%
Woontoren met 40 appartementen	€ 4.924.405,00	€ 5.179.953,88	5%
Allardhof te buren	€ 2.414.120,00	€ 2.904.671,53	20%
17 appartementen Oudenbosch	€ 1.622.132,71	€ 1.713.252,40	6%
49 seniorenappartementen Dronten	€ 4.386.824,70	€ 5.577.792,45	27%
15 stadswoningen Maastricht	€ 2.396.184,79	€ 2.944.541,78	23%
Appartementen Winterswijk	€ 4.228.964,40	€ 5.531.734,93	31%
A.ifc	€ 2.267.617,99	€ 2.146.674,34	-5%
B.ifc	€ 3.712.303,91	€ 2.703.719,54	-27%
C.ifc	€ 2.511.492,54	€ 2.771.172,88	10%
D.ifc	€ 2.620.614,11	€ 2.799.981,48	7%
E.ifc	€ 3.315.522,11	€ 2.736.052,49	-17%
F.ifc	€ 2.929.378,37	€ 2.062.372,84	-30%
G.ifc	€ 989.262,70	€ 1.721.203,55	74%
H.ifc	€ 540.782,18	€ 817.820,01	51%
I.ifc	€ 441.135,74	€ 839.490,85	90%
J.ifc	€ 1.888.289,88	€ 1.858.531,98	-2%
K.ifc	€ 3.560.412,73	€ 3.651.175,18	3%
L.ifc	€ 681.174,41	€ 1.302.045,02	91%

Table 11: Predicted prices of dataset B

Name	Actual price	Predicted price	Error in %
28 appartementen De Haere	€ 3.380.015,00	€ 2.933.956,56	-13%
Appartementencomplex de barchaan	€ 2.641.105,00	€ 2.029.967,85	-23%

44 apartementen de parkwachters	€ 5.665.615,00	€ 5.913.629,92	4%
29 appartementen boven winkels	€ 4.623.110,00	€ 4.006.705,04	-13%
Woontoren met 40 appartementen	€ 4.924.405,00	€ 4.873.810,76	-1%
Allardhof te buren	€ 2.414.120,00	€ 2.519.090,69	4%
17 appartementen Oudenbosch	€ 1.622.132,71	€ 2.189.812,24	35%
49 seniorenappartementen Dronten	€ 4.386.824,70	€ 4.473.685,05	2%
Appartementen Winterswijk	€ 4.228.964,40	€ 4.406.424,78	4%
B.ifc	€ 3.712.303,91	€ 3.338.541,54	-10%
C.ifc	€ 2.511.492,54	€ 2.774.966,32	10%
D.ifc	€ 2.620.614,11	€ 2.736.779,24	4%
E.ifc	€ 3.315.522,11	€ 3.076.425,48	-7%
J.ifc	€ 1.888.289,88	€ 2.660.717,79	41%

6.6. Conclusion data analytics

In the previous chapter already appeared that the quality of the BIM models is an issue for automated quantity take off with the purpose of data mining. In this chapter a new issue was found. After applying the data mining algorithms on the dataset, no usable relations between the data could be found.

The proposed reason for the issue was that the dataset was too small and that there was too much variation in the design of the buildings as both apartments and houses were used in the dataset. To test this assumption, addition data was collected from project analyses available in databases. These additional data have been used to create two new datasets. One dataset, where all the data of both the original data and the new data were combined and another dataset with only one type of buildings in the dataset, namely apartments.

These new datasets were tested by applying the previously used linear regression algorithm. From the results of the algorithm it could be concluded that the made assumptions were correct. The dataset with all the data showed that more data give a more consistent result, where the dataset with only apartments shows that the accuracy of the prediction are acceptable for conceptual cost estimation if only one type of buildings is used in the dataset.

Part I: Problem statement

Part II: Theoretical review

Part III: Practical implementation

Part IV: Conclusions and recommendations

This page is intentionally left blank

7. Discussion

The main objective of this research was to evaluate the effectiveness of a knowledge management system using BIM data and the data mining process. To realise this goal, two different approaches have been used, a literature study and a practical implementation. At first the literature study has been conducted. This study has shown the theoretical potential of such system. However during the practical implementation approach, it was found that the implementation of such system encountered some problems. To solve this problem additional data was collected, which resulted in a useful model for cost estimation.

Part of the scientific relevance of this research is that it has shown an alternative way to use building information modelling. In the recent years more attention has been paid towards using BIM as the knowledge source in knowledge management, as shown in the literature review. However little of these researches has resulted in practical implementation. Also this research has shown some of the implementation problems with the current state of BIM models. It has been made clear that there is a gap between the scientific potential of BIM models and the potential for implementation at this point. More attention is necessary toward improving the quality and correctness of data in BIM models if the AEC industry want to use the full potential of BIM.

The practical relevance of this research is that it has shown what improvements are needed for BIM to reach its potential for the application of data mining process. As will be discussed in the next chapters, it is necessary to focus on improving the way BIM models are shared and how they are modelled. Over the whole building process the BIM model should be kept up to date and shared in predefined format.

This page is intentionally left blank

8. Conclusions

In order to answer the research problem, five research questions were defined. In this chapter the answers to these research questions are given. After answering these research questions the research problem will be answered.

Research question 1: Which are the steps the data mining process consists of?

The data mining process can be divided in two parts and has a total of nine steps. The first part is the data engineering part. It starts by understanding and defining the problem. In this research the problem of the data mining process is to create knowledge to predict the building costs of residential buildings in the early phase of the design process. After the problem is made clear, the data which will be used in the data mining process need to be collected and extracted into a useable dataset. For this research BIM models with corresponding costs data were collected and the data of the BIM models were extracted to a dataset in form of a quantity take off. The next step is to clean and explore the data. As the quality of the data in this research appeared to be quite low, a lot of cleaning had to be done. Following the data exploration the dataset had to be converted into a subset which will perform best with the used algorithms. As the dataset available in this research was very small, it was chosen to test the performance of multiple subsets. The subsets were created based on a ranking method for feature selection. After the data have been engineered the data mining process continues to the next part, data analysing.

This part starts with the engineering of the algorithm to be used. Because the goal of this research is to create new knowledge for cost estimation a machine learning algorithm is needed. It was chosen to use both a regression algorithm as a classification algorithm. This was chosen because both have been successfully applied in the past to predict cost estimation for building costs. This process is continued with the running of the algorithm. In case of this research both the regression as the classification algorithm were run. The next step is defined as refining the data and problem. As the evaluation of the performance of the data mining algorithms was that no reliable relations were found within the dataset and it was concluded that the most probable cause of this was that the dataset was not large enough and the used building project were to different, additional data has been collected. This data has been used to create two new datasets that were applied in the data mining algorithm. One dataset which exists of all data both from the old dataset as from the newly gathered data, showed that adding more data to the dataset results in more consistent results. The other dataset also exist of both the old dataset and the newly acquired data but it only had the projects that were apartments to have more homogenous content in the dataset. The results of the application of the data mining algorithm on this dataset showed that useful relations can be found if the dataset contains data of building from a similar type. The last step of the data mining process is to use the results. To test the usefulness of the results, the results were used to estimate the costs of the buildings within the dataset. Although the results were close enough to use the generated cost estimation model to calculate a conceptual estimate for building project, the cost estimation model was only tested on the dataset that was used to create the model, so it should be noted that it is yet unclear how precise the model will be on unseen data.

Research question 2: Is the IFC data format a suitable BIM data format to extract a dataset for the data mining process?

In the literature review it was shown how Ma et al. (2013) used the IFC model to extract a quantity take off in the format of the Chinese mandatory specification of cost estimation in the tender phase. This suggested that the IFC data format was indeed a suitable BIM data format to extract a dataset for the data mining process. This was confirmed during the practical implementation. The part of the BIM models that were modelled correctly could be extracted from the model automatically using an extraction script. Due to limitations in the script the information embedded within the BIM model about the area of the different spaces within the building could not be extracted. However this is not a result of the IFC data format but a result of the applied script.

Research question 3: Is the current quality of the BIM models good enough to automatically extract a dataset for the data mining process?

A lot of effort has been made to create an extraction script that can automatically extract the dataset from BIM models. However to get the extractor script to extract an acceptable dataset a lot of code had to be written especially for one of the IFC models. In one of the IFC models three different IfcBuildingStorey elements were created for every floor level. This makes the automatic interpretation of floor levels not possible without writing a code for the script that can interpret multiple IfcBuildingStorey elements for one floor level. Because of the similarity between the height of the different IfcBuildingStorey elements, it was not possible to analyse and merge the levels by height. As a solution, some code has been written that allows the script to interpret the IfcBuildingStoreys for the IFC model that had this problem, but for other IFC models with a similar problem the script will not be able to interpret this correctly without changing the code.

As already concluded in the previous part of the research, another problem with the IFC quality is the export of the BIM authoring tool model into the IFC model. All but one of the BIM models were modelled with Revit, the other model with Allplan. Of all those models four contained the necessary information about the organisation that exported the BIM file. All of them were the contractor of the building projects. Another five of them were exported by the author of the presented research.

One of the mentioned problem with the IFC models was that three different IfcElements were found to represent a stair, namely: IfcStair, IfcFurnishingElement, IfcBuildingElementProxy. This is a problem because as the name IfcFurnishingElement suggests, this type of IfcElement is a generalization of all furniture related objects (buildingSMART, nd). Another problem found was that more than one level of height per building storey was used in the BIM model. This is not necessarily a problem if only one of the levels of height was modelled as the main level and the other levels of the storey as a sub level. However this was not the case and therefore the BIM model was exported to the IFC data format with too many building storeys. All these problems could be solved if all the organisations within the building project that model within the BIM model would work according to the BIM Basis ILS.

The loss of necessary information is also a problem. In good practise some building elements like walls are modelled as one building element composed of different construction products (see figure 11). As this is a relatively easy way of modelling and it even saves time compared to modelling all the different construction parts separately, it is plausible that the building

elements were modelled as one element although they are composed of different construction products. In the IFC models however, all these different construction products were already decomposed to standard IfcElements. This makes the interpretation of these elements much harder. In most exporters provided by the BIM authoring tools, this problem can be solved by changing the setting of the exporter.

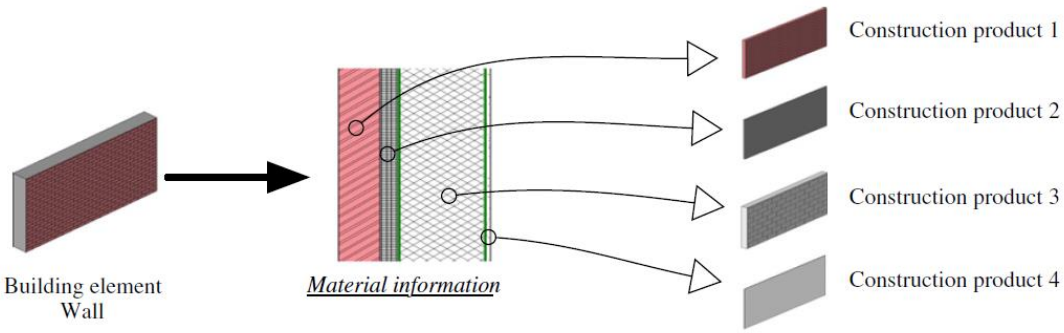

Figure 10: Building elements composing of different construction products (source: Ma et al. (2013)

Research question 4: Can the data mining process be used to predict the costs of residential buildings in the early stages of the design process?

The research question: *Can the data mining process be used to predict the costs of residential buildings in the early stages of the design process?* has been answered both via literature review as via practical implementation. However the answers differ. Based on the literature it can be concluded that it is indeed possible to use the data mining process to predict the costs of residential buildings in the early stage of the design process. The BIM model can contain enough information about the buildings. With an extractor this knowledge embedded in the model can be made explicit and with a machine learning algorithm this knowledge can be used to create new information that allows us to predict the costs of buildings.

In the practical implementation part of this research however it became clear that there are two implementation problems. First of all the quality of the BIM models does not meet the same level as is necessary for automatic extraction of the required dataset, as discussed in the answering of the previous research question. Another problem is the amount of BIM models available. The data mining process needs data to analyse and generally the more data it can analyse the better the data mining process is able to find patterns within the data. It has been shown that 12 BIM models do not provide enough data to find these patterns. So, even when the AEC industry would create BIM models of the quality necessary for data mining, also a lot of models are needed of a specific building type with the above mentioned quality.

In this research it was chosen to add additional data from public available project analyses to create a larger dataset. This data proved to be very useful and it resulted in a cost estimation model that was precisely enough to calculate the conceptual building costs of the buildings within the dataset.

Main research question: “Can the use of a knowledge management system using BIM data and the data mining process create a more efficient process of cost estimation for residential buildings?”

The literature review has shown that the data mining process is an effective tool to create a knowledge management system. It is especially useful because it can be made into an automated computer process, where little maintenance is required. The literature review has also shown that BIM data have a great potential to serve as the data source. These are created in almost every project and they contain all the physical information of the buildings. These data can be converted to the dataset necessary for the data mining process by creating a quantity take off. In previous researches could be found that automated quantity take off from BIM models is possible. From this all it can be concluded that a data mining process can efficiently estimate the costs of residential buildings and therefore used as a knowledge management system for cost estimation.

However during the practical implementation it became clear that there are both qualitative and quantitative problems with the available BIM models. As mentioned above additional data were added to test viability of the data mining process for cost calculation for residential buildings. With this additional data the data mining process could predict the costs of buildings within the dataset precisely enough to be a conceptual cost estimate. Nonetheless until the qualitative and quantitative problems of BIM data are solved the knowledge management system will still need the manual input of additional data to be able to estimate the costs of residential buildings.

9. Recommendations

In this chapter the recommendations are given. First the recommendations for implementation are discussed. Thereafter the recommendations for further research are given.

9.1. Recommendations for implementation

9.1.1. Implementation of the IFD format

It has been concluded that the biggest obstacle to overcome before the proposed knowledge management system can be implemented is the quality of the BIM models.

Part of the problem is that for the IFC format in order to be an effective format to export data to, it allows various data to be shared in various ways. If the receiver of the information, in this research the extractor, wants to be sure to be able to utilize the information he receives, the sender and receiver need to be in agreement which and how the information is shared (Bell & Bjørkhaug, 2006). For this purpose buildingSMART already created the International Framework of Dictionaries (IFD). The Dutch implementation of this is the Nederlandse Conceptenbibliotheek voor de bouw (CB-NL). This can be used to standardise the way information within IFC is shared, and thus allows to use the proposed agreements to create IFC models that allow for automated data extraction. So part of the solution for this problem already exists, however the solution was not used in the development of the BIM models that were available in this research. Consequentially it is recommended that the IFD is implemented within all organisations within a building project. To facilitate this, the government should start demanding that IFD is used like the Dutch government has done with IFC in their DBFMO contracts.

9.1.2. Implementation of the Information Delivery Manual(IDM)

Another way to ensure better quality of BIM models and more specifically IFC models is to use Information Delivery Manuals (IDM). IDM can be used for the documentation of existing or new processes and describes the associated information that needs to be exchanged between parties. The main purpose of IDM is to make sure the relevant data are communicated via a structured way such as the data can be interpreted by the software on the receiving side (buildingsmart, nd.). For the Dutch AEC industry the 'BIM basis informatieleveringsspecificatie (ILS)' is defined and it has already been adopted by many companies. As has been shown during the practical application most of the problems of the quality of the IFC models would have been prevented if the BIM Basis ILS would have been used.

9.1.3. BIM as-built models

Even if the IFD format or another ontology and the IDM would systematically be applied not all quality problems of the BIM models would be solved. The IFD format would not prevent that the BIM models are not modelled as built. In order to accomplish this the BIM model should be updated consequently during the whole building process. This can only be ensured if all companies involved would have an incentive to update the BIM models. As the information that can be extracted from the BIM models to a dataset is limited to the information located within the BIM models, it is very important for the application of data mining on BIM models that all the information is modelled.

9.1.4. Use of additional data

Next to the quality of the BIM models also the number of BIM models available is a problem for the proposed knowledge management system. To overcome this problem in this research additional data has been added to the dataset. If one wants to implement the proposed knowledge management system within the organisation the dataset can be manually enlarged with data from projects of which no viable BIM model is available. Although it is desirable to have little to no human effort needed to maintain the KMS it allows the KMS to be implemented. With the current trends of BIM modelling a lot more BIM data will come available in the upcoming years, and thus the BIM data can eventually replace the use of additional data.

9.2. Recommendation for further research

9.2.1. Quality of BIM models

In the recommendations for implementation it is suggested that the quality of the BIM models can be improved by implementing an ontology. However the focus of this research was aimed at the implementation of BIM models in a knowledge management system. More research is needed to provide more ways to improve the quality of BIM models.

9.2.2. Other goal of KMS

This research has focussed on the implementation of a KMS for cost prediction. However other goals for a KMS system based on BIM data and the data mining process are possible. For example a goal for a KMS system could be to predict the energy consumption of buildings.

References

- Aalst, W. M. P. van der, Stoffele, M., & Wamelink, J. W. F. (2003). Case handling in construction. *Automation in Construction*, *12*(3), 303-320.
- Abeljaber, R., Ioannou, C., Maor, D., Razo, R., & Tribolet, J. (1998) *Knowledge management of corporate intranets.* Retrieved on 17 october 2016 from: http://web.mit.edu/ecom/www/Project98/G4/
- Alavi, M., & Leidner, D. E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS quarterly*, *25*(1), 107-136.
- Al-Gahtani, K., & Ghani, S. R. (2010, March). Knowledge management tools: adaption in construction management practices. In *Engineering Systems Management and Its Applications (ICESMA), 2010 Second International Conference on* (pp. 1-6). IEEE.
- Almana, A. M., & Aksoy, M. S. (2014). An Overview of Inductive Learning Algorithms. *International Journal of Computer Applications*, 88(4), 975–8887.
- Argote, L. & Ingram P. (2000). Knowledge Transfer: A Basis for Competitive Advantage in Firms. *Organizational Behavior and Human Decision Processes*, 82(1), 150–169.
- Barney, B. (2016). *Introduction to Parallel Computing.* Retrieved on 17 march 2017 from: https://computing.llnl.gov/tutorials/parallel_comp/
- Bell, H., & Bjørkhaug, L. (2006, August). A buildingSMART ontology. In *Proceedings of the 2006 ECPPM Conference* (pp. 185-190). sn.
- Bhargav, D., Koskela, L., & Dave, B. (2009). Collaborative Knowledge Management – A Construction Case Study. *Automation in Construction*, 18(7), 894–902. Retrieved from http://eprints.hud.ac.uk/25837/
- Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. Owolabi, H., Alaka, H and Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, 30 (2016). pp. 500-521.
- Böhms, M., Mommers, B., Baggen, M., Bakker, J., Manoku, E., Nierop, M. van, Nagtegaal, R., Beetz, J., Brink, L. van den & Oldfield, J. (2013). *Nederlandse Conceptenbibliotheek voor de bouw (CB-NL) – Modeling Guide.* Amersfoort: Geonovum
- Botter, R., (2013) *Efficiënt begroten met historische projectgegevens.* Eindhoven: TU Eindhoven. Fac. Bouwkunde
- Brown, A., & Adams, J. (2000). Measuring the effect of project management on construction outputs: a new approach. *International Journal of Project Management*, 18, 327-335.
- BuildingSMART (n.d.). *Ifc Overview*. Retrieved 2016, December 12, from http://www.buildingsmart-tech.org/specifications/ifc-overview
- BuildingSMART (n.d.). *IfcObject*. Retrieved 2016, December 12, from http://www.buildingsmart-tech.org/ifc/IFC2x4/rc2/html/schema/ifckernel/lexical/ifcobject.htm
- BuildingSMART (n.d.). *IfcProduct.* Retrieved 2016, December 12, from http://www.buildingsmart-tech.org/ifc/IFC2x3/TC1/html/ifckernel/lexical/ifcproduct.htm

- BuildingSMART (n.d.). *IfceElement.* Retrieved 2016, December 12, from http://www.buildingsmart-tech.org/ifc/IFC2x3/TC1/html/ifcproductextension/lexical/ifcelement.htm
- BuildingSMART (n.d.). *IfcBuildingElement*. Retrieved 2016, December 12, from http://www.buildingsmart-tech.org/ifc/IFC2x3/TC1/html/ifcproductextension/lexical/ifcbuildingelement.htm
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). *Power failure: why small sample size undermines the reliability of neuroscience*. Nature Reviews Neuroscience, 14(5), 365-376.
- Calkins, D. E., Egging, N., Scholz, C., & Gomez-Levi, G. (2000). *AutoDSS: A System Level KBE Tool for Vehicle Product Definition (No. 2000-01-1351)*. SAE Technical Paper.
- Campbell, D.A. (2006, 11 October) Modeling Rules, Architecture week. Retrieved on 17 march 2017 from: http://www.architectureweek.com/2006/1011/tools_1-1.html
- Charlesraj, V. P. C. (2014, January). Knowledge-based building information modeling (K-BIM) for facilities management. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction* (Vol. 31, p. 1). Vilnius Gediminas Technical University, Department of Construction Economics & Property.
- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, *19*(2), 171-209.
- Cheung, F. K., Rihan, J., Tah, J., Duce, D., & Kurul, E. (2012). Early stage multi-level cost estimation for schematic BIM models. *Automation in Construction*, *27*, 67-77.
- Clayton, M. J., Ellis, C. D., Woo, J., Clayton, M. J., Johnson, R. E., Flores, B. E., & Ellis, C. (2004). Dynamic Knowledge Map : Reusing experts ' tacit knowledge in the AEC industry. *Automation in Construction 13(2), 203-207.*
- Coakes, E. W., Coakes, J. M., & Rosenberg, D. (2008). Co-operative work practices and knowledge sharing issues: A comparison of viewpoints. *International Journal of Information Management*, *28*(1), 12-25.
- Das, A., & Kempe, D. (2008, May). Algorithms for subset selection in linear regression. In *Proceedings of the fortieth annual ACM symposium on Theory of computing* (pp. 45-54). ACM.
- Dave, B., & Koskela, L. (2009). Collaborative knowledge management - A construction case study. *Automation in construction*, *18*(7), 894-902.
- Davenport, T. H., De Long, D. W., & Beers, M. C. (1997). Building successful knowledge management projects. *Center for business innovation working paper*, *4*.
- Davtalaba, O., & Delgadob, J. L. (2014). Benefits of 6D BIM for Facilities Management Departments for Construction Projects–A Case Study Approach. *Southern California, University of Southern California*
- Deshpande, A., Azhar, S., & Amireddy, S. (2014). A framework for a BIM-based knowledge management system. *Procedia Engineering*, *85*, 113-122.
- Deveraja, H. (2015). *Perspective: Knowledge based Engineering (KBE)*. Bangalore: Infosys. Retrieved from Infosys.
- Ding, L. Y., Zhong, B. T., Wu, S., & Luo, H. B. (2016). Construction risk knowledge management in BIM using ontology and semantic web technology. *Safety science*, *87*, 202-213.
- Dittrich, J., & Quiané-Ruiz, J. A. (2012). Efficient big data processing in Hadoop MapReduce. *Proceedings of the VLDB Endowment*, *5*(12), 2014-2015.

- Eastman, C. M., Teicholz, P., Sacks, R., & Liston, K. (2011). *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors*. John Wiley & Sons.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), 37.
- Grover, R., & Froese, T. M. (2016). Knowledge Management in Construction Using a SocioBIM Platform: A Case Study of AYO Smart Home Project. *Procedia Engineering*, *145*, 1283-1290.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, *3*(Mar), 1157-1182.
- Hämäläinen, W. (2006). *Descriptive and predictive modelling techniques for educational technology.* Licentiate thesis, Department of Computer Science, Joensuu, Finland: University of Joensuu.
- Ho, S. P., Tserng, H. P., & Jan, S. H. (2013). Enhancing knowledge sharing management using BIM technology in construction. *The Scientific World Journal*, 2013. http://dx.doi.org/10.1155/2013/170498
- Hoogeveen, N. (2015). *Building Knowledge Modelling, from BIM data to Artificial Intelligent Knowledge Systems.* Delft: TU Delft Faculty Civil Engineering and Geosciences.
- Howitt, D., & Cramer, D. (2005). *Introduction to statistics in psychology.* Essex, England: Pearson Education.
- Huang, M. J., Chen, M. Y., & Lee, S. C. (2007). Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis. *Expert systems with applications*, *32*(3), 856-867.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, *22*(4), 679-688.
- IBM (n.d.). *The Four V's of Big Data.* Retrieved 2016, October 18, from http://www.ibmbigdatahub.com/infographic/four-vs-big-data
- IBM Knowledge Centre (2017). *What is distributed computing.* Retrieved on 17 march 2017 from: http://www.ibm.com/support/knowledgecenter/SSAL2T_7.1.0/com.ibm.cics.tx.doc/concepts/c_wht_is_distd_comptg.html
- International Organization for Standardization. (2007). *Building construction -- Organization of information about construction works -- Part 3: Framework for object-oriented information* (ISO Standard No. 12006-3:2007). Retrieved from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=38706
- Jin, R., Cho, K., Hyun, C., & Son, M. (2012). MRA-based revised CBR model for cost prediction in the early stage of construction projects. *Expert Systems with Applications*, *39*(5), 5214-5222.
- Johannessen, J. A., Olaisen, J., & Olsen, B. (2001). Mismanagement of tacit knowledge: the importance of tacit knowledge, the danger of information technology, and what to do about it. *International journal of information management*, *21*(1), 3-20.
- John, G. H. (1997). *Enhancements to the data mining process* (Doctoral dissertation, Stanford USA: Stanford university).

- Kamsu-Foguem, B., & Abanda, F. H. (2015). Experience modeling with graphs encoded knowledge for construction industry. *Computers in Industry*, *70*, 79-88.
- Kanapeckiene, L., Kaklauskas, A., Zavadskas, E. K., & Seniut, M. (2010). *Integrated knowledge management model and system for construction projects. Engineering Applications of Artificial Intelligence*. https://doi.org/10.1016/j.engappai.2010.01.030
- Khosrowshahi, F. (2011). Innovation in artificial neural network learning: Learn-On-Demand methodology. *Automation in Construction*, *20*(8), 1204-1210.
- Kock, N. F., McQueen, R. J., & Corner, J. L. (1997). The nature of data, information and knowledge exchanges in business processes: implications for process improvement and organizational learning. *The Learning Organization*, *4*(2), 70-80.
- Konukcu, S., & Koseoglu, O. (2012, October). *Knowledge management through BIM in construction supply chains.* In The 29th International Conference–Beirut, Lebanon (pp. 17-19).
  Koskela, L. (2000). *An exploration towards a production theory and its application to construction*. VTT Technical Research Centre of Finland. Helsinki: University of Technology (Espoo, Finland)
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variaty. *Application Delivery Strategies.* Stamford: META Group Inc.
- Li, S., Harner, E. J., & Adjeroh, D. A. (2011). Random KNN feature selection-a fast and stable alternative to Random Forests. *BMC bioinformatics*, *12*(1), 450.
- Lin, Y. C., Wang, L. C., & Tserng, H. P. (2006). Enhancing knowledge exchange through web map-based knowledge management system in construction: Lessons learned in Taiwan. *Automation in Construction*, *15*(6), 693-705.
- Lin, Y. C. (2014). Construction 3D BIM-based knowledge management system: a case study. *Journal of Civil Engineering and Management*, *20*(2), 186-200.
- Linoff, G. S., & Berry, M. J. (2011). *Data mining techniques: for marketing, sales, and customer relationship management*. Third edition, Indianapolis: John Wiley & Sons.
- Ma, Z., Wei, Z., & Zhang, X. (2013). Semi-automatic and specification-compliant cost estimation for tendering of building projects based on IFC data of design model. *Automation in Construction*, *30*, 126-135.
- Mathew, P. A., Dunn, L. N., Sohn, M. D., Mercado, A., Custudio, C., & Walter, T. (2015). Big-data for building energy performance: Lessons from assembling a very large national database of building energy use. *Applied Energy*, 140, 85-93.
- Meadati, P., & Irizarry, J. (2010, April). BIM–a knowledge repository. In *Proceedings of the 46th Annual International Conference of the Associated Schools of Construction, Retrieved November* (Vol. 12, p. 2010).
- Mi, C., Qian, H., Liu, S., & Chang, Z. (2008, June). Study on case retrieving in case-based reasoning based on grey incidence theory and its application in bank regulation. In *Fuzzy Systems, 2008. FUZZ-IEEE 2008.(IEEE World Congress on Computational Intelligence). IEEE International Conference on* (pp. 1530-1533). IEEE.
- Motawa, I., & Almarshad, A. (2013). A knowledge-based BIM system for building maintenance. *Automation in Construction*, *29*, 173-182.
- National BIM Standard (NBIMS) (n.d.). *What is BIM?* Retrieved on 19 march 2017 from: https://www.nationalbimstandard.org/faqs#faq1
- Nonaka, I. (1991) *The Knowledge creating Company Harvard Business Review, **69** (6 Nov-Dec): 96–104*

- Nonaka, I., & Lewin, A. Y. (1994). A Dynamic Theory of Organizational Knowledge Creation, 5(1), 14–37. Retrieved from http://www.jstor.org/stable/2635068
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation.* Oxford: Oxford university press.
- Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, Ba and leadership: a unified model of dynamic knowledge creation. *Long range planning*, *33*(1), 5-34.
- Phaobunjong, K. (2002). *Parametric cost estimating model for conceptual cost estimating of building construction projects* (Doctoral dissertation). Austin USA: the University of Texas
- Pozzali, A., & Viale, R. (2015). "Cognition, Types of "Tacit Knowledge" and Technology Transfer" *In Cognitive Economics: New Trends*. Published online: 08 Mar 2015; 205-224.
- Quinlan, J. R. (1987). Simplifying decision trees. *International journal of man-machine studies*, *27*(3), 221-234.
- Redmond, A., Hore, A., Alshawi, M., & West, R. (2012). Exploring how information exchanges can be enhanced through Cloud BIM. *Automation in Construction*, *24*, 175-183.
- Rezgui, Y., Hopfe, C. J., & Vorakulpipat, C. (2010). Generations of knowledge management in the architecture, engineering and construction industry: An evolutionary perspective. *Advanced Engineering Informatics*, *24*(2), 219-228.
- Sayad, S., (2010). *K Nearest Neighbors*. Retrieved from: http://chem-eng.utoronto.ca/~datamining/Presentations/KNN.pdf
- Seni, G., & Elder, J. F. (2010). Ensemble methods in data mining: improving accuracy through combining predictions. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 2(1), 1-126.
- Shen, Z., & Issa, R. R. (2010). Quantitative evaluation of the BIM-assisted construction detailed cost estimates. *Journal of Information Technology in Construction* ITcon), Vol. 15, pg. 234-257.
- Smith, E. A. (2001). The role of tacit and explicit knowledge in the workplace. *Journal of knowledge Management*, *5*(4), 311-321.
- Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, *22*(1), 97-103.
- Stewart, R. A., & Spencer, C. A. (2006). Six-sigma as a strategy for process improvement on construction projects: a case study. *Construction Management and Economics*, *24*(4), 339-348.
- Sweet, S., & Pitt-Catsouphes, M. (2010). Talent Pressures and the Aging Workforce. Retrieved on March 16, 2017 from: http://www.bc.edu/content/dam/files/research_sites/agingandwork/pdf/publications/TMISR05_Construction.pdf
- The R Foundation (n.d.). *What is R, Introduction to R.* Retrieved on 17 march 2017 from: https://www.r-project.org/about.html
- Tibaut, A., & Jakoša, D. (n.d.). Development of Knowledge Model for Construction Projects. https://doi.org/10.1007/978-3-319-21009-4_19
- Vonk, A.S., Wilde, J.J. de & Groot, T.C. de (2016). Indexcijfers Woningbouw, 2de kwartaal 2016. Den Haag. Bouwkostenindex

- Vukovic, S., Delibasic, B., Uzelac, A., & Suknovic, M. (2012). A case-based reasoning model that uses preference theory functions for credit scoring. *Expert Systems with Applications*, *39*(9), 8389-8395.
- Walker, D. H., & Walker, D. H. (2016). Reflecting on 10 years of focus on innovation, organisational learning and knowledge management literature in a construction project management context. *Construction Innovation*, *16*(2), 114-126. No Title. (1999), (June), 1–81.
- Ward, J. S., & Barker, A. (2013). Undefined by data: a survey of big data definitions. *arXiv preprint arXiv:1309.5821*. Fife, Scotland. University of St. Andrews
- Weippert, A., & Kajewski, S. L. (2007). Value-adding Knowledge Management (KM) Framework: 'An AEC industry perspective'. *CIB W065/055 Commissions: Transformation through Construction*, 1-11.
- Williams, G., (2010) *Data mining desktop survival guide.* Retrieved 2017, januari 8, from: http://datamining.togaware.com/survivor/dmsurvivor.html
- Witten, I. H., & Frank, E. (2000). *Practical Machine Learning Tools and Techniques with Java implementations, Data Mining*. San Francisco, CA: Morgan Kaufmann Publishers
- Woo, J. H., Clayton, M. J., Johnson, R. E., Flores, B. E., & Ellis, C. (2004). Dynamic Knowledge Map: reusing experts' tacit knowledge in the AEC industry. *Automation in construction*, *13*(2), 203-207.
- Yew Wong, K. (2005). Critical success factors for implementing knowledge management in small and medium enterprises. *Industrial Management & Data Systems*, *105*(3), 261-279.

Appendix

Appendix 1: Structure of the IFC data format

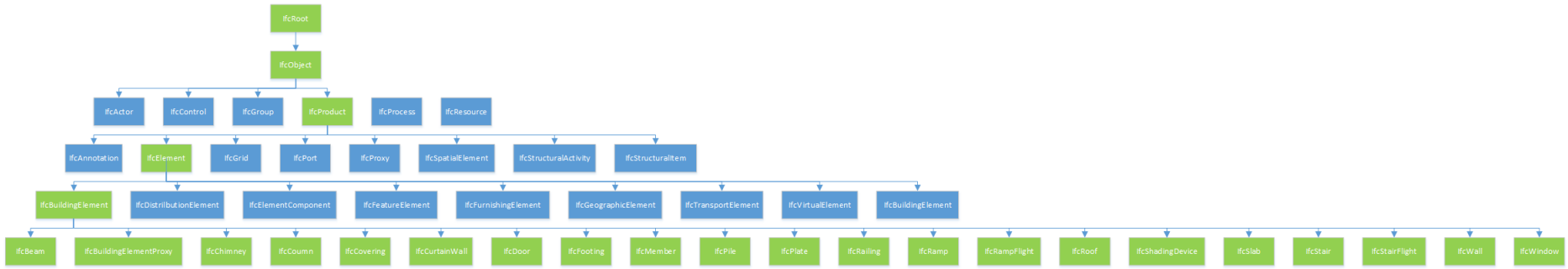

Figure 11: Structure of the IFC data format

Appendix 2: The extractor script

Start script

0_start_script_new.py

```
import glob
import subprocess
import sys
import os
import os.path
import csv
import datetime

# Get directory of script_dir
script_dir: (os.path.dirname(os.path.realpath(__file__)))

# Creating file to write
timestamp_date: datetime.datetime.now().strftime("%Y%m%d")
timestamp_time: datetime.datetime.now().strftime("%H%M%S")

save_path: script_dir+"/"+timestamp_date+"/"
if not os.path.exists(save_path):
    os.makedirs(save_path)

dataset_csv: os.path.join(save_path, timestamp_time+"_dataset.csv")

with open(dataset_csv, "wb") as f_csv:
    csvWriter: csv.writer(f_csv)
    csvWriter.writerow(["Filename", "areaGroundFloor",
"areaUpperFloor", "areaRoof", "countStair", "areaExternalWalls",
"areaExternalOpenings", "areaInternalWalls", "countInternalDoors"])

for fn in glob.glob("*.ifc"):
    # if fn not in processed_files:
    print fn
    subprocess.call([sys.executable, "1_QTO_new.py", fn,
dataset_csv])
    print "File processed"
```

Actual extractor

1_QTO_new.py

```
import os
import os.path
import glob
import sys
import math

from operator import itemgetter

import OCC.gp
import OCC.Geom
import OCC.Bnd
import OCC.BRepBndLib
import OCC.BRep
import OCC.BRepPrimAPI
import OCC.BRepAlgoAPI
import OCC.BRepBuilderAPI
```

```
17   import OCC.BRepAlgo
18   import OCC.TopOpeBRepTool
19   import OCC.ShapeExtend
20   import OCC.GProp
21   import OCC.BRepGProp
22   import OCC.GC
23   import OCC.ShapeAnalysis
24   import OCC.TopTools
25   import OCCUtils
26
27   from OCC.TopoDS import topods
28
29   import ifcopenshell
30   import ifcopenshell.geom
31
32   import operator
33
34   import csv
35
36   import re
37
38   # Specify to return pythonOCC shapes from
ifcopenshell.geom.create_shape()
39   settings: ifcopenshell.geom.settings()
40   settings.set(settings.USE_PYTHON_OPENCASCADE, True)
41
42
43   # Some helper functions to map to the list of walls
44   def create_shape(elem):
45       return ifcopenshell.geom.create_shape(settings, elem)
46
47
48   def calc_volume(s):
49       props: OCC.GProp.GProp_GProps()
50       OCC.BRepGProp.brepgprop_VolumeProperties(s.geometry, props)
51       return props.Mass()
52
53
54   def calc_area(s):
55       props: OCC.GProp.GProp_GProps()
56       OCC.BRepGProp.brepgprop_SurfaceProperties(s.geometry, props)
57       return props.Mass()
58
59
60   def calc_length(LProps):
61       props: OCC.GProp.GProp_GProps()
62       OCC.BRepGProp.brepgprop.LinearProperties(LProps.geometry, props)
63       return props.Mass()
64
65
66   def count_windows(product):
67       if product.is_a("IfcWindow"):
68           global ifcwindow
69           ifcwindow += 1
70
71
72   def count_doors(product):
73       if product.is_a("IfcDoor"):
74           global ifcdoor
75           ifcdoor += 1
76
```

```
77
78   def count_stories(product):
79       if product.is_a("IfcBuildingStorey"):
80           global ifcstorey
81           ifcstorey += 1
82
83
84   def create_levels(product):
85       if product.is_a("IfcBuildingStorey"):
86           level: product.Elevation
87           if level is not None:
88               floor_levels.append((level))
89           else:
90               level: 0
91               floor_levels.append((level))
92
93
94   def get_bounding_box_minimal(bbox):
95       bbmin: [0.] * 3;
96       bbmax: [0.] * 3
97       bbmin[0], bbmin[1], bbmin[2], bbmax[0], bbmax[1], bbmax[2]:
bbox.Get()
98       return OCC.gp.gp_Pnt(*map(lambda xy: (xy[0]), zip(bbmin,
bbmax)))
99
100
101  def get_bounding_box_maximal(bbox):
102      bbmin: [0.] * 3;
103      bbmax: [0.] * 3
104      bbmin[0], bbmin[1], bbmin[2], bbmax[0], bbmax[1], bbmax[2]:
bbox.Get()
105      return OCC.gp.gp_Pnt(*map(lambda xy: (xy[1]), zip(bbmin,
bbmax)))
106
107
108  def get_zvalue(product):
109      if product.Representation is not None:
110          shape: ifcopenshell.geom.create_shape(settings,
product).geometry
111          bbox: OCC.Bnd.Bnd_Box()
112          OCC.BRepBndLib.brepbndlib_Add(shape, bbox)
113          bounding_box_minimal: get_bounding_box_minimal(bbox)
114          bounding_box_maximal: get_bounding_box_maximal(bbox)
115          return (bounding_box_minimal.Z(), bounding_box_maximal.Z())
116
117  def get_xvalue(product):
118      if product.Representation is not None:
119          shape: ifcopenshell.geom.create_shape(settings,
product).geometry
120          bbox: OCC.Bnd.Bnd_Box()
121          OCC.BRepBndLib.brepbndlib_Add(shape, bbox)
122          bounding_box_minimal: get_bounding_box_minimal(bbox)
123          bounding_box_maximal: get_bounding_box_maximal(bbox)
124          return (bounding_box_minimal.X(), bounding_box_maximal.X())
125
126  def get_yvalue(product):
127      if product.Representation is not None:
128          shape: ifcopenshell.geom.create_shape(settings,
product).geometry
129          bbox: OCC.Bnd.Bnd_Box()
130          OCC.BRepBndLib.brepbndlib_Add(shape, bbox)
```

```
131             bounding_box_minimal: get_bounding_box_minimal(bbox)
132             bounding_box_maximal: get_bounding_box_maximal(bbox)
133             return (bounding_box_minimal.Y(), bounding_box_maximal.Y())
134
135     def get_values(product):
136         if product.Representation is not None:
137             shape: ifcopenshell.geom.create_shape(settings,
product).geometry
138             bbox: OCC.Bnd.Bnd_Box()
139             OCC.BRepBndLib.brepbndlib_Add(shape, bbox)
140             bounding_box_minimal: get_bounding_box_minimal(bbox)
141             bounding_box_maximal: get_bounding_box_maximal(bbox)
142             return (bounding_box_minimal.X(), bounding_box_maximal.X(),
bounding_box_minimal.Y(), bounding_box_maximal.Y(),
bounding_box_minimal.Z(), bounding_box_maximal.Z())
143
144     # creating temporary file
145     script_dir: (os.path.dirname(os.path.realpath(__file__)))
146     save_path_temp: script_dir + "/temp/"
147     if not os.path.exists(save_path_temp):
148         os.makedirs(save_path_temp)
149
150     # Open the IFC file using IfcOpenShell
151     ifc_file: ifcopenshell.open(sys.argv[1])
152     name_file: str(ifc_file)
153
154
155     print "Calculating Height of Elements"
156
157     products: ifc_file.by_type("IfcBuildingStorey")
158
159     levels: []
160
161     for product in products:
162         product_str: str(product)
163         product_string: product_str.lower()
164         count: product_string.count(",")
165         abc: product_string.split(",")
166         tempH: abc[-1]
167         if count > 10 :
168             placement: abc[6]
169         else:
170             placement: abc[5]
171         name: abc[2]
172         tempSplit: tempH.split(")")
173         h: tempSplit[0]
174         h1: float(h)
175         height: h1 / 1000
176         if height > -0.5:
177             level: [placement, height, name]
178             levels.append(level)
179
180     levels.sort(key=lambda x: x[1])
181
182     products: ifc_file.by_type("IfcLocalPlacement")
183
184     localPlacement: []
185
186     for product in products:
187
188         product_str: str(product)
```

```
189         product_string: product_str.lower()
190         abc: product_string.split(",")
191         temp1: abc[0]
192         temp1Split: temp1.split("(")
193         temp: temp1Split[0]
194         temp0: temp.split("=")
195         temp1: temp0[0]
196         temp2: temp1Split[1]
197
198         temp: [temp1, temp2]
199
200
201         localPlacement.append(temp)
202
203     for index, level in enumerate(levels):
204         test1: level[0]
205         for item in localPlacement:
206             test2: item[1]
207             if test2:= test1:
208                 tempList: []
209                 test3: item[0]
210                 tempList.append(test3)
211                 if any(option in tempList for option in level): continue
212                 level: level + tempList
213                 levels[index]: level
214
215     products: ifc_file.by_type("IfcWall")
216     WallProductsPerLevel: []
217     for level in levels:
218         list: []
219         WallProductsPerLevel.append(list)
220
221     for product in products:
222         product_str: str(product)
223         product_string: product_str.lower()
224         abc: product_string.split(",")
225         temp: abc[-3]
226         for index, level in enumerate(levels):
227             if temp in level:
228                 WallProductsPerLevel[index].append(product)
229
230     products: ifc_file.by_type("IfcSlab")
231     SlabProductsPerLevel: []
232     SlabAreaPerLevel: []
233     for level in levels:
234         list: []
235         SlabProductsPerLevel.append(list)
236
237     for product in products:
238         product_str: str(product)
239         product_string: product_str.lower()
240         abc: product_string.split(",")
241         temp: abc[-4]
242         for index, level in enumerate(levels):
243             if temp in level:
244                 SlabProductsPerLevel[index].append(product)
245
246     print "Calculating IfcSlabs"
247
248     for index, list in enumerate(SlabProductsPerLevel):
249         products: list
```

```
250
251         # vocabulary IfcSlab
252         wideSlabFloor_options: ["gewapend beton"]
253         hollowCoreSlab_options: ["pbh prefabvloer", "kanaalplaat",
"leidingplaat"]
254         groundFloor_options: ["begane grondvloer", "gewapende betonvloer
berging"]
255         ribbedSlab_options: ["ribcassette bg-vloer", "tempex bg-vloer"]
256         prefabSlab_options: ["balkon", "gallerijplaat"]
257         creeds_options: ["afwerkvloer", "cementdekvloer"]
258         insulationSlab_options: ["isolatie"]
259         roof_options: ["dak"]
260         slab_ignore_options: ["tegelvloer", "vergaarbak", "gras",
"asfalt", "4e verdiepingsvloer"]
261
262         # attributes IfcSlab
263         wideSlabFloor_area: 0
264         hollowCoreSlab_area: 0
265         groundFloor_area: 0
266         ribbedSlab_area: 0
267         prefabSlab_area: 0
268         creeds_area: 0
269         insulationSlab_area: 0
270         roof_area: 0
271         undefinedSlab_area: 0
272         VolumeSlabs: 0
273
274         z_levels: []
275
276         area_groundFloor: 0
277         area_RoofFlat: 0
278         areaIgnore_options: ["isolatie"]
279
280
281         for product in products:
282             z: get_zvalue(product)
283             if z is not None:
284                 z0: z[0]
285                 z1: z[1]
286                 thickness: z1 - z0
287                 if thickness < 0.15: continue
288                 # Create geometry for these windows
289                 shape: create_shape(product)
290
291                 # Calculate their volumes
292                 volume: calc_volume(shape)
293                 abs_volume: abs(volume)
294                 VolumeSlabs += abs_volume
295
296                 area: abs_volume / thickness
297
298                 product_str: str(product)
299                 product_string: product_str.lower()
300                 if any(option in product_string for option in
wideSlabFloor_options):
301                     wideSlabFloor area += area
302                 elif any(option in product_string for option in
hollowCoreSlab_options):
303                     hollowCoreSlab_area += area
304                 elif any(option in product_string for option in
groundFloor_options):
```

```
                    groundFloor_area += area
                elif any(option in product_string for option in ribbedSlab_options):
                    ribbedSlab_area += area
                elif any(option in product_string for option in prefabSlab_options):
                    prefabSlab_area += area
                elif any(option in product_string for option in creeds_options):
                    creeds_area += area
                elif any(option in product_string for option in insulationSlab_options):
                    insulationSlab_area += area
                elif any(option in product_string for option in roof_options):
                    roof_area += area
                elif any(option in product_string for option in slab_ignore_options):
                    pass
                else:
                    undefinedSlab_area += area

        totalarea: wideSlabFloor_area + hollowCoreSlab_area + groundFloor_area + ribbedSlab_area + prefabSlab_area + roof_area + undefinedSlab_area

        level: levels[index]
        nameLevel: level[2]
        heightLevel: level[1]
        listArea: []
        listArea.append(totalarea)
        SlabAreaPerLevel: SlabAreaPerLevel + listArea

    sumSlabArea: sum(SlabAreaPerLevel)

    if sumSlabArea != 0:
        for index, item in enumerate(levels):
            i: index
            i2: i + 1
            try:
                NextLevel: levels[i2]
                heightToNextLevel: NextLevel[1] - item[1]

                if 1.85 <= heightToNextLevel <= 1.95 or heightToNextLevel < 0.11 or 2.524 < heightToNextLevel < 2.526:
                    SlabArea: SlabAreaPerLevel[i] + SlabAreaPerLevel[i2]
                    SlabAreaPerLevel[i]: 0
                    SlabAreaPerLevel[i2]: SlabArea

            except IndexError:
                pass

        SlabAreaPerLevel2: []

        for area in SlabAreaPerLevel:
            if area > 0:
                SlabAreaPerLevel2.append(area)

        areaRoof: 0
        try:
            for index, item in enumerate(SlabAreaPerLevel2):
```

```
357                 index2: index + 1
358                 roof: item - SlabAreaPerLevel2[index2]
359                 if roof > 0:
360                     areaRoof += roof
361         except IndexError:
362             pass
363
364         sumSlabArea: sum(SlabAreaPerLevel)
365
366         areaGroundFloor: SlabAreaPerLevel2[0]
367         areaRoof += SlabAreaPerLevel2[-1]
368         areaUpperFloor: sumSlabArea - areaGroundFloor - areaRoof
369
370     else:
371         SlabAreaPerLevel: []
372         for index, list in enumerate(SlabProductsPerLevel):
373             products: list
374
375             # vocabulary IfcSlab
376             slab_ignore_options: ["tegelvloer", "vergaarbak", "gras",
"asfalt", "4e verdiepingsvloer", "ontgraving"]
377
378             # attributes IfcSlab
379             wideSlabFloor_area: 0
380             hollowCoreSlab_area: 0
381             groundFloor_area: 0
382             ribbedSlab_area: 0
383             prefabSlab_area: 0
384             creeds_area: 0
385             insulationSlab_area: 0
386             roof_area: 0
387             undefinedSlab_area: 0
388             VolumeSlabs: 0
389
390             z_levels: []
391
392             area_groundFloor: 0
393             area_RoofFlat: 0
394             areaIgnore_options: ["isolatie"]
395
396             for product in products:
397                 z: get_zvalue(product)
398                 if z is not None:
399                     z0: z[0]
400                     z1: z[1]
401                     thickness: z1 - z0
402                     # Create geometry for these windows
403                     shape: create_shape(product)
404
405                     # Calculate their volumes
406                     volume: calc_volume(shape)
407                     abs_volume: abs(volume)
408                     VolumeSlabs += abs_volume
409
410                     area: abs_volume / thickness
411
412                     product_str: str(product)
413                     product_string: product_str.lower()
414                     if any(option in product_string for option in
slab_ignore_options):
415                         pass
```

```
416                     else:
417                         undefinedSlab_area += area
418
419
420
421             totalarea: undefinedSlab_area
422             level: levels[index]
423             nameLevel: level[2]
424             heightLevel: level[1]
425             listArea: []
426             listArea.append(totalarea)
427             if totalarea > 0:
428                 SlabAreaPerLevel: SlabAreaPerLevel + listArea
429
430         areaRoof: 0
431         try:
432             for index, item in enumerate(SlabAreaPerLevel):
433                 index2: index + 1
434                 roof: item - SlabAreaPerLevel[index2]
435                 if roof > 0:
436                     areaRoof += roof
437         except IndexError:
438             pass
439
440         sumSlabArea: sum(SlabAreaPerLevel)
441
442         areaGroundFloor: SlabAreaPerLevel[0]
443         areaRoof += SlabAreaPerLevel[-1]
444         areaUpperFloor: sumSlabArea - areaGroundFloor - areaRoof
445
446     products1: ifc_file.by_type("IfcBuildingElementProxy")
447     products2: ifc_file.by_type("IfcFurnishingElement")
448     products: products1 + products2
449
450
451     stair_options: ["nl0-dht_nl+1-oht_2kwart", "trap blok b 2", "trap
blok a 1", "vwn_houten trap-dk"]
452     stair_ignore: ["trapleuning","traphek"]
453     countStair: 0
454
455     for product in products:
456         product_str: str(product)
457         product_string: product_str.lower()
458         if any(option in product_string for option in stair_options):
459             countStair += 1
460
461     listOfStairs: []
462     products: ifc_file.by_type("IfcStair")
463
464     for product in products:
465         product_str: str(product)
466         product_string: product_str.lower()
467
468         abc: product_string.split(",", 3)
469         abc2: abc[2]
470         if any(option in product_string for option in stair_ignore):
continue
471         listOfStairs.append(abc2)
472     countStair += len(listOfStairs)
473
474     print "Calculating IfcWalls"
```

```
475  products: ifc_file.by_type("IfcWall")
476
477  # vocabulary IfcWall
478  prefabWall_options: ["prefab beton_100", "prefab beton-90"]
479  insulationWall_options: ["isolatie", "steenwol", "rockpanel",
"isover", "kingspan"]
480  brickwork_options: ["metselwerk", "mets.", "baksteen", "terca",
"uitstekende kop", "vlakke kop", "mw 100"]
481  limestone_options: ["kalkzandsteen", "kzst."]
482  timberFrame_options: ["hsbwand", "hsb wand", "hsb_100mm",
"hsb_150mm", "hsb_214mm", "21hsb_150"]
483  cellularConcrete_options: ["cellenbeton", "gasbeton", "ytong", "gas
beton"]
484  gypsum_options: ["gibo"]
485  insideWall_options: ["binnenwand"]
486  ignore_options: ["ms-wand", "privacyscherm", "gaas met hedera",
"wandtegels", "staal_10mm", "gasbeton_100/g4_600", "beplating", "31
paneel_kozijn",
487                     "41_beplating_15", "21_kozijnrekje 80mm",
"20_zink", "stucwerk", "21_stijl en regelwerk 18+71", "betonnen
kantplank_100mm", "multiplex",
488                     "foamglass", "berging", "schutting", "hekwerk",
"tegelwand", "prefabplint", "fundering", "houtregel"]
489
490  prefabWall_area: 0
491  insulationWall_area: 0
492  brickwork_area: 0
493  limestone_area: 0
494  timberFrame_area: 0
495  cellularConcrete_area: 0
496  gypsum_area= 0
497  insideWall_area: 0
498  undefinedWall_area: 0
499
500  limestone_cs36_300: 0
501  limestone_cs36_214: 0
502  limestone_cs36_150: 0
503  limestone_cs36_120: 0
504  limestone_cs36_100: 0
505  limestone_300: 0
506  limestone_214: 0
507  limestone_150: 0
508  limestone_120: 0
509  limestone_100: 0
510  limestone_undefined: 0
511
512  VolumeWalls: 0
513  errorCount_wall: 0
514
515  products: ifc_file.by_type("IfcWall")
516
517  for product in products:
518      try:
519          x: get_xvalue(product)
520          y: get_yvalue(product)
521          z: get_zvalue(product)
522
523          x0: x[0]
524          x1: x[1]
525
526          y0: y[0]
```

```
527             y1: y[1]
528
529             z0: z[0]
530             z1: z[1]
531
532             l: x1 - x0
533             b: y1 - y0
534             h: z1 - z0
535
536             if l > 0.5 and b > 0.5:
537                 length: math.sqrt((l*l)+(b*b))
538             elif l < b:
539                 length: b
540             else:
541                 length: l
542
543             area: length * h
544
545             shape: create_shape(product)
546             volume: calc_volume(shape)
547             abs_volume: abs(volume)
548             VolumeWalls += abs_volume
549
550             product_str: str(product)
551             product_string: product_str.lower()
552             found: 0
553             if any(option in product_string for option in
prefabWall_options):
554                 prefabWall_area += area
555             elif any(option in product_string for option in
insulationWall_options):
556                 insulationWall_area += area
557             elif any(option in product_string for option in
brickwork_options):
558                 brickwork_area += area
559             elif any(option in product_string for option in
limestone_options):
560                 limestone_area += area
561                 if "cs36" in product_string:
562                     if "300mm" in product_string or "e300" in
product_string:
563                         limestone_cs36_300 += area
564                     elif "214mm" in product_string or "e214" in
product_string:
565                         limestone_cs36_214 += area
566                     elif "150mm" in product_string or "e150" in
product_string:
567                         limestone_cs36_150 += area
568                     elif "120mm" in product_string or "e120" in
product_string:
569                         limestone_cs36_120 += area
570                     elif "100mm" in product_string or "e100" in
product_string:
571                         limestone_cs36_100 += area
572                     else:
573                         limestone_undefined += area
574                 else:
575                     if "300mm" in product_string or "e300" in
product_string:
576                         limestone_300 += area
```

```
577                     elif "214mm" in product_string or "e214" in
product_string:
578                         limestone_214 += area
579                     elif "150mm" in product_string or "e150" in
product_string:
580                         limestone_150 += area
581                     elif "120mm" in product_string or "e120" in
product_string:
582                         limestone_120 += area
583                     elif "100mm" in product_string or "e100" in
product_string:
584                         limestone_100 += area
585                     else:
586                         limestone_undefined += area
587             elif any(option in product_string for option in
insideWall_options):
588                 insideWall_area += area
589             elif any(option in product_string for option in
timberFrame_options):
590                 timberFrame_area += area
591             elif any(option in product_string for option in
cellularConcrete_options):
592                 cellularConcrete_area += area
593             elif any(option in product_string for option in
gypsum_options):
594                 gypsum_area += area
595             elif any(option in product_string for option in
ignore_options):
596                 pass
597             else:
598                 undefinedWall_area += area
599
600         except Exception as ex:
601             template: "An exception of type {0} occured.
Arguments:\n{1!r}"
602             message: template.format(type(ex).__name__, ex.args)
603             errorCount_wall += 1
604
605     areaExternalWalls: brickwork_area+ timberFrame_area
606     areaAllInternalWalls:  limestone_area + cellularConcrete_area +
gypsum_area + insideWall_area + undefinedWall_area + prefabWall_area
607     areaInternalWalls: areaAllInternalWalls - areaExternalWalls
608
609     if areaInternalWalls < 0:
610         areaInternalWalls: areaAllInternalWalls
611
612     print "Calculating IfcDoors"
613
614     products: ifc_file.by_type("IfcDoor")
615
616     # vocabulary IfcDoor
617     externalDoors_option: ["fit_deur", "31_do_deur", "vision trend
merk", "31_deur", "31_stoldeur", "deur-hout_by", "deur-hout_stoldeur",
"buitendeur", "garagedeur"]
618
619     # attributes IfcDoor
620     externalDoors_area: 0
621     countInternalDoors: 0
622
623
624     for product in products:
```

```
625         product_str: str(product)
626         product_string: product_str.lower()
627         abc: product_string.split(",")
628
629         h1: abc[-2]
630         h2: h1.split(".")
631         h3: h2[0]
632         b1: abc[-1]
633         b2: b1.split(".")
634         b3: b2[0]
635
636         h: float(h3) / 1000
637         b: float(b3) / 1000
638         area: h * b
639         if any(option in product_string for option in
externalDoors option):
640             externalDoors_area += area
641
642         else:
643             if area > 1.5:
644                 countInternalDoors += 1
645
646     areaExternalWalls += externalDoors_area
647
648
649     print "Calculating IfcWindows"
650
651     products: ifc_file.by_type("Ifcwindow")
652
653     # attributes IfcDoor
654     externalWindow_area: 0
655
656     for product in products:
657         product_str: str(product)
658         product_string: product_str.lower()
659         abc: product_string.split(",")
660
661         h1: abc[-2]
662         h2: h1.split(".")
663         h3: h2[0]
664         b1: abc[-1]
665         b2: b1.split(".")
666         b3: b2[0]
667
668         h: float(h3) / 1000
669         b: float(b3) / 1000
670         area: h * b
671         externalWindow_area += area
672
673     areaExternalWalls += externalWindow_area
674
675     areaExternalOpenings: externalDoors_area +externalWindow_area
676
677     with open(sys.argv[2], "ab") as f_csv:
678         csvWriter: csv.writer(f_csv)
679         csvWriter.writerow([(sys.argv[1]), areaGroundFloor,
areaUpperFloor, areaRoof, countStair, areaExternalWalls,
areaExternalOpenings, areaInternalWalls, countInternalDoors])
```

Appendix 3: Code of the k-Nearest Neighbour

k-Nearest Neighbour

```
setwd("C:/test")
require(9825)
mydata: read.csv("dataset2.csv", header: TRUE, sep: ";", dec: ",")
set.seed(1)
gp <- runif(nrow(mydata))
mydata2 <- mydata[order(gp),]
normalize <- function(x) { return( (x - min(x)) / (max(x) - min(x)) ) }
data_n <- as.data.frame(lapply(mydata2[,c(2,3,4,5,6,7)], normalize))
data_train <- data_n[1:9, 1:5]
data_test <- data_n[10:12, 1:5]
data_train_target <- mydata2[1:9, 7]
data_test_target <- mydata2[10:12, 7]
m1 <- knn(train=data_train, test=data_test, cl=data_train_target, k=1)
m3 <- knn(train=data_train, test=data_test, cl=data_train_target, k=3)
m5 <- knn(train=data_train, test=data_test, cl=data_train_target, k=5)
data_train_target
data_test_target
m1
m3
m5
```

Appendix 4: Additional project data

Nieuwbouw 22 rijenwoning Dronten (source: casadata.nl)

Bouwkosten nieuwbouw 22 rijenwoningen, 3-laags met zadeldak, met gemiddeld 112 m2 bvo, 288 m3 bi (tabel p82.00.010)

Kosten in euro per eenheid (peildatum: 20 maart 2017)

Soort project	Omschrijving project	Bijzonderheid	Eenheid	Excl. BTW Totaal	Incl. BTW Eindtotaal
Nieuwbouw 22 rijenwoningen, 3-laags, koop en huur. Gemiddeld 112 m2 bvo, 288 m3 bi. Totaal project 2.471 m2 bvo en 6.329 m3 inhoud	Geïndexeerde kosten	Bouwkosten totaal	totaal	1.657.280,80	2.005.309,76
			woning	75.330,95	91.150,44
			m2 bvo	670,69	811,54
			m3 bi	261,86	316,84
		Waarvan bouwkundig werk	totaal	1.074.706,48	1.300.394,84
			woning	48.850,29	59.108,86

			m2 bvo	434,93	526,26
			m3 bi	169,81	205,47
		Waarvan installaties	totaal	293.271,70	354.858,75
			woning	13.330,53	16.129,94
			m2 bvo	118,69	143,61
			m3 bi	46,34	56,07
		Waarvan overige bouwkosten	totaal	289.302,62	350.056,18
			woning	13.150,12	15.911,64
			m2 bvo	117,08	141,67
			m3 bi	45,71	55,31
	Originele bouwkosten op origineel prijspeil april 2006	Totaal bouwkosten	totaal	1.372.483,00	1.660.704,43
			woning	62.385,59	75.486,57
		Waarvan bouwkundig werk	totaal	890.022,00	1.076.926,62
		Waarvan installaties	totaal	242.874,00	293.877,54
		Waarvan overige bouwkosten	totaal	239.587,00	289.900,27

Kosten

Uitgangspunt

Alle kosten in deze tabel zijn gebaseerd op de oorspronkelijke bouwkosten op het originele prijspeil van april 2006 van de aanbesteding. De projecthoeveelheden betreffen dit referentieproject.
Het gaat om de nieuwbouw van 22 rijenwoningen, die als koop- en huurwoningen in een project met 67 woningen in De Gilden in Dronten zijn gerealiseerd.

Kosten exclusief

De kosten zijn exclusief terreinkosten en bijkomende investeringskosten, zoals ontwerp- en advieskosten en de leges van (bouw)vergunningen.

Kosten inclusief

De kosten bouwkundig werk zijn inclusief sanitaire voorzieningen en bergingen.

Projectomschrijving

De rechthoekige woningen zijn gefundeerd op betonnen funderingsbalken op prefab betonnen palen. L = 4,5 m d = 290 mm. De gevel is volledig opgetrokken in metselwerk en gevuld met meranti buitenkozijnen met geïsoleerd glas en holonite vensterbanken.
Het dak bestaat uit een prefab houten dakconstructie, gedekt met betonpannen en voorzien van waterslagprofielen, dakkappen en schoorstenen. Hemelwaterafvoer via mastgoten.
Als begane grondvloer is een ribcasseteplaten en als verdiepingsvloer kanaalplaatvloeren toegepast, afgewerkt als

cementdekvloeren. Vloer in badkamer en toiletten zijn betegeld. Binnenwanden in kalkzandsteen en GIBO wanden, gevuld met stalen binnenkozijnen met opdekdeuren. Binnenwanden zijn afgewerkt met stucwerk en in de badkamer en toiletten betegeld.
Het plafond is in de woningen afgewerkt met spuitwerk en in de bergingen met houtwolcementplaat. Houten vaste trappen met aftimmering van trapgaten en lepe hoeken. Standaard sanitair inrichting in badkamers en toiletten.
Gasinstallatie voor CV-ketels. Verwarming met CV-ketel en radiatoren en mechanische ventilatie met warmteterugwinning units. Gangbaar leidingwerk voor elektrische installaties. Brandbeveiliging met optische rookmelders.

Algemene projectgegevens
De woningen hebben een nokhoogte van 7,3 m en een goothoogte van 5 meter..
De bruto inhoud van de 22 woningen is 6.329 m3 totaal ofwel gemiddeld 287,7 m3 per woning. De bruto vloeroppervlakte is totaal 2.471 m2 bvo ofwel gemiddeld 112,3 m2 bvo per woning.
Het gebouw bevat drie bouwlagen en heeft geen bouwlaag -peil.
De bruto dakoppervlakte is 1.177 m2 totaal (53,5 m2 per woning) en de bruto geveloppervlakte is 1.291 m2 (58,7 m2 per woning), waarvan 317 m2 met gevelopeningen is gevuld (25%).
In verhouding tot de bruto vloeroppervlakte is de geveloppervlakte 52% en is de omsluiting van gevel en dak samen 100%.

Bouwpartners
De informatie is samengesteld met medewerking van Huls Bouwkostenmanagement in Zwolle.
De opdrachtgever is Oost Flevoland Woondiensten in Dronten.
De architect is 19 het Atelier in Zwolle.
De aannemer is B.J. van der Steeg in Genemuiden.

Samenstelling gegevens
Disclaimer
Iedereen mag deze informatie raadplegen, afdrukken en gebruiken.
Casadata cv en de informatieverstrekkers stellen de informatie naar beste weten en kunnen zo zorgvuldig mogelijk samen. Casadata cv adviseert iedereen nadrukkelijk om niet uitsluitend deze informatie te gebruiken, maar deze altijd in samenhang met andere informatie te controleren.
Casadata cv en de informatieverstrekkers aanvaarden dan ook geen enkele aansprakelijkheid voor eventuele schade, die het gevolg is van handelingen of beslissingen, die voortkomen of gebaseerd zijn op door Casadata cv verstrekte informatie.

Nieuwbouw 31 rijenwoningen (source: casadata.nl)

Bouwkosten nieuwbouw 31 rijenwoningen, 2-laags met plat dak, met gemiddeld 121 m2 bvo, 360 m3 bi (tabel p82.00.009)

Kosten in euro per eenheid (peildatum: 20 maart 2017)

Soort project	Omschrijving project	Bijzonderheid	Eenheid	Excl. BTW Totaal	Incl. BTW Eindtotaal
Nieuwbouw 31 rijenwoningen, 2-laags. Gemiddeld 121 m2 bvo en 360 m3 bruto inhoud. Totaal project 3.757 m2 bvo en 11.153 m3 bi	Geïndexeerde kosten	Bouwkosten totaal	totaal	3.216.985,14	3.892.552,02
			woning	103.773,71	125.566,19
			m2 bvo	856,26	1.036,08
			m3 bi	288,44	349,01
	Originele bouwkosten op origineel prijspeil november 2004	Totaal bouwkosten	totaal	2.573.000,00	3.113.330,00
			woning	83.000,00	100.430,00

Kosten

Uitgangspunt

Alle kosten in deze tabel zijn gebaseerd op de oorspronkelijke bouwkosten op het originele prijspeil van november 2004 van de aanbesteding. De projecthoeveelheden betreffen dit referentieproject.
Het gaat om de nieuwbouw van 31 rijenwoningen in Rotterdam.

Kosten exclusief

De kosten zijn exclusief terreinafwerking en bijkomende investeringskosten, zoals ontwerp- en advieskosten en de leges van (bouw)vergunningen.

Projectomschrijving

Het nieuwbouwproject bestaat uit vier blokken met 31 eengezinswoningen. Elke ruimte heeft verdiepingshoge kozijnen, die overal even groot zijn. Daardoor kan men de ramen en deuren onderling uitwisselen. Door het verschuiven van de aluminium zonneschermen op de zuidgevel bepalen de bewoners zelf hun privacy en het zonlicht.
De buitenwanden zijn uitgevoerd in oranje metselwerk. In het gevelvlak zijn vierkante hemelwaterafvoeren opgenomen.
In de gevelopeningen zijn Sapupira hardhouten kozijnen met FSC keurmerk geplaatst.
De woningen zijn voorzien van standaard sanitair en standaard installaties voor gas, water en afvoeren.

Algemene projectgegevens

De woningen hebben een breedte van 5,4 m en het totaal van de vier woningblokken is 167,4 meter breed.
De bruto inhoud van de 31 woningen is 11.153 m3 totaal ofwel gemiddeld 359,8 m3 per woning. De bruto vloeroppervlakte is totaal 3.757 m2 bvo ofwel gemiddeld 121,2 m2 bvo per woning.
Het gebouw bevat twee bouwlagen en heeft geen bouwlaag -peil.
De bruto dakoppervlakte is 1.995 m2 totaal (64,4 m2 per woning) en de bruto geveloppervlakte is 2.632 m2 (85 m2 per woning), waarvan 770 m2 met gevelopeningen is gevuld (29%).
In verhouding tot de bruto vloeroppervlakte is de geveloppervlakte 70% en is de omsluiting van gevel en dak samen 123%.

Bouwpartners

De informatie is samengesteld met medewerking van Architectenbureau K2 te Amsterdam.
De opdrachtgever is Vestia Rotterdam Zuid.
De architect is Architectenbureau K2 te Amsterdam.
De aannemer is GN bouw b.v. in Dordrecht.

Samenstelling gegevens

Disclaimer

Iedereen mag deze informatie raadplegen, afdrukken en gebruiken.
Casadata cv en de informatieverstrekkers stellen de informatie naar beste weten en kunnen zo zorgvuldig mogelijk samen. Casadata cv adviseert iedereen nadrukkelijk om niet uitsluitend deze informatie te gebruiken, maar deze altijd in samenhang met andere informatie te controleren.
Casadata cv en de informatieverstrekkers aanvaarden dan ook geen enkele aansprakelijkheid voor eventuele schade, die het gevolg is van handelingen of beslissingen, die voortkomen of gebaseerd zijn op door Casadata cv verstrekte informatie.

Nieuwbouw 14 rijenwoningen Groenoord (source: casadata.nl)

Bouwkosten nieuwbouw 14 rijenwoningen, 4-laags met plat dak, met gemiddeld 195 m2 bvo, 591 m3 bi (tabel p83.00.010)

Kosten in euro per eenheid (peildatum: 20 maart 2017)

Soort project	Omschrijving project	Bijzonderheid	Eenheid	Excl. BTW Totaal	Incl. BTW Eindtotaal
Nieuwbouw 14 rijenwoningen, 4-laags. Gemiddeld 195 m2 bvo en 591 m3 bi per woning. Totaal project 2.730 m2 bvo en 8.270 m3 bi	Geïndexeerde kosten	Bouwkosten totaal	totaal	2.173.509,93	2.629.947,02
			woning	155.250,71	187.853,36
			m2 bvo	796,16	963,35
			m3 bi	262,82	318,01
	Originele bouwkosten op origineel prijspeil mei 2006	Totaal bouwkosten	totaal	1.800.000,00	2.178.000,00
			woning	128.571,43	155.571,43

Kosten

Uitgangspunt

Alle kosten in deze tabel zijn gebaseerd op de oorspronkelijke bouwkosten op het originele prijspeil van mei 2006 van de aanbesteding. De projecthoeveelheden betreffen dit referentieproject.
Het gaat om de nieuwbouw van 14 rijenwoningen in een reconstructiewijk uit de jaren zestig in Rotterdam zijn gerealiseerd.

Kosten exclusief

De kosten zijn exclusief terreinafwerking en bijkomende investeringskosten, zoals ontwerp- en advieskosten en de leges van (bouw)vergunningen.

Projectomschrijving

Door de smalle beukbreedte van 3,90 m van de woningen zijn er geen extra stabiliteitsvoorzieningen nodig, waardoor de woningen volledig flexibel zijn. Ook zijn er afwijkende doorsneden mogelijk, waardoor een kamer kan vervallen voor een dakterras of voor een extra hoogte in keuken, werk- of slaapkamer. Door de centrale plaats van de leidingschacht is op elke verdieping een extra badkamer of keukenblok mogelijk. De betonstructuur is op strategische plaatsen wapeningsvrij, zodat men later woningen of delen ervan kan samenvoegen.
Op de betonnen fundering is het skelet van het gebouw in gietbouw opgetrokken met een betonnen dak. Er zijn houtskeletbouw gevelelementen toegepast met kozijnen in combinatie met IBS vezelcementplaat. Het betonnen dak is afgewerkt met bitumineuze dakafwerking.
Er is gangbaar sanitair geplaatst. Ook zijn de woningen voorzien van gangbare woninginstallaties en leidingwerk voor water, binnenriolering, elektra en verlichting, verwarming en ventilatie. Er zijn zonnecollectoren geplaatst.

Algemene projectgegevens

De woningen hebben een breedte van 3,9 m en het totale woningblok is 54,6 meter breed.
De bruto inhoud van de 14 woningen is 8.270 m3 totaal ofwel gemiddeld 591 m3 per woning. De bruto vloeroppervlakte is totaal 2.730 m2 bvo ofwel gemiddeld 195 m2 bvo per woning.
Het gebouw bevat vier bouwlagen en heeft geen bouwlaag -peil.
De bruto dakoppervlakte is 645 m2 totaal (46 m2 per woning) en de bruto geveloppervlakte is 1.628 m2 (116 m2 per woning), waarvan 652 m2 met gevelopeningen is gevuld (40%).
In verhouding tot de bruto vloeroppervlakte is de geveloppervlakte 60% en is de omsluiting van gevel en dak samen 83%.

Bouwpartners

De informatie is samengesteld met medewerking van Architectenbureau K2 te Amsterdam.
De opdrachtgever is Wijk Ontwikkelingsmaatschappij Groenoord te Rotterdam.
De architect is Architectenbureau K2 te Amsterdam.
De aannemer is Riz bouw b.v. in IJsselstein.

Samenstelling gegevens

Disclaimer

Iedereen mag deze informatie raadplegen, afdrukken en gebruiken.
Casadata cv en de informatieverstrekkers stellen de informatie naar beste weten en kunnen zo zorgvuldig mogelijk samen. Casadata cv adviseert iedereen nadrukkelijk om niet uitsluitend deze informatie te gebruiken, maar deze altijd in samenhang met andere informatie te controleren.
Casadata cv en de informatieverstrekkers aanvaarden dan ook geen enkele aansprakelijkheid voor eventuele schade, die het gevolg is van handelingen of beslissingen, die voortkomen of gebaseerd zijn op door Casadata cv verstrekte informatie.

28 appartementen De Haere (source: bouwkosten.nl)

BOUWkosten

28 APPARTEMENTEN "DE HAERE" TE HEINO

TABEL | OMSCHRIJVING | AFBEELDINGEN

Omschrijving

Projectomschrijving

Op een voormalig schoolterrein in Heino wordt een woonproject van 28 huurappartementen met daarachter 8 geschakelde huurwoningen voor 55+'ers gerealiseerd met energielabel A.

Het appartementencomplex is opgezet in een U-vorm waarvan de twee uitlopers naar de doorgaande straat wijzen en een openbaar hofje omsluiten. Op de begane grond bevinden zich 8 appartementen met een eigen buitenruimte. De eerste verdieping en de tweede verdieping met kap bevatten elk 10 appartementen. Er bevinden zich 2 liften in het complex. De appartementen op de begane grond zijn allleen toegankelijk via de openbare weg. De overige appartementen zijn toegankelijk vanaf de centrale hal, welke is afgesloten. Bezoekers melden zich via de videofoon.

Aan de achterzijde van het appartementencomplex liggen aan een rustige woonstraat 8 seniorenwoningen. Deze woningen hebben 1 bouwlaag met een dakopbouw met eigen buitenruimte.

De appartementen en de laagbouwwoningen hebben een klassieke uitstraling door het gebruik van een eenvoudige rode baksteen –een geribbelde en een tweemaal zo grote gladde steen – die afwisselend in verticale stroken worden gebruikt.

De kozijnen zijn van kunststof met HR++ beglazing. De zuidoost- en zuidwestzijde zijn voorzien van elektrisch bedienbare zonwering.

De badkamer, keuken en toilet zijn (deels) betegeld. Overige wanden zijn behangklaar gemaakt. Plafonds: spuitwerk (spack), kleur wit. Vloer: afwerkvloer (cementdekvloer) behalve in de badkamer waar de vloer is betegeld.

Iedere woning/appartement is voorzien van een gebalanceerd ventilatiesysteem, een zonnecollector (PV) en een zonneboiler voor warmtapwater. Op het dak van het appartementencomplex bevinden zich ook zonnecollectoren (PV) welke elektriciteit levert voor de lift en de verlichting van de algemene ruimten.

Algemene projectgegevens

Samenstelling: Huls Bouwkostenmanagement
Opdrachtgever: Salland Wonen te Raalte
Architect: Van Den Berg IJsselmuiden te IJsselmuiden
Hoofdaannemer: Aannemersbedrijf J. Valk en Zn. B.V. te Daarle

Peildatum originele kosten: juni 2012
Start uitvoering: augustus 2012
Oplevering: november 2013
Bouwtijd: 15 kalendermaanden
Aanbestedingsvorm: bouwteam
Risicoregeling van toepassing: n.v.t.

Bruto-inhoud: 10.493 m3
Alle kosten zijn geïndexeerd.

Kosten- en gegevenstabellen

KOSTEN OP NIVEAU 1 GEHEEL BOUWWERK | BOUWFYSISCHE KENMERKEN | HOEVEELHEDEN | VORMFACTOREN
KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS | KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS, HOOFDINDELING | KOSTEN OP NIVEAU 3 ELEMENTEN
OMSCHRIJVING ELEMENTEN | RUIMTESOORTEN, TOTAAL | RUIMTESOORTEN, SPECIFICATIE

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	2.285.185,00		2.285.185,00
3	Installaties		565.105,00	565.105,00

4	Vaste inrichtingen		103.935,00	103.935,00
5	Terrein	12.535,00		12.535,00
	Totaal directe bouwkosten	2.297.720,-	669.040,-	2.966.760,00
6	Algemene bouwkosten	223.840,00		223.840,00
	Subtotaal	2.521.560,-	669.040,-	3.190.600,-
6	Algemene bedrijfskosten	133.390,00		133.390,00
	Subtotaal	2.654.950,-	669.040,-	3.323.990,-
6	Winst en risico	56.025,00		56.025,00
	Totaal bouwkosten (excl. BTW)	2.710.975,-	669.040,-	3.380.015,00

Bouwfysische kenmerken

Beganegrondvloer	R =	3,50	$m^2.K/W$
Geveiconstructie	R =	5,00	$m^2.K/W$
Glas	U =	1,10	$W/(m^2.K)$
Dakconstructie	R =	5,00	$m^2.K/W$
Energie Prestatie Coëfficiënt		0,60	

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	3	stuks	Dakoppervlakte, bruto	1.322	m^2
Maximale hoogte	11,00	m	Dakoppervlakte	1.322	m^2
Terrein			Dakopeningen		m^2
Terreinoppervlakte	2.520	m^2	Gevels		
Bebouwde terreinoppervlakte	1.165	m^2	Gevelomtrek	284	m
Onbebouwde terreinoppervlakte	1.355	m^2	Buitenwandoppervlakte, bruto	2.092	m^2
Oppervlakteverharding	215	m^2	Buitenwanden	1.486	m^2
Aantal parkeerplaatsen		stuks	Buitenwandopeningen	606	m^2
Ruimtedelen			Binnenwanden		
Bruto-vloeroppervlakte (BVO)	3.302	m^2	Binnenwandoppervlakte, bruto	4.303	m^2
Netto-vloeroppervlakte (NVO)	2.708	m^2	Binnenwanden	3.684	m^2
Installatieoppervlakte	36	m^2	Binnenwandopeningen	619	m^2
Verkeersoppervlakte	630	m^2	Trapconstructies		
Nuttige oppervlakte	2.042	m^2	Trappen en hellingen	4	stuks
Niet-toegankelijke ruimte		m^2	Balustrades en leuningen	34	m
Tarra-oppervlakte	594	m^2	Installaties		
Gebruiksoppervlakte	2.763	m^2	Klimaatinstallaties (koeling)		kW
Bruto-inhoud (BI)	10.493	m^3	Klimaatinstallaties (verwarming)		kW
Skelet			Warmtepomp		kW
Buitenwanden (constructief)	1.276	m^2	Zonnecollector		kW
Binnenwanden (constructief)	899	m^2	Vloerverwarming		m^2
Beganegrondvloeren (constructief)	1.165	m^2	Warmtelichamen		stuks
Verdiepingsvloeren (constructief)	2.137	m^2	Wandverwarming		m^2
Dakvloeren (constructief)	1.322	m^2	Liften	2	stuks

29 appartementen boven winkels (source: bouwkosten.nl)

BOUWkosten

29 APPARTEMENTEN BOVEN WINKELS TE ZWOLLE

TABEL | OMSCHRIJVING | AFBEELDINGEN

Omschrijving

Projectomschrijving

De N.V. Wonen boven Winkels Zwolle heeft in het centrum van de stad circa 800 m² nieuwe winkelruimte met daarboven in totaal 29 appartementen weten te realiseren. Daartoe is het overgrote deel van de eerdere bebouwing –na uitvoerige inventarisatie, bouwhistorisch onderzoek en modellenstudies– gesloopt. Het project is van forse omvang en heeft gevels aan drie straten: de Diezerstraat, Korte Smeden en Nieuwstraat.

De nieuwbouw moest binnen de voorschriften van het bestemmingsplan blijven; te weten drie lagen en een kap. De winkel op de begane grond wordt door de corporatie als één grote ruimte verhuurd, maar de winkelruimte is op te splitsen in meerdere kleine eenheden. Verdeeld over de bovenliggende drie verdiepingen een combinatie van 2 tweekamerappartementen, 12 driekamerappartementen en 15 studio's (eenkamerappartementjes à 25 m²).

De meeste wooneenheden liggen aan het binnenterrein en zijn met de rug tegen het aangrenzende achterliggende pand gebouwd. Ze hebben geen balkon, maar de bewoners kunnen buiten zitten op de galerij of op het binnenterrein.

In eerste instantie zou de hoofdingang aan de Nieuwstraat komen. Daar is uiteindelijk het noodtrappenhuis geplaatst. De ingang, het trappenhuis en de lift zijn nu gelegen aan de Korte Smeden. Voor de bewoners is daar ook de ingang van hun gemeenschappelijke ondergrondse fietsenberging geplaatst. Elk van de nieuwe of gerenoveerde gevels is aan de straatzijde anders van maat en kleur. Hierdoor sluiten deze aan bij de variëteit aan bestaande gevels in dat deel van de binnenstad.

Algemene projectgegevens
Samenstelling: Huls Bouwkostenmanagement
Opdrachtgever: N.V. Wonen Winkels Zwolle te Zwolle
Architect: IM architecten te Deventer
Hoofdaannemer: Salverda te 't Harde

Peildatum originele kosten: oktober 2006
Start uitvoering: september 2007
Oplevering: december 2008
Bouwtijd: 52 werkbare weken
Aanbestedingsvorm: bouwteam
Risicoregeling van toepassing: Nee

Bruto-inhoud: 11.688 m³
Alle kosten zijn geïndexeerd.

Kosten- en gegevenstabellen

KOSTEN OP NIVEAU 1 GEHEEL BOUWWERK | BOUWFYSISCHE KENMERKEN | HOEVEELHEDEN | VORMFACTOREN
KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS | KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS, HOOFDINDELING | KOSTEN OP NIVEAU 3 ELEMENTEN
OMSCHRIJVING ELEMENTEN | RUIMTESOORTEN, TOTAAL | RUIMTESOORTEN, SPECIFICATIE

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	3.042.910,00		3.042.910,00
3	Installaties		443.810,00	443.810,00

4	Vaste inrichtingen		64.665,00	64.665,00
5	Terrein			
	Totaal directe bouwkosten	3.042.910,-	508.480,-	3.551.385,00
6	Algemene bouwkosten	719.690,00		719.690,00
	Subtotaal	3.762.595,-	508.480,-	4.271.075,-
6	Algemene bedrijfskosten	239.385,00		239.385,00
	Subtotaal	4.001.980,-	508.480,-	4.510.460,-
6	Winst en risico	112.655,00		112.655,00
	Totaal bouwkosten (excl. BTW)	4.114.630,-	508.480,-	4.623.110,00

Bouwfysische kenmerken

Beganegrondvloer	R =	4,00	$m^2.K/W$
Gevelconstructie	R =	3,00	$m^2.K/W$
Glas	U =	1,70	$W/(m^2.K)$
Dakconstructie	R =	4,00	$m^2.K/W$
Energie Prestatie Coëfficiënt		0,96	

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	5	stuks	Dakoppervlakte, bruto	990	m^2
Maximale hoogte	15,55	m	Dakoppervlakte	979	m^2
Terrein			Dakopeningen	11	m^2
Terreinoppervlakte	990	m^2	Gevels		
Bebouwde terreinoppervlakte	990	m^2	Gevelomtrek	176	m
Onbebouwde terreinoppervlakte		m^2	Buitenwandoppervlakte, bruto	1.981	m^2
Oppervlakteverharding		m^2	Buitenwanden	1.416	m^2
Aantal parkeerplaatsen		stuks	Buitenwandopeningen	565	m^2
Ruimtedelen			Binnenwanden		
Bruto-vloeroppervlakte (BVO)	3.511	m^2	Binnenwandoppervlakte, bruto	1.411	m^2
Netto-vloeroppervlakte (NVO)	3.013	m^2	Binnenwanden	1.024	m^2
Installatieoppervlakte	63	m^2	Binnenwandopeningen	387	m^2
Verkeersoppervlakte	694	m^2	Trapconstructies		
Nuttige oppervlakte	2.256	m^2	Trappen en hellingen	12	stuks
Niet-toegankelijke ruimte		m^2	Balustrades en leuningen	49	m
Tarra-oppervlakte	498	m^2	Installaties		
Gebruiksoppervlakte	3.006	m^2	Klimaatinstallaties (koeling)		kW
Bruto-inhoud (BI)	11.688	m^3	Klimaatinstallaties (verwarming)		kW
Skelet			Warmtepomp		kW
Buitenwanden (constructief)	1.981	m^2	Zonnecollector		kW
Binnenwanden (constructief)	1.024	m^2	Vloerverwarming		m^2
Beganegrondvloeren (constructief)	990	m^2	Warmtelichamen		stuks
Verdiepingsvloeren (constructief)	2.521	m^2	Wandverwarming		m^2
Dakvloeren (constructief)	990	m^2	Liften		stuks
Hoofddraagconstructie	3.511	m^2			

44 appartementen de parkwachters (source: bouwkosten.nl)

BOUWkosten

44 APPARTEMENTEN DE PARKWACHTERS TE BOSKOOP

TABEL | OMSCHRIJVING | AFBEELDINGEN

Omschrijving

Projectomschrijving

Aan de Parklaan in Boskoop is een nieuw bouwproject gerealiseerd, bestaande uit 44 sociale huurwoningen. Deze drie- en vierkamerappartementen – met woonkamers van 37 tot 48 m^2 – zijn verdeeld over twee torens: 'De Parkwachters'. Het geheel ademt de sfeer van de jaren dertig van de vorige eeuw.

Het basisontwerp is eenvoudig: per bouwlaag staan vier woningen haaks op elkaar gegroepeerd rondom een liftschacht en een trappenhuis. De twee torens zijn op de begane grond verbonden door middel van bergingen. De hoogste toren heeft zeven lagen en 24 appartementen, de ander toren zes lagen en 20 appartementen. In elke toren bevindt zich een lift; één met zes en één met zeven stopplaatsen.

De appartementen in de woontorens op de begane grond beschikken over een terras, de overige appartementen hebben een balkon. De woningen hebben energielabel A, zijn levensloopbestendig en toegankelijk voor rolstoelgebruikers. Vanwege een hogedrukgasleiding in de buurt is het trappenhuis voorzien van druk- en brandwerende beglazing.

De kozijnen van de bergingen zijn van hout, de overige kozijnen zijn van kunststof. In de appartementen zijn de natte ruimten voorzien van keramische tegels met daarboven spuitwerk. De overige wanden zijn behangklaar. Installaties: HR-combiketel met kamerthermostaat. Gescheiden afvoersysteem. De appartementen worden verwarmd via radioatoren. In de badkamer bevindt zich een designradiator en basis sanitair inclusief tweede toilet.

Algemene projectgegevens

Samenstelling: Winket bv | Huisvestingseconomie, bouwkosten en bestekken
Opdrachtgever: Trifolium Woondiensten / Woonpartners Midden-Holland te Boskoop / Waddinxveen
Architect: Mullener en Mullener te Amsterdam
Hoofdaannemer: Van Wijnen te Stolwijk

Peildatum originele kosten: november 2011
Start uitvoering: april 2012
Oplevering: juli 2013
Bouwtijd: 240 werkbare werkdagen
Aanbestedingsvorm: bouwteam
Risicoregeling van toepassing: nee

Bruto-inhoud: 18.236 m3
Alle kosten zijn geïndexeerd.

Kosten- en gegevenstabellen

KOSTEN OP NIVEAU 1 GEHEEL BOUWWERK | BOUWFYSISCHE KENMERKEN | HOEVEELHEDEN | VORMFACTOREN
KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS | KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS, HOOFDINDELING | KOSTEN OP NIVEAU 3 ELEMENTEN
OMSCHRIJVING ELEMENTEN

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	3.569.430,00		3.569.430,00
3	Installaties		1.005.495,00	1.005.495,00
4	Vaste inrichtingen		113.740,00	113.740,00
5	Terrein	85.290,00		85.290,00

	Totaal directe bouwkosten	3.654.725,-	1.119.235,-	4.773.960,00
6	Algemene bouwkosten	402.455,00		402.455,00
	Subtotaal	4.057.180,-	1.119.235,-	5.176.415,-
6	Algemene bedrijfskosten	311.615,00		311.615,00
	Subtotaal	4.368.795,-	1.119.235,-	5.488.030,-
6	Winst en risico	177.590,00		177.590,00
	Totaal bouwkosten (excl. BTW)	4.546.380,-	1.119.235,-	5.665.615,00

Bouwfysische kenmerken

Beganegrondvloer	R =	3,50	$m^2.K/W$
Gevelconstructie	R =	3,50	$m^2.K/W$
Glas	U =	1,65	$W/(m^2.K)$
Dakconstructie	R =	4,00	$m^2.K/W$
Energie Prestatie Coëfficiënt		0,80	

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	6	stuks	Dakoppervlakte, bruto	1.368	m^2
Maximale hoogte	23,00	m	Dakoppervlakte	1.368	m^2
Terrein			Dakopeningen		m^2
Terreinoppervlakte	3.178	m^2	Gevels		
Bebouwde terreinoppervlakte	1.158	m^2	Gevelomtrek	1.503	m
Onbebouwde terreinoppervlakte	2.020	m^2	Buitenwandoppervlakte, bruto	4.358	m^2
Oppervlakteverharding	2.020	m^2	Buitenwanden	2.797	m^2
Aantal parkeerplaatsen	60	stuks	Buitenwandopeningen	1.561	m^2
Ruimtedelen			Binnenwanden		
Bruto-vloeroppervlakte (BVO)	6.766	m^2	Binnenwandoppervlakte, bruto	6.545	m^2
Netto-vloeroppervlakte (NVO)	6.022	m^2	Binnenwanden	5.467	m^2
Installatieoppervlakte	88	m^2	Binnenwandopeningen	1.078	m^2
Verkeersoppervlakte	1.056	m^2	Trapconstructies		
Nuttige oppervlakte	4.778	m^2	Trappen en hellingen	19	stuks
Niet-toegankelijke ruimte	100	m^2	Balustrades en leuningen	545	m
Tarra-oppervlakte	744	m^2	Installaties		
Gebruiksoppervlakte	5.616	m^2	Klimaatinstallaties (koeling)		kW
Bruto-inhoud (BI)	18.236	m^3	Klimaatinstallaties (verwarming)	880	kW
Skelet			Warmtepomp		kW
Buitenwanden (constructief)	2.290	m^2	Zonnecollector		kW
Binnenwanden (constructief)	1.689	m^2	Vloerverwarming		m^2
Beganegrondvloeren (constructief)	1.045	m^2	Warmtelichamen	352	stuks
Verdiepingsvloeren (constructief)	4.976	m^2	Wandverwarming		m^2
Dakvloeren (constructief)	1.368	m^2	Liften	2	stuks
Hoofddraagconstructie	6.766	m^2			

Vormfactoren

Allardhof te buren (source: bouwkosten.nl)

BOUWkosten

ALLARDHOF TE BUREN

TABEL | OMSCHRIJVING | AFBEELDINGEN

Omschrijving

Projectomschrijving

In Buren aan de Appelgaard stond een verouderd appartementencomplex. Kleurrijk Wonen (voorheen Stichting Volkshuisvesting Tiel) heeft dit complex gesloopt. Jeanne Dekkers Architectuur heeft 41 bijzondere appartementen ontworpen die in de plaats van het complex worden gebouwd: project Allardhof. Kleurrijk Wonen heeft ervoor gekozen het project Allardhof in twee fasen te realiseren. Fase 1: 25 huurappartementen, uitgewerkt in deze projectanalyse. Fase 2 bevat 16 koopappartementen.

Buren kent een dorpsachtig karakter. Als snel ontstond hierdoor het idee van een "dorp in een dorp", een mooie manier van wonen in een groene omgeving. De appartementen zijn verdeeld over meerdere kleinschalige gebouwen die in hoogte variëren en rondom een gezamenlijke binnentuin staan. De appartementen zijn gebouwd aan de Willem Alexanderstraat in Buren. Het hoogste gebouw telt drie verdiepingen. Het basismateriaal voor de gevel – een lichte en donkere baksteen – komt voor in de omliggende wijk. Door het echter uniform bij alle bouwdelen toe te passen, geeft het de nieuwbouw zijn eigen identiteit. Parkeren wordt opgelost aan de randen van het plan en onder twee gebouwen, waardoor de tuin vrij blijft van autoverkeer. De daken zijn allemaal bedekt met Sedum wat voor een mooi uitzicht zorgt vanuit de hogere appartementen. Alle appartementen hebben een terras of balkon en maken gebruik van de gezamenlijke binnentuin. Deze centrale tuin ligt in het hart van het plan is tevens een nieuwe, openbare plek voor de bestaande wijk. Door de ecologische inrichting, de mos-sedumdaken en de nieuwe vleermuiskasten is de tuin een bijzonder aangename plek om te verblijven voor mens én dier.

Algemene projectgegevens
Samenstelling: Huls Bouwkostenmanagement
Opdrachtgever: Stichting Volkshuisvesting Tiel te Tiel
Architect: Jeanne Dekkers Architectuur te Delft
Hoofdaannemer: Bouwbedrijf Berghege te Oss

Peildatum originele kosten: oktober 2013
Start uivoering: april 2014
Oplevering: januari 2015
Bouwtijd: 10 maanden
Aanbestedingsvorm: openbaar
Risicoregeling van toepassing: nee

Bruto-inhoud: 9.108 m3
Alle kosten zijn geïndexeerd.

Kosten- en gegevenstabellen

KOSTEN OP NIVEAU 1 GEHEEL BOUWWERK | BOUWFYSISCHE KENMERKEN | HOEVEELHEDEN | VORMFACTOREN
KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS | KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS, HOOFDINDELING | KOSTEN OP NIVEAU 3 ELEMENTEN
OMSCHRIJVING ELEMENTEN | RUIMTESOORTEN, TOTAAL | RUIMTESOORTEN, SPECIFICATIE

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	1.603.205,00		1.603.205,00
3	Installaties		472.765,00	472.765,00
4	Vaste inrichtingen		41.295,00	41.295,00
5	Terrein	22.735,00		22.735,00

	Totaal directe bouwkosten	1.625.945,-	514.060,-	2.140.005,00
6	Algemene bouwkosten	165.550,00		165.550,00
	Subtotaal	1.791.495,-	514.060,-	2.305.555,-
6	Algemene bedrijfskosten	89.725,00		89.725,00
	Subtotaal	1.881.220,-	514.060,-	2.395.280,-
6	Winst en risico	18.845,00		18.845,00
	Totaal bouwkosten (excl. BTW)	1.900.065,-	514.060,-	2.414.120,00

Bouwfysische kenmerken

Beganegrondvloer	R =	5,00	$m^2.K/W$
Geveiconstructie	R =	5,00	$m^2.K/W$
Glas	U =	1,10	$W/(m^2.K)$
Dakconstructie	R =	5,00	$m^2.K/W$
Energie Prestatie Coëfficiënt		0,60	

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	3	stuks	Dakoppervlakte, bruto	1.256	m^2
Maximale hoogte	10,71	m	Dakoppervlakte	1.256	m^2
Terrein			Dakopeningen		m^2
Terreinoppervlakte	7.360	m^2	Gevels		
Bebouwde terreinoppervlakte	1.220	m^2	Gevelomtrek	359	m
Onbebouwde terreinoppervlakte	6.104	m^2	Buitenwandoppervlakte, bruto	3.653	m^2
Oppervlakteverharding		m^2	Buitenwanden	3.088	m^2
Aantal parkeerplaatsen		stuks	Buitenwandopeningen	565	m^2
Ruimtedelen			Binnenwanden		
Bruto-vloeroppervlakte (BVO)	2.533	m^2	Binnenwandoppervlakte, bruto	2.537	m^2
Netto-vloeroppervlakte (NVO)	2.044	m^2	Binnenwanden	1.916	m^2
Installatieoppervlakte	33	m^2	Binnenwandopeningen	621	m^2
Verkeersoppervlakte	282	m^2	Trapconstructies		
Nuttige oppervlakte	1.729	m^2	Trappen en hellingen	6	stuks
Niet-toegankelijke ruimte		m^2	Balustrades en leuningen	184	m
Tarra-oppervlakte	489	m^2	Installaties		
Gebruiksoppervlakte	2.224	m^2	Klimaatinstallaties (koeling)		kW
Bruto-inhoud (BI)	9.108	m^3	Klimaatinstallaties (verwarming)	140	kW
Skelet			Warmtepomp		kW
Buitenwanden (constructief)	645	m^2	Zonnecollector		kW
Binnenwanden (constructief)	792	m^2	Vloerverwarming		m^2
Beganegrondvloeren (constructief)	1.220	m^2	Warmtelichamen		stuks
Verdiepingsvloeren (constructief)	1.313	m^2	Wandverwarming		m^2
Dakvloeren (constructief)	1.256	m^2	Liften	1	stuks
Hoofddraagconstructie	2.533	m^2			

Vormfactoren

Appartementencomplex de barchaan (source: bouwkosten.nl)

BOUWkosten

APPARTEMENTENCOMPLEX DE BARCHAAN TE ALMELO

TABEL | OMSCHRIJVING | AFBEELDINGEN

Omschrijving

Projectomschrijving
Op een voormalig fabrieksterrein verrezen koop- en huurwoningen, een appartementencomplex en een winkelcentrum. Een van deze gebouwen is De Barchaan, gebouwd in de jaren 30-stijl.
Het gebouw bestaat uit 4 bouwlagen. Op de begane grond aan twee zijden bevindt zich een commerciële ruimte van 360 m^2. Deze is vrij in te delen in kleinere segmenten. Verdeeld over 3 verdiepingen bevinden zich 14 luxe appartementen waar mensen met een verstandelijke beperking zelfstandig kunnen wonen. Tevens is er voor deze mensen een gemeenschappelijke ruimte beschikbaar voor onder andere dagactiviteiten.
Alle appartementen zijn 65 m^2 en 78 m^2 groot, zijn twee aan twee gespiegeld en hebben een eigen voordeur naar een inpandige galerij met lift en trappenhuis. Een lichtkoepel in het dak van de galerij en glazen lamellen voorzien voor voldoende daglicht. Het appartement bevat een woonkamer, keuken, slaapkamer, badkamer, een berging/hobbyruimte en een ruime loggiabalkon. Tevens bevat het een elux-systeem; een loos leidingsysteem voor alle toekomstige domotica-voorzieningen. De onderlinge scheidingswand van de gespiegelde wooneenheden bestaat uit metalstud wat in de toekomst relatief eenvoudig te verwijderen is, waardoor er één appartement met een woonoppervlakte van 130 m^2 ontstaat. De twee wooneenheden hebben een gemeenschappelijke meterkast en verwarmingsinstallatie. Eén van de twee wooneenheiden bevat een separaat gastentoilet.
Op de begane grond bevindt zich een technische ruimte met een warmtepomp voor een lage-temperatuurvloerverwarming en een koelsysteem voor het gehele complex.

Algemene projectgegevens
Samenstelling: Huls Bouwkostenmanagement
Opdrachtgever: Woningstichting St. Joseph te Almelo
Architect: 4D Architecten te Enter
Hoofdaannemer: Nijhuis Bouw te Rijssen

Peildatum originele kosten: juli 2011
Start uitvoering: juni 2012
Oplevering: juli 2013
Bouwtijd: 13 maanden
Aanbestedingsvorm: Bouwteam
Risicoregeling van toepassing: n.v.t.

Bruto-inhoud: 6.185 m^3
Alle kosten zijn geïndexeerd.

Kosten- en gegevenstabellen

KOSTEN OP NIVEAU 1 GEHEEL BOUWWERK | BOUWFYSISCHE KENMERKEN | HOEVEELHEDEN | VORMFACTOREN
KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS | KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS, HOOFDINDELING | KOSTEN OP NIVEAU 3 ELEMENTEN
OMSCHRIJVING ELEMENTEN | RUIMTESOORTEN, TOTAAL | RUIMTESOORTEN, SPECIFICATIE

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	1.757.955,00		1.757.955,00
3	Installaties		252.605,00	252.605,00
4	Vaste inrichtingen		75.860,00	75.860,00
5	Terrein	25.540,00		25.540,00

	Totaal directe bouwkosten	1.783.490,-	328.465,-	2.111.955,00
6	Algemene bouwkosten	304.350,00		304.350,00
	Subtotaal	2.087.845,-	328.465,-	2.416.310,-
6	Algemene bedrijfskosten	154.120,00		154.120,00
	Subtotaal	2.241.965,-	328.465,-	2.570.430,-
6	Winst en risico	70.675,00		70.675,00
	Totaal bouwkosten (excl. BTW)	2.312.640,-	328.465,-	2.641.105,00

Bouwfysische kenmerken

Beganegrondvloer	R =	3,00	$m^2.K/W$
Geveiconstructie	R =	7,00	$m^2.K/W$
Glas	U =	1,10	$W/(m^2.K)$
Dakconstructie	R =	7,00	$m^2.K/W$
Energie Prestatie Coëfficiënt			

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	4	stuks	Dakoppervlakte, bruto	656	m^2
Maximale hoogte	14,45	m	Dakoppervlakte	645	m^2
Terrein			Dakopeningen	11	m^2
Terreinoppervlakte	880	m^2	Gevels		
Bebouwde terreinoppervlakte	599	m^2	Geveiomtrek	132	m
Onbebouwde terreinoppervlakte	281	m^2	Buitenwandoppervlakte, bruto	1.528	m^2
Oppervlakteverharding		m^2	Buitenwanden	954	m^2
Aantal parkeerplaatsen		stuks	Buitenwandopeningen	574	m^2
Ruimtedelen			Binnenwanden		
Bruto-vloeroppervlakte (BVO)	1.649	m^2	Binnenwandoppervlakte, bruto	1.495	m^2
Netto-vloeroppervlakte (NVO)	1.344	m^2	Binnenwanden	1.258	m^2
Installatieoppervlakte	39	m^2	Binnenwandopeningen	237	m^2
Verkeersoppervlakte	219	m^2	Trapconstructies		
Nuttige oppervlakte	1.086	m^2	Trappen en hellingen	3	stuks
Niet-toegankelijke ruimte		m^2	Balustrades en leuningen	29	m
Tarra-oppervlakte	305	m^2	Installaties		
Gebruiksoppervlakte	1.364	m^2	Klimaatinstallaties (koeling)		kW
Bruto-inhoud (BI)	6.185	m^3	Klimaatinstallaties (verwarming)		kW
Skelet			Warmtepomp		kW
Buitenwanden (constructief)	1.409	m^2	Zonnecollector		kW
Binnenwanden (constructief)	361	m^2	Vloerverwarming		m^2
Beganegrondvloeren (constructief)	599	m^2	Warmtelichamen		stuks
Verdiepingsvloeren (constructief)	1.050	m^2	Wandverwarming		m^2
Dakvloeren (constructief)	656	m^2	Liften		stuks
Hoofddraagconstructie	1.649	m^2			

Vormfactoren

Woontoren met 40 appartementen (source: bouwkosten.nl)

BOUWkosten

WOONTOREN MET 40 APPARTEMENTEN TE ZWOLLE

TABEL | OMSCHRIJVING | AFBEELDINGEN

Omschrijving

Projectomschrijving
De entree van de wijk Holtenbroek wordt sinds kort gemarkeerd door een stevig, donkergrijs getint blok met koopappartementen. Op de begane grond van het gebouw bevinden zich naast de horecaruimte de hoofdentree van het gebouw en de toegang tot de bergingen. Aan de achterzijde van het gebouw zijn inpandig de lift en het trappenhuis gesitueerd.

De woontoren bevat veertig driekamerappartementen. Met per etage een compacte indeling van vijf appartementen waarvan één op elke hoek en één in het midden tegenover de lift en het trappenhuis. De driekamerappartementen hebben een woonkamer en twee slaapkamers met een oppervlak van 115 tot 118 m^2 BVO. Per etage heeft elk appartement een andere plattegrond. Elk appartement heeft een groot inpandig balkon dat vanuit de woning via de glazen schuifdeuren te bereiken is.

Er is gekozen voor donkergrijze baksteenbekleding en prefab onderdelen omdat de naast het appartementengebouw gelegen seniorenwoningen van vier lagen eveneens deze gevelbekleding hebben. Aan drie zijden van het gebouw zijn, voor elk appartement afzonderlijk en evenwijdig aan de gevel, verschuifbare schermen met oranjegekleurde, zonwerende doeken van circa 1,50 m breed geplaatst. Sommige zijn geplaatst voor de balkons, andere voor de ramen van de woonkamers. Omdat de schermen door de bewoners afzonderlijk te bedienen zijn, krijgt de gevel een variërende aanblik.

Algemene projectgegevens
Samenstelling: Huls Bouwkostenmanagement
Opdrachtgever: Projectteam Holtenbroek 1 te Zwolle
Architect: de Zwarte Hond te Rotterdam
Hoofdaannemer: Moes Bouwgroep te Zwolle

Peildatum originele kosten: februari 2006
Start uitvoering: april 2007
Oplevering: november 2007
Bouwtijd: 225 werkbare werkdagen
Aanbestedingsvorm: op uitnodiging (4)
Risicoregeling van toepassing: n.v.t.

Bruto-inhoud: 17.057 m^3
Alle kosten zijn geïndexeerd.

Kosten- en gegevenstabellen

KOSTEN OP NIVEAU 1 GEHEEL BOUWWERK | BOUWFYSISCHE KENMERKEN | HOEVEELHEDEN | VORMFACTOREN
KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS | KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS, HOOFDINDELING | KOSTEN OP NIVEAU 3 ELEMENTEN
OMSCHRIJVING ELEMENTEN | RUIMTESOORTEN, TOTAAL | RUIMTESOORTEN, SPECIFICATIE

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	3.398.160,00		3.398.160,00
3	Installaties		644.550,00	644.550,00
4	Vaste inrichtingen		72.020,00	72.020,00
5	Terrein	27.375,00		27.375,00

	Totaal directe bouwkosten	3.425.535,-	716.565,-	4.142.100,00
6	Algemene bouwkosten	462.375,00		462.375,00
	Subtotaal	3.887.910,-	716.565,-	4.604.475,-
6	Algemene bedrijfskosten	319.930,00		319.930,00
	Subtotaal	4.207.840,-	716.565,-	4.924.405,-
6	Winst en risico			
	Totaal bouwkosten (excl. BTW)	4.207.840,-	716.565,-	4.924.405,00

Bouwfysische kenmerken

Beganegrondvloer	R =	2,50	$m^2.K/W$
Gevelconstructie	R =	3,00	$m^2.K/W$
Glas	U =	1,20	$W/(m^2.K)$
Dakconstructie	R =	3,00	$m^2.K/W$
Energie Prestatie Coëfficiënt		1,00	

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	9	stuks	Dakoppervlakte, bruto	713	m^2
Maximale hoogte	27,50	m	Dakoppervlakte	712	m^2
Terrein			Dakopeningen	1	m^2
Terreinoppervlakte	1.808	m^2	Gevels		
Bebouwde terreinoppervlakte	673	m^2	Gevelomtrek	121	m
Onbebouwde terreinoppervlakte	1.134	m^2	Buitenwandoppervlakte, bruto	3.603	m^2
Oppervlakteverharding	1.121	m^2	Buitenwanden	1.999	m^2
Aantal parkeerplaatsen	49	stuks	Buitenwandopeningen	1.604	m^2
Ruimtedelen			Binnenwanden		
Bruto-vloeroppervlakte (BVO)	5.680	m^2	Binnenwandoppervlakte, bruto	6.304	m^2
Netto-vloeroppervlakte (NVO)	4.568	m^2	Binnenwanden	5.115	m^2
Installatieoppervlakte	30	m^2	Binnenwandopeningen	1.189	m^2
Verkeersoppervlakte	1.240	m^2	Trapconstructies		
Nuttige oppervlakte	3.294	m^2	Trappen en hellingen	8	stuks
Niet-toegankelijke ruimte	4	m^2	Balustrades en leuningen	813	m
Tarra-oppervlakte	1.112	m^2	Installaties		
Gebruiksoppervlakte	4.090	m^2	Klimaatinstallaties (koeling)		kW
Bruto-inhoud (BI)	17.057	m^3	Klimaatinstallaties (verwarming)		kW
Skelet			Warmtepomp		kW
Buitenwanden (constructief)	475	m^2	Zonnecollector		kW
Binnenwanden (constructief)	3.046	m^2	Vloerverwarming		m^2
Beganegrondvloeren (constructief)	673	m^2	Warmtelichamen		stuks
Verdiepingsvloeren (constructief)	5.007	m^2	Wandverwarming		m^2
Dakvloeren (constructief)	713	m^2	Liften	1	stuk
Hoofddraagconstructie	5.680	m^2			

Vormfactoren

Totaaloverzicht nieuwbouw (Kulmann, 2007)

Totaaloverzicht nieuwbouw; oorspronkelijke peildatum

Woongebouwen

Nieuwbouw - Woongebouwen in het algemeen

SfB-gebouw-code	Project	Aantal		Peildatum		Kosten	
		m^2 BVO	m^3 BI			Peildatum	Geïndexeerd
81	WONINGEN IN HET ALGEMEEN						
	17 appartementen met winkelruimte te Oudenbosch	1.608	4.838	Apr. 2002	Totaal	1.359.660,-	1.515.415,-
					Per m^2 BVO	846,-	942,-
					Per m^3 BI	281,-	313,-
	49 seniorenappartementen te Dronten	6.723	18.587	Apr. 2002	Totaal	3.677.005,-	4.098.220,-
					Per m^2 BVO	547,-	610,-
					Per m^3 BI	198,-	220,-
	15 stadswoningen te Maastricht	2.567	8.110	Okt. 2002	Totaal	2.039.045,-	2.243.440,-
					Per m^2 BVO	794,-	874,-
					Per m^3 BI	251,-	277,-
	Vrijstaande woning te Zundert	174	632	Aug. 2003	Totaal	259.170,-	277.025,-
					Per m^2 BVO	1.490,-	1.590,-
					Per m^3 BI	410,-	438,-
	Vrijstaande woning te Delfgauw	294	705	Nov. 2002	Totaal	307.835,-	338.690,-
					Per m^2 BVO	1.045,-	1.150,-
					Per m^3 BI	437,-	480,-
	Vrijstaande woning te Zaltbommel	223	646	Mei 2003	Totaal	214.810,-	234.385,-
					Per m^2 BVO	963,-	1.050,-
					Per m^3 BI	333,-	363,-
	Appartementen en winkels te Winterswijk	5.661	19.910	Mei 2003	Totaal	3.562.415,-	3.841.300,-
					Per m^2 BVO	629,-	679,-
					Per m^3 BI	179,-	193,-
	Vrijstaande woning te Schoonrewoerd	286	750	Mrt. 2005	Totaal	272.570,-	286.605,-
					Per m^2 BVO	953,-	1.000,-
					Per m^3 BI	363,-	382,-
84	BIJZONDERE WOONGEBOUWEN						
	Studentenhuisvesting te Dronten	3.580	10.728	Dec. 2003	Totaal	1.994.210,-	2.131.785,-
					Per m^2 BVO	557,-	595,-
					Per m^3 BI	186,-	199,-

17 appartementen Oudenbosch (source: Kuhlmann, 2007)

denbosch

appartementen in drie
bergingen en een aan-

winkelcentrum, waarin
n in 2 blokken naast el-
kort blok van 2 woon-
kelruimte aanwezig,
mgegeven wordt.
iche en toilet, apart
slaapkamer en hobby-
voor effectieve wo-
De badkamer is alleen
komen niet uit op een
eruggelegd waardoor

r	Overige aannemers	Totaal bouwkosten
		1.023.440,-
	213.005,-	213.005,-
	49.280,-	49.280,-
		42.255,-
	262.285,-	1.327.980,-
		46.570,-
	262.285,-	1.374.550,-
		54.985,-
	262.285,-	1.429.635,-
		85.780,-
	262.285,-	1.515.415,-

81 Woningen in het algemeen

17 appartementen met winkelruimte te Oudenbosch

Bouwfysische kenmerken

Begane grondvloer	R =	3,50	m^2.K/W
Gevelconstructie	R =	3,00	m^2.K/W
Glas	U =	1,80	W/(m^2.K)
Dakconstructie	R =	3,50	m^2.K/W
Energie Prestatie Coëfficiënt			

Hoeveelheden

Algemeen		
Aantal bouwlagen	3	stuks
Terrein		
Terreinoppervlakte	3.456	m^2
Bebouwde oppervlakte	585	m^2
Onbebouwde oppervlakte	2.871	m^2
Oppervlakteverharding	2.304	m^2
Aantal parkeerplaatsen	65	stuks
Ruimtedelen		
Bruto-vloeroppervlakte (BVO)	1.608	m^2
Netto-vloeroppervlakte (NVO)	1.359	m^2
Installatieoppervlakte	17	m^2
Verkeersoppervlakte	353	m^2
Nuttige oppervlakte	989	m^2
Niet-toegankelijke ruimte		m^2
Tara-oppervlakte	249	m^2
Gebruiksoppervlakte	1.308	m^2
Bruto-inhoud	4.838	m^3
Skelet		
Buitenwanden (constructief)	916	m^2
Binnenwanden (constructief)	1.246	m^2
Begane grondvloeren (constructief)	487	m^2
Verdiepingsvloeren (constructief)	976	m^2
Dakvloeren (constructief)	554	m^2
Hoofddraagconstructie	1.608	m^2

Daken		
Dakoppervlakte, bruto	755	m^2
Dakoppervlakte	637	m^2
Dakopeningen	117	m^2
Gevels		
Gevelomtrek	232	m
Buitenwandoppervlakte, bruto	1.389	m^2
Buitenwanden	1.009	m^2
Buitenwandopeningen	380	m^2
Binnenwanden		
Binnenwandoppervlakte, bruto	2.121	m^2
Binnenwanden	1.790	m^2
Binnenwandopeningen	331	m^2
Trapconstructies		
Trappen en hellingen	4	stuks
Balustrades en leuningen	142	m
Installaties		
Klimaatinstallaties (koeling)		kW
Klimaatinstallaties (verwarming)		kW
Warmtepomp		kW
Zonnecollector		kW
Vloerverwarming		m^2
Warmtelichamen		stuks
Wandverwarming		m^2
Liften	1	stuk

Vormfactoren

Bruto-inhoud/ bruto-vloeroppervlakte (BVO)	=	4.838	/	1.608	=	3,01
Netto-vloeroppervlakte (NVO)/ bruto-vloeroppervlakte (BVO)	=	1.359	/	1.608	=	0,84
Tara-oppervlakte/ bruto-vloeroppervlakte (BVO)	=	249	/	1.608	=	0,16
Gebruiksoppervlakte/ bruto-vloeroppervlakte (BVO)	=	1.308	/	1.608	=	0,81
Nuttige oppervlakte/ bruto-vloeroppervlakte (BVO)	=	989	/	1.608	=	0,62
Bebouwde oppervlakte/ bruto-vloeroppervlakte (BVO)	=	585	/	1.608	=	0,36
Buitenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	1.389	/	1.608	=	0,86
Dakoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	755	/	1.608	=	0,47
Buitenwand- en dakoppervlak/ bruto-vloeroppervlakte (BVO)	=	2.144	/	1.608	=	1,33
Binnenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	2.121	/	1.608	=	1,32
Bebouwde oppervlakte/ nuttige oppervlakte	=	585	/	989	=	0,59
Bruto-vloeroppervlakte (BVO)/ nuttige oppervlakte	=	1.608	/	989	=	1,63
Buitenwandoppervlakte, bruto/ nuttige oppervlakte	=	1.389	/	989	=	1,40
Dakoppervlakte, bruto/ nuttige oppervlakte	=	755	/	989	=	0,76
Buitenwand- en dakoppervlak/ nuttige oppervlakte	=	2.144	/	989	=	2,17
Binnenwandoppervlakte, bruto/ nuttige oppervlakte	=	2.121	/	989	=	2,14
Bebouwde oppervlakte/ gebruiksoppervlakte	=	585	/	1.308	=	0,45
Bruto-vloeroppervlakte (BVO)/ gebruiksoppervlakte	=	1.608	/	1.308	=	1,23
Buitenwandoppervlakte, bruto/ gebruiksoppervlakte	=	1.389	/	1.308	=	1,06
Dakoppervlakte, bruto/ gebruiksoppervlakte	=	755	/	1.308	=	0,58
Buitenwand- en dakoppervlak/ gebruiksoppervlakte	=	2.144	/	1.308	=	1,64
Binnenwandoppervlakte, bruto/ gebruiksoppervlakte	=	2.121	/	1.308	=	1,62

81

49 seniorenappartementen Dronten (source: Kuhlmann, 2007)

81 Woningen in het algemeen

49 seniorenappartementen te Dronten

Projectomschrijving
Het appartementencomplex 'Het Tussendek' staat in Dronten op de grens van het centrum, winkelgebied en de nabij gelegen woonwijken. Het complex heeft ook voorzieningen zoals een gezondheidscentrum, de bibliotheek, het cultureel centrum De Meerpaal en het openbaar vervoer in de onmiddellijke omgeving. Het is dan ook bij uitstek geschikt voor senioren en daarom ook mede toegankelijk voor personen die een scootmobiel gebruiken.
Het complex bestaat uit drie woonvleugels die een gemeenschappelijke tuin omsluiten. Op de begane grond bevinden zich de hoofdentree, een recreatieruimte, één berging voor ieder appartement en een aantal appartementen.
Over de eerste tot en met de derde verdieping zijn de resterende appartementen verdeeld en per verdieping is er een bergingsruimte voor scootmobielen aanwezig. De op de omgeving uitkijkende appartementen hebben middels een loggia hun eigen omsloten buitenruimte.
De entrees van de woningen zijn te bereiken via galerijen, gelegen aan de omsloten binnenruimte met zicht op de gemeenschappelijke tuin. Verder completeren voorzieningen zoals videofoon, lift en laagdrempeligheid (maximaal 20 mm) het geheel. De parkeerplaatsen met de plaatsen voor mindervaliden zijn direct bij de hoofdingang aanwezig.
Door zijn verfrissende architectuur, uitgevoerd in een combinatie van metselwerk, stucwerk en verschillende staalbeplatingen, vormt het complex een passende aanvulling in de vernieuwing van het centrumgebied.

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	2.644.705,-		2.644.705,-
3	Installaties		721.810,-	721.810,-
4	Vaste inrichtingen		124.790,-	124.790,-
5	Terrein	45.875,-		45.875,-
	Totaal bouwkosten elementen	2.690.580,-	846.595,-	3.537.175,-
6	Algemene bouwkosten (bouwproject)	232.205,-		232.205,-
	Subtotaal	2.922.785,-	846.595,-	3.769.380,-
6	Algemene bedrijfskosten (bouwbedrijf)	328.840,-		328.840,-
	Subtotaal	3.251.625,-	846.595,-	4.098.220,-
6	Winst en risico (bouwbedrijf)			
	Totaal bouwkosten exclusief BTW	3.251.625,-	846.595,-	4.098.220,-

nanagement

oondiensten te

Seckel v.d. Dob-
ten te Hengelo

ddrijf BV te

ele kosten

rkdagen

:587 m³
peindexeerc.

81 Woningen in het algemeen

49 seniorenappartementen te Dronten

Bouwfysische kenmerken

Begaanegrondvloer	R =	4,00	m².K/W
Gevelconstructie	R =	3,05	m².K/W
Glas	U =	1,60	W/(m².K)
Dakconstructie	R =	4,50	m².K/W
Energie Prestatie Coëfficiënt		0,94	

Hoeveelheden

Algemeen			**Daken**		
Aantal bouwlagen	4	stuks	Dakoppervlakte, bruto	1.446	m²
Terrein			Dakoppervlakte	1.445	m²
Terreinoppervlakte	3.812	m²	Dakopeningen	1	m²
Bebouwde oppervlakte	1.581	m²	**Gevels**		
Overbouwde oppervlakte	2.231	m²	Gevelomtrek	284	m
Oppervlakteverharding	526	m²	Buitenwandoppervlakte, bruto	3.425	m²
Aantal parkeerplaatsen	2	stuks	Buitenwanden	2.203	m²
Ruimtelijks			Buitenwandopeningen	1.222	m²
Bruto-vloeroppervlakte (BVO)	6.723	m²	Binnenwanden		
Netto-vloeroppervlakte (NVO)	5.941	m²	Binnenwandoppervlakte, bruto	4.919	m²
Installatieoppervlakte	82	m²	Binnenwanden	3.715	m²
Verkeersoppervlakte	1.444	m²	Binnenwandopeningen	1.204	m²
Nuttige oppervlakte	4.415	m²	Trapconstructies		
Niet-toegankelijke ruimte		m²	Trappen en hellingen	3	stuks
Tuin-oppervlakte	782	m²	Balustrades en leuningen	386	m
Gebruiksoppervlakte	5.313	m²	**Installaties**		
Bruto-inhoud	18.587	m³	Klimaatinstallaties (koeling)		kW
Skelet			Klimaatinstallaties (verwarming)		kW
Buitenwanden (constructief)	563	m²	Warmtepomp		kW
Binnenwanden (constructief)	2.209	m²	Zonnecollector		kW
Begaanegrondvloeren (constructief)	1.463	m²	Vloerverwarming		m²
Verdiepingvloeren (constructief)	4.785	m²	Warmtelichamen		stuks
Dakvloeren (constructief)	1.231	m²	Wandverwarming		m²
Draagconstructie	6.723	m²	Liften	1	stuk

Vormfactoren

Bruto-inhoud/ bruto-vloeroppervlakte (BVO)	=	18.587	/	6.723	=	2,76
Netto-vloeroppervlakte (NVO)/ bruto-vloeroppervlakte (BVO)	=	5.941	/	6.723	=	0,88
Tuin-oppervlakte/ bruto-vloeroppervlakte (BVO)	=	782	/	6.723	=	0,12
Gebruiksoppervlakte/ bruto-vloeroppervlakte (BVO)	=	5.313	/	6.723	=	0,79
Nuttige oppervlakte/ bruto-vloeroppervlakte (BVO)	=	4.415	/	6.723	=	0,66
Bebouwde oppervlakte/ bruto-vloeroppervlakte (BVO)	=	1.581	/	6.723	=	0,24
Buitenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	3.425	/	6.723	=	0,51
Dakoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	1.446	/	6.723	=	0,22
Buitenwand- en dakoppervlak/ bruto-vloeroppervlakte (BVO)	=	4.871	/	6.723	=	0,72
Binnenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	4.919	/	6.723	=	0,73
Bebouwde oppervlakte/ nuttige oppervlakte	=	1.581	/	4.415	=	0,36
Bruto-vloeroppervlakte (BVO)/ nuttige oppervlakte	=	6.723	/	4.415	=	1,52
Buitenwandoppervlakte, bruto/ nuttige oppervlakte	=	3.425	/	4.415	=	0,78
Dakoppervlakte, bruto/ nuttige oppervlakte	=	1.446	/	4.415	=	0,33
Buitenwand- en dakoppervlak/ nuttige oppervlakte	=	4.871	/	4.415	=	1,10
Binnenwandoppervlakte, bruto/ nuttige oppervlakte	=	4.919	/	4.415	=	1,11
Bebouwde oppervlakte/ gebruiksoppervlakte	=	1.581	/	5.313	=	0,30
Bruto-vloeroppervlakte (BVO)/ gebruiksoppervlakte	=	6.723	/	5.313	=	1,27
Buitenwandoppervlakte, bruto/ gebruiksoppervlakte	=	3.425	/	5.313	=	0,64
Dakoppervlakte, bruto/ gebruiksoppervlakte	=	1.446	/	5.313	=	0,27
Buitenwand- en dakoppervlak/ gebruiksoppervlakte	=	4.871	/	5.313	=	0,92
Binnenwandoppervlakte, bruto/ gebruiksoppervlakte	=	4.919	/	5.313	=	0,93

15 stadwoningen Maastricht (source: Kuhlmann, 2007)

Projectomschrijving
Het project betreft de nieuwbouw van 15 stadswoningen in 3 lagen. De woningen maken onderdeel uit van de herontwikkeling van een voormalig bedrijventerrein in de stad. De wijk heet nu Alegonda park en bestaat uit 107 woningen.
In de woningen liggen op de begane grond de entree, keuken, woonkamer en een erker die doorloopt naar de 1e etage. Op de 1e etage is een woon-/werkkamer, met vide en een slaapkamer. Op de bovenste halve verdieping zijn 2 slaapkamers.
Er is een uitgebreid optiepakket waarmee de woningen naar wens ingedeeld kunnen worden. Zo kan de 1e etage een doorlopende woonkamer worden. In de pui aan de achtergevel is een tussenstuk aangebracht waartegen een tussenmuur geplaatst kan worden. Er zijn daar 2 radiatoren geplaatst en 2 draaikiepramen, zodat beide kamers - als die optie wordt gekozen - gelucht kunnen worden. In alle opties zit de trap in dezelfde sparing. Daar zijn ook de leidingschachten aangebracht. De bovenste halve verdieping kan worden uitgebreid.

Samenstelling
Winket voor de bouw te Roosendaal
Opdrachtgever
Janssen de Jong Bouw Plancoördinatie bv te Son en Breugel
Architect
Architectenbureau Humlé Neuhof te Maastricht
Hoofdaannemer
Janssen de Jong Bouw te Kerkrade

Peildatum originele kosten
Oktober 2002
Start uitvoering
Januari 2003
Oplevering
Februari 2004
Bouwtijd
220 dagen

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	1.445.030,-		1.445.030,-
3	Installaties		224.145,-	224.145,-
4	Vaste inrichtingen		69.490,-	69.490,-
5	Terrein	27.440,-		27.440,-
	Totaal bouwkosten elementen	1.472.470,-	293.635,-	1.766.105,-
6	Algemene bouwkosten (bouwproject)	224.300,-		224.300,-
	Subtotaal	1.696.770,-	293.635,-	1.990.405,-
6	Algemene bedrijfskosten (bouwbedrijf)	144.945,-		144.945,-
	Subtotaal	1.841.720,-	293.635,-	2.135.350,-
6	Winst en risico (bouwbedrijf)	108.090,-		108.090,-
	Totaal bouwkosten exclusief BTW	1.949.805,-	293.635,-	2.243.440,-

81

81 Woningen in het algemeen

15 stadswoningen te Maastricht

Bouwfysische kenmerken

Beganegrondvloer	R =	3,00	$m^2.K/W$
Gevelconstructie	R =	3,00	$m^2.K/W$
Glas	U =	1,60	$W/(m^2.K)$
Dakconstructie	R =	3,00	$m^2.K/W$
Energie Prestatie Coëfficiënt			

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	3	stuks	Dakoppervlakte, bruto	1.467	m^2
Terrein			Dakoppervlakte	1.467	m^2
Terreinoppervlakte	2.313	m^2	Dakopeningen		m^2
Bebouwde oppervlakte	1.211	m^2	Gevels		
Onbebouwde oppervlakte	1.102	m^2	Gevelomtrek	139	m
Oppervlakteverharding	200	m^2	Buitenwandoppervlakte, bruto	2.096	m^2
Aantal parkeerplaatsen	15	stuks	Buitenwanden	1.450	m^2
Ruimtedelen			Buitenwandopeningen	646	m^2
Bruto-vloeroppervlakte (BVO)	2.567	m^2	Binnenwanden		
Netto-vloeroppervlakte (NVO)	1.922	m^2	Binnenwandoppervlakte, bruto	3.081	m^2
Installatieoppervlakte	16	m^2	Binnenwanden	2.682	m^2
Verkeersoppervlakte	302	m^2	Binnenwandopeningen	399	m^2
Nuttige oppervlakte	1.604	m^2	Trapconstructies		
Niet-toegankelijke ruimte		m^2	Trappen en hellingen	31	stuks
Tarra-oppervlakte	645	m^2	Balustrades en leuningen	261	m
Gebruiksoppervlakte	2.163	m^2	Installaties		
Bruto-inhoud	8.110	m^3	Klimaatinstallaties (koeling)		kW
Skelet			Klimaatinstallaties (verwarming)		kW
Buitenwanden (constructief)	1.338	m^2	Warmtepomp		kW
Binnenwanden (constructief)	1.785	m^2	Zonnecollector		kW
Beganegrondvloeren (constructief)	981	m^2	Vloerverwarming		m^2
Verdiepingsvloeren (constructief)	1.412	m^2	Warmtelichamen		stuks
Dakvloeren (constructief)	1.393	m^2	Wandverwarming		m^2
Hoofddraagconstructie	2.567	m^2	Liften		stuks

Vormfactoren

81

Bruto-inhoud/ bruto-vloeroppervlakte (BVO)	=	8.110	/	2.567	=	3,16
Netto-vloeroppervlakte (NVO)/ bruto-vloeroppervlakte (BVO)	=	1.922	/	2.567	=	0,75
Tarra-oppervlakte/ bruto-vloeroppervlakte (BVO)	=	645	/	2.567	=	0,25
Gebruiksoppervlakte/ bruto-vloeroppervlakte (BVO)	=	2.163	/	2.567	=	0,84
Nuttige oppervlakte/ bruto-vloeroppervlakte (BVO)	=	1.604	/	2.567	=	0,62
Bebouwde oppervlakte/ bruto-vloeroppervlakte (BVO)	=	1.211	/	2.567	=	0,47
Buitenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	2.096	/	2.567	=	0,82
Dakoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	1.467	/	2.567	=	0,57
Buitenwand- en dakoppervlak/ bruto-vloeroppervlakte (BVO)	=	3.563	/	2.567	=	1,39
[illegible]	=	3.081	/	2.567	=	1,20

Appartementen Winterswijk (source: Kuhlmann, 2007)

81

Samenstelling
Huls Bouwkostenmanagement te Zwolle
Opdrachtgever
De Woonplaats
Woningcorporatie te Groenlo
Architect
Architektenburo Ghijsen Partners B.V. te Deventer
Hoofdaannemer
Bouwcombinatie WAM en Van Duren / Thomasson Dura V.O.F. te Winterswijk

Peildatum originele kosten
Mei 2003

81 Woningen in het algemeen

Appartementen en winkels te Winterswijk

Projectomschrijving
Een voormalig terrein van Gemeentewerken aan de Spoorstraat Winterwijk heeft een nieuwe bestemming gekregen. Er is plaatsgemaakt voor een supermarkt van 1.490 m^2, een aantal winkels en daarboven twee lagen met koop- en huurappartementen en zelfstandige woningen voor bewoners met een zorgindicatie.
Naast dit gebouw is een parkeerterrein aangelegd. De locatie is gedeeltelijk afgezoomd met een muur die de achtertuinen van de omliggende woningen afschermt van dit parkeerterrein. De winkels op de begane grond worden casco opgeleverd. De constructie bestaat in hoofdzaak uit betonkolommen, betonschijven en enkele betonwanden op de begane grond met betonbalken onder de eerste verdiepingsvloer. Op de verdieping woningscheidende betonwanden. De beganegrondvloer is een kanaalplaatvloer. De verdiepings- en dakvloeren zijn bekistingsplaatvloeren. Op de tweede verdieping prefab betonnen galerijplaten. Op de eerste verdieping worden de woningen bereikt door looppaden op het dak. Het middengebied van het dak op de eerste verdieping wordt ingericht als daktuin.

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	2.445.320,-		2.445.320,-
3	Installaties		432.195,-	432.195,-
4	Vaste inrichtingen		65.075,-	65.075,-
5	Terrein	56.635,-		56.635,-
	Totaal bouwkosten elementen	2.501.955,-	497.270,-	2.999.225,-
6	Algemene bouwkosten (bouwproject)	464.025,-		464.025,-
	Subtotaal	2.965.980,-	497.270,-	3.463.250,-
6	Algemene bedrijfskosten (bouwbedrijf)	248.170,-		248.170,-
	Subtotaal	3.214.150,-	497.270,-	3.711.420,-
6	Winst en risico (bouwbedrijf)	129.880,-		129.880,-
	Totaal bouwkosten exclusief BTW	3.344.030,-	497.270,-	3.841.300,-

81 Woningen in het algemeen

Appartementen en winkels te Winterswijk

Bouwfysische kenmerken

Beganegrondvloer	R =	2,50	m^2.K/W
Gevelconstructie	R =	3,00	m^2.K/W
Glas	U =	1,20	W/(m^2.K)
Dakconstructie	R =	3,00	m^2.K/W
Energie Prestatie Coëfficiënt		0,75 / 2,15	

Hoeveelheden

Algemeen		
Aantal bouwlagen	3	stuks
Terrein		
Terreinoppervlakte	7.260	m^2
Bebouwde oppervlakte	2.451	m^2
Onbebouwde oppervlakte	4.809	m^2
Oppervlakteverharding		m^2
Aantal parkeerplaatsen		stuks
Ruimtedelen		
Bruto-vloeroppervlakte (BVO)	5.661	m^2
Netto-vloeroppervlakte (NVO)	4.820	m^2
Installatieoppervlakte	52	m^2
Verkeersoppervlakte	603	m^2
Nuttige oppervlakte	4.165	m^2
Niet-toegankelijke ruimte		m^2
Terra-oppervlakte	841	m^2
Gebruiksoppervlakte	4.735	m^2
Bruto-inhoud	19.910	m^3
Skelet		
Buitenwanden (constructief)		m^2
Binnenwanden (constructief)	1.233	m^2
Beganegrondvloeren (constructief)	2.451	m^2
Verdiepingsvloeren (constructief)	3.210	m^2
Dakvloeren (constructief)	2.451	m^2
Hoofddragconstructie	5.661	m^2

Daken		
Dakoppervlakte, bruto	2.451	m^2
Dakoppervlakte	2.440	m^2
Dakopeningen	11	m^2
Gevels		
Gevelomtrek	206	m
Buitenwandoppervlakte, bruto	2.554	m^2
Buitenwanden	1.938	m^2
Buitenwandopeningen	616	m^2
Binnenwanden		
Binnenwandoppervlakte, bruto	2.521	m^2
Binnenwanden	1.672	m^2
Binnenwandopeningen	849	m^2
Trapconstructies		
Trappen en hellingen	11	stuks
Balustrades en leuningen	168	m
Installaties		
Klimaatinstallaties (koeling)		kW
Klimaatinstallaties (verwarming)		kW
Warmtepomp		kW
Zonnecollector		kW
Vloerverwarming	1.949	m^2
Warmtelichamen		stuks
Wandverwarming		m^2
Liften	2	stuks

Vormfactoren

Bruto-inhoud/ bruto-vloeroppervlakte (BVO)	=	19.910	/	5.661	= 3,52
Netto-vloeroppervlakte (NVO)/ bruto-vloeroppervlakte (BVO)	=	4.820	/	5.661	= 0,85
Terra-oppervlakte/ bruto-vloeroppervlakte (BVO)	=	841	/	5.661	= 0,15
Gebruiksoppervlakte/ bruto-vloeroppervlakte (BVO)	=	4.735	/	5.661	= 0,84
Nuttige oppervlakte/ bruto-vloeroppervlakte (BVO)	=	4.165	/	5.661	= 0,74
Bebouwde oppervlakte/ bruto-vloeroppervlakte (BVO)	=	2.451	/	5.661	= 0,43
Buitenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	2.554	/	5.661	= 0,45
Dakoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	2.451	/	5.661	= 0,43
Buitenwand- en dakoppervlak/ bruto-vloeroppervlakte (BVO)	=	5.005	/	5.661	= 0,88
Binnenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	2.521	/	5.661	= 0,45
Bebouwde oppervlakte/ nuttige oppervlakte	=	2.451	/	4.165	= 0,59
Bruto-vloeroppervlakte (BVO)/					

Appendix 5: Price index

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2002	89.47	89.64	89.81	89.98	90.43	90.87	91.32	91.33	91.34	91.35	90.69	90.02
2003	89.36	89.67	89.99	90.3	90.37	90.45	90.52	90.58	90.64	90.7	90.9	91.11
2004	91.31	91.29	91.28	91.26	91.58	91.91	92.23	92.49	92.74	93	93.18	93.37
2005	93.55	93.6	93.66	93.71	93.81	93.91	94.01	94.15	94.28	94.42	94.55	94.68
2006	94.81	95	95.19	95.38	95.65	95.91	96.18	96.67	97.16	97.65	98.43	99.22
2007	100	100.22	100.43	100.65	101.01	101.37	101.73	102.01	102.29	102.57	102.81	103.04
2008	103.28	103.99	104.71	105.42	105.61	105.79	105.98	105.76	105.53	105.31	104.88	104.45
2009	104.02	104.01	104.01	104	103.56	103.13	102.69	102.69	102.7	102.7	102.16	101.63
2010	101.09	101.06	101.04	101.01	100.95	100.9	100.84	100.66	100.47	100.29	100.06	99.82
2011	99.59	99.89	100.2	100.5	100.61	100.71	100.82	100.42	100.02	99.62	99.71	99.8
2012	99.89	99.98	100.07	100.16	100.18	100.19	100.21	100.24	100.26	100.29	100.34	100.38
2013	100.43	100.55	100.68	100.8	100.8	100.8	100.8	101.05	101.31	101.56	101.89	102.21
2014	102.54	102.5	102.45	102.41	102.52	102.64	102.75	102.89	103.03	103.17	103.52	103.86
2015	104.21	104.31	104.42	104.52	104.57	104.63	104.68	104.8	104.93	105.05	105.28	105.51
2016	105.74	105.93	106.12	106.31	106.38	106.45	106.52	106.63	106.75	106.86	107.02	107.19
2017	107.35											

Figure 12: Price index residential buildings (source: Vonk, Wilde & Groot, 2016)

Appendix 6: Graphical representation performance measures database A

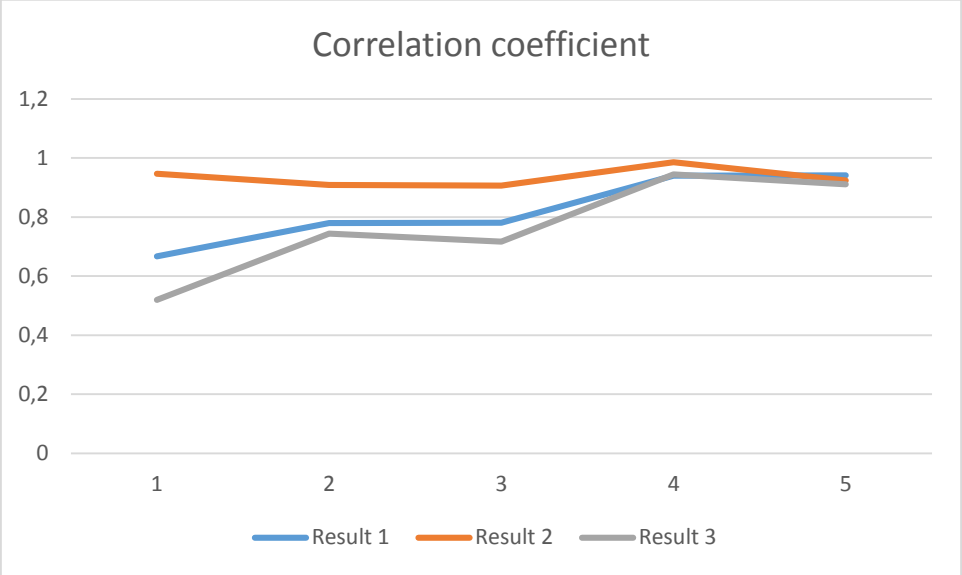

Figure 13: Correlation coefficient of database A

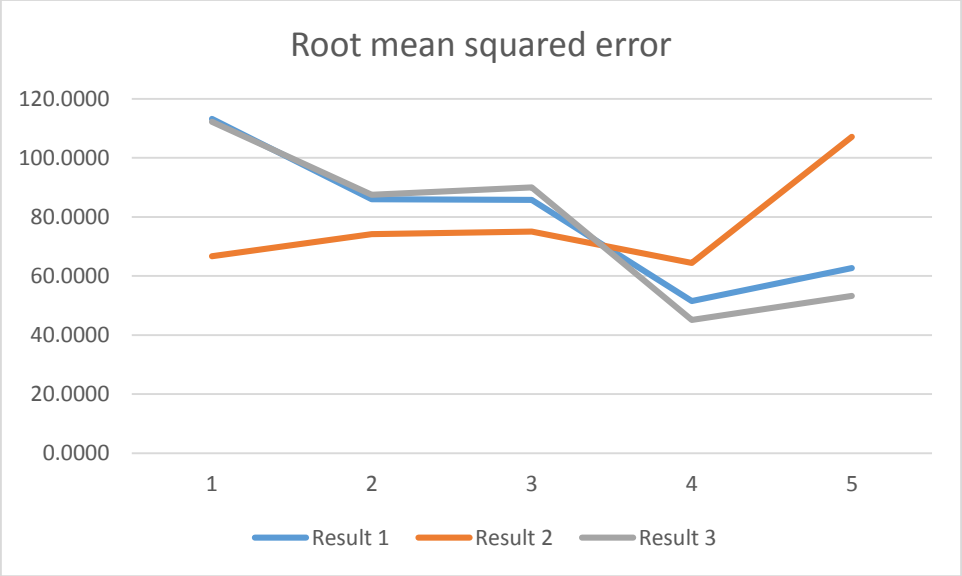

Figure 14: Root mean squared error of database A

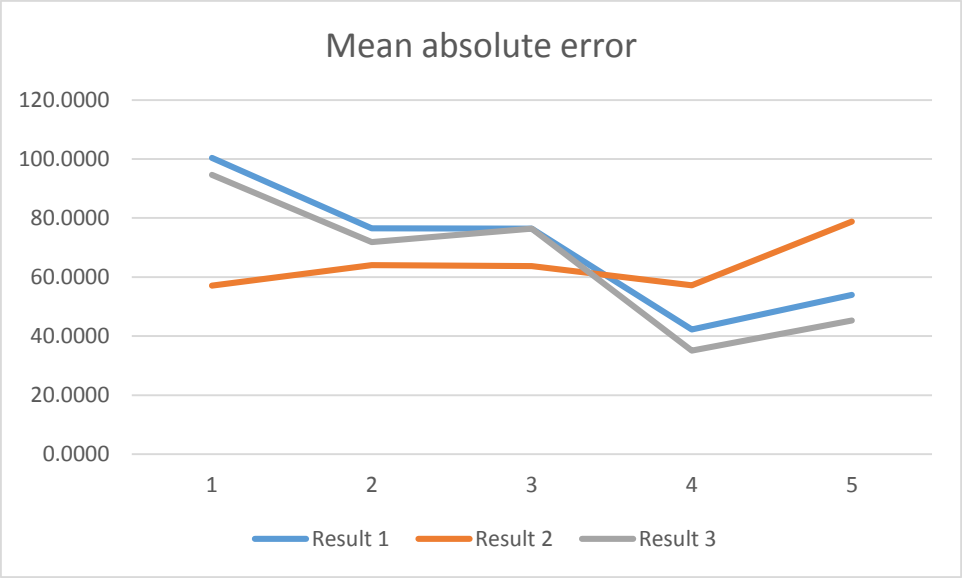

Figure 15: Mean absolute error of database A

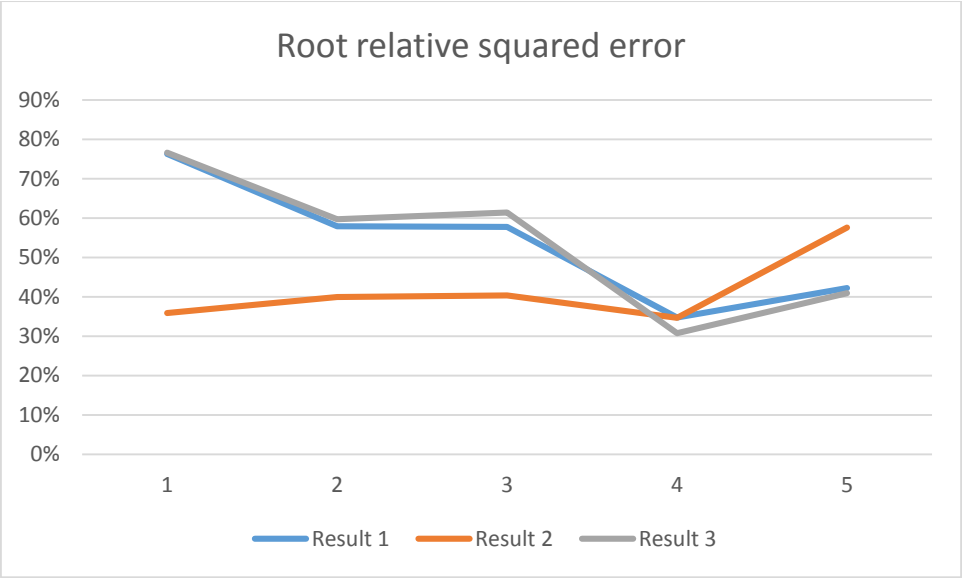

Figure 16: Root relative squared error of database A

Appendix 7: Graphical representation performance measures database B

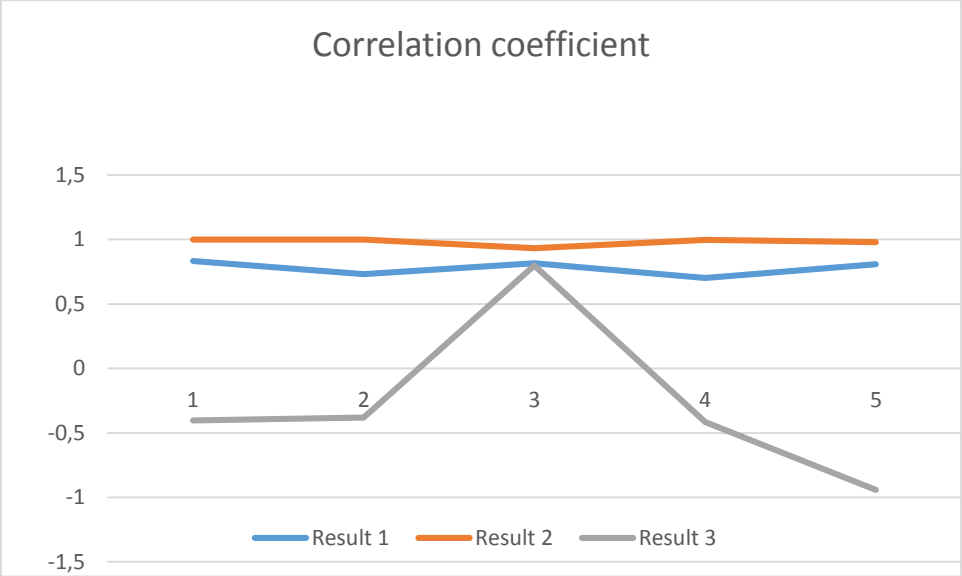

Figure 17: Correlation coefficient of database B

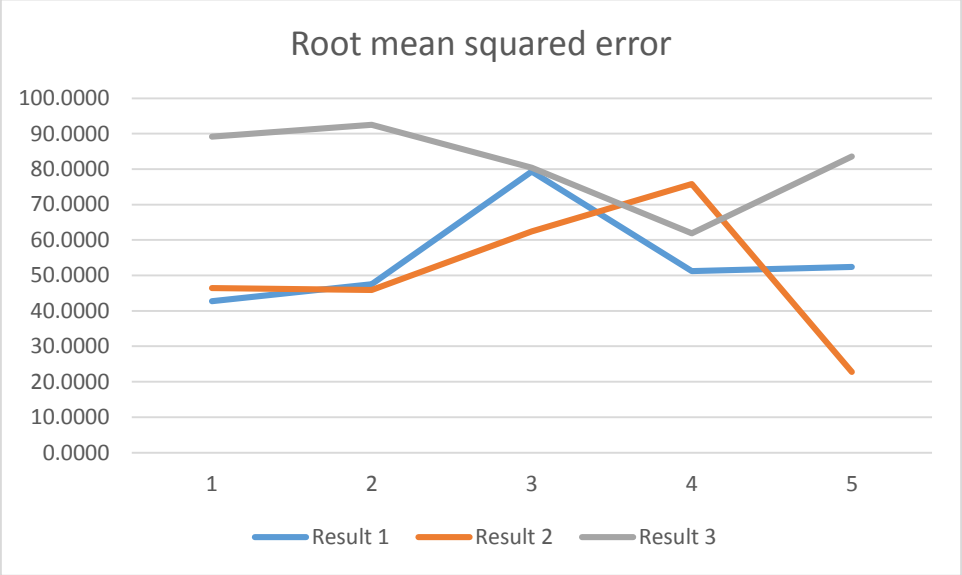

Figure 18: Root mean squared error of database B

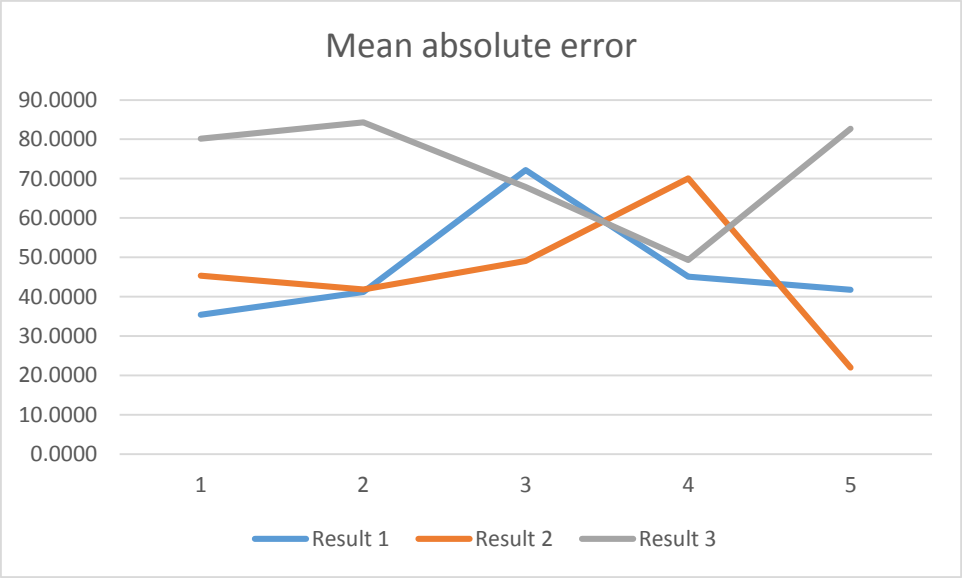

Figure 19: Mean absolute error of database B

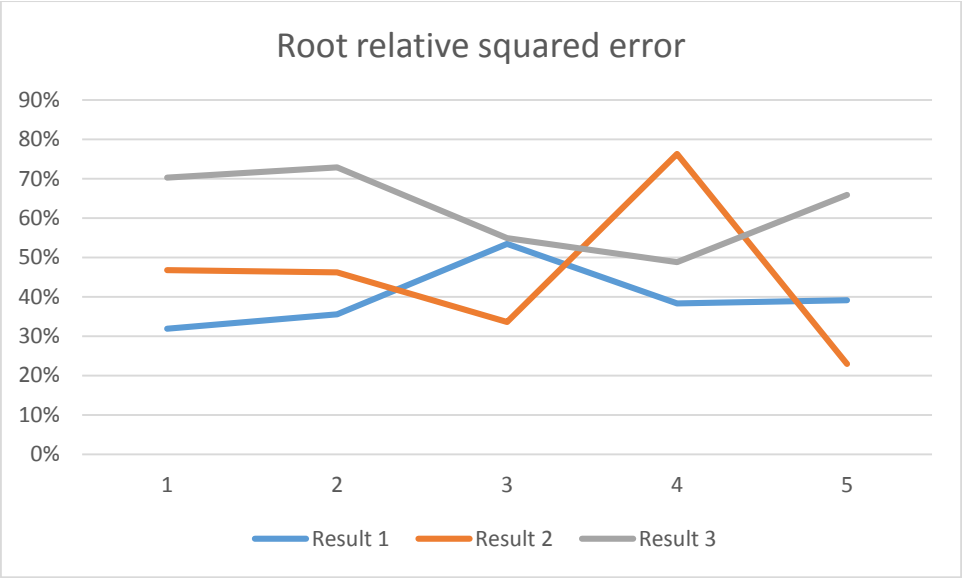

Figure 20: Root relative squared error of database B

This page is intentionally left blank

Colofon

Title: BIM Based Cost Estimation Knowledge Management

Subtitle: The application of a Knowledge Management System by combining BIM models and the data mining process to estimate the costs of residential buildings

Author:	Y. (Yonne) Rekveld
Student number:	0916003
E-mail:	yrekveld@gmail.com
University:	Eindhoven University of Technology
Graduation program:	Master Construction Management and Engineering

Graduation date: 11th of May 2017

Graduation committee:
Prof. Dr. Ir. B. (Bauke) de Vries
University supervisor (chairman graduation committee)
b.d.vries@tue.nl

Dr. Dipl.-Ing. J. (Jakob) Beetz
University supervisor
j.beetz@tue.nl

Prof. Dr. C.C.P. (Chris) Snijders
University supervisor
c.c.p.snijders@tue.nl

Company:
IBIS BV
Groot Handelsgebouw
Stationsplein 45, unit A6.002
3013 AK Rotterdam

J. (Jeroen) van de Laar
External graduation supervisor
j.van.de.laar@ibis.nl

T. (Toine) Bullens
External graduation supervisor
t.bullens@brink.nl

This page is intentionally left blank

Summary

Due to a change in what leading clients demand from the architecture, engineering and construction (AEC) industry, the AEC industry has to work more efficient. A possible solution is to apply knowledge management (KM) within the organisation. However, the problem with traditional KM techniques used in the AEC industry is that they demand too much maintenance and they only store but do not share information within the organisation. With the development of Building Information Modelling (BIM), a new possibility to create a knowledge management system has emerged. BIM provides a unique source of information as it generates and manages the data created during the building life cycle. This unique source of information can be converted into knowledge by means of data mining.

The problem this research has focused on, is that the current design process of the AEC industry is inefficient. One of the reasons is that the knowledge management systems (KMS) within the AEC industry do not meet the requirements necessary to be effective. The design process is very broad and has multiple tasks that can be made more efficient with KM. As this research is based on the limitations of the research done by Hoogeveen (2015), the KMS developed in this research has the same tasks, namely cost prediction in the early stage of the design process.

Hoogeveen (2015) proved part of this concept by using a data mining approach to predict the costs of datacentres, however he did not use the actual BIM models in his data mining process. In this research BIM models are used for automated data extraction for the data mining process. This data mining process is used to create a knowledge management system for conceptual cost estimation for residential buildings.

Based on this problem definition the following main research question has been drafted: “Can the use of a knowledge management system using BIM data and the data mining process create a more efficient process of cost estimation for residential buildings?” To answer these question a knowledge management system was created based on a literature review, which was tested with collected data.

In the literature review it was found that knowledge management can be defined as a tool set for the automation of deductive or inherent relationships between information objects, corporate users and business processes. To understand knowledge management, at first it had to be understood what knowledge is. Knowledge is defined as refined information, which means that it has meaning. Knowledge can be divided in two forms, tacit or explicit knowledge. The main difference between tacit and explicit knowledge is that explicit knowledge is documented, where tacit knowledge is implied. However, for knowledge to be useful for KM it needed to be made explicit. Especially because the goal of KM is to create new knowledge. This can only be done by combining multiple forms of explicit knowledge.

The suggested source for KM in this research is BIM. BIM is an intelligent simulation of architecture that exhibits the following six key characteristics, digital, spatial (3D), measurable, comprehensive, accessible and durable (Campbell 2006). The main product of BIM is a BIM model which is a virtual prototype of the designed building. A BIM model is characterized by building components that include data that describe how they behave and are consistent and non-redundant data (Eastman et al., 2011). There are two types of BIM models, a BIM model with a proprietary file format and a BIM model with an universal supported standard, generally the IFC file format. In this research the IFC data format was used.

To convert the data within the IFC models to knowledge the data mining process was used. Applied to the problem of converting data within the IFC model to knowledge in form of a system to estimate the costs of residential buildings in the early stage of the design process, the data mining process consists of the following steps:

1. Understanding and defining the problem
2. Collect necessary data
3. Convert non-IFC BIM models to the IFC data format
4. Convert the IFC models to a dataset
5. Add the corresponding costs
6. Clean the dataset
7. Select a useful subset
8. Engineer a useful data mining algorithm
9. Use the algorithm
10. Evaluate results with numerical prediction measures
11. Redefine data and problem
12. Use the generated model for cost prediction

The main obstacle within this research was the extraction of the information from BIM models into a dataset. In theory it should be possible to extract all physical data of the buildings from the BIM models. However during the practical implementation, the current quality of the BIM models provided problems, which have been solved by creating an extraction script that was made specific to overcome these problems.

However after applying the data mining algorithms on the dataset, no usable relations between the data could be found. It was proposed that the dataset was too small and that there was too much variation in the design of the buildings as both apartments and houses were used in the dataset. To test this assumption, addition data were collected from project analyses available in databases. These additional data have been used to create two new datasets. One dataset, where all the data of both the original data and the new data were combined and another dataset with only one type of buildings in the dataset, namely apartments.

These new datasets were tested by applying the same data mining process. From the results it could be concluded that the made assumptions were correct. The dataset with all the data showed that more data give a more consistent result, where the dataset with only apartments shows that the accuracy of the prediction are acceptable for conceptual cost estimation if only one type of building is used in the dataset.

To test the usefulness of the results, the results were used to estimate the costs of the buildings within the dataset. Although the results were close enough to use the generated cost estimation model to calculate a conceptual estimate for building project, the cost estimation model was only tested on the dataset that was used to create the model, so it should be noted that it is yet unclear how precise the model will be on unseen data.

Answering the research question, literature review has shown that the data mining process is an effective tool to create a knowledge management system and that BIM data have a great potential to serve as the data source. However during the practical implementation it became clear that there are both qualitative and quantitative problems with the available BIM models. Additional data were added to test viability of the data mining process for cost calculation for

residential buildings. With this additional data the data mining process could predict the costs of buildings within the dataset precisely enough to be a conceptual cost estimate. Nonetheless until the qualitative and quantitative problems of BIM data are solved the knowledge management system will still need the manual input of additional data to be able to estimate the costs of residential buildings.

This page is intentionally left blank

Samenvatting

Door een veranderde vraag binnen de bouwsector worden bouwbedrijven gedwongen om efficiënter te werken. Een mogelijke oplossing is het toepassen van kennismanagement (KM). Echter de traditionele kennismanagementtechnieken die gebruikt worden binnen de bouw voldoen niet. De reden is dat ze teveel onderhoud vragen en dat de kennis alleen kan worden opgeslagen en niet gedeeld worden binnen de organisatie. Door de opkomst van Building Information Modelling (BIM) zijn er nieuwe mogelijkheden ontstaan om een KM systeem te maken. BIM is namelijk een unieke bron van informatie gezien het data genereert en beheert gedurende de hele levensloop van het gebouw.

Dit onderzoek richt zich op het huidig ontwerpproces in de bouwsector, welke inefficiënt is. Een van de redenen hiervoor is dat de kennismanagementsystemen niet voldoen aan de eisen die nodig zijn om effectief te zijn. Gezien de omvang van het ontwerpproces zijn er veel taken te onderscheiden in dit proces. Dit onderzoek zal zich focussen op het calculeren van de bouwkosten in het vroege stadium van het ontwerpproces. Hiervoor is gekozen omdat dit onderzoek voortborduurt op een eerder onderzoek van Hoogeveen (2015) en daarbij verder ingaat op de beperkingen van zijn onderzoek.

Hoogeveen (2015) bewees gedeeltelijk de werking van een kennismanagementsysteem dat gebruik maakt van BIM, door data mining toe te passen op data over de bouw van datacenters. Hij maakte daarbij echter geen gebruik van de daadwerkelijke BIM modellen. In dit onderzoek worden BIM modellen gebruikt om op geautomatiseerde wijze data uit de BIM modellen te halen en deze te gebruiken in het data mining proces. Daarbij zal dit proces gebruikt worden als kennismanagementsysteem voor de conceptuele kostencalculatie voor woningen.

Gebaseerd op de bovenstaande probleemstelling is de volgende onderzoeksvraag opgesteld: “Kan het gebruiken van een kennismanagementsysteem, dat BIM data en het data mining proces gebruikt, zorgen voor een efficiënter kostencalculatieproces voor woningen?” Om deze vraag te beantwoorden is een kennismanagementsysteem ontwikkeld, gebaseerd op de kennis uit een literatuurstudie. Deze is daarna getest met projectdata die tijdens dit onderzoek verzameld zijn.

In de literatuurstudie wordt kennismanagement gedefinieerd als een set met handvaten voor het vinden van relaties tussen informatieobjecten, personen en bedrijfsprocessen. Om kennismanagement te begrijpen moet eerst begrepen worden wat kennis is. Kennis is gedefinieerd als verfijnde informatie die een betekenis heeft. Kennis kan gescheiden worden in twee vormen, ‘impliciete’ en ‘expliciete’ kennis. Het grootste verschil tussen de twee vormen van kennis is dat ‘expliciete’ kennis gedocumenteerd is, waar bij ‘impliciete’ kennis er vanuit gegaan wordt dat deze er is. Voor kennismanagement is alleen ‘expliciete’ kennis nuttig, zeker gezien het doel is nieuwe kennis te genereren. Dit kan namelijk alleen door het combineren van ‘expliciete’ kennis.

In dit onderzoek wordt voorgesteld om BIM te gebruiken als bron voor kennismanagement. BIM is een intelligente simulatie van de architectuur dat de volgende karaktereigenschappen heeft: digitaal, ruimtelijk (3D), meetbaar, allesomvattend, toegankelijk en duurzaam (Campbell, 2006). Het gebruik van BIM resulteert in een BIM model. Een BIM model is een virtueel prototype van het ontworpen gebouw. Er zijn twee typen BIM modellen, BIM modellen met een bestandsformaat van de leverancier en BIM modellen die een open

standaard als bestandsformaat hebben, vaak het IFC bestandsformaat. In dit onderzoek wordt gebruik gemaakt van het IFC bestandsformaat.

Voor het omzetten van de data in het IFC model naar kennis om kosten te calculeren is het data mining proces toegepast. Dit proces bestaat uit de volgende stappen:

1. Probleem begrijpen en definiëren
2. Verzamelen van de benodigde data
3. Omzetten van alle BIM modellen naar IFC modellen
4. Omzetten van IFC modellen naar een dataset
5. Kosten van gebouwen toevoegen aan dataset
6. Opschonen van de dataset
7. Een bruikbare subset selecteren
8. Ontwikkelen van een bruikbare data mining algoritme
9. Gebruik maken van het algoritme
10. Evalueren van resulteren met statistische indicatoren
11. Herdefiniëren van de data en het probleem
12. Het gegenereerde model gebruiken voor kostencalculaties

Het grootste knelpunt in dit onderzoek was het geautomatiseerd omzetten van de informatie in BIM modellen naar een dataset. In de theorie wordt omschreven dat het mogelijk is om alle fysieke informatie over de gebouwen in een BIM model geautomatiseerd om te zetten naar een dataset. Echter gedurende de praktische implementatie bleek dat de huidige kwaliteit van de BIM modellen problematisch is voor het geautomatiseerd omzetten van de data in BIM modellen naar een dataset. Om dit knelpunt te op te lossen is een script geschreven dat specifiek de problemen oplost voor de modellen die beschikbaar waren tijdens dit onderzoek.

Een ander probleem kwam aan het licht tijdens het gebruik van de gegenereerde dataset in het data mining algoritme, aangezien er geen bruikbare relaties in de data gevonden konden worden. Verondersteld werd dat dit probleem ontstond doordat de dataset te klein is en dat de dataset meerdere type woningen, appartementen en huizen, bevat. Om deze veronderstelling te testen zijn extra data verzameld uit publiekelijk toegankelijke projectanalyses van bouwprojecten. Deze extra data zijn gebruikt om twee datasets te creëren. Een dataset die bestaat uit alle data van zowel de oude dataset als de nieuw verworven data en een dataset die ook bestaat uit de oude dataset en de nieuw verworven data, echter met alleen de data van appartementen.

Deze datasets zijn gebruikt in hetzelfde data mining proces. Uit de resultaten kan worden geconcludeerd dat de gemaakte aannames correct zijn. De dataset met alle data gaf zeer consistente resultaten, waar de dataset met alleen appartementen liet zien dat het gegenereerde model precies genoeg de kostencalculatie kan maken voor conceptuele kostencalculaties.

Om te testen in hoeverre het gegenereerde model bruikbaar is, is het gegenereerde model toegepast op de dataset. Hoewel het model kostencalculaties maakte die dicht genoeg bij de daadwerkelijke kosten zitten voor een conceptuele kostenraming, is het model alleen getest op data die ook gebruikt zijn om het model te genereren. Het is daarom nog onbekend hoe precies het model zal werken op nieuwe, ongebruikte data.

Om de onderzoeksvraag te beantwoorden, de literatuurstudie heeft aangetoond dat het data mining proces een effectief hulpmiddel is voor het creëren van een kennismanagement systeem en dat de BIM data veel potentie hebben als databron. Echter, tijdens de praktische implementatie werd duidelijk dat er problemen zijn met zowel de hoeveelheid beschikbare BIM modellen als met de kwaliteit van de beschikbare BIM modellen. Door extra data te verzamelen was het toch mogelijk om het data mining proces succesvol toe te passen en bruikbare resultaten te krijgen. Echter, zolang de kwantitatieve en kwalitatieve problemen met BIM modellen niet zijn opgelost kost het handmatig invoeren van de data te veel tijd om het calculatieproces efficiënter te maken.

This page is intentionally left blank

Abstract

Due to a change in what leading clients demand from the architecture, engineering and construction (AEC) industry, the AEC industry has to work more efficient. Knowledge management (KM) has a great potential to make the design process of the AEC industry more efficient. Nonetheless the traditional knowledge management systems (KMS) do not meet the requirements necessary to be effective. Therefore new types of KMS are needed. A possible solution is to combine the potential of BIM and data mining into a KMS. The main scope of this research is to develop a KMS based on BIM and data mining and test the effectiveness of this KMS.

One of the most important aspects of the data mining process is the dataset that is analysed. To create such a dataset from BIM models a tool is created to automate the extraction of data from the BIM models into a dataset useful for data mining. This has led to insights about the current state of both the quality and quantity of BIM models. After the application of the data mining algorithms on the dataset it is suggested that the amount of available BIM models has to increase for the KMS to be successful.

To test this statement, additional data has been collected from publically available data. These data have been added to the dataset and the data mining process has been applied on the newly created dataset. It could be concluded that if the quality of the BIM models meets the necessary standard, the KMS system can be successful as long as enough BIM data is available and this BIM data is about buildings of a similar type.

This page is intentionally left blank

Preface & Acknowledgment

It is with great proud that I hereby present my graduation thesis written to finish the master program Construction Management and Engineering at Eindhoven University of Technology. After finishing my bachelor in the Built Environment at the Avans University, I started this master track and have really enjoyed it. Before getting into the content of the report, I would like to express my gratitude toward the people who have contributed to making this research possible.

I want to thank Jakob Beetz and Chris Snijders for their feedback and the time they have made to support me during the period of my research. They have given me the guidance I needed to successfully finish my graduation project. I would also like to thank Ibis BV for the opportunity to do my research at the company. The support of the team has really helped me during my research. Especially I would like to thank Jeroen van de Laar and Toine Bullens for being my external supervisors. Moreover, I would also like to thank Thomas Krijnen with his help with programming the python script. Without his help I would not have been able to create the script necessary for my research.

I hope that every reader will enjoy reading and learning from this thesis report.

Yonne Rekveld
Eindhoven, May 2017

This page is intentionally left blank

Abbreviations

AEC:	Architecture, Engineering and Construction
(A)NN:	(Artificial) Neural Network
AI:	Artificial Intelligence
BIM:	Building Information Modelling
ILS:	Informatieleveringsspecificatie (Information Delivery Manual)
CAD:	Computer Aided Design
CB-NL:	Nederlands conceptenbibliotheek voor de bouw
CBR:	Case-Based Reasoning
DBFMO:	Design, Build, Finance, Maintain, Operate
GD:	Generative Design
ICT:	Information and Communication Technology
IFC:	Industry Foundation Classes
KBE:	Knowledge-Based Engineering
KBS:	Knowledge-Based System
KDD:	Knowledge Discovery in Databases
KE:	Knowledge Engineering
KM:	Knowledge Management
KMS:	Knowledge Management System
K-NN:	K-Nearest Neighbour
SVM:	Support Vector Machine
TCO:	Total Cost of Ownership

This page is intentionally left blank

Contents

Colofon 3
Summary 5
Samenvatting 9
Abstract 13
Preface & Acknowledgment 15
Abbreviations 17
Contents 19
List of figures, tables and equations 21
List of figures 21
List of tables 21
List of equations 21
1. Introduction 25
1.1. Problem definition 27
1.1.1. Problem analysis and limitations 27
1.1.2. Problem definition 27
1.1.3. Research questions 27
1.2. Research approach 28
1.3. Practical and theoretical relevance 29
1.3.1. Theoretical relevance 29
1.3.2. Practical relevance 29
1.4. Reading Guide 29
2. Introduction literature review 33
3. Knowledge management 35
3.1. Definition of knowledge management 35
3.2. From data to knowledge 35
3.3. Tacit and explicit knowledge 36
3.4. Knowledge conversion and creation 36
3.5. Goal of knowledge management 37
3.6. Current state of AEC knowledge management 38
3.7. Conclusion knowledge management 40
4. BIM data as knowledge source 41
4.1. Big data 41
4.1.1. Big Data Engineering and Analytics 42
4.1.2. Data mining 43
4.1.3. Machine learning 44
4.1.4. Evaluation of the data mining performance 46
4.2. Building Information Modelling 47
4.2.1. What is BIM? 47
4.2.2. Industry Foundation Classes 48
4.2.3. Improvements of the shared BIM Models 49
4.2.4. BIM for cost estimation 50
4.3. Conclusions BIM as a knowledge source 51

4.4. Conclusion of literature review 52
5. System Design 57
5.1. Data collection and extraction 57
5.2. Data cleaning and Exploration 65
5.3. Data engineering 67
5.3.1. Sub dataset selection 67
5.3.2. Normalization 68
5.4. Conclusions data engineering 68
6. Data analytics 71
6.1. Algorithm engineering 71
6.2. Running the algorithm 71
6.3. Evaluation of result 73
6.4. Redefining data 73
6.5. Using the new datasets 75
6.6. Conclusion data analytics 78
7. Discussion 81
8. Conclusions 83
9. Recommendations 87
9.1. Recommendations for implementation 87
9.1.1. Implementation of the IFD format 87
9.1.2. Implementation of the Information Delivery Manual(IDM) 87
9.1.3. BIM as-built models 87
9.1.4. Use of additional data 88
9.2. Recommendation for further research 88
9.2.1. Quality of BIM models 88
9.2.2. Other goal of KMS 88
References 89
Appendix 95

List of figures, tables and equations

List of figures

Figure 1: Research model 28
Figure 2: SECI model of knowledge conversion 37
Figure 3: Supervised Learning compared to Unsupervised Learning 45
Figure 4: Proposed model 53
Figure 5: From input data to CSV 59
Figure 6: Structure of extractor script 62
Figure 7: Distinction between the preferred and actual working of the script 63
Figure 8: Model without structural floor 66
Figure 9: Correlation plot 68
Figure 10: Building elements composing of different construction products 85
Figure 11: Structure of the IFC data format 96
Figure 12: Price index residential buildings 137
Figure 13: Correlation coefficient of database A 138
Figure 14: Root mean squared error of database A 138
Figure 15: Mean absolute error of database A 139
Figure 16: Root relative squared error of database A 139
Figure 17: Correlation coefficient of database B 140
Figure 18: Root mean squared error of database B 140
Figure 19: Mean absolute error of database B 141
Figure 20: Root relative squared error of database B 141

List of tables

Table 1: Data mining process according to Fayyad et al. and John 43
Table 2: Building elements to be extracted 60
Table 3: Overview of building elements with the corresponding IfcElement 61
Table 4: Overview of the extracted dataset 64
Table 6: Performance of the linear regression algorithm 72
Table 7: Performance of the k-nearest neighbour algorithm 72
Table 8: Additional data from databases 74
Table 9: Performance of the linear algorithm on the dataset A 75
Table 10: Performance of the linear algorithm on dataset B 76
Table 11: Predicted prices of dataset A 77
Table 12: Predicted prices of dataset B 77

List of equations

Equation 1: Correlation coefficient 46
Equation 2: Root mean squared error 47
Equation 3: Mean absolute error 47
Equation 4: Root relative squared error 47
Equation 5: Euclidean distance 68
Equation 6: Normalization 68

Equation 7: Price normalization ……72

Part I: Problem statement

Part II: Theoretical review

Part III: Practical implementation

Part IV: Conclusions and recommendations

This page is intentionally left blank

1. Introduction

Leading clients in the architecture engineering and construction (AEC) industry are increasingly demanding higher quality products at a low cost. They demand a high reliability in quality and delivery on the required date (Brown & Adams, 2000; Stewart & Spencer, 2006). To achieve this change in demand, the construction time of the building projects has to be reduced while the quality has to be improved. As a result of this, companies within the construction process are forced to improve their efficiency as it is the ratio between input and output.

Because of the unique nature of every construction project, the construction process is difficult to standardize. Every project needs different solutions (Van der Aalst et al., 2002). Although this might be true, only 20% of all engineering activities done by experts within the building project are tasks with no pre-existing solution (Koskela, 2000). One of the reasons why this percentage is so low is because the knowledge gained within the organization is not shared. This leads to the re-engineering of a design solution while the solution already exists within the organization. Calkins, Egging and Scholz (2000) stated that when the architecture, engineering and construction (AEC) industry would achieve to systematically managing the acquired knowledge, they could use this knowledge for 80% of all the design tasks, which would lead to a reduction of process time.

The reuse of knowledge would improve the efficiency of the construction process. But in order to be able to reuse this knowledge, this knowledge has to be captured and stored. One of the solutions to achieve this is by using an effective Knowledge Management System (KMS) (Konukcu & Koseoglu, 2012).

Many companies around the world recognize the advantage of Knowledge Management (KM). KM will not only lead to a more efficient process, it is also crucial for companies to sustain their competitive advantage (Alavi & Leidner, 1999). This is the case in the upcoming years, when a large percentage of the skilled workforce will retire from work, which will lead to a lot of lost knowledge. Researchers at the Sloan Center on Aging & Work at Boston College (Sweet & Pitt-Catsouphes, 2010) examined the effect of the aging workforce within the construction industry. In this research they concluded that one of the major concerns is how to transfer the existing knowledge to lesser experienced workforce. Given that KM provides a way to transfer this knowledge to sustain competitive advantage and leads to more efficient processes, it is hard to see why KM is rarely implemented adequately within construction firms.

The reason behind this is that traditional KM techniques used in the AEC industry have some shortcomings (Deshpande, Azhar & Amireddy, 2014). According to a research of Wong (2006) one of the most critical success factors of KM is that the chosen KMS is a very dynamic system that can easily capture and transfer knowledge and requires little maintenance. Most techniques used in the AEC industry, like standards and best practise guides and post practice reviews, require paper based documentation. This requires a lot of KM maintenance and it only achieves that knowledge is captured, not shared within the organisation (Dave & Koskela, 2009).

As stated above, KM is only effective when the system demands little maintenance effort and is easy to access, both to capture and to store knowledge. This is why data from Building Information Modelling (BIM) have a great potential for KM because “BIM is the process of generating and managing data during the building life cycle” (Jan, Ho, & Tserng, 2013) and it

is already implemented in lots of building projects (Ho, Tserng & Jan, 2013). As the result is that BIM data are a unique source for KM as they provide a way to capture and store the information created during a project.

As stated by Jan, Ho and Tserng (2013), the development in BIM provides a way to share information about projects. When BIM is applied adequately within a project, the project uses a central database where all information about the project generated by the different stakeholders within the process is stored. As this database is accessible to every project member with an internet connection, the data within the database can also be easily updated and transferred to other stakeholders.

Although BIM allows stakeholders to share and store information about projects, it stores no knowledge. However data can be transformed into information and information into knowledge according to the widely accepted theory of knowledge hierarchy (Kebede, 2010). This knowledge can be used to create new knowledge when it is made explicit and combined with a other form of explicit knowledge (Nonaka, 1994). This means that if one is able to convert the information stored in the BIM database, in the form of BIM models, to explicit knowledge it can be used to gain new knowledge, by finding relationships between the different forms of knowledge that are made explicit.

Hoogeveen (2015) recognised this potential of converting BIM data into knowledge. In his research Hoogeveen used the data mining process, to create new knowledge in the form of a model to estimate the costs of building projects. The data mining process is a process in which information, in the form of data which are located in databases, is converted into knowledge. This is also called knowledge discovery in databases (KDD) (John, 1997).

Due to limitations in his research, Hoogeveen (2015) did not actually extract the information used in the process from BIM data. He used the data in the form of PDF files and extracted these data manually. This is problematic because for a KMS, in order to be effective, the system has to require little maintenance. So for a KMS to be effective, the information located within the BIM data needs to be extracted into knowledge automatically. Another limitation is the kind of building projects used in his research. In his research Hoogeveen used information about datacentres and converted this into knowledge in the form of a model that can predict the total costs of ownership for datacentres. As datacentres have relatively simple designs and the designs do not significantly differ from each other, they are pretty uniform. This makes datacentres uniquely qualified for cost prediction based on data mining (Hoogeveen, 2015), but unlikely to deliver general knowledge that allows usage outside the domain of datacentres. Given both these limitations, the presented study will focus on the creation of a KMS that automatically extracts information from BIM models and will use this KMS for cost prediction of residential buildings.

In Summary, due to a change in what leading clients demand from the AEC industry, the AEC industry has to work more efficient. A possible solution is to apply KM within the organisation. However, the problem with traditional KM techniques used in the AEC industry is that they demand too much maintenance and they only store but do not share information within the organisation. With the development of BIM, a new possibility to create a KMS has emerged. BIM provides a unique source of information as it generates and manages the data created during the building life cycle. This unique source of information can be converted into knowledge by means of data mining. Hoogeveen (2015) proved part of this concept by using a data mining approach to predict the costs of datacentres. This study will focus on the

unproven part, to extract the information in BIM models into knowledge automatically. It will also explore whether or not this concept is also applicable to predict the costs of buildings with a less general design as datacentres.

1.1. Problem definition

1.1.1. Problem analysis and limitations

The problem this research will focus on, is that the current design process of the AEC industry is inefficient. One of the reasons is that the knowledge management systems (KMS) within the AEC industry do not meet the requirement necessary to be effective. Theoretically a combination of BIM and the data mining process can provide a KMS that is effective and make the design process more efficient. The design process is very broad and has multiple tasks that can be made more efficient with KM. As this research is based on the limitations of the research done by Hoogeveen (2015), the KMS will focus on the same tasks, namely cost prediction in the early stage of the design process.

Although a combination of BIM and the data mining process can theoretically provide a KMS, the data mining process has to be able to recognise patterns within the data. This concept has only be proven with a very standardised type of building, specifically datacentres. To test if this concept can also be used for less standardised building types, another building type should be used. For this research BIM data of residential buildings were available. Therefore this KMS will be focussed on residential buildings.

The objective of this research is:

"To realize a more efficient process of cost estimation for residential buildings".

1.1.2. Problem definition

Based on the research object the problem definition is defined:

"Can the use of a knowledge management system using BIM data and the data mining process create a more efficient process of cost estimation for residential buildings?"

1.1.3. Research questions

In order to achieve the research objective of this research and to answer its problem definition the following research questions are specified.

Research question 1:

- Which are the steps the data mining process consists of?

As the goal of this research is to use the data mining process as a KMS it is useful to know what the steps are within the data mining process.

Research question 2:

- Does the IFC data format contain all the necessary data to extract a dataset for the data mining process?

The IFC data format will be used as the BIM standard for this research. This was chosen because this makes it possible to use BIM models created with all BIM authoring tools that are generally used in the AEC industry as they all support the model to be exported to the IFC data format. Also the IFC data format is becoming the Dutch BIM standard as both

companies are starting to mutually agree on using the IFC data format is the only format to share BIM models (BIM basic information delivery manual, 2016) and the government is starting to demand the IFC data format for integrated contracts such as Design, Build, Finance, Maintain and Operate (DBFMO) contracts (Central Government Real Estate Agency, 2013). Although this makes the IFC data format the logical format to use it still needs to be determined if all required information can be extracted from the IFC data format.

Research question 3:

- Is the current BIM models modelled with a high enough quality to automatically extract a dataset for the data mining process?

The effectiveness of the data mining process depends on the quality of the dataset. This dataset is extracted from the BIM models and therefore if the quality of the BIM models is not high enough, the data mining process will not work.

Research question 4:

- Can the data mining process be used to predict the cost of residential buildings in the early stages of the design process?

1.2. Research approach

This research consist of both a theoretical and practical part. In the theoretical part a knowledge management system will be suggested to create a cost estimation model for residential buildings, based on the literature. The use of this system will be tested in the practical implementation. After both parts are carried out it will be concluded if the proposed system is effective.

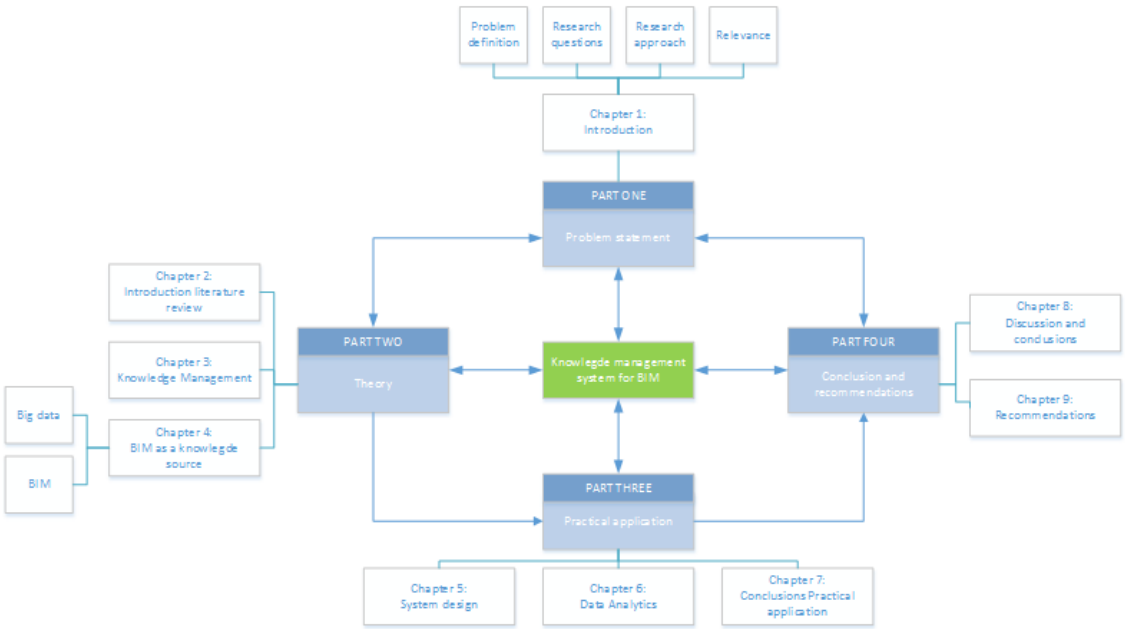

Figure 1: Research model

1.3. Practical and theoretical relevance

1.3.1. Theoretical relevance

The theoretical relevance of this research is to further explore the current possibilities and limitations of knowledge management using data mining and BIM data. This research continues the research of Hoogeveen (2015) in which BIM data and data mining were used to predict the costs of a building project. It aims to give knowledge about the possibilities outside the domain of datacentres. It also explores the potential of the current BIM models to extract data automatically by using an extraction tool for cost prediction and the potential improvement that should be made in the quality of the BIM models before it can be successful.

1.3.2. Practical relevance

The practical relevance of this research is to explore the possibility of an alternative way to create a cost estimation in the early phase of the design process. If the KMS can be automated the labour time for cost estimation will be greatly reduce. This leaves more time to do tasks that add value for the client.

For a data mining process to work effectively, the input data have to be of a high enough quality. In this research the quality of the BIM models is addressed. It will provide insight in the current quality of the BIM models and in how this quality can be improved to be useful in a data mining process.

1.4. Reading Guide

This research consists of both a theoretical approach and a practical implementation. The report starts with an introduction of the problem and the derived research questions. Thereafter a literature review is given, which is followed by a practical implementation with system development and application. The report ends with the conclusions and recommendations.

Chapter 1 elaborates the motivation of the research. The research objective, research problem and research questions are given. Also the research approach, theoretical and practical relevance and the reading guide are given.

Chapter 2 gives an introduction to the literature review. It explains in more detail how the literature research is structured.

Chapter 3 provides an overview in knowledge management. It elaborates what knowledge management is, and what the current state of knowledge management is within the AEC industry.

Chapter 4 gives more background information about the combination of big data and BIM. It explains how big data and data mining are related to each other and how the data mining process is structured. It also provides a review of BIM and explains the potential of BIM as a source for data mining. In addition it also explores the Industry Foundation Classes.

Chapter 5 describes the implementation of the data engineering part of the data mining process. It explains in more detail how data engineering should be executed. During this chapter also an overview of the practical implementation problems identified during data engineering is given.

Chapter 6 describes the implementation of the data analytic part of the data mining process. It explains in more detail how data analytics should be executed. For this chapter also an overview of the problems with the practical implementation is given.

Chapter 7 discusses a critical evaluation of the presented research.

Chapter 8 provides the conclusions of this research. In this chapter the research questions and problem statement are answered.

Chapter 9 elaborates the recommendations for implementations and the possibilities for further research in combination of data mining, BIM data and knowledge management.

Part I: Problem statement

Part II: Theoretical review

Part III: Practical implementation

Part IV: Conclusions and recommendations

This page is intentionally left blank

2. Introduction literature review

In the problem statement part of this research the research questions were defined. Some of these research questions can be answered by the literature review. In this part of the report the research question *Which are the steps the data mining process consists of?* will be answered.

As this research is a continuation on the research done by Hoogeveen (2015) most of the literature used in his research can be used for the literature review of this research. Especially on the topic of knowledge management much of the literature review is based on Hoogeveen. This literature has been verified by finding additional literature and by critically reviewing the used literature. For other topics the literature of Hoogeveen (2015) is also useful, however as this research has a different focus more research on these topics was necessary.

The literature review will start with a chapter about knowledge management. This chapter will explain what knowledge is and how it can be managed. Thereafter it will be explained how BIM data can be used as a knowledge source when using the data mining process.

This page is intentionally left blank

3. Knowledge management

The change from an industrial society to a knowledge based society has let knowledge to be the most important resource of a company. With this movement it has become more important to share knowledge within a company (Johannessen, Olaisen & Olsen, 2001). The reuse of existing organizational knowledge attained by previous experiences can reduce time spent on problem solving and increase the quality of work and therefore results in competitive advantage. This makes managing the knowledge within the organization very important. It is especially important for AEC-based organizations, because of the high amount of engineering tasks, as these are highly knowledge and experience driven (Deveraja, 2015). As knowledge is the most important asset of companies, companies must leverage their knowledge in order to maintain their sustainable competitive advantage.

3.1. Definition of knowledge management

Knowledge management (KM) is difficult to articulate and quantify because it contains elements of disciplines of both “hard” and “soft” sciences (Abaljaber et al., 1998). There is no consensus on what KM is. As the research of MIT (Abaljaber et al., 1998) shows, different articles give different solutions of KM. In the same research a small alteration of the definitions of Frappaolo and Toms (1997) is proposed: "KM is a tool set for the automation of deductive or inherent relationships between information objects, corporate users and business processes".

As knowledge management is about knowledge it is important that there is an understanding of what knowledge is and how it can be obtained. In the following section it will be made clear where knowledge comes from, how it can be defined and which types of knowledge exist.

3.2. From data to knowledge

An understanding of the concept of knowledge and knowledge taxonomies is important because theoretical developments in the knowledge management area are influenced by the distinction among the different types of knowledge (Alavi & Leidnet, 2001). The concepts of data, information and knowledge are closely related (Kock et al. 1997), and it is commonly known that knowledge has a higher level than information, and information has a higher level than data (Tuomi 1999). Data are symbols that represent the properties of objects and events without any added interpretation or analysis. They simply exist and have no significance beyond its existence (in and of itself). They can exist in any form, usable or not. They do not have meaning of their selves (Ackoff, 1989; Ackoff, 1999). Davenport and Prusak (2000) indicated that “data are a set of discrete, objective facts about events”, and “provide no judgment or interpretation and no sustainable basis of action”. Data are syntactic entities and patterns without meaning, and exist in usable or non-usable forms without any significance beyond the existence (Aamodt and Nygård 1995; Bellinger et al. 2004). Data have no meaningful relation to anything else, since they are missing a context (Uriarte, 2008).

Information is data that have been given meaning by way of relational connection. This meaning can be useful, but it does not have to be so. Aamond and Nygård (1995) explain information as structured data with meanings, which is generated from the interpretation process of data. Ackoff (1990) defined information as “data that are processed to be useful, providing answers to ‘who’, ‘what’, ‘where’, and ‘when’ questions”.

Knowledge is refined information. Knowledge is the appropriate collection of information, in such way that its intention is to be useful. Knowledge is a deterministic process. When someone "memorizes" information, then they have amassed knowledge. This knowledge has a useful meaning to them, but it does not provide for, in and of itself, an integration such as when it would infer further knowledge (Ackoff, 1990; Aamond and Nygård, 1995).

In summary, data are a carrier and storage of information and knowledge, and a media for information exchange and knowledge transfer (Kock et al. 1997). Kock et al. (1997) indicated that information is descriptive and related to the past and the present, while knowledge can be used to predict the future within a certain limit. The role of knowledge is to facilitate the processes of transforming data into information through data interpretation, deriving new information from existing through elaboration, and acquiring new knowledge through learning (Aamodt and Nygård 1995).

3.3. Tacit and explicit knowledge

Nonaka (1994) has identified two dimensions of knowledge in organizations: tacit and explicit. According to Nonaka, the tacit knowledge is rooted in action, experience, and involvement in a specific context. Tacit knowledge is comprised of both cognitive and technical elements (Nonaka, 1994). The cognitive element refers to an individual's mental model consisting of mental maps, beliefs, paradigms and viewpoints. The technical component consists of concrete know-how, crafts and skills that apply to a specific context. Tacit knowledge consists of professional expertise, individual insights and experience, and creative solutions (Pozzali & Viale, 2015).

Junnarkar and Brown (1998) suggest that "tacit knowledge is that which is implied but not actually documented". In other words knowledge can be tacit not because one is unable to articulate it, but because it has not been documented yet. This perspective is very useful because it suggests that some tacit knowledge may be more valuable when made explicit than other. The goal of knowledge management would not be to explicate all tacit knowledge but to assess first the existing tacit knowledge and determine that which has the most value before trying to make it explicit.

The explicit knowledge is articulated, codified and communicated in symbolic form and/or natural language. Most explicit knowledge exists in forms of technical or academic documents, like manuals, mathematical expressions, copyright and patents. This ''know-what'' or systematic knowledge is readily communicated and shared through printed documents, electronic methods and other formal means. Explicit knowledge is technical and requires a level of academic knowledge or understanding that is gained through formal education, or structured study. Explicit knowledge is carefully codified, stored in a hierarchy of databases and is accessed with high quality, reliable, fast information retrieval systems. Once codified, explicit knowledge assets can be reused to solve many similar types of problems or connect people with valuable, reusable knowledge (Smith, 2001). Sharing processes often require major monetary investments in the infrastructure needed to support and fund information technology (Hansen et al., 1999).

3.4. Knowledge conversion and creation

Next to the tacit-explicit distinction of knowledge another distinction between dimensions of knowledge was identified by Nonaka (1994). These dimensions, individual and collective (or

social) knowledge, in combination with the tacit-explicit dimension can be used to distinguish different kinds of knowledge conversion and creation.

Nonaka (1991) dimensioned four types of knowledge conversion on the SECI model. These four types are socialization, externalization, combination and internalization. 'Socialization' is converting individual tacit knowledge to group tacit knowledge. In the type 'externalization' tacit knowledge is made explicit. The type 'combination' is conversing separate explicit knowledge to systematic explicit knowledge and 'internalization' is the conversion from explicit knowledge to tacit knowledge.

	To tactic knowledge	to explicit knowledge
From tactic knowledge	Socialization	Externalization
From explicit knowledge	Internalization	Combination

Figure 2: SECI model of knowledge conversion (Nonaka, 1991)

As this research has as goal to convert information saved in BIM data, which is explicit, to new, well documented knowledge in the form of a tool to predict the costs of a building, this research will focus on the type 'combination'.

3.5. Goal of knowledge management

In the previous four sections, knowledge and knowledge management have been discussed. It has been made clear what knowledge and knowledge management are and why knowledge management is important. However what the actual goal of knowledge management is, has not been discussed yet. The goal of knowledge management (KM) is to connect knowledge providers and knowledge seekers to provide value creation and create sustainable competitive advantage (Abaljaber et al., 1998; Alavi & Leidner, 2001). Sustainable competitive advantage can be achieved through resources that are valuable, rare and imperfectly imitable. This resources can be property-based or knowledge based. Property based resources are legally controlled by a specific firm where knowledge based assets are protected because they are often subtle or difficult to understand or copied by outside observers.

In a study by Davenport, De Long and Beers (1997) four business objectives that fulfil the goal of KM are identified, namely:

1. To capture knowledge;
2. To improve knowledge access;
3. To enhance the knowledge environment;
4. To manage knowledge as an asset.

Capturing knowledge can be achieved by creating KM repositories. These consist of structured documents with knowledge embedded in them, stored in a way that they may be easily retrieved. Better access to knowledge can be facilitated by improving the processes of knowledge transfer between individuals and between organizations. By proactively facilitating and rewarding knowledge creation, transfer and use of an enhanced knowledge environment can be achieved. Knowledge should also be managed as an asset, and can be achieved in

multiple ways. For example some companies include their intellectual capital in the balance sheet, where other leverage their knowledge assets to generate new income or reduce costs.

3.6. Current state of AEC knowledge management

In the section below a review of the current state of knowledge management within the AEC industry is given. Much is based on the review given by Hoogeveen (2015). Also additional literature was found, both to verify the work done by Hoogeveen and to expand the review.

Knowledge is one of the most important assets of an AEC firm (Deshpande, Azhar, & Amireddy, 2014; Dave & Koskela, 2009; Woo et. Al, 2004) and since most of the firms knowledge is created during the lifecycle of a project, managing this generated knowledge is important. AEC firms mostly used to rely on individual tacit knowledge and experiences gained through previous projects. Since the last decade, more attention has been given toward KM (Al-Gahtani & Ghani, 2010; Weippert & Kajewski, 2007; Kanapeckien et al., 2010).

AEC firms have both explicit and tacit knowledge to manage (Woo, Clayton, Johnson, Flores, & Ellis, 2004). In the past, most research in the AEC industry has been focused on explicit knowledge. Explicit knowledge examples in the AEC are specifications, reports, procedure manuals, drawings and documents management systems (Al-Gahtani & Ghani, 2010).

Around 2004, a shift in attention from managing explicit knowledge to managing tacit knowledge can be found (Kamsu-Foguem & Abanda, 2015). The management of tacit knowledge requires other techniques, such as lessons learned to record personal tacit experiences, post project reviews, databases of best practices, web-based communities, decision support systems, collaborative internets, and Communities of Practice (CoP) (Deshpande, Azhar, & Amireddy, 2014; Tibaut & Jakosa, 2015; Woo, Clayton, Johnson, Flores, & Ellis, 2004; Al- Gahtani & Ghani, 2010; Kanapeckiene, Kaklauskas, Zavadskas, & Seniut, 2010).

Techniques like post project reviews were introduced to convert tacit knowledge to explicit knowledge. However these techniques are very time consuming and therefore unlikely to be effective. Even when the reviews are done properly, these documents are often stored in archives, serving no purpose and are not easily and universally accessible (Dave & Koskela, 2009). Recently, most of the scientific research has focussed on managing tactic knowledge and mostly human interaction (Al-Gahtani and Ghani, 2010). This movement in research seems to be directed towards managing tacit knowledge for processes and experiences (Woo, Clayton, Johnson, Flores, & Ellis, 2004; Dave & Koskela, 2009).

Woo et al (2004) focused on developing a web based knowledge navigator. It allows experts to search for knowledge and facilitates communication with other experts. Hence it enables socialisation. Dave and Koskela (2009) stated that the general consensus in the literature is that the AEC industry fails to retain project knowledge for future reuse. They also develop a system to enable social interaction between experts to capture tacit knowledge in the form of experiences, and make it explicit in the form of best practices with a database.

Lately, different tools of KM are integrated with BIM. Many of these tools are used to graphically represent the captured knowledge. Ho, Tserng and Jan (2013) developed a BIM-based knowledge model where experiences are graphic representations via objects and aspects of BIM models. Liu et al (2013) developed a similar model, where tacit experts'

knowledge was made explicit by using lessons learned techniques. These were linked to a BIM model, to improve the sharing of knowledge between engineers.

Lin (2014) created a 3D Construction BIM-based Knowledge Management system, called the CBIMKM system. This BIM-based KM system focused the representation of knowledge for all participants in a project. The CBIMKM system is similar to traditional expert 'yellow pages' and document management module, where stakeholders can search on a topic, via a search engine, to obtain the knowledge in the form of a BIM object with an explanation provided by the design expert.

Grover and Froese (2016) assessed the usefulness of a BIM-based social platform to manage tacit knowledge. Such application, which in generally is called SocioBIM, facilitates the sharing of collective skills, expertise, understanding and knowledge throughout the life cycle of the project. Again, the BIM application is to graphically represent the captured knowledge.

Ding, Zhong, Wu and Luo (2016) suggested a combination of BIM, ontology and semantic web technology to manage the knowledge of construction risks. Their suggested method facilitates the knowledge reuse during the construction risk analysis process. The risk database is created via documents and the BIM model serves as a graphical representation. So again the BIM part of the KM tool is to visualise the knowledge. However it uses the information within the BIM model to link automatically the existing database and the object within the building design to analyse the risks involved in the building process.

In the last couple of years, the implementation of big data in tools for KM has emerged. Bilal et al. (2016) presented a review of the current state of big data within the AEC industry. In this review a couple of (possible) KM tools were presented. Generative design (GD) is one if the possible application given. GD aims to automatically generate a design based on specified design objects. However this idea is still very conceptual. Another possible application is the generating knowledge using clash detection within BIM to provide guidelines for early design review. Again this is still very conceptual and much development within the field is necessary.

A more feasible way for implementation in short term is to predict the performance of building. As this predictions have a high dimensional dataset, big data application can provide solutions to analyse the performances (Bilal et al., 2016; Mathew, Dunn, Sohn et al., 2015). In the next chapter it will become clear why the application of big data is suitable for this application.

Another feasible implementation is to analyse the social networking with big data. As mentioned above, tools have been created with BIM to provide an application platform for experts within an AEC project. This communication can be analysed with big data for automatically classification of best practises mentioned in the social networking tool (Bidal et al., 2016).

Bidal et al. (2016) also mentioned possibilities for the future. Most of them were combining BIM model with big data to create new knowledge. Cost prediction with big data however was not mentioned in the review of Bidal et al. (2016). But in a graduation thesis Botter (2013) created a model to calculate project costs based on historical project data. Botter reviewed multiple different techniques to calculate the project costs. Although the term big data is never mentioned within the research report the used techniques to analyse the data are similar to the big data techniques. As a conclusion the usefulness of historical project data to

calculate the project costs was stated. Another conclusion was that BIM should be integrated in the process to make it less time consuming.

These findings on the trend of KM focus in the AEC can be confirmed by Rezgui et al. (2010). The authors did research regarding the evolution of KM in the AEC sector. They stated that the evolution of KM in the AEC could be divided in three generations. The first generation focused on capturing explicit knowledge via document-based knowledge (e.g. CAD drawings). The second generation was directed on the codification of knowledge in the form of BIM and associated data. In the third generation, the use of BIM was vacated and the focus was converted to capturing tacit knowledge, this resulted in systems to encourage the sharing of tacit knowledge. Walker (2016) also did a research on the evolution of KM in the AEC sector and concluded that with the emerge of big data, the new competitive advantage can be created through big data. This new possibility has led to the development of tools that use big data in KM.

Summarizing, the development of KM within the AEC industry can be defined in four generations, the three generations as defined by Rezqui et al. (2010) and the new generations defined by Walker. The first generation aimed on capturing explicit knowledge, the second generation used the emerging of BIM to capture both tacit and explicit knowledge. In the third generation the focus was on capturing tacit knowledge. With the new possibilities of big data new possibilities for KM were discovered. A new generation is emerging, which offers many possibilities to generate new knowledge. Especially the combination of BIM models and big data has a lot of potential for KM. However to date, little practical application has been created.

3.7. Conclusion knowledge management

In this chapter it has been shown that KM is very important. KM can be defined as a tool set for the automation of deductive or inherent relationships between information objects, corporate users and business processes. To understand knowledge management, at first it has to be understood what knowledge is. Knowledge is defined as refined information, which means that it has meaning. Knowledge can be divided in two forms, tacit or explicit knowledge. The main difference between tacit and explicit knowledge is that explicit knowledge is documented, where tacit knowledge is implied. However, for knowledge to be useful for KM it needs to be made explicit. Especially because the goal of KM is to create new knowledge. This can only be done by combining multiple forms of explicit knowledge. In the AEC industry great attention has been paid on KM. Rezqui et al. (2010) defined three generations of knowledge management. In this research a fourth generation is suggested, KM with big data. Already multiple researches have been conducted to explore this subject. However few of those researches resulted in practical implementation.

4. BIM data as knowledge source

In the last decade, Building Information Models (BIM) have been successfully used to enhance the performance of AEC projects. BIM is a "shared knowledge resource for information about a facility forming a reliable basis for decisions during its life-cycle; defined as existing from earliest conception to demolition" (NBIMS, nd). It has improved communication of the design between various stakeholders by enabling the identification of clashes ahead of time, by enabling the simulation of the construction sequence and improving the communication between various craft subcontractors and the general contractor (Deshpande, Azhar & Amireddy, 2014). Building information models are inherently parametric, data-rich, object based representations of the facility being designed and constructed. Alternately, building information models can be conceptualized as centralized, interconnected data stores which can contain design and some construction information about the various disciplines within a construction project. This centralized and integrated nature of the design information can potentially provide a very context rich platform for the capture, storage and dissemination of the knowledge generated during the design and construction processes. One of the requirements of an effective knowledge management system is its ability in communicating and preserving knowledge effectively across various stages of a construction project (Dave and Koskela, 2009). Because BIM models can be used over the whole span of the construction project and even evolve and are able to capture the knowledge as soon as the knowledge is created, BIM models are uniquely qualified as a knowledge source (Deshpande, Azhar & Amireddy, 2014).

Although BIM models are great knowledge sources, the knowledge within these models is not explicit. Therefore BIM models can be seen as sources of embedded knowledge. Embedded knowledge is knowledge locked in processes, products or artefacts (Argote & Ingram, 2000). Even though embedded knowledge can have an explicit form, such as BIM models, the knowledge itself is not explicit, the implications of the embedded knowledge are not immediately clear. (Gable & Blackwell, 2001).

To be able to use the embedded knowledge as a source for knowledge management, the knowledge itself has to be made explicit and usable. To do so, big data techniques will be used. What big data and the associated technique are will be explained in the sections below.

4.1. Big data

Big data is an abstract concept. At present, although the importance of big data has been generally recognized, people still have different opinions on its definition. In general, big data means the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time. There have been considerable discussions from both industry and academia on the definition of big data (Chen, Mao & Liu, 2014). One of the most cited definitions of Big Data is that of Doug Laney from the then-META (now Gartner) group. This definition consists of three V's, Volume, Velocity and Variety (Laney, 2001). Later this definition has been expanded to include a fourth V: Veracity (Ward & Barker, 2013).

- Volume: How much data?
- Velocity: How fast data are processed?
- Variety: The various types of data
- Veracity: How accurate are the data in predicting business value?

Many companies like IBM use these V's to explain big data (IBM, 2016). Although these definition is widely adopted there is also criticism. Hurwitz et al (2013) stated: "While it is convenient to simplify big data into the four V's, it can be misleading and overly simplistic. For example, you may be managing a relatively small amount of very disparate, complex data or you may be processing a huge volume of very simple data. That simple data may be all structured or all unstructured".

Ward and Barker (2013) conducted a research on the definitions used for Big Data. They concluded that there are many different definitions, which all include different factors and they created a new definition which takes all those definitions into account. "Big data is a term describing the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning" (Ward & Barker, 2013).

4.1.1. Big Data Engineering and Analytics

As the terms storage and analysis already suggest big data has two complementary aspects, Big Data Engineering (BDE) and Big Data Analytics (BDA) (Bidal et al., 2016). BDE is the domain that primarily is concerned with supporting the relevant data storage and processing activities needed for the analytics. BDE can be divided in two subdomains:

- Big Data processing
- Big Data storage

An important part of the research problem of the research is to use BIM models as input for data mining. This part of the process belongs to data processing. The literature discusses big data processing for databases in the order of terabytes and even petabytes (Dittrich & Quiané –Ruiz, 2012). These databases can be handled with data processing engines that allow parallel and distributed computation. In parallel computing many calculation or the execution of processes are carried out simultaneously. It divides the overall (large) problem in smaller ones that can be solved at the same time (Barney, 2016). In distributed computing the overall problem is also divided. However in distributed computing the calculations are carried out on multiple systems (IBM, 2017).

However the database in this research is much smaller. The problem in this research is to convert the embedded information in the BIM model to a dataset with a suitable file format that can be used in the analytics of the data. This file format has to be supported by the software package used during the data analytics. More about this in chapter 5.

Big Data Analytics relates to the tasks responsible for extracting the knowledge to drive decision-making. It mostly deals with the principles, processes and techniques to understand the Big Data (Bidal et al., 2016). In their research Bidal et al. (2016) performed a review of Big Data in de AEC industry. They divided BDA within the AEC industry in:

- *Statistics*
- *Data mining*
- *Machine learning*

Statistics is the study of collecting, analysing, and drawing conclusions from the data, with the primary focus on selecting the right tools and techniques at every data analysis stage. Right from the data collections, to efficiently analysing it, and then inferring or formulating conclusions out of it, all of these steps come under the scope of statistics (Bidal et al., 2016).

Data mining and machine learning are terms often used for the same thing, but there is an essential difference. Data mining is concerned with the process of automatic or semiautomatic exploration and analysis, of large volumes of data, to discover meaningful patterns or rules (Bidal et al., 2016). Machine learning (ML) is defined by Arthur Samuel (1957) as a "Field of study that gives computers the ability to learn without being explicitly programmed". ML uses algorithms that can learn from and make predictions on data (Kohavi & Provost, 1998). In the data mining process machine learning is one of the techniques that can be used to find the pattern or rules in data.

In this research, the goal is to extract the explicit knowledge embedded in the data to useful explicit knowledge. This is often called knowledge discovery in databases (KDD) or data mining. What data mining is and how machine learning can be implemented in this process will be explained in the following section.

4.1.2. Data mining

Data mining or knowledge discovery in databases (KDD) is the process of discovering advantageous patterns in data (John, 1997) (Fayyad, Piatetsky-Shapiro & Smyth, 1996). In the commercial field this whole process is called data mining. This in contrary with academia, where data mining is just a single step in the whole process of finding patterns (John, 1997). John proposed that academia should use the definition of the commercial field. Therefore in this research data mining is adopted as the used term.

The data mining process is “an iterative process involving several steps, beginning with the understanding and definition of a problem and ending with the analysis of result and a strategy for using the results to gain advantage (John, 1997).

Both John (1997) and Fayyad et al. (1996) describe data mining as the same process, although they split the process in different steps and use different terms. In the table below an overview of the steps according to both John and Fayyad is given.

In both researches the step that actually runs the algorithm was called data mining. However data mining is very broad. One of the sub classes of data mining is machine learning. Because this research will use machine learning as the data mining algorithm, this term is already

Table 1: Data mining process according to Fayyad et al. (1996) and John (1997)

	Fayyad et al. (1996)	**John (1997)**
Big data Engineer	**1.** Data selection **2.** Data preprocessing **3.** Data transformation	1. Data collection and extraction 2. Data cleaning and exploration 3. Data engineering
Big Data Analytics	**4.** Machine Learning	4. Algorithm engineering 5. Running the Machine Learning algorithm
	5. Interpretation/Evaluation	6. Preliminary Evaluation of Results 7. Refining the Data and Problem 8. Using the Results

implemented in the overview above. Why this research uses machine learning will become clear in the next section of this report.

4.1.3. Machine learning

The machine learning field has many techniques and algorithms. To get a better understanding about the different methods available, it is helpful to categorize them. One of the most used ways to categorize machine learning methods is supervised and unsupervised learning (Ng, 2016b).

In supervised machine learning a dataset with 'right answers' is given to the algorithm. It can be used to solve regression and classification problems. In figure 3 the supervised learning picture is a classification problem. A classification problem predicts a discrete valued output. Using the known values for the variables the algorithm can predict if the outcome of the case is likely to be for example true or false (Ng, 2016a). In unsupervised machine learning a dataset is given which is not labelled into different categories. Unsupervised learning is used to find structure in datasets. One example, which is used in figure 3, is the clustering algorithm (Ng, 2016b). The difference between supervised and unsupervised learning becomes clear with figure 3. In supervised learning the answer, a circle or a cross, is known. The algorithm will learn based on x_1 and x_2 values when the answer is a circle or a cross. When presented with a known x_1 and x_2 but an unknown answer it will be able to predict the most likely answer. In unsupervised learning the answer is not known, however based on the x_1 and x_2 it is able to create two clusters.

As the goal is to learn from datasets with known output variables, this research is focused on supervised learning. Although it is now known that a supervised learning method is used in this research, there are still a lot of methods to choose from. Therefore it is necessary to divide supervised learning even further into new categories. A logical way to divide them is in the preferred output value. Supervised learning has two main categories, regression and classification.

Regression

Regression is the method of ML that is concerned about predicting the numerical value of a target variable based on input variables. For instance, estimating the costs of the design based on design specifications (Bilal et al., 2016). Weiss and Indurkhya (1995) state that the problem of approximating the values of a continuous variable is described in the statistical literature as regression. Given samples of output (response) variable y and input (predictor) variables x, the regression task is to find a mapping y. Based on this, regression is likely to be a useful technique for this research.

The most simple form of regression is linear regression, where one variable, the explanatory variable x, is used to predict the dependable variable y. Multiple linear regressions, as the name already suggests, is a technique where more than one explanatory variable is used to predict the dependable variable and is commonly used as a regression approach. As the costs of a building are dependable on more than one variable, multiple linear regression should be considered in the system design of this research.

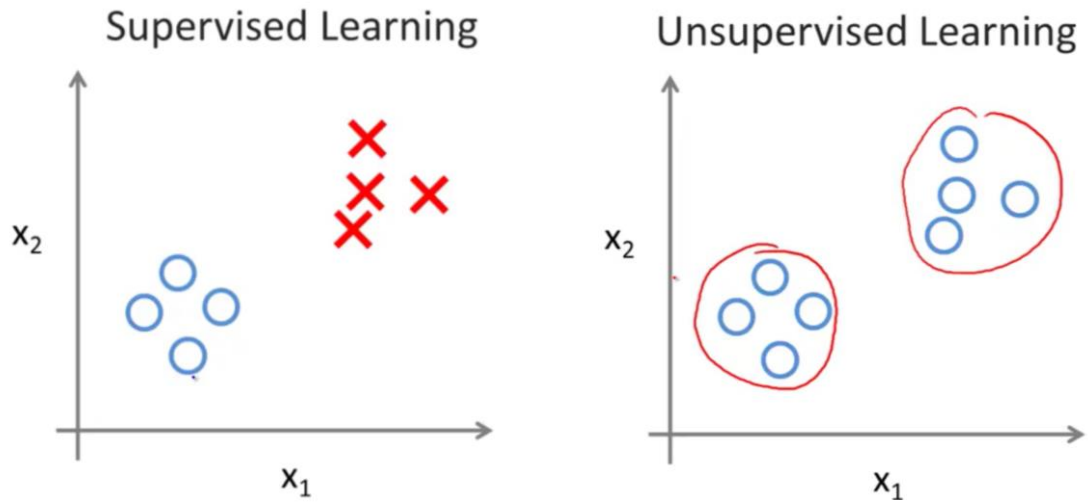

Figure 3: Supervised Learning compared to Unsupervised Learning (source: Ng, 2016b, edited)

Neural Networks are also a method that can be used as a regression technique. Although often used as a classifier, due to the relationship the logistic regression (Andew, 2011) neural network are able to solve regression problems. Berry and Linoff (2011) showed the use of neural networks to estimate the value of a house. They have been successfully used as a multivariate non-linear analytical tool, and are known to be highly effective in recognizing patterns from noisy and complex data, and estimating non-linear relationships (Khosrowshahi, 2011).

Classification

As mentioned before, besides regression there is another main category of machine learning, namely classification. Within classification there are a number of methods that can be chosen to predict costs. One of this methods is Case Based Reasoning. CBR is a data-mining technique that remembers similar situations applied to the solution of previous problems, and uses the information and knowledge from such situations to solve a new problem (Aamodt & Plaza, 1994). This technique does not require a clear model for problem-solving; rather, establishing cases is an important task in problem-solving. Since the model can be established by defining key attributes that express cases, this technique makes it easy to establish a model (Jin et al., 2012). The CBR process has four steps:

1. Retrieve: This phase inquires about and extracts the previous case most similar to the current one from a case base;
2. Reuse: This phase reuses the information and knowledge from the retrieved case for problem-solving;
3. Revise: If the retrieved case is not suitable for solving the new problem, this phase analyses the difference(s) between the new problem and the retrieved case, then revises the retrieved case accordingly;
4. Retain: This phase stores in the case base the solution proposed from the retrieved case so that it can be used in future problems

According to Vukovic et al. (2012) three techniques are usually used in the retrieval phase, namely nearest neighbour, inductive learning, and knowledge-guide. Inductive learning allows for the identification of training data or earlier knowledge patterns and similarities which are then extracted as generalized rules (Almana & Aksoy, 2014).

The most known implementation of inductive learning is the decision tree. A decision tree uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility (Quilant, 1987). Knowledge-guided approach applies existing domain and experimental knowledge to find suitable cases. Although this method is believed to be superior to the other two, knowledge-guided is difficult to apply because such knowledge often cannot be successfully captured and represented (Mi, Qian, Liu & Chang, 2008). Nearest neighbour is the most used technique, as it is easy to uses (Vukovic et al., 2012) (Sayad, 2010). It can used for classifying a new case by searching for the cases in the database that have the most similar features and assigning the class of its nearest neighbours to the new case (Vukovic et al., 2012).

K-Nearest Neighbour is a modified, extended version of Nearest Neighbour. The method gives better results because it searches for a k number of closest cases, where k is normally an odd number, and proposes the case that is most frequently found. It also allows for similarity. Similarity is the uses of weighted features. This means that the features that are the most relevant have higher weights and therefore have more influence on the predicted outcome as less relevant features (Vukovic et al., 2012).

Both classification and regression are possible for cost prediction. If classification is used the machine learning algorithm will determine the predicted cost based on the costs of the buildings most similar to the building the prediction is made for. In regression the total cost prediction is based on the sum of the input variables multiplied by their weights. More about the implementation of machine learning algorithms in the chapter 6, system analytics.

4.1.4. Evaluation of the data mining performance

As shown in table 1, one of the steps of the data mining process is to evaluate the performance of the data mining algorithm on the dataset. This performance can be tested with statistical evaluation measures. There are two types of measures. One type of measures evaluates the performance of classification and the other type evaluates the performance of numerical prediction (Witten & Frank, 2005). Although a classification algorithm is used in this research the outcome of both the regression algorithm as the classification algorithm will be a numerical prediction. Therefore the evaluation measures that can be applied on numerical prediction will be used in this research.

To evaluate the performance, the correlation coefficient, the root mean squared error (RMSE) and, mean absolute error (MAE) the root relative squared error (RRSE) are used. A correlation coefficient shows the linear dependence between the predicted value and the actual value. The result of the correlation coefficient is an index value between one and minus one, where one shows perfect positive dependence and minus one perfect negative dependence. If the value is zero there is no dependence. The most used correlation coefficient is the Pearson correlation coefficient (*r*) (Howitt & Cramer, 2005). The Pearson correlation coefficient is calculated with the equation:

$$Correlation\ coefficient = \frac{\sum(p-\bar{p})(a-\bar{a})}{\sqrt{\sum(p-\bar{p})^2}\sqrt{\sum(a-\bar{a})^2}} \quad (1)$$

$p = predicted\ value$
$\bar{p} = average\ of\ predicted\ values$
$a = actual\ value$
$\bar{a} = average\ of\ actual\ value$

The RMSE is one of the most used measures in statistics. It is used to calculate the difference between the calculated value and the actual value. An alternative to RMSE is the mean absolute error. The biggest difference between the measures is that because the RMSE is squared larger discrepancies are penalised harder. The RMSE can be calculated with the equation 2 and the MAE with equation 3.

$$Root\ mean\ squared\ error = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_n - p_n)^2} \quad (2)$$

$$Mean\ absolute\ error = \frac{1}{n}\sum_{i=1}^{n}|a_n - p_n| \quad (3)$$

The Fourth used measure to evaluate the performance is the root relative squared error (RRSE). It shows the performance of the algorithm against an more simple predictor, the average of all actual variables. If the value is below 100% the algorithm performs better than a prediction based on the average value. The closer the value comes to 0% the better the algorithm performs. The equation for the RRSE is:

$$Root\ relative\ squared\ error = \sqrt{\frac{\sum(p_n - a_n)^2}{\sum(a_n - \bar{a})^2}} \quad (4)$$

4.2. Building Information Modelling

As explained in the previous paragraphs, the BIM data will be used as a source for knowledge. Before BIM data can be used as knowledge, at first should be explained what BIM data are and where they come from. BIM data are the resulting data of using Building Information Modelling (BIM) techniques and tools in the building process.

4.2.1. What is BIM?

Building Information Modelling is a promising development within the architecture, engineering and construction (AEC) industries. With BIM technology, it is possible to create a digital and accurate virtual model of a building. When BIM is used correctly within the whole construction process the computer-generated model contains precise geometry and relevant data needed to support the construction, fabrication and procurement activities needed to realize the building (Eastman et al., 2011). But not only during the realization BIM technologies can be useful, even in the operation and maintenance phase of the building, BIM technologies can be very beneficial (Davtalab & Delgado, 2014).

A useful definition of BIM was given by Campbell (2006). BIM is an intelligent simulation of architecture that exhibits the following six key characteristics:

- Digital;
- Spatial (3D);
- Measurable (quantifiable, dimension-able, and query-able);
- Comprehensive (encapsulating and communicating design intent, building performance, constructability, and include sequential and financial aspects of means and methods);
- Accessible (to the entire AEC/ owner team through an interoperable and intuitive interface);
- Durable (usable through all phases of a facility's life).

One of the important tools of BIM Authoring Software, which is used as a tool to create a BIM model. Instead of drawing lines with a CAD program to create a 2D visualization of the building, BIM Authoring Software allows the user to create an object based 3D model. This not only allows the user to view the model in both 2D and 3D, it also allows to add other relevant data to object. For instance it is possible to add the function and material of the object (Eastman et al., 2011).

A BIM model is characterized by building components that include data that describe how they behave and are consistent and non-redundant data. Building components are objects that have digital representations and data about what they are. They can be associated with computable graphics, data attributes and parametric rules. These components also describe how they behave. This makes it possible to create analyses of the building and use it in work processes. This BIM model also contains coordinated data.

Next to the BIM model, another very important part of BIM is the interoperability between members of the building project team, to ensure that every member of the team has access to the latest project data. It is possible to allow every member to have access to all the data, these data need to be:

- Real time data exchange
- Share in a predefined format

A cloud based server is the most used way to ensure that project data is both shared in real time as it is accessible from every location.

For the predefined format there are two primary approaches, (1) use a proprietary file format and therefore stay within one software vendor's product (or use another product that is allowed by the vendor) or (2) use different vendors that can exchange data using non-proprietary file format that is an universal supported standard. The advantage of the first approach is that it allows for tighter integration among products in multiple directions. For example, a change in one model results in a change in all other linked models. The obvious disadvantage is that every member of the project team is forced to use the programs of the specific vendor. This could potentially cost a lot of money considering both licensing and training. The second approach would solve the disadvantage of the first approach. The disadvantage of this approach is that the current universal standard, IFC, is not designed to carry all relevant data.

Although the disadvantage of not being able to carry all relevant data, in this research the IFC standard will be used. The necessary data are collected via different clients of IBIS BV and therefore the acquired data will be created with software from different vendors. This makes the first approach not possible. Although this will result in some loss of data, it also has an advantage. As the tool created for the research uses a universal standard, the tool is reusable in other researches. More about the universal standard IFC in the section below.

4.2.2. Industry Foundation Classes

The Industry Foundation Classes (IFC) are the open and neutral data format for openBIM. (BuildingSMART, nd). It is developed and maintained by buildingSMART International. The IFC specification is a neutral data format to describe, exchange and share information typically

used within the building and facility management industry sector. It provides a data model structure for sharing data across various applications used in the building.

As a BIM model is characterized by building components, so is an IFC model. In an IFC model these building components are called IfcObjects. “An IfcObject is the generalization of any semantically treated thing or process” (buildingSmart, nd). These objects are divided in six subtypes, see figure 4.

Although the other five subtypes can contain a lot of information, for this research only the IfcProduct subtype is interesting because everything related to geometric or spatial context is an IfcProduct (buildingsmart, nd). An IfcProduct is also divided in subtypes (see also figure 4). As an IfcElement is a “generalization of all components that make up an AEC product” (buildingsmart, n.d.), only this subtype is relevant for this research. From the subtypes that holds an IfcElement, the IfcBuildingElement types are the types we are interested in as they are the “major functional part of a building, examples are floor, roof and wall” (buildingSmart, nd).

4.2.3. Improvements of the shared BIM Models

In the exchange of information between sender and receiver three factors are important to make sure the receiver is able to interpret the information received as the sender intents (Bell & Bjørkhaug, 2006). The first part is the format in which the information is exchanged. As already discussed, in the presented research this will be the IFC data format. The other two parts are a specification of which information to exchange and when it needs to be exchanged and a mutual understanding of what the exchanged information actually is.

The IFC data format allows various data to be shared in various ways. If the receiver of the information, in this research the extractor, wants to be sure it is able to utilize the information he receives, the sender and receiver need to be in agreement which and how the information is shared (Bell & Bjørkhaug, 2006). To ensure the agreement of how the information is shared buildingSMART created the International Framework of Dictionaries (IFD). Where the IFC format is very broad to support all the different BIM authoring tools, IFD format “specifies a language-independent information model which can be used for the development of dictionaries used to store or provide information about construction works. It enables classification systems, information models, object models and process models to be referenced from within a common framework” (International Organization for Standardization [ISO], 2007). It is a language to define ontologies, “which defines concepts and semantic relations between these concepts” (Böhms et al., 2013). The Nederlandse Conceptenbibliotheek voor de bouw (CB-NL) is the Dutch implementation of IFD. The ontologies created can be used to standardise the way information within IFC is shared. For the presented research the standardised way of sharing information is useful as the effectiveness of the extraction tool created in the practical application part of this research is highly influenced by the script ability to interpret the information within the IFC data.

Another way to ensure better quality of BIM models is to have an understanding of which information to share. The Information Delivery Manuals (IDM) can be used for the documentation of existing or new processed and describes the associated information that needs to be exchanged between parties. The main purpose of IDM is to make sure the relevant data are communicated via a structured way such as the data can be interpreted by the software on the receiving side (buildingsmart, nd.). For the Dutch AEC industry the ‘BIM basis

informatieleveringsspecificatie (BIM Basis ILS)' is defined and it has already been adopted by many companies. As the 'BIM basis ILS' has been introduced in 2016 and it started with just 14 organisations within the AEC industry it is highly unlikely that the used BIM models within the presented research are in compliance with the BIM basis ILS. In the practical implementation part of this research it will be discussed what the effect would have been if the BIM Basis ILS would have been used on all IFC.

4.2.4. BIM for cost estimation

In the previous years, a lot of research has been done in the combination of BIM and cost estimation. As a BIM model is a virtual model of the building, it is a unique source of information in the building process. If modelled correctly, a BIM model comprehends all the building elements and the relations of these building elements to each other. Most research has aimed to use BIM models as a source for quantity take off (QTO) and the use of this QTO to determine the cost estimate based on unit rates. This research also tend to focus on BIM cost estimate in the difference phases of the building process. Both these research topics could provide useful information for this research.

Part of the research on the topic of BIM for cost estimation focusses on the use of BIM tools to assist in the process of QTO. The estimator uses a virtual model of the building to estimate the QTO manually and the BIM tools only allow for a more efficient process (Shen & Issa, 2010). Examples of commercial application of this research are the software packages used in the Netherlands, like IBIS-TRAD, Vico Office and BIM ncalc (Van de Laar, personal communication, September 1, 2016; Construsoft, 2017; Knaan Bouwcomputing, 2017). All these packages are focussed on improving the process of the traditional way of creating a cost estimate, by using BIM to determine the quantities faster and with more accuracy. As this process still requires a lot of human effort, it cannot provide solutions for the presented research (Shen & Issa, 2010).

There are also researches done to automate the cost estimation. Ma, Wei and Zhang (2013) proposed a model for such automatic cost estimation. The goal of this research was to have a bill of quantities in compliances with GB50500-2008, the Chinese national mandatory specification of cost estimation for tendering of building projects. The used data format of these BIM models is IFC. In their research Ma et al. (2013) divided the IFC format in three useful parts for cost estimation. One part expresses the decomposition of building elements into construction products. Another part contains three types of information of the construction products, i.e. product type, geometric information and material information. The last part expresses accessory, opening and intersection relationships between the building elements as well as the construction products. They created a prototype QTO software application and tested the prototype's performance on more than 3000 construction products. 67% of the construction products were automatically classified and extracted to a QTO. For the other 33% it was necessary to add information manually that was lacking from the IFC model. From this research can be concluded that it is possible to extract all the necessary information for a cost estimation from the IFC data format. It can also be concluded that in some cases not all necessary information is modelled and therefore information is missing in IFC models.

Research by Cheung et al. (2012) demonstrate the use of a tool to evaluate the functional, economical properties of a building as well as the performance, during the building design. The economical properties are automatically calculated using the measurements that are

extracted from the BIM model. Four levels of cost estimation are suggested, where the level should be selected according to the level of information available. The first level is calculated based on the site boundary and massing of the building. In the second level the available information is updated with the number of floors and storey heights. On this level, default specifications of building elements are presumed to be present in the model. The third level consists of defining spaces with vertical elements. For this level the available information is updated with the floor plan of the building. In the highest level, information, like building material, is provided for the building elements.

4.3. Conclusions BIM as a knowledge source

BIM has been widely adopted in the AEC industry as it has enhanced the performance of projects. As a result BIM models are available which contain a lot of embedded knowledge. However this knowledge needs to be made explicit. In this research it is suggested that the data mining process could be used to make this knowledge explicit and even create new knowledge in the form of a cost estimation tool to estimate the building costs in the early stages of the building process.

The data mining process can be divided in a data engineering and a data analysing part. In the data engineering part the available data are converted to a usable dataset for the data analysing part. The part of data analysing consists of using algorithms to analyse the dataset and to interpret the result. One of the types of algorithms that can be used is the machine learning type. Machine learning (ML) algorithms can be used to learn from and make predictions on data. There are two categories within ML. The supervised ML category uses a dataset with known output variables to make predictions on new data. Unsupervised ML is used to find structure within the dataset. This research will implement supervised ML because our goal is to make a prediction on building costs. Within supervised ML again two categories can be found, regression and classification. As both are useful to predict costs, from both categories an algorithm will be used in this research.

The data mining process needs input data in order to be able to make predictions. As BIM has been implemented within most of the AEC industry and one of the results of BIM is a BIM model with all the information about the building, these BIM models have a great potential as a knowledge source. A universal standard for sharing BIM models is the IFC format. It provides a data model structure. For this research the physical information of the building model is interesting. All these are located within the IfcBuildingElement class.

This research closely relates to the research of cost estimation with BIM. As BIM contains all the physical information of the buildings, these information can be extracted in the form of a quantity take off. Cheung et al. (2012) have demonstrated a tool that allows for cost estimation directly from the BIM model. They suggest that during the design of the building four levels of cost estimation can be made. The first level is a cost estimation based on the site boundary and massing of the building. In the second level the model is updated with information about the amount of floors and story heights. For the third level also the spaces and vertical elements need to be defined. The most detailed level consists of a cost estimation where also specific information, like building materials, is used. This separation of levels is very useful as it allows for a method that bases cost estimation on the level of available information.

Ma Wei and Zhang (2013) showed the real potential of automatic QTO from BIM models. In their research they were able to (semi) automatic extract the QTO in compliance to the Chinese mandatory specification of cost estimation in the tender phase.

4.4. Conclusion of literature review

Based on the information obtained during the literature review, it is now known how BIM models can be used as the data source for KM. By using the data mining process one should be able to convert the knowledge embedded in BIM models to useful explicit knowledge for cost estimation. The steps necessary to convert BIM models with corresponding building costs to a cost estimation model are:

1. Understanding and defining the problem
2. Collect necessary data
3. Convert non-IFC BIM models to the IFC data format
4. Convert the IFC models to a dataset
5. Add the corresponding costs
6. Clean the dataset
7. Select a useful subset
8. Engineer a useful data mining algorithm
9. Use the algorithm
10. Evaluate results with numerical prediction measures
11. Redefine data and problem
12. Use the generated model for cost prediction

On the next page the proposed model is given, this model combines the enumeration given above with the previously mentioned SECI model on page 37.

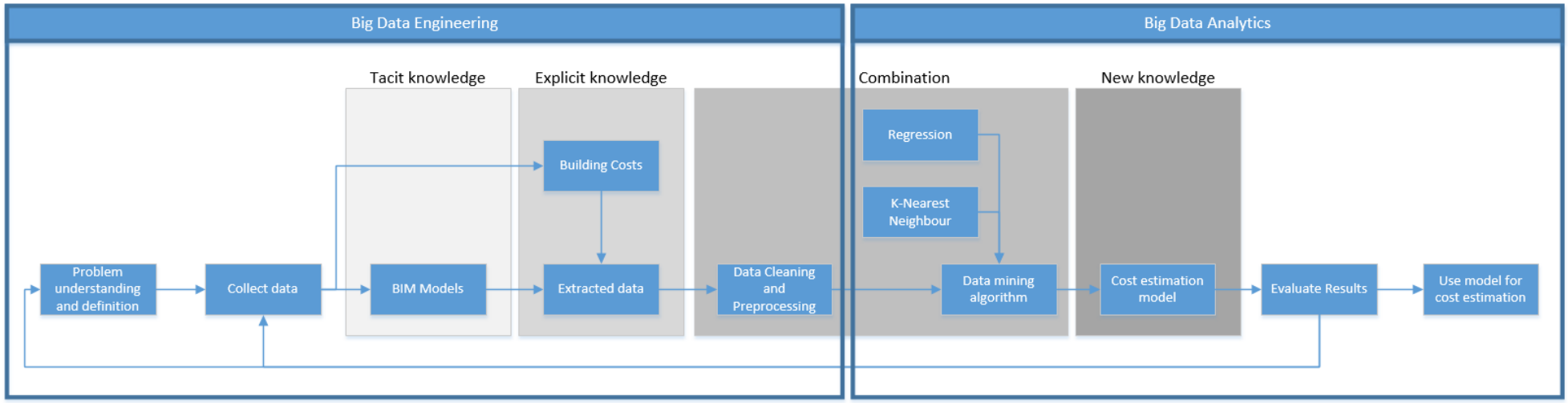

Figure 4: Proposed model

This page is intentionally left blank

Part I: Problem statement

Part II: Theoretical review

Part III: Practical implementation

Part IV: Conclusions and recommendations

This page is intentionally left blank

5. System Design

From the literature review it is known that the process of discovering advantageous patterns in data is called data mining (John, 1997). This process has been adjusted to fit the presented research and consists of the following steps.

1. Understanding and defining the problem
2. Collect necessary data
3. Convert non-IFC BIM models to the IFC data format
4. Convert the IFC models to a dataset
5. Add the corresponding costs
6. Clean the dataset
7. Select a useful subset
8. Engineer a useful data mining algorithm
9. Use the algorithm
10. Evaluate results with numerical prediction measures
11. Redefine data and problem
12. Use the generated model for cost prediction

The first step of the data mining process, understanding and defining the problem, has been dealt with in the previous part of the research. This part of the research will focus on the other steps of the process. In this chapter steps 2 to 7 are discussed. In chapter 6 the steps 8 to 12 will be explained.

5.1. Data collection and extraction

The second step of the data mining process is data collection and extraction. As discussed in the literature review, the BIM and costs data are already created for other purposes and stored in diverse databases. These data need to be collected and stored in a single database. For this research, construction companies in the Netherlands were asked to provide their BIM and matching costs data. This has resulted in 15 sets of data, each containing one building. Three of these sets were not suitable for this research because the type of buildings that were built in the projects were industrial constructions as opposed to the 12 other sets that were residential buildings.

Although 12 suitable sets of data were collected, these data are not directly usable for data mining. First the usable data need to be extracted and stored in a format that can be accessed by the data mining algorithm (John, 1997). Thus first it has to be determined in what format the original data are stored and in what format the machine learning algorithms need to analyse the data.

In the literature review is explained that BIM data could be shared in either a proprietary file format or non-proprietary file format. Also was explained that the format used in this research would be a universal standard format and therefore a non-proprietary format, because it allows all the provided BIM data to be used. As IFC is the most widely accepted non-proprietary BIM data format, IFC is used as the file format in this research.

The data collected had both types of formats as in advance there was no demand given to the construction companies in which format the BIM data should be provided. The reason for not demanding the data to be provided in the IFC format was because it would require more effort from the construction company and therefore it could reduce the chance of cooperation by

those companies. It was not worthwhile taking this risk as converting the format from the BIM authoring tool to IFC would only require access to the BIM authoring tool and this would also take time. Both were available in this research.

From the collected datasets, five of them were provided in the original format created by the BIM authoring tool and seven were provided in the IFC format. From the five provided in the original format all were provided in the RVT format, which is the original format of BIM data created in Revit, a product of Autodesk.

The costs data were provided in two file formats. One was provided in PDF, the rest was provided in XTBF, the file format used by Ibis TRAD, the building costs calculation package of Ibis BV (A. Lassche, personal communication, September 14, 2017). It was chosen to use only to total costs of the building. This choice was made based on the amount of data available. As this is a small amount, it is necessary to keep it as generalized as possible. As only the total costs data need to be determined, it was chosen to convert these data manually to the required format.

Data extraction

In the data extraction process the necessary information is extracted from the input data and written to the output data (see figure 6). As the input data are already known, a selection needs to be made on the output data and how these data are extracted. As the goal of the extraction process is to get the necessary output, first the output data will be selected. Afterwards a selection will be made of how to extract the necessary data from the IFC data file (input data).

To extract the usable data from the collected data, there needs to be a file format to store the usable data into it. This file format needs to be suitable to be used with the machine learning algorithm. It is also necessary that the programming language used to extract the usable data from the collected data can write to this specific file format, preferably as a standard option.

The machine learning algorithms will be provided by R. R is an open source software environment for statistical computing. One of the many applications of R is data mining (The R foundation, nd). The most used file format for R is a simple text format. To separate the text filed a separator is used. A comma is an easy and useful separator. A simple text file format with a comma as a separator is called a CSV, an abbreviation of Comma Separated Value. CSV is a standard file format often used to exchange data between application and a good option for importing data into R (Togaware, nd).

As it is known that we need to convert IFC to CSV, it is possible to determine how the data can be extracted. There are multiple commercial and non-commercial tools available for the interpretation of IFC models. Solibri Viewer is one of these tools that allows users to view IFC models. It is also able to determine properties of the IfcObject, like area, thickness and volume. However it does not allow for automatic data extraction to a data format that can be used with a machine learning algorithm. As mentioned before, it is important that as much as possible is automated in the knowledge management system because the more people have to do this themselves, the lower the chance of actual implementation. Because of this it was chosen to program a new extraction tool for this research.

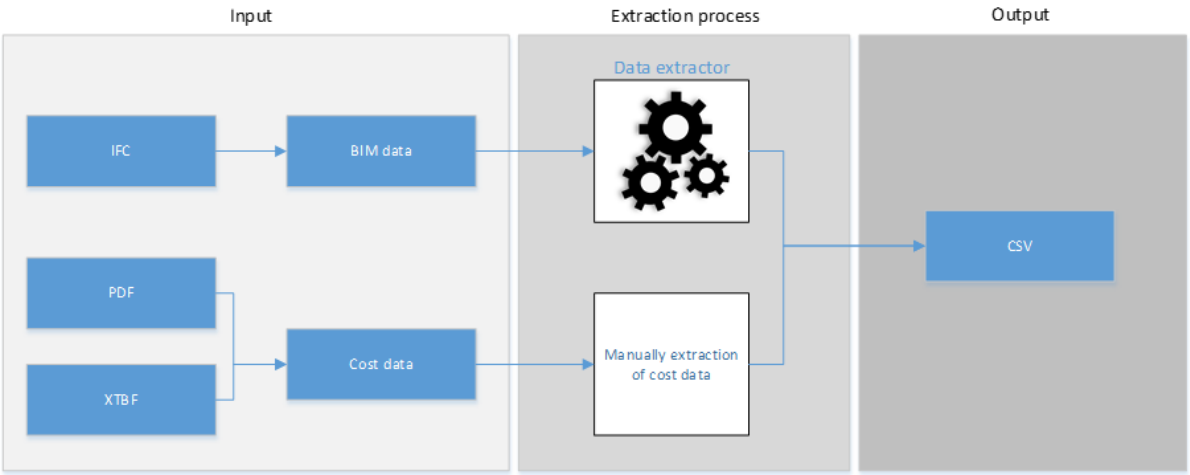

Figure 5: From input data to CSV

The programming language that is used is Python in combination with IfcOpenShell and PythonOCC, the programming language used in the research by Krijnen (2015). This was chosen because IfcOpenShell is one of the very few libraries that allow programming access to IFc model. And because Krijnen very willingly shared his knowledge, this was used in the development of the extracting script.

The Python script uses the IfcOpenShell module to interpret the object based IFC data. It sorts the data by IfcElement. This is done by creating a Python list based on the IfcElement that the extractor needs to extract using the '*by_type*' function of the module. As the extraction script needs to be able to understand what kind of elements it is extracting from the IFC file, it is necessary to program for every element how it can be classified, mostly by IfcElement and how it can calculate or interpret the necessary information to extract. Therefore it is necessary to select possible relevant IfcElements to be extracted.

The goal of the extractor is to create a CSV format file that contains the dataset to be used in the data mining algorithm. Essentially this file would contain a quantity take off (QTO) of the IFC file. Based on the literature research, the cost estimation can be determined by two different types of quantity take-off. Either a very detailed level of extraction as Ma et al. (2013) used or a very general extraction as used in the research of Cheung et al. (2012). This extracted quantity take off would be used as the dataset of the big data process. The one suggested by Ma et al. (2013) would result in a very detailed dataset, in which every element of the building is extracted. This would give a very precise outcome when estimating the costs of the building with known unit rates. However, in this research the unit rates are calculated through the machine learning algorithm. For this a statistical correlation has to be found between the different buildings. With the amount of data available it is more likely to find this correlation with general elements. Therefore the QTO suggested by Cheung et al. (2012) is chosen. Based on the research of Cheung et. al (2012) the following components need to be extracted:

Table 2: Building elements to be extracted

Group elements	Elements		Unit
Substructure	Substructure	Ground floor area	m^2
Superstructure	Frame	Area of all floors	m^2
	Upper Floors	Same as frame only without ground floor	m^2
	Roof	Area of roof as viewed from plan	m^2
	Stairs and ramps	Number of stairs within building	no
	External walls	Area of external surface of the external walls	m^2
	Windows and external doors	Area of external surface of the windows and external doors	m^2
	Internal walls and partitions	The vertical area of the inner walls	m^2
	Internal doors	Amount of inner doors	no
	Wall finishes	Vertical surface area of the walls that needs finishing, area per wall finish type	m^2
	Floor finishes	Upper surface area of the floors, area per floor finish type	m^2
	Ceiling finishes	Lower surface area of ceilings, area per ceiling finish type	m^2
Fittings and furnishings	Fitting and furnishings	Total area for fitting and furnishings	m^2
Services	Services	Total area for services	m^2

Based on the level of detail within the BIM models, it is not possible to determine the finishing of the wall, floor and ceiling, the fittings and furnishings and services of the buildings. In only four of the BIM models information about the finishing and the fittings and furnishings were available. For services it was even more problematic as for only six IFC models the spaces were defined and even if the spaces were defined only little information was available about the quality of service within the spaces. Therefore these elements are ignored in this research.

The substructure, frame, upper floors and roof could be determined by calculating the area of the IfcSlabs located on the corresponding floor(s). The stairs can normally be determined by the IfcStair element. However in some IFC models the stairs are exported to IfcBuildingElementProxy or IfcFurnishingElement. Using a filter on these IFC elements the amount of stairs could be determined. The external and internal walls can be extracted with the IfcWall elements, and the doors with the IfcDoor elements. Both needed a filter to determine if they are located inside the building or located in the external wall. The windows can be determined with IfcWindow.

Table 3: Overview of building elements with the corresponding IfcElement

Group elements	Elements		IfcElement
Substructure	Substructure	Ground floor area	IfcSlab
Superstructure	Frame	Area of all floors	IfcSlab
	Upper Floors	Same as frame only without ground floor	IfcSlab
	Roof	Area of roof as viewed from plan	IfcSlab
	Stairs	Number of stairs within building	IfcBuildingElementProxy/ IfcFurnishingElement/ IfcStair
	External walls	Area of external surface of the external walls	IfcWall
	Windows and external doors	Area of external surface of the windows and external doors	IfcWindow and IfcDoor
	Internal walls and partitions	The vertical area of the inner walls	IfcWall
	Internal doors	Amount of inner doors	IfcDoor
	Wall finishes	Total vertical surface area of the walls that need finishing	IfcWall
	Floor finishes	Total upper surface area of the floors	IfcSlab
	Ceiling finishes	Total lower surface area of ceilings	IfcSlab

This all results in a dataset with the following attributes: area of ground floor, area of upper Floors, area of roofs, amount of stairs, area of external walls, area of external openings, area internal walls, amount of internal doors.

The process of creating the extractor script contained a lot of trial and error. The biggest challenge was the data management of Python. In the original script, based on the scripts of Van Strien (2015) and Krijnen (2015) all the objects in the IFC file were imported and converted to shapes at the same time, which allow for interpreting the objects as related to each other. The problem however was the amount of objects that the Python script needed to interpret. As Van Stien (2015) and Krijnen (2015) both used a small IFC model called IfcOpenHouse to test their script and this research uses IFC models from real construction projects, the amount of objects is significantly higher in the latter. The IfcOpenHouse model contains only 47 IfcProducts in contrary to the average of 6255 IfcProducts that the models used in this

research contain. As a result the original script crashed due to the limited amount of memory available to run the script.

This problem was eventually solved by changing the order in the script. Instead of creating all shapes directly, first a list was made of elements needed to calculate a particular attribute. An overview of the element needed per attribute can be found in table 3. Because for the floor area the floor level is also important, the list of IfcSlabs was split by floor levels using the link between the IfcSlab and IfcBuildingStorey.

For some elements the location of the element matters. For example for the IfcWall it matters if it is an inside or an outside wall. In most IFC models this can be determined with the IsExternal property in the PSet_WallCommon. However as one of the IFC models did not contain the PSet_WallCommon this was not used. A filter was used based on name and material. Because all the outside walls in the IFC models are made of brickwork a filter was used to find all IfcSlabs with material type brickwork. In the theory it was already discussed that there are initiatives to ensure that the proper information is shared within the IFC model. The BIM Basis ILS prescribes the use of Pset_WallCommon to share information about the properties of the wall. If this would have been implemented by the exportation of all BIM models to the IFC data format, the Pset_WallCommon could have been used and therefore

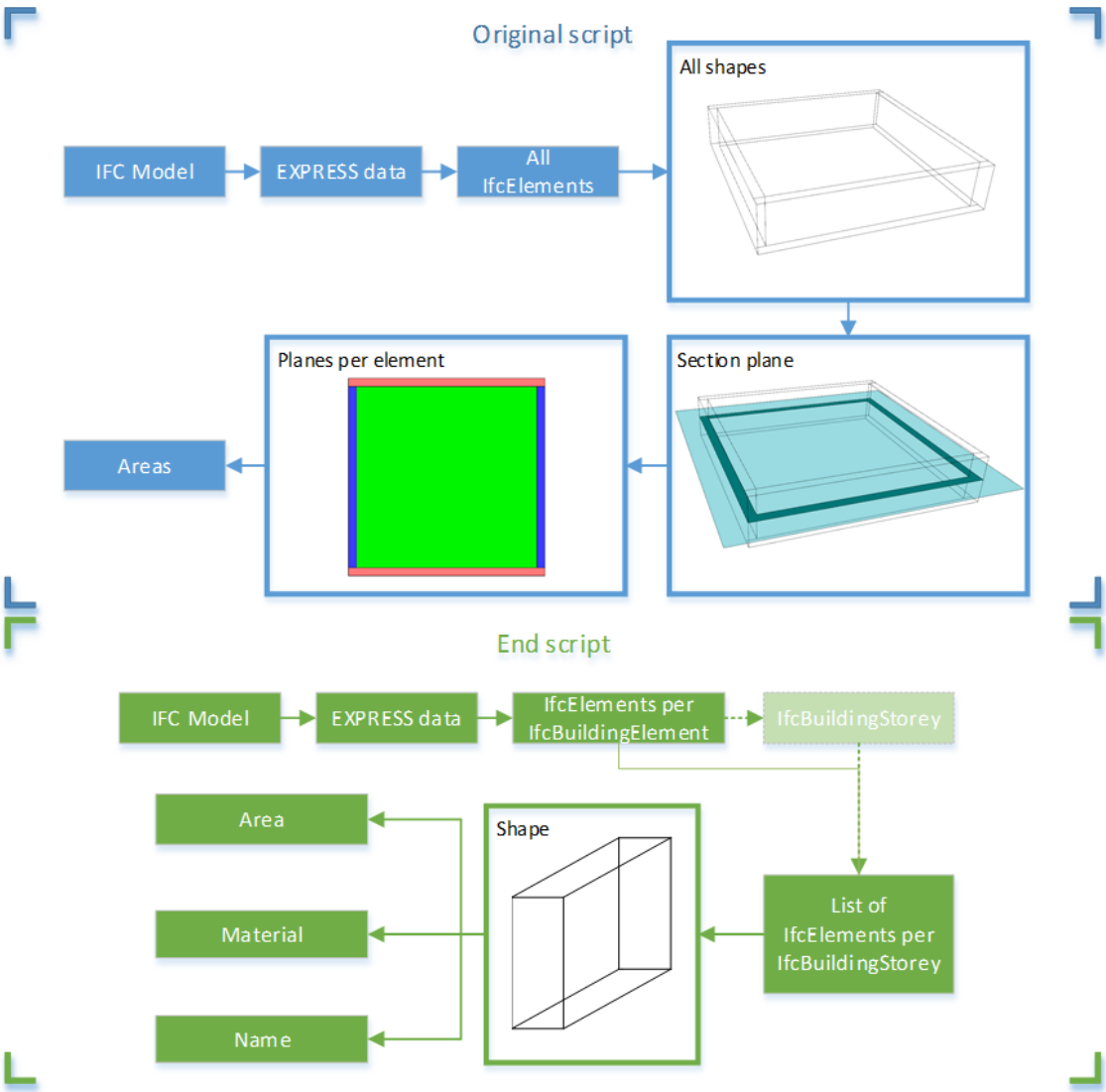

Figure 6: Structure of extractor script

the script would have worked also with external walls that are not made of brickwork.

It should be noted that although changing the order of the script made the script work, it also means a loss of information. The new script analyses every element individually. Therefore most of the information about how the elements are located to each other is lost. This made it impossible to analyse for example the area of particular rooms based on building elements. In the left part of figure 8 it is shown how the area of a room could be determined if the preferred script would have worked. On the right part of this figure the working of the actual script is shown. In the table below, the dataset as created by the extractor can be found. Notable is the 'Count Stairs' attribute as two of the return values are zero. This is because no information about stairs could be found in those IFC models. It is very common in data mining that the dataset is cleaned after extraction. More about this in the next section.

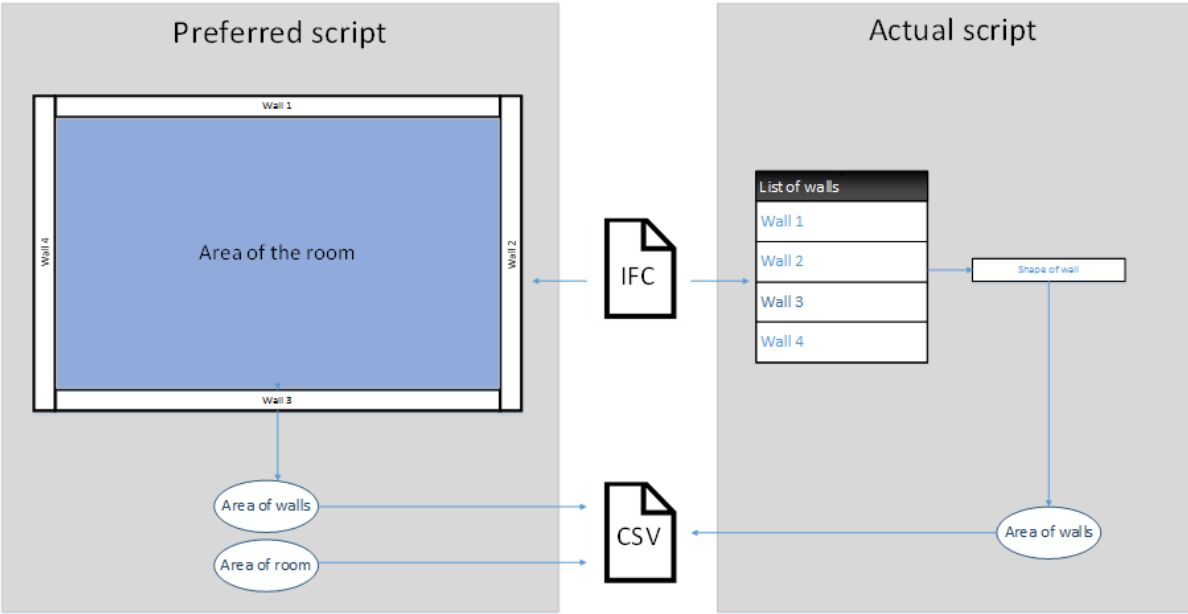

Figure 7: Distinction between the preferred and actual working of the script

Table 4: Overview of the extracted dataset

Filename	Area Ground Floor	Area Upper Floor	Area Roof	Count Stair	Area External Walls	Area External Openings	Area Internal Walls	Count Internal Doors
A.ifc	842	652	842	0	2902	726	3452	252
B.ifc	584	2101	654	9	2614	668	3352	466
C.ifc	595	1984	880	4	3607	673	2326	421
D.ifc	682	1948	850	4	3224	690	2894	426
E.ifc	567	2234	615	5	2405	647	2962	176
F.ifc	880	696	880	26	1829	500	2608	130
G.ifc	569	931	603	20	1984	315	1989	100
H.ifc	280	186	321	12	920	149	1147	54
I.ifc	233	359	252	8	960	130	724	40
J.ifc	374	1374	475	6	1891	414	1936	294
K.ifc	1496	364	1496	32	6560	2017	2846	177
L.ifc	432	575	436	0	1307	279	674	102

5.2. Data cleaning and Exploration

In a traditional data mining process, the amount of data in the database is very large. As large databases contain a lot of samples, the chance of errors in the created dataset is very likely. In order to be successful, it is required that the system is robust regarding errors in dataset (John, 1997).

For large datasets, a sophisticated method is needed to test the robustness of the datasets. This is necessary because of its size, it is not really possible to check the data manually. The dataset created by the extractor however is small. Although it is preferred to have a large database, the small size of the used database allows us to check the extracted data by hand.

In this research most of the data cleaning is actually done by the extractor. For the structure of the thesis however, it has been chosen to separate the extraction and cleaning of the data. The way the set data was checked has been done by comparing the extracted data with the data in the original costs estimate. Unfortunately the data appeared to be anything but robust. For almost every project the extracted data did not match the data from the original cost estimate. One of the most important reasons is that the IFC models were not made for this purpose. Most of them were once created by the architect and never updated in the further process. Even though the architect also created the specifications of the project, many cases were found in which the IFC model did not match the made specifications.

One of the most commonly found errors in the data is mislabelling of the IfcObjects. As explained in the literature review IfcObjects are the generalisation of the components in an IFC model. The most often mislabelling was the mislabelling of stairs. In the IFC models the stairs were either labelled as IfcStair, IfcFurnisingElement or IfcBuildingProxyElement. As explained in the previous chapter, in the different models the stairs were modelled with three different IFC elements, IfcStair, IfcBuildingElementProxy and IfcFurnishingElement.

This is a problem because as the name IfcFurnishingElement suggests, this type of IfcElement is a generalization of all furniture related objects (buildingSMART, nd). This is even more the case for IfcBuildingElementProxy as this is “a proxy definition that provides the same functionality as an IfcBuildingElement, but without having a defined meaning of the special type of building element, it represents” (buildingSMART, nd). As a result a query on the name of the IfcFurnishingElements and IfcBuildingElementProxy is necessary to find the stairs within all of the IfcFurnishingElements. This is especially hard because as there is no standardisation in the naming of IfcElements and therefore it is not necessarily the case that the word ‘stair’ or a translation of this word for example the word ‘trap’ in Dutch, is within the name of the IFC element.

In only two models the stair was modelled as an IfcStair. Five models had the stairs modelled as an IfcFurniture with the word ‘trap’ in the name of the elements. In three of the models the used element for the stair was IfcObject and there was no indication in the name that is was a stair. Therefore it was necessary to query on the used name, ‘nl0-dht_nl+1-oht_2kwart,’ to find the stair. In the dataset, twice the extractor returned zero for the attribute Count Stairs. In these IFC models the stairs were not modelled. For the dataset to be robust the actual values are updated manually.

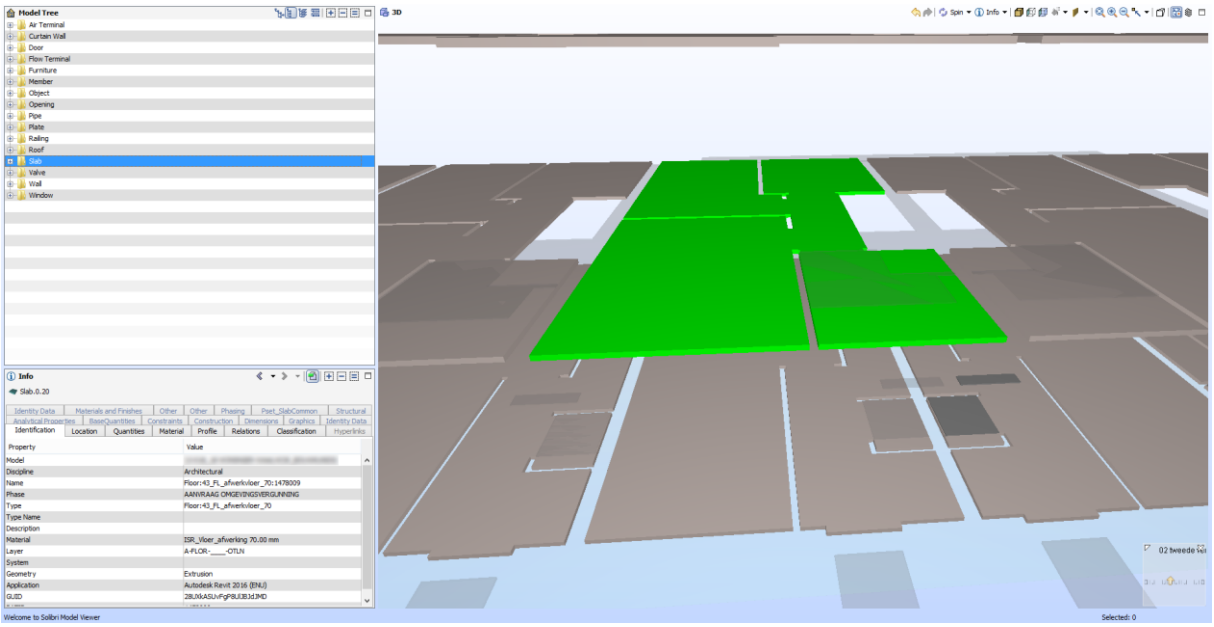

Figure 8: Model without structural floor

The general problem of mislabelling the IfcObjects could be solved with the use of either an IDM, more specific the BIM Basis ILS for the Nederlands or an IFD. The BIM Basis ILS prescribes the use of correct entities and the Nederlandse Revit Afspraken even describes how to use a Mapping Table within Revit to make sure the correct entities are exported. The Nederlandse Revit Afspraken is a document created by Pijffers (2016) in compliance with BuildingSMART Benelux which is the managing organisation of the BIM Basis ILS.

During the exploration of the IFC models two projects were found in which no structural floors were defined using IfcSlab (see figure 9). In both cases only the screed floors and roof insulation were modelled as IfcSlab elements. After some exploration, for one of the models the structural floor could be found as a IfcBuildingElementProxy. After comparing the other IFC model with the cost files, it was found that the structural floor was a wide slab floor. Because of this, the extractor needed extra code to check the presence of different types of floors and to use the appropriate type for the specific model. As already explained in the BIM Basis ILS is prescribed that the appropriate entity should be used, in this case a IfcSlab. This would have solved the problem for one of the two models. However for the other model this would not have solved it as the structural floor was never modelled.

In the dataset the floor area is important for different attributes. As it is used for different attributes calculating the total floor area is not enough. The area is needed per floor. To be able to calculate the area of a floor, the script needed to be able to interpret on which floor the IfcSlab is located, as the IfcSlabs were used to calculated the floor area. As IfcSlabs are linked to a IfcBuildingStorey and IfcBuildingStorey's are used to model different floor levels, in most IFC models a combination of IfcSlabs and IfcBuildingStorey was enough to interpret the area per floor level. However in some models, the IfcSlabs were not linked to the correct IfcBuildingStorey. For these IfcSlabs an offset was used to give them the correct height. For this problem also additional code was written (see appendix 2, lines 332-346) to get the correct output from the extractor. The code however is very model specific and therefore would not work with other IFC models that present the same kind of problem. If the model would have been made in accordance to the BIM Basis ILS this problem would have been solved as the BIM Basis ILS only allows for one main level for every building storey.

In his dissertation, John (1997) said about the data: “in the end the data stored in a database is no better than the typing skill of the data-entry clerk or the decisions made by experts that are recorded in the database” (pp. 51-52). As most IFC models are created by architects to show their design and limitedly to never updated during the rest of the design phase, few decisions made by experts are recorded in the data. In the normal building process this is not really a problem, as most contractors still use the 2D drawings and the specifications to create a cost estimate. But as the AEC industry is slowly starting to use the possibilities of BIM data, the accuracy of the BIM models will become more important. As the quality of the IFC files is problematic for automated cost estimation based on IFC files, we can already conclude that with the current state of the AEC industry regarding BIM, it is not yet possible to use the BIM files for automated knowledge management. However if the BIM Basis ILS would be implemented much of the problems that occurred would be solved.

5.3. Data engineering

5.3.1. Sub dataset selection

The data engineering part of the data mining process consists of a large number of possible activities. One of the most important activities is to select a subset of attributes that will be used in the rest of the data mining process (John, 1997).

The subset selection has already taken place in the data extraction phase. This is because every attribute that is extracted from the IFC data, had to be programmed into the data extraction script created for this research, as explained in section 7.1. Therefore already choices were made of which types of objects to extract from the data and thus which possible attributes would likely be relevant.

However in the dataset there are still 8 independent variables compared to 12 observations. This is already a high number of independent variables compared to the number of observations. This becomes even more problematic when the dataset needs to be split in a trainings dataset and a test dataset. To reduce the amount of independent variables a method is needed to select the most relevant variables. This is called feature selection. To evaluate the dataset, first a correlation plot is made of the dependent variables and all the independent variables, see figure 10. It shows that most of the attributes have a very low correlation. According to Guyon and Elisseeff a variable ranking method is the best option to use whenever the data is noisy. The result of the ranking method is the following ranking:

1. AreaInternalWalls
2. AreaExternalOpenings
3. AreaRoof
4. CountInternalDoors
5. AreaUpperFloor
6. AreaGroundFloor
7. AreaExternalWalls
8. CountStairs

Therefore it has been chosen to apply stepwise regression for the regression model. In stepwise regression every step of a variable is added to the algorithm as long as the performance of the algorithms increases.

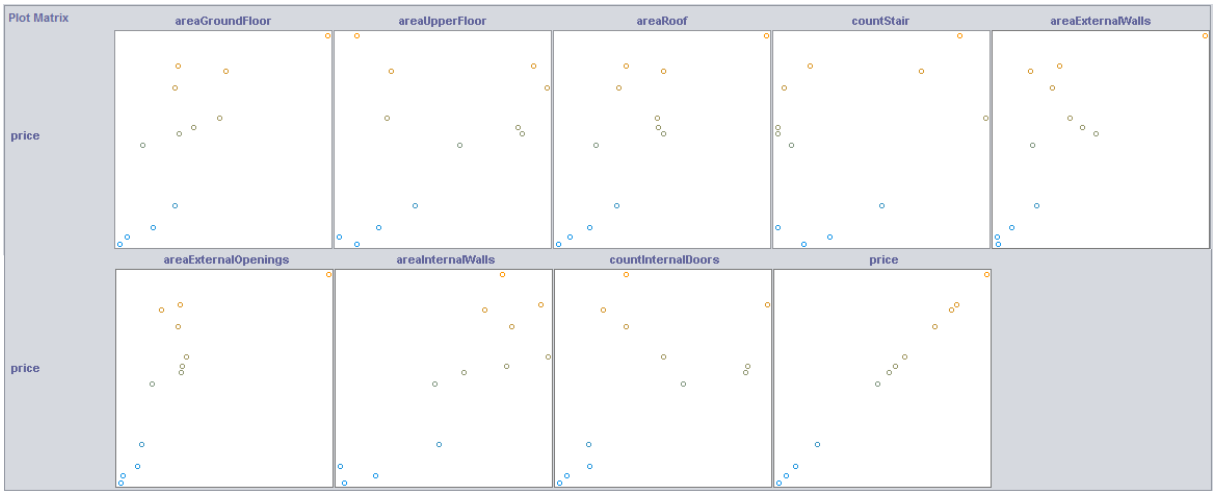

Figure 9: Correlation plot

5.3.2. Normalization

The classification algorithm k-Nearest Neighbour uses the Euclidean distance between the variables to calculate the closest neighbour. The Euclidean distance (*d)* is calculated with equation 5. It calculates the straight-line difference between two points. Therefore it is necessary to scale all variables to a similar scale. Otherwise the variables with high values would have a higher influence on the dataset as opposed to variables with low values.

$$Euclidean\ distance: \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \qquad (5)$$

The variables can be normalized using equation 6. Via this equation all variables are transformed into variables between zero and one. Via this equation the range within a variable is restricted between a zero and a one. Therefore the relative distance between a variable with a small range of values has the same influence as the relative distance between variables with a high range of values.

$$Normalization = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (6)$$

5.4. Conclusions data engineering

In this chapter the data engineering part of the data mining process is discussed. Much attention has been paid towards the extraction of the information from BIM models into a dataset. It has become clear that there is a difference between the theoretical potential of the BIM models and the actual potential of the BIM models in the current state of the AEC industry. As discussed in the literature review, theoretically it should be possible to extract all physical data of the buildings from the BIM models. However the current quality of the BIM models has presented problems.

First of all, too often the models are not a real representation of the actual building. For example in the collected BIM models part of the building were found missing. Another problem with the quality of the BIM models is the mislabelling of the building element type. Elements were modelled as walls while they were for example part of the window frame. This presents a problem in automatic QTO as the extractor script interprets the object as the wrong type and thus extracts an incorrect QTO.

As discussed in the literature both how and which information that is shared with the IFC model can be standardised with an IDM and an IFD. Using this standardisation would solve most of the problems that occurred with the extraction of the QTO.

In the literature review it was made clear what the theoretical strength of the proposed concept is. In this chapter however it was shown what the weaknesses of the concept are. The whole system depends on the quality of the BIM models and it has been shown that at this point the models available in real life situation do not meet the level of quality needed for the system to work automatically.

This page is intentionally left blank

6. Data analytics

In this chapter, it is explained how the different machine learning algorithms are applied and what the results are. Additionally the conclusions of the practical application are discussed. It should be noted that the dataset is very small and therefore not really useful for data mining. However in this chapter it is explored if some useful relationships can be found.

6.1. Algorithm engineering

In the algorithm engineering step the appropriate machine learning algorithm is chosen or if necessary created. For this research two different machine learning algorithms are chosen, as discussed in the literature review. These two algorithms are multiple linear regression and k-Nearest Neighbour. These algorithms were chosen to test the performance of both a regression algorithm as a classification algorithm, which are the two types of machine learning algorithms.

The software package R will be used for the data mining analytics. R has a large library of algorithms that can be used. Additionally a graphical user interface (GUI) for data mining called Rattle has been created by Togaware. This GUI provides an interface for R which has the advantage that it creates the code necessary to run the algorithm in R and thus prevents unnecessary coding. For this reason the Rattle package is used. Unfortunately not all algorithms that are available in software package R are available within the Rattle GUI. This is also the case for k-Nearest Neighbour. Consequently the code for running the algorithm k-Nearest Neighbour has been created during the presented research. This code can be found in appendix 3.

6.2. Running the algorithm

As discussed in the literature review the performance of the algorithm can be tested with statistical evaluation measures. To evaluate the performance, the correlation coefficient (CC), the root mean squared error (RMSE), the mean absolute error (MAE) and the root relative squared error (RRSE) are used.

For the linear regression algorithm only the input variables need to be determined before the algorithm can be run. In the previous chapter, feature selection was used to determine a ranking for the relevance of the variables. To determine the optimal amount of attributes used, stepwise regression is applied.

To test the performance of the model, the dataset is randomly split into a training set and a test set. The split percentage has been set on 75%. Three different test- and training sets have been applied as can be seen in table 5. The table shows the result of the algorithm is very dependent on the randomly selected training and test set. Therefore it can be concluded that the amount of data available is too little to have any significant meaning.

Table 5: Performance of the linear regression algorithm

Amount of variables	Measures	Result 1	Result 2	Result 3
AreaInternalWalls	CC	0.723	0.9838	-0.9976
	RMSE	560617.4449	505518.9929	627267.4708
	MAE	481468.0006	466434.9967	508912.5596
	RRSE	48.0045 %	36.2489 %	64.5285 %
AreaInternalWalls	CC	*0.7870*	0.9902	-0.9909
AreaExternalOpenings	RMSE	*433203.4989*	555776.8736	621719.9478
	MAE	*402421.8829*	435928.7036	570460.5939
	RRSE	*37.0943 %*	39.8527 %	63.9578 %
AreaInternalWalls	CC	0.6855	0.9674	-0.8468
AreaExternalOpenings	RMSE	459702.867	959054.6101	791664.2357
AreaRoof	MAE	428591.8218	825663.6968	614111.1752
	RRSE	39.3634 %	68.7702 %	81.4404 %
AreaInternalWalls	CC	0.4980	0.9142	-0.8131
AreaExternalOpenings	RMSE	544428.3547	355500.1695	785137.176
AreaRoof	MAE	513398.0693	401117.7355	1103176.0896
CountInternalDoors	RRSE	46.6183 %	40.1181 %	113.4863 %

For the k-Nearest neighbour algorithms besides the amount of variables also the amount of nearest neighbours used to predict the value are important. As the k is the amount of samples used to determine the costs based on their average, the result of the algorithm is highly influenced by the number of k's used. In this research the following k's are evaluated, k: 1, k: 3 and k: 5. For all three k's the performance was evaluated on the dataset. In table 6 the performance of the algorithm with k: 3 and only areaExternalOpenings as variable is shown.

Although the difference in outcome between the three results is relatively the same for RMSE MAE and RRSE, the difference in the correlation coefficient indicates that also the k-nearest neighbour is not useful on the data. The result of the first and third result indicate that the area of external openings is positively correlated with the costs of the building, where the second result indicates that the correlation is negative.

Table 6: Performance of the k-nearest neighbour algorithm

Measure	Result 1	Result 2	Result 3
Correlation coefficient	0.9425	-0.7737	0.9994
Root mean squared error	521440.4598	401838.3491	626107.6
Mean absolute error	460706.2222	347361.6667	473364.6667
Root relative squared error	44.6499 %	41.3381 %	44.8958 %

6.3. Evaluation of result

Based on the result in the previous chapter it can be concluded that the dataset is indeed not suitable to apply data mining. The main reason is that the dataset is too small. If a dataset is small the dataset has low statistical power and the research is therefore unlikely to result in reprehensive conclusions (Button, Ioannidis, Mokrysz, Nosek, Flint, Robinson & Munafò, 2013). As only 12 projects were used in the evaluation, this is certainly the case. However in the current situation not many firms in the AEC industry will have many more BIM models of one type of buildings. So even though Hoogeveen (2015) proved that the concept works theoretically, for the concept to work within real-world situations, more BIM models need to be generated. And as already concluded, these BIM models need to have a higher level of quality as the current standard. Another problem with the dataset is that because of the small amount of projects no difference was made between apartment buildings and houses. Although they both have the same function their designs differentiate.

6.4. Redefining data

In the last part it was concluded that the dataset was not suitable for data mining. Therefore additional data about residential buildings have been collected and added to the dataset. These data were found using three sources, namely:

- Bouwkosten: nieuwbouw en renovatie projectanalyses (Kuhlmann, 2007)
- Bouwkosten.nl (2017)
- Casadata.nl (2017)

On the next page the collected data are given. As not all of the data contained information about the internal walls or the internal doors, these variables have been ignored. In appendix 4 the used sources can be found.

To compare the prices of the projects it is important to normalize the prices to the same reference point in time. As some of the collected data were about projects in 2002, this step was important to be able to compare the prices accurately. The used formula to normalize the price is:

$$Price\ normalization = Price_{original} * \frac{PriceIndex_{2017}}{PriceIndex_{original\ date}} \quad (7)$$

In appendix 5 the price index from the last 15 years can be found.

The old dataset was not suitable for data mining because it had had two problems, it was small and it contained two types of residential buildings, namely apartments and houses. Therefore it was chosen to create two new datasets. One dataset with all available data of both the old and new data. The other dataset contains data from both the old and new data, but only from projects that were apartments. It was chosen to create a dataset for apartments and not houses as there was more data available about apartments than about houses. In the remainder of this research the dataset with all the data is called dataset A and the dataset with only apartments is called dataset B. In the next part both datasets will be used in the algorithm to evaluate which of the two datasets perform best.

Table 7: Additional data from databases

Project name	Source	Price index	Area Ground Floor	Area Upper Floor	Area Roof	Count Stair	Area External Walls	Area External Openings	Original price	Normalized price
Nieuwbouw 14 rijenwoningen Groenoord	casadata.nl	1-3-2017	645	2085	645	42	1628	652	€ 2.178.000,00	€ 2.178.000,00
Nieuwbouw 22 rijenwoningen Dronten	casadata.nl	1-3-2017	1177	1294	1177	44	1291	317	€ 2.005.310,00	€ 2.005.310,00
Nieuwbouw 31 rijenwoningen	casadata.nl	1-3-2017	1995	1762	1995	31	2632	770	€ 3.892.552,00	€ 3.892.552,00
28 appartementen De Haere	Bouwkosten.nl	1-3-2017	1165	2137	1322	4	2092	606	€ 3.380.015,00	€ 3.380.015,00
Appartementencomplex de barchaan	Bouwkosten.nl	1-3-2017	599	1050	656	3	1528	574	€ 2.641.105,00	€ 2.641.105,00
44 apartementen de parkwachters	Bouwkosten.nl	1-3-2017	1045	4976	1368	19	4358	1561	€ 5.665.615,00	€ 5.665.615,00
29 appartementen boven winkels	Bouwkosten.nl	1-3-2017	990	2521	990	12	1981	565	€ 4.623.110,00	€ 4.623.110,00
Woontoren met 40 appartementen	Bouwkosten.nl	1-3-2017	673	5007	713	8	3603	1604	€ 4.924.405,00	€ 4.924.405,00
Allardhof te buren	Bouwkosten.nl	1-3-2017	1220	1313	1256	6	3653	565	€ 2.414.120,00	€ 2.414.120,00
17 appartementen Oudenbosch	Kuhlmann (2007)	1-4-2002	487	976	554	4	1389	380	€ 1.359.660,00	€ 1.622.132,71
49 seniorenappartementen Dronten	Kuhlmann (2007)	1-4-2002	1463	4785	1231	3	3425	1222	€ 3.677.005,00	€ 4.386.824,70
15 stadswoningen Maastricht	Kuhlmann (2007)	1-10-2002	981	1412	1393	31	2096	646	€ 2.039.045,00	€ 2.396.184,79
Appartementen Winterswijk	Kuhlmann (2007)	1-5-2002	2451	3210	2451	11	2554	616	€ 3.562.415,00	€ 4.228.964,40

6.5. Using the new datasets

For the new datasets the same methods have been applied as were for the old dataset. The input variables were determined with a ranking method. And again the dataset has been randomly split into a training set and a test set, where 75% was put into the training set and 25% into the test set. Also the same methods were used to test the performance of the algorithm, which are:

- the correlation coefficient (CC),
- the root mean squared error (RMSE)
- the mean absolute error (MAE)
- root relative squared error (RRSE)

The performance of the data mining process on the new datasets can be found in table 8 and 9. In table 8 the performance is measured on the dataset A. In comparison to the old dataset this dataset has 13 more building projects in the dataset, which is an increase of 108%. Table 9 shows the performance of the linear algorithm on the dataset B. This dataset has almost the same number of projects, 14 in the new dataset and 12 in the old dataset.

Both tables show an improvement compared to the performance of the old dataset. Especially in table 8 where the consistency in performance, based on different training- and test sets, has greatly increased in all performance measured.

Table 8: Performance of the linear algorithm on the dataset A

Amount of variables	Measures	Result 1	Result 2	Result 3
AreaUpperFloor	CC	0.6671	0.9467	0.5195
	RMSE	1131852.4198	666785.1939	1122752.2352
	MAE	1003933.5539	571196.4198	945815.3132
	RRSE	76.2631 %	35.8792 %	76.5787 %
AreaUpperFloor	CC	0.7798	0.9081	0.7441
AreaExternalOpenings	RMSE	859523.2126	742378.2462	875311.6464
	MAE	764947.4453	641098.7321	718483.882
	RRSE	57.9138 %	39.9468 %	59.7017 %
AreaUpperFloor	CC	0.7808	0.9064	0.7166
AreaExternalOpenings	RMSE	857395.0371	750482.6377	900271.0147
AreaRoof	MAE	763929.6517	637432.6853	764562.7303
	RRSE	57.7704 %	40.3829 %	61.4041 %
AreaUpperFloor	CC	0.9398	0.9854	0.9441
AreaExternalOpenings	RMSE	515213.5166	644344.289	450966.3423
AreaRoof	MAE	422329.6066	571994.0504	350937.6706
AreaGroundFloor	RRSE	34.7146 %	34.6717 %	30.7587 %
AreaUpperFloor	CC	0.9409	0.924	0.9105
AreaExternalOpenings	RMSE	626445.8633	1070879.5792	532660.4909
AreaRoof	MAE	539378.4266	788590.4438	452562.9005
AreaGroundFloor	RRSE	42.2093 %	57.6233 %	40.9216 %
AreaExternalWalls				

Table 9: Performance of the linear algorithm on dataset B

Amount of variables	Measures	Result 1	Result 2	Result 3
AreaUpperFloor	CC	0.832	0.9986	-0.4039
	RMSE	426982.2271	464236.7724	892102.2197
	MAE	354557.2523	453382.0412	801657.1538
	RRSE	31.9167 %	46.736 %	70.2887 %
AreaUpperFloor	CC	0.7304	0.9996	-0.3797
AreaExternalOpenings	RMSE	475461.7763	458896.0863	925018.2828
	MAE	411819.3968	418881.0927	843154.0895
	RRSE	35.5405 %	46.1984 %	72.8822 %
AreaUpperFloor	CC	0.8158	0.9324	0.7974
AreaExternalOpenings	RMSE	793180.8622	624127.2974	804384.9282
CountStairs	MAE	721734.6241	491184.1518	678728.1529
	RRSE	53.4438 %	33.5838 %	54.8641 %
AreaUpperFloor	CC	0.7015	0.9967	-0.4164
AreaExternalOpenings	RMSE	512596.1218	758006.4997	619014.072
CountStairs	MAE	451168.1877	700553.4729	493597.3885
AeraExternalWalls	RRSE	38.3162 %	76.3107 %	48.7721 %
AreaUpperFloor	CC	0.8081	0.979	-0.9428
AreaExternalOpenings	RMSE	523895.9156	227580.5032	835725.0506
CountStairs	MAE	418128.2352	220203.8128	827008.3429
AreaExternalWalls	RRSE	39.1609 %	22.9112 %	65.8468 %
AreaGroundFloor				

Based on the performance shown in the tables 8 and 9 the best performing linear algorithm was chosen. For dataset A, the best performing algorithm is the one with four independent variables. This is because the performance indicator is generally the best when four variables are used. This has been made more clear in a graphical representation, which can be found in appendix 6. This results in the following formula:

$$Predicted\ price = 244587{,}58 + 598.53 * areaUpperFloor + 691{,}24 * areaExternalOpening + 562{,}00 * areaRoof + 637{,}53 * areaGroundFloor \quad (8)$$

For dataset B, the performance is best when three variables are used. Although only 2 of the four indicators used perform best with three algorithms, as the RMSE and MAE perform better with four variables, the consistency in performance between the different training- and test sets makes three variables the best option. See also appendix 7 for the graphical representation. This results in the following formula:

$$Predicted\ price = 1306976.57 + 763{,}49 * areaUpperFloor + 94219{,}90 * countStair - 629{,}50 * areaExternalOpenings \quad (9)$$

The formula 8 and 9 have been used to predict the prices of the projects within the dataset (see table 10 and 11). The predicted prices are compared to the actual prices to determine the accuracy of the algorithm. According to Phaobunjong (2002) the expected accuracy for a definitive estimate of a building project is within a range of +15% to – 5% and the expected range of a conceptual estimate should be within +50% to -30%.

The margin of error of the predicted prices in table 10 and 11 are too high for a definitive cost estimate. However for a conceptual estimate the algorithms perform acceptable. Only four of the predicted prices are not within the acceptable range for a conceptual estimate. All of these are predicted based on dataset A and are houses.

Where the performance indicators of dataset A have shown that the consistency increases as the dataset becomes larger, the predicted prices of dataset B shows the increased accuracy when only the same type of residential buildings are used. Calculated based on the absolute error in percentage of the predicted price, the average error is 12.4%. As the dataset A has an average error of 25.6%, the dataset B is clearly more accurate.

Table 10: Predicted prices of dataset A

Name	Actual price	Predicted price	Error in %
Nieuwbouw 14 rijenwoning Groenoord	€ 2.178.000,00	€ 2.716.914,40	25%
Nieuwbouw 22 rijenwoning Dronten	€ 2.005.310,00	€ 2.650.061,01	32%
Nieuwbouw 31 rijenwoning	€ 3.892.552,00	€ 4.224.521,49	9%
28 appartementen De Haere	€ 3.380.015,00	€ 3.428.232,40	1%
Appartementencomplex de barchaan	€ 2.641.105,00	€ 2.020.371,10	-24%
44 apartementen de parkwachters	€ 5.665.615,00	€ 5.736.949,05	1%
29 appartementen boven winkels	€ 4.623.110,00	€ 3.331.576,18	-28%
Woontoren met 40 appartementen	€ 4.924.405,00	€ 5.179.953,88	5%
Allardhof te buren	€ 2.414.120,00	€ 2.904.671,53	20%
17 appartementen Oudenbosch	€ 1.622.132,71	€ 1.713.252,40	6%
49 seniorenappartementen Dronten	€ 4.386.824,70	€ 5.577.792,45	27%
15 stadswoningen Maastricht	€ 2.396.184,79	€ 2.944.541,78	23%
Appartementen Winterswijk	€ 4.228.964,40	€ 5.531.734,93	31%
A.ifc	€ 2.267.617,99	€ 2.146.674,34	-5%
B.ifc	€ 3.712.303,91	€ 2.703.719,54	-27%
C.ifc	€ 2.511.492,54	€ 2.771.172,88	10%
D.ifc	€ 2.620.614,11	€ 2.799.981,48	7%
E.ifc	€ 3.315.522,11	€ 2.736.052,49	-17%
F.ifc	€ 2.929.378,37	€ 2.062.372,84	-30%
G.ifc	€ 989.262,70	€ 1.721.203,55	74%
H.ifc	€ 540.782,18	€ 817.820,01	51%
I.ifc	€ 441.135,74	€ 839.490,85	90%
J.ifc	€ 1.888.289,88	€ 1.858.531,98	-2%
K.ifc	€ 3.560.412,73	€ 3.651.175,18	3%
L.ifc	€ 681.174,41	€ 1.302.045,02	91%

Table 11: Predicted prices of dataset B

Name	Actual price	Predicted price	Error in %
28 appartementen De Haere	€ 3.380.015,00	€ 2.933.956,56	-13%
Appartementencomplex de barchaan	€ 2.641.105,00	€ 2.029.967,85	-23%

44 apartementen de parkwachters	€ 5.665.615,00	€ 5.913.629,92	4%
29 appartementen boven winkels	€ 4.623.110,00	€ 4.006.705,04	-13%
Woontoren met 40 appartementen	€ 4.924.405,00	€ 4.873.810,76	-1%
Allardhof te buren	€ 2.414.120,00	€ 2.519.090,69	4%
17 appartementen Oudenbosch	€ 1.622.132,71	€ 2.189.812,24	35%
49 seniorenappartementen Dronten	€ 4.386.824,70	€ 4.473.685,05	2%
Appartementen Winterswijk	€ 4.228.964,40	€ 4.406.424,78	4%
B.ifc	€ 3.712.303,91	€ 3.338.541,54	-10%
C.ifc	€ 2.511.492,54	€ 2.774.966,32	10%
D.ifc	€ 2.620.614,11	€ 2.736.779,24	4%
E.ifc	€ 3.315.522,11	€ 3.076.425,48	-7%
J.ifc	€ 1.888.289,88	€ 2.660.717,79	41%

6.6. Conclusion data analytics

In the previous chapter already appeared that the quality of the BIM models is an issue for automated quantity take off with the purpose of data mining. In this chapter a new issue was found. After applying the data mining algorithms on the dataset, no usable relations between the data could be found.

The proposed reason for the issue was that the dataset was too small and that there was too much variation in the design of the buildings as both apartments and houses were used in the dataset. To test this assumption, addition data was collected from project analyses available in databases. These additional data have been used to create two new datasets. One dataset, where all the data of both the original data and the new data were combined and another dataset with only one type of buildings in the dataset, namely apartments.

These new datasets were tested by applying the previously used linear regression algorithm. From the results of the algorithm it could be concluded that the made assumptions were correct. The dataset with all the data showed that more data give a more consistent result, where the dataset with only apartments shows that the accuracy of the prediction are acceptable for conceptual cost estimation if only one type of buildings is used in the dataset.

Part I: Problem statement

Part II: Theoretical review

Part III: Practical implementation

Part IV: Conclusions and recommendations

This page is intentionally left blank

7. Discussion

The main objective of this research was to evaluate the effectiveness of a knowledge management system using BIM data and the data mining process. To realise this goal, two different approaches have been used, a literature study and a practical implementation. At first the literature study has been conducted. This study has shown the theoretical potential of such system. However during the practical implementation approach, it was found that the implementation of such system encountered some problems. To solve this problem additional data was collected, which resulted in a useful model for cost estimation.

Part of the scientific relevance of this research is that it has shown an alternative way to use building information modelling. In the recent years more attention has been paid towards using BIM as the knowledge source in knowledge management, as shown in the literature review. However little of these researches has resulted in practical implementation. Also this research has shown some of the implementation problems with the current state of BIM models. It has been made clear that there is a gap between the scientific potential of BIM models and the potential for implementation at this point. More attention is necessary toward improving the quality and correctness of data in BIM models if the AEC industry want to use the full potential of BIM.

The practical relevance of this research is that it has shown what improvements are needed for BIM to reach its potential for the application of data mining process. As will be discussed in the next chapters, it is necessary to focus on improving the way BIM models are shared and how they are modelled. Over the whole building process the BIM model should be kept up to date and shared in predefined format.

This page is intentionally left blank

8. Conclusions

In order to answer the research problem, five research questions were defined. In this chapter the answers to these research questions are given. After answering these research questions the research problem will be answered.

Research question 1: Which are the steps the data mining process consists of?

The data mining process can be divided in two parts and has a total of nine steps. The first part is the data engineering part. It starts by understanding and defining the problem. In this research the problem of the data mining process is to create knowledge to predict the building costs of residential buildings in the early phase of the design process. After the problem is made clear, the data which will be used in the data mining process need to be collected and extracted into a useable dataset. For this research BIM models with corresponding costs data were collected and the data of the BIM models were extracted to a dataset in form of a quantity take off. The next step is to clean and explore the data. As the quality of the data in this research appeared to be quite low, a lot of cleaning had to be done. Following the data exploration the dataset had to be converted into a subset which will perform best with the used algorithms. As the dataset available in this research was very small, it was chosen to test the performance of multiple subsets. The subsets were created based on a ranking method for feature selection. After the data have been engineered the data mining process continues to the next part, data analysing.

This part starts with the engineering of the algorithm to be used. Because the goal of this research is to create new knowledge for cost estimation a machine learning algorithm is needed. It was chosen to use both a regression algorithm as a classification algorithm. This was chosen because both have been successfully applied in the past to predict cost estimation for building costs. This process is continued with the running of the algorithm. In case of this research both the regression as the classification algorithm were run. The next step is defined as refining the data and problem. As the evaluation of the performance of the data mining algorithms was that no reliable relations were found within the dataset and it was concluded that the most probable cause of this was that the dataset was not large enough and the used building project were to different, additional data has been collected. This data has been used to create two new datasets that were applied in the data mining algorithm. One dataset which exists of all data both from the old dataset as from the newly gathered data, showed that adding more data to the dataset results in more consistent results. The other dataset also exist of both the old dataset and the newly acquired data but it only had the projects that were apartments to have more homogenous content in the dataset. The results of the application of the data mining algorithm on this dataset showed that useful relations can be found if the dataset contains data of building from a similar type. The last step of the data mining process is to use the results. To test the usefulness of the results, the results were used to estimate the costs of the buildings within the dataset. Although the results were close enough to use the generated cost estimation model to calculate a conceptual estimate for building project, the cost estimation model was only tested on the dataset that was used to create the model, so it should be noted that it is yet unclear how precise the model will be on unseen data.

Research question 2: Is the IFC data format a suitable BIM data format to extract a dataset for the data mining process?

In the literature review it was shown how Ma et al. (2013) used the IFC model to extract a quantity take off in the format of the Chinese mandatory specification of cost estimation in the tender phase. This suggested that the IFC data format was indeed a suitable BIM data format to extract a dataset for the data mining process. This was confirmed during the practical implementation. The part of the BIM models that were modelled correctly could be extracted from the model automatically using an extraction script. Due to limitations in the script the information embedded within the BIM model about the area of the different spaces within the building could not be extracted. However this is not a result of the IFC data format but a result of the applied script.

Research question 3: Is the current quality of the BIM models good enough to automatically extract a dataset for the data mining process?

A lot of effort has been made to create an extraction script that can automatically extract the dataset from BIM models. However to get the extractor script to extract an acceptable dataset a lot of code had to be written especially for one of the IFC models. In one of the IFC models three different IfcBuildingStorey elements were created for every floor level. This makes the automatic interpretation of floor levels not possible without writing a code for the script that can interpret multiple IfcBuildingStorey elements for one floor level. Because of the similarity between the height of the different IfcBuildingStorey elements, it was not possible to analyse and merge the levels by height. As a solution, some code has been written that allows the script to interpret the IfcBuildingStoreys for the IFC model that had this problem, but for other IFC models with a similar problem the script will not be able to interpret this correctly without changing the code.

As already concluded in the previous part of the research, another problem with the IFC quality is the export of the BIM authoring tool model into the IFC model. All but one of the BIM models were modelled with Revit, the other model with Allplan. Of all those models four contained the necessary information about the organisation that exported the BIM file. All of them were the contractor of the building projects. Another five of them were exported by the author of the presented research.

One of the mentioned problem with the IFC models was that three different IfcElements were found to represent a stair, namely: IfcStair, IfcFurnishingElement, IfcBuildingElementProxy. This is a problem because as the name IfcFurnishingElement suggests, this type of IfcElement is a generalization of all furniture related objects (buildingSMART, nd). Another problem found was that more than one level of height per building storey was used in the BIM model. This is not necessarily a problem if only one of the levels of height was modelled as the main level and the other levels of the storey as a sub level. However this was not the case and therefore the BIM model was exported to the IFC data format with too many building storeys. All these problems could be solved if all the organisations within the building project that model within the BIM model would work according to the BIM Basis ILS.

The loss of necessary information is also a problem. In good practise some building elements like walls are modelled as one building element composed of different construction products (see figure 11). As this is a relatively easy way of modelling and it even saves time compared to modelling all the different construction parts separately, it is plausible that the building

elements were modelled as one element although they are composed of different construction products. In the IFC models however, all these different construction products were already decomposed to standard IfcElements. This makes the interpretation of these elements much harder. In most exporters provided by the BIM authoring tools, this problem can be solved by changing the setting of the exporter.

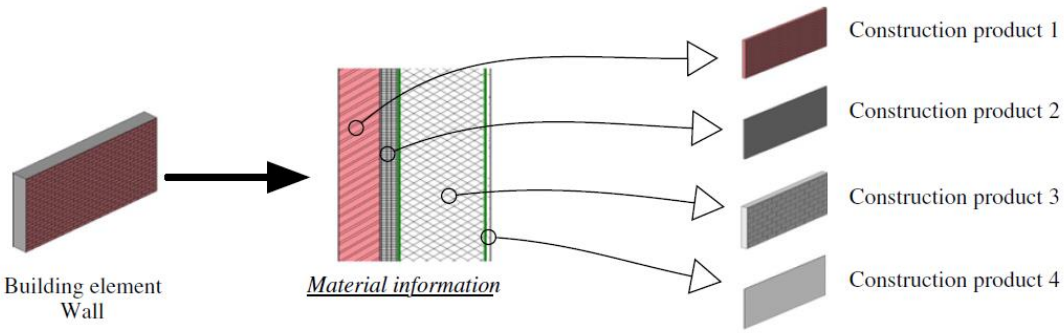

Figure 10: Building elements composing of different construction products (source: Ma et al. (2013)

Research question 4: Can the data mining process be used to predict the costs of residential buildings in the early stages of the design process?

The research question: *Can the data mining process be used to predict the costs of residential buildings in the early stages of the design process?* has been answered both via literature review as via practical implementation. However the answers differ. Based on the literature it can be concluded that it is indeed possible to use the data mining process to predict the costs of residential buildings in the early stage of the design process. The BIM model can contain enough information about the buildings. With an extractor this knowledge embedded in the model can be made explicit and with a machine learning algorithm this knowledge can be used to create new information that allows us to predict the costs of buildings.

In the practical implementation part of this research however it became clear that there are two implementation problems. First of all the quality of the BIM models does not meet the same level as is necessary for automatic extraction of the required dataset, as discussed in the answering of the previous research question. Another problem is the amount of BIM models available. The data mining process needs data to analyse and generally the more data it can analyse the better the data mining process is able to find patterns within the data. It has been shown that 12 BIM models do not provide enough data to find these patterns. So, even when the AEC industry would create BIM models of the quality necessary for data mining, also a lot of models are needed of a specific building type with the above mentioned quality.

In this research it was chosen to add additional data from public available project analyses to create a larger dataset. This data proved to be very useful and it resulted in a cost estimation model that was precisely enough to calculate the conceptual building costs of the buildings within the dataset.

Main research question: “Can the use of a knowledge management system using BIM data and the data mining process create a more efficient process of cost estimation for residential buildings?”

The literature review has shown that the data mining process is an effective tool to create a knowledge management system. It is especially useful because it can be made into an automated computer process, where little maintenance is required. The literature review has also shown that BIM data have a great potential to serve as the data source. These are created in almost every project and they contain all the physical information of the buildings. These data can be converted to the dataset necessary for the data mining process by creating a quantity take off. In previous researches could be found that automated quantity take off from BIM models is possible. From this all it can be concluded that a data mining process can efficiently estimate the costs of residential buildings and therefore used as a knowledge management system for cost estimation.

However during the practical implementation it became clear that there are both qualitative and quantitative problems with the available BIM models. As mentioned above additional data were added to test viability of the data mining process for cost calculation for residential buildings. With this additional data the data mining process could predict the costs of buildings within the dataset precisely enough to be a conceptual cost estimate. Nonetheless until the qualitative and quantitative problems of BIM data are solved the knowledge management system will still need the manual input of additional data to be able to estimate the costs of residential buildings.

9. Recommendations

In this chapter the recommendations are given. First the recommendations for implementation are discussed. Thereafter the recommendations for further research are given.

9.1. Recommendations for implementation

9.1.1. Implementation of the IFD format

It has been concluded that the biggest obstacle to overcome before the proposed knowledge management system can be implemented is the quality of the BIM models.

Part of the problem is that for the IFC format in order to be an effective format to export data to, it allows various data to be shared in various ways. If the receiver of the information, in this research the extractor, wants to be sure to be able to utilize the information he receives, the sender and receiver need to be in agreement which and how the information is shared (Bell & Bjørkhaug, 2006). For this purpose buildingSMART already created the International Framework of Dictionaries (IFD). The Dutch implementation of this is the Nederlandse Conceptenbibliotheek voor de bouw (CB-NL). This can be used to standardise the way information within IFC is shared, and thus allows to use the proposed agreements to create IFC models that allow for automated data extraction. So part of the solution for this problem already exists, however the solution was not used in the development of the BIM models that were available in this research. Consequentially it is recommended that the IFD is implemented within all organisations within a building project. To facilitate this, the government should start demanding that IFD is used like the Dutch government has done with IFC in their DBFMO contracts.

9.1.2. Implementation of the Information Delivery Manual(IDM)

Another way to ensure better quality of BIM models and more specifically IFC models is to use Information Delivery Manuals (IDM). IDM can be used for the documentation of existing or new processes and describes the associated information that needs to be exchanged between parties. The main purpose of IDM is to make sure the relevant data are communicated via a structured way such as the data can be interpreted by the software on the receiving side (buildingsmart, nd.). For the Dutch AEC industry the 'BIM basis informatieleveringsspecificatie (ILS)' is defined and it has already been adopted by many companies. As has been shown during the practical application most of the problems of the quality of the IFC models would have been prevented if the BIM Basis ILS would have been used.

9.1.3. BIM as-built models

Even if the IFD format or another ontology and the IDM would systematically be applied not all quality problems of the BIM models would be solved. The IFD format would not prevent that the BIM models are not modelled as built. In order to accomplish this the BIM model should be updated consequently during the whole building process. This can only be ensured if all companies involved would have an incentive to update the BIM models. As the information that can be extracted from the BIM models to a dataset is limited to the information located within the BIM models, it is very important for the application of data mining on BIM models that all the information is modelled.

9.1.4. Use of additional data

Next to the quality of the BIM models also the number of BIM models available is a problem for the proposed knowledge management system. To overcome this problem in this research additional data has been added to the dataset. If one wants to implement the proposed knowledge management system within the organisation the dataset can be manually enlarged with data from projects of which no viable BIM model is available. Although it is desirable to have little to no human effort needed to maintain the KMS it allows the KMS to be implemented. With the current trends of BIM modelling a lot more BIM data will come available in the upcoming years, and thus the BIM data can eventually replace the use of additional data.

9.2. Recommendation for further research

9.2.1. Quality of BIM models

In the recommendations for implementation it is suggested that the quality of the BIM models can be improved by implementing an ontology. However the focus of this research was aimed at the implementation of BIM models in a knowledge management system. More research is needed to provide more ways to improve the quality of BIM models.

9.2.2. Other goal of KMS

This research has focussed on the implementation of a KMS for cost prediction. However other goals for a KMS system based on BIM data and the data mining process are possible. For example a goal for a KMS system could be to predict the energy consumption of buildings.

References

- Aalst, W. M. P. van der, Stoffele, M., & Wamelink, J. W. F. (2003). Case handling in construction. *Automation in Construction*, *12*(3), 303-320.
- Abeljaber, R., Ioannou, C., Maor, D., Razo, R., & Tribolet, J. (1998) *Knowledge management of corporate intranets.* Retrieved on 17 october 2016 from: http://web.mit.edu/ecom/www/Project98/G4/
- Alavi, M., & Leidner, D. E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS quarterly*, *25*(1), 107-136.
- Al-Gahtani, K., & Ghani, S. R. (2010, March). Knowledge management tools: adaption in construction management practices. In *Engineering Systems Management and Its Applications (ICESMA), 2010 Second International Conference on* (pp. 1-6). IEEE.
- Almana, A. M., & Aksoy, M. S. (2014). An Overview of Inductive Learning Algorithms. *International Journal of Computer Applications*, 88(4), 975–8887.
- Argote, L. & Ingram P. (2000). Knowledge Transfer: A Basis for Competitive Advantage in Firms. *Organizational Behavior and Human Decision Processes*, 82(1), 150–169.
- Barney, B. (2016). *Introduction to Parallel Computing.* Retrieved on 17 march 2017 from: https://computing.llnl.gov/tutorials/parallel_comp/
- Bell, H., & Bjørkhaug, L. (2006, August). A buildingSMART ontology. In *Proceedings of the 2006 ECPPM Conference* (pp. 185-190). sn.
- Bhargav, D., Koskela, L., & Dave, B. (2009). Collaborative Knowledge Management – A Construction Case Study. *Automation in Construction*, 18(7), 894–902. Retrieved from http://eprints.hud.ac.uk/25837/
- Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. Owolabi, H., Alaka, H and Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, 30 (2016). pp. 500-521.
- Böhms, M., Mommers, B., Baggen, M., Bakker, J., Manoku, E., Nierop, M. van, Nagtegaal, R., Beetz, J., Brink, L. van den & Oldfield, J. (2013). *Nederlandse Conceptenbibliotheek voor de bouw (CB-NL) – Modeling Guide.* Amersfoort: Geonovum
- Botter, R., (2013) *Efficiënt begroten met historische projectgegevens.* Eindhoven: TU Eindhoven. Fac. Bouwkunde
- Brown, A., & Adams, J. (2000). Measuring the effect of project management on construction outputs: a new approach. *International Journal of Project Management*, 18, 327-335.
- BuildingSMART (n.d.). *Ifc Overview*. Retrieved 2016, December 12, from http://www.buildingsmart-tech.org/specifications/ifc-overview
- BuildingSMART (n.d.). *IfcObject*. Retrieved 2016, December 12, from http://www.buildingsmart-tech.org/ifc/IFC2x4/rc2/html/schema/ifckernel/lexical/ifcobject.htm
- BuildingSMART (n.d.). *IfcProduct.* Retrieved 2016, December 12, from http://www.buildingsmart-tech.org/ifc/IFC2x3/TC1/html/ifckernel/lexical/ifcproduct.htm

- BuildingSMART (n.d.). *IfceElement.* Retrieved 2016, December 12, from http://www.buildingsmart-tech.org/ifc/IFC2x3/TC1/html/ifcproductextension/lexical/ifcelement.htm
- BuildingSMART (n.d.). *IfcBuildingElement*. Retrieved 2016, December 12, from http://www.buildingsmart-tech.org/ifc/IFC2x3/TC1/html/ifcproductextension/lexical/ifcbuildingelement.htm
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). *Power failure: why small sample size undermines the reliability of neuroscience*. Nature Reviews Neuroscience, 14(5), 365-376.
- Calkins, D. E., Egging, N., Scholz, C., & Gomez-Levi, G. (2000). *AutoDSS: A System Level KBE Tool for Vehicle Product Definition (No. 2000-01-1351)*. SAE Technical Paper.
- Campbell, D.A. (2006, 11 October) Modeling Rules, Architecture week. Retrieved on 17 march 2017 from: http://www.architectureweek.com/2006/1011/tools_1-1.html
- Charlesraj, V. P. C. (2014, January). Knowledge-based building information modeling (K-BIM) for facilities management. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction* (Vol. 31, p. 1). Vilnius Gediminas Technical University, Department of Construction Economics & Property.
- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, *19*(2), 171-209.
- Cheung, F. K., Rihan, J., Tah, J., Duce, D., & Kurul, E. (2012). Early stage multi-level cost estimation for schematic BIM models. *Automation in Construction*, *27*, 67-77.
- Clayton, M. J., Ellis, C. D., Woo, J., Clayton, M. J., Johnson, R. E., Flores, B. E., & Ellis, C. (2004). Dynamic Knowledge Map : Reusing experts ' tacit knowledge in the AEC industry. *Automation in Construction 13(2), 203-207.*
- Coakes, E. W., Coakes, J. M., & Rosenberg, D. (2008). Co-operative work practices and knowledge sharing issues: A comparison of viewpoints. *International Journal of Information Management*, *28*(1), 12-25.
- Das, A., & Kempe, D. (2008, May). Algorithms for subset selection in linear regression. In *Proceedings of the fortieth annual ACM symposium on Theory of computing* (pp. 45-54). ACM.
- Dave, B., & Koskela, L. (2009). Collaborative knowledge management - A construction case study. *Automation in construction*, *18*(7), 894-902.
- Davenport, T. H., De Long, D. W., & Beers, M. C. (1997). Building successful knowledge management projects. *Center for business innovation working paper*, *4*.
- Davtalaba, O., & Delgadob, J. L. (2014). Benefits of 6D BIM for Facilities Management Departments for Construction Projects–A Case Study Approach. *Southern California, University of Southern California*
- Deshpande, A., Azhar, S., & Amireddy, S. (2014). A framework for a BIM-based knowledge management system. *Procedia Engineering*, *85*, 113-122.
- Deveraja, H. (2015). *Perspective: Knowledge based Engineering (KBE)*. Bangalore: Infosys. Retrieved from Infosys.
- Ding, L. Y., Zhong, B. T., Wu, S., & Luo, H. B. (2016). Construction risk knowledge management in BIM using ontology and semantic web technology. *Safety science*, *87*, 202-213.
- Dittrich, J., & Quiané-Ruiz, J. A. (2012). Efficient big data processing in Hadoop MapReduce. *Proceedings of the VLDB Endowment*, *5*(12), 2014-2015.

- Eastman, C. M., Teicholz, P., Sacks, R., & Liston, K. (2011). *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors*. John Wiley & Sons.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), 37.
- Grover, R., & Froese, T. M. (2016). Knowledge Management in Construction Using a SocioBIM Platform: A Case Study of AYO Smart Home Project. *Procedia Engineering*, *145*, 1283-1290.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, *3*(Mar), 1157-1182.
- Hämäläinen, W. (2006). *Descriptive and predictive modelling techniques for educational technology.* Licentiate thesis, Department of Computer Science, Joensuu, Finland: University of Joensuu.
- Ho, S. P., Tserng, H. P., & Jan, S. H. (2013). Enhancing knowledge sharing management using BIM technology in construction. *The Scientific World Journal*, 2013. http://dx.doi.org/10.1155/2013/170498
- Hoogeveen, N. (2015). *Building Knowledge Modelling, from BIM data to Artificial Intelligent Knowledge Systems.* Delft: TU Delft Faculty Civil Engineering and Geosciences.
- Howitt, D., & Cramer, D. (2005). *Introduction to statistics in psychology.* Essex, England: Pearson Education.
- Huang, M. J., Chen, M. Y., & Lee, S. C. (2007). Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis. *Expert systems with applications*, *32*(3), 856-867.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, *22*(4), 679-688.
- IBM (n.d.). *The Four V's of Big Data.* Retrieved 2016, October 18, from http://www.ibmbigdatahub.com/infographic/four-vs-big-data
- IBM Knowledge Centre (2017). *What is distributed computing.* Retrieved on 17 march 2017 from: http://www.ibm.com/support/knowledgecenter/SSAL2T_7.1.0/com.ibm.cics.tx.doc/concepts/c_wht_is_distd_comptg.html
- International Organization for Standardization. (2007). *Building construction -- Organization of information about construction works -- Part 3: Framework for object-oriented information* (ISO Standard No. 12006-3:2007). Retrieved from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=38706
- Jin, R., Cho, K., Hyun, C., & Son, M. (2012). MRA-based revised CBR model for cost prediction in the early stage of construction projects. *Expert Systems with Applications*, *39*(5), 5214-5222.
- Johannessen, J. A., Olaisen, J., & Olsen, B. (2001). Mismanagement of tacit knowledge: the importance of tacit knowledge, the danger of information technology, and what to do about it. *International journal of information management*, *21*(1), 3-20.
- John, G. H. (1997). *Enhancements to the data mining process* (Doctoral dissertation, Stanford USA: Stanford university).

- Kamsu-Foguem, B., & Abanda, F. H. (2015). Experience modeling with graphs encoded knowledge for construction industry. *Computers in Industry*, *70*, 79-88.
- Kanapeckiene, L., Kaklauskas, A., Zavadskas, E. K., & Seniut, M. (2010). *Integrated knowledge management model and system for construction projects. Engineering Applications of Artificial Intelligence*. https://doi.org/10.1016/j.engappai.2010.01.030
- Khosrowshahi, F. (2011). Innovation in artificial neural network learning: Learn-On-Demand methodology. *Automation in Construction*, *20*(8), 1204-1210.
- Kock, N. F., McQueen, R. J., & Corner, J. L. (1997). The nature of data, information and knowledge exchanges in business processes: implications for process improvement and organizational learning. *The Learning Organization*, *4*(2), 70-80.
- Konukcu, S., & Koseoglu, O. (2012, October). *Knowledge management through BIM in construction supply chains.* In The 29th International Conference–Beirut, Lebanon (pp. 17-19).
  Koskela, L. (2000). *An exploration towards a production theory and its application to construction*. VTT Technical Research Centre of Finland. Helsinki: University of Technology (Espoo, Finland)
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variaty. *Application Delivery Strategies.* Stamford: META Group Inc.
- Li, S., Harner, E. J., & Adjeroh, D. A. (2011). Random KNN feature selection-a fast and stable alternative to Random Forests. *BMC bioinformatics*, *12*(1), 450.
- Lin, Y. C., Wang, L. C., & Tserng, H. P. (2006). Enhancing knowledge exchange through web map-based knowledge management system in construction: Lessons learned in Taiwan. *Automation in Construction*, *15*(6), 693-705.
- Lin, Y. C. (2014). Construction 3D BIM-based knowledge management system: a case study. *Journal of Civil Engineering and Management*, *20*(2), 186-200.
- Linoff, G. S., & Berry, M. J. (2011). *Data mining techniques: for marketing, sales, and customer relationship management*. Third edition, Indianapolis: John Wiley & Sons.
- Ma, Z., Wei, Z., & Zhang, X. (2013). Semi-automatic and specification-compliant cost estimation for tendering of building projects based on IFC data of design model. *Automation in Construction*, *30*, 126-135.
- Mathew, P. A., Dunn, L. N., Sohn, M. D., Mercado, A., Custudio, C., & Walter, T. (2015). Big-data for building energy performance: Lessons from assembling a very large national database of building energy use. *Applied Energy*, 140, 85-93.
- Meadati, P., & Irizarry, J. (2010, April). BIM–a knowledge repository. In *Proceedings of the 46th Annual International Conference of the Associated Schools of Construction, Retrieved November* (Vol. 12, p. 2010).
- Mi, C., Qian, H., Liu, S., & Chang, Z. (2008, June). Study on case retrieving in case-based reasoning based on grey incidence theory and its application in bank regulation. In *Fuzzy Systems, 2008. FUZZ-IEEE 2008.(IEEE World Congress on Computational Intelligence). IEEE International Conference on* (pp. 1530-1533). IEEE.
- Motawa, I., & Almarshad, A. (2013). A knowledge-based BIM system for building maintenance. *Automation in Construction*, *29*, 173-182.
- National BIM Standard (NBIMS) (n.d.). *What is BIM?* Retrieved on 19 march 2017 from: https://www.nationalbimstandard.org/faqs#faq1
- Nonaka, I. (1991) *The Knowledge creating Company Harvard Business Review, **69** (6 Nov-Dec): 96–104*

- Nonaka, I., & Lewin, A. Y. (1994). A Dynamic Theory of Organizational Knowledge Creation, 5(1), 14–37. Retrieved from http://www.jstor.org/stable/2635068
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation.* Oxford: Oxford university press.
- Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, Ba and leadership: a unified model of dynamic knowledge creation. *Long range planning*, *33*(1), 5-34.
- Phaobunjong, K. (2002). *Parametric cost estimating model for conceptual cost estimating of building construction projects* (Doctoral dissertation). Austin USA: the University of Texas
- Pozzali, A., & Viale, R. (2015). "Cognition, Types of "Tacit Knowledge" and Technology Transfer" *In Cognitive Economics: New Trends*. Published online: 08 Mar 2015; 205-224.
- Quinlan, J. R. (1987). Simplifying decision trees. *International journal of man-machine studies*, *27*(3), 221-234.
- Redmond, A., Hore, A., Alshawi, M., & West, R. (2012). Exploring how information exchanges can be enhanced through Cloud BIM. *Automation in Construction*, *24*, 175-183.
- Rezgui, Y., Hopfe, C. J., & Vorakulpipat, C. (2010). Generations of knowledge management in the architecture, engineering and construction industry: An evolutionary perspective. *Advanced Engineering Informatics*, *24*(2), 219-228.
- Sayad, S., (2010). *K Nearest Neighbors*. Retrieved from: http://chem-eng.utoronto.ca/~datamining/Presentations/KNN.pdf
- Seni, G., & Elder, J. F. (2010). Ensemble methods in data mining: improving accuracy through combining predictions. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 2(1), 1-126.
- Shen, Z., & Issa, R. R. (2010). Quantitative evaluation of the BIM-assisted construction detailed cost estimates. *Journal of Information Technology in Construction* ITcon), Vol. 15, pg. 234-257.
- Smith, E. A. (2001). The role of tacit and explicit knowledge in the workplace. *Journal of knowledge Management*, *5*(4), 311-321.
- Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, *22*(1), 97-103.
- Stewart, R. A., & Spencer, C. A. (2006). Six-sigma as a strategy for process improvement on construction projects: a case study. *Construction Management and Economics*, *24*(4), 339-348.
- Sweet, S., & Pitt-Catsouphes, M. (2010). Talent Pressures and the Aging Workforce. Retrieved on March 16, 2017 from: http://www.bc.edu/content/dam/files/research_sites/agingandwork/pdf/publications/TMISR05_Construction.pdf
- The R Foundation (n.d.). *What is R, Introduction to R.* Retrieved on 17 march 2017 from: https://www.r-project.org/about.html
- Tibaut, A., & Jakoša, D. (n.d.). Development of Knowledge Model for Construction Projects. https://doi.org/10.1007/978-3-319-21009-4_19
- Vonk, A.S., Wilde, J.J. de & Groot, T.C. de (2016). Indexcijfers Woningbouw, 2de kwartaal 2016. Den Haag. Bouwkostenindex

- Vukovic, S., Delibasic, B., Uzelac, A., & Suknovic, M. (2012). A case-based reasoning model that uses preference theory functions for credit scoring. *Expert Systems with Applications*, *39*(9), 8389-8395.
- Walker, D. H., & Walker, D. H. (2016). Reflecting on 10 years of focus on innovation, organisational learning and knowledge management literature in a construction project management context. *Construction Innovation*, *16*(2), 114-126. No Title. (1999), (June), 1–81.
- Ward, J. S., & Barker, A. (2013). Undefined by data: a survey of big data definitions. *arXiv preprint arXiv:1309.5821*. Fife, Scotland. University of St. Andrews
- Weippert, A., & Kajewski, S. L. (2007). Value-adding Knowledge Management (KM) Framework: 'An AEC industry perspective'. *CIB W065/055 Commissions: Transformation through Construction*, 1-11.
- Williams, G., (2010) *Data mining desktop survival guide.* Retrieved 2017, januari 8, from: http://datamining.togaware.com/survivor/dmsurvivor.html
- Witten, I. H., & Frank, E. (2000). *Practical Machine Learning Tools and Techniques with Java implementations, Data Mining*. San Francisco, CA: Morgan Kaufmann Publishers
- Woo, J. H., Clayton, M. J., Johnson, R. E., Flores, B. E., & Ellis, C. (2004). Dynamic Knowledge Map: reusing experts' tacit knowledge in the AEC industry. *Automation in construction*, *13*(2), 203-207.
- Yew Wong, K. (2005). Critical success factors for implementing knowledge management in small and medium enterprises. *Industrial Management & Data Systems*, *105*(3), 261-279.

Appendix

Appendix 1: Structure of the IFC data format

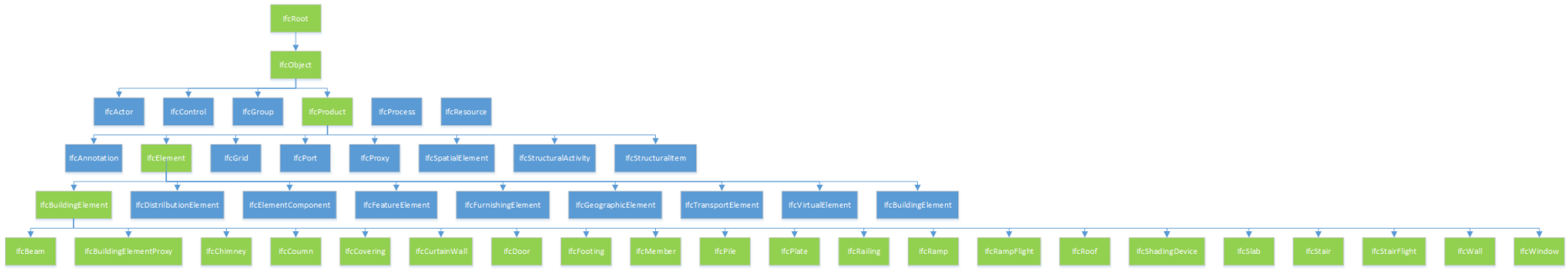

Figure 11: Structure of the IFC data format

Appendix 2: The extractor script

Start script

0_start_script_new.py

```
import glob
import subprocess
import sys
import os
import os.path
import csv
import datetime

# Get directory of script_dir
script_dir: (os.path.dirname(os.path.realpath(__file__)))

# Creating file to write
timestamp_date: datetime.datetime.now().strftime("%Y%m%d")
timestamp_time: datetime.datetime.now().strftime("%H%M%S")

save_path: script_dir+"/"+timestamp_date+"/"
if not os.path.exists(save_path):
    os.makedirs(save_path)

dataset_csv: os.path.join(save_path, timestamp_time+"_dataset.csv")

with open(dataset_csv, "wb") as f_csv:
    csvWriter: csv.writer(f_csv)
    csvWriter.writerow(["Filename", "areaGroundFloor",
"areaUpperFloor", "areaRoof", "countStair", "areaExternalWalls",
"areaExternalOpenings", "areaInternalWalls", "countInternalDoors"])

for fn in glob.glob("*.ifc"):
    # if fn not in processed_files:
    print fn
    subprocess.call([sys.executable, "1_QTO_new.py", fn,
dataset_csv])
    print "File processed"
```

Actual extractor

1_QTO_new.py

```
import os
import os.path
import glob
import sys
import math

from operator import itemgetter

import OCC.gp
import OCC.Geom
import OCC.Bnd
import OCC.BRepBndLib
import OCC.BRep
import OCC.BRepPrimAPI
import OCC.BRepAlgoAPI
import OCC.BRepBuilderAPI
```

```
17   import OCC.BRepAlgo
18   import OCC.TopOpeBRepTool
19   import OCC.ShapeExtend
20   import OCC.GProp
21   import OCC.BRepGProp
22   import OCC.GC
23   import OCC.ShapeAnalysis
24   import OCC.TopTools
25   import OCCUtils
26
27   from OCC.TopoDS import topods
28
29   import ifcopenshell
30   import ifcopenshell.geom
31
32   import operator
33
34   import csv
35
36   import re
37
38   # Specify to return pythonOCC shapes from
ifcopenshell.geom.create_shape()
39   settings: ifcopenshell.geom.settings()
40   settings.set(settings.USE_PYTHON_OPENCASCADE, True)
41
42
43   # Some helper functions to map to the list of walls
44   def create_shape(elem):
45       return ifcopenshell.geom.create_shape(settings, elem)
46
47
48   def calc_volume(s):
49       props: OCC.GProp.GProp_GProps()
50       OCC.BRepGProp.brepgprop_VolumeProperties(s.geometry, props)
51       return props.Mass()
52
53
54   def calc_area(s):
55       props: OCC.GProp.GProp_GProps()
56       OCC.BRepGProp.brepgprop_SurfaceProperties(s.geometry, props)
57       return props.Mass()
58
59
60   def calc_length(LProps):
61       props: OCC.GProp.GProp_GProps()
62       OCC.BRepGProp.brepgprop.LinearProperties(LProps.geometry, props)
63       return props.Mass()
64
65
66   def count_windows(product):
67       if product.is_a("IfcWindow"):
68           global ifcwindow
69           ifcwindow += 1
70
71
72   def count_doors(product):
73       if product.is_a("IfcDoor"):
74           global ifcdoor
75           ifcdoor += 1
76
```

```
77
78   def count_stories(product):
79       if product.is_a("IfcBuildingStorey"):
80           global ifcstorey
81           ifcstorey += 1
82
83
84   def create_levels(product):
85       if product.is_a("IfcBuildingStorey"):
86           level: product.Elevation
87           if level is not None:
88               floor_levels.append((level))
89           else:
90               level: 0
91               floor_levels.append((level))
92
93
94   def get_bounding_box_minimal(bbox):
95       bbmin: [0.] * 3;
96       bbmax: [0.] * 3
97       bbmin[0], bbmin[1], bbmin[2], bbmax[0], bbmax[1], bbmax[2]:
bbox.Get()
98       return OCC.gp.gp_Pnt(*map(lambda xy: (xy[0]), zip(bbmin,
bbmax)))
99
100
101  def get_bounding_box_maximal(bbox):
102      bbmin: [0.] * 3;
103      bbmax: [0.] * 3
104      bbmin[0], bbmin[1], bbmin[2], bbmax[0], bbmax[1], bbmax[2]:
bbox.Get()
105      return OCC.gp.gp_Pnt(*map(lambda xy: (xy[1]), zip(bbmin,
bbmax)))
106
107
108  def get_zvalue(product):
109      if product.Representation is not None:
110          shape: ifcopenshell.geom.create_shape(settings,
product).geometry
111          bbox: OCC.Bnd.Bnd_Box()
112          OCC.BRepBndLib.brepbndlib_Add(shape, bbox)
113          bounding_box_minimal: get_bounding_box_minimal(bbox)
114          bounding_box_maximal: get_bounding_box_maximal(bbox)
115          return (bounding_box_minimal.Z(), bounding_box_maximal.Z())
116
117  def get_xvalue(product):
118      if product.Representation is not None:
119          shape: ifcopenshell.geom.create_shape(settings,
product).geometry
120          bbox: OCC.Bnd.Bnd_Box()
121          OCC.BRepBndLib.brepbndlib_Add(shape, bbox)
122          bounding_box_minimal: get_bounding_box_minimal(bbox)
123          bounding_box_maximal: get_bounding_box_maximal(bbox)
124          return (bounding_box_minimal.X(), bounding_box_maximal.X())
125
126  def get_yvalue(product):
127      if product.Representation is not None:
128          shape: ifcopenshell.geom.create_shape(settings,
product).geometry
129          bbox: OCC.Bnd.Bnd_Box()
130          OCC.BRepBndLib.brepbndlib_Add(shape, bbox)
```

```
131             bounding_box_minimal: get_bounding_box_minimal(bbox)
132             bounding_box_maximal: get_bounding_box_maximal(bbox)
133             return (bounding_box_minimal.Y(), bounding_box_maximal.Y())
134
135     def get_values(product):
136         if product.Representation is not None:
137             shape: ifcopenshell.geom.create_shape(settings,
product).geometry
138             bbox: OCC.Bnd.Bnd_Box()
139             OCC.BRepBndLib.brepbndlib_Add(shape, bbox)
140             bounding_box_minimal: get_bounding_box_minimal(bbox)
141             bounding_box_maximal: get_bounding_box_maximal(bbox)
142             return (bounding_box_minimal.X(), bounding_box_maximal.X(),
bounding_box_minimal.Y(), bounding_box_maximal.Y(),
bounding_box_minimal.Z(), bounding_box_maximal.Z())
143
144     # creating temporary file
145     script_dir: (os.path.dirname(os.path.realpath(__file__)))
146     save_path_temp: script_dir + "/temp/"
147     if not os.path.exists(save_path_temp):
148         os.makedirs(save_path_temp)
149
150     # Open the IFC file using IfcOpenShell
151     ifc_file: ifcopenshell.open(sys.argv[1])
152     name_file: str(ifc_file)
153
154
155     print "Calculating Height of Elements"
156
157     products: ifc_file.by_type("IfcBuildingStorey")
158
159     levels: []
160
161     for product in products:
162         product_str: str(product)
163         product_string: product_str.lower()
164         count: product_string.count(",")
165         abc: product_string.split(",")
166         tempH: abc[-1]
167         if count > 10 :
168             placement: abc[6]
169         else:
170             placement: abc[5]
171         name: abc[2]
172         tempSplit: tempH.split(")")
173         h: tempSplit[0]
174         h1: float(h)
175         height: h1 / 1000
176         if height > -0.5:
177             level: [placement, height, name]
178             levels.append(level)
179
180     levels.sort(key=lambda x: x[1])
181
182     products: ifc_file.by_type("IfcLocalPlacement")
183
184     localPlacement: []
185
186     for product in products:
187
188         product_str: str(product)
```

```
189         product_string: product_str.lower()
190         abc: product_string.split(",")
191         temp1: abc[0]
192         temp1Split: temp1.split("(")
193         temp: temp1Split[0]
194         temp0: temp.split("=")
195         temp1: temp0[0]
196         temp2: temp1Split[1]
197
198         temp: [temp1, temp2]
199
200
201         localPlacement.append(temp)
202
203     for index, level in enumerate(levels):
204         test1: level[0]
205         for item in localPlacement:
206             test2: item[1]
207             if test2:= test1:
208                 tempList: []
209                 test3: item[0]
210                 tempList.append(test3)
211                 if any(option in tempList for option in level): continue
212                 level: level + tempList
213                 levels[index]: level
214
215     products: ifc_file.by_type("IfcWall")
216     WallProductsPerLevel: []
217     for level in levels:
218         list: []
219         WallProductsPerLevel.append(list)
220
221     for product in products:
222         product_str: str(product)
223         product_string: product_str.lower()
224         abc: product_string.split(",")
225         temp: abc[-3]
226         for index, level in enumerate(levels):
227             if temp in level:
228                 WallProductsPerLevel[index].append(product)
229
230     products: ifc_file.by_type("IfcSlab")
231     SlabProductsPerLevel: []
232     SlabAreaPerLevel: []
233     for level in levels:
234         list: []
235         SlabProductsPerLevel.append(list)
236
237     for product in products:
238         product_str: str(product)
239         product_string: product_str.lower()
240         abc: product_string.split(",")
241         temp: abc[-4]
242         for index, level in enumerate(levels):
243             if temp in level:
244                 SlabProductsPerLevel[index].append(product)
245
246     print "Calculating IfcSlabs"
247
248     for index, list in enumerate(SlabProductsPerLevel):
249         products: list
```

```
250
251         # vocabulary IfcSlab
252         wideSlabFloor_options: ["gewapend beton"]
253         hollowCoreSlab_options: ["pbh prefabvloer", "kanaalplaat",
"leidingplaat"]
254         groundFloor_options: ["begane grondvloer", "gewapende betonvloer
berging"]
255         ribbedSlab_options: ["ribcassette bg-vloer", "tempex bg-vloer"]
256         prefabSlab_options: ["balkon", "gallerijplaat"]
257         creeds_options: ["afwerkvloer", "cementdekvloer"]
258         insulationSlab_options: ["isolatie"]
259         roof_options: ["dak"]
260         slab_ignore_options: ["tegelvloer", "vergaarbak", "gras",
"asfalt", "4e verdiepingsvloer"]
261
262         # attributes IfcSlab
263         wideSlabFloor_area: 0
264         hollowCoreSlab_area: 0
265         groundFloor_area: 0
266         ribbedSlab_area: 0
267         prefabSlab_area: 0
268         creeds_area: 0
269         insulationSlab_area: 0
270         roof_area: 0
271         undefinedSlab_area: 0
272         VolumeSlabs: 0
273
274         z_levels: []
275
276         area_groundFloor: 0
277         area_RoofFlat: 0
278         areaIgnore_options: ["isolatie"]
279
280
281         for product in products:
282             z: get_zvalue(product)
283             if z is not None:
284                 z0: z[0]
285                 z1: z[1]
286                 thickness: z1 - z0
287                 if thickness < 0.15: continue
288                 # Create geometry for these windows
289                 shape: create_shape(product)
290
291                 # Calculate their volumes
292                 volume: calc_volume(shape)
293                 abs_volume: abs(volume)
294                 VolumeSlabs += abs_volume
295
296                 area: abs_volume / thickness
297
298                 product_str: str(product)
299                 product_string: product_str.lower()
300                 if any(option in product_string for option in
wideSlabFloor_options):
301                     wideSlabFloor area += area
302                 elif any(option in product_string for option in
hollowCoreSlab_options):
303                     hollowCoreSlab_area += area
304                 elif any(option in product_string for option in
groundFloor_options):
```

```
                    groundFloor_area += area
                elif any(option in product_string for option in ribbedSlab_options):
                    ribbedSlab_area += area
                elif any(option in product_string for option in prefabSlab_options):
                    prefabSlab_area += area
                elif any(option in product_string for option in creeds_options):
                    creeds_area += area
                elif any(option in product_string for option in insulationSlab_options):
                    insulationSlab_area += area
                elif any(option in product_string for option in roof_options):
                    roof_area += area
                elif any(option in product_string for option in slab_ignore_options):
                    pass
                else:
                    undefinedSlab_area += area

        totalarea: wideSlabFloor_area + hollowCoreSlab_area + groundFloor_area + ribbedSlab_area + prefabSlab_area + roof_area + undefinedSlab_area

        level: levels[index]
        nameLevel: level[2]
        heightLevel: level[1]
        listArea: []
        listArea.append(totalarea)
        SlabAreaPerLevel: SlabAreaPerLevel + listArea

    sumSlabArea: sum(SlabAreaPerLevel)

    if sumSlabArea != 0:
        for index, item in enumerate(levels):
            i: index
            i2: i + 1
            try:
                NextLevel: levels[i2]
                heightToNextLevel: NextLevel[1] - item[1]

                if 1.85 <= heightToNextLevel <= 1.95 or heightToNextLevel < 0.11 or 2.524 < heightToNextLevel < 2.526:
                    SlabArea: SlabAreaPerLevel[i] + SlabAreaPerLevel[i2]
                    SlabAreaPerLevel[i]: 0
                    SlabAreaPerLevel[i2]: SlabArea

            except IndexError:
                pass

        SlabAreaPerLevel2: []

        for area in SlabAreaPerLevel:
            if area > 0:
                SlabAreaPerLevel2.append(area)

        areaRoof: 0
        try:
            for index, item in enumerate(SlabAreaPerLevel2):
```

```
357                 index2: index + 1
358                 roof: item - SlabAreaPerLevel2[index2]
359                 if roof > 0:
360                     areaRoof += roof
361         except IndexError:
362             pass
363
364         sumSlabArea: sum(SlabAreaPerLevel)
365
366         areaGroundFloor: SlabAreaPerLevel2[0]
367         areaRoof += SlabAreaPerLevel2[-1]
368         areaUpperFloor: sumSlabArea - areaGroundFloor - areaRoof
369
370     else:
371         SlabAreaPerLevel: []
372         for index, list in enumerate(SlabProductsPerLevel):
373             products: list
374
375             # vocabulary IfcSlab
376             slab_ignore_options: ["tegelvloer", "vergaarbak", "gras",
"asfalt", "4e verdiepingsvloer", "ontgraving"]
377
378             # attributes IfcSlab
379             wideSlabFloor_area: 0
380             hollowCoreSlab_area: 0
381             groundFloor_area: 0
382             ribbedSlab_area: 0
383             prefabSlab_area: 0
384             creeds_area: 0
385             insulationSlab_area: 0
386             roof_area: 0
387             undefinedSlab_area: 0
388             VolumeSlabs: 0
389
390             z_levels: []
391
392             area_groundFloor: 0
393             area_RoofFlat: 0
394             areaIgnore_options: ["isolatie"]
395
396             for product in products:
397                 z: get_zvalue(product)
398                 if z is not None:
399                     z0: z[0]
400                     z1: z[1]
401                     thickness: z1 - z0
402                     # Create geometry for these windows
403                     shape: create_shape(product)
404
405                     # Calculate their volumes
406                     volume: calc_volume(shape)
407                     abs_volume: abs(volume)
408                     VolumeSlabs += abs_volume
409
410                     area: abs_volume / thickness
411
412                     product_str: str(product)
413                     product_string: product_str.lower()
414                     if any(option in product_string for option in
slab_ignore_options):
415                         pass
```

```
416                     else:
417                         undefinedSlab_area += area
418
419
420
421             totalarea: undefinedSlab_area
422             level: levels[index]
423             nameLevel: level[2]
424             heightLevel: level[1]
425             listArea: []
426             listArea.append(totalarea)
427             if totalarea > 0:
428                 SlabAreaPerLevel: SlabAreaPerLevel + listArea
429
430         areaRoof: 0
431         try:
432             for index, item in enumerate(SlabAreaPerLevel):
433                 index2: index + 1
434                 roof: item - SlabAreaPerLevel[index2]
435                 if roof > 0:
436                     areaRoof += roof
437         except IndexError:
438             pass
439
440         sumSlabArea: sum(SlabAreaPerLevel)
441
442         areaGroundFloor: SlabAreaPerLevel[0]
443         areaRoof += SlabAreaPerLevel[-1]
444         areaUpperFloor: sumSlabArea - areaGroundFloor - areaRoof
445
446     products1: ifc_file.by_type("IfcBuildingElementProxy")
447     products2: ifc_file.by_type("IfcFurnishingElement")
448     products: products1 + products2
449
450
451     stair_options: ["nl0-dht_nl+1-oht_2kwart", "trap blok b 2", "trap
blok a 1", "vwn_houten trap-dk"]
452     stair_ignore: ["trapleuning","traphek"]
453     countStair: 0
454
455     for product in products:
456         product_str: str(product)
457         product_string: product_str.lower()
458         if any(option in product_string for option in stair_options):
459             countStair += 1
460
461     listOfStairs: []
462     products: ifc_file.by_type("IfcStair")
463
464     for product in products:
465         product_str: str(product)
466         product_string: product_str.lower()
467
468         abc: product_string.split(",", 3)
469         abc2: abc[2]
470         if any(option in product_string for option in stair_ignore):
continue
471         listOfStairs.append(abc2)
472     countStair += len(listOfStairs)
473
474     print "Calculating IfcWalls"
```

```
475  products: ifc_file.by_type("IfcWall")
476
477  # vocabulary IfcWall
478  prefabWall_options: ["prefab beton_100", "prefab beton-90"]
479  insulationWall_options: ["isolatie", "steenwol", "rockpanel",
"isover", "kingspan"]
480  brickwork_options: ["metselwerk", "mets.", "baksteen", "terca",
"uitstekende kop", "vlakke kop", "mw 100"]
481  limestone_options: ["kalkzandsteen", "kzst."]
482  timberFrame_options: ["hsbwand", "hsb wand", "hsb_100mm",
"hsb_150mm", "hsb_214mm", "21hsb_150"]
483  cellularConcrete_options: ["cellenbeton", "gasbeton", "ytong", "gas
beton"]
484  gypsum_options: ["gibo"]
485  insideWall_options: ["binnenwand"]
486  ignore_options: ["ms-wand", "privacyscherm", "gaas met hedera",
"wandtegels", "staal_10mm", "gasbeton_100/g4_600", "beplating", "31
paneel_kozijn",
487                     "41_beplating_15", "21_kozijnrekje 80mm",
"20_zink", "stucwerk", "21_stijl en regelwerk 18+71", "betonnen
kantplank_100mm", "multiplex",
488                     "foamglass", "berging", "schutting", "hekwerk",
"tegelwand", "prefabplint", "fundering", "houtregel"]
489
490  prefabWall_area: 0
491  insulationWall_area: 0
492  brickwork_area: 0
493  limestone_area: 0
494  timberFrame_area: 0
495  cellularConcrete_area: 0
496  gypsum_area= 0
497  insideWall_area: 0
498  undefinedWall_area: 0
499
500  limestone_cs36_300: 0
501  limestone_cs36_214: 0
502  limestone_cs36_150: 0
503  limestone_cs36_120: 0
504  limestone_cs36_100: 0
505  limestone_300: 0
506  limestone_214: 0
507  limestone_150: 0
508  limestone_120: 0
509  limestone_100: 0
510  limestone_undefined: 0
511
512  VolumeWalls: 0
513  errorCount_wall: 0
514
515  products: ifc_file.by_type("IfcWall")
516
517  for product in products:
518      try:
519          x: get_xvalue(product)
520          y: get_yvalue(product)
521          z: get_zvalue(product)
522
523          x0: x[0]
524          x1: x[1]
525
526          y0: y[0]
```

```
527             y1: y[1]
528
529             z0: z[0]
530             z1: z[1]
531
532             l: x1 - x0
533             b: y1 - y0
534             h: z1 - z0
535
536             if l > 0.5 and b > 0.5:
537                 length: math.sqrt((l*l)+(b*b))
538             elif l < b:
539                 length: b
540             else:
541                 length: l
542
543             area: length * h
544
545             shape: create_shape(product)
546             volume: calc_volume(shape)
547             abs_volume: abs(volume)
548             VolumeWalls += abs_volume
549
550             product_str: str(product)
551             product_string: product_str.lower()
552             found: 0
553             if any(option in product_string for option in
prefabWall_options):
554                 prefabWall_area += area
555             elif any(option in product_string for option in
insulationWall_options):
556                 insulationWall_area += area
557             elif any(option in product_string for option in
brickwork_options):
558                 brickwork_area += area
559             elif any(option in product_string for option in
limestone_options):
560                 limestone_area += area
561                 if "cs36" in product_string:
562                     if "300mm" in product_string or "e300" in
product_string:
563                         limestone_cs36_300 += area
564                     elif "214mm" in product_string or "e214" in
product_string:
565                         limestone_cs36_214 += area
566                     elif "150mm" in product_string or "e150" in
product_string:
567                         limestone_cs36_150 += area
568                     elif "120mm" in product_string or "e120" in
product_string:
569                         limestone_cs36_120 += area
570                     elif "100mm" in product_string or "e100" in
product_string:
571                         limestone_cs36_100 += area
572                     else:
573                         limestone_undefined += area
574                 else:
575                     if "300mm" in product_string or "e300" in
product_string:
576                         limestone_300 += area
```

```
577                     elif "214mm" in product_string or "e214" in
product_string:
578                         limestone_214 += area
579                     elif "150mm" in product_string or "e150" in
product_string:
580                         limestone_150 += area
581                     elif "120mm" in product_string or "e120" in
product_string:
582                         limestone_120 += area
583                     elif "100mm" in product_string or "e100" in
product_string:
584                         limestone_100 += area
585                     else:
586                         limestone_undefined += area
587             elif any(option in product_string for option in
insideWall_options):
588                 insideWall_area += area
589             elif any(option in product_string for option in
timberFrame_options):
590                 timberFrame_area += area
591             elif any(option in product_string for option in
cellularConcrete_options):
592                 cellularConcrete_area += area
593             elif any(option in product_string for option in
gypsum_options):
594                 gypsum_area += area
595             elif any(option in product_string for option in
ignore_options):
596                 pass
597             else:
598                 undefinedWall_area += area
599
600         except Exception as ex:
601             template: "An exception of type {0} occured.
Arguments:\n{1!r}"
602             message: template.format(type(ex).__name__, ex.args)
603             errorCount_wall += 1
604
605     areaExternalWalls: brickwork_area+ timberFrame_area
606     areaAllInternalWalls:  limestone_area + cellularConcrete_area +
gypsum_area + insideWall_area + undefinedWall_area + prefabWall_area
607     areaInternalWalls: areaAllInternalWalls - areaExternalWalls
608
609     if areaInternalWalls < 0:
610         areaInternalWalls: areaAllInternalWalls
611
612     print "Calculating IfcDoors"
613
614     products: ifc_file.by_type("IfcDoor")
615
616     # vocabulary IfcDoor
617     externalDoors_option: ["fit_deur", "31_do_deur", "vision trend
merk", "31_deur", "31_stoldeur", "deur-hout_by", "deur-hout_stoldeur",
"buitendeur", "garagedeur"]
618
619     # attributes IfcDoor
620     externalDoors_area: 0
621     countInternalDoors: 0
622
623
624     for product in products:
```

```
625         product_str: str(product)
626         product_string: product_str.lower()
627         abc: product_string.split(",")
628
629         h1: abc[-2]
630         h2: h1.split(".")
631         h3: h2[0]
632         b1: abc[-1]
633         b2: b1.split(".")
634         b3: b2[0]
635
636         h: float(h3) / 1000
637         b: float(b3) / 1000
638         area: h * b
639         if any(option in product_string for option in
externalDoors option):
640             externalDoors_area += area
641
642         else:
643             if area > 1.5:
644                 countInternalDoors += 1
645
646     areaExternalWalls += externalDoors_area
647
648
649     print "Calculating IfcWindows"
650
651     products: ifc_file.by_type("Ifcwindow")
652
653     # attributes IfcDoor
654     externalWindow_area: 0
655
656     for product in products:
657         product_str: str(product)
658         product_string: product_str.lower()
659         abc: product_string.split(",")
660
661         h1: abc[-2]
662         h2: h1.split(".")
663         h3: h2[0]
664         b1: abc[-1]
665         b2: b1.split(".")
666         b3: b2[0]
667
668         h: float(h3) / 1000
669         b: float(b3) / 1000
670         area: h * b
671         externalWindow_area += area
672
673     areaExternalWalls += externalWindow_area
674
675     areaExternalOpenings: externalDoors_area +externalWindow_area
676
677     with open(sys.argv[2], "ab") as f_csv:
678         csvWriter: csv.writer(f_csv)
679         csvWriter.writerow([(sys.argv[1]), areaGroundFloor,
areaUpperFloor, areaRoof, countStair, areaExternalWalls,
areaExternalOpenings, areaInternalWalls, countInternalDoors])
```

Appendix 3: Code of the k-Nearest Neighbour

k-Nearest Neighbour

```
setwd("C:/test")
require(9825)
mydata: read.csv("dataset2.csv", header: TRUE, sep: ";", dec: ",")
set.seed(1)
gp <- runif(nrow(mydata))
mydata2 <- mydata[order(gp),]
normalize <- function(x) { return( (x - min(x)) / (max(x) - min(x)) ) }
data_n <- as.data.frame(lapply(mydata2[,c(2,3,4,5,6,7)], normalize))
data_train <- data_n[1:9, 1:5]
data_test <- data_n[10:12, 1:5]
data_train_target <- mydata2[1:9, 7]
data_test_target <- mydata2[10:12, 7]
m1 <- knn(train=data_train, test=data_test, cl=data_train_target, k=1)
m3 <- knn(train=data_train, test=data_test, cl=data_train_target, k=3)
m5 <- knn(train=data_train, test=data_test, cl=data_train_target, k=5)
data_train_target
data_test_target
m1
m3
m5
```

Appendix 4: Additional project data

Nieuwbouw 22 rijenwoning Dronten (source: casadata.nl)

Bouwkosten nieuwbouw 22 rijenwoningen, 3-laags met zadeldak, met gemiddeld 112 m2 bvo, 288 m3 bi (tabel p82.00.010)

Kosten in euro per eenheid (peildatum: 20 maart 2017)

Soort project	Omschrijving project	Bijzonderheid	Eenheid	Excl. BTW Totaal	Incl. BTW Eindtotaal
Nieuwbouw 22 rijenwoningen, 3-laags, koop en huur. Gemiddeld 112 m2 bvo, 288 m3 bi. Totaal project 2.471 m2 bvo en 6.329 m3 inhoud	Geïndexeerde kosten	Bouwkosten totaal	totaal	1.657.280,80	2.005.309,76
			woning	75.330,95	91.150,44
			m2 bvo	670,69	811,54
			m3 bi	261,86	316,84
		Waarvan bouwkundig werk	totaal	1.074.706,48	1.300.394,84
			woning	48.850,29	59.108,86

			m2 bvo	434,93	526,26
			m3 bi	169,81	205,47
		Waarvan installaties	totaal	293.271,70	354.858,75
			woning	13.330,53	16.129,94
			m2 bvo	118,69	143,61
			m3 bi	46,34	56,07
		Waarvan overige bouwkosten	totaal	289.302,62	350.056,18
			woning	13.150,12	15.911,64
			m2 bvo	117,08	141,67
			m3 bi	45,71	55,31
	Originele bouwkosten op origineel prijspeil april 2006	Totaal bouwkosten	totaal	1.372.483,00	1.660.704,43
			woning	62.385,59	75.486,57
		Waarvan bouwkundig werk	totaal	890.022,00	1.076.926,62
		Waarvan installaties	totaal	242.874,00	293.877,54
		Waarvan overige bouwkosten	totaal	239.587,00	289.900,27

Kosten

Uitgangspunt

Alle kosten in deze tabel zijn gebaseerd op de oorspronkelijke bouwkosten op het originele prijspeil van april 2006 van de aanbesteding. De projecthoeveelheden betreffen dit referentieproject.
Het gaat om de nieuwbouw van 22 rijenwoningen, die als koop- en huurwoningen in een project met 67 woningen in De Gilden in Dronten zijn gerealiseerd.

Kosten exclusief

De kosten zijn exclusief terreinkosten en bijkomende investeringskosten, zoals ontwerp- en advieskosten en de leges van (bouw)vergunningen.

Kosten inclusief

De kosten bouwkundig werk zijn inclusief sanitaire voorzieningen en bergingen.

Projectomschrijving

De rechthoekige woningen zijn gefundeerd op betonnen funderingsbalken op prefab betonnen palen. L = 4,5 m d = 290 mm. De gevel is volledig opgetrokken in metselwerk en gevuld met meranti buitenkozijnen met geïsoleerd glas en holonite vensterbanken.
Het dak bestaat uit een prefab houten dakconstructie, gedekt met betonpannen en voorzien van waterslagprofielen, dakkappen en schoorstenen. Hemelwaterafvoer via mastgoten.
Als begane grondvloer is een ribcasseteplaten en als verdiepingsvloer kanaalplaatvloeren toegepast, afgewerkt als

cementdekvloeren. Vloer in badkamer en toiletten zijn betegeld. Binnenwanden in kalkzandsteen en GIBO wanden, gevuld met stalen binnenkozijnen met opdekdeuren. Binnenwanden zijn afgewerkt met stucwerk en in de badkamer en toiletten betegeld.
Het plafond is in de woningen afgewerkt met spuitwerk en in de bergingen met houtwolcementplaat. Houten vaste trappen met aftimmering van trapgaten en lepe hoeken. Standaard sanitair inrichting in badkamers en toiletten.
Gasinstallatie voor CV-ketels. Verwarming met CV-ketel en radiatoren en mechanische ventilatie met warmteterugwinning units. Gangbaar leidingwerk voor elektrische installaties. Brandbeveiliging met optische rookmelders.

Algemene projectgegevens
De woningen hebben een nokhoogte van 7,3 m en een goothoogte van 5 meter..
De bruto inhoud van de 22 woningen is 6.329 m3 totaal ofwel gemiddeld 287,7 m3 per woning. De bruto vloeroppervlakte is totaal 2.471 m2 bvo ofwel gemiddeld 112,3 m2 bvo per woning.
Het gebouw bevat drie bouwlagen en heeft geen bouwlaag -peil.
De bruto dakoppervlakte is 1.177 m2 totaal (53,5 m2 per woning) en de bruto geveloppervlakte is 1.291 m2 (58,7 m2 per woning), waarvan 317 m2 met gevelopeningen is gevuld (25%).
In verhouding tot de bruto vloeroppervlakte is de geveloppervlakte 52% en is de omsluiting van gevel en dak samen 100%.

Bouwpartners
De informatie is samengesteld met medewerking van Huls Bouwkostenmanagement in Zwolle.
De opdrachtgever is Oost Flevoland Woondiensten in Dronten.
De architect is 19 het Atelier in Zwolle.
De aannemer is B.J. van der Steeg in Genemuiden.

Samenstelling gegevens
Disclaimer
Iedereen mag deze informatie raadplegen, afdrukken en gebruiken.
Casadata cv en de informatieverstrekkers stellen de informatie naar beste weten en kunnen zo zorgvuldig mogelijk samen. Casadata cv adviseert iedereen nadrukkelijk om niet uitsluitend deze informatie te gebruiken, maar deze altijd in samenhang met andere informatie te controleren.
Casadata cv en de informatieverstrekkers aanvaarden dan ook geen enkele aansprakelijkheid voor eventuele schade, die het gevolg is van handelingen of beslissingen, die voortkomen of gebaseerd zijn op door Casadata cv verstrekte informatie.

Nieuwbouw 31 rijenwoningen (source: casadata.nl)

Bouwkosten nieuwbouw 31 rijenwoningen, 2-laags met plat dak, met gemiddeld 121 m2 bvo, 360 m3 bi (tabel p82.00.009)

Kosten in euro per eenheid (peildatum: 20 maart 2017)

Soort project	Omschrijving project	Bijzonderheid	Eenheid	Excl. BTW Totaal	Incl. BTW Eindtotaal
Nieuwbouw 31 rijenwoningen, 2-laags. Gemiddeld 121 m2 bvo en 360 m3 bruto inhoud. Totaal project 3.757 m2 bvo en 11.153 m3 bi	Geïndexeerde kosten	Bouwkosten totaal	totaal	3.216.985,14	3.892.552,02
			woning	103.773,71	125.566,19
			m2 bvo	856,26	1.036,08
			m3 bi	288,44	349,01
	Originele bouwkosten op origineel prijspeil november 2004	Totaal bouwkosten	totaal	2.573.000,00	3.113.330,00
			woning	83.000,00	100.430,00

Kosten

Uitgangspunt

Alle kosten in deze tabel zijn gebaseerd op de oorspronkelijke bouwkosten op het originele prijspeil van november 2004 van de aanbesteding. De projecthoeveelheden betreffen dit referentieproject.
Het gaat om de nieuwbouw van 31 rijenwoningen in Rotterdam.

Kosten exclusief

De kosten zijn exclusief terreinafwerking en bijkomende investeringskosten, zoals ontwerp- en advieskosten en de leges van (bouw)vergunningen.

Projectomschrijving

Het nieuwbouwproject bestaat uit vier blokken met 31 eengezinswoningen. Elke ruimte heeft verdiepingshoge kozijnen, die overal even groot zijn. Daardoor kan men de ramen en deuren onderling uitwisselen. Door het verschuiven van de aluminium zonneschermen op de zuidgevel bepalen de bewoners zelf hun privacy en het zonlicht.
De buitenwanden zijn uitgevoerd in oranje metselwerk. In het gevelvlak zijn vierkante hemelwaterafvoeren opgenomen.
In de gevelopeningen zijn Sapupira hardhouten kozijnen met FSC keurmerk geplaatst.
De woningen zijn voorzien van standaard sanitair en standaard installaties voor gas, water en afvoeren.

Algemene projectgegevens

De woningen hebben een breedte van 5,4 m en het totaal van de vier woningblokken is 167,4 meter breed.
De bruto inhoud van de 31 woningen is 11.153 m3 totaal ofwel gemiddeld 359,8 m3 per woning. De bruto vloeroppervlakte is totaal 3.757 m2 bvo ofwel gemiddeld 121,2 m2 bvo per woning.
Het gebouw bevat twee bouwlagen en heeft geen bouwlaag -peil.
De bruto dakoppervlakte is 1.995 m2 totaal (64,4 m2 per woning) en de bruto geveloppervlakte is 2.632 m2 (85 m2 per woning), waarvan 770 m2 met gevelopeningen is gevuld (29%).
In verhouding tot de bruto vloeroppervlakte is de geveloppervlakte 70% en is de omsluiting van gevel en dak samen 123%.

Bouwpartners

De informatie is samengesteld met medewerking van Architectenbureau K2 te Amsterdam.
De opdrachtgever is Vestia Rotterdam Zuid.
De architect is Architectenbureau K2 te Amsterdam.
De aannemer is GN bouw b.v. in Dordrecht.

Samenstelling gegevens

Disclaimer

Iedereen mag deze informatie raadplegen, afdrukken en gebruiken.
Casadata cv en de informatieverstrekkers stellen de informatie naar beste weten en kunnen zo zorgvuldig mogelijk samen. Casadata cv adviseert iedereen nadrukkelijk om niet uitsluitend deze informatie te gebruiken, maar deze altijd in samenhang met andere informatie te controleren.
Casadata cv en de informatieverstrekkers aanvaarden dan ook geen enkele aansprakelijkheid voor eventuele schade, die het gevolg is van handelingen of beslissingen, die voortkomen of gebaseerd zijn op door Casadata cv verstrekte informatie.

Nieuwbouw 14 rijenwoningen Groenoord (source: casadata.nl)

Bouwkosten nieuwbouw 14 rijenwoningen, 4-laags met plat dak, met gemiddeld 195 m2 bvo, 591 m3 bi (tabel p83.00.010)

Kosten in euro per eenheid (peildatum: 20 maart 2017)

Soort project	Omschrijving project	Bijzonderheid	Eenheid	Excl. BTW Totaal	Incl. BTW Eindtotaal
Nieuwbouw 14 rijenwoningen, 4-laags. Gemiddeld 195 m2 bvo en 591 m3 bi per woning. Totaal project 2.730 m2 bvo en 8.270 m3 bi	Geïndexeerde kosten	Bouwkosten totaal	totaal	2.173.509,93	2.629.947,02
			woning	155.250,71	187.853,36
			m2 bvo	796,16	963,35
			m3 bi	262,82	318,01
	Originele bouwkosten op origineel prijspeil mei 2006	Totaal bouwkosten	totaal	1.800.000,00	2.178.000,00
			woning	128.571,43	155.571,43

Kosten

Uitgangspunt

Alle kosten in deze tabel zijn gebaseerd op de oorspronkelijke bouwkosten op het originele prijspeil van mei 2006 van de aanbesteding. De projecthoeveelheden betreffen dit referentieproject.
Het gaat om de nieuwbouw van 14 rijenwoningen in een reconstructiewijk uit de jaren zestig in Rotterdam zijn gerealiseerd.

Kosten exclusief

De kosten zijn exclusief terreinafwerking en bijkomende investeringskosten, zoals ontwerp- en advieskosten en de leges van (bouw)vergunningen.

Projectomschrijving

Door de smalle beukbreedte van 3,90 m van de woningen zijn er geen extra stabiliteitsvoorzieningen nodig, waardoor de woningen volledig flexibel zijn. Ook zijn er afwijkende doorsneden mogelijk, waardoor een kamer kan vervallen voor een dakterras of voor een extra hoogte in keuken, werk- of slaapkamer. Door de centrale plaats van de leidingschacht is op elke verdieping een extra badkamer of keukenblok mogelijk. De betonstructuur is op strategische plaatsen wapeningsvrij, zodat men later woningen of delen ervan kan samenvoegen.
Op de betonnen fundering is het skelet van het gebouw in gietbouw opgetrokken met een betonnen dak. Er zijn houtskeletbouw gevelelementen toegepast met kozijnen in combinatie met IBS vezelcementplaat. Het betonnen dak is afgewerkt met bitumineuze dakafwerking.
Er is gangbaar sanitair geplaatst. Ook zijn de woningen voorzien van gangbare woninginstallaties en leidingwerk voor water, binnenriolering, elektra en verlichting, verwarming en ventilatie. Er zijn zonnecollectoren geplaatst.

Algemene projectgegevens

De woningen hebben een breedte van 3,9 m en het totale woningblok is 54,6 meter breed.
De bruto inhoud van de 14 woningen is 8.270 m3 totaal ofwel gemiddeld 591 m3 per woning. De bruto vloeroppervlakte is totaal 2.730 m2 bvo ofwel gemiddeld 195 m2 bvo per woning.
Het gebouw bevat vier bouwlagen en heeft geen bouwlaag -peil.
De bruto dakoppervlakte is 645 m2 totaal (46 m2 per woning) en de bruto geveloppervlakte is 1.628 m2 (116 m2 per woning), waarvan 652 m2 met gevelopeningen is gevuld (40%).
In verhouding tot de bruto vloeroppervlakte is de geveloppervlakte 60% en is de omsluiting van gevel en dak samen 83%.

Bouwpartners

De informatie is samengesteld met medewerking van Architectenbureau K2 te Amsterdam.
De opdrachtgever is Wijk Ontwikkelingsmaatschappij Groenoord te Rotterdam.
De architect is Architectenbureau K2 te Amsterdam.
De aannemer is Riz bouw b.v. in IJsselstein.

Samenstelling gegevens

Disclaimer

Iedereen mag deze informatie raadplegen, afdrukken en gebruiken.
Casadata cv en de informatieverstrekkers stellen de informatie naar beste weten en kunnen zo zorgvuldig mogelijk samen. Casadata cv adviseert iedereen nadrukkelijk om niet uitsluitend deze informatie te gebruiken, maar deze altijd in samenhang met andere informatie te controleren.
Casadata cv en de informatieverstrekkers aanvaarden dan ook geen enkele aansprakelijkheid voor eventuele schade, die het gevolg is van handelingen of beslissingen, die voortkomen of gebaseerd zijn op door Casadata cv verstrekte informatie.

28 appartementen De Haere (source: bouwkosten.nl)

BOUWkosten

28 APPARTEMENTEN "DE HAERE" TE HEINO

TABEL | OMSCHRIJVING | AFBEELDINGEN

Omschrijving

Projectomschrijving

Op een voormalig schoolterrein in Heino wordt een woonproject van 28 huurappartementen met daarachter 8 geschakelde huurwoningen voor 55+'ers gerealiseerd met energielabel A.

Het appartementencomplex is opgezet in een U-vorm waarvan de twee uitlopers naar de doorgaande straat wijzen en een openbaar hofje omsluiten. Op de begane grond bevinden zich 8 appartementen met een eigen buitenruimte. De eerste verdieping en de tweede verdieping met kap bevatten elk 10 appartementen. Er bevinden zich 2 liften in het complex. De appartementen op de begane grond zijn allleen toegankelijk via de openbare weg. De overige appartementen zijn toegankelijk vanaf de centrale hal, welke is afgesloten. Bezoekers melden zich via de videofoon.

Aan de achterzijde van het appartementencomplex liggen aan een rustige woonstraat 8 seniorenwoningen. Deze woningen hebben 1 bouwlaag met een dakopbouw met eigen buitenruimte.

De appartementen en de laagbouwwoningen hebben een klassieke uitstraling door het gebruik van een eenvoudige rode baksteen –een geribbelde en een tweemaal zo grote gladde steen – die afwisselend in verticale stroken worden gebruikt.

De kozijnen zijn van kunststof met HR++ beglazing. De zuidoost- en zuidwestzijde zijn voorzien van elektrisch bedienbare zonwering.

De badkamer, keuken en toilet zijn (deels) betegeld. Overige wanden zijn behangklaar gemaakt. Plafonds: spuitwerk (spack), kleur wit. Vloer: afwerkvloer (cementdekvloer) behalve in de badkamer waar de vloer is betegeld.

Iedere woning/appartement is voorzien van een gebalanceerd ventilatiesysteem, een zonnecollector (PV) en een zonneboiler voor warmtapwater. Op het dak van het appartementencomplex bevinden zich ook zonnecollectoren (PV) welke elektriciteit levert voor de lift en de verlichting van de algemene ruimten.

Algemene projectgegevens

Samenstelling: Huls Bouwkostenmanagement
Opdrachtgever: Salland Wonen te Raalte
Architect: Van Den Berg IJsselmuiden te IJsselmuiden
Hoofdaannemer: Aannemersbedrijf J. Valk en Zn. B.V. te Daarle

Peildatum originele kosten: juni 2012
Start uitvoering: augustus 2012
Oplevering: november 2013
Bouwtijd: 15 kalendermaanden
Aanbestedingsvorm: bouwteam
Risicoregeling van toepassing: n.v.t.

Bruto-inhoud: 10.493 m3
Alle kosten zijn geïndexeerd.

Kosten- en gegevenstabellen

KOSTEN OP NIVEAU 1 GEHEEL BOUWWERK | BOUWFYSISCHE KENMERKEN | HOEVEELHEDEN | VORMFACTOREN
KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS | KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS, HOOFDINDELING | KOSTEN OP NIVEAU 3 ELEMENTEN
OMSCHRIJVING ELEMENTEN | RUIMTESOORTEN, TOTAAL | RUIMTESOORTEN, SPECIFICATIE

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	2.285.185,00		2.285.185,00
3	Installaties		565.105,00	565.105,00

4	Vaste inrichtingen		103.935,00	103.935,00
5	Terrein	12.535,00		12.535,00
	Totaal directe bouwkosten	2.297.720,-	669.040,-	2.966.760,00
6	Algemene bouwkosten	223.840,00		223.840,00
	Subtotaal	2.521.560,-	669.040,-	3.190.600,-
6	Algemene bedrijfskosten	133.390,00		133.390,00
	Subtotaal	2.654.950,-	669.040,-	3.323.990,-
6	Winst en risico	56.025,00		56.025,00
	Totaal bouwkosten (excl. BTW)	2.710.975,-	669.040,-	3.380.015,00

Bouwfysische kenmerken

Beganegrondvloer	R =	3,50	$m^2.K/W$
Geveiconstructie	R =	5,00	$m^2.K/W$
Glas	U =	1,10	$W/(m^2.K)$
Dakconstructie	R =	5,00	$m^2.K/W$
Energie Prestatie Coëfficiënt		0,60	

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	3	stuks	Dakoppervlakte, bruto	1.322	m^2
Maximale hoogte	11,00	m	Dakoppervlakte	1.322	m^2
Terrein			Dakopeningen		m^2
Terreinoppervlakte	2.520	m^2	Gevels		
Bebouwde terreinoppervlakte	1.165	m^2	Gevelomtrek	284	m
Onbebouwde terreinoppervlakte	1.355	m^2	Buitenwandoppervlakte, bruto	2.092	m^2
Oppervlakteverharding	215	m^2	Buitenwanden	1.486	m^2
Aantal parkeerplaatsen		stuks	Buitenwandopeningen	606	m^2
Ruimtedelen			Binnenwanden		
Bruto-vloeroppervlakte (BVO)	3.302	m^2	Binnenwandoppervlakte, bruto	4.303	m^2
Netto-vloeroppervlakte (NVO)	2.708	m^2	Binnenwanden	3.684	m^2
Installatieoppervlakte	36	m^2	Binnenwandopeningen	619	m^2
Verkeersoppervlakte	630	m^2	Trapconstructies		
Nuttige oppervlakte	2.042	m^2	Trappen en hellingen	4	stuks
Niet-toegankelijke ruimte		m^2	Balustrades en leuningen	34	m
Tarra-oppervlakte	594	m^2	Installaties		
Gebruiksoppervlakte	2.763	m^2	Klimaatinstallaties (koeling)		kW
Bruto-inhoud (BI)	10.493	m^3	Klimaatinstallaties (verwarming)		kW
Skelet			Warmtepomp		kW
Buitenwanden (constructief)	1.276	m^2	Zonnecollector		kW
Binnenwanden (constructief)	899	m^2	Vloerverwarming		m^2
Beganegrondvloeren (constructief)	1.165	m^2	Warmtelichamen		stuks
Verdiepingsvloeren (constructief)	2.137	m^2	Wandverwarming		m^2
Dakvloeren (constructief)	1.322	m^2	Liften	2	stuks

29 appartementen boven winkels (source: bouwkosten.nl)

BOUWkosten

29 APPARTEMENTEN BOVEN WINKELS TE ZWOLLE

TABEL | OMSCHRIJVING | AFBEELDINGEN

Omschrijving

Projectomschrijving

De N.V. Wonen boven Winkels Zwolle heeft in het centrum van de stad circa 800 m² nieuwe winkelruimte met daarboven in totaal 29 appartementen weten te realiseren. Daartoe is het overgrote deel van de eerdere bebouwing –na uitvoerige inventarisatie, bouwhistorisch onderzoek en modellenstudies– gesloopt. Het project is van forse omvang en heeft gevels aan drie straten: de Diezerstraat, Korte Smeden en Nieuwstraat.

De nieuwbouw moest binnen de voorschriften van het bestemmingsplan blijven; te weten drie lagen en een kap. De winkel op de begane grond wordt door de corporatie als één grote ruimte verhuurd, maar de winkelruimte is op te splitsen in meerdere kleine eenheden. Verdeeld over de bovenliggende drie verdiepingen een combinatie van 2 tweekamerappartementen, 12 driekamerappartementen en 15 studio's (eenkamerappartementjes à 25 m²).

De meeste wooneenheden liggen aan het binnenterrein en zijn met de rug tegen het aangrenzende achterliggende pand gebouwd. Ze hebben geen balkon, maar de bewoners kunnen buiten zitten op de galerij of op het binnenterrein.

In eerste instantie zou de hoofdingang aan de Nieuwstraat komen. Daar is uiteindelijk het noodtrappenhuis geplaatst. De ingang, het trappenhuis en de lift zijn nu gelegen aan de Korte Smeden. Voor de bewoners is daar ook de ingang van hun gemeenschappelijke ondergrondse fietsenberging geplaatst. Elk van de nieuwe of gerenoveerde gevels is aan de straatzijde anders van maat en kleur. Hierdoor sluiten deze aan bij de variëteit aan bestaande gevels in dat deel van de binnenstad.

Algemene projectgegevens
Samenstelling: Huls Bouwkostenmanagement
Opdrachtgever: N.V. Wonen Winkels Zwolle te Zwolle
Architect: IM architecten te Deventer
Hoofdaannemer: Salverda te 't Harde

Peildatum originele kosten: oktober 2006
Start uitvoering: september 2007
Oplevering: december 2008
Bouwtijd: 52 werkbare weken
Aanbestedingsvorm: bouwteam
Risicoregeling van toepassing: Nee

Bruto-inhoud: 11.688 m³
Alle kosten zijn geïndexeerd.

Kosten- en gegevenstabellen

KOSTEN OP NIVEAU 1 GEHEEL BOUWWERK | BOUWFYSISCHE KENMERKEN | HOEVEELHEDEN | VORMFACTOREN
KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS | KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS, HOOFDINDELING | KOSTEN OP NIVEAU 3 ELEMENTEN
OMSCHRIJVING ELEMENTEN | RUIMTESOORTEN, TOTAAL | RUIMTESOORTEN, SPECIFICATIE

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	3.042.910,00		3.042.910,00
3	Installaties		443.810,00	443.810,00

4	Vaste inrichtingen		64.665,00	64.665,00
5	Terrein			
	Totaal directe bouwkosten	3.042.910,-	508.480,-	3.551.385,00
6	Algemene bouwkosten	719.690,00		719.690,00
	Subtotaal	3.762.595,-	508.480,-	4.271.075,-
6	Algemene bedrijfskosten	239.385,00		239.385,00
	Subtotaal	4.001.980,-	508.480,-	4.510.460,-
6	Winst en risico	112.655,00		112.655,00
	Totaal bouwkosten (excl. BTW)	4.114.630,-	508.480,-	4.623.110,00

Bouwfysische kenmerken

Beganegrondvloer	R =	4,00	$m^2.K/W$
Gevelconstructie	R =	3,00	$m^2.K/W$
Glas	U =	1,70	$W/(m^2.K)$
Dakconstructie	R =	4,00	$m^2.K/W$
Energie Prestatie Coëfficiënt		0,96	

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	5	stuks	Dakoppervlakte, bruto	990	m^2
Maximale hoogte	15,55	m	Dakoppervlakte	979	m^2
Terrein			Dakopeningen	11	m^2
Terreinoppervlakte	990	m^2	Gevels		
Bebouwde terreinoppervlakte	990	m^2	Gevelomtrek	176	m
Onbebouwde terreinoppervlakte		m^2	Buitenwandoppervlakte, bruto	1.981	m^2
Oppervlakteverharding		m^2	Buitenwanden	1.416	m^2
Aantal parkeerplaatsen		stuks	Buitenwandopeningen	565	m^2
Ruimtedelen			Binnenwanden		
Bruto-vloeroppervlakte (BVO)	3.511	m^2	Binnenwandoppervlakte, bruto	1.411	m^2
Netto-vloeroppervlakte (NVO)	3.013	m^2	Binnenwanden	1.024	m^2
Installatieoppervlakte	63	m^2	Binnenwandopeningen	387	m^2
Verkeersoppervlakte	694	m^2	Trapconstructies		
Nuttige oppervlakte	2.256	m^2	Trappen en hellingen	12	stuks
Niet-toegankelijke ruimte		m^2	Balustrades en leuningen	49	m
Tarra-oppervlakte	498	m^2	Installaties		
Gebruiksoppervlakte	3.006	m^2	Klimaatinstallaties (koeling)		kW
Bruto-inhoud (BI)	11.688	m^3	Klimaatinstallaties (verwarming)		kW
Skelet			Warmtepomp		kW
Buitenwanden (constructief)	1.981	m^2	Zonnecollector		kW
Binnenwanden (constructief)	1.024	m^2	Vloerverwarming		m^2
Beganegrondvloeren (constructief)	990	m^2	Warmtelichamen		stuks
Verdiepingsvloeren (constructief)	2.521	m^2	Wandverwarming		m^2
Dakvloeren (constructief)	990	m^2	Liften		stuks
Hoofddraagconstructie	3.511	m^2			

44 appartementen de parkwachters (source: bouwkosten.nl)

BOUWkosten

44 APPARTEMENTEN DE PARKWACHTERS TE BOSKOOP

TABEL | OMSCHRIJVING | AFBEELDINGEN

Omschrijving

Projectomschrijving

Aan de Parklaan in Boskoop is een nieuw bouwproject gerealiseerd, bestaande uit 44 sociale huurwoningen. Deze drie- en vierkamerappartementen – met woonkamers van 37 tot 48 m^2 – zijn verdeeld over twee torens: 'De Parkwachters'. Het geheel ademt de sfeer van de jaren dertig van de vorige eeuw.

Het basisontwerp is eenvoudig: per bouwlaag staan vier woningen haaks op elkaar gegroepeerd rondom een liftschacht en een trappenhuis. De twee torens zijn op de begane grond verbonden door middel van bergingen. De hoogste toren heeft zeven lagen en 24 appartementen, de ander toren zes lagen en 20 appartementen. In elke toren bevindt zich een lift; één met zes en één met zeven stopplaatsen.

De appartementen in de woontorens op de begane grond beschikken over een terras, de overige appartementen hebben een balkon. De woningen hebben energielabel A, zijn levensloopbestendig en toegankelijk voor rolstoelgebruikers. Vanwege een hogedrukgasleiding in de buurt is het trappenhuis voorzien van druk- en brandwerende beglazing.

De kozijnen van de bergingen zijn van hout, de overige kozijnen zijn van kunststof. In de appartementen zijn de natte ruimten voorzien van keramische tegels met daarboven spuitwerk. De overige wanden zijn behangklaar. Installaties: HR-combiketel met kamerthermostaat. Gescheiden afvoersysteem. De appartementen worden verwarmd via radioatoren. In de badkamer bevindt zich een designradiator en basis sanitair inclusief tweede toilet.

Algemene projectgegevens

Samenstelling: Winket bv | Huisvestingseconomie, bouwkosten en bestekken
Opdrachtgever: Trifolium Woondiensten / Woonpartners Midden-Holland te Boskoop / Waddinxveen
Architect: Mullener en Mullener te Amsterdam
Hoofdaannemer: Van Wijnen te Stolwijk

Peildatum originele kosten: november 2011
Start uitvoering: april 2012
Oplevering: juli 2013
Bouwtijd: 240 werkbare werkdagen
Aanbestedingsvorm: bouwteam
Risicoregeling van toepassing: nee

Bruto-inhoud: 18.236 m3
Alle kosten zijn geïndexeerd.

Kosten- en gegevenstabellen

KOSTEN OP NIVEAU 1 GEHEEL BOUWWERK | BOUWFYSISCHE KENMERKEN | HOEVEELHEDEN | VORMFACTOREN
KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS | KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS, HOOFDINDELING | KOSTEN OP NIVEAU 3 ELEMENTEN
OMSCHRIJVING ELEMENTEN

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	3.569.430,00		3.569.430,00
3	Installaties		1.005.495,00	1.005.495,00
4	Vaste inrichtingen		113.740,00	113.740,00
5	Terrein	85.290,00		85.290,00

	Totaal directe bouwkosten	3.654.725,-	1.119.235,-	4.773.960,00
6	Algemene bouwkosten	402.455,00		402.455,00
	Subtotaal	4.057.180,-	1.119.235,-	5.176.415,-
6	Algemene bedrijfskosten	311.615,00		311.615,00
	Subtotaal	4.368.795,-	1.119.235,-	5.488.030,-
6	Winst en risico	177.590,00		177.590,00
	Totaal bouwkosten (excl. BTW)	4.546.380,-	1.119.235,-	5.665.615,00

Bouwfysische kenmerken

Beganegrondvloer	R =	3,50	$m^2.K/W$
Gevelconstructie	R =	3,50	$m^2.K/W$
Glas	U =	1,65	$W/(m^2.K)$
Dakconstructie	R =	4,00	$m^2.K/W$
Energie Prestatie Coëfficiënt		0,80	

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	6	stuks	Dakoppervlakte, bruto	1.368	m^2
Maximale hoogte	23,00	m	Dakoppervlakte	1.368	m^2
Terrein			Dakopeningen		m^2
Terreinoppervlakte	3.178	m^2	Gevels		
Bebouwde terreinoppervlakte	1.158	m^2	Gevelomtrek	1.503	m
Onbebouwde terreinoppervlakte	2.020	m^2	Buitenwandoppervlakte, bruto	4.358	m^2
Oppervlakteverharding	2.020	m^2	Buitenwanden	2.797	m^2
Aantal parkeerplaatsen	60	stuks	Buitenwandopeningen	1.561	m^2
Ruimtedelen			Binnenwanden		
Bruto-vloeroppervlakte (BVO)	6.766	m^2	Binnenwandoppervlakte, bruto	6.545	m^2
Netto-vloeroppervlakte (NVO)	6.022	m^2	Binnenwanden	5.467	m^2
Installatieoppervlakte	88	m^2	Binnenwandopeningen	1.078	m^2
Verkeersoppervlakte	1.056	m^2	Trapconstructies		
Nuttige oppervlakte	4.778	m^2	Trappen en hellingen	19	stuks
Niet-toegankelijke ruimte	100	m^2	Balustrades en leuningen	545	m
Tarra-oppervlakte	744	m^2	Installaties		
Gebruiksoppervlakte	5.616	m^2	Klimaatinstallaties (koeling)		kW
Bruto-inhoud (BI)	18.236	m^3	Klimaatinstallaties (verwarming)	880	kW
Skelet			Warmtepomp		kW
Buitenwanden (constructief)	2.290	m^2	Zonnecollector		kW
Binnenwanden (constructief)	1.689	m^2	Vloerverwarming		m^2
Beganegrondvloeren (constructief)	1.045	m^2	Warmtelichamen	352	stuks
Verdiepingsvloeren (constructief)	4.976	m^2	Wandverwarming		m^2
Dakvloeren (constructief)	1.368	m^2	Liften	2	stuks
Hoofddraagconstructie	6.766	m^2			

Vormfactoren

Allardhof te buren (source: bouwkosten.nl)

BOUWkosten

ALLARDHOF TE BUREN

TABEL | OMSCHRIJVING | AFBEELDINGEN

Omschrijving

Projectomschrijving

In Buren aan de Appelgaard stond een verouderd appartementencomplex. Kleurrijk Wonen (voorheen Stichting Volkshuisvesting Tiel) heeft dit complex gesloopt. Jeanne Dekkers Architectuur heeft 41 bijzondere appartementen ontworpen die in de plaats van het complex worden gebouwd: project Allardhof. Kleurrijk Wonen heeft ervoor gekozen het project Allardhof in twee fasen te realiseren. Fase 1: 25 huurappartementen, uitgewerkt in deze projectanalyse. Fase 2 bevat 16 koopappartementen.

Buren kent een dorpsachtig karakter. Als snel ontstond hierdoor het idee van een "dorp in een dorp", een mooie manier van wonen in een groene omgeving. De appartementen zijn verdeeld over meerdere kleinschalige gebouwen die in hoogte variëren en rondom een gezamenlijke binnentuin staan. De appartementen zijn gebouwd aan de Willem Alexanderstraat in Buren. Het hoogste gebouw telt drie verdiepingen. Het basismateriaal voor de gevel – een lichte en donkere baksteen – komt voor in de omliggende wijk. Door het echter uniform bij alle bouwdelen toe te passen, geeft het de nieuwbouw zijn eigen identiteit. Parkeren wordt opgelost aan de randen van het plan en onder twee gebouwen, waardoor de tuin vrij blijft van autoverkeer. De daken zijn allemaal bedekt met Sedum wat voor een mooi uitzicht zorgt vanuit de hogere appartementen. Alle appartementen hebben een terras of balkon en maken gebruik van de gezamenlijke binnentuin. Deze centrale tuin ligt in het hart van het plan is tevens een nieuwe, openbare plek voor de bestaande wijk. Door de ecologische inrichting, de mos-sedumdaken en de nieuwe vleermuiskasten is de tuin een bijzonder aangename plek om te verblijven voor mens én dier.

Algemene projectgegevens
Samenstelling: Huls Bouwkostenmanagement
Opdrachtgever: Stichting Volkshuisvesting Tiel te Tiel
Architect: Jeanne Dekkers Architectuur te Delft
Hoofdaannemer: Bouwbedrijf Berghege te Oss

Peildatum originele kosten: oktober 2013
Start uivoering: april 2014
Oplevering: januari 2015
Bouwtijd: 10 maanden
Aanbestedingsvorm: openbaar
Risicoregeling van toepassing: nee

Bruto-inhoud: 9.108 m3
Alle kosten zijn geïndexeerd.

Kosten- en gegevenstabellen

KOSTEN OP NIVEAU 1 GEHEEL BOUWWERK | BOUWFYSISCHE KENMERKEN | HOEVEELHEDEN | VORMFACTOREN
KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS | KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS, HOOFDINDELING | KOSTEN OP NIVEAU 3 ELEMENTEN
OMSCHRIJVING ELEMENTEN | RUIMTESOORTEN, TOTAAL | RUIMTESOORTEN, SPECIFICATIE

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	1.603.205,00		1.603.205,00
3	Installaties		472.765,00	472.765,00
4	Vaste inrichtingen		41.295,00	41.295,00
5	Terrein	22.735,00		22.735,00

	Totaal directe bouwkosten	1.625.945,-	514.060,-	2.140.005,00
6	Algemene bouwkosten	165.550,00		165.550,00
	Subtotaal	1.791.495,-	514.060,-	2.305.555,-
6	Algemene bedrijfskosten	89.725,00		89.725,00
	Subtotaal	1.881.220,-	514.060,-	2.395.280,-
6	Winst en risico	18.845,00		18.845,00
	Totaal bouwkosten (excl. BTW)	1.900.065,-	514.060,-	2.414.120,00

Bouwfysische kenmerken

Beganegrondvloer	R =	5,00	$m^2.K/W$
Geveiconstructie	R =	5,00	$m^2.K/W$
Glas	U =	1,10	$W/(m^2.K)$
Dakconstructie	R =	5,00	$m^2.K/W$
Energie Prestatie Coëfficiënt		0,60	

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	3	stuks	Dakoppervlakte, bruto	1.256	m^2
Maximale hoogte	10,71	m	Dakoppervlakte	1.256	m^2
Terrein			Dakopeningen		m^2
Terreinoppervlakte	7.360	m^2	Gevels		
Bebouwde terreinoppervlakte	1.220	m^2	Gevelomtrek	359	m
Onbebouwde terreinoppervlakte	6.104	m^2	Buitenwandoppervlakte, bruto	3.653	m^2
Oppervlakteverharding		m^2	Buitenwanden	3.088	m^2
Aantal parkeerplaatsen		stuks	Buitenwandopeningen	565	m^2
Ruimtedelen			Binnenwanden		
Bruto-vloeroppervlakte (BVO)	2.533	m^2	Binnenwandoppervlakte, bruto	2.537	m^2
Netto-vloeroppervlakte (NVO)	2.044	m^2	Binnenwanden	1.916	m^2
Installatieoppervlakte	33	m^2	Binnenwandopeningen	621	m^2
Verkeersoppervlakte	282	m^2	Trapconstructies		
Nuttige oppervlakte	1.729	m^2	Trappen en hellingen	6	stuks
Niet-toegankelijke ruimte		m^2	Balustrades en leuningen	184	m
Tarra-oppervlakte	489	m^2	Installaties		
Gebruiksoppervlakte	2.224	m^2	Klimaatinstallaties (koeling)		kW
Bruto-inhoud (BI)	9.108	m^3	Klimaatinstallaties (verwarming)	140	kW
Skelet			Warmtepomp		kW
Buitenwanden (constructief)	645	m^2	Zonnecollector		kW
Binnenwanden (constructief)	792	m^2	Vloerverwarming		m^2
Beganegrondvloeren (constructief)	1.220	m^2	Warmtelichamen		stuks
Verdiepingsvloeren (constructief)	1.313	m^2	Wandverwarming		m^2
Dakvloeren (constructief)	1.256	m^2	Liften	1	stuks
Hoofddraagconstructie	2.533	m^2			

Vormfactoren

Appartementencomplex de barchaan (source: bouwkosten.nl)

BOUWkosten

APPARTEMENTENCOMPLEX DE BARCHAAN TE ALMELO

TABEL | OMSCHRIJVING | AFBEELDINGEN

Omschrijving

Projectomschrijving
Op een voormalig fabrieksterrein verrezen koop- en huurwoningen, een appartementencomplex en een winkelcentrum. Een van deze gebouwen is De Barchaan, gebouwd in de jaren 30-stijl.
Het gebouw bestaat uit 4 bouwlagen. Op de begane grond aan twee zijden bevindt zich een commerciële ruimte van 360 m^2. Deze is vrij in te delen in kleinere segmenten. Verdeeld over 3 verdiepingen bevinden zich 14 luxe appartementen waar mensen met een verstandelijke beperking zelfstandig kunnen wonen. Tevens is er voor deze mensen een gemeenschappelijke ruimte beschikbaar voor onder andere dagactiviteiten.
Alle appartementen zijn 65 m^2 en 78 m^2 groot, zijn twee aan twee gespiegeld en hebben een eigen voordeur naar een inpandige galerij met lift en trappenhuis. Een lichtkoepel in het dak van de galerij en glazen lamellen voorzien voor voldoende daglicht. Het appartement bevat een woonkamer, keuken, slaapkamer, badkamer, een berging/hobbyruimte en een ruime loggiabalkon. Tevens bevat het een elux-systeem; een loos leidingsysteem voor alle toekomstige domotica-voorzieningen. De onderlinge scheidingswand van de gespiegelde wooneenheden bestaat uit metalstud wat in de toekomst relatief eenvoudig te verwijderen is, waardoor er één appartement met een woonoppervlakte van 130 m^2 ontstaat. De twee wooneenheden hebben een gemeenschappelijke meterkast en verwarmingsinstallatie. Eén van de twee wooneenheiden bevat een separaat gastentoilet.
Op de begane grond bevindt zich een technische ruimte met een warmtepomp voor een lage-temperatuurvloerverwarming en een koelsysteem voor het gehele complex.

Algemene projectgegevens
Samenstelling: Huls Bouwkostenmanagement
Opdrachtgever: Woningstichting St. Joseph te Almelo
Architect: 4D Architecten te Enter
Hoofdaannemer: Nijhuis Bouw te Rijssen

Peildatum originele kosten: juli 2011
Start uitvoering: juni 2012
Oplevering: juli 2013
Bouwtijd: 13 maanden
Aanbestedingsvorm: Bouwteam
Risicoregeling van toepassing: n.v.t.

Bruto-inhoud: 6.185 m^3
Alle kosten zijn geïndexeerd.

Kosten- en gegevenstabellen

KOSTEN OP NIVEAU 1 GEHEEL BOUWWERK | BOUWFYSISCHE KENMERKEN | HOEVEELHEDEN | VORMFACTOREN
KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS | KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS, HOOFDINDELING | KOSTEN OP NIVEAU 3 ELEMENTEN
OMSCHRIJVING ELEMENTEN | RUIMTESOORTEN, TOTAAL | RUIMTESOORTEN, SPECIFICATIE

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	1.757.955,00		1.757.955,00
3	Installaties		252.605,00	252.605,00
4	Vaste inrichtingen		75.860,00	75.860,00
5	Terrein	25.540,00		25.540,00

	Totaal directe bouwkosten	1.783.490,-	328.465,-	2.111.955,00
6	Algemene bouwkosten	304.350,00		304.350,00
	Subtotaal	2.087.845,-	328.465,-	2.416.310,-
6	Algemene bedrijfskosten	154.120,00		154.120,00
	Subtotaal	2.241.965,-	328.465,-	2.570.430,-
6	Winst en risico	70.675,00		70.675,00
	Totaal bouwkosten (excl. BTW)	2.312.640,-	328.465,-	2.641.105,00

Bouwfysische kenmerken

Beganegrondvloer	R =	3,00	$m^2.K/W$
Geveiconstructie	R =	7,00	$m^2.K/W$
Glas	U =	1,10	$W/(m^2.K)$
Dakconstructie	R =	7,00	$m^2.K/W$
Energie Prestatie Coëfficiënt			

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	4	stuks	Dakoppervlakte, bruto	656	m^2
Maximale hoogte	14,45	m	Dakoppervlakte	645	m^2
Terrein			Dakopeningen	11	m^2
Terreinoppervlakte	880	m^2	Gevels		
Bebouwde terreinoppervlakte	599	m^2	Geveiomtrek	132	m
Onbebouwde terreinoppervlakte	281	m^2	Buitenwandoppervlakte, bruto	1.528	m^2
Oppervlakteverharding		m^2	Buitenwanden	954	m^2
Aantal parkeerplaatsen		stuks	Buitenwandopeningen	574	m^2
Ruimtedelen			Binnenwanden		
Bruto-vloeroppervlakte (BVO)	1.649	m^2	Binnenwandoppervlakte, bruto	1.495	m^2
Netto-vloeroppervlakte (NVO)	1.344	m^2	Binnenwanden	1.258	m^2
Installatieoppervlakte	39	m^2	Binnenwandopeningen	237	m^2
Verkeersoppervlakte	219	m^2	Trapconstructies		
Nuttige oppervlakte	1.086	m^2	Trappen en hellingen	3	stuks
Niet-toegankelijke ruimte		m^2	Balustrades en leuningen	29	m
Tarra-oppervlakte	305	m^2	Installaties		
Gebruiksoppervlakte	1.364	m^2	Klimaatinstallaties (koeling)		kW
Bruto-inhoud (BI)	6.185	m^3	Klimaatinstallaties (verwarming)		kW
Skelet			Warmtepomp		kW
Buitenwanden (constructief)	1.409	m^2	Zonnecollector		kW
Binnenwanden (constructief)	361	m^2	Vloerverwarming		m^2
Beganegrondvloeren (constructief)	599	m^2	Warmtelichamen		stuks
Verdiepingsvloeren (constructief)	1.050	m^2	Wandverwarming		m^2
Dakvloeren (constructief)	656	m^2	Liften		stuks
Hoofddraagconstructie	1.649	m^2			

Vormfactoren

Woontoren met 40 appartementen (source: bouwkosten.nl)

BOUWkosten

WOONTOREN MET 40 APPARTEMENTEN TE ZWOLLE

TABEL | OMSCHRIJVING | AFBEELDINGEN

Omschrijving

Projectomschrijving
De entree van de wijk Holtenbroek wordt sinds kort gemarkeerd door een stevig, donkergrijs getint blok met koopappartementen. Op de begane grond van het gebouw bevinden zich naast de horecaruimte de hoofdentree van het gebouw en de toegang tot de bergingen. Aan de achterzijde van het gebouw zijn inpandig de lift en het trappenhuis gesitueerd.

De woontoren bevat veertig driekamerappartementen. Met per etage een compacte indeling van vijf appartementen waarvan één op elke hoek en één in het midden tegenover de lift en het trappenhuis. De driekamerappartementen hebben een woonkamer en twee slaapkamers met een oppervlak van 115 tot 118 m^2 BVO. Per etage heeft elk appartement een andere plattegrond. Elk appartement heeft een groot inpandig balkon dat vanuit de woning via de glazen schuifdeuren te bereiken is.

Er is gekozen voor donkergrijze baksteenbekleding en prefab onderdelen omdat de naast het appartementengebouw gelegen seniorenwoningen van vier lagen eveneens deze gevelbekleding hebben. Aan drie zijden van het gebouw zijn, voor elk appartement afzonderlijk en evenwijdig aan de gevel, verschuifbare schermen met oranjegekleurde, zonwerende doeken van circa 1,50 m breed geplaatst. Sommige zijn geplaatst voor de balkons, andere voor de ramen van de woonkamers. Omdat de schermen door de bewoners afzonderlijk te bedienen zijn, krijgt de gevel een variërende aanblik.

Algemene projectgegevens
Samenstelling: Huls Bouwkostenmanagement
Opdrachtgever: Projectteam Holtenbroek 1 te Zwolle
Architect: de Zwarte Hond te Rotterdam
Hoofdaannemer: Moes Bouwgroep te Zwolle

Peildatum originele kosten: februari 2006
Start uitvoering: april 2007
Oplevering: november 2007
Bouwtijd: 225 werkbare werkdagen
Aanbestedingsvorm: op uitnodiging (4)
Risicoregeling van toepassing: n.v.t.

Bruto-inhoud: 17.057 m^3
Alle kosten zijn geïndexeerd.

Kosten- en gegevenstabellen

KOSTEN OP NIVEAU 1 GEHEEL BOUWWERK | BOUWFYSISCHE KENMERKEN | HOEVEELHEDEN | VORMFACTOREN
KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS | KOSTEN OP NIVEAU 2 ELEMENTENCLUSTERS, HOOFDINDELING | KOSTEN OP NIVEAU 3 ELEMENTEN
OMSCHRIJVING ELEMENTEN | RUIMTESOORTEN, TOTAAL | RUIMTESOORTEN, SPECIFICATIE

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	3.398.160,00		3.398.160,00
3	Installaties		644.550,00	644.550,00
4	Vaste inrichtingen		72.020,00	72.020,00
5	Terrein	27.375,00		27.375,00

	Totaal directe bouwkosten	3.425.535,-	716.565,-	4.142.100,00
6	Algemene bouwkosten	462.375,00		462.375,00
	Subtotaal	3.887.910,-	716.565,-	4.604.475,-
6	Algemene bedrijfskosten	319.930,00		319.930,00
	Subtotaal	4.207.840,-	716.565,-	4.924.405,-
6	Winst en risico			
	Totaal bouwkosten (excl. BTW)	4.207.840,-	716.565,-	4.924.405,00

Bouwfysische kenmerken

Beganegrondvloer	R =	2,50	$m^2.K/W$
Gevelconstructie	R =	3,00	$m^2.K/W$
Glas	U =	1,20	$W/(m^2.K)$
Dakconstructie	R =	3,00	$m^2.K/W$
Energie Prestatie Coëfficiënt		1,00	

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	9	stuks	Dakoppervlakte, bruto	713	m^2
Maximale hoogte	27,50	m	Dakoppervlakte	712	m^2
Terrein			Dakopeningen	1	m^2
Terreinoppervlakte	1.808	m^2	Gevels		
Bebouwde terreinoppervlakte	673	m^2	Gevelomtrek	121	m
Onbebouwde terreinoppervlakte	1.134	m^2	Buitenwandoppervlakte, bruto	3.603	m^2
Oppervlakteverharding	1.121	m^2	Buitenwanden	1.999	m^2
Aantal parkeerplaatsen	49	stuks	Buitenwandopeningen	1.604	m^2
Ruimtedelen			Binnenwanden		
Bruto-vloeroppervlakte (BVO)	5.680	m^2	Binnenwandoppervlakte, bruto	6.304	m^2
Netto-vloeroppervlakte (NVO)	4.568	m^2	Binnenwanden	5.115	m^2
Installatieoppervlakte	30	m^2	Binnenwandopeningen	1.189	m^2
Verkeersoppervlakte	1.240	m^2	Trapconstructies		
Nuttige oppervlakte	3.294	m^2	Trappen en hellingen	8	stuks
Niet-toegankelijke ruimte	4	m^2	Balustrades en leuningen	813	m
Tarra-oppervlakte	1.112	m^2	Installaties		
Gebruiksoppervlakte	4.090	m^2	Klimaatinstallaties (koeling)		kW
Bruto-inhoud (BI)	17.057	m^3	Klimaatinstallaties (verwarming)		kW
Skelet			Warmtepomp		kW
Buitenwanden (constructief)	475	m^2	Zonnecollector		kW
Binnenwanden (constructief)	3.046	m^2	Vloerverwarming		m^2
Beganegrondvloeren (constructief)	673	m^2	Warmtelichamen		stuks
Verdiepingsvloeren (constructief)	5.007	m^2	Wandverwarming		m^2
Dakvloeren (constructief)	713	m^2	Liften	1	stuk
Hoofddraagconstructie	5.680	m^2			

Vormfactoren

Totaaloverzicht nieuwbouw (Kulmann, 2007)

Totaaloverzicht nieuwbouw; oorspronkelijke peildatum

Woongebouwen

Nieuwbouw - Woongebouwen in het algemeen

SfB-gebouw-code	Project	Aantal		Peildatum		Kosten	
		m^2 BVO	m^3 BI			Peildatum	Geïndexeerd
81	WONINGEN IN HET ALGEMEEN						
	17 appartementen met winkelruimte te Oudenbosch	1.608	4.838	Apr. 2002	Totaal	1.359.660,-	1.515.415,-
					Per m^2 BVO	846,-	942,-
					Per m^3 BI	281,-	313,-
	49 seniorenappartementen te Dronten	6.723	18.587	Apr. 2002	Totaal	3.677.005,-	4.098.220,-
					Per m^2 BVO	547,-	610,-
					Per m^3 BI	198,-	220,-
	15 stadswoningen te Maastricht	2.567	8.110	Okt. 2002	Totaal	2.039.045,-	2.243.440,-
					Per m^2 BVO	794,-	874,-
					Per m^3 BI	251,-	277,-
	Vrijstaande woning te Zundert	174	632	Aug. 2003	Totaal	259.170,-	277.025,-
					Per m^2 BVO	1.490,-	1.590,-
					Per m^3 BI	410,-	438,-
	Vrijstaande woning te Delfgauw	294	705	Nov. 2002	Totaal	307.835,-	338.690,-
					Per m^2 BVO	1.045,-	1.150,-
					Per m^3 BI	437,-	480,-
	Vrijstaande woning te Zaltbommel	223	646	Mei 2003	Totaal	214.810,-	234.385,-
					Per m^2 BVO	963,-	1.050,-
					Per m^3 BI	333,-	363,-
	Appartementen en winkels te Winterswijk	5.661	19.910	Mei 2003	Totaal	3.562.415,-	3.841.300,-
					Per m^2 BVO	629,-	679,-
					Per m^3 BI	179,-	193,-
	Vrijstaande woning te Schoonrewoerd	286	750	Mrt. 2005	Totaal	272.570,-	286.605,-
					Per m^2 BVO	953,-	1.000,-
					Per m^3 BI	363,-	382,-
84	BIJZONDERE WOONGEBOUWEN						
	Studentenhuisvesting te Dronten	3.580	10.728	Dec. 2003	Totaal	1.994.210,-	2.131.785,-
					Per m^2 BVO	557,-	595,-
					Per m^3 BI	186,-	199,-

17 appartementen Oudenbosch (source: Kuhlmann, 2007)

denbosch

appartementen in drie
bergingen en een aan-

winkelcentrum, waarin
n in 2 blokken naast el-
kort blok van 2 woon-
kelruimte aanwezig,
mgegeven wordt.
iche en toilet, apart
slaapkamer en hobby-
voor effectieve wo-
De badkamer is alleen
komen niet uit op een
eruggelegd waardoor

r	Overige aannemers	Totaal bouwkosten
		1.023.440,-
	213.005,-	213.005,-
	49.280,-	49.280,-
		42.255,-
	262.285,-	1.327.980,-
		46.570,-
	262.285,-	1.374.550,-
		54.985,-
	262.285,-	1.429.635,-
		85.780,-
	262.285,-	1.515.415,-

81 Woningen in het algemeen

17 appartementen met winkelruimte te Oudenbosch

Bouwfysische kenmerken

Begane grondvloer	R =	3,50	m^2.K/W
Gevelconstructie	R =	3,00	m^2.K/W
Glas	U =	1,80	W/(m^2.K)
Dakconstructie	R =	3,50	m^2.K/W
Energie Prestatie Coëfficiënt			

Hoeveelheden

Algemeen		
Aantal bouwlagen	3	stuks
Terrein		
Terreinoppervlakte	3.456	m^2
Bebouwde oppervlakte	585	m^2
Onbebouwde oppervlakte	2.871	m^2
Oppervlakteverharding	2.304	m^2
Aantal parkeerplaatsen	65	stuks
Ruimtedelen		
Bruto-vloeroppervlakte (BVO)	1.608	m^2
Netto-vloeroppervlakte (NVO)	1.359	m^2
Installatieoppervlakte	17	m^2
Verkeersoppervlakte	353	m^2
Nuttige oppervlakte	989	m^2
Niet-toegankelijke ruimte		m^2
Tara-oppervlakte	249	m^2
Gebruiksoppervlakte	1.308	m^2
Bruto-inhoud	4.838	m^3
Skelet		
Buitenwanden (constructief)	916	m^2
Binnenwanden (constructief)	1.246	m^2
Begane grondvloeren (constructief)	487	m^2
Verdiepingsvloeren (constructief)	976	m^2
Dakvloeren (constructief)	554	m^2
Hoofddraagconstructie	1.608	m^2

Daken		
Dakoppervlakte, bruto	755	m^2
Dakoppervlakte	637	m^2
Dakopeningen	117	m^2
Gevels		
Gevelomtrek	232	m
Buitenwandoppervlakte, bruto	1.389	m^2
Buitenwanden	1.009	m^2
Buitenwandopeningen	380	m^2
Binnenwanden		
Binnenwandoppervlakte, bruto	2.121	m^2
Binnenwanden	1.790	m^2
Binnenwandopeningen	331	m^2
Trapconstructies		
Trappen en hellingen	4	stuks
Balustrades en leuningen	142	m
Installaties		
Klimaatinstallaties (koeling)		kW
Klimaatinstallaties (verwarming)		kW
Warmtepomp		kW
Zonnecollector		kW
Vloerverwarming		m^2
Warmtelichamen		stuks
Wandverwarming		m^2
Liften	1	stuk

Vormfactoren

Bruto-inhoud/ bruto-vloeroppervlakte (BVO)	=	4.838	/	1.608	=	3,01
Netto-vloeroppervlakte (NVO)/ bruto-vloeroppervlakte (BVO)	=	1.359	/	1.608	=	0,84
Tara-oppervlakte/ bruto-vloeroppervlakte (BVO)	=	249	/	1.608	=	0,16
Gebruiksoppervlakte/ bruto-vloeroppervlakte (BVO)	=	1.308	/	1.608	=	0,81
Nuttige oppervlakte/ bruto-vloeroppervlakte (BVO)	=	989	/	1.608	=	0,62
Bebouwde oppervlakte/ bruto-vloeroppervlakte (BVO)	=	585	/	1.608	=	0,36
Buitenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	1.389	/	1.608	=	0,86
Dakoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	755	/	1.608	=	0,47
Buitenwand- en dakoppervlak/ bruto-vloeroppervlakte (BVO)	=	2.144	/	1.608	=	1,33
Binnenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	2.121	/	1.608	=	1,32
Bebouwde oppervlakte/ nuttige oppervlakte	=	585	/	989	=	0,59
Bruto-vloeroppervlakte (BVO)/ nuttige oppervlakte	=	1.608	/	989	=	1,63
Buitenwandoppervlakte, bruto/ nuttige oppervlakte	=	1.389	/	989	=	1,40
Dakoppervlakte, bruto/ nuttige oppervlakte	=	755	/	989	=	0,76
Buitenwand- en dakoppervlak/ nuttige oppervlakte	=	2.144	/	989	=	2,17
Binnenwandoppervlakte, bruto/ nuttige oppervlakte	=	2.121	/	989	=	2,14
Bebouwde oppervlakte/ gebruiksoppervlakte	=	585	/	1.308	=	0,45
Bruto-vloeroppervlakte (BVO)/ gebruiksoppervlakte	=	1.608	/	1.308	=	1,23
Buitenwandoppervlakte, bruto/ gebruiksoppervlakte	=	1.389	/	1.308	=	1,06
Dakoppervlakte, bruto/ gebruiksoppervlakte	=	755	/	1.308	=	0,58
Buitenwand- en dakoppervlak/ gebruiksoppervlakte	=	2.144	/	1.308	=	1,64
Binnenwandoppervlakte, bruto/ gebruiksoppervlakte	=	2.121	/	1.308	=	1,62

81

49 seniorenappartementen Dronten (source: Kuhlmann, 2007)

81 Woningen in het algemeen

49 seniorenappartementen te Dronten

Projectomschrijving
Het appartementencomplex 'Het Tussendek' staat in Dronten op de grens van het centrum, winkelgebied en de nabij gelegen woonwijken. Het complex heeft ook voorzieningen zoals een gezondheidscentrum, de bibliotheek, het cultureel centrum De Meerpaal en het openbaar vervoer in de onmiddellijke omgeving. Het is dan ook bij uitstek geschikt voor senioren en daarom ook mede toegankelijk voor personen die een scootmobiel gebruiken.
Het complex bestaat uit drie woonvleugels die een gemeenschappelijke tuin omsluiten. Op de begane grond bevinden zich de hoofdentree, een recreatieruimte, één berging voor ieder appartement en een aantal appartementen.
Over de eerste tot en met de derde verdieping zijn de resterende appartementen verdeeld en per verdieping is er een bergingsruimte voor scootmobielen aanwezig. De op de omgeving uitkijkende appartementen hebben middels een loggia hun eigen omsloten buitenruimte.
De entrees van de woningen zijn te bereiken via galerijen, gelegen aan de omsloten binnenruimte met zicht op de gemeenschappelijke tuin. Verder completeren voorzieningen zoals videofoon, lift en laagdrempeligheid (maximaal 20 mm) het geheel. De parkeerplaatsen met de plaatsen voor mindervaliden zijn direct bij de hoofdingang aanwezig.
Door zijn verfrissende architectuur, uitgevoerd in een combinatie van metselwerk, stucwerk en verschillende staalbeplatingen, vormt het complex een passende aanvulling in de vernieuwing van het centrumgebied.

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	2.644.705,-		2.644.705,-
3	Installaties		721.810,-	721.810,-
4	Vaste inrichtingen		124.790,-	124.790,-
5	Terrein	45.875,-		45.875,-
	Totaal bouwkosten elementen	2.690.580,-	846.595,-	3.537.175,-
6	Algemene bouwkosten (bouwproject)	232.205,-		232.205,-
	Subtotaal	2.922.785,-	846.595,-	3.769.380,-
6	Algemene bedrijfskosten (bouwbedrijf)	328.840,-		328.840,-
	Subtotaal	3.251.625,-	846.595,-	4.098.220,-
6	Winst en risico (bouwbedrijf)			
	Totaal bouwkosten exclusief BTW	3.251.625,-	846.595,-	4.098.220,-

nanagement

oondiensten te

Seckel v.d. Dob-
ten te Hengelo

ddrijf BV te

ele kosten

rkdagen

:587 m³
peindexeerc.

81 Woningen in het algemeen

49 seniorenappartementen te Dronten

Bouwfysische kenmerken

Begaanegrondvloer	R =	4,00	m².K/W
Gevelconstructie	R =	3,05	m².K/W
Glas	U =	1,60	W/(m².K)
Dakconstructie	R =	4,50	m².K/W
Energie Prestatie Coëfficiënt		0,94	

Hoeveelheden

Algemeen			**Daken**		
Aantal bouwlagen	4	stuks	Dakoppervlakte, bruto	1.446	m²
Terrein			Dakoppervlakte	1.445	m²
Terreinoppervlakte	3.812	m²	Dakopeningen	1	m²
Bebouwde oppervlakte	1.581	m²	**Gevels**		
Overbouwde oppervlakte	2.231	m²	Gevelomtrek	284	m
Oppervlakteverharding	526	m²	Buitenwandoppervlakte, bruto	3.425	m²
Aantal parkeerplaatsen	2	stuks	Buitenwanden	2.203	m²
Ruimtelijks			Buitenwandopeningen	1.222	m²
Bruto-vloeroppervlakte (BVO)	6.723	m²	Binnenwanden		
Netto-vloeroppervlakte (NVO)	5.941	m²	Binnenwandoppervlakte, bruto	4.919	m²
Installatieoppervlakte	82	m²	Binnenwanden	3.715	m²
Verkeersoppervlakte	1.444	m²	Binnenwandopeningen	1.204	m²
Nuttige oppervlakte	4.415	m²	Trapconstructies		
Niet-toegankelijke ruimte		m²	Trappen en hellingen	3	stuks
Tuin-oppervlakte	782	m²	Balustrades en leuningen	386	m
Gebruiksoppervlakte	5.313	m²	**Installaties**		
Bruto-inhoud	18.587	m³	Klimaatinstallaties (koeling)		kW
Skelet			Klimaatinstallaties (verwarming)		kW
Buitenwanden (constructief)	563	m²	Warmtepomp		kW
Binnenwanden (constructief)	2.209	m²	Zonnecollector		kW
Begaanegrondvloeren (constructief)	1.463	m²	Vloerverwarming		m²
Verdiepingvloeren (constructief)	4.785	m²	Warmtelichamen		stuks
Dakvloeren (constructief)	1.231	m²	Wandverwarming		m²
Draagconstructie	6.723	m²	Liften	1	stuk

Vormfactoren

Bruto-inhoud/ bruto-vloeroppervlakte (BVO)	=	18.587	/	6.723	=	2,76
Netto-vloeroppervlakte (NVO)/ bruto-vloeroppervlakte (BVO)	=	5.941	/	6.723	=	0,88
Tuin-oppervlakte/ bruto-vloeroppervlakte (BVO)	=	782	/	6.723	=	0,12
Gebruiksoppervlakte/ bruto-vloeroppervlakte (BVO)	=	5.313	/	6.723	=	0,79
Nuttige oppervlakte/ bruto-vloeroppervlakte (BVO)	=	4.415	/	6.723	=	0,66
Bebouwde oppervlakte/ bruto-vloeroppervlakte (BVO)	=	1.581	/	6.723	=	0,24
Buitenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	3.425	/	6.723	=	0,51
Dakoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	1.446	/	6.723	=	0,22
Buitenwand- en dakoppervlak/ bruto-vloeroppervlakte (BVO)	=	4.871	/	6.723	=	0,72
Binnenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	4.919	/	6.723	=	0,73
Bebouwde oppervlakte/ nuttige oppervlakte	=	1.581	/	4.415	=	0,36
Bruto-vloeroppervlakte (BVO)/ nuttige oppervlakte	=	6.723	/	4.415	=	1,52
Buitenwandoppervlakte, bruto/ nuttige oppervlakte	=	3.425	/	4.415	=	0,78
Dakoppervlakte, bruto/ nuttige oppervlakte	=	1.446	/	4.415	=	0,33
Buitenwand- en dakoppervlak/ nuttige oppervlakte	=	4.871	/	4.415	=	1,10
Binnenwandoppervlakte, bruto/ nuttige oppervlakte	=	4.919	/	4.415	=	1,11
Bebouwde oppervlakte/ gebruiksoppervlakte	=	1.581	/	5.313	=	0,30
Bruto-vloeroppervlakte (BVO)/ gebruiksoppervlakte	=	6.723	/	5.313	=	1,27
Buitenwandoppervlakte, bruto/ gebruiksoppervlakte	=	3.425	/	5.313	=	0,64
Dakoppervlakte, bruto/ gebruiksoppervlakte	=	1.446	/	5.313	=	0,27
Buitenwand- en dakoppervlak/ gebruiksoppervlakte	=	4.871	/	5.313	=	0,92
Binnenwandoppervlakte, bruto/ gebruiksoppervlakte	=	4.919	/	5.313	=	0,93

15 stadwoningen Maastricht (source: Kuhlmann, 2007)

Projectomschrijving
Het project betreft de nieuwbouw van 15 stadswoningen in 3 lagen. De woningen maken onderdeel uit van de herontwikkeling van een voormalig bedrijventerrein in de stad. De wijk heet nu Alegonda park en bestaat uit 107 woningen.
In de woningen liggen op de begane grond de entree, keuken, woonkamer en een erker die doorloopt naar de 1e etage. Op de 1e etage is een woon-/werkkamer, met vide en een slaapkamer. Op de bovenste halve verdieping zijn 2 slaapkamers.
Er is een uitgebreid optiepakket waarmee de woningen naar wens ingedeeld kunnen worden. Zo kan de 1e etage een doorlopende woonkamer worden. In de pui aan de achtergevel is een tussenstuk aangebracht waartegen een tussenmuur geplaatst kan worden. Er zijn daar 2 radiatoren geplaatst en 2 draaikiepramen, zodat beide kamers - als die optie wordt gekozen - gelucht kunnen worden. In alle opties zit de trap in dezelfde sparing. Daar zijn ook de leidingschachten aangebracht. De bovenste halve verdieping kan worden uitgebreid.

Samenstelling
Winket voor de bouw te Roosendaal
Opdrachtgever
Janssen de Jong Bouw Plancoördinatie bv te Son en Breugel
Architect
Architectenbureau Humlé Neuhof te Maastricht
Hoofdaannemer
Janssen de Jong Bouw te Kerkrade

Peildatum originele kosten
Oktober 2002
Start uitvoering
Januari 2003
Oplevering
Februari 2004
Bouwtijd
220 dagen

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	1.445.030,-		1.445.030,-
3	Installaties		224.145,-	224.145,-
4	Vaste inrichtingen		69.490,-	69.490,-
5	Terrein	27.440,-		27.440,-
	Totaal bouwkosten elementen	1.472.470,-	293.635,-	1.766.105,-
6	Algemene bouwkosten (bouwproject)	224.300,-		224.300,-
	Subtotaal	1.696.770,-	293.635,-	1.990.405,-
6	Algemene bedrijfskosten (bouwbedrijf)	144.945,-		144.945,-
	Subtotaal	1.841.720,-	293.635,-	2.135.350,-
6	Winst en risico (bouwbedrijf)	108.090,-		108.090,-
	Totaal bouwkosten exclusief BTW	1.949.805,-	293.635,-	2.243.440,-

81

81 Woningen in het algemeen

15 stadswoningen te Maastricht

Bouwfysische kenmerken

Beganegrondvloer	R =	3,00	$m^2.K/W$
Gevelconstructie	R =	3,00	$m^2.K/W$
Glas	U =	1,60	$W/(m^2.K)$
Dakconstructie	R =	3,00	$m^2.K/W$
Energie Prestatie Coëfficiënt			

Hoeveelheden

Algemeen			Daken		
Aantal bouwlagen	3	stuks	Dakoppervlakte, bruto	1.467	m^2
Terrein			Dakoppervlakte	1.467	m^2
Terreinoppervlakte	2.313	m^2	Dakopeningen		m^2
Bebouwde oppervlakte	1.211	m^2	Gevels		
Onbebouwde oppervlakte	1.102	m^2	Gevelomtrek	139	m
Oppervlakteverharding	200	m^2	Buitenwandoppervlakte, bruto	2.096	m^2
Aantal parkeerplaatsen	15	stuks	Buitenwanden	1.450	m^2
Ruimtedelen			Buitenwandopeningen	646	m^2
Bruto-vloeroppervlakte (BVO)	2.567	m^2	Binnenwanden		
Netto-vloeroppervlakte (NVO)	1.922	m^2	Binnenwandoppervlakte, bruto	3.081	m^2
Installatieoppervlakte	16	m^2	Binnenwanden	2.682	m^2
Verkeersoppervlakte	302	m^2	Binnenwandopeningen	399	m^2
Nuttige oppervlakte	1.604	m^2	Trapconstructies		
Niet-toegankelijke ruimte		m^2	Trappen en hellingen	31	stuks
Tarra-oppervlakte	645	m^2	Balustrades en leuningen	261	m
Gebruiksoppervlakte	2.163	m^2	Installaties		
Bruto-inhoud	8.110	m^3	Klimaatinstallaties (koeling)		kW
Skelet			Klimaatinstallaties (verwarming)		kW
Buitenwanden (constructief)	1.338	m^2	Warmtepomp		kW
Binnenwanden (constructief)	1.785	m^2	Zonnecollector		kW
Beganegrondvloeren (constructief)	981	m^2	Vloerverwarming		m^2
Verdiepingsvloeren (constructief)	1.412	m^2	Warmtelichamen		stuks
Dakvloeren (constructief)	1.393	m^2	Wandverwarming		m^2
Hoofddraagconstructie	2.567	m^2	Liften		stuks

Vormfactoren

81

Bruto-inhoud/ bruto-vloeroppervlakte (BVO)	=	8.110	/	2.567	=	3,16
Netto-vloeroppervlakte (NVO)/ bruto-vloeroppervlakte (BVO)	=	1.922	/	2.567	=	0,75
Tarra-oppervlakte/ bruto-vloeroppervlakte (BVO)	=	645	/	2.567	=	0,25
Gebruiksoppervlakte/ bruto-vloeroppervlakte (BVO)	=	2.163	/	2.567	=	0,84
Nuttige oppervlakte/ bruto-vloeroppervlakte (BVO)	=	1.604	/	2.567	=	0,62
Bebouwde oppervlakte/ bruto-vloeroppervlakte (BVO)	=	1.211	/	2.567	=	0,47
Buitenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	2.096	/	2.567	=	0,82
Dakoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	1.467	/	2.567	=	0,57
Buitenwand- en dakoppervlak/ bruto-vloeroppervlakte (BVO)	=	3.563	/	2.567	=	1,39
[illegible]	=	3.081	/	2.567	=	1,20

Appartementen Winterswijk (source: Kuhlmann, 2007)

81

Samenstelling
Huls Bouwkostenmanagement te Zwolle
Opdrachtgever
De Woonplaats
Woningcorporatie te Groenlo
Architect
Architektenburo Ghijsen Partners B.V. te Deventer
Hoofdaannemer
Bouwcombinatie WAM en Van Duren / Thomasson Dura V.O.F. te Winterswijk

Peildatum originele kosten
Mei 2003

81 Woningen in het algemeen

Appartementen en winkels te Winterswijk

Projectomschrijving
Een voormalig terrein van Gemeentewerken aan de Spoorstraat Winterwijk heeft een nieuwe bestemming gekregen. Er is plaatsgemaakt voor een supermarkt van 1.490 m^2, een aantal winkels en daarboven twee lagen met koop- en huurappartementen en zelfstandige woningen voor bewoners met een zorgindicatie.
Naast dit gebouw is een parkeerterrein aangelegd. De locatie is gedeeltelijk afgezoomd met een muur die de achtertuinen van de omliggende woningen afschermt van dit parkeerterrein. De winkels op de begane grond worden casco opgeleverd. De constructie bestaat in hoofdzaak uit betonkolommen, betonschijven en enkele betonwanden op de begane grond met betonbalken onder de eerste verdiepingsvloer. Op de verdieping woningscheidende betonwanden. De beganegrondvloer is een kanaalplaatvloer. De verdiepings- en dakvloeren zijn bekistingsplaatvloeren. Op de tweede verdieping prefab betonnen galerijplaten. Op de eerste verdieping worden de woningen bereikt door looppaden op het dak. Het middengebied van het dak op de eerste verdieping wordt ingericht als daktuin.

Kosten op niveau 1 Geheel bouwwerk

NEN 2634	Omschrijving	Hoofdaannemer	Overige aannemers	Totaal bouwkosten
2	Bouwkundige werken	2.445.320,-		2.445.320,-
3	Installaties		432.195,-	432.195,-
4	Vaste inrichtingen		65.075,-	65.075,-
5	Terrein	56.635,-		56.635,-
	Totaal bouwkosten elementen	2.501.955,-	497.270,-	2.999.225,-
6	Algemene bouwkosten (bouwproject)	464.025,-		464.025,-
	Subtotaal	2.965.980,-	497.270,-	3.463.250,-
6	Algemene bedrijfskosten (bouwbedrijf)	248.170,-		248.170,-
	Subtotaal	3.214.150,-	497.270,-	3.711.420,-
6	Winst en risico (bouwbedrijf)	129.880,-		129.880,-
	Totaal bouwkosten exclusief BTW	3.344.030,-	497.270,-	3.841.300,-

81 Woningen in het algemeen

Appartementen en winkels te Winterswijk

Bouwfysische kenmerken

Beganegrondvloer	R =	2,50	m^2.K/W
Gevelconstructie	R =	3,00	m^2.K/W
Glas	U =	1,20	W/(m^2.K)
Dakconstructie	R =	3,00	m^2.K/W
Energie Prestatie Coëfficiënt		0,75 / 2,15	

Hoeveelheden

Algemeen		
Aantal bouwlagen	3	stuks
Terrein		
Terreinoppervlakte	7.260	m^2
Bebouwde oppervlakte	2.451	m^2
Onbebouwde oppervlakte	4.809	m^2
Oppervlakteverharding		m^2
Aantal parkeerplaatsen		stuks
Ruimtedelen		
Bruto-vloeroppervlakte (BVO)	5.661	m^2
Netto-vloeroppervlakte (NVO)	4.820	m^2
Installatieoppervlakte	52	m^2
Verkeersoppervlakte	603	m^2
Nuttige oppervlakte	4.165	m^2
Niet-toegankelijke ruimte		m^2
Terra-oppervlakte	841	m^2
Gebruiksoppervlakte	4.735	m^2
Bruto-inhoud	19.910	m^3
Skelet		
Buitenwanden (constructief)		m^2
Binnenwanden (constructief)	1.233	m^2
Beganegrondvloeren (constructief)	2.451	m^2
Verdiepingsvloeren (constructief)	3.210	m^2
Dakvloeren (constructief)	2.451	m^2
Hoofddragconstructie	5.661	m^2

Daken		
Dakoppervlakte, bruto	2.451	m^2
Dakoppervlakte	2.440	m^2
Dakopeningen	11	m^2
Gevels		
Gevelomtrek	206	m
Buitenwandoppervlakte, bruto	2.554	m^2
Buitenwanden	1.938	m^2
Buitenwandopeningen	616	m^2
Binnenwanden		
Binnenwandoppervlakte, bruto	2.521	m^2
Binnenwanden	1.672	m^2
Binnenwandopeningen	849	m^2
Trapconstructies		
Trappen en hellingen	11	stuks
Balustrades en leuningen	168	m
Installaties		
Klimaatinstallaties (koeling)		kW
Klimaatinstallaties (verwarming)		kW
Warmtepomp		kW
Zonnecollector		kW
Vloerverwarming	1.949	m^2
Warmtelichamen		stuks
Wandverwarming		m^2
Liften	2	stuks

Vormfactoren

Bruto-inhoud/ bruto-vloeroppervlakte (BVO)	=	19.910	/	5.661	= 3,52
Netto-vloeroppervlakte (NVO)/ bruto-vloeroppervlakte (BVO)	=	4.820	/	5.661	= 0,85
Terra-oppervlakte/ bruto-vloeroppervlakte (BVO)	=	841	/	5.661	= 0,15
Gebruiksoppervlakte/ bruto-vloeroppervlakte (BVO)	=	4.735	/	5.661	= 0,84
Nuttige oppervlakte/ bruto-vloeroppervlakte (BVO)	=	4.165	/	5.661	= 0,74
Bebouwde oppervlakte/ bruto-vloeroppervlakte (BVO)	=	2.451	/	5.661	= 0,43
Buitenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	2.554	/	5.661	= 0,45
Dakoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	2.451	/	5.661	= 0,43
Buitenwand- en dakoppervlak/ bruto-vloeroppervlakte (BVO)	=	5.005	/	5.661	= 0,88
Binnenwandoppervlakte, bruto/ bruto-vloeroppervlakte (BVO)	=	2.521	/	5.661	= 0,45
Bebouwde oppervlakte/ nuttige oppervlakte	=	2.451	/	4.165	= 0,59
Bruto-vloeroppervlakte (BVO)/					

Appendix 5: Price index

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2002	89.47	89.64	89.81	89.98	90.43	90.87	91.32	91.33	91.34	91.35	90.69	90.02
2003	89.36	89.67	89.99	90.3	90.37	90.45	90.52	90.58	90.64	90.7	90.9	91.11
2004	91.31	91.29	91.28	91.26	91.58	91.91	92.23	92.49	92.74	93	93.18	93.37
2005	93.55	93.6	93.66	93.71	93.81	93.91	94.01	94.15	94.28	94.42	94.55	94.68
2006	94.81	95	95.19	95.38	95.65	95.91	96.18	96.67	97.16	97.65	98.43	99.22
2007	100	100.22	100.43	100.65	101.01	101.37	101.73	102.01	102.29	102.57	102.81	103.04
2008	103.28	103.99	104.71	105.42	105.61	105.79	105.98	105.76	105.53	105.31	104.88	104.45
2009	104.02	104.01	104.01	104	103.56	103.13	102.69	102.69	102.7	102.7	102.16	101.63
2010	101.09	101.06	101.04	101.01	100.95	100.9	100.84	100.66	100.47	100.29	100.06	99.82
2011	99.59	99.89	100.2	100.5	100.61	100.71	100.82	100.42	100.02	99.62	99.71	99.8
2012	99.89	99.98	100.07	100.16	100.18	100.19	100.21	100.24	100.26	100.29	100.34	100.38
2013	100.43	100.55	100.68	100.8	100.8	100.8	100.8	101.05	101.31	101.56	101.89	102.21
2014	102.54	102.5	102.45	102.41	102.52	102.64	102.75	102.89	103.03	103.17	103.52	103.86
2015	104.21	104.31	104.42	104.52	104.57	104.63	104.68	104.8	104.93	105.05	105.28	105.51
2016	105.74	105.93	106.12	106.31	106.38	106.45	106.52	106.63	106.75	106.86	107.02	107.19
2017	107.35											

Figure 12: Price index residential buildings (source: Vonk, Wilde & Groot, 2016)

Appendix 6: Graphical representation performance measures database A

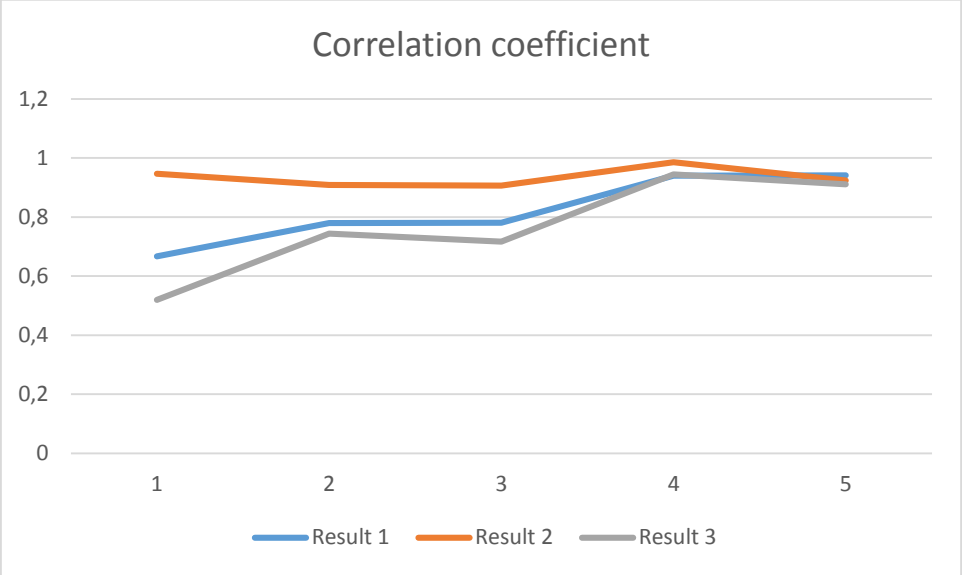

Figure 13: Correlation coefficient of database A

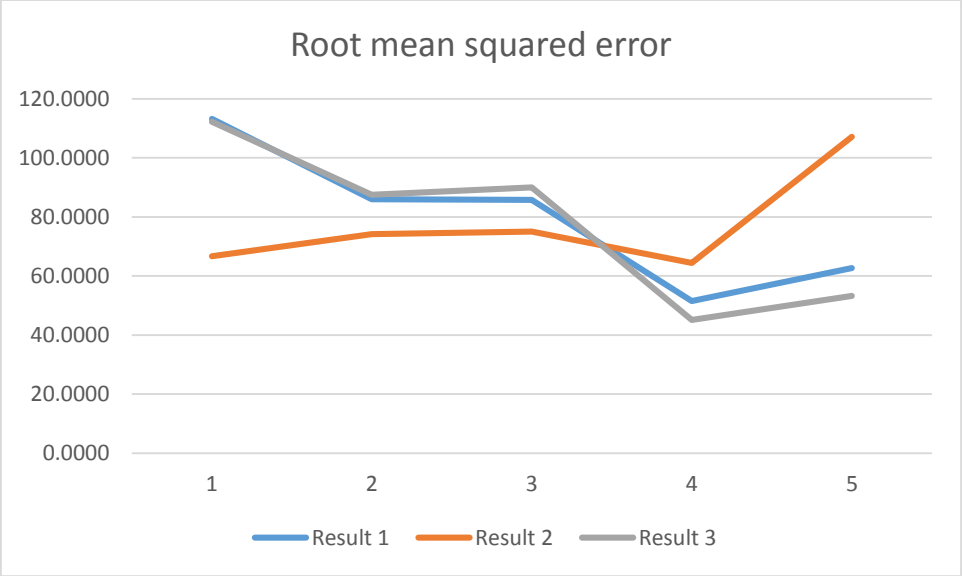

Figure 14: Root mean squared error of database A

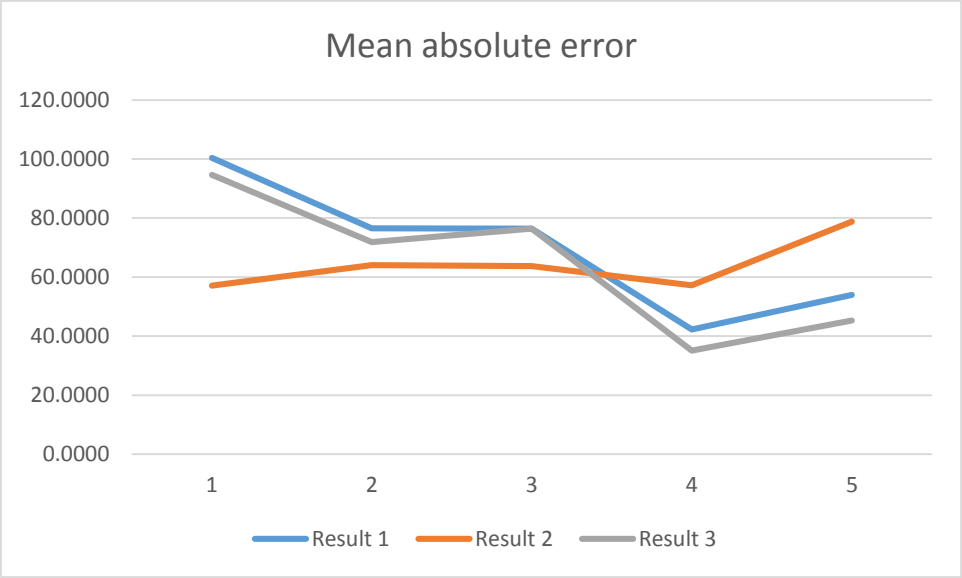

Figure 15: Mean absolute error of database A

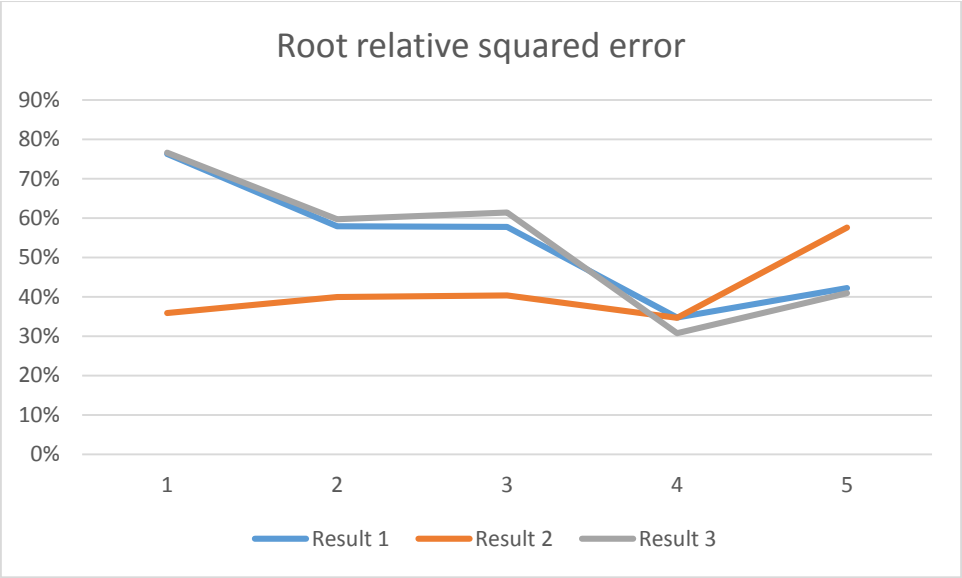

Figure 16: Root relative squared error of database A

Appendix 7: Graphical representation performance measures database B

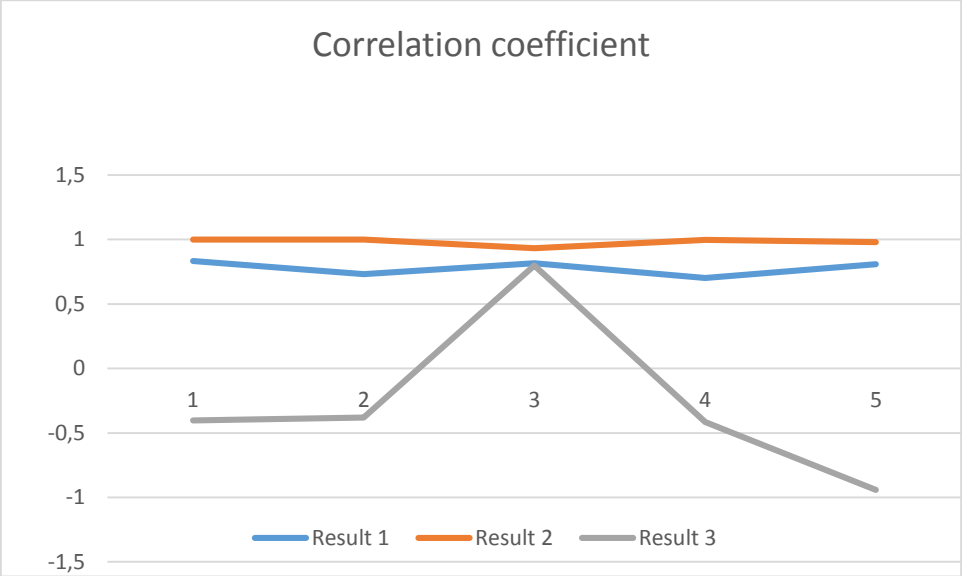

Figure 17: Correlation coefficient of database B

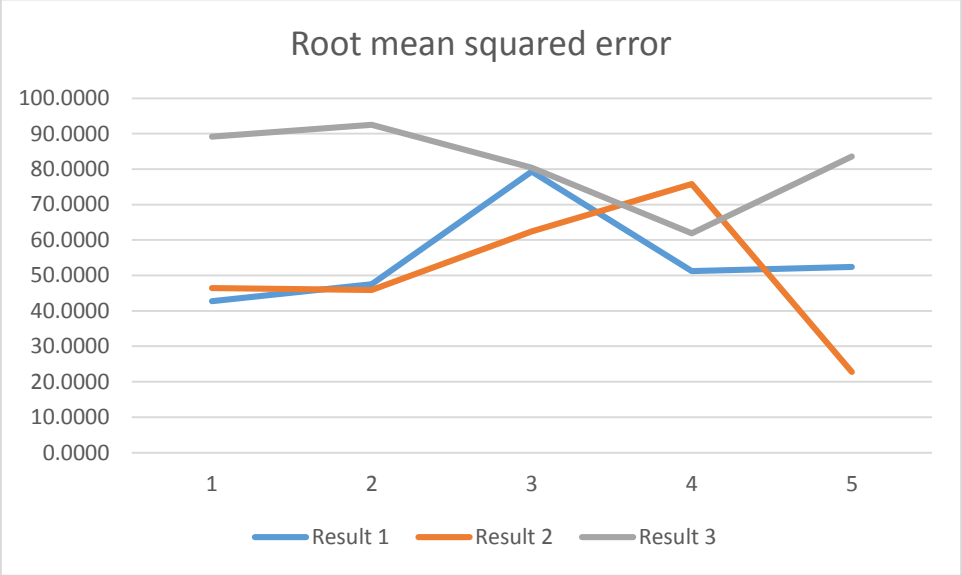

Figure 18: Root mean squared error of database B

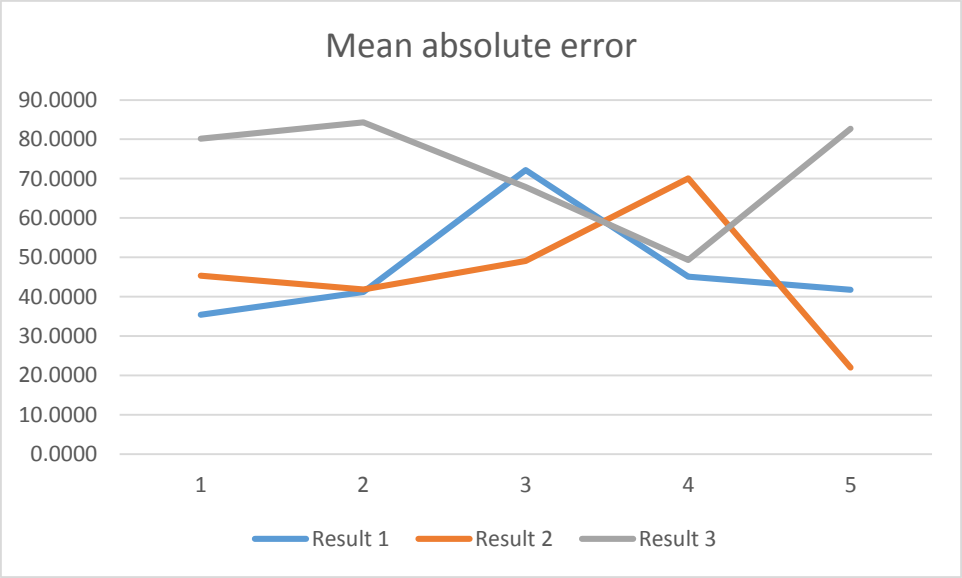

Figure 19: Mean absolute error of database B

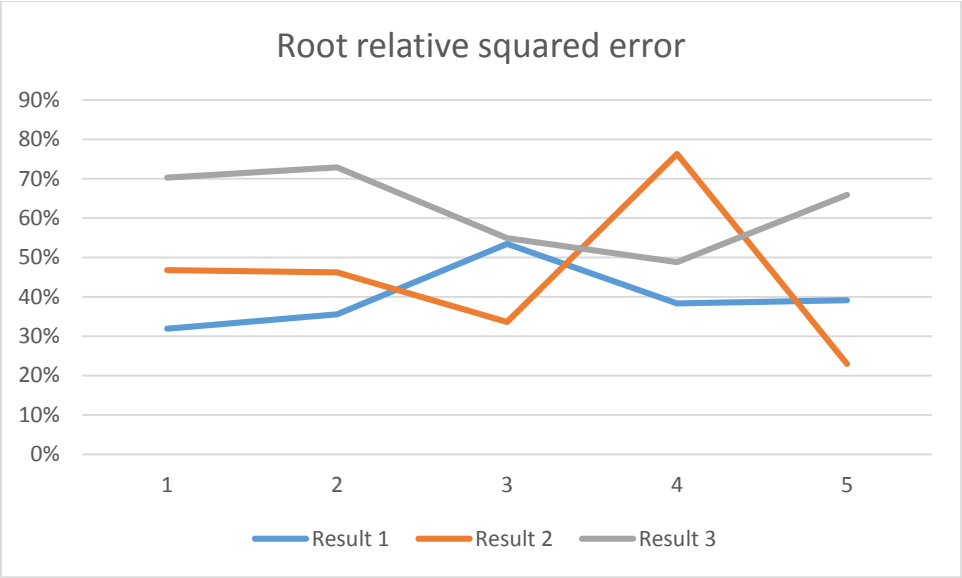

Figure 20: Root relative squared error of database B