Chair of Computational Modeling and Simulation
Department of Civil, Geo and Environmental Engineering
Technical University of Munich

Data Mining within the as-performed construction process

Scientific work to obtain the degree

Master of Science (M.Sc.)

at the Department of Civil, Geo and Environmental Engineering of the Technical University of Munich.

Supervised by Prof. Dr.-Ing. André Borrmann
Dr.-Ing. Alexander Braun
Chair of Computational Modeling and Simulation

Submitted by Fabian Pfitzner [redacted]
e-Mail: fabian.pfitzner@tum.de

Submitted on 24.09.2021

Abstract

In contrast to planning, construction has hardly changed in the last decade concerning digital progress. The execution on the construction site is still intransparent and complex due to the high number of different processes. However, the need for optimisation is immense: Operations can be accelerated, disruptions avoided, and damage prevented. To this end, there is a lack of support from digital methods within executed construction processes. In the context of this master's thesis, a developed data framework for investing existing processes of construction execution follows. The classification for the organisation and evaluation of the individual methods plays a central role. In consideration of this, the implemented data pipeline represents the "as-performed construction process" methodology. Eventually, individual processes are schematically examined based on the system created. With the help of designing a prototype, the potential for optimisation is analysed. The use of sensors (e.g. Bluetooth Low Energy (BLE) chips) applied on the construction site provides a suitable solution. All introduced prototypes are illustrated and verified within diverse case studies. The data obtained is then validated, evaluated and integrated into the Data Mining (DM) process. In summary, this thesis provides the foundation for further developing existing construction processes and collecting data for future analyses.

Zusammenfassung

Im Gegensatz zur der Planung hat sich das Bauen im letzten Jahrzehnt hinsichtlich des digitalen Fortschritts kaum verändert. Der Ablauf auf der Baustelle ist durch die hohe Anzahl unterschiedlicher Prozesse immer noch intransparent und komplex. Der Optimierungsbedarf ist jedoch hoch: Abläufe können beschleunigt, Störungen vermieden und Schäden vorgebeugt werden. Hierfür fehlt es an Unterstützung durch digitale Methoden in den ausgeführten Bauprozessen. Im Rahmen dieser Masterarbeit wird ein Datenframework zur Erfassung bestehender Prozesse der Bauausführung entwickelt. Dabei spielt die Klassifizierung zur Organisation und Bewertung der einzelnen Methoden eine zentrale Rolle. Die implementierte Datenpipeline stellt vor diesem Hintergrund die Methodik des "Bauausführungsprozess" dar. Auf Basis des erstellten Systems werden schließlich einzelne Prozesse schematisch untersucht. Mit Hilfe des Entwurfs eines Prototyps wird das Optimierungspotenzial analysiert. Der Einsatz von Sensoren (z.B. BLE-Chips), die auf der Baustelle verwendet werden, stellt eine geeignete Lösung dar. Alle vorgestellten Prototypen werden in verschiedenen Fallstudien veranschaulicht und verifiziert. Die gewonnenen Daten werden anschließend validiert, ausgewertet und in den DM-Prozess integriert. Zusammenfassend liefert diese Arbeit die Grundlage für die Weiterentwicklung bestehender Bauprozesse und die Erhebung von Daten für zukünftige Analysen.

Contents

List of Figures

List of Tables

Acronyms

ACE	Architecture Engineering & Construction
AI	Artificial Intelligence
AOA	Angle of Arrival
BDA	Big Data Analytics
BDE	Big Data Engineering
BI	Business Intelligence
BIM	Building Information Modeling
BLE	Bluetooth Low Energy
Cell-ID	Cell Identity
CRISP	Cross Industry Standard Process for Data Mining
DM	Data Mining
EIA	Environmental Impact Assessment
FTP	File Transfer Protocol
GIS	Geo Information System
GPS	Global Positioning System
IFC	Industry Foundation Classes
IoT	Internet of Things
IP	Internet Protocol
ISO	International Organisation for Standardisation
JSON	JavaScript Object Notation
KDD	Knowledge Discovery in Databases
KPI	Key Performance Indicator
LAN	Local Area Network
MAC	Media Access Code
MEP	Mechanical, Electrical and Plumbing engineering
MQTT	Message Queuing Telemetry Transport
NCV	Net Cashflow Value
NPV	Net Present Value
OGC	Open Geospatial Consortium
RF	Radio Frequency
RFID	Radio Frequency Identification
RSSI	Received Signal Strength Indication
SEMMA	Sample, Explore, Modify, Model and Assess
TDOA	Time Difference of Arrival
TLS	Terrestrial Laser scanning
TM	Text Mining
TOA	Time of Arrival
UWB	Ultra-wideband
WLAN	Wireless Local Area Network

Chapter 1

Introduction

1.1 Motivation

While considering the planning sector of the construction industry, many modern digitalisation methods such as Building Information Modeling (BIM) are presently being applied. These state-of-the-art methods have largely established themselves in the planning process, invoking significant optimisation potential. However, when examining the execution area of the construction industry, these advances are not yet to be found here. One of the main reasons for this is that the construction process has received less attention in preceding research than the final product, the building. On the one hand, a well-developed methodology and corresponding tools exist for planning processes with BIM. Yet, on the other hand, a complementary methodological approach is lacking on the construction site. For example, there are hardly any studies on the as-performed construction process, which represent the implemented construction methods, processes, and resource inputs, which would enable an analysis of potentially avoidable delays and costs. Research in this area is in its infancy worldwide.

This unexploited potential requires going beyond the simple application of established 4D modelling and looking at construction planning and execution processes in much greater detail than it is done today to make detailed statements about performance in construction execution. For precisely this purpose, an approach that has been adopted in many industries was selected: Data Mining (DM). A detailed inspection of this procedure is necessary for understanding the possibilities of analysing the ever-increasing amount of data. Consequently, this thesis deals with the theoretical points of view of DM and illustrates the scope of the process alongside a practical construction-related example, introducing project-related Key Performance Indicator (KPI) values and giving insights into unresolved research questions.

1.2 Digital Twinning

A **digital twin** is a digital replication of entities that enable data to be seamlessly transmitted between the physical and virtual world [1]. More, digital twinning illustrates the process of designing digital twins. Specific key characteristics can define digital twins:

- **Unique identifier:** Digital twins can provide communication with their real twin.
- **Sensors:** Digital twins can replicate the senses of their real twin.

Figure 1.1: Example of digital twin created by the Microsoft Hololens 2

- **Artificial Intelligence (AI):** Digital twins can make decisions according to their real twin.
- **Communication:** Digital twins can connect with the natural environment.
- **Representation:** Digital twins can represent the actual environment in a specified way.
- **Trust:** Real twins can trust their digital twins while interacting with them.
- **Privacy:** Digital twins must protect the privacy of their real twin.

El Saddik [1] describes multiple representations of digital twins, like Augmented, Virtual, and Mixed Reality, Haptic, Robotics, Cloud Computing, the Internet, Wearables and AI. Eventually, one of the discussed realisation methods is Internet of Things (IoT) devices. Data can be transferred from users to their apparatus by sending information about the environment through IoT sensors. This is of high interest when taking a closer inspection at monitoring construction sites. Each sensor device needed has to be accessible somehow. A flexible and reliable approach is connecting these devices to remote servers using IoT capabilities. The following section will have a closer look at the IoT.

1.3 Internet of Things

Nowadays, a large number of devices can communicate via Internet Protocol (IP) networks. The Internet of Things describes a giant network of connected devices, which can regularly collect data about how IoT objects are applied and analyse surrounding environments. IoT devices use sensors for different purposes to provide those functionalities, like measuring temperature, air quality, speed, acceleration, signal strength, and more [2].

The IoT is defined by fundamental characteristics [2]:

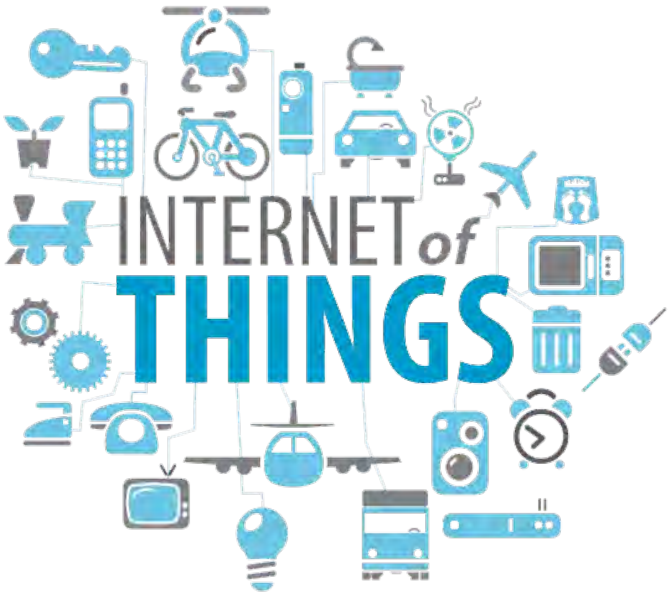


Figure 1.2: Internet of Things [2]

- **Interconnectivity:** A broad spectrum of objects can be connected to the IoT network. Examples are mobile devices, smartwatches, light bulbs, different sensors, etc.

- **Things-related services:** The offered services depend on the device. A smart-watch, for example, is usually capable of tracking the heart rate of its user.

- **Heterogenenity:** IoT devices can appear in all kinds of hardware.

- **Dynamic changes:** The state of the IoT object can change dynamically, e.g. from sleep to wake, low temperature to high temperature, etc.

- **Size:** The vast number of devices needs to be managed and maintained.

- **Security:** A crucial point of IoT devices is safety. The data must be secured safely, which can be done by using security protocol, for example.

- **Connectivity:** IoT devices need access to a local network to be connected to the internet and ultimately to produce data.

Eventually, the won data needs to be dumped to cloud platforms. IoT platforms are getting data from various sources, can perform data analytics, and extract information. Ultimately, IoT devices are adopted to automate processes and improve workflows. Which, in a sense, symbolises the fundamentals of Data Mining.

1.4 Internet of Things and Building Information Modeling

Recent studies have proven major opportunities for an IoT platform and BIM combination [3]–[5]. Today, IoT based systems have found much adaptation in building services and building automation systems during the operational stage of a building. However, the implementation of these IoT technologies is still missing at the construction stage of a building and, as a consequence, in the BIM method.

Tang, Huber, Akinci, *et al.* [3] emphasised the desire to improve construction and operation efficiencies and demonstrated that BIM and IoT integration are still in their worldwide infancy. More, they illustrated new application areas and common design patterns for the integration of BIM and IoT devices current frontiers and predicted future research directions. Some of the proposed future research directions are described in the following.

First, a great demand for web services for BIM and IoT integration to make BIM models stateful, and thus real-time interactive was identified. This includes real-time model updates based on IoT device readings, real-time information acquisition and control, ubiquitous monitoring and crowdsourcing monitoring and integration of other cutting-edge technologies. Yet, the desire for information integration and management standards in the Architecture Engineering & Construction (ACE) industry and interoperability between IoT devices and BIM cities was also emphasised. More, the need for cloud-computing concepts within the ACE industry was displayed, by illustrating crucial parts, like the lack of real-time Big Data analytics, BIM-IoT standards, IoT integration portals and methodologies, as most of the current sensors and BIM integration research are not yet connected with the cloud [3].

Indeed, this covers a lot of different IoT and BIM-related topics, showing the tremendous research gap. Still, some progress has been achieved in recent years. Dave, Buda, Nurminen, *et al.* [4] for example, show a conceptual framework using open standards and Industry Foundation Classes (IFC) models, integrating the built environment with IoT sensors.

In addition, the Open Geospatial Consortium (OGC) created an open-source, web-based IoT platform called the Sensor-things API. The SensorThings API contains two main functionalities, Sensing and Tasking. The Sensing part enables collecting metadata from various IoT systems and can be used to manage significant amounts of data [6]. On the other side, the tasking part provides parameterising functionalities for IoT devices, like individual sensors and actuators, composite in-situ platforms, mobile and wearable devices, or even pilotless systems platforms like drones, satellites, and autonomous vehicles [7].

More, Isikdag [5] shows in his research work how essential BIM and IoT data for GIS-based monitoring is and provides a method merging information of digital buildings with information received from sensors. This hugely benefits emergency response, urban surveillance and urban monitoring of smart buildings.

Past research projects ultimately show enormous potential when it comes to the integration of IoT and BIM. Among other things, this also applies to progress control within the as-performed construction process.

1.5 Research objectives

From the potentials mentioned above, the following research questions can be defined for this project:

- What is the basis for calculating construction-related KPIs?
- How can the as-performed construction process be detected automatically to a large extent without the need for personnel?
- What for the construction process suitable IoT technologies exist and how can they be joined with BIM information?
- How can the DM process be integrated into the construction industry?

1.6 Reading guide

The thesis is structured as defined in the following:

- chapter 2 gives an overview of the DM topic, provides information about specific DM terms and procedures, and shortly discusses methods of data analysis
- chapter 3 explains essential KPIs of the construction industry while deriving them to the as-performed construction process
- chapter 4 illustrates the adaptation of DM fields and procedures within the construction industry in-depth by showing DM related research work and concepts adapted accordingly
- chapter 5 displays a case study, showing different monitoring techniques, capturing the present state of a construction site-project, representing a crucial step of the DM process
- chapter 6 presents a case study, developing a positioning system for analysing workflows of the as-performed construction process
- chapter 7 provides a discussion of the case studies results
- chapter 8 recaps the overall findings during this thesis and gives a short outlook on further proceedings and future work

Chapter 2

Data Mining

2.1 Big Data

The volume of data has increased rapidly across all sectors in recent years. Big tech companies can process more than 20 petabytes (10^{15} bytes) a day. Not only the software industry treats high amounts of data, but also the construction industry deals with large data sets. BIM-Models can quickly reach the size of 50 GB. With increasing usages of sensors and devices, the amount can also grow. However, the capability to handle enormous amounts of data highly influences a company's performance and thus generates competitive advantages [8].

Big Data means controlling a considerable amount and variety of data from different sources. Structure and origin are often unknown and are usually characterised by the term "Multimedia". Five key attributes can describe Big Data [9]:

- Volume: Size of data files
- Variety: Diversity of data formats
- Velocity: Static or dynamic data
- Value: The increasing value of a company through data
- Veracity: Specific algorithms to classify vague data

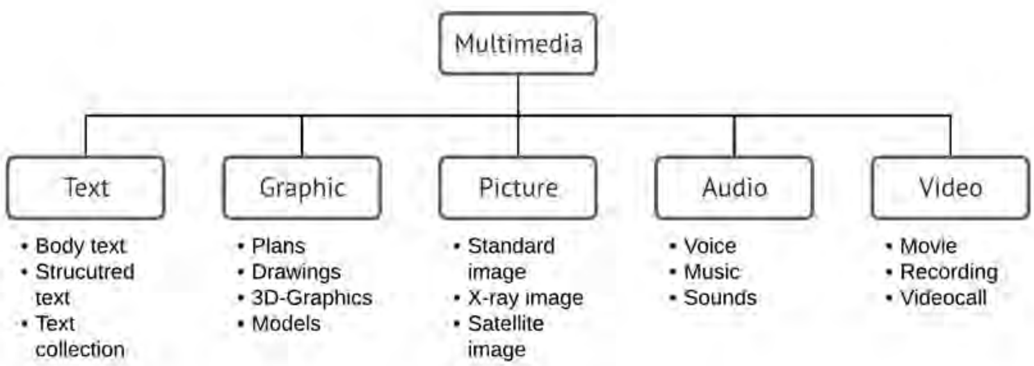


Figure 2.1: Multimedia, similar to D'Onofrio and Meier [9]

Construction data is primarily large, heterogeneous and dynamic, though it can help optimise construction operations and gain competitive advantages. Yet, the data has to be managed and understood. Therefore, Big Data accommodates **Big Data Engineering (BDE)** and **Big Data Analytics (BDA)**. BDE includes providing data storage and processing activities for analytics. The data storage can either be horizontally or vertically scaled. Horizontal means in this aspect allocating data on multiple servers. Vertical scaling, however, remains data on single-based configurations with upgraded server hardware [8]. An example for storing Big Data is using NoSQL databases, briefly discussed in the next paragraph.

NoSQL Databases are an often-used practice in data-intensive industries. NoSQL stands for "Not Only SQL" and is an improved method of traditional data management. With minor schema, they provide more scalability, performance and flexibility [8].

NoSQL Databases connect the SQL referential data schema with additional techniques like parallel data storing that increase database performance tremendously [10]. There are four standard formats to store data in a NoSQL database:

- Tree
- Columnar
- Graph
- Key-Value pair

An example of a NoSQL Database is MongoDB. This database is capable of storing data in any format and content. It is perfect for not only storing vast amounts of data files but also doing fast queries. The database is structured as a dictionary with Key-Value pairs, which means each object is associated with a unique attribute. This structure makes mapping data easy for developers. Even though MongoDB can handle many data files, it is limited to specific file sizes, making it not ideal for all use-cases [10]. Still, MongoDB has already been used for storing BIM data of building models for classifying processing through MapReduce. MongoDB also works with IFC. Nonetheless, the hierarchy must be adapted to support efficient query execution [8].

Going back to the BDE topic, the second step is about data processing. **Data processing** can be done by a large number of different models, Bilal, Oyedele, Qadir, *et al.* [8] present MapReduce and Directed Acyclic Graphs. Those are tools for dividing data into small fragments, which can be executed on a cluster and later re-aggregated for a final result [10].

Big Data Analytics is essentially bound for deriving information and patterns from the output of BDA [8]. Commonly, statistical methods and DM are used for these tasks. The following section gives a closer look at the practices of data analytics.

2.2 Fundamentals of Data Mining

2.2.1 Basics

A definition of Data Mining

The term Data Mining is used in quite different manners, which can lead to uncertainties. Cleve and Lämmel [11] define it as a synonym of Knowledge Discovery in Databases (KDD). However, it can also be interpreted as a single process of the KDD, as Bilal, Oyedele, Qadir, *et al.* [8] describe it. In Bissantz and Hagedorn [12] DM is seen as a pattern detection process, where specific rules can be derived from. Adams, Blunt, Hand, *et al.* [13] explain the process of seeking exciting or valuable information within large databases as DM. Yet, Hand [14] sets DM as discovery of structures in data sets. In fact, multiple definitions are used to describe the Data Mining process. To prevent misunderstandings, DM will be defined in this master thesis as follows: Data Mining describes the holistic process of extracting knowledge from a large amount of data and generating useful information.

Essential terms of Data Mining

Data is defined as a small unit of information. It can be unstructured (e.g. pictures, text), semi-structured (e.g. websites) and structured (e.g. relational databases, spreadsheets). Hence, the Data Mining process can be classified in Text Mining (unstructured data), Web Mining (semi-structured data) and Data Mining (structured data) at closer inspection. **Information** can be seen as the use-case-specific interpretation of data, which is related to a particular process. Therefore it is covered with a timestamp. **Knowledge** on the other hand, is information, which only can be used with expert capabilities [11].

Models are used to structure and manage data in Data Mining. A Model describes a large-scale summary set of data. The main objective of DM is detecting patterns. A **Pattern** is a local structure, referring to a small (depending on the context) number of objects. In many cases, it has an unusual structure and a not-trivial relationship to the dataset [13].

Business Intelligence (BI) is a method to manage the knowledge of companies efficiently based on DM. Yet, it is not only for gaining competitive advantages but also to survive in rapidly growing markets. The significant tasks of Business Intelligence are knowledge acquisition, knowledge management and knowledge processing. It can directly influence decisions with specific applications like Online Analytical Processing (OLAP), Management Information Systems (MIS) and Executive Information Systems (EIS). BI further covers all tasks used for the decision-making process, like, for example, systems to present, store, and manage data [11].

Data Warehouses are assembly points for external data, which are already cumulated and filtered. The main objective is to discover new structures in existing data. Data Warehouses consist of many databases and are often topic-related. To correctly assemble the data, Data Warehouses are required to be unified and time-oriented. Ultimately, consistent data growth is desired to improve pattern finding. Similar to BI, Data Warehouses contain decision-relevant data. As follows, there is an apparent close connection to BI. Indeed, Data Warehouses also can include tools for data analysing or use external interfaces [11].

2.2.2 Categorisation of data

Data can occur in many different ways and forms. As Data Mining methods require primarily structured and processed data, it is helpful to classify the data into specific types and ultimately to understand large amounts of data.

High-level

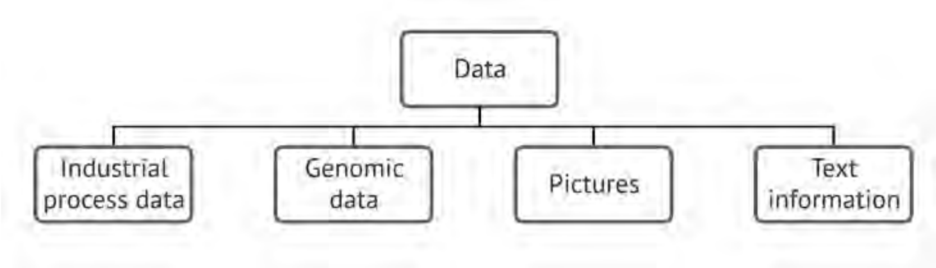


Figure 2.2: High-level categorisation

At the beginning of a data mining process, there are often unknown clusters and dependencies. In general, the following types of data can be applied on a high-level: **Industrial process data, sales data**, **genomic data**, **pictures** and **text information** [11]. **Industrial data** covers all data gained by sensors. Different sensors can be used to analyse working conditions and processes. **Sales data** is a vital factor to describe the current economic state of companies. This data is often already available and does not need many preparations. However, it is challenging to access sales data as most firms keep their data non-public. **Genomic data** is related to an integrative field of biology used to increase knowledge about the human body. **Pictures** and **text data** are often lower-level data with less information. Still, merging pictures and text with other data can be helpful to validate discovered patterns [11].

Low-level

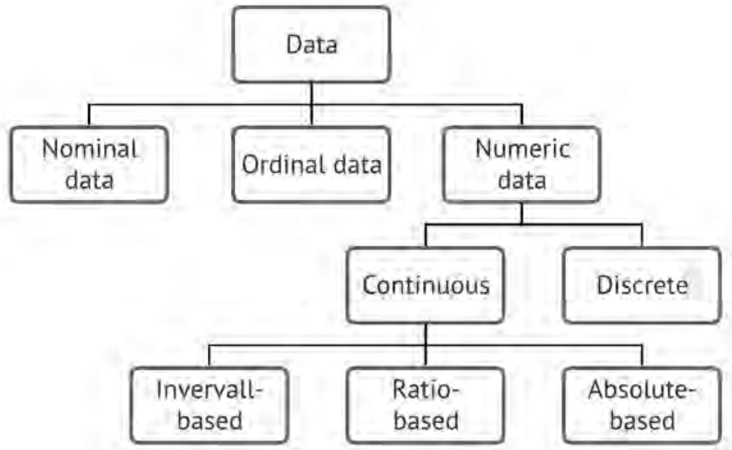


Figure 2.3: Low-level categorisation of data

Data can be categorised at a lower level into three different classes. Cleve and Lämmel [11] show different standards to organise data. The first criteria to classify data is order and calculability. It is distinguished between **nominal**, **ordinal** and **numeric** data. **Nominal data** has no order at all. An example of that would be the hair colour or country of a person. **Ordinal data** can be sorted according to specific rules, for instance, income level, education level. **Numeric data** can be used for calculating. It is further divided into discrete and continuous data types. Discrete data types are finite numbers to describe attributes, for example (e.g. shoe size). Continuous data types are numeric types that can adopt any numerical values within the predefined scale. Moreover, continuous data types are subdivided into interval-based with random origin (e.g. temperature), ratio-based with natural source (e.g. distance) and absolute-based data with no unit (e.g. the number of children).

2.2.3 Steps of the data analysis procedure

Fayyad-Model

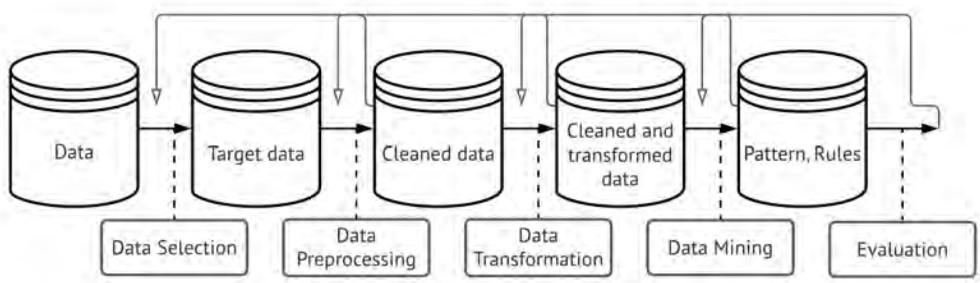


Figure 2.4: Data Mining process similar to Fayyad, Piatetsky-Shapiro, Smyth, *et al.* [15]

A common way to describe the Data Mining process is the Fayyad Model. It is divided into five key steps: Data Selection, Data Preprocessing, Data Transformation, Data Mining and Evaluation. The first step, Data Selection, determines the choice of appropriate data. This stage implicates analysing what data is available and what data can be obtained. The second step involves handling missing and incorrect data, which can appear due to particular circumstances. All issues must be removed to further work with the data, and missing values must be generated based on specific algorithms. The third step is about transforming data into suitable data formats. Therefore a specific dependency on units is needed. The Data Mining happens within the fourth step. The preprocessed and transformed data gets analysed. For that process, tools like decision trees or clusters are used. The last step involves the actual interpretation of the observed patterns. This is often done by consolidating experts and technical knowledge. The stage tremendously determines if specific processes can be improved or not [11].

CRISP Data Mining Model

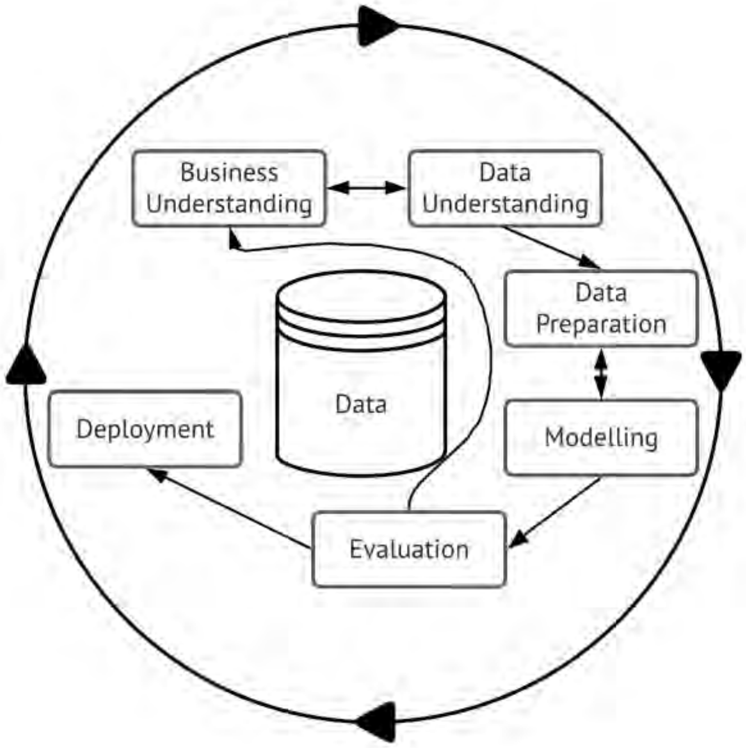


Figure 2.5: Data Mining process according to CRISP similar to Cleve and Lämmel [11]

The CRISP (Cross Industry Standard Process for Data Mining) model faces an industry-based approach. The model consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment. The first step covers task understanding. Therefore, expert knowledge is required to evaluate the actual state, quantify the resources, risks and costs and determine success criteria. The second stage is about understanding the data. Here, it must be clarified which data is desired and available. Further, semantic is key to interpreting the data. Within the third step, the data is prepared. Data preparation means selecting and transferring data into tables, sorting out incorrect data, handling incomplete data and transforming data for following DM processes. Once the data preparation is complete, the Data Mining step happens. Similar to Fayyad models, decision trees and clusters are created to structure the data. Depending on the tasks, the procedure and parameters are chosen. After conducting experiments, the model can be refined and improved. The evaluation process is responsible for rating the result according to economic value. Ultimately, an error analysis is done, which can also lead to a step back. Once the evaluation process is done, the deployment stage arises. This includes planning of integrating the achieved DM results. The main aim is to embed the development into the ongoing company process [11].

Looking at engineering-based tasks, which are often related to industries, this approach is promising for future data analysis. Still, it is inspiring to look at a different method implemented by the SAS Institute.

SEMMA-Model

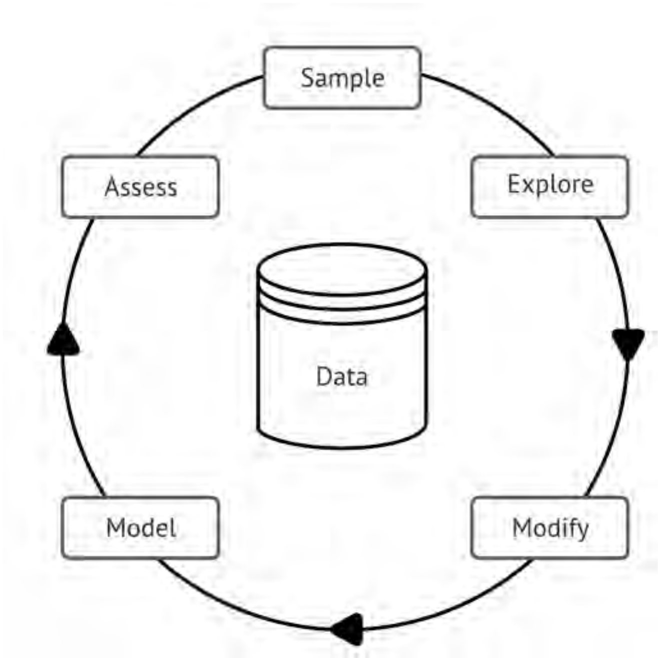


Figure 2.6: Data Mining process according to SEMMA similar to Kurgan and Musilek [16]

The SEMMA model stands for Sample, Explore, Modify, Model and Assess. It is a further procedure for Data Mining. The "Sample" step is for collecting data. Within the next stage, "Explore", the data is inspected and checked for data quality and comprehension. The third step covers modifying the data to improve the data quality and transformation for further usage. Within the "Model" step, an analysis and model building is done. The final step, "Assess", is responsible for evaluating the results [11].

2.3 Requirements for Data Mining

To process a fundamental DM method and not perform classical statistical methods, DM has some predefined attributes which need to be fulfilled [12]:

1. The presentation of the data must be in a clear, structured, readable form.
2. The processed results must be generated with a certain degree of security.
3. The gained information is compared to the existing system, not trivial.
4. The processing times of the algorithms are not too time-consuming manner.
5. Generating advantages to competitors with intelligent data analysis is an essential aim of DM.

More, it is essential to have automatic data processing within DM. Manual data management would be highly time-consuming and not expedient. Thus, the methods used must be adaptive and sequential. Before interpreting patterns out of data, the data has to be preprocessed. Unknown structures and dependencies of many data have to be assigned to classes, known or unknown ahead. This requires a structured and well-prepared data model [13].

2.4 Data Mining techniques

There are different approaches to structure, classify, and evaluate data within the DM step of the KDD process. Indeed, covering all procedures would extend the scope of this thesis. However, it must be taken a closer look at some common methods to understand further proceedings.

In general, it can be distinguished between describing methods and predicting methods. Part of the description is clustering and association analysis. The prediction method, however, covers classification and regression. Classification refers to nominal values, regression refers to numeric values, explained in section 2.2.2. A basic approach to differentiate between describing and predicting techniques is to classify those into unsupervised and supervised learning methods. Unsupervised means in this manner using undeclared variables. Thus, no forecasts are made, meaning there is no comparison to actual patterns. Supervised learning methods, on the other hand, are comparing extracted information with existing, labelled datasets. Therefore, they are comparing input variables to output variables to estimate the influence of the gained knowledge. More, the model learns how the result can fit the actual data [9].

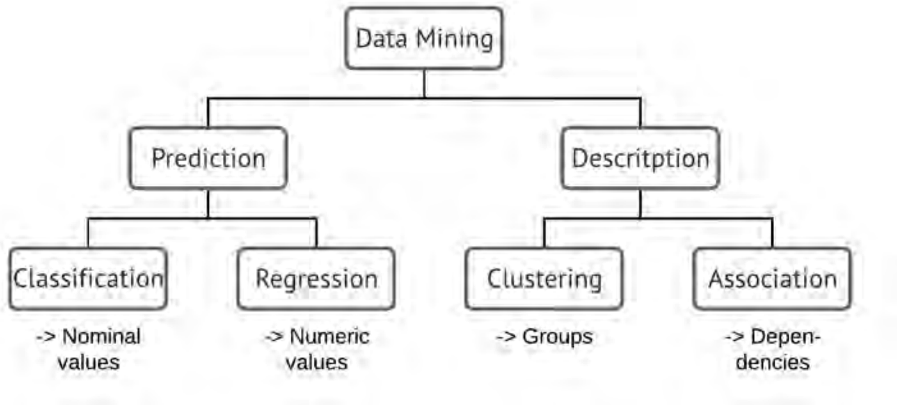


Figure 2.7: Data Mining methods similar to D'Onofrio and Meier [9]

2.4.1 Cluster analysis

The cluster analysis uses proximity to split data into groups. The following scheme is applied: The similarity of objects of a class should be large, yet the similarity between objects of different classes should be small. It can be distinguished between two different approaches: Conventional and conceptional cluster analysis. Within the conceptional approach, a high degree of proximity measurements is done. The proximity can be measured via specific distance methods, like the euclidean norm or manhattan distance. An approach for comparing different types of objects is, for example, to determine the distance 0 with equal attributes and 1 with different attributes. Further, the meaning of the elements can be extended by adding weight. The conceptional clustering, however, is based on rules for each class. A group of rules build a concept, which can create logical links to the object's attribute. A concept can be created by implementing a cluster system, and explicit assignment rules can be given [12].

2.4.2 Association

The objective of the association is to analyse which events are appearing at the same time. Association requires a set of items, for example, a group of measured data. This means a certain amount must be available to create rules. These rules can be described by a certain degree of confidence and support. Confidence is declaring the probability in which cases the rule applies. Support is declaring the probability in which cases the rule is applicable. In summary, confidence and support define the importance of a rule [9].

2.4.3 Decision trees

Decision trees are widely known methods for evaluating extensive data sets. Yet, they require an existing data pool with predefined classes. With this existing data set, rules can derive and be displayed in a structured tree. Those trees can also be extended with new objects. The procedure to create decision trees is called Decision Tree Learners and consist of the following steps [9]:

1. Define the classification variable.
2. Segment necessary amount of training sets.
3. Split training sets up into homogenous groups by creating knots of the tree with a degree of homogeneity.
4. Continue separation until the predefined degree of uniformity is reached in all knots.
5. Validate decision tree with actual data.

Decision trees are a good tool for deriving statements quite quickly from a given data set. Still, the data must be classified before, which can sometimes be challenging.

2.5 Challenges of Data Mining

A common problem of DM is **incorrect data**. This data can increase not only computing time but also hamper the DM process. Therefore, the efficiency and performance of the constructed algorithms must be high. Another critical point of DM is the **lack of expert knowledge**, which is crucial to interpret the extracted data and create rules. Further, it must be clarified how to handle uncertainties. Thus, a scale of error should be introduced before setting up the project. Another common problem is the **triviality** treatment. Sometimes the gained knowledge is already existing. At other times the developed statement is only based on a single element, which isn't meaningful [12].

A further challenge relates to the handling of very **large databases** with high dimensionality. This can tremendously extend computing time and also may require expensive hardware and high performing algorithms. Thus, it is necessary to do a validation step after the typical DM processes. This should verify the quality of the discovered knowledge, check the performance, prove human factors and detect triviality [17].

Chapter 3

KPIs of construction projects

3.1 KPI values at the project level

For the DM process and further the evaluation step, it is essential to understand extracted patterns and classify them. Therefore, a certain degree of comparability must be there. Yet, a common way for investigating industrial processes are KPI values. Their purpose is to represent a factor, which empowers measuring project and organisational performance. This section will provide a deeper insight into usual KPIs within the ACE industry. Indeed, the as-performed construction process can not influence all presented KPIs. Thus, a later selection will follow.

Chan and Chan [18] present a variety of KPIs which already have been adapted to the construction industry. The proposed KPI values from their research have also been validated within four case studies, confidently utilising them.

In general, it can be distinguished between two groups of KPIs: Those which can be estimated by a mathematical formula and those which are based on subjective opinions. The main groups will further be analysed to give a short overview. The most perceptible value is construction time.

3.1.1 Time

According to the initial schedule, each process has a predefined period in which it needs to be fulfilled. Steadiness is not always the case, and delays can appear. Thus, it has to be differentiated between reasonable completion date and project commencement date [18].

$$Construction\ time = Practical\ completion\ date - Project\ commencement\ date \tag{3.1}$$

Yet, the construction time can be utilised for measuring other time-related KPI values like the speed of construction and time variation.

$$Speed\ of\ construction = \frac{Gross\ floor\ area\ (m^2)}{Construction\ time\ (\frac{days}{weeks})} \tag{3.2}$$

$$Time\ variation = \frac{Construction\ time\ -\ Revised\ contract\ period}{Revised\ contract\ period} \tag{3.3}$$

3.1.2 Cost

Most companies are aiming to gain profit out of construction projects. Thus, costs are a crucial part of understanding how well a project is doing. It ultimately summarises all products, processes and modifications that can appear until the completion of the project [18]. There are some common ways to measure cost, the most common one yet is to estimate the unit costs:

$$Unit\ cost = \frac{Final\ contract\ sum}{Gross\ floor\ area\ (m^2)} \tag{3.4}$$

3.1.3 Value

As already mentioned in section 3.1.2 most construction projects are profit-oriented. Thus, it makes sense to have an indicator for measuring profit. Chan and Chan [18] propose the Net Present Value (NPV), which consists of the Net Cashflow Value (NCV) and a discount rate r. According to these values, the current cash flow of a construction project can be estimated.

$$NPV = \sum_t \frac{NCV_t}{(1+r)^t} \tag{3.5}$$

3.1.4 Health and safety

Whether a project is successful or not can also be determined by the number of accidents during construction time. Indeed, safety management is a crucial part of the overall as-performed construction process and shouldn't be neglected since most accidents happen during this stage. In general, safety rules are therefore negotiated before a construction project starts. Eventually, a factor of success could be comparing the actual state with the contracted state by calculating the accident rate of the project [18].

$$Accident\ rate = \frac{Total\ No.\ of\ reportable\ construction\ site\ accidents}{Total\ No.\ of\ workers\ employed\ or\ person - Hours\ worked\ on\ a\ specific\ project} \tag{3.6}$$

3.1.5 Environmental performance

Construction projects have a significant influence on the environment. Since environmental impacts are a global concern, there are specific rules to be followed. The International Organisation for Standardisation (ISO) provides a guideline on environmental management, also known as the ISO14000. Another framework that can be considered while measuring the environmental impact of a building is the Environmental Impact Assessment (EIA). The EIA covers projects in ecologically sensitive areas, like construction projects with special scientific interests. In summary, the environmental performance can be estimated by applying the ISO 14000, the EIA score and the total number of complaints received during the construction stage [18].

3.1.6 Quality and Functionality

The measurement of quality and functionality is often subjective, and therefore can only be interpreted as a vague success indicator. Still, there are different methods to estimate the quality of a building. One way is to see the building as a product, which convinces the customer to purchase due to its quality. Another way to determine the quality success is to check whether the technical specifications are fulfilled, either by representing a good standard or a proper procedure. Functionality is closely related to quality and thus also relying on a subjective evaluation. A common way to measure the functionality is to control if the building provides the intended functions and fulfils the expectations of the project participant [18].

3.1.7 Satisfaction of involved people

The satisfaction of involved people can also determine the success of a project. Nonetheless, this has two challenging aspects. First, the satisfaction of users and participants is most likely subjective, decreasing the expressiveness. However, the level of satisfaction from the client, design team, leader and construction team can be considered an indicator for measuring project collaboration success. Second, users' satisfaction can only be measured when the project is completed, making it challenging to provide a statement in an early stage [18].

Even though the participant's satisfaction is challenging to determine and sometimes a bit vague, it strongly influences future project collaborations and thus should not be neglected.

3.2 KPI values at the process level

Looking at the previously described standard KPI values, it quickly becomes apparent that some are more relevant for the as-performed construction process, and others less. Therefore, an investigation was done to see if a further analysis has to be taken in that field.

3.2.1 Deriving project KPIs to as-performed construction processes

The first question is how specific factors influence the construction process. In the following, the KPIs mentioned above are derived from the construction process and analysed.

Initially, **costs** have no direct influence on the construction process itself. However, they play a decisive role, as they are strongly linked to many other factors. For example, with low investment, less personnel, fewer vehicles, poorer materials can be used, ultimately leading to more unsatisfactory process performance, making costs a relevant factor. As all costs are well-documented and archived during a construction process, it has investigation potential.

Above all, **time** is the factor that influences the construction process most. Yet, time is closely related to the costs. If we look at the cost structure of a construction project, for example, it quickly becomes clear that many costs are dependent on the duration of individual construction processes. A typical characteristic of the construction industry is that a large part of the work builds on each other. Even if only small parts of the work are delayed, the time frame of the entire project can be significantly extended, leading to higher costs in total. Since each construction project has a detailed schedule, there are many opportunities to analyse individual process durations.

The overall **value** of the project and ultimately the profit can also strongly influence the performance of a construction process. A good example is prefabricated elements, which significantly accelerate the construction process but are only worthwhile if a sufficiently good cost calculation is made. Yet, it isn't easy to detect the overall value of a project since it depends on various factors.

The **health and safety** of workers also have a significant impact on the construction process itself. Accidents can cause enormous delays, reducing the performance of a construction process. Indeed, the health and safety field of construction workers was already covered in a research project using the planning stage of a construction project to detecting safety hazards in a BIM model and proposing measures by Zhang, Teizer, Lee, *et al.* [19]. However, these rules have to find adaptation during the as-performed construction process and must therefore be inspected. Thus, safety zones have to be defined, and monitoring has to take place. Still, as construction projects often cover extensive areas, it is hardly possible to encounter all incidents.

More, **environmental** regulations can be decisive for the duration of a construction process. Khalfan [20] presents a sustainability management activity zone, aiming to bring changes for improvement for the current day-to-day construction processes by using specific incentives, among other things. Additionally, environmental regulations often set the framework for materials that can be used, and are directly linked to other processes, thus indirectly influencing the construction process. Materials, however, directly influence building processes, as they determine the duration of a construction process. A good example of this is using prefabricated materials, which positively impacts the performance of the construction process since the required construction time is much less. Still, it is challenging to estimate the overall environmental impact due to the variety of materials used during a construction project.

Quality and functionality also influence the performance of individual construction processes. For example, both are responsible for whether work is done twice if the original work does not meet the required quality. In addition, the quality of a construction process can be affected by various factors, as pointed out by Arditi and Gunaydin [21]. However, in some cases, the quality can be subjective, e.g. building design, which makes it challenging to measure.

A final point that should not be neglected and also has an indirect influence on the building process is the **satisfaction of the people involved**. In most cases, these are the workers who carry out the work. Workers have a moderate influence on the performance of a construction process. Yet, the motivation of a worker is difficult to detect and thus offers fewer possibilities for investigation.

Table 3.1: Influence of construction project KPIs on as-performed construction processes

KPI	Influence on construction process	Detection potential
Cost	medium	high
Time	high	high
Value	high	low
Health and Safety	high	medium
Environmental	medium	medium
Quality and functionality	medium	medium
Workers satisfaction	medium	low

3.2.2 Scope for investigating individual processes

As previously described, it is essential to inspect specific processes in detail to evaluate their efficiency. Thus, it is crucial to collect a large amount of data to make target-oriented statements.

However, the most crucial parts of the project related KPIs seem to be time and costs. When analysing a construction process, it is often not so easy to determine the costs. This is sometimes because the costs depend on various factors, such as the process duration, material costs, transport costs, etc. Due to the opacity of the amount of data, the overview is quickly lost, which leads to high inaccuracies in the calculation and cost estimation.

Indeed, many studies have been taken to improve the cost estimation process. Abanda, Kamsu-Foguem, and Tah [22] for example, show an automated cost estimation process using the semantic of BIM models, which improves the cost checking workflow. A further cost estimation process, discussed by Shen and Issa [23], showed the relation of physical quantities and process-related costs. Similar to costs, there have also been studies taken to optimise process time. Jaselskis and El-Misalami [24] created an approach to reduce required scanning time by implementing an Radio Frequency Identification (RFID) system, showing significant optimisation potential in this area. Walker [25] developed a systematic method for measuring construction time performance, enabling comparisons between individual project performance and best practices worldwide. Eventually, it became clear that process time improvement is a highly complex issue with many different influences.

In order to get an overview of the time, costs and other KPI values of individual processes and the enormous, ever-growing amount of data, it makes, therefore, sense to look at DM methods that already have been adapted within the construction industry. In the following chapter, data mining in the construction industry will therefore be discussed in more detail.

Chapter 4

Data Mining in the construction industry

4.1 General overview

To further understand the need for Data Mining in the construction industry, a closer look has to be taken at the current status of DM inside the construction industry. In recent years, Yan, Yang, Peng, *et al.* [26] observed an increasing trend of DM related research projects. This tendency is closely associated with the increasing challenges of digital transformation, which need to be overcome. More, the rapid digital development of digitalisation also presents opportunities for digital transformation. With the beginning of 2016, the number of research publications has more than doubled, as illustrated in fig. 4.1. The geographical distribution shows that most of the research work was done in China. The increasing amount of construction-related data mining projects is intimately correlated to the rapid growth of the Chinese construction industry [26].

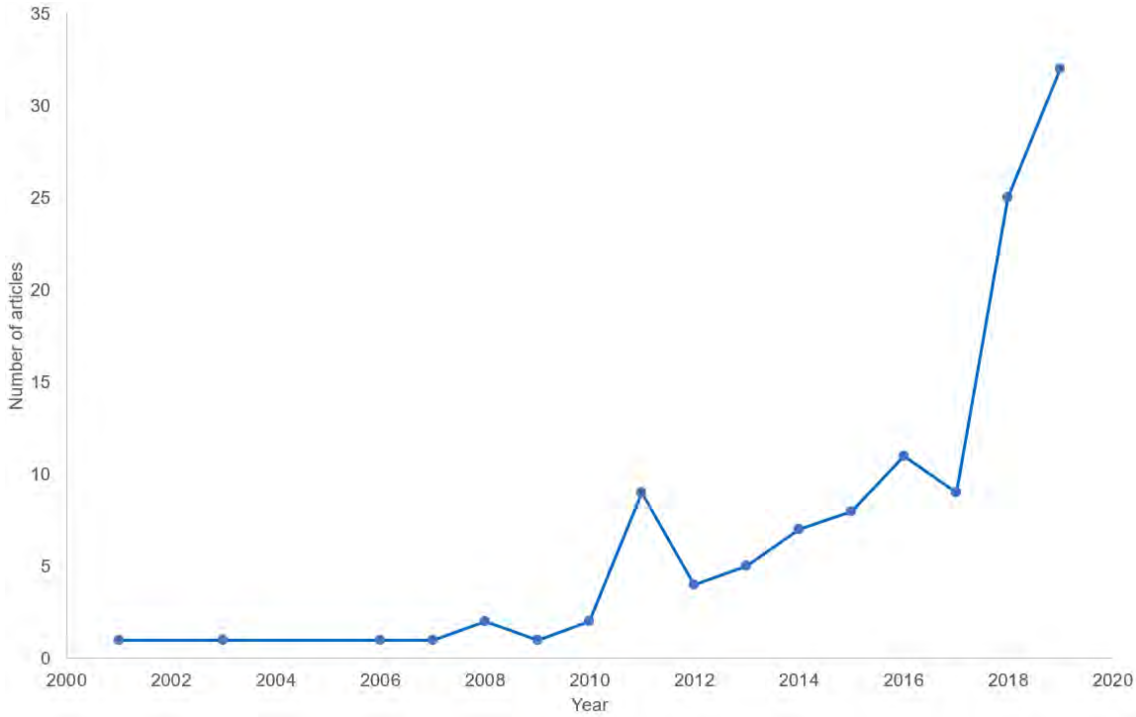


Figure 4.1: Published construction-related DM articles, similar to [26]

4.2 Application areas of Data Mining within the as-performed construction process

Data Mining is applied in almost all application areas of the construction industry. The research is distributed among energy, building occupancy and occupant behaviour, cost management, material performance, safety management, textual knowledge discovery, framework establishment and building design [26].

4.2.1 Energy

DM has been applied in many research projects to predict the building's energy consumption and identify patterns analogously. It has been pointed out to be an effective way for energy management. Consequently, many studies were taken in this area, which further substantiated that DM methods outperform conventional methods in most aspects. Those innovations enable potential opportunities, like using deep learning models, expanding the performance of DM models and ultimately avoiding the high cost and privacy concerns by using non-innovative approaches [26].

4.2.2 Building occupancy

A closely related factor to energy consumption is building occupancy. Yet, traditional approaches do not predict building occupancy. Thus, DM methods have found much application in this field. The data is mostly gathered by sensors of different devices, leading to a tremendous amount of raw data. Using IoT devices with DM methods provides a new scale of opportunities. Deep learning approaches, combining statistical methods with DM techniques, helped to identify common activities. Ultimately, relationships between occupancy and building energy consumption could be analysed, and various building energy levels could be compared [26].

4.2.3 Material performance

One of the most vital materials of the construction industry is concrete. The properties of concrete have a high impact on the stability and reliability of building structures. Yet, the estimation of these properties is done by experiments, which is costly and time-consuming. Recent studies have shown that it is possible to predict concrete parameters and thus the material's performance using DM techniques. The outputs of the DM models have been compared with experimental results for proof validation. These new approaches make the before-mentioned experiments obsolete in the future and reduce the number of delays caused by outdated procedures [26].

4.2.4 Building design

Many decisions made at early planning stages have an enormous impact on buildings overall environmental performance. Thus, some research projects propose adopting DM techniques to analyse the performance of multiple different building designs. Eventually, data analysis can help to save energy, reduce waste and save costs in future stages [26].

While looking at the as-performed construction process, exciting research topics are cost management, safety management, textual knowledge discovery, and framework establishment. Further insight on these points will follow.

4.2.5 Cost estimation

Cost estimation is essential for the successful implementation of a site project. Primarily cost-forecasting models have found adaptation in this area. DM classification algorithms can predict the level of cost overrun, which is very useful for managing and structuring financials in construction companies. Further research in the construction field has shown that it is also possible to estimate construction costs by considering economic variables and indexes. Estimating construction costs can be a substantial competitive advantage, making Data Mining a valuable tool for increasing the company's efficiency [26].

4.2.6 Safety Management

Safety management is not only crucial for worker security but also to prevent financial loss. Additionally, it is vital to identify causes for accidents. The past has shown that qualitative methods are time-consuming and subjective. An excellent way to secure accident causes is to use DM methods. Injury-related factors can be discovered in the early stages and eliminated. DM approaches in safety management are hazard identification, incident prediction, risk monitoring, and safety indicator development. In brief, DM within safety monitoring helps identify the correlation between causes and accidents [26].

4.2.7 Text Mining of construction-related documents

Text Mining (TM) is a procedure to process text data to mine valuable patterns, trends and rules. Typically an unstructured text is transformed into a structured representation with many marks or words using preprocessing techniques. Adding a weight factor to each term can compare the importance of documents against others. Certainly, TM can support discovering information in construction project documents. Many studies applied TM techniques for deriving potential knowledge in unstructured text data. Research projects have shown that knowledge was mined with TM methods from BIM Logs, construction contracts, work orders, building quality complaints, post-project reviews and change orders [26].

4.2.8 Data frameworks for handling BIM data

Data frameworks cover the entire DM process and can be used to solve construction-related problems. Research has been done to understand the application capability within the construction industry of these frameworks. This can help to analyse data more efficiently in the construction business. Much research was done on establishing DM frameworks in the construction industry, like cloud-based frameworks for handling massive BIM-Data, building automation systems, solving interoperability problems between business and project management in construction industries, and automatically identifying building operation strategies [26].

4.3 Monitoring the as-performed construction state

A crucial step of DM is Business Understanding, explained in section 2.2.3, which highly determines the content of a DM project. Thus, it makes sense to look at specific ways to gain data from construction sites. Construction monitoring is a common way to extract information from site projects and has found application in many research projects [27]–[31]. There are different methods to capture the current state of a construction project, which will be closer inspected in the following.

4.3.1 Laser scanning

Terrestrial Laser scanning (TLS) is a usual approach for generating point-cloud models, which has gained high importance in recent years.

It is done by a three-step procedure, as shown by Tang, Huber, Akinci, *et al.* [3]: The first step involves data collection and is the only stage where the laser scanner comes into action. In this step, point measurements of the construction site are collected from significant locations. Once it is completed, the second step can begin, which is data preprocessing. A set of points, also known as point clouds, are combined with one coordinate system. The last stage is related to modelling the BIM representation. The final step is crucial since the building models require rich semantic information, while information of a point cloud is relatively low-level.

Taking a closer look at this process, the similarities to the DM process quickly become apparent. More, the Scan-to-BIM process, as it is called in some research projects [28], can be seen as a part of the DM process, with stages of Business Understanding, Data Preparation, and Modelling.

Bosché [27] pointed out the lack of solutions for accurately and efficiently tracking the 3D status of buildings and emphasised how critical this status is for successful management of construction buildings. Thus, he presented a possibility for 3D status tracking while introducing a 3D Computer-Aided Design model.

In the last years, laser scanning has proven to be a promising approach for construction monitoring. Even though it might be complicated to process data, methods, as mentioned above, have been developed for efficient 3D object detection. More, the generated models have been combined with 4D object recognition systems [32], making it a perfect opportunity under the scope of DM.

As pointed out in many research projects, laser scanners have high accuracy and can achieve mm-precision [27], [32].

Laser scans have been used for different purposes as shown in [29], some of them are:

- Automated quality assessment of construction projects
- Tracking of the progress of individual structural components
- Dimensional compliance of Mechanical, Electrical and Plumbing engineering (MEP) systems
- Tracking progress on MEP systems
- Inspection tasks

4.3.2 Image-based Monitoring

The methodology to generate image-based point clouds can be broken down into two stages. First, using the produced images during the as-performed construction process, a point cloud can be generated. Images can, for example, be made by using a drone or preinstalling a camera on the crane. Second, the generated point cloud needs to be transformed to a BIM model using, for example, methods mentioned in section 4.3.1 [33].

However, the point cloud generation stage differs from the laser scanning approach. Braun, Tuttas, Borrmann, *et al.* [30] define four steps to generate a point cloud with photogrammetric material. The first step is **Data acquistion**. This includes placing a camera on-site, creating a storage system and start the image taking process. It is essential to cover the complete site to include all processes.

Once the collection of data is complete, it can be continued with the second step, **Orientation**. This stage involves creating fix points to determine the orientation of all images, which provides a scale and adjustment in space by specifying the corresponding standard deviations. Within the following step, **Image matching**, 3D points are calculated by overlapping disparity map images and then using triangulation algorithms. The information, like Image-ID, RGB-colour and accuracy, are saved for every point. The last step, **Co-registration**, includes measuring the images to calculate transformation parameters. However, it is only performed if the model coordinates and control point coordinates are not in a standard construction reference frame [30].

4.4 Monitoring of construction processes

Aside from monitoring the current progress of a construction project, it is also essential to look at some processes in more detail. Park and Brilakis [31] showed how tracking data of workers using video frame detection methods is helpful for tasks like productivity management, dynamic sequence analysis, and safety management. Yet, the lack of detection methods and the challenges of positioning systems were also pointed out as unsolved topics.

Similarly to tracking workers, tracking material on construction projects can improve project performance and be an indicator for creating KPI values. Song, Haas, and Caldas [34] showed an approach using RFID tagged materials and Global Positioning System (GPS) sensors to gain information about construction materials automatically. This method then can be used to quickly identify and locate certain materials during the as-performed construction process.

Yet, there are many ways to perform a positioning procedure, which will be closer inspected in the following paragraphs.

4.4.1 Positioning systems

As already shown in other publications, there is a wide variety of different monitoring technologies [35]–[39]. Analog to the current research work, a further summary was accomplished. The survey allows comparing the systems and ultimately making a suitable selection, focusing on the intended use. The following section provides a variety of existing positioning systems.

Radio Frequency Identification

RFID is a technology using electromagnetic fields to store and retrieve data. The system's main objective is the automatic tracking of individual people and objects. In general, there can be distinguished between two basic types: active and passive systems. Passive systems are fixed on locations to calculate the live position of a receiver, while the tags are in no need of an external battery. Active systems, however, use receivers to read the location of a battery-powered tag. Passive systems are limited by range. Active systems are more expensive due to the increasing amount of RFID readers. The accuracy of RFID depends on the setup and varies between one and five meters. The RFID technology is primarily used in complex indoor environments. For outdoor use, it is commonly combined with GPS to cover large open areas [35]–[37].

RFID comes with reliable performance and a strong anti-interference. The downside, however, is the need for RFID readers and tags, which might be costly [38].

An example of implementing this technology on a construction site would be using a network structure of an RFID tracking system. Chai, Wu, Zhao, *et al.* [40] provide an RFID tracking system consisting of three main components: active RFID tags, RFID readers and a tracking server. The tag is powered by a battery and transmits radio signals to the readers. All the data gets passed to a tracking server, which processes the readings and calculates the object's position. Active tags were used to achieve better coverage in large industrial areas.

Global Positioning System

GPS uses the triangulation method, based on satellites and receivers, to gain the position of an object. GPS technology can be found in various devices like smartphones, tablets and other portable devices. On the construction site, however, it is mainly used for tracking the location of trucks, materials and cyclic activities of equipment [35], [37], [38].

GPS is one of the widely used positioning technology in outdoor environments. The limitation comes with indoor usage since it is impossible to have a line-of-sight transmission between receivers and satellites. According to Gu, Lo, and Niemegeers [39] GPS can not give exact location information in indoor space. Thus it is required to extend the system with other location-based services. Torrent and Caldas [37] propose combining GPS with RFID to support data collection while using advanced localisation technologies to obtain an improved location result of the objective based on the collected data. Further, it has been demonstrated that the aggregation of GPS and RFID results in a highly accurate result, as pointed out by Li, Chan, Wong, *et al.* [35].

Bluetooth Low Energy

BLE locating systems are connected to a mobile device, sending data to a remote control server. The position of the target can be estimated based on different algorithms. The sensors are mostly applied in indoor environments [41], [42].

Since construction sites are dynamic and noisy environments with much interference, it can come to unreliable Received Signal Strength Indication (RSSI) values. To overcome this problem Park, Cho, and Ahn [42] were deciding to use so-called probability approaches instead of mathematical positioning methods.

Generally, Bluetooth chipsets are low cost, which makes the BLE positioning system relatively attractive. One main limitation of a Bluetooth-based positioning system is the suffering accuracy with a complex and changing indoor situation. Also important to notice is the delay in calculating the position, which may take around ten to thirty seconds [39].

Ultra-wideband

UWB is part of the Radio Frequency (RF) location systems. The UWB technology uses the time-of-flight measurement technique and two-way ranging mechanism to calculate the distance between target and anchor. A multilateration algorithm, similar to the GPS technique, and at least four anchors are needed to estimate the target's location [43], [44].

UWB systems are mostly used in indoor environments as they provide more accurate results due to the distortion resistant behaviour. In parallel, the penetration capability has proven to be strong. Yet, on large open spaces, the performance of the UWB system decreases. An additional issue comes with the distance between the tags, which will also influence the accuracy of the UWB technique. It has been observed that obstacles can reduce the precision of the system. Considering the scope of this thesis, "Data-Mining within the as-performed construction process", the deployment of UWB requires the connection of a Local Area Network (LAN) to the receivers, which might not be available at early construction stages. Other limitations are potentially high costs, large size hardware and no smartphone support [35], [38].

Rashid, Louis, and Fiawoyife [44] show an example UWB-based localisation system used for a smart home environment. The system consists of four anchors and one hand-held clicker, including the UWB localisation tag. The clicker, carried by the user, transmits the position and orientation information, while a user-interaction with electrical devices, provided as a BIM-based virtual model, is made possible.

Vision Analysis

Vision-based positioning systems consist of multiple cameras. There is no need for the target to wear any tag. Visual tracking algorithms are required to obtain the position of the object. Video monitoring is mainly utilised in indoor environments. The vision-based technique is mainly used to analyse workforce productivity, workforce training and safety monitoring [35], [45], [46].

Vision analysis can cover large areas and are reasonably low cost. The accuracy of camera-based systems reaches into the meter range. One of the limiting factors is the surrounding environment. Dynamic environments with changing light and background colour may affect the accuracy of the system. It is also important to point out that implementing video monitoring systems can be challenging due to privacy protection laws [35], [36], [38].

An example of using the vision-based localisation technique for safety monitoring and progress analysis on the construction site is given by Yang, Arif, Vela, *et al.* [45]. A stationary camera and a moving camera were used to detect motion and people. The target localisation was done by training a colour model to discover the body regions of people.

Wireless Local Area Network

WLAN based localisation systems use the signal strength of an Radio Frequency wave to estimate the position of an object. Different localisation techniques like Cell Identity (Cell-ID), Time of Arrival (TOA), Time Difference of Arrival (TDOA), Angle of Arrival (AOA), and signal strength categories are classified within WLAN positioning. While most methods are hard to implement or inaccurate, a distance-to-signal-strength relationship system, so-called RSSI, is commonly used [47].

WLAN can reuse the existing WLAN structure by using the embedded sensors on smartphones. This makes implementing the system relatively low cost. Nonetheless, a constant stable connection between the target and the WLAN is needed, making it harder to track persons. The accuracy is within the meter scale and decreases as the target changes from static to dynamic. It may also be affected by manufactured materials such as concrete [35], [38].

A case study of Woo, Jeong, Mok, *et al.* [47] showed an approach of using active RFID tags and a Wi-Fi-based positioning system to track workers at the Guangzhou MTR tunnel construction site. It was found that the system was sufficiently robust under harsh conditions.

Ultrasound

Ultrasound positioning systems use ultrasound signals to locate an object. Distance measuring and trilateration algorithms can be used to estimate the position of an object. The ultrasound moves slowly enough for a device to sense the first arrival, preventing multipath interference [48].

Ultrasound tracking techniques are developed mainly for indoor use. The localisation system benefits from a high accuracy, which goes beyond the centimetre scale. With increasing distance, however, the precision decreases. This limits the range between one and fifteen meters, making it nearly impossible to use the system in large open spaces. Another critical point is the limited penetration capability of ultrasound. As ultrasound waves can not penetrate walls, they also may be reflected by other signals and noises [35], [38].

Table 4.1: Comparison of positioning systems

Technology	Range	Accuracy	Limitations	Cost
RFID	large	1-5 m	outdoor environment	moderate
GPS	very large	1-10 m	indoor environment	low
BLE	moderate	2-5 m	sensitive to other signals	low
UWB	moderate	0.5-1 m	obstructions	high
Vision Analysis	moderate	0.5 m	surrounding environment	low
WLAN	no data yet	1.5-5 m	network stability	low
Ultrasound	moderate	<0.1 m	reflecting signals	moderate

4.4.2 Suitability of positioning technologies for construction site use

A closer look at the on-site construction conditions is needed to specify which previously described localisation techniques could be applied. Construction sites are often dynamic environments with rough surroundings and changing lighting. Therefore the assembling of the system has to be considered, and flexible build ups might be prefered. Since there are many obstructions, holding a continuous line-of-site to the target object is not always possible. Workers or materials are in an ongoing flow, which can prevent a stable vision line. Construction sites are often segregated into indoor and outdoor spaces. For covering the majority of the construction field, the positioning system must function under indoor and outdoor conditions. Even though it is doubtful that there will be an existing Wi-Fi network or comparable network infrastructure during the shell construction phase, the network accessibility of the monitoring systems need to be inspected. To meet the purpose of this project, which involves DM, a large amount of data is necessary, leading to the preference of a dynamic tracking method. It is therefore also relevant to check whether the systems are capable of dynamic tracking. Tracking multiple workers is crucial to get meaningful results. For this reason, the positioning system should allow various targets. Since the meter range is sufficient for determining the position of a person, no highly accurate systems are needed. Nevertheless, the system should be able to provide reliable results. Interference is often present on the construction site due to different materials, obstacles or other sources. The systems have to be checked and evaluated accordingly. As the project is to occur within the framework of a research project, the costs must be manageable. This also implies that it should not be too demanding and time-consuming to install the tracking system. Further, a system with less infrastructure and preferably smartphone utilisation is favoured.

Suitability for dynamic, rough indoor and outdoor environments

At first, a closer look has been taken at the capability of the individual positioning technologies to perform under dynamic indoor and outdoor conditions. Thus, it is also essential to cover the historical application of the tracking techniques.

RFID systems are commonly used in complex indoor building environments as they provide an interference-resistant and flexible solution [35], [39]. Further, it has been shown that the RFID technology was already successfully implemented on construction sites, as RF suffers less from interference [38]. Summing up, the RFID technology is a suitable system for dynamic, rough indoor and outdoor environments.

GPS technology has already proven a lot to be an excellent outdoor localisation system. Still, GPS is struggling with indoor usage since it is hardly possible to achieve accurate position estimations [38], [39]. At construction sites, it has been used for tracking vehicles [35], [36]. An idea for using it on-site would be attaching a GPS receiver to each construction item. It has been proven that this attempt is not a feasible option at all since the costs of most GPS devices are high, and obstructions can easily block the weak GPS signals. To overcome this problem, storing the location of each component on hand-held units would be possible but still would lead to increasing labour intensive and manual updates [37]. Another approach to face indoor construction environments are indoor global positioning technologies. While those systems can improve indoor tracking accuracy, it still is an expensive and rarely used solution [49]. The GPS is, without any doubt, an adaptable tracking solution for outdoor environments. Nonetheless, GPS technologies still suffer in indoor environments and are not recommended inside a building environment.

BLE systems are easy to set up and robust, making them an ideal choice for dynamic construction environments. Further, there is no need for external processing support or configuration updates. Workers can be equipped with wearable battery-powered tags, which last for multiple years. Depending on the systems infrastructure, BLE can be used for indoor and outdoor purposes [50]. It has been observed that it is an accurate positioning technology, although it might interfere with multi-obstacle environments [35], [39]. After all, Bluetooth systems have found many applications for simple tracking tasks in construction [36].

UWB has been used a lot in complex building indoor environments such as hospitals and warehouse facilities. Further, it also has been tested outdoors. In dynamic changing environments, however, the system seems to be less accurate, as occlusion and the surrounding environment impact the accuracy [35], [38], [39]. Still, it has been demonstrated to be a working solution for monitoring the construction progress. Indeed getting information on placement resources such as workers, parts, and heavy equipment worked with the UWB system [36]. UWB is a feasible option for indoor and outdoor environments. When it comes to dynamic environments, the system may struggle.

A standard tracking system for complex indoor and outdoor environments is vision analysis. As video monitoring is performing seamlessly under static conditions, the technique suffers from dynamic environments. Occlusion and changing backgrounds decrease the accuracy of the positioning technology [35], [38], [39]. Nevertheless, vision analysis has been used on construction sites for tracking workers, parts and heavy equipment. The dynamic environment is a big problem, even though overcoming the background issues with specific processing algorithms is possible [36], [38]. Since camera position systems are hardly useable in dynamic environments, the system is not suitable for construction sites, although system usage may be preferred in indoor and outdoor areas.

The WLAN based positioning technology is typically used in indoor environments related to existing network infrastructure [39]. Yet, it is the most inaccurate positioning technique in complex, dynamic environments [35]. Still, the WLAN positioning technology found application at the construction site for personal tracking in indoor environments as it has promising potential [36], [38], [49]. However, it is not a preferred option for construction sites for dynamic and outdoor environments, the small accuracy and the missing infrastructure in outdoor areas are insurmountable odds.

Ultrasound systems were tested as indoor positioning technology on construction sites. It pointed out to be challenging to apply the system in a dynamic environment. Further, the ultrasound waves can not penetrate walls and can also be distorted by reflecting signals [35], [39]. Yet, it can be generally used on dynamic outdoor and indoor construction sites, but it still needs a lot of development [38].

Line of sight, hidden targets, multiple target limitation

The RFID technology can track multiple materials and workers [35]. It does not require a line-of-sight implementation [49]. Further, active RFID tags operate without a line of sight at long distances, which makes them one of the preferred options for identification processes [37].

GPS needs a line of sight between satellite and GPS receiver [35], [37], [39]. However, it can track several measurement devices simultaneously [36].

Since UWB belongs to the RF family, no line of sight is required [39]. Nonetheless, a missing line of sight may impact accuracy due to multipath reflection and absorption [49]. There is still a lot to be discussed about the line of sight influence on UWB as shown in Heydariaan, Mohammadmoradi, and Gnawali [51]. Tracking multiple tags at the same time will decrease the accuracy of UWB [35].

For vision analysis, a line of sight is required. It can come to difficulties with multiple targets because all targets need to be in the range of the visual angle. An advantage of camera monitoring systems is that there is no need for the target to carry any device [39].

WLAN and Bluetooth positioning systems do not require line-of-sight implementation [49]. Further, it is possible to track multiple targets in a wireless network [39].

When we look at ultrasound positioning systems, there is no line of sight transmission between tracked targets and detectors required at first [39]. There is a high risk of multipath reflection and absorption in environments with many obstructions, which will decrease accuracy when there is no line of sight between transmitters and mobile receivers [49]. Depending on the ultrasound positioning system, it is possible to track multiple devices.

Application of existing network structure

In early construction phases, there might not be a network structure available. Still, it is crucial to consider the impact of this aspect, as an existing network structure can be a huge advantage for some location systems since deployment effort and cost decreases.

When combined with WLAN, RFID can use existing network structures [35]. Most RFID systems, however, have their own locating structure, which reduces the usage of the Wi-Fi network only for connecting to a server [38].

Both GPS and vision analysis are not capable of using existing network structures.

Analog to RFID UWB, Bluetooth and Ultrasound technologies have their own locating system structure [39]. Nevertheless, the deployment of the data may need a connection to a local area network [35]. Thus, an existing network structure can combine several gateways and push gathered data on a cloud network.

The WLAN positioning technique can use an existing network infrastructure, shown multiple times [35], [38], [39], [49]. Summing up, it can be noticed that only the WLAN positioning technology benefits from the existing network structure. As already mentioned at the beginning of this subsection, it is unsure if Wi-Fi infrastructure is already available. Considering this, not so much weight should be given to the application of existing network structures.

Dynamic tracking method

To get relevant, real-time results, a dynamic tracking method is needed.

As each RFID tag can be identified and located, it is possible to track locations [39]. Combining it with other technologies, like GPS and WLAN, many components can be live tracked within a short amount of time, covering large areas [43].

The GPS technology provides a real-time tracking method, which can also be used on mobile devices [35], [38]. Nonetheless, it must be differentiated between the purpose of the positioning of elements as explained by Vähä, Heikkilä, Kilpeläinen, *et al.* [36] and live monitoring targets shown by Torrent and Caldas [37], as it requires different GPS devices.

In general, it is possible to create a live positioning system with BLE technology. However, it depends on the beacon setup. Shorter distances between beacons and receivers are needed to avoid long response times. Large areas might be divided into smaller subareas to gain dynamic positioning [41].

The UWB indoor positioning system was developed for real-time tracking, making it a suitable method [35]–[37], [39]. The same applies to ultrasound systems.

Similar to GPS, vision analysis can be used for dynamic tasks, such as the safety monitoring of workers [35], [39] and static tracking tasks, such as measuring distances [36].

For the WLAN tracking technique, real-time positioning is possible. However, the error increases for a moving object [35]. This makes the system more suitable for static tasks.

In summary, RFID, GPS, UWB, vision analysis and ultrasound technologies are suitable for dynamic tracking tasks. WLAN and Bluetooth are not recommendable for live tracking on construction sites, as they have their previously mentioned limits.

Coverage of large areas

When we think of construction sites, we have to consider the building area and the land around them.

Only active RFID tags can cover large areas since passive tags have smaller ranges [39]. Indeed active tags can communicate beyond hundred meters [37]. Further, RFID systems are often combined with GPS in outdoor environments, which increases range [35]. It is required to select at least one active RFID method to provide a capable system at a construction site. Combining it with GPS would be the best solution.

As already mentioned in the paragraph before, GPS is a widespread technique to cover vast outdoor areas [35], [36], [38]. Taking construction areas into account, GPS is suitable for scanning construction components in extensive areas [37]. In total, GPS provides an excellent choice for covering large outdoor areas and should also be considered for supporting other positioning systems.

The range of UWB technology depends on the individual system specifications. The Ubisense locating system, based on UWB, can, for example, areas up to four hundred square meters [39]. Other systems, however, only offer a range of between one and three hundred metres [36]. Nonetheless, it has been tested that it is possible to cover large open areas in construction environments up to 65.000 square meters. Yet, it comes with respectively less accuracy [35]. In conclusion, UWB is capable of covering moderate areas. However, the systems aren't intended for significant open area usage. This makes it a less prefered option, considering open construction fields.

Camera-based positioning systems can cover relatively large areas [35], [38], [39]. Still, it must be taken into account that cost and effort tremendously increase by enlarging the area. Even though it is possible to cover large construction fields with visual locating systems, it might not be the most suitable choice.

WLAN and Bluetooth positioning systems provide both a moderate range. As for Bluetooth, it was shown ranges to hundred meters, WLAN is even possible to cover an area around three hundred square meters inside a building [39]. With Bluetooth 5.0, greater areas can be covered [50]. Looking at extensive outdoor areas, deploying those two systems might become difficult. As already mentioned in section 4.4.2 it is a significant effort to set up the systems infrastructure in large open construction areas. More, an increasing number of beacons and areas with mesh routers are needed.

For example, some ultrasound positioning systems, like the Active Bat System, can cover areas up to a thousand square meters. Yet, a large number of receivers is needed, depending on the system implementation [35], [39]. Placing hundreds of receivers exceeds the limits of a construction project, making ultrasound positioning systems a less preferable choice for open area coverage.

Accuracy to meter range

When it comes to positioning at construction sites, decent precision is needed to get meaningful results. Thus, the error of the locating system should be less than two meters.

RFID systems can obtain good results with an error range from two to five meters [37]–[39], [49]. Further, accuracy can be improved with different algorithms [35]. The RFID positioning system is a decent option accuracy wise, as it fulfils the before mentioned requirements.

GPS systems, however, are not known for estimating the most accurate results. Usually, a GPS device can provide an accuracy of about ten meters [35]–[37]. Nonetheless, real-time correction methods can be adapted to obtain precise GPS positions [37]. This can improve the precision to one meter. Yet, it is crucial to notice that accuracy decreases in indoor environments, as pointed out in section 4.4.2. In summary, GPS are not the first choice when it comes to a precise location method. As it can be merged with other positioning technologies, there is still much room for adaptation on construction sites.

Bluetooth locating systems can gather an error around two meters [36], [39]. Withal accuracy decreases if the distance between the beacons becomes greater than ten meters [41]. In this way, the preciseness of a BLE system depends on network density. So far, it has been used for improving the accuracy of GPS systems [35]. On that account, Bluetooth technology fulfils the requirements for accuracy on construction sites.

UWB positioning systems can obtain accuracy up to the centimetre scale [36], [38], [39]. Still, it has to be considered that the error increases in large areas [35]. Notwithstanding the above, UWB provides an excellent choice accuracy wise for construction sites.

The accuracy of camera-based positioning technologies varies [36]. It is usually possible to achieve an error within the meter range [38]. Nonetheless, the focus on an object can be lost when its appearance changes [35]. Errorless tracking might become challenging with visual analysis since the background conditions can change all the time. Hence another tracking option should be preferred.

The WLAN positioning technology can reach accuracy to one meter [36], [38], [39], [49]. However, the target must be connected to the network [35]. On construction sites, it can be challenging to comply with a stable connection at all times, making WLAN based positioning techniques not the optimal choice.

Ultrasound is the most precise positioning technique. It can reach an accuracy to the centimetre scale [35], [38], [39], [49]. When placed correctly, ultrasound systems can quickly fulfil the accuracy requirements on construction sites, making it the optimal choice for precision. Still, it is crucial to remember that such precision is not necessarily needed on construction sites.

Interference resistance

A high interference resistance is required to reckon a seamless positioning technology. Hence, all systems are checked on the aspect mentioned above.

The RFID technology seems pretty resistant overall. Still, RF positioning systems can suffer from the multipath distortion of radio signals reflected by indoor walls [39]. Steel elements can also affect accuracy [35]. It is crucial to note that RFID systems have a higher resistance than other positioning technologies [39]. Overall, RFID location systems epitomise a suitable solution for job sites in terms of interference resistance.

The impedance of GPS devices can decrease when another radiation is present [39]. GPS systems are hardly useable in dense urban areas since signals are blocked by high buildings and other signals interfere with the GPS sensors [35]. As already mentioned in section 4.4.2, GPS devices need to be in a constant connection with satellites, which makes them somewhat unsuitable for construction sites.

UWB systems provide a robust anti-interference ability and strong penetrability [37]. There are methodologies to filter interfering signals from the original signal. Thus there is less multipath distortion [39]. Similar to the RFID technology, UWB systems suffer from the downsides of RF, like being affected by metal effects and obstructions at a construction site [35]. Further, it has been noticed that weather conditions may impact accuracy [37]. Summing up, the UWB positioning system maintains a stable impedance for job sites, even though some signals may interfere with the devices.

While camera-based systems are not affected by surrounding signals, it still suffers from illumination, background colour and the dynamic environment [35], [38], [39]. The aforementioned makes it hard to implement cameras on construction sites, as it is susceptible to many environmental effects.

Various influence sources deteriorate the accuracy of WLAN and Bluetooth technologies [39]. The frequency, signal strength, device, and orientation of wireless networks interfere with the locating system [35]. The existing Wi-Fi structure can interfere with the Bluetooth beacons. However, Bluetooth 5.0 improved this circumstance [50]. As we have many disturbing signals on the construction site, an appropriate placing of the anchors must be considered.

Ultrasound positioning systems suffer most from interfering signals. The missing ability to penetrate obstructions, illumination, occlusion, and scale variation decline the resistance of the ultrasound technology [35]. The reflection between tags and receivers also degrades accuracy, as it can change wave properties such as time, angle, or phase of arrival [49]. The reflection can be caused by other noise sources such as jangling metal objects or crisp packets [39]. In consideration of the before mentioned, ultrasound is not an appropriate positioning system for job sites.

Infrastructure implementation effort

Since construction sites are not forever, installing a positioning system should not be too costly and time-consuming.

RFID systems require numerous infrastructure components, like tags, location antennas, location processors and servers. Those components have to be fixed in different locations, making the installation rather time-consuming [39]. Yet, the amount of infrastructure needed depends on implementing an active or passive system [35]. An infrastructure-less data collection process is possible, though, when combining RFID with GPS. Torrent and Caldas [37] show an approach where a GPS anchor determines its coordinates. In contrast, the connected RFID receiver simultaneously detects the RFID-tagged components at the construction site. Based on the data collection and specific locating methods positioning the tagged materials were possible. Implementing an RFID infrastructure on job sites might be too expensive and too much effort, so a combination of RFID with another technology might be a working solution.

GPS can be added to various devices or is already build in, like in smartphones or similar apparatus [38], [39]. Infrastructure is only needed if GPS technology is combined with other positioning systems [35]. In contrast to other locating systems, GPS provides an optimal solution for construction sites regarding framework implementation.

Bluetooth chipsets are low cost. The BLE positioning systems deploying effort depends on the user's needs. For minor purposes, it can be installed with minimum infrastructure [50]. Yet, an infrastructure with fixed access points in different places is needed [39]. As it is reasonably low effort setting up a BLE system, it can be applied on construction sites.

The UWB technology depends upon sensors in fixed locations. The targets tags are wireless and lightweight, and a LAN connection is required [35], [36], [39]. As UWB requires a dense network of stationary receivers, coming with large-size hardware and no smartphone support, it is expensive [37], [38]. All of these aspects above are deficient for construction sites, making the UWB positioning system ineligible on this matter.

Vision-based positioning systems do not require tags. However, a camera infrastructure is needed [35], [39]. The cameras might be low cost. Still, it has to be considered that arranging tasks and calibration are time-consuming tasks [36]. Compared to the implementation effort of other locating systems, the camera-based system comes in small, making it a considerable choice for construction sites.

When there is an existing Wi-Fi infrastructure, the WLAN positioning technology can reuse the network to track targets [39]. Locating can be done by either tracking tags, or smartphones, which comes both with reasonable cost [35], [36], [38]. However, Wi-Fi infrastructure is frequently not available at construction sites, making the above benefits obsolete. In conclusion, the suitability of WLAN systems depends on the existing network structure. Assuming that most job sites don't have an extant infrastructure, WLAN-based approaches are not the most convenient methods.

The ultrasound positioning system comes with a costly and time-consuming installation, as many receivers are required to cover large areas for tracking multiple targets [39]. Due to deployment and maintenance over the lifetime, ultrasound technology is high-cost [49]. Ultimately, the ultrasound systems are not the most fitting real-time locating solution for construction sites because of the high implementation effort.

Table 4.2: Suitability of positioning technologies on construction sites

	Dynamic Environments	Line of Sight	Existing Network	Dynamic tracking	Site coverage	Accuracy	Interference
RFID	0	+	−	+	−	+	+
GPS	−	−	−	+	+	−	−
UWB	−	+	−	+	0	+	+
Vision Analysis	−	−	−	+	0	+	+
WLAN	0	+	+	+	0	+	+
Bluetooth	+	+	+	+	0	+	+
Ultrasound	−	+	−	+	0	+	−

Summary

After looking at the most critical aspects of construction sites, it has been pointed out that every positioning solution has its downsides. However, some systems have been proven to be more suitable than others. GPS sensors are not a practical solution due to the simple fact of missing indoor positioning capabilities. High privacy restriction laws make it impossible to place cameras and vision analysis systems on the job site. WLAN and ultrasound technologies are an exciting approach, yet the lack of finished market products encumbers acquiring an interconnected system. Both technologies are under ongoing development, and with increasing demand in construction monitoring systems, future improvements are to be expected. Today, the WLAN and Ultrasound positioning systems are not yet market-ready and usable for our purposes.

RFID, UWB and BLE are the most promising monitoring technologies. Even though UWB is a positioning technology that has demonstrated its worth in other industries, it doesn't come into consideration due to the following reason: The UWB solutions on the market are targeting large companies with much higher capital. This doesn't meet the demand of this research project.

The availability of many more BLE than RFID positioning systems on the market makes BLE a more attractive solution. Further, BLE systems are complete sets with better working components. According to the above mentioned, we decided to go for a BLE positioning system.

4.4.3 Positioning Methods

Some basic principles are needed to determine the position of an object. First of all, it is important to determine on which basis the positioning is to be carried out and which parameters are required. In the following, different positioning methods are therefore described.

RSSI distance estimation

The first parameter, which determines how close an object is to its receiver, is a signal-strength indicator called the RSSI value. It measures how strong the signal from the target to the receiver is. While the signal strength is highly dependent on how far a target is away from the receiver, this relation can be used to estimate the distance between the two objects. Dong and Dargie [52] explain in their study how distance is related to the RSSI value. With a couple of experiments they came to the following equation:

$$d = 10^{\frac{RSSI-A}{-10*n}} \tag{4.1}$$

This is a well-known method in the positioning field and is thus used by many devices to estimate an object's distance quickly. However, this approach doesn't provide the most accurate result since it is highly dependent on the signal strength, which is in some cases, measured incorrectly. Consequently, the following chapter describes methods to overcome the interference issues.

Kalman filtering

When it comes to RSSI to distance estimation, an often related problem is noise [52]. A general approach to fix noise issues is adopting Kalman Filters. The Kalman filter is a linear, unbiased, and minimum error variance recursive algorithm to optimally estimate the unknown state of a dynamic system from noisy data taken in real-time [53]. It has found wide adaption within certain signal-strength based positioning approaches and thus provides further optimisation advantages. Ultimately, the filtered values lead to more precise positioning, making it indispensable for positioning methods.

Triangulation

Triangulation is a well-studied approach for positioning. It is also working with RSSI-based values, good examples are shown by Wang, Yang, Zhao, *et al.* [54] and Shang, Su, Wang, *et al.* [55]. Triangulation is a standard methodology when it comes to GPS tracking. Triangulation on a base level works the following way. At first, at least three fix-points are needed. The distances to the target device are determined based on the fixpoints. With the distances, a system of equations can be set up based on the circular equations of the individual fixed points. By solving this system of equations, the coordinates of the point can be obtained.

The circular equation is presented in the following way:

$$(x - x_n)^2 + (y - y_n)^2 = r_n^2 \tag{4.2}$$

Using this equation for all three points, we end up with the following formulas:

$$(x - x_1)^2 + (y - y_1)^2 = r_1^2 \tag{4.3}$$

$$(x - x_2)^2 + (y - y_2)^2 = r_2^2 \tag{4.4}$$

$$(x - x_3)^2 + (y - y_3)^2 = r_3^2 \tag{4.5}$$

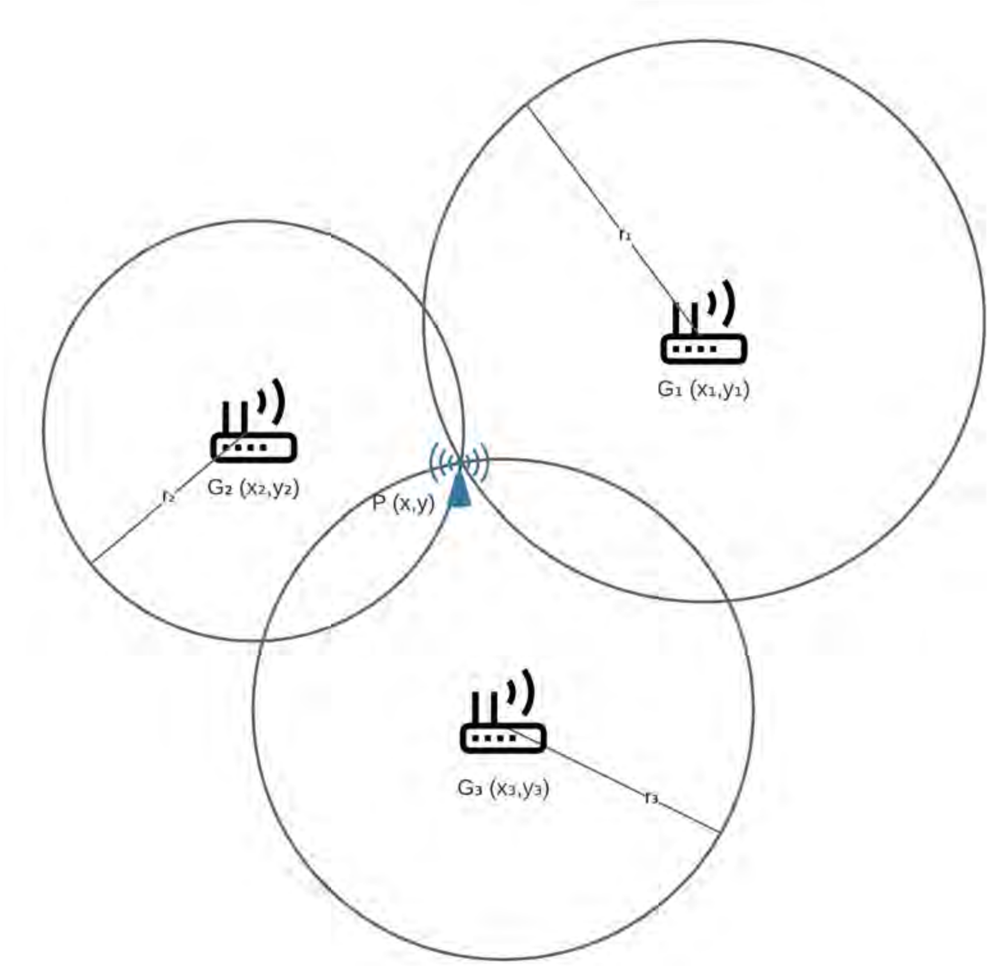


Figure 4.2: Triangulation method

Those formulas can be formed into a system of equations for estimating a point's x and y values.

$$2\begin{bmatrix} x_3 - x_1 & y_3 - y_1 \\ x_3 - x_2 & y_3 - y_2 \end{bmatrix} \begin{bmatrix} x_u \\ y_u \end{bmatrix} = \begin{bmatrix} (r_1^2 - r_3^2) - (x_1^2 - x_3^2) - (y_1^2 - y_3^2) \\ (r_2^2 - r_3^2) - (x_2^2 - x_3^2) - (y_2^2 - y_3^2) \end{bmatrix} \tag{4.6}$$

Fingerprinting

Fingerprinting is a method often used for indoor positioning. The method can be split up into two parts: Offline and Online. First, a site survey has to be done to collect sample signal strength data of the grid points. The measured RSSI values represent the fingerprint of the knot. The second step is the online phase. Once the grid of points is filled with fingerprints, sample vectors can be measured. The measured point is then compared to the existing grid of points stored on a database, and ultimately, the closest fix-point will be returned. A simple way to evaluate the nearest spot is using the Euclidean norm [56].

Summary

This subsection pointed out the most familiar positioning methods within indoor positioning. However, there are also other positioning techniques, which weren't covered here. Yet, using triangulation and fingerprinting methods, suitable results can be gained, making these methods enough for our research purposes. With these techniques, specific processes can be monitored and thus extend the DM of the as-performed construction process.

While all of these methods seem to be promising data sources for construction sites, the data still has to be handled. The following section will analyse what challenges might occur during data inspection and the overall DM process in the construction industry.

4.5 Challenges of Data Mining in the construction industry

One of the enormous burdens of DM in the construction industry is **data security**. Indoor environments can provide great insights for understanding human behaviour. However, they are frequently hard to access. Thus, it is vital to inform participants early and keep data perturbed and encrypted. Another challenge, similar to general DM, is **poor data quality**. Moreover, data often contain misleading values, human errors and machine errors. Hence, a lot of time needs to be sacrificed for cleaning up incorrect data. When it comes to interpreting extracted patterns of data, there is often a problem with knowledge interpretation. This difficulty is closely related to the small availability of skilled professionals. Ultimately, the efficiency of the DM process has high improvement potential while increasing the analysis of the received results. In many research projects, it has been pointed out that proposed models were **missing validation**. As construction projects are often surrounded by **different contexts** (location, climate, project type, etc.), transferring the acquired knowledge to other projects is problematic. Most studies require more cases to obtain more reliable and convincing results [26].

Chapter 5

Case Study: Monitoring the as-performed construction process

5.1 Building Lab: State of the art spaces for the future of construction

The Data Mining research project is related to collaboration with the Innovationsmanagement Bau GmbH and the Bayrische Bauindustrie e.V., making it possible to access specific construction projects and use them to gain data sensors. One project utilised during this master thesis is the Building Lab in Regensburg, explained in this section. The

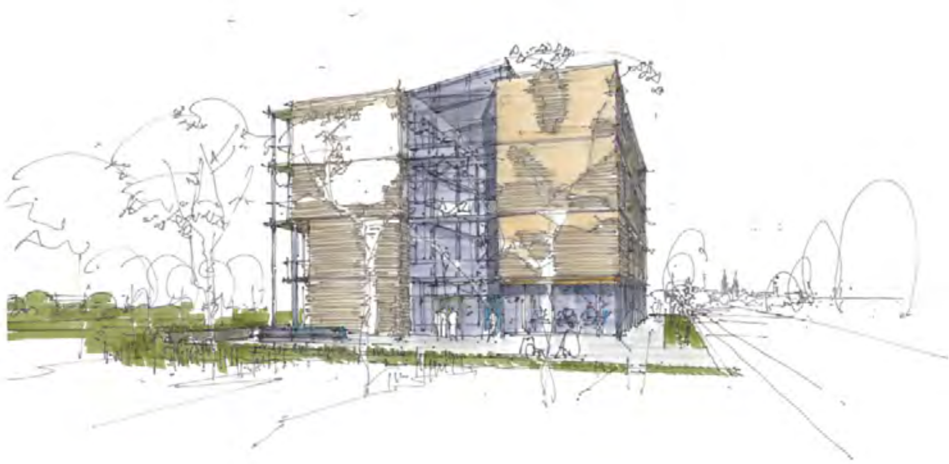


Figure 5.1: A draft of Building Lab Regensburg [57]

Bayrische Bauindustrie e.V. plans to build a centre for innovation in Regensburg, called the Building Lab, extending the Tech Campus of the University of Regensburg. The building symbolises a direct connection between science and the ACE industry. It provides space for demonstration and development of the latest building technologies, education, and a dormitory for fellow research. Presented building technologies are AI, 3D printing and BIM. The Building Lab is located in the centre of one of the most vital construction know-how regions in Europe. Nowhere else are so many such strong construction companies so close together located as in Eastern Bavaria [57]. There already have been many student projects involved with this project, showing its importance [57].

Ultimately, this project provides a perfect opportunity to test research prototypes and collect essential data. Since this is an actual construction project, we have access to all project-related documents, like the BIM model shown in fig. 5.2 and fig. 5.3. The before

Figure 5.2: A draft of Building Lab Regensburg [57]

illustrated BIM data can be extracted to create a digital twin explained in the following section.

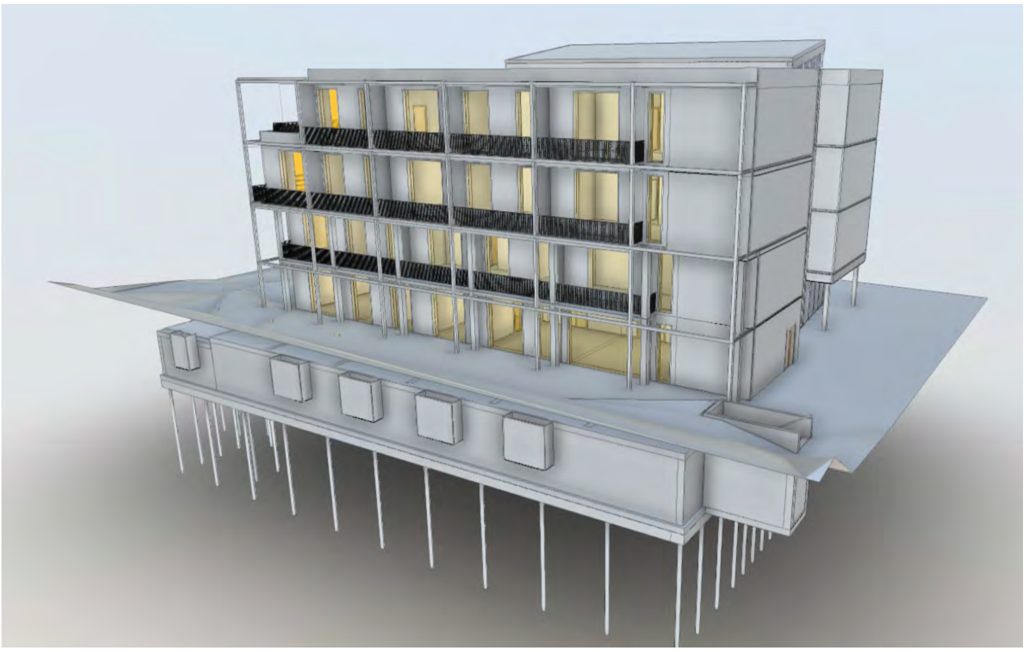


Figure 5.3: The BIM model of Building Lab Regensburg

As explained in section 2.2.3 the DM process according to the CRISP model starts with Business Understanding. Therefore, data sources have to be found and implemented on the site project. In this chapter, many methods are demonstrated, which provide possibilities to acquire data from construction sites.

Figure 5.4: The BIM model of Building Lab Regensburg

5.2 Laser scanning

As shown in section 4.3.1 laser scanning is an essential part of capturing the as-performed construction process and thus also part of DM. Therefore, we decided to perform scans of the site project every month.

For capturing the current state of the construction project, a Faro X130 scanner, shown in fig. 5.5 was used, which is a not too heavy, easily portable system. Each time a concept had to be developed before a scan could be carried out. This includes where the scanner would be placed and at what frequency the scan would be conducted. While the accuracy and frequency of the scans greatly influence the duration of the whole process, they must be considered in advance.

Figure 5.5: Laser scanner

In the following, the first scan is described as an example. Since little of the building exists at the beginning of the construction project, it is relatively easy to scan. However, as the building grows and becomes more complex, the number of scans increases dramatically, making the scanning process much more difficult. By conducting laser scans from five different locations, shown in fig. 5.6, we were able to scan the complete construction site. The laser scanner then prepossesses the points. With specific software projects of Faro, we eventually were able to reference a point cloud of the construction site state with over 72 million points. The first result is displayed in fig. 5.7.

Laser scanning is carried out during the project and thus provides a good overview of the overall course of the project. The recording of the actual state makes it possible to do various statements later, such as deviations between plan and actual, placement of materials and changes made during the execution phase.

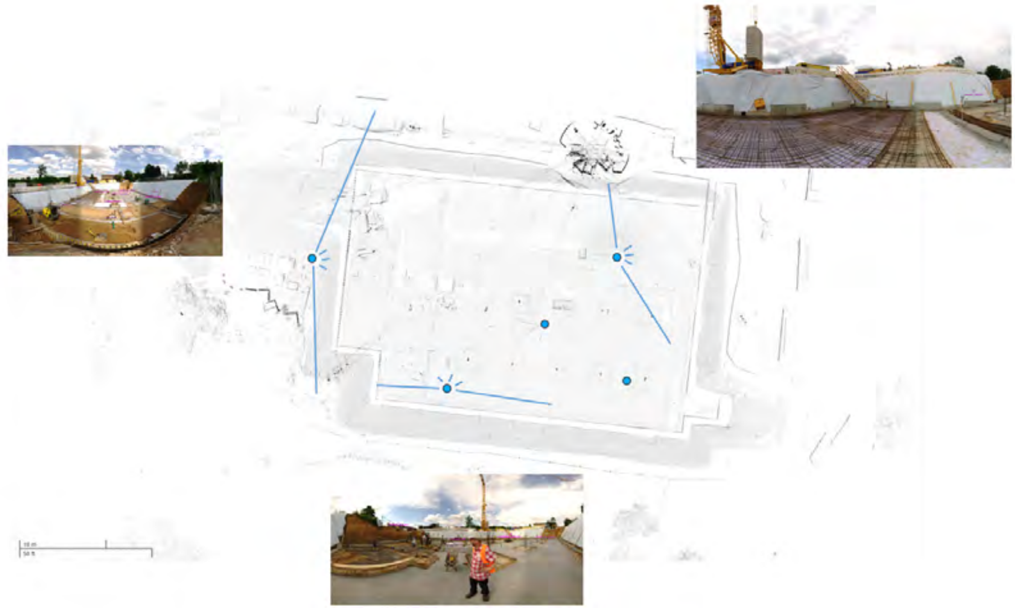


Figure 5.6: Five-point scan with Faro X130

Figure 5.7: Point cloud of laser-scanner after referencing

Since it already has become clear that the Laser-scanning approach is very time consuming, we decided to extend the construction state capturing with photogrammetric point clouds. Creating image-based point clouds has the downside of not receiving accuracy to mm-scale, but can be performed way faster, and therefore is also a promising approach, which will be investigated in the next section.

5.3 Image-based monitoring

Similar to section 4.3.1 the current progress of a construction project can also be monitored with photogrammetric approaches. Considering the Building Lab site, we decided to go for two different approaches. The drone scans are also followed every month.

5.3.1 Drone monitoring

First, the drone provides the flexibility to see the construction site from a bird's eye view. Among other things, this makes it easier to inspect all processes and to capture the construction site fully. With a certain number of images, point clouds can be generated using the approach described in section 4.3.2. To not take pictures permanently, it makes sense to create a video of the entire flight and extract the pictures from it afterwards. The images can then be used to create a photogrammetric point cloud.

With this approach, there is another possibility to record the current status of the construction project, which makes it possible to achieve comparisons soon. Furthermore, the number of data increases, making the DM process's modelling step much more precisely.

The first result of a photogrammetric-based point cloud is shown in fig. 5.9.

Figure 5.8: Monitoring the construction site with a drone

Figure 5.9: Point cloud created of drone

5.3.2 Camera monitoring

Second, we decided to place crane cameras to monitor the construction progress continuously.

The fig. 5.10 shows how the cameras were placed. Three cameras are located at different angles of the crane, while one is positioned on the neighbour building. The wide-spread placement allows capturing the complete site project and therefore monitoring the construction process in much greater detail.

Industrial climbers were hired to place the cameras, as shown in fig. 5.11. They fixed the cameras in the appropriate places and laid the LAN cables to the cameras. After aligning the system and setting up the individual cameras, the entire system was ready for use.

The cameras take a picture every thirty seconds and send it to an File Transfer Protocol (FTP) server hosted by us. Sample images of the cameras are shown in fig. 5.12, fig. 5.13 and fig. 5.14. The collected set of images can then be used to create a point cloud later in the project, providing capturing the as-performed construction status. Further, the pictures can also be adopted for other use-cases, such as checking the location of materials, detecting specific processes and validating the BLE positioning system, which will be analysed in the following.

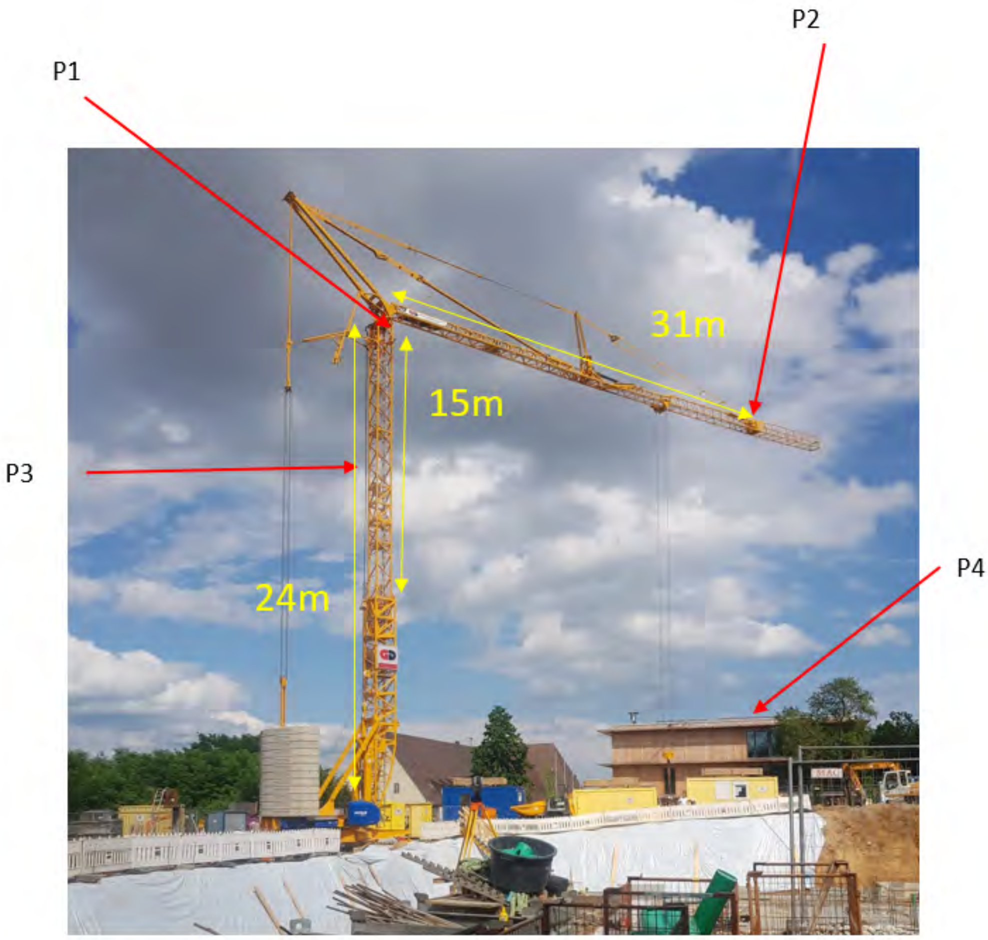


Figure 5.10: Positions of crane cameras

Figure 5.11: Industrial climber placing the crane camera

Figure 5.12: Sample image of crane camera 1

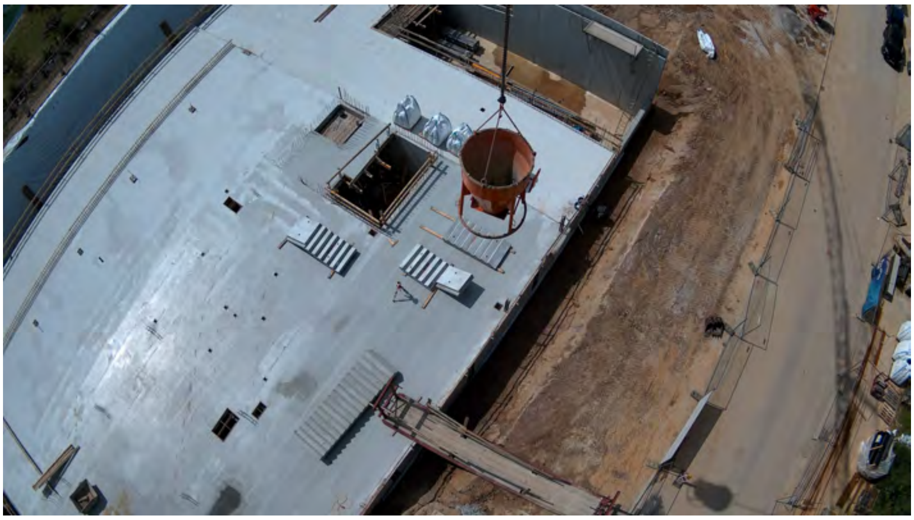


Figure 5.13: Sample image of crane camera 2

Figure 5.14: Sample image of crane camera 3

5.4 Positioning system

Eventually, a further step to monitor individual construction processes is frequently checking where materials and workers are located. Thus, the approach mentioned before requires a positioning system, which can perform under construction site conditions. Shown in section 4.4.2, creating and implementing a positioning system can be time-consuming and expensive. However, it also became clear that there are no ready-to-use solutions for tracking objects as the state of now. This research niche is leading to a couple of questions, which are required to be solved before a suitable positioning system can be assembled on the actual construction site:

- Which components are needed to enable location tracking of both workers and materials?
- What are the limitations and challenges of using a positioning system?
- How is positioning data handled?

The next chapter will consequently inspect the topic of creating and implementing a positioning system.

Chapter 6

Case Study: A positioning system used for Data Mining within the as-performed construction process

6.1 Boundary conditions of the case study

As pointed out in section 4.4.2, implementing a positioning system on construction sites can be challenging. More, it has become clear now that many positioning systems are not suitable for construction sites.

Nevertheless, it has also been shown that some systems can certainly prove their worth on construction sites. A factor that should not be neglected is the cost of the research project. The research project, presented in chapter 5 has a limited budget, so the cost of the positioning system plays an important role. Some firms offer complete BLE or UWB systems, but firstly on an end-user-oriented level and at a price that only large companies can afford. Latter is due to the hardware required and the software solutions needed, which must be implemented and maintained. Another criterion to achieve good results of positioning analysis was the challenges that arise with data mining methods. Data mining requires, as described under section 2.3 the highest possible data quality to gain meaningful results. Therefore, it is vital to check the system's data quality in advance to support later data preparation.

Within section 4.4.2 some systems were described that are suitable for the construction site conditions. However, as discussed in section 4.4.2, the choice fell on one system due to significantly differing costs of the various system hardware. The chosen BLE system is described in more detail below.

6.2 Creating a BLE positioning system

According to the challenges explained in section 6.1 a couple of steps need to be taken to make this project work. First, suitable BLE hardware had to be chosen and configured. Second, software that can handle the information of the hardware had to be implemented, full-filling all requirements for DM, explained in section 2.3. The third and final step involved testing the system, making it useable for construction sites. Therefore, we decided to test the positioning system in a well-known environment before taking it on the construction site.

6.2.1 System design

In general, the structure of a BLE system can be displayed quite straightforwardly. The system contains four key components: Beacon, gateway, IoT server, and database server.

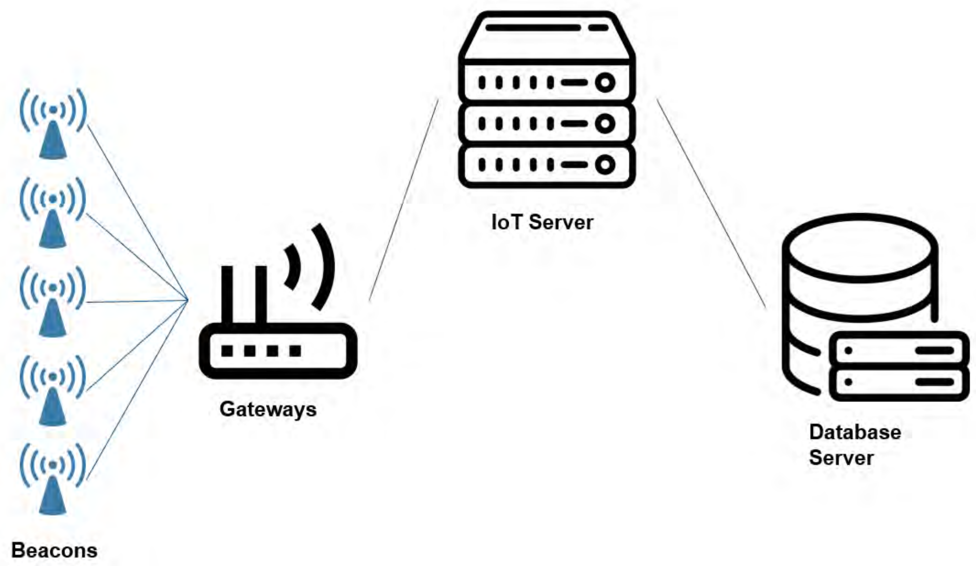


Figure 6.1: Structure of the implemented BLE system

The **beacons** are in constant communication with the gateways and correspond to the "senders" of the system. More, the beacons are meant to be carried around by the site workers or be attached to specific materials. There are different beacon types for different use-cases. Some beacons can measure, for example, signal strength, temperature, acceleration and more. A beacon continuously communicates with all gateways which are in range. They also have different protocols which can be used to gain a certain kind of data. As we need to protect personal data, we decided to make anonymisation as high as possible by attaching the beacons to helmets. Since the selection among the workers of the construction helmet is made randomly, there is no possibility to assign the beacon to a person. Beacons can be used for different use-cases, like measuring temperature or acceleration. For the described use-case, however, it is most important to figure out a worker's position. Therefore, the beacons continuously send a protocol with the signal strength (RSSI) inside.

The **gateway** creates a Bluetooth 5.0 network, which allows communication for all BLE devices. The gateways are permanently fixed at precise locations. In a predefined time interval, the gateways are receiving messages from the beacons. These messages get combined into a gateway protocol which is then sent to an IoT server. The gateway ultimately represents a communication bridge between beacons and IoT servers. Using the RSSI-value of the beacon corresponding to the gateway, we can calculate the distance between the beacon and the gateways. With several gateways, positioning can be carried out.

The **IoT server** builds the communication bridge between gateway and database server. The protocols from the gateways are first sent to the IoT server, where they get processed and then get forwarded. The communication happens via status requests, like Message Queuing Telemetry Transport (MQTT) requests, which are sent from the database server.

The **database server** is responsible for processing, transforming and storing the received data. More, it gains values like timestamp, signal strength, the beacon's and the gateway's Media Access Code (MAC) address for later allocation. Each database entry refers to the specific gateway, which it was sent from. A query can finally be sent to the database server to receive the exact location at a particular time.

The described system can store data and positions for the later DM step, but it also checks the positions in live time. Even though real-time monitoring may not be needed for data analysis in the first step, it can be very useful for calibrating the system and prechecking the results.

6.2.2 Creating a data pipeline for receiving protocols

As mentioned in section 6.2.1, the data from the beacons gets via gateways and IoT server send to the database server. Therefore a data pipeline including software infrastructure was needed. The software framework was implemented by using Microsoft Visual C# in the .NET Framework 4.7.1 environment. Following steps for running a data pipeline are needed:

1. Registration of beacons
2. Connecting the gateways to the IoT servers
3. Requesting status updates from the IoT servers

Registration of beacons

The beacons can be identified via a MAC address. The MAC address is a unique combination of letters and numbers, also registered in the gateways protocol. Before enabling a beacon, it is crucial to store the MAC address. A common way to save such information is to use JavaScript Object Notation (JSON) configuration files, which the program can read.

Once the beacon is connected to the gateway and protocols are sent to the IoT server and stored at the database server, the beacons can immediately be identified by the MAC address.

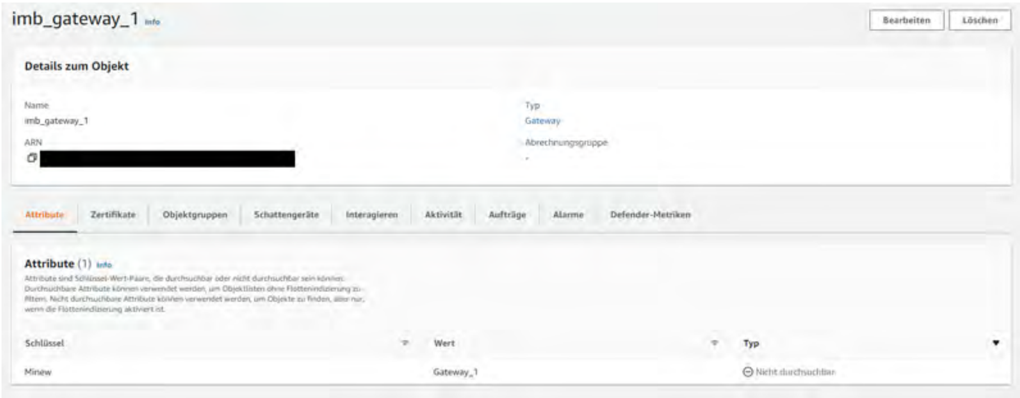


Figure 6.2: The interface of the AWS IoT service

Connecting the gateways to the IoT server

The second step involves connecting the gateway to the IoT server. Therefore, two things are needed: An IoT server endpoint domain and a safety certificate. Both can be added via the gateway device panel, which can be accessed in the same local network.

A well-known IoT server provider is Amazon, with Amazon Web Service (AWS). One of the main advantages of using such a cloud computing service is avoiding investment and server costs. Further, it provides high flexibility and existing IoT infrastructure, including accessible IoT endpoints [58]. Since creating a stand-alone IoT server requires expert knowledge and time-consuming effort, those server providers present an excellent alternative.

On the AWS interface, shown in fig. 6.2, you first register an IoT device and then use the given IoT endpoint domain to create a connection between the device and the server. The connection only works with a security certificate, which must be downloaded via the AWS service and stored to the gateway. The certificate is also required later to accept the status requests from the database server. When both steps are done, the gateway can continuously send protocols to the IoT server.

Requesting status updates from the IoT servers

The third part of the data pipeline covers the aggregation of the database server. As already mentioned in section 6.2.2 the software handling the database also must have access to the security certificates. Thus, all certificates are loaded to the program once the database server is started. Then, the database server can receive the protocols from the IoT server by sending a status request. The database server needs to be connected via a client via the IoT endpoint to send requests. Therefore, it uses the IoT endpoint domain generated on the AWS server. The IoT server sends an MQTT protocol back to the database server when the status request is accepted. The following code section represents all before mentioned steps:

1. Providing security certificates

2. Connecting via client to the IoT server

3. Subscribing on status updates from the IoT devices

Connecting to the IoT server

```
var caCert = gateway.CaCertificate;
var clientCert = gateway.ClientCertificate;
var client = new MqttClient(iotEndpoint, brokerPort, true, caCert,
    clientCert, MqttSslProtocols.TLSv1_2);
var clientId = Guid.NewGuid().ToString();
client.Connect(clientId);
client.MqttMsgSubscribed += Client_MqttMsgSubscribed;
client.MqttMsgPublishReceived += Client_MqttMsgPublishReceived;

client.Subscribe(new[] { gateway.TopicStatus }, new[] { MqttMsgBase.
    QOS_LEVEL_AT_LEAST_ONCE });
```

After that, the database server processes the protocol. The processing step will be explained in more detail in the following section.

6.2.3 Processing gateway protocol data

Once the MQTT protocol reaches the database server, it has to be processed first. This is related to certain aspects. Firstly the structure of the IoT protocols is tremendously nested, which makes it impossible to access the required information. Secondly, some data stored inside the protocol, like the beacon UUID or battery, is unnecessary for the positioning process. Hence, to avoid saving unnecessary data, this data has to be excluded. However, the biggest benefit of creating a self-implemented structure is that it covers data understanding and preparation steps within the DM process. Fundamental information can quickly be found and evaluated, attributes like date and time are correctly assigned, and additional elements like rough distance estimations can easily be added.

JSON sample protocol


```
{
  "topicFilter": "/gw/ac233fc0a70c/status",
  "qos": 0,
  "messages": [
    {
      "format": "json",
      "topic": "/gw/ac233fc0a70c/status",
      "timestamp": 1629117957502,
      "payload": [
        {
          "timestamp": "2021-08-16T14:45:56.894Z",
          "type": "Gateway",
          "mac": "AC233FC0A70C",
          "gatewayFree": 87,
```

```
                "gatewayLoad": 1.2
            },
            {
                "timestamp": "2021-08-16T14:45:56.894Z",
                "type": "iBeacon",
                "mac": "AC233F73C88B",
                "bleName": "",
                "ibeaconUuid": "E2C56DB5DFFB48D2B060D0F5A71096E0",
                "ibeaconMajor": 0,
                "ibeaconMinor": 0,
                "rssi": -66,
                "ibeaconTxPower": -59,
                "battery": 0
            }
        ]
    }
...
}
```

Converting MQTT protocol data

Since MQTT protocol supports JSON file formats, data can be extracted similar to JSON file reading. First, the received JSON string must be deserialised and converted to a machine-readable Protocol object.

Deserializing MQTT protocols

```
List<Protocol> protocols = JsonSerializer.Deserialize<IEnumerable<
    Protocol>>(json).ToList();
```

After all the information is accessible, further structuring can be done. In the first step, we are assigning all relevant information to a GateWay class. This includes the internal GatewayID, MAC, Date, Time and ConnectedBeacons of the Gateway.

Gateway class

```
public class GateWay
{
    public int GatewayID { get; set; }
    public string MAC { get; set; }
    public string Date { get; set; }
    public string Time { get; set; }
    public IEnumerable<Beacon> ConnectedBeacons { get; set; }
    ...
}
```

All information regarding the beacon gets similarly stored. The beacon also contains a MAC address, the corresponding RSSI value, the tx power (calibration value of the beacon), and a rough distance estimation to the gateway.

Gateway class

```
public class Beacon
{
    public string MAC { get; set; }
    public double RAWRSSI { get; set; }
    public double TxPower { get; set; }
    public double RawDistance { get; set; }
    ...
}
```

With both classes and the generated protocol, gateway entries can be generated.

Generating gateway entries

```
var dataManager = new DataManager(jsonReader.GatewayProtocol);
var gatewayPublished = dataManager.TransformProtocol();
```

And are ultimately stored in a database.

Store to database

```
dataprocessing.PushToDatabase<GateWay>(gatewayPublished);
```

Distance estimation based on RSSI values

As already seen in section 6.2.3 a rough distance estimation gets stored to the beacon class. Yet, the gateway protocols are not providing a distance value per se. The RSSI distance estimation, seen in section 4.4.3 is a well-known approach to roughly estimate the distance from the RSSI value.

The RSSI values can be extracted from the protocol, making distance conversion a suitable approach. The "txPower" variable describes the calibration of the beacon, which depends on the indoor environment. The "n" variable describes the loss of signal and is also depending on the indoor environment. Adapted to the C# environment, the formula, explained in section 4.4.3, looks as shown below.

RSSI to distance

```
public static double Distance(double RSSI, double txPower, double n)
{
    return Math.Pow(10, (rssi - txPower) / ((-10) * n));
}
```

This method is called while generating the beacon class and gets immediately assigned to each beacon. As it might not be the perfect way to estimate a precise position, it is a good reference indicator, which can later be used to post-processing the estimated location.

Storing gateway entries

Looking back at section 6.2.3 storing gateway entries to a database was briefly discussed. This will be inspected more precise now. In section 2.1 a short introduction about NoSQL databases was given. MongoDB is a common NoSQL database used in this project. The following shows how creating a database and pushing objects to it works.

With integrated NuGet packages, the MongoDB database can be built relatively quickly. First, a client has to be created. Then the client can make or get the database.

Creating a database

```
public void StartMongoDB(string name)
{
    var client = new MongoClient();
    this.database = client.GetDatabase(name);
}
```

Once a database is built, collections can be created or accessed by a string variable. With the assigned collection, data can be pushed to the database.

Creating a database

```
public void PushToDatabase<T>(T item, string collectionName)
{
    var collection = this.database.GetCollection<T>(collectionName);
    collection.InsertOne(item);
}
```

All received and converted protocols are stored in this manner. The database provides extensive storing capabilities and easy access via queries, which will be discussed in more detail in the following section.

Extracting data from database

For later processing and DM steps, the data needs to be extracted from the database. So-called queries can do this. If, for instance, a closer inspection on the fifth of August is needed between twelve and one a clock, a query would look the following way.

Query of MongoDB

```
{$and:[{Time:{$gt:"12:00:00" }},{Time: {$lt: "13:00:00"}}, {Date:
   "05.08.2021"}]}
```

All keywords start with a $ sign statement. An attribute of an object can be accessed by putting the MongoDB request in curly brackets. However, this is not the only way to get desired results, yet, it is also possible to access the database via the API. Still, the above-explained procedure stays the same.

Once the desired data is received, it can be stored in a JSON file, which the implemented software can read. The last step contains transforming the received RSSI values to actual points using triangulation and fingerprinting. The following section explains these processes in detail.

6.2.4 Generating Locationpoints

Kalman filtering of RSSI values

Since Kalman filters are a highly complex topic, as briefly discussed in section 4.4.3, we are using a library that provides the basic functionalities of Kalman filtering. It has been used quite commonly in the BLE environment and thus provides an optimal tool for decreasing the noise of RSSI data. The filter sorts out incorrect RSSI values and adapts the received values to get an overall more fitting result. The Kalman filter is applied in the following way:

Filtering RSSI values

```
public DataFilter(IEnumerable<double> measurements)
{
    var filter = new UKF();
    foreach (var m in measurements)
    {
        filter.Update(new[] { m });
        Estimations.Add(filter.getState()[0]);
    }
}
```

Once filtering is finished, the data filter will return the collection of filtered values. The cleaned list can be used for triangulation methods which will be explained in the following section.

6.2.5 Position estimation with triangulation

The formulas of triangulation, shown in section 4.4.3, were also applied in code, which is presented in the following way:

Generating a 2D point with triangulation

```
public Point GetCoordinates()
{
    var matrix = DenseMatrix.OfArray(new double[,] {
    {p3.X -p1.X, p3.Y -p1.Y },
    { p3.X-p2.X, p3.Y-p2.Y} });

    var vector = Vector<double>.Build.DenseOfArray(new double[] {
    (Math.Pow(dis1,2)-Math.Pow(dis3,2))-(Math.Pow(p1.X,2)-Math.Pow(p3.X
        ,2))-(Math.Pow(p1.Y,2)-Math.Pow(p3.Y,2)),
```

```
    (Math.Pow(dis2,2)-Math.Pow(dis3,2))-(Math.Pow(p2.X,2)-Math.Pow(p3.X
        ,2))-(Math.Pow(p2.Y,2)-Math.Pow(p3.Y,2))});

    var solve = matrix.Inverse() * vector * 0.5;


    return new Point(solve.ElementAt(0), solve.ElementAt(1));
}
```

Yet, the equation can be extended by adopting additional libraries to use multiple fix points. Even though this seems like a reasonable approach to get positioning done, the noise of the signal strength still causes some issues. As the distance estimation is not always 100% precise, it can highly influence the precision of the triangulation algorithms. For example, if the object seems to be closer to a gateway than it is, but the distances to other gateways stay the same, it ultimately gets miss-located. Since slight distance deviations have a high impact on triangulation methods, it makes sense to also look at an additional positioning approach.

6.2.6 Position estimation with fingerprinting

As explained in section 4.4.3, two steps need to be taken to perform the fingerprinting method. Looking at the described approach, this was carried out in the following way. First, a setup to measure specific points needed to be created and second, implementing a workflow for measuring and storing RSSI values had to be done. The first step just needed some creativity and was, in our case, performed by marking specific points on the ground and a chair as seen in fig. 6.3.

Implementing this into the BLE network requires a method to read in RSSI values from a specific location and a database. The RSSI values can be read in by using the in section 6.2.3 discussed procedures. Each grid point is scanned for a predefined amount of time. Then, the collection of RSSI values is flattened and filtered by a Kalman Filter, shown in section 4.4.3. Storing the values works the same as in section 6.2.3 explained. Finally, the calibration procedure is implemented according to the above mentioned.

Saving gridpoints for fingerprinting

```
var server = new AmazonServer();
var rssiFixPoint = server.StartCalibration(beacon);
point.RssiValues = rssiFixPoint;
SavePointToDataBase(point, beacon);
```

Once the calibration is finished, the second step of fingerprinting follows. First, the RSSI values need to be extracted from the database. The separation is accomplished within the step shown in section 6.2.3. Second, RSSI fingerprints must be compared with the grid points, which requires nearest neighbour algorithms. As already mentioned above, the

Figure 6.3: Fingerprinting setup

Euclidean norm is a common approach for finding the nearest value. By comparing the distances, the most suitable fixpoint can be evaluated by implementing the following:

Getting the nearest neighbour

```
public Point GetClosestPoint(double rssiValue, List<Point> fixPoints,
    int gatewayIndex)
{
    Point closestPoint = null;
    double closestDistance = 0.0;

    for (var i = 0; i < fixPoints.Count(); i++)
    {
        var fixPoint = fixPoints[i];
        var rssiFixPoint = fixPoint.RssiValues.ElementAt(gatewayIndex);

        var difference = Math.Abs(rssiValue - rssiFixPoint);

        if (i == 0 || difference < closestDistance)
```

```
            {
                closestDistance = difference;
            }
        }

        return closestPoint;
}
```

Depending on the scale of the grid, precise positioning can be done. Yet, it involves high effort and also may not always work due to RSSI noise.

6.3 Testing the implemented positioning system

We decided to use a room at the Technical University of Munich as the first test facility. The location allowed for quick calibration and testing due to easy accessibility and open cooperation with the involved people. The tested room is an open-space office of the university. A wide variety of sensor systems have already been tested within this facility, and the people working there are accordingly open to new ideas.

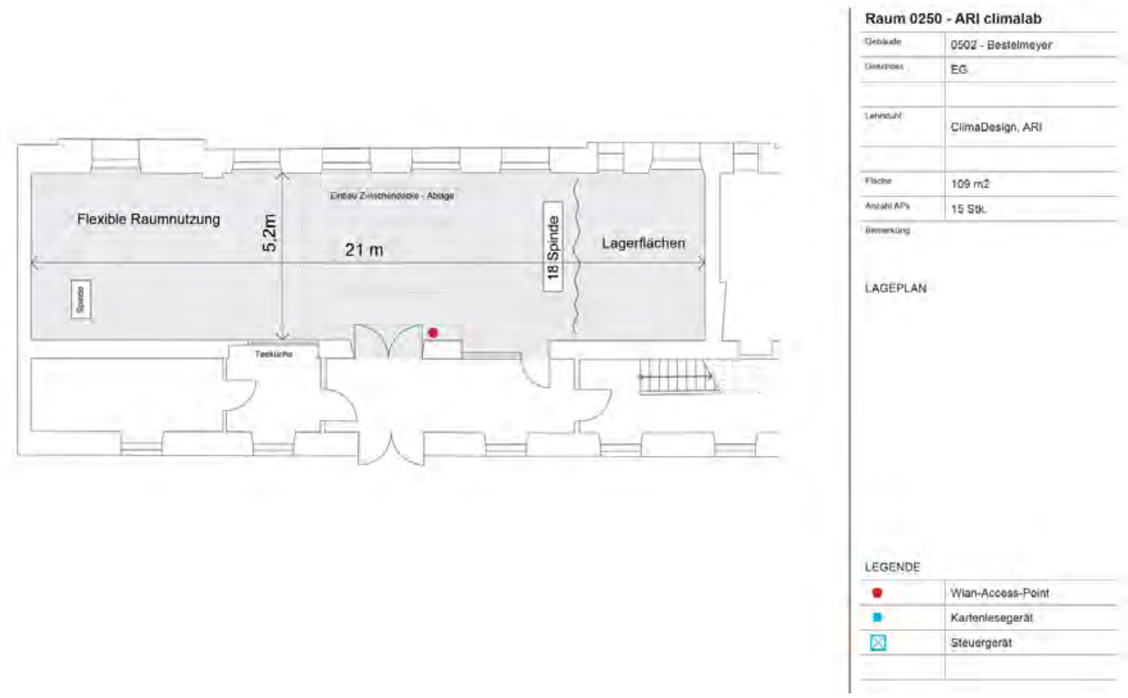


Figure 6.4: Plan of BLE system's test location [59]

The room, shown in fig. 6.4, has a total area of 109 square meters and is also providing obstacles, thus reflecting the open spaces representing the shell of the building site.

Two things were tested during the case study within this area. First, a long-term experiment was conducted to learn about the overall behaviour of the BLE system. The beacons were worn permanently by the people. The duration of the test was one month. The aim was to gain initial experience with the BLE sensors and to identify strengths and weaknesses in practice. The results will be analysed in the discussion chapter.

Second, an accuracy test was conducted by estimating the positions of a series of location points. The experiment was build up as follows: Two different beacons and three gateways were used for the test. The gateways were attached to the ceiling at known positions. A series of markings with known coordinates were fixed in a one-meter grid on the floor. The RSSI values for the fingerprinting method were then read in and stored in a database.

A series of measuring points were scanned at meter intervals to calibrate the system. The duration of each scan was one minute. The data collected was incorporated into a Kalman filter, resulting in the map, shown here fig. 6.5.

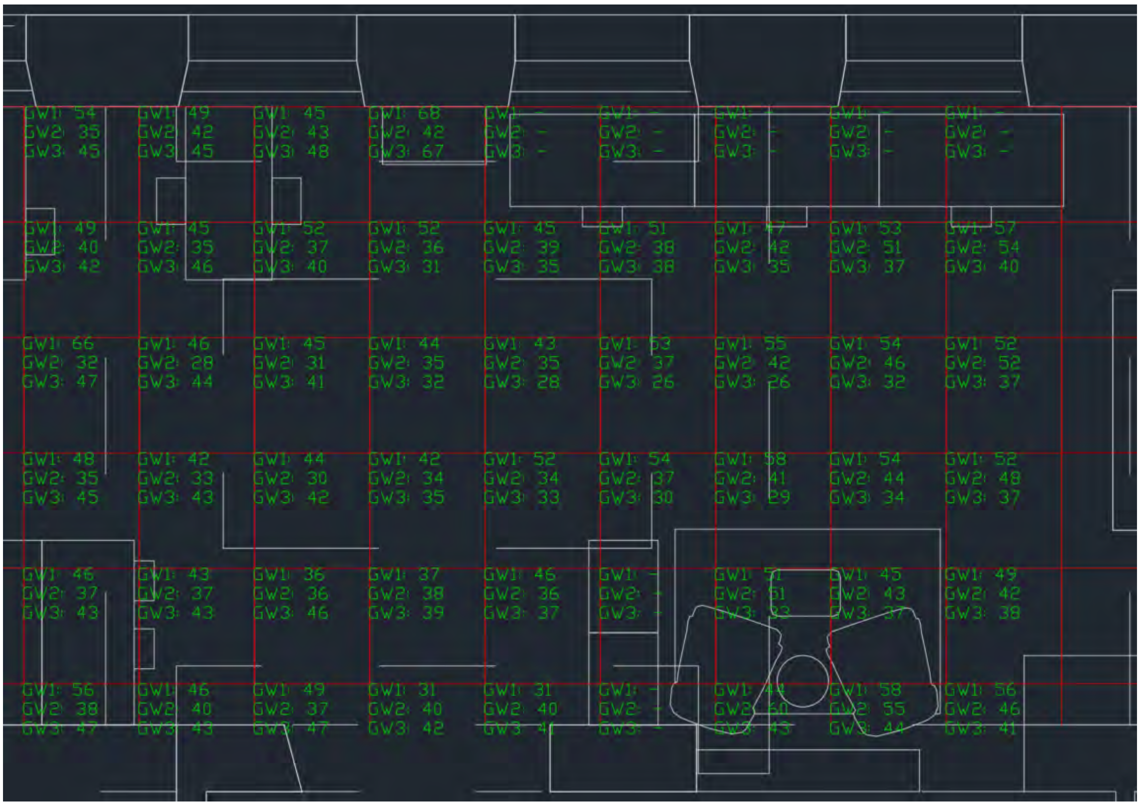
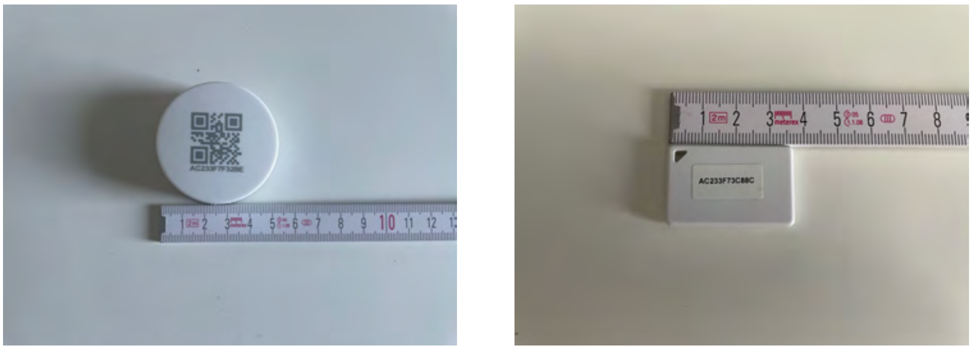


Figure 6.5: Created RSSI map of the investigated area

The RSSI map could then be used to determine where a beacon is located. Since there are different beacons for different purposes, a similar RSSI map had to be created for each beacon type.

By conducting the study, hundred randomly selected measuring points were obtained while using two different beacons, see fig. 6.6. One is a beacon that is more suitable for materials and offers a greater range. The other is a beacon that is small and portable and can be attached to a key chain.

Of the hundred positioning tests, fifty of the measuring points were determined using triangulation methods, discussed in section 4.4.3. The other fifty were determined using fingerprinting methods, seen in section 4.4.3.



(a) Material beacon (b) Person beacon

Figure 6.6: Different types of beacons

The tested points were randomly selected to ensure comparability between the different methods. The fingerprinting and the triangulation method was carried out once for each point. In each case, a beacon was placed on the marker and then scanned for half a minute. The RSSI values were seemingly recorded during this time and used to determine the points. The implemented program then outputs the determined point according to fig. 6.7.

A test criterion should further determine the recognition of movement patterns for workers and materials. For pattern recognition, it is sufficient for our research work if we achieve an accuracy of 1.5 meters, on which the results are eventually to be checked.

```
Successfully subscribed to the AWS IoT topic.
Connected to AWS IoT gateway 3 with client id: 0d385e81-a941-45ee-9cd1-d317d8781f

Successfully subscribed to the AWS IoT topic.
Finished calibrating.
Corrected RSSI 1:-46
Corrected RSSI 1:-46
Corrected RSSI 1:-35
FIXPOINT checked:
ID: 87
X: 14
Y: 2
Press 0 to calibrate system.
Press 1 to start tracking and enable RTLS server connection.
Press 2 to read file.
Press 3 to generate points from file.
Press 4 to check fixpoint.
Press 5 to leave program.
4
Choose Beacon to Calibrate:
Press 0 for beacon MAC: AC233F73C88B
Press 1 for beacon MAC: AC233F73C88C
Press 2 for beacon MAC: AC233F7F32B9
2
Calibrating Started.
Please keep beacon at the same position for 0,5 minutes.
AWS IoT Dotnet message publisher starting..

Connected to AWS IoT gateway 1 with client id: d9664b55-8746-4ae0-8355-e01a3e0887

Successfully subscribed to the AWS IoT topic.
Connected to AWS IoT gateway 2 with client id: 27c6d87b-5a17-4519-ab04-dc982c4f78

Successfully subscribed to the AWS IoT topic.
Connected to AWS IoT gateway 3 with client id: b8a434e5-4782-476a-8339-4c2cb46dbb

Successfully subscribed to the AWS IoT topic.
[#####-----]  57% \
```

Figure 6.7: Calibration window of the implemented positioning system

Chapter 7

Discussion

This chapter focuses on explaining and evaluating the findings to understand the meaning, importance, and relevance of the gained result during the case studies. More, this discussion gives a short retrospective on research questions pointed out in chapter 1.

7.1 Laser scanning

The results of the laser scans provide a precise representation of the actual state of the building site. More, point clouds include the outer shell of a building and all interior spaces and the infrastructure. The scanned object builds the basis for all other processes that include data related to the construction process. Regular scanning of the building site is therefore essential to be able to classify other data.

As mentioned under chapter 3, the individual construction processes are very strongly dependent on each other. Thus, the creator of the DM project must be aware of all input parameters to classify the data accordingly. An example of the strong dependency of various processes is the workflow of window installation. Construction workers can not place a window without a wall. The latter clearly emphasises that without the data basis of the existing walls, it is simply impossible to capture and evaluate the data of the interior finishing work, as the information is difficult to classify. In summary, laser scanning is an excellent method to record the actual state of the construction site and forms a critical data basis to make correct statements for future process evaluations.

However, laser scans also have their limitations. The time required for a scanning process increases steadily with the construction project's progress and the structure's complexity. Sufficient human resources and a sophisticated concept can optimise the scanning step to a certain extent. However, there is one more step left to transform the scanned information to point clouds. Referencing is a time-consuming process that increases with the number of scans and with the complexity of the building. Despite the high accuracy of the scans, the results should be treated with caution. The outcomes will be discussed in more detail below.

When looking at the first results of the laser scan, it quickly becomes apparent that the unprocessed data is challenging for further use. An example of this shows fig. 7.1.

Five essential points are highlighted on the graphic, are reviewed individually now. Point one shows that moving objects are duplicated or cannot be reproduced in detail. The second point illustrates that the detailed images of the laser scan can also be misleading. The exact surface structure cannot be identified precisely, which sometimes causes

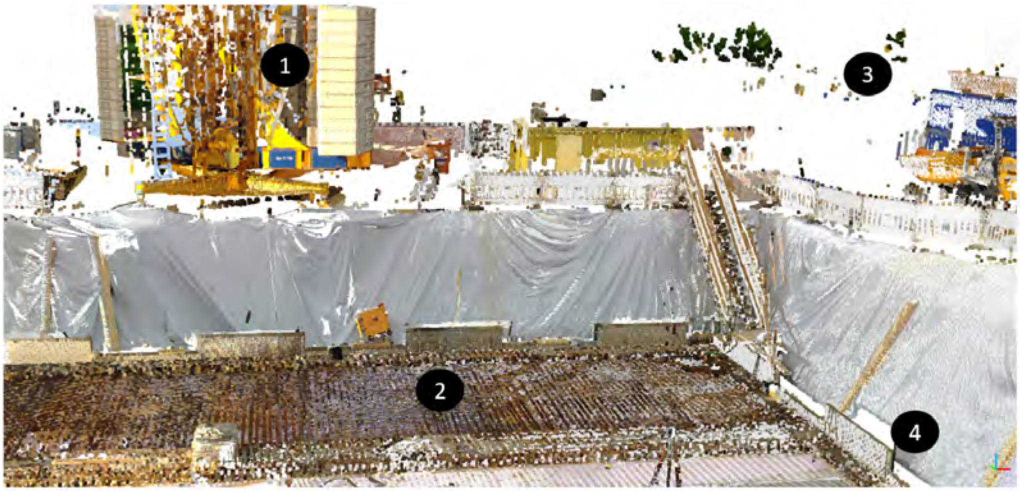


Figure 7.1: Point cloud result before data preparation

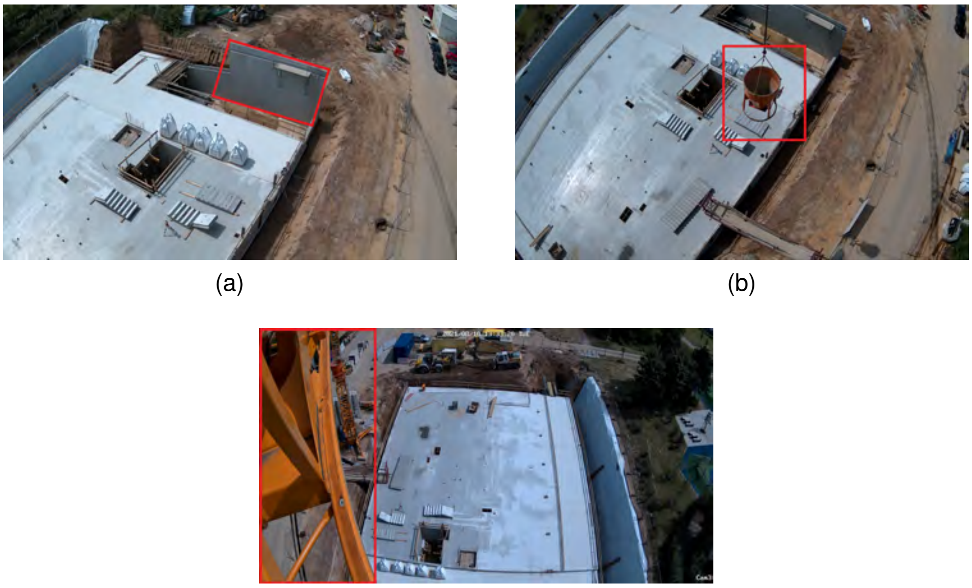
distortions. Another decisive criterion is found with point three: Objects of the surroundings are recorded unintentionally, which leads to noisy areas. Lastly, point four pictures that certain areas are covered by obstacles and thus cannot be recorded.

When considering all these problems, it quickly can be realised that such a model cannot be transferred unprocessed into the DM model building step. Such flaws must be corrected and processed to achieve correct results finally. It makes, therefore, sense to conduct additional data sources to be able to evaluate and sort out data. Additionally, the progress of the site is also covered by image-based monitoring methods, which was described in section 4.3.2 and will be discussed in the following.

7.2 Image-based monitoring

As explained in section 4.3.2, image-based monitoring is another approach to generate point clouds from the construction site and, therefore, also a way to capture the current progress of the construction project. While being less precise than a laser-scan based point cloud, the photogrammetric procedure has the significant advantage of being much faster. Indeed, more accelerated scanning means capturing a site can be done in several minutes by flying over the site with a drone. Similarly, the effort is relatively low considering the crane cameras. No additional effort needs to be taken despite placing the cameras and creating a data pipeline, which is already done.

Yet, both systems have their limitations, which is going to be discussed in the next paragraph. First, the drone can only capture areas visible from the sky, making it impossible to cover indoor areas. Having no indoor capturing option is indeed a big problem when looking at the Data Mining approach. As explained in section 7.1, the point cloud of the current construction state builds the basis of all upcoming process data. Many processes are in an indoor environment during the as-performed construction process. Those processes require knowledge of the ongoing inside construction state, which a drone or



(a) (b)

(c)

Figure 7.2: Crane images before data preparation

a crane camera cannot cover. More, it is essential to point out that all areas covered by obstacles are not included within the photogrammetric point cloud.

The following example investigates this problem in more detail. The different perspectives of the cameras, shown in section 4.3.2, have various issues. For example, the picture of the first crane camera fig. 7.2 shows that a part of the crane will be permanently in the picture. The overlapping part has to be rendered out afterwards while maintaining the image format. The second crane camera, shown in fig. 7.2, illustrates that the lifting cylinder will be permanently in the picture. Here, too, subsequent processing is necessary. With the third camera perspective fig. 7.2, there is not necessarily an obstacle in the picture. Still, the camera cannot capture particular objects entirely from above due to the lower position, which increases the proportion of obscured areas. Ultimately, similar to the laser-scanning data, the images were taken raw and must be post-processed.

Analogous to the laser scanning approach, the data of the photogrammetric methods are to be preprocessed to be used in upcoming DM steps. Eventually, combining both of these procedures makes sense to ultimately reduce errors and improve data quality.

In conclusion, different procedures for capturing the actual state of the construction site provide the broadest possible bandwidth of generated data. Further, the weaknesses of the individual methods balance each other out and thus build a good foundation for the analysis of different processes.

7.3 Positioning System

The positioning system was extensively tested, as described in section 6.3 by conducting hundred single positioning tests.

As two different kinds of beacons were used by different people in all experiments, shown in fig. 6.6, it makes sense to look at the triangulation and fingerprinting experiments individually now and later compare them. The beacons are defined as follows:

- Beacon 1: Location beacon, primarily used for material tracking
- Beacon 2: Tag beacon, primarily used for human tracking

At first, a closer look is taken at the results of the fingerprinting experiment. 72 % of the results measured with Beacon 1 achieved an accuracy of more than 1,50 meters. The median of all deviations was determined to be 1,00 meters, and the mean value had a value of 1,54 meters. Even though these are presumably exact results, only 44,4 % of the correct position were recognised by two or more gateways. Beacon 2 worked likewise to Beacon 1. 68 % of the results measured achieved an accuracy of more than 1,50 meters. Analogous to Beacon 1, the median of Beacon 2 was estimated at 1,00 meters. However, the overall spread of the results was slightly smaller, with an average of 1,49 meters.

Second, the results of the triangulation method are now analysed. Here, only 36% of the results conducted with Beacon 1 had an overall accuracy of 1,50 meters or better. The median of the deviation of the triangulation test with Beacon 1 was estimated at 1,68 meters, the mean value with 2,13 meters. Beacon 2 performed quite similarly to Beacon 1, showing that 36,00 % of the gained results had an accuracy of more than 1,50 meters. Yet, both the median and mean values were estimated slightly higher with average deviations of 1,84 and 2,13 meters. Ultimately the results are summarised in table 7.1.

The entire results of the fingerprinting test can be found in table A.1 and table A.2. The triangulation test results are shown in table A.3 and table A.4. In addition, the target and actual point ID, coordinate value and deviation for all fixpoints are listed in the tables mentioned above. A conclusive comparison shows that 64% of the results with fingerprinting methods were less deviant, and therefore more accurate, for both beacons.

Table 7.1: Results of fingerprinting and triangulation tests

	Fingerprinting		**Triangulation**	
	Beacon 1	Beacon 2	Beacon 1	Beacon 2
Accuracy <1.5 m	72 %	68 %	36 %	36 %
Median deviation	1,00 m	1,00 m	1,68 m	1,84 m
Mean Value deviation	1,54 m	1,49 m	2,11 m	2,13 m

Several general things became evident in the course of testing the long-term performance and the accuracy of the positioning system. First, the beacons have very small variations in their results and can therefore both be used without limitations due to different accuracies. Second, the fingerprinting results proved to be more accurate than the triangulation results in this test case scenario. Third, the precision of the system is high enough to monitor construction workers and materials sufficiently. However, there is room for improvement. Fourth, there were no significant deviations or inaccuracies in both methods. Fifth, the data pipeline functioned without any significant problems, even if the power or network supply to the gateways was interrupted for a short time. Sixth, the participants of the long-term haven't had any problems with carrying the beacons. More, by the time the involved people forgot they were carrying them. Seventh, the battery life of the beacons is not an issue. Eight, even though the database got over two million entries, accessing the database was possible without noticeable waiting times.

Eventually, an investigation on the ups and downs of the fingerprinting and triangulation method is done in the following. First, looking at the fingerprinting procedure, it became evident that the initial setup is way more time-consuming than the triangulation procedure. Each grid point needs to be measured and saved to a database which can take a while depending on the number of grid points. The dependence mentioned above also makes the fingerprinting method in extensive areas hardly implementable or less accurate when creating a coarser grid. Another downside of the fingerprinting approach is that it only covers the scanned grid area. Latter highly influences the accuracy of the system since a coarser grid will accordingly be less precise. However, the fingerprinting approach also showed many positive aspects, following now. First, the fingerprinting method does not need any high computational complex analysis procedures since it is based on measured values. Not having error-prone calculations prevents inaccuracies or mistakes taken from signal-strength deviations and neglected boundary conditions. Indeed, the previously mentioned problem of wrongly calculated results of triangulation procedures leads to another advantage of the fingerprinting method: higher accuracy by using a few gateways. Finally, fingerprinting provides no completely unusable and completely wrong values. The low susceptibility to errors is related to the possibility that having high deviations in fingerprinting is low because only values estimated before can be displayed.

Table 7.2: Advantages and disadvantages of the fingerprinting method

Advantages	**Disadvantages**
Computational effort low	Time-consuming
More accurate, when using smaller number of gateways	Difficult with extensive areas
No highly inaccurate results	Covers only grid area

Aside from the fingerprinting method, the triangulation procedure was also tested in the case study. The experiments showed the upsides and downsides of this procedure quite clearly. First, the results proved to be less accurate than the fingerprinting method, which is related to two circumstances. The first one is that triangulation methods reach their limits when RSSI values fluctuate more strongly, resulting in that in some places without an accurate result.

The second aspect is that triangulation strongly depends on the calibration value of the beacons (RSSI value at 1 meter) and an environmental variable, which can be influenced by many factors, as explained in section 4.4.3. In our test scenario, this proved to be worse for the triangulation method. Looking at the RSSI map, shown in fig. 6.5, it quickly becomes evident that the RSSI values fluctuate very strongly and are completely various in some zones. The high volatility makes it challenging to calculate the distance between beacons and gateways by specific rules and thus gaining precise triangulation results.

Interference can also lead to tremendous wrong results, e.g. if the signal is completely blocked by an obstacle, which is another disadvantage of the triangulation method. The cases illustrated above can be prevented by relying on fingerprinting results. Yet, the triangulation method has also a lot of advantages when adapted correctly. First, it is not time-consuming, as there is no need for measuring the positioning points before. Second, with a higher number of gateways and good calibration, triangulation can handle larger areas with presumably less effort. Having a scalable approach can be a major advantage when looking at the in section 4.4.2, described large construction site areas. Ultimately, the entire area can be covered by triangulation, making it independent from a grid.

Table 7.3: Advantages and disadvantages of the triangulation method

Advantages	**Disadvantages**
Easier to cover large, extensive areas	Highly depending on RSSI values
Can cover entire areas, independent of grid	Highly depending on System calibration
No highly inaccurate results	More error-prone

In summary, it became apparent that three gateways are not enough for both accurate triangulation and fingerprinting. More gateways are needed for both triangulation and fingerprinting positioning methods to achieve higher accuracies.

Nevertheless, the positioning system produced promising results during the long-term test scenario. fig. 7.3 shows a screenshot of a floor plan read in from Auto-CAD into which location points of a specific time-span have been imported. Visualising the data also enables identifying points of interest where people were more frequently present.

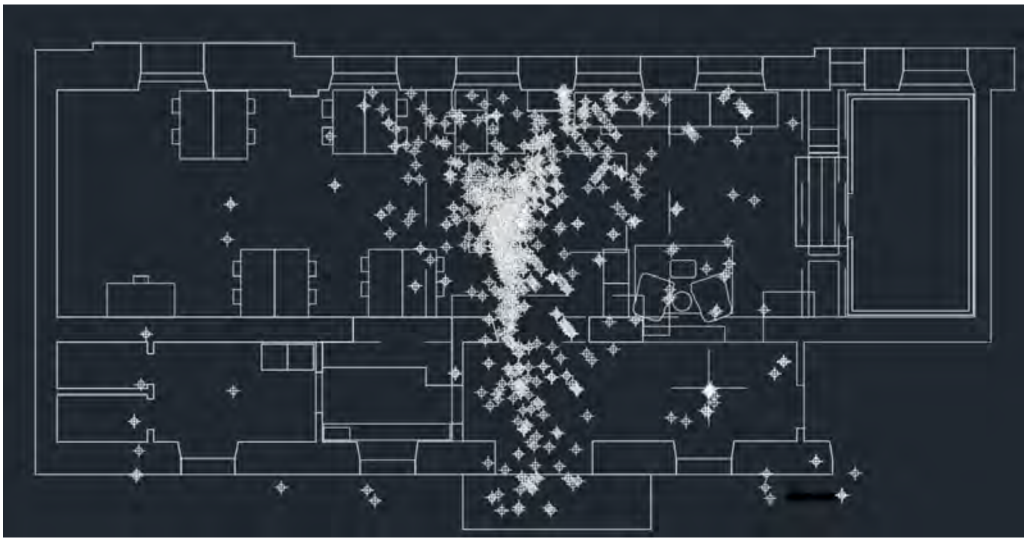


Figure 7.3: Results of a positioning system

With the collected findings, essential points can now be established that play a central role when using the BLE system on the construction site.

- Placement of the gateways and calibration of the system is crucial for later data quality
- Power and internet are essential

- A more extensive number of gateways is required for functioning triangulation and fingerprinting methods

All listed points can prevent several problems expected to occur on the construction site in the future. Still, it must also be considered that the system can only be fully perfected when used on the construction site.

Combining triangulation with fingerprinting

As discussed earlier, triangulation and fingerprinting have different weak points. However, both methods also have huge benefits when it comes to indoor positioning. Combining both methods makes sense to gain a certain degree of accuracy within highly complex indoor environments. Moreover, the compound of both helps determine incorrect values and further improves the system's stability. The latter can be implemented by using both methods to gain a point and then double-checking the result. If both estimated points are in the exact location, the calculated position is correct with a high probability. If both estimated points are not in the same location, the algorithm must investigate further. The system primarily relies on the fingerprinted values and thus obtains the exact position from the database. Still, the system gets controlled by additional boundary conditions, involving floorplans with predefined areas and predefined paths: For example, if the point is located inside a wall, it can be safely removed from the list.

In conclusion, this approach gives a further opportunity to improve accuracy. Presumably, the algorithm can also be extended by adding more boundary conditions learned from other construction-related DM procedures.

Chapter 8

Conclusion

8.1 Summary

In this thesis, a general overview of given Data Mining approaches of the construction industry was presented. A closer investigation was done by creating two case studies, looking at the current status of data analysis within the as-performed construction process.

The following points can now be summarised, also considering the initial questions raised under section 1.5. First, a brief insight into the most important KPI values of the construction industry was presented. Then these were inherited from the as-performed construction process and tested for their suitability. In the chapter 4, it was explained in detail what the current status of DM is in the construction industry, which methods exist and which IoT systems are suitable for construction sites. Furthermore, two case studies pointed out which potentials DM pilots to the construction industry and how an actual implementation of the DM processes can be realised. The fundamental steps were demonstrated schematically alongside monitoring a site project and the development of a positioning system. Nevertheless, it became clear that DM consists of many processes enabling the preparation of data analysis and emphasising the dependence on the proposed data system. Processes before gaining data have to be well planned and organised, systems have to be structured and arranged, and data needs to be understood and prepared. Ultimately, this thesis demonstrated how a data pipeline can work and what resources are needed for the corresponding infrastructure.

8.1.1 Data Mining steps covered within the case studies

Since this work represents DM procedures within the construction industry, it is now inviting to look at the DM steps that have been taken, followed up by actions that need to be taken in future. As defined in section 2.2.3, there are different models for the KDD approach. However, as we were working with industry-related projects, the application of the CRISP Data Mining procedure made the most sense considering construction sites, as shown in chapter 4. Eventually, the following paragraph will derive the findings to the first step of the CRISP model: Business Understanding.

Business Understanding

Business Understanding, shown in section 2.2.3, involves getting an expert in a specific field and preparing the complete DM process since the quality of the result is highly dependent on the assumptions made before.

At first, a general understanding of construction-related KPIs was given, showing a short introduction to the topic of DM in the construction industry. These indicators create the foundation of the DM project by supporting the perception of construction-related challenges and listing up improvable areas.

Various elements of the Business Understanding stage are seen, looking at the first case study, described in chapter 5. The first stage of the CRISP model involved creating a concept for placing the crane cameras and conducting laser scans and drone flights. More, for the construction site, fitting cameras and laser scanners had to be chosen, which was pointed out in chapter 5. By developing a concept, the origin of this data has become clear, which enables a conscious use of the gained data.

All of the previously mentioned systems, like the cameras, were in a ready-to-use state. No investigation had to be done on implementing additional software components. However, by creating a positioning system, things looked a bit differently. Similarly, adaptable choices had to be compared, yet, the positioning hardware was missing suitable software. Therefore, the sections section 4.4.1 and section 4.4.2 presented a variety of different monitoring technologies and determined a suitable positioning system. After analysing all available positioning systems, an exhaustive suitability test was carried out according to the site conditions. Some positioning technologies have proven to be a promising approach. Still, with the harsh site conditions, other systems are completely ineligible. A good example shows the UWB system: As explained in section 4.4.1, it is one of the most precise systems and thus used in a lot of industrial projects. However, while closer inspecting the site conditions in section 4.4.2 it quickly becomes clear that the system suffers in dynamic environments and require a high amount of sensors, making it inappropriate for construction sites. Without the Business Understanding step, the fundamentals of the data analysis approach would be, in this case, vaguely insufficient. Conclusively, neglecting the Business Understanding step can deteriorate results and thus the quality of the DM process.

Data Understanding

The data understanding step of the CRISP model further emphasises how essential interpreting the gained information is. As discussed in section 2.2.3 this phase involves, on the one hand, clarifying which data is desired and obtainable.

In the case of laser scanning, we were able to accumulate a large number of points. The number of measuring points depends on the type of scanner and defines the quality of the point cloud. However, little can be done with the individual measuring points, missing particularly semantic information, resulting in further processing.

According to the first case study with the cameras, a similar circumstance was observed. The data is based on images taken at specific times by either the fixed cameras or the drone. Analogous to the scanned measuring points, the images seem initially meaningless. Only through the correct processing and linking of essential information knowledge can be extracted, eventually leading to various benefits.

When looking at the positioning system at the construction site, the primary data to be desired are paths of workers, time-oriented positions, and locations of materials, as illustrated in chapter 4. They can, however, be summarised by general position results. On the other hand, though, it must be revealed which data is already there. As explained in section 6.2.3 at first, there is only RSSI data available. Since the usability of RSSI values regarding other DM steps is excessively low, boundary conditions and transformations are needed. Therefore, specific methods, like triangulation and fingerprinting, were shown in section 4.4.3 and section 4.4.3 to ultimately get substantial results, which summarises in a sense the Data Understanding stage. Having a predefined coordinate system in a floor plan with a list of points, for example, those can be attached to the construction plan, and desired locations can be seen.

Yet, creating location points, shown in section 6.2.4 is only possible due to an RSSI to point transformation, as illustrated above. Even though the data seems to be primarily structured for data analysis completing the Data understanding stage, one step is left to be covered.

Data Preparation

Before the obtained data from diverse sources can be utilised to create data models, the data must be prepared by conducting a further step. The Data Preparation stage was therefore embedded in all case studies.

Within the step of data preparation, the recorded points had to be transformed into point clouds. A standard methodology, the referencing, was described under section 4.3.1. The afore-mentioned procedure was done while creating a point cloud by using specific software. Nevertheless, the data mining process is iterative, which means that the individual steps are taken up several times. Thus, the under section 7.1 described problems of the point cloud have to be replaced with creative approaches and other data.

The images also have to be prepared to create point clouds, which includes, among other things, correcting the overcast elements and adjusting the angle of view. Furthermore, additional information can be extracted from the images, analysed differently for several processes.

As for now, the positioning system can deliver a collection of points stored in a database, which might, at first sight, be enough to continue with the modelling step of the DM process. The following example will explain why this is not the case.

Looking at the described RSSI filtering process in section 4.4.3, it quickly becomes clear why data preparation is of high importance. Not filtering the signal strength values would lead to substantial inaccurate positioning results since RSSI has a high fluctuation, as described in section 7.3. In fact, using those incorrect values for triangulation would lead to non-sense positioning, which cannot be interpreted by any machine or human. The lack of preparation would result in a loss of data and a waste of time since collecting data takes time. To be able to run data analysis algorithms, the data must be prepared as much as possible. Incorrect or incomplete data can tremendously change results, resulting in data that cannot be processed.

Another example is the placement of points, which are supposedly displayed being in a wall. From a human perspective, this makes no sense, which requires sorting out those points. Yet, as a machine does not have human intelligence, it can't distinguish between supposedly correct and incorrect data. Thus, the machine must be provided with boundary conditions for sorting out false values. A solution to this problem is comparing values of different positioning algorithms, which was covered in more detail in section 7.3 and is still to be extended. Further, this also explains why different positioning methods are used, as it improves accuracy and prepares data for other processes.

Inspecting the list of points generated from distinct positioning algorithms and stored in a database collection, it quickly becomes apparent that points can't be accessed without performing queries. More, to create statements or find patterns in the data, the information must be in an easy to read form. section 6.2.3 gave a brief insight into how the data gets processed from a database with gateway entries to a list of points.

Whether it is filtering RSSI values, setting boundary conditions for location points, or bringing collections of points in a useable form, data preparation appears in many processes, as shown above. Only when this step is completed correctly, additional DM processes can follow.

8.1.2 Summary

Specific steps of the DM process had to be covered, as shown above. First, at the beginning of a DM project, Business Understanding is significant since it determines the overall topic. Data Understanding and Data Preparation are then required to perform a meaningful data analysis in later stages. While some examples pointed out how essential these steps are, it is crucial to mention that the steps above cover more processes, which weren't shown in section 8.1.1, still representing a part of DM. Eventually, as described in section 2.2.3, some steps are missing from the CRISP model, which will be discussed in the next section.

8.2 Future Work and Outlook

8.2.1 Data Mining steps to be followed up

Once the data preparation is complete further steps need to be taken, as explained in section 2.2.3. Common approaches for the Modeling stage are decision trees and clusters, ultimately practising AI. However, this stage requires an actual amount of data, which by this time is in progress. Due to the limited period of this thesis, this step will follow within the scope of an ongoing research project, "Data Mining within the as-performed construction process", which will be described in more detail in the last section. The same applies to the evaluation and deployment stages. Even though the data analysis steps haven't been covered to all extend, it makes sense to look at areas and workflows that can be inspected in the future.

8.2.2 Future research fields of Data Mining methods in the construction industry

Many things provide future research potential performing DM in the construction industry. Looking back at the recent challenges of DM within the as-performed construction process, pointed out in section 4.5, it quickly becomes apparent that present studies have to be generalised as Yan, Yang, Peng, *et al.* [26] propose. Due to the high replication potential of specific processes, creating a data framework for knowledge discovery would make sense. Ultimately, this could lead to standardising processes, which can then be embedded in DM techniques described in section 2.4.

As explained in section 2.2.1, DM can conclude structured, semi-structured and unstructured data. However, most studies focus only on numerical and textual data. The scope of the analysis can be extended by using advanced techniques, like machine learning and neural networks, for extracting knowledge of non-traditional data. Yet, more research is needed on improving existing DM methods. Indeed, there is massive potential in adapting advanced DM techniques (e.g., graph mining and deep learning), Big Data processing techniques (e.g., MapReduce), Big Data storage techniques (e.g. Hadoop distributed file system and NoSQL databases), and other mainstream computing techniques (e.g., cloud computing, visual analytics). Another region for future research holds DM for sustainable construction. Energy, safety management, and green buildings are within the scope of sustainable construction and provide huge sustainable improvement potential. Examples are indoor air quality, waste management resource management and more [26].

8.2.3 Project outlook

Further steps illustrated above also need to be taken within the framework of the research project mentioned above. As already described under section 8.2.1 the collection and processing of data are followed by a far-reaching data analysis to optimise existing construction processes. The latter also requires a far-reaching, fine-granular process description of the individual construction processes with subsequent integration into BIM models. The processes and the inherent activities then form an important basis for calculating corresponding KPI values. After identifying activities and patterns in the data, it also makes sense to compare actual and target values of the construction execution process. The results obtained from this can be generalised and bundled into a data platform. Ultimately, it has become apparent that future steps are required within the DM framework of the as-performed construction process to present significant information of individual processes to the end-user in a suitable form.

Appendix A

Additional information to the Case Study of the implemented positioning system

The full source code of the BLE-Network can be found here: GIT Repository BLE-Network

Table A.1: Fingerprinting test with a location beacon

Fixpoint (ID) Target	x	y	Fixpoint (ID) Actual	x	y	Deviation	Criterion met
78	12	5	78	12	5	0,00	YES
49	8	0	56	9	1	1,41	YES
80	13	1	75	12	2	1,41	YES
61	10	0	61	10	0	0,00	YES
66	10	5	72	11	5	1,00	YES
70	11	3	71	11	4	1,00	YES
74	12	1	82	13	3	2,24	NO
49	8	0	62	10	1	2,24	NO
87	14	2	76	12	3	2,24	NO
82	13	3	71	11	4	2,24	NO
98	16	1	87	14	2	2,24	NO
51	8	2	50	8	1	1,00	YES
60	9	5	60	9	5	0,00	YES
77	12	4	65	10	4	2,00	NO
88	14	3	93	15	2	1,41	YES
76	12	3	76	12	3	0,00	YES
61	10	0	54	8	5	5,39	NO
75	12	2	76	12	3	1,00	YES
53	8	4	58	9	3	1,41	YES
100	16	3	101	16	4	1,00	YES
57	9	2	63	10	2	1,00	YES
94	15	3	70	11	3	4,00	NO
99	16	2	100	16	3	1,00	YES
50	8	1	56	9	1	1,00	YES
64	10	3	65	10	4	1,00	YES

Table A.2: Fingerprinting test with a tag beacon

Fixpoint (ID) Target	x	y	Fixpoint (ID) Actual	x	y	Deviation	Criterion met
58	9	3	59	9	4	1,00	YES
80	13	1	74	12	1	1,00	YES
78	12	5	62	10	1	4,47	NO
92	15	1	86	14	1	1,00	YES
55	9	0	58	9	3	3,00	NO
98	16	1	98	16	1	0,00	YES
100	16	3	86	14	1	2,83	NO
88	14	3	87	14	2	1,00	YES
69	11	2	70	11	3	1,00	YES
92	15	1	95	15	4	3,00	NO
102	16	5	102	16	5	0,00	YES
92	15	1	86	14	1	1,00	YES
66	10	5	65	10	4	1,00	YES
68	11	1	75	12	2	1,41	YES
63	10	2	57	9	2	1,00	YES
60	9	5	64	10	3	2,24	NO
98	16	1	98	16	1	0,00	YES
102	16	5	101	16	4	1,00	YES
65	10	4	71	11	4	1,00	YES
57	9	2	58	9	3	1,00	YES
87	14	2	64	10	3	4,12	NO
62	10	1	54	8	5	4,47	NO
75	12	2	76	12	3	1,00	YES
86	14	1	86	14	1	0,00	YES
80	13	1	86	14	1	1,00	YES

Table A.3: Triangulation test with a location beacon

Fixpoint (ID) Target	x	y	Actual x	Actual y	Deviation	Criterion met
78	12	5	11,8	3,9	1,12	YES
49	8	0	9,9	1,5	2,42	NO
80	13	1	12,7	2	1,04	YES
61	10	0	9,4	2,4	2,47	NO
66	10	5	11,4	3,8	1,84	NO
70	11	3	11,5	2,4	0,78	YES
74	12	1	11,2	1,6	1,00	YES
49	8	0	9	5,4	5,49	NO
87	14	2	13,5	2,5	0,71	YES
82	13	3	13,4	2,5	0,64	YES
98	16	1	14	2,5	2,50	NO
51	8	2	10,3	1,5	2,35	NO
60	9	5	11,8	3,3	3,28	NO
77	12	4	12,3	3,2	0,85	YES
88	14	3	12,1	1,5	2,42	NO
76	12	3	11,9	2,7	0,32	YES
61	10	0	10,4	1,5	1,55	NO
75	12	2	11,6	2,4	0,57	YES
53	8	4	11,3	2,7	3,55	NO
100	16	3	9,4	1,5	6,77	NO
57	9	2	11	1,5	2,06	NO
94	15	3	13,2	3,2	1,81	NO
99	16	2	12,1	2,8	3,98	NO
50	8	1	10	1,5	2,06	NO
64	10	3	11,5	2,3	1,66	NO

Table A.4: Triangulation test with a tag beacon

Fixpoint (ID) Target	x	y	Actual x	Actual y	Deviation	Criterion met
58	9	3	9,4	4,2	1,26	YES
80	13	1	9,4	5,3	5,61	NO
78	12	5	9,4	5,3	2,62	NO
92	15	1	13,4	1,5	1,68	NO
55	9	0	9,4	0	0,40	YES
98	16	1	13,4	1,6	2,67	NO
100	16	3	13,8	1,6	2,61	NO
88	14	3	13,4	1,6	1,52	NO
69	11	2	9,4	1,5	1,68	NO
92	15	1	13,5	1,6	1,62	NO
102	16	5	14	5,4	2,04	NO
92	15	1	13,4	5	4,31	NO
66	10	5	13,3	5,4	3,32	NO
68	11	1	9,4	1,5	1,68	NO
63	10	2	9,4	1,5	0,78	YES
60	9	5	13,2	5,4	4,22	NO
98	16	1	13,4	5	4,77	NO
102	16	5	14,4	5	1,60	NO
65	10	4	13,4	5,3	3,64	NO
57	9	2	9,4	2	0,40	YES
87	14	2	13,4	1,5	0,78	YES
62	10	1	9,5	1,5	0,71	YES
75	12	2	13,4	1,6	1,46	YES
86	14	1	13,4	1,5	0,78	YES
80	13	1	13,4	1,5	0,64	YES

References

[1] A. El Saddik, “Digital twins: The convergence of multimedia technologies”, *IEEE multimedia*, vol. 25, no. 2, pp. 87–92, 2018.

[2] K. K. Patel, S. M. Patel, *et al.*, “Internet of things-iot: Definition, characteristics, architecture, enabling technologies, application & future challenges”, *International journal of engineering science and computing*, vol. 6, no. 5, 2016.

[3] P. Tang, D. Huber, B. Akinci, R. Lipman, and A. Lytle, “Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques”, *Automation in construction*, vol. 19, no. 7, pp. 829–843, 2010.

[4] B. Dave, A. Buda, A. Nurminen, and K. Främling, “A framework for integrating bim and iot through open standards”, *Automation in Construction*, vol. 95, pp. 35–45, 2018.

[5] U. Isikdag, “Bim and iot: A synopsis from gis perspective”, *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 40, p. 33, 2015.

[6] S. Liang, C.-Y. Huang, and T. Khalafbeigi, “Ogc sensorthings api part 1: Sensing, version 1.0.”, 2016.

[7] S. Liang and T. Khalafbeigi, “Ogc sensorthings api part 2–tasking core, version 1.0.”, 2019.

[8] M. Bilal, L. O. Oyedele, J. Qadir, K. Munir, S. O. Ajayi, O. O. Akinade, H. A. Owolabi, H. A. Alaka, and M. Pasha, “Big data in the construction industry: A review of present status, opportunities, and future trends”, *Advanced engineering informatics*, vol. 30, no. 3, pp. 500–521, 2016.

[9] S. D’Onofrio and A. Meier, *Big Data Analytics - Grundlagen, Fallbeispiele und Nutzungspotenziale*. Wiesbaden: Springer Fachmedien Wiesbaden, 2021, ISBN: 978-3-658-32235-9.

[10] P. Ghavami, *Big Data Management: Data Governance Principles for Big Data Analytics*, ser. Business & economics. Walter de Gruyter GmbH, 2020, ISBN: 9783110662917. [Online]. Available: https://books.google.de/books?id=vNTAzAEACAAJ.

[11] J. Cleve and U. Lämmel, *Data Mining*, ser. De Gruyter Studium Series. Oldenbourg Wissenschaftsverl., 2014, ISBN: 9783486713916. [Online]. Available: https://books.google.de/books?id=4i2nngEACAAJ.

[12] N. Bissantz and J. Hagedorn, “Data mining (datenmustererkennung)”, *Wirtschaftsinformatik*, vol. 51, no. 1, pp. 139–144, 2009.

[13] N. M. Adams, G. Blunt, D. J. Hand, and M. G. Kelly, “Data mining for fun and profit”, *Statistical Science*, vol. 15, no. 2, pp. 111–131, 2000.

[14] D. J. Hand, “Principles of data mining”, *Drug safety*, vol. 30, no. 7, pp. 621–622, 2007.

[15] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, *et al.*, “Knowledge discovery and data mining: Towards a unifying framework.”, in *KDD*, vol. 96, 1996, pp. 82–88.

[16] L. A. Kurgan and P. Musilek, “A survey of knowledge discovery and data mining process models”, *The Knowledge Engineering Review*, vol. 21, no. 1, pp. 1–24, 2006.

[17] H. M. Chung and P. Gray, “Data mining”, *Journal of management information systems*, vol. 16, no. 1, pp. 11–16, 1999.

[18] A. P. Chan and A. P. Chan, “Key performance indicators for measuring construction success”, *Benchmarking: an international journal*, 2004.

[19] S. Zhang, J. Teizer, J.-K. Lee, C. M. Eastman, and M. Venugopal, “Building information modeling (bim) and safety: Automatic safety checking of construction models and schedules”, *Automation in construction*, vol. 29, pp. 183–195, 2013.

[20] M. M. Khalfan, “Managing sustainability within construction projects”, *Journal of Environmental Assessment Policy and Management*, vol. 8, no. 01, pp. 41–60, 2006.

[21] D. Arditi and H. M. Gunaydin, “Factors that affect process quality in the life cycle of building projects”, *Journal of construction engineering and management*, vol. 124, no. 3, pp. 194–203, 1998.

[22] F. H. Abanda, B. Kamsu-Foguem, and J. Tah, “Bim–new rules of measurement ontology for construction cost estimation”, *Engineering Science and Technology, an International Journal*, vol. 20, no. 2, pp. 443–459, 2017.

[23] Z. Shen and R. R. Issa, “Quantitative evaluation of the bim-assisted construction detailed cost estimates”, 2010.

[24] E. J. Jaselskis and T. El-Misalami, “Implementing radio frequency identification in the construction process”, *Journal of construction engineering and management*, vol. 129, no. 6, pp. 680–688, 2003.

[25] D. H. Walker, “An investigation into construction time performance”, *Construction Management and Economics*, vol. 13, no. 3, pp. 263–274, 1995.

[26] H. Yan, N. Yang, Y. Peng, and Y. Ren, “Data mining in the construction industry: Present status, opportunities, and future trends”, *Automation in Construction*, vol. 119, p. 103 331, 2020.

[27] F. Bosché, “Automated recognition of 3d cad model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction”, *Advanced engineering informatics*, vol. 24, no. 1, pp. 107–118, 2010.

[28] F. Bosché, M. Ahmed, Y. Turkan, C. T. Haas, and R. Haas, “The value of integrating scan-to-bim and scan-vs-bim techniques for construction monitoring using laser scanning and bim: The case of cylindrical mep components”, *Automation in Construction*, vol. 49, pp. 201–213, 2015.

[29] H. Son, F. Bosché, and C. Kim, “As-built data acquisition and its use in production monitoring and automated layout of civil infrastructure: A survey”, *Advanced Engineering Informatics*, vol. 29, no. 2, pp. 172–183, 2015.

[30] A. Braun, S. Tuttas, A. Borrmann, and U. Stilla, “A concept for automated construction progress monitoring using bim-based geometric constraints and photogrammetric point clouds.”, *J. Inf. Technol. Constr.*, vol. 20, no. 5, pp. 68–79, 2015.

[31] M.-W. Park and I. Brilakis, “Construction worker detection in video frames for initializing vision trackers”, *Automation in Construction*, vol. 28, pp. 15–25, 2012.

[32] Y. Turkan, F. Bosche, C. T. Haas, and R. Haas, “Automated progress tracking using 4d schedule and 3d sensing technologies”, *Automation in construction*, vol. 22, pp. 414–421, 2012.

[33] A. Braun and A. Borrmann, “Combining inverse photogrammetry and bim for automated labeling of construction site images for machine learning”, *Automation in Construction*, vol. 106, p. 102 879, 2019.

[34] J. Song, C. T. Haas, and C. H. Caldas, “Tracking the location of materials on construction job sites”, *Journal of construction engineering and management*, vol. 132, no. 9, pp. 911–918, 2006.

[35] H. Li, G. Chan, J. K. W. Wong, and M. Skitmore, “Real-time locating systems applications in construction”, *Automation in Construction*, vol. 63, pp. 37–47, 2016.

[36] P. Vähä, T. Heikkilä, P. Kilpeläinen, M. Järviluoma, and E. Gambao, “Extending automation of building construction—survey on potential sensor technologies and robotic applications”, *Automation in construction*, vol. 36, pp. 168–178, 2013.

[37] D. G. Torrent and C. H. Caldas, “Methodology for automating the identification and localization of construction components on industrial projects”, *Journal of computing in civil engineering*, vol. 23, no. 1, pp. 3–13, 2009.

[38] Z. Ma, S. Cai, N. Mao, Q. Yang, J. Feng, and P. Wang, “Construction quality management based on a collaborative system using bim and indoor positioning”, *Automation in Construction*, vol. 92, pp. 35–45, 2018.

[39] Y. Gu, A. Lo, and I. Niemegeers, “A survey of indoor positioning systems for wireless personal networks”, *IEEE Communications surveys & tutorials*, vol. 11, no. 1, pp. 13–32, 2009.

[40] J. Chai, C. Wu, C. Zhao, H.-L. Chi, X. Wang, B. W.-K. Ling, and K. L. Teo, “Reference tag supported rfid tracking using robust support vector regression and kalman filter”, *Advanced Engineering Informatics*, vol. 32, pp. 1–10, 2017.

[41] Y.-W. Lin and C.-Y. Lin, “An interactive real-time locating system based on bluetooth low-energy beacon network”, *Sensors*, vol. 18, no. 5, p. 1637, 2018.

[42] J. Park, Y. K. Cho, and C. R. Ahn, “A wireless tracking system integrated with bim for indoor construction applications”, in *Construction Research Congress 2016*, 2016, pp. 2660–2668.

[43] W. Chantaweesomboon, C. Suwatthikul, S. Manatrinon, K. Athikulwongse, K. Kaemarungsi, R. Ranron, and P. Suksompong, “On performance study of uwb real time locating system”, in *2016 7th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES)*, IEEE, 2016, pp. 19–24.

[44] K. M. Rashid, J. Louis, and K. K. Fiawoyife, “Wireless electric appliance control for smart buildings using indoor location tracking and bim-based virtual environments”, *Automation in Construction*, vol. 101, pp. 48–58, 2019.

[45] J. Yang, O. Arif, P. A. Vela, J. Teizer, and Z. Shi, “Tracking multiple workers on construction sites using video cameras”, *Advanced Engineering Informatics*, vol. 24, no. 4, pp. 428–434, 2010.

[46] J. Gong and C. H. Caldas, “An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations”, *Automation in Construction*, vol. 20, no. 8, pp. 1211–1226, 2011.

[47] S. Woo, S. Jeong, E. Mok, L. Xia, C. Choi, M. Pyeon, and J. Heo, “Application of wifi-based indoor positioning system for labor tracking at construction sites: A case study in guangzhou mtr”, *Automation in Construction*, vol. 20, no. 1, pp. 3–13, 2011.

[48] W.-S. Jang and M. J. Skibniewski, “Embedded system for construction asset tracking combining radio and ultrasound signals”, *Journal of Computing in Civil Engineering*, vol. 23, no. 4, pp. 221–229, 2009.

[49] S. Taneja, A. Akcamete, B. Akinci, J. H. Garrett Jr, L. Soibelman, and E. W. East, “Analysis of three indoor localization technologies for supporting operations and maintenance field tasks”, *Journal of Computing in Civil Engineering*, vol. 26, no. 6, pp. 708–719, 2012.

[50] J. M. Gómez-de-Gabriel, J. A. Fernández-Madrigal, A. López-Arquillos, and J. C. Rubio-Romero, “Monitoring harness use in construction with ble beacons”, *Measurement*, vol. 131, pp. 329–340, 2019.

[51] M. Heydariaan, H. Mohammadmoradi, and O. Gnawali, “Toward standard non-line-of-sight benchmarking of ultra-wideband radio-based localization”, in *2018 IEEE Workshop on Benchmarking Cyber-Physical Networks and Systems (CPSBench)*, IEEE, 2018, pp. 19–24.

[52] Q. Dong and W. Dargie, “Evaluation of the reliability of rssi for indoor localization”, in *2012 International Conference on Wireless Communications in Underground and Confined Areas*, IEEE, 2012, pp. 1–6.

[53] C. K. Chui, G. Chen, *et al.*, *Kalman filtering*. Springer, 2017.

[54] Y. Wang, X. Yang, Y. Zhao, Y. Liu, and L. Cuthbert, “Bluetooth positioning using rssi and triangulation methods”, in *2013 IEEE 10th Consumer Communications and Networking Conference (CCNC)*, IEEE, 2013, pp. 837–842.

[55] F. Shang, W. Su, Q. Wang, H. Gao, and Q. Fu, “A location estimation algorithm based on rssi vector similarity degree”, *International Journal of Distributed Sensor Networks*, vol. 10, no. 8, p. 371 350, 2014.

[56] K. Kaemarungsi and P. Krishnamurthy, “Modeling of indoor positioning systems based on location fingerprinting”, in *Ieee Infocom 2004*, IEEE, vol. 2, 2004, pp. 1012–1022.

[57] Bayerischer Bauindustrieverband e.V., *Building lab - bauen neu gedacht*, https://www.bauindustrie-bayern.de/building-lab, Accessed: 2021-08-18.

[58] Amazon, *Informationen zu aws*, https://aws.amazon.com/de/about-aws/, Accessed: 2021-08-26.

[59] TUM Chair of Building Technology and Climate Responsive Design, *Plan of climalab*, https://www.ar.tum.de/fileadmin/w00bfl/www/03_Fakultaet/Orientierung/Plaene_2018/Bestelmeyer/Raum_0250___ARI_Clima_Lab.pdf, Accessed: 2021-09-13.

Appendix B

Declaration

I hereby affirm that I have independently written the thesis submitted by me and have not used any sources or aids other than those indicated.

Location, Date, Signature