Machine learning in the design phase of construction projects

Lunds Tekniska Högskola

Erik Hiekkala

June 25, 2020

LUND
UNIVERSITY

Lunds Universitet, Lunds Tekniska Högskola
Institutionen för bygg- och miljöteknologi, Byggproduktion

Telephone: +46 46 222 74 21
Webbsite: www.bekon.lth.se

ISRN LUTVDG/TVBP-20/????-EN

Abstract

Machine learning has the potential to improve processes and revolutionize economies on a global level and that is also true for the construction industry. Machine learning has started to get attention in recent years as computer power have increased and the cost has reduced. The purpose with machine learning is train an algorithm to process large quantities of data for essential information. This can be applied to varies field and industries and as the societies are gathering data and are being digitized, the applications for machine learning will only grow. Machine learning has yet be applied on a large scale in the construction sector but that will inevitable change as the benefits becomes to great to ignore. Many of the large Swedish companies have started to look at digitizing their activities, which would be an important milestone for machine learning. However many are also unaware of machine learning and how to incorporate it into the business, hampering its progress in the construction sector. The result conformed the the idea about all parties are recognise the promise with digitizing but currently on a low level, the amount of data is not sufficient to train algorithms on satisfying level.

Sammanfattning

I takt med att datorer har blivit kraftfullare och effektivare har möjligheterna med att tillämpa datorverktyg i ekonomin växt. Mer och mer kan man börja läsa om maskininlärning och om hur den kan användas för att förbättra processer och leda till högre tillväxt. Byggsektorn har lidit av låg tillväxt under en längre tid och maskininlärning kan vara en del av lösning till ökad produktivitet. Maskininlärning innebär att man låta algoritmer behandla data för att lära sig och se mönster, koppling och att kunna göra prognoser inom ett område. Fördelarna här är att algoritmer kan behandla stora mängder data, något som hade varit svårt för människor att göra. Att gå igenom datan innebär att kunskap kan utvinnas, vilket kan gör ett företag/organisation konkurrenskraftig. Hur maskininlärning kan tillämpas för att göra projekteringsfasen mer effektiv i termer om tid, finanser och projektera för öka säkerhet i produktionen har det undersöks lite om. Resultatet visade att om med osäkerheterna i svaren att byggföretagen generellt inte är lika digitaliserade som resten av samhället och att tillämpningen av maskininlärningen kan bli svår innan tillräckliga mängder data har samlats, men att digitaliseringen kommer att ske då alla anser att fördelarna är för stora för att kunna ignorera.

Förord

Examensarbetet är det sista moment som studenter vid Lunds Tekniska Högskola får göra innan man examineras. Detta avslutar en fem års lång ingenjörsutbildning på universitet. Jag skulle vilja börja med att tacka min handledare, Radhlinah Aulin för all hjälp under arbetets gång med att diskutera och lägga upp planer för arbetet, det har varit till mycket hjälp för mig.

Jag skulle vilja fortsätta med att tacka allt folk som har ställt upp på undersökning. Tack för att ni tog er tiden till att skriva, ringa, skriva mail och dela med er av informationen som mitt examensarbete behövde. Sist skulle jag vilja tacka alla vänner som har stöttat och varit med och bollat idéer. Det stödet har varit värdefullt under dessa 20 veckor som examensarbetet har pågått i.

Erik Hiekkala

Contents

1 Introduction

1.1 Background

The construction industry in Sweden makes up a large part of the GDP as it is an important part of the economy but it has not been as innovative and has a lower productivity development compared to the rest of the economy [1][2]. The industry has simply been slow to incorporate new technique and as a result lags behind in the development [1]. This is something which require a change.

With the progress made in transistor in integrated circuit in the last decades and with the introduction of more powerful computers, an emerging field known as machine learning has started to make its entrance in the society [3][4]. The promises with machine learning are many, aided with the rise of data thanks to digitization, increased access to big data, increasingly powerful computers making it seems inevitable that machine learning would not make it into the construction industry [5][4][6]. The question should rather be when then if [5]. Integrating machine learning into the business strategy will help to streamline companies and the organisations overall performance. Improved performance are likely to make companies more competitive on the market [7].

Machine learning can be applied in the industrial production to make the production lines more efficient, reducing the operating costs etc [7]. With machine learning, it would be possible to use an algorithm to execute monotonous task which can be less efficient if performed by humans. There are many fields where machine learning can be applied to help improve the economy [8]. Examples of application are agriculture, brain machine interfaces, computer vision, machine translation, medical diagnosis, optimization, search engines etc [9][10][11]. Because an algorithm doesn't need rest, it can work with the same efficiency around the clock and with a low operation cost. As the world is starting to become more digitized it appears inevitable that data and algorithms will play a bigger roll in people's everyday life in the future. Given the numerous applications for machine learning in the economy as a whole, it will be integrated into every sector in the economy sooner or later when The Internet of Things starts to emerges [12]. The question is if the construction will adapt the technology together with the other sectors in the

economy or fall behind [13].

1.2 Problem

As societies develops and new technology emerges, new challenges as well as opportunities arise for the construction industry. Technology like machine learning have be applied in several industries in recent years as the technology have matured and become available on the market [14][15]. Utilizing machine learning is expected to have a profound impact on the way data is being processed [16]. With trained algorithms it is now possible to handle larger quantities of data than before, and to take decision based on the analysis made by the algorithm in a much shorter time span than previously [16].

There are several problems ahead before machine learning can be utilized in the construction industry. One big obstacle to overcome is the construction sectors pace of including new technology [13]. Construction companies are in general sluggish to use new technology, which hampers the economical progress [13]. How raw data shall be used in machine learning to train an algorithm is a another problem. To acquire this technology and be able to use it requires knowledge in the field of machine learning and coding, this competence is something most construction companies do not have and that gives the construction companies two choices, to acquire the competence through an initiative to build up an internal department for it or to hire it from an external consult [1]. The cost for acquiring is also something to be considered. Even though machine learning can help to make the design phase more productive, it can be costly in the initial stage with recruiting computer experts and investing in a project with the objective to create an algorithm for the design phase. This also takes time which is a factor to consider.

1.3 Aim

The aim with this master thesis is present a theoretical framework for how machine learning can be utilized in the design phase in a construction project make it more efficient in terms of capital, safety and time. The paper will beyond that evaluate if constructions companies are likely to utilize this technology in the future. Hiring architects to design, engineers to calculate and economists to estimate profitability costs both time and capital. Machine

learning can improve a numerous things of this processes and will only be better with time as more and more data is collected. The ambition is to have an understanding on machine learning and its role (or lack of role) in the design phase in the construction industry for the future. To achieve the aim of the study the following research questions will be addressed.

1. What is machine learning in construction design?
2. Can machine learning result in i) reducing costs, ii) faster working process and iii) help to design and create safer work places during the design phase?
3. Has the technique matured enough for machine learning to be utilized in a meaningful way and is it difficult to acquire the technology for machine learning in construction design?

1.4 Delimitation

Because of the vastness of both the construction industry and machine learning, some limitations have been made in order for this master thesis to be possible within the time frame of 20 weeks. The thesis will not evaluate any long term consequences of machine learning but only to focus on how machine learning can can help the construction companies to streamline the design phase in the near future in order to increase margin, safety and decrease time needed for a project.

The master thesis will only evaluate the larger parties ability to acquire the technology as including small companies will be too time consuming since they are less likely to have the time, capability and money to spend on machine learning.

1.5 Target group

Since the master thesis will focus on machine learning in the construction industry, the primary target group will be parties in the construction industry. This includes both private and public companies as well as computer engineers that are interested in exploiting an undeveloped digital market [13].

1.6 Disposition

Chapter	Description
1. Introduction	The introduction is to introduce machine learning and the purpose with it. Here will the background and problems be highlighted as well the aim of this paper.
2. Theory	The chapter presents the theory behind machine learning, what it is, different forms of training and how it can be utilized by different organizations to create more value and improve efficiency within an organization. The chapter will further present how machine learning can be used in the construction sector.
3. Method	Third chapter presents a framework for how the method will be presented and executed in light of the studies objectives.
4. Result	The results will present the answers from the survey and will be anonymously presented.
5. Discussion	The chapter will discuss on several levels the theory of machine learning and the results from the survey. How machine learning can change a business is important to understand for the aims of the master thesis as well as understanding how the future for machine learning will play out.
6. Conclusion	The final chapter will present conclusions drawn from the process of collecting data. Given the magnitude of information, there will be suggestions of continued studies which the master thesis was unable to process.

2 Theory

2.1 What is machine learning?

Machine learning was first coined as an expression by a American professor, named Arthur Lee Samuel in 1959 and is the scientific study of algorithms and statistical models [17]. The idea with machine learning is that giving an algorithm data and training it will let the algorithm find and understands pattern [18]. The algorithm creates a model based on the supplied data and through iterative testing, the model may perfect itself, i.e. as the algorithm learns it becomes more precise in it's predictions [18]. The data can have one or more variables that the algorithm will train to understand, and the variables are usually called features [19]. The algorithm can through learning and processing of data solve problems that are too difficult for conventional program to handle. Improving its performance over time as the algorithm is processing new data [20]. This can be visualised in figure 1.

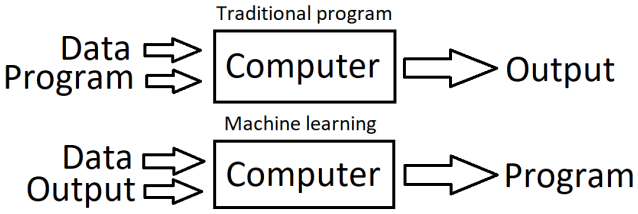

Figure 1: Showing the difference between traditional programming and machine learning. Source: towardsdatascience, 2019

How the process of training an algorithm can be seen in three levels, representation, evaluation and optimization [21]

- Representation: The algorithm creates a model which learns the connection between the input data and the labeled output after processing the data.

- Evaluation: After creating a number of models, the best model (with the most accurate predictions) are chosen and tested with new data to see if the model(s) work with unknown data.
- Optimization: Here the model(s)with the best performance is chosen and are trained with more diversified data to be optimized.

Besides the process of training an algorithm, preparing the data is a process in it self [21]. The data could be split up into two groups, the training data and the evaluation data [19]. The purpose with this is that after the training is complete, the model should be tested to see if it is accurate in its predictions, based on data that it has never seen before [19]. The ratio between these groups should be around 80-20% in the favour of the training data, though this is just a rough estimate [19]. Before the data can be used for training, it needs to be prepared [20]. Preparing data usually follows these steps [22]

- Selecting a subset of data: Select data which helps the algorithm to train and understand a problem
- Incorporate the data into groups: The amount of data can be easier to handle when sorted into groups, data should therefore be sorted after similarities
- Choose the format and modeling: Choosing the format for the data, some algorithm needs the data to be organized in special order.
- Remove and/or replace data that are missing or blank: Remove any unnecessary data and control that the data is complete.
- Normalize the data: Transforming all data to a shared scale so that the measurement is the same for all the data sets

The data is crucial for training the algorithm and how well the algorithm performance depends strongly on the data [23]. Given the importance data now has for companies and industries, data will play a bigger roll in the future [24].

The algorithm is trained using a learning model [20]. The model used for the training should be well suited and adapted to the task for being successfully, as different models are training the algorithm for very different objective [20]. More about learning models in chapter 2.6. There are many types of algorithms and some of the more common are in the list below [25]

- Regression algorithms: The algorithm uses linear and logistic regression to predict an output. All regression algorithm models follows the same general pattern.
- Decision trees: The algorithm has a set of decision rules and makes recommendations using classification.
- Instance-based algorithms: Classification is used to estimate how probable it is for a data point to belong to a group or not, based on its proximity to the other data points.
- Clustering algorithms: The algorithm works by putting all data into groups based on similarities.
- Artificial neural network: The algorithm uses artificial neurons that are connected with each other to mimic the human brain and is effective for image recognition and computer vision [21].

These algorithms suits different needs and therefore it is important to choose a right algorithm for the problem that shall be solved [26]. Overall, the process for training an algorithm, using machine learning, can be more or less identified into eight steps [21]

- Identify and bring forth the necessary data: Understand what kind of data is relevant for the process
- Prepare data: The data needs to be clean, secured and governed for usage
- Select the algorithm: Decide what type of model is best suited for the purpose of machine learning.
- Train the algorithm: The algorithm is trained to create models of that chosen type in the previously step. The models process through the data sets and are slowly increasing the the accuracy of the predictions.
- Evaluate performance: Evaluate the models and choose the model with the best performance for deployment
- Deploy models: The model is deployed.
- Prediction: The model starts to make predictions on new data.

- Evaluate the predictions: The model is evaluated and the data that the model has used is feed back for improvement. With a continuous flow of data, the model can continue to increase its accuracy.

This process can be seen more as a cycle rather than as a line [21]. Deploying the model and letting it collecting more data, the model can improve the accuracy of its predictions and learn without human assistance [21].

2.2 Why now?

The recent years have seen lots of progress in machine learning and AI, resulting in a large quantity of capital being invested into tech companies [21]. There are several things that have enable the progress of machine learning [27]. More powerful processors, declining cost of storage and management of data, clustering of computer, more commercial data sets and availability of algorithms on commercial platforms [21]. As can be seen i figure 2, the process power in computers have increase enormous in the last decades, though the pace of increasing the numbers of transistors into a chip have been slowing down compared to Moore's law [28][29]. However, this trend of

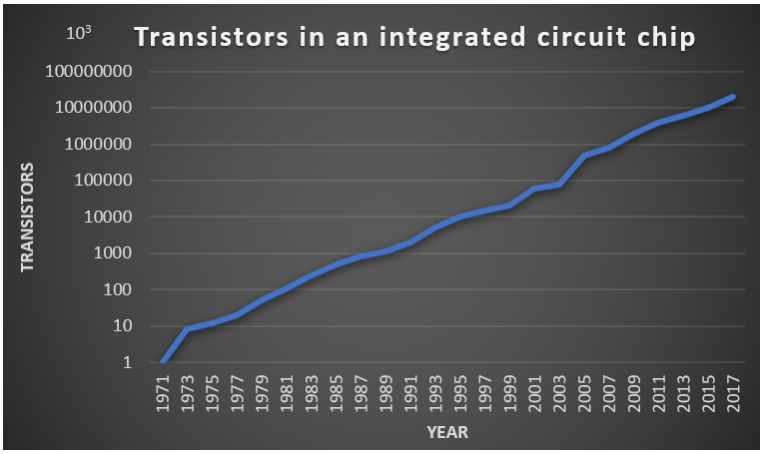

Figure 2: Displays the numbers of transistors in integrated chips. Source: Our world in data, 2018

using machine learning is likely to continue and will result in transformation for industries and businesses with the rise of more powerful computers [4].

The access to data is an equally important reason for the emergence of machine learning [21], and the quantities of data have exploded in the last years [30]. This trend of massive data creation seems to continue as showed in figure 3 [30]. This is a goldmine for corporations as the data will help them to improve their businesses [16]. Today, there are numerous application on the market for machine learning such as, chatbots, process optimization etc [31]. The industry for machine learning is expected to grown at a very high rate over the next few years [31].

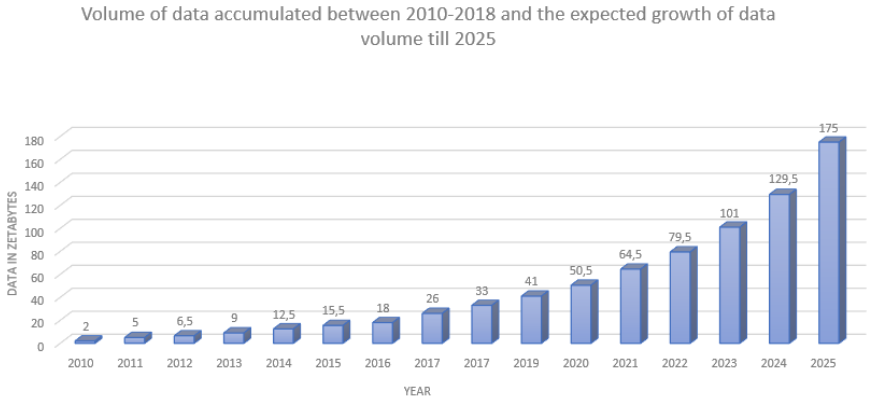

Figure 3: Displays how the amount of data that have been created the last years and how it is expected to grow till 2025. Source: Statista, 2018

2.3 Algorithms

Depending on the purpose with machine learning, different methods can be applied to create an algorithm, and different way can be used to train it [26]. The algorithm itself is nothing more that instructions for a computer on how to handle data [21]. Compared to other algorithms, machine learning algorithms are different in the sense that they have flipped the process, it is the data (output) that creates the model [21]. The more data, the better the algorithm becomes [21].Below are some of the most commonly know algorithms (some of which have already been discussed) [21][32] [33]

- Linear regression: The algorithm models the relationship between input and output using a linear function, Y=f(X)
- Logistic regression: This algorithm is used when the variable is binary, meaning the value are discrete
- Classification: A classification algorithm is a prediction in what class/group a data points belong to [34]
- Dimensionality reduction: The algorithm reduces the numbers of features in the data sets [35]
- Support vector machines: Support vector machines or SVM tries to separate the data points using a line or a hyperplane so that it will be easier to classified the data
- Decision trees: The algorithm uses a tree-like model for the decision making and the consequences of the options
- Naive Bayes: The algorithm calculate the probability that an event will happen if another action has happened, using Bayes's theorem
- k-means: The algorithm will group the data into groups with similarities

One type of algorithm that is relative new in practise, but seems promising, is artificial neural network [36]. A neural network works by processing an input vector through a field of artificial neurons [26]. These neurons are connected and are thought to mimic the human brain [26]. It is however despite the similarities a much less complex version compared to the human brain [26]. The input vector(s) are processed by the artificial neurons which are divided into "hidden layer" [26]. How many numbers of hidden layers depends on the complexity of the algorithm and every hidden layer have a finite number of parameters [36]. After being processed, the information is sent to the output layer which gather all the data from the different neurons and sends out an output as figure 4 shows [36].

Each neuron has a weight that is adjusted as the model is changing, and all outputs are change with this weight [36]. A layer of only one neuron is called a Perceptron [37].

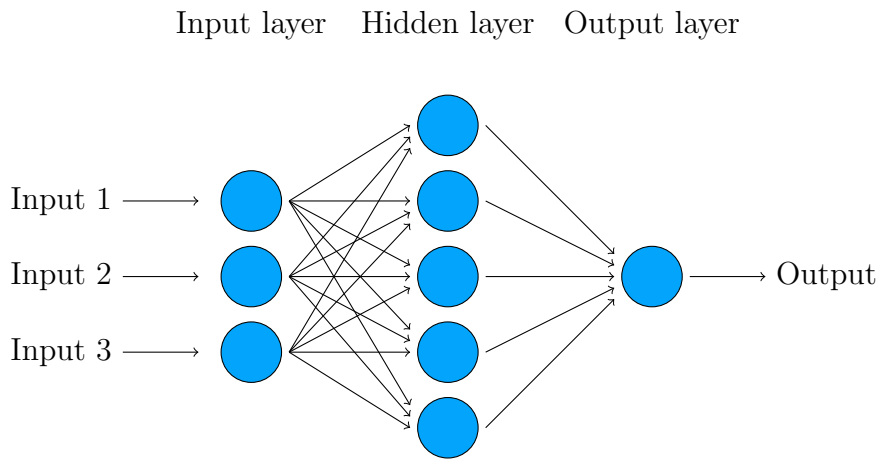

Figure 4: How information is processed through a neural network. Source: IBM, 2017

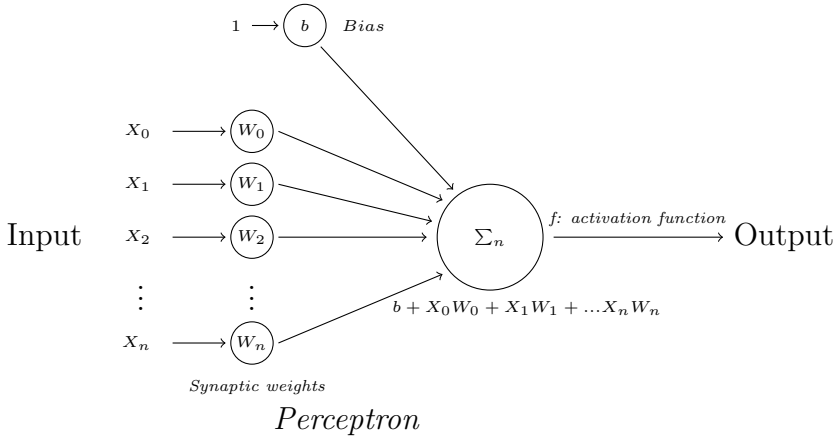

Figure 5: How a single neuron works. Source: towarddatascience, 2019

Figure 5 describes how a single neuron works, the x:s are input and multiplied with the weight which is applied to each artificial synapse and the weight can be seen as the strength for a node [36]. The value b is a bias value, giving the neuron the ability to shift the activation function up and

down [36]. The activation function decides if a neuron shall be activated or not by adding the sum of all weighted inputs and bias [36].

Doing so allows the output to be non linear, which is the point with artificial neural networks [36]. When the output has passed through (called a forward pass) and is compared to the actual value, an error value or a cost function is sent back (called backward pass) through the neurons with a value of one half of the squared difference between the output and the actual value [36]. Repeating this pattern reduces the error and the algorithm is trained to perform better and give answers with a higher precision [36]. The process of converging the parameters in the network to minimize error in the system is called gradient descent [27]. The purpose with ANN is to perform deeper analysis than just organizing data, this includes computer vision, speech recognition, self driving cars, medical diagnosis etc [36]. Deep learning which is often call a sub-discipline is a machine learning methods that uses an artificial neural network [21].

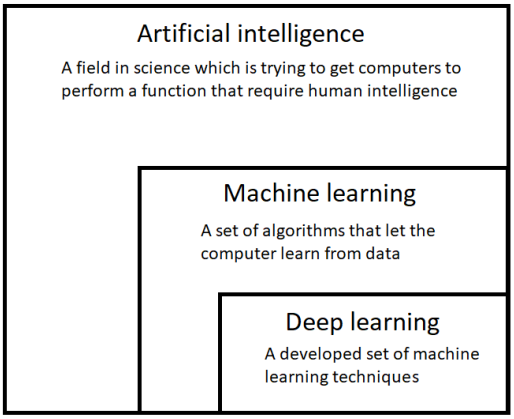

Figure 6: How the different areas relate to each other. Source: Autodesk BIM 360, 2017

It is a technique that uses both supervised and unsupervised machine learning with a large quantities of data and have many more hidden layers compared to a ordinary neural network [21]. Unlike other machine learning algorithms,

it learns the same way as humans do, by examples and is a sub category to AI [36]. How the different areas are connected can be seen in figure 6. The application for deep learning are many and especially in areas where other forms of machine learning algorithms performs poorly, such as speech recognition, prediction of machine malfunction, self-driving cars, language translations, healthcare, fraud detection, virtual assistants etc [38][21][36]. There are however challenges with deep learning [27]. For one, training with ANN requires a lot of data, which may not always be available in sufficient amount [27]. Secondly, using reinforcement learning with a reward-based purpose can lead to the agent acting unethically in its pursuit for rewards, which in the real world would have a negative impact if deployed [27].

2.4 Features

All models depends on a or more features, which can be seen as a variable or a common factor in the data [18]. These are necessary because they can be seen as a measurement that helps to form the model [18]. Examples of features could be age in a population, colour on objects, alcohol content in drinks etc [18][19]. It is important to choose as few features as possible but enough to cover all variables to explain the situation which the algorithm faces [39]. Unnecessary features results in more work for the algorithm without the corresponding output [39]. Irrelevant features may even affect the model negatively if chosen poorly [40]. Features is not necessary for every form of learning models for machine learning, which will be discuss in chapter 2.6.

2.5 Building Information Model

Building Information Modeling, or BIM for short, is a process for incorporating all information about a project in all phases [41]. BIM can be seen as a digital platform to create and handle all information regarding the planning, design, construction and collaboration between parties for a project [42]. The tools in BIM can help people to see different views of the projects geometry and attributes which helps the personal to visualise the the structure [43]. This makes it possible for anyone involved in the project to get access to real time information through the cyber space, known as a common data environment [42]. For successfully use of BIM, 5 things are necessary for it to work properly and effective [44]

- Information: There are two types of data, models and documents. Models are the digital data that describe an object (facilities, roads, cottage etc). Documents are images, films, drawing etc.
- People: Using BIM often includes different kinds of roles for the personal responsible for its creation. There is often a BIM director, BIM manager, BIM consultant and BIM technologist
- Policies: This is the frame work for decision making to let the work proceed smoothly and without friction
- Processes: This is the order of which activities will happen in
- Technologies: This is the soft- and hardware required to manage the processes in BIM

These things allows the work process to be effective and gives the team a better overlook of the visual, planning, cost, coordination, operational data etc [44]. The result of using BIM can be reduced rework, reduced project duration and reduced errors throughout the project [44].

2.5.1 Why is BIM important for machine learning?

Given that BIM is collecting enormous amount of data through the digitization of the documents, this would be a perfect place to utilize machine learning [45]. Machine learning would be able to classify every element and process all information about a project, giving it the ability to make improvements on the model building process [46]. Accordingly to a survey from Kreo, a majority of construction companies takes more than four weeks to prepare a bid [46]. This can be reduce when using machine learning to generate different designs alternatives which reduces the human errors [46]. The result will be more cost-efficient, reduce time at the design phase, more realistic time planes and improved margins [47]. This trend of gathering data is likely to continue as the technology improves and they can handle more data [47]. There are companies today that are working with introducing machine learning to BIM, Autodesk is such a company [48]. Autodesk wants to reduce accidents by training algorithms for risk management using data from Classic BIM 360 field amongst other [49].

BAM is an another construction company with a presence in over 30 countries who wants to use machine learning to improve its workflow [50]. As BAM has a large quantity of data to use, they are hoping to use machine learning to identify risks and safety issues in projects [50]. By knowing the risk, they can be avoided and reduced in probability and severity, leading to less costs [50].

2.6 Different forms of learning models for algorithms

Training an algorithm can be done with multiple models, which can be categorized into groups, the most common are supervised, unsupervised, semi-supervised (which is a combination of supervised and unsupervised machine learning) and reinforcement machine learning [51]. This is shown by figure 12 together with some common algorithms [52].

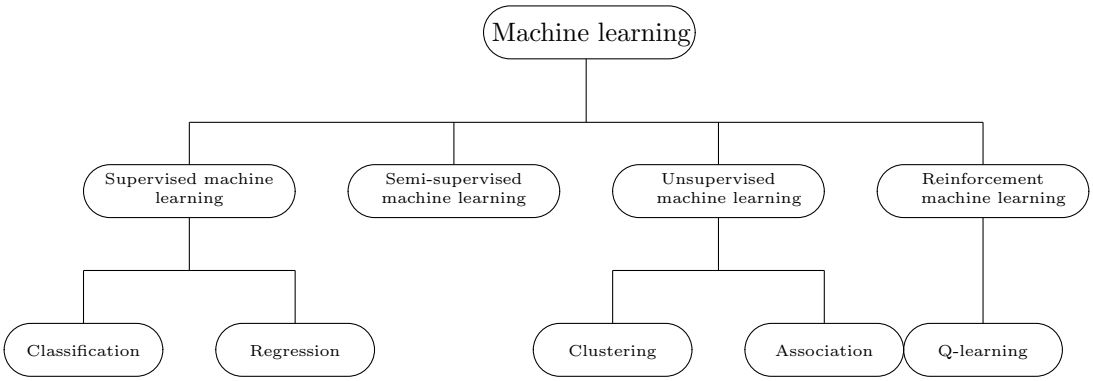

Figure 7: Different types of machine learning models with some common algorithms. Source: towardsdatascience, 2019

2.6.1 Supervised machine learning

Supervised machine learning algorithms use classified data (organized, not random data) to analyze and create a mathematical model [53]. This model can be of different types, such as decision trees, support vector machines or regression analysis to mention a few [54]. For all the different types of algorithms trained in supervised machine learning form, data are inserted and the algorithm predict output values which the algorithm can compare

with the correct answer and modify its model [54]. It can be seen as a function of Y=(X) where X is the input data and Y is the answer [55]. This is shown in figure 16. The desired output depends on the feature chosen for the

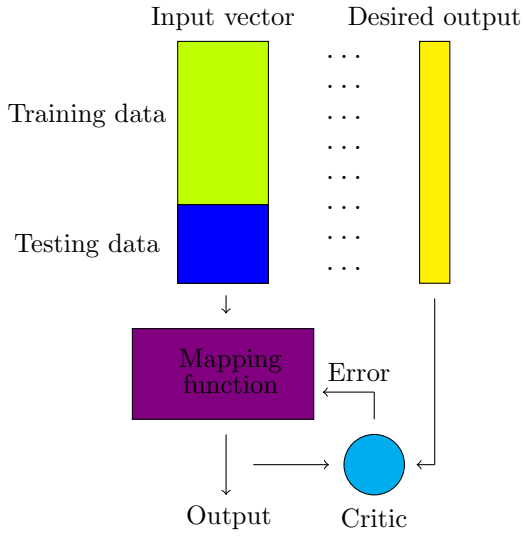

Figure 8: How supervised machine learning works. Source: IBM, 2017

algorithm and it will train until an acceptable level of accurate prediction has been achieved [54]. For supervised machine learning, one problem with using this learning method is overfitting, meaning that the algorithm can not generalize well from new data [56]. The algorithm may perform excellent on the data that it has been given but is unable to predict effectively on new data, because the model do not include variance to an adequate degree [57]. However having too much variance results in a low bias for the model [56]. Bias is when a algorithm has limited flexibility to understand important patterns from the data, the algorithm is consistent but inaccurate [56]. Variance on the other hand is an algorithms ability to be sensitive to specific sets in the data and is on average accurate but inconsistent, the opposite compared with bias [56]. It can be seen as a trade-off between bias and variance, where both should be as high as possible [57].

The reason for the trade-off in the first place is due to the fact that low variance algorithms are less complex and have a few variables to depend on [56]. A complex algorithm tends to be too flexibility, making it inconsistent,

this is shown in figure 9 [56]. That is why the data is divided into two sets, the training data and the testing data [54]. The testing data will be used for see how well the model responds to new data [54].

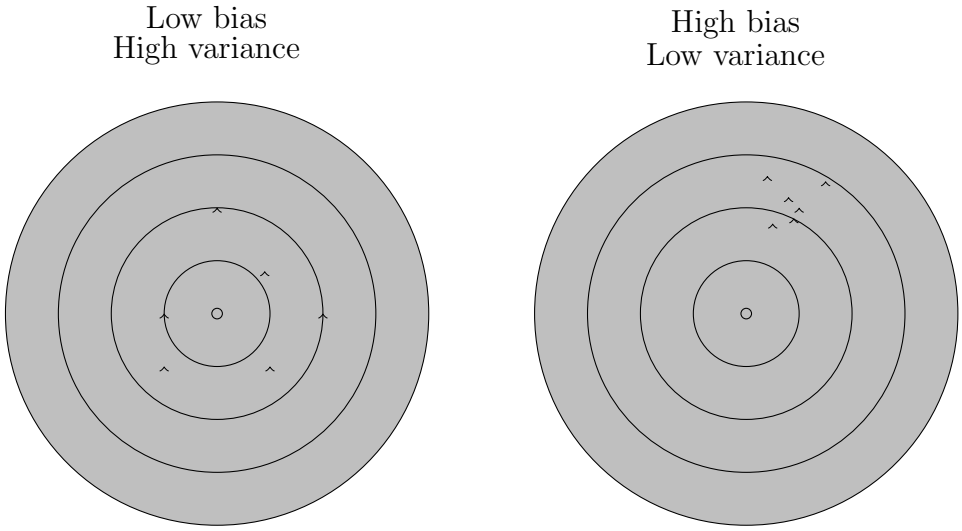

Figure 9: The trade-off between bias and variance. Source: Elitedatascience, 2016

For supervised machine learning, the are many types of different algorithms used for supervised machine learning [54]. The more common are described below [54].

- Linear regression: The algorithm tries to find the linear relationship in the data which can be described as Y=f(x) [58]
- Decision trees: An algorithm using decision trees shows visually the decisions with a probability from classification [26]. The way the virtual decisions branches out like a tree have led to the name [59]
- Artificial neural network: This algorithm works by processing an input vector through artificial neurons, which are divided into layers [26]
- Classification: A classification algorithm is predicting in what class/group a data points belong to [34]. This can be used for decision trees, k-means etc [34]

2.6.2 Unsupervised machine learning

Unsupervised machine learning is different from supervised machine learning in that algorithms here are feed with data, but the data are not classified nor labeled [26]. The purpose is to study how the algorithm tries way to create a function/model to discover hidden structures and label the data [51]. The objective is to create a function/model with the ability to categorize new unlabeled data based on the model, this is visualised in figure 10. Unlike supervised machine learning, unsupervised machine learning has no known output or features [60]. This means it is possible to find patterns or connections which were not expected, or completely irrelevant if the data had not been prepared properly [39].

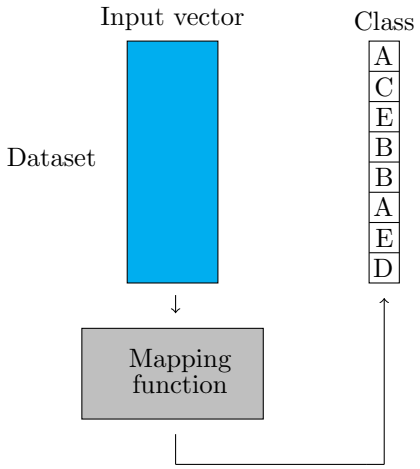

Figure 10: How unsupervised machine learning works. Source: IBM, 2017

Unsupervised machine learning can be applied to several situations where there are large amount of data that needs to be filtered and sorted [39]. Example of this can be association mining, meaning to detect relations between different variables in the data, and latent variable models, to remove unnecessary data or reduce the number of features/dimensions [61]. There are many common algorithms that can be used in unsupervised learning [39].

- K-means clustering: The algorithm is divides the data (n numbers of observations) into k numbers of clusters, which is shown in figure 11 [62]. This can be applied in the detection of cancer [63]

- Clustering methods: The algorithms tries to find groups in the data with similar data points [39]. This is shown in figure 12
- Latent variable models: The algorithm works to find latent variables in the data to detected hidden structures [39]
- Unsupervised neural networks: This algorithm works the same as artificial neural network, the only different is that the data processed by the neural network is unlabeled [39]

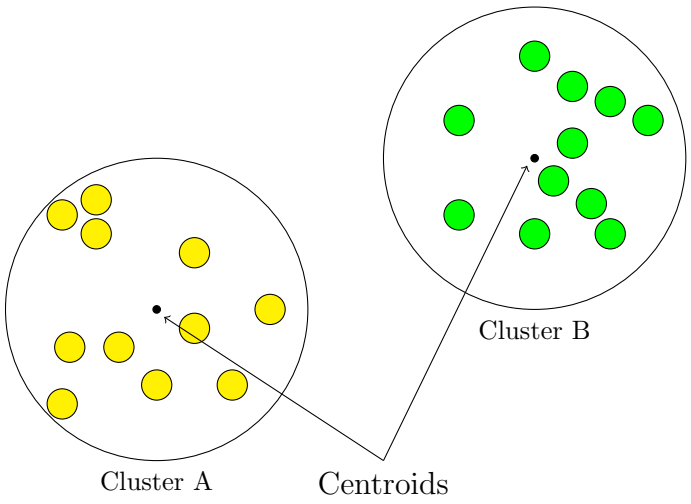

Figure 11: How k-means clustering works. Source: towardsdatascience, 2018

One problem with unsupervised machine learning is that is often does not work properly [39]. Another problem is that it can be hard to evaluate the result from the algorithm since no labeled data is available [60]. Clustering is for example hard to evaluate but can be very useful in practise [64].

2.6.3 Semi-supervised machine learning

Between supervised and unsupervised machine learning is semi-supervised machine learning [51]. The algorithm uses both labeled and unlabeled data to improve the algorithm [65]. Usually a small amount of labeled data and a large amount of unlabeled data, using this ratio allows the models to improve the the learning accuracy [51].

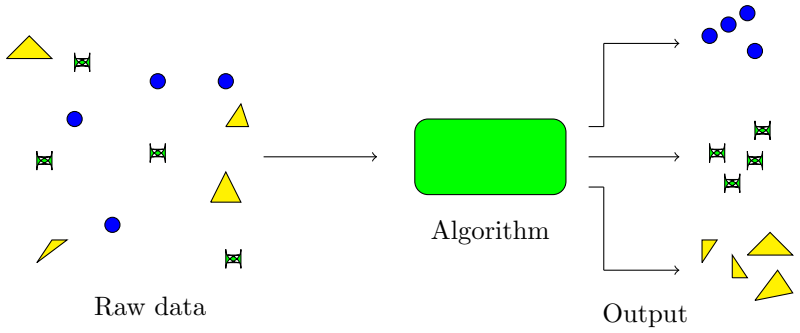

Figure 12: One way to use clustering methods. Source: Data fair, 2019

Semi-supervise machine learning is applied in situations where there are large quantities of unlabeled data and since it takes time and money to label, using semi-supervised is an effective approach [65]. This can be in a bank where there is a need to build data security [66]. Semi-supervised machine learning has had a resurgence during the last years because of its ability to reduce the error rate but is overall a relative new phenomenon [65].

2.6.4 Reinforcement machine learning

Reinforcement machine learning is the fourth way to train an algorithm and is when a algorithm interacts with the environment and learns by trials and error [26]. The algorithm or the agent performs actions freely to execute a task, which, if successful results in a reward and penalty if failure occur [67]. The reinforcement learning can be more closely connected to supervised learning than unsupervised as figure 13 shows.

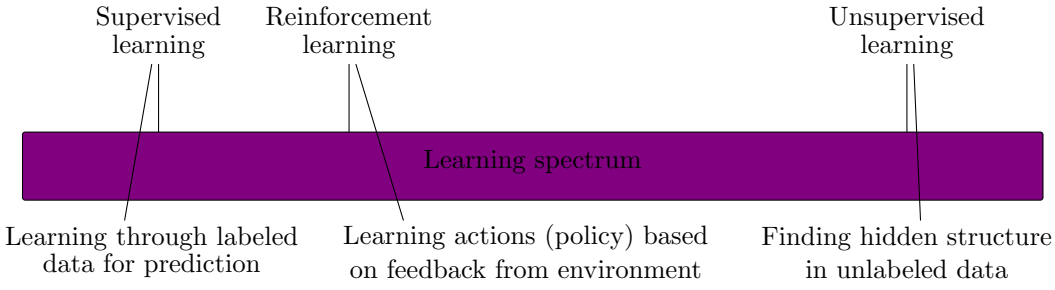

Figure 13: The learning spectrum for machine learning. Source: IBM, 2017

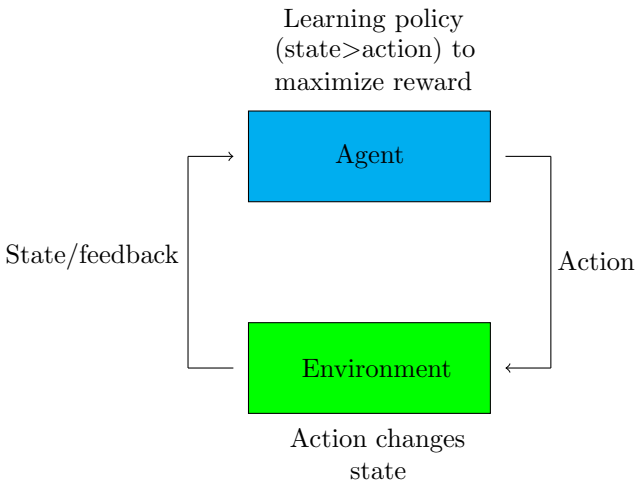

Figure 14: The loop of how an algorithm/agent learns through interaction with the environment. Source: IBM, 2017

The algorithm is not using raw data, but is learned to identify actions in an environment and every action is connected to a state [26]. The endgame is for it is to has a optimized state-to-action as figure 14 shows. The advantaged with this is that it can be used in situations where a model of an environment is known but a solution is not [68]. This can typically be robotics, gaming, self driving cars, traffic light as well as search engine, virtual personal assistants, email spam filtering etc [69][70]. One good example on this is the experiment OpenAI tried, where they used reinforcement learning in a hide-and-seek game for agents [71]. The agents could act freely in a closed environment and after many iterations of testing, they started to learn different strategies on their own to best win rewards [71].

There are many different algorithms for reinforcement machine learning and on of the more common is Q-learning algorithm [72]. A Q-learning algorithm is one form of reinforcement learning, in which Q values are embodied for every state-action pair [26]. This helps the algorithm to understand which path to take to achieve an objective and to receive a reward [26]. These Q values are then updated for every state-action until the algorithm reaches it goal [26]. This type of learning is well suited for gaining knowledge and to take decisions in uncertain environments [26]. This is depicted in figure 15.

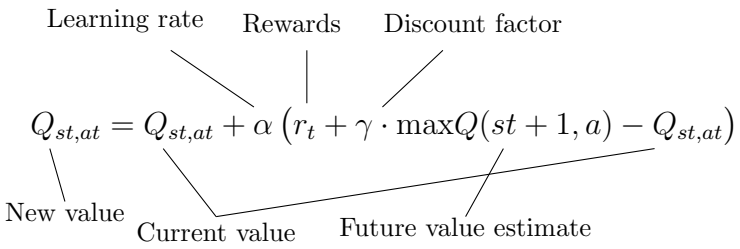

Figure 15: How a Q-learning algorithm works in general. Source: IBM, 2017

2.7 Machine learning in the industry/business sector

Machine learning can be applied to many fields and areas in the business and industrialise sector where data is needed to the processed [73]. A list is shown on some of the applications for machine learning [73][32][74]

- Risk management in financial services
- Medical services
- Geographic image processing
- Facial recognition
- weather data analysis
- Epidemiology
- Spam filtering
- Speech recognition
- Information retrieval
- Virtual personal assistant

Though machine learning not yet has had it entrance into the wider market of the construction industry, it has in two other industries started to become adopted [75][13][76]. Because of the difficulties with finding materials on machine learning in the construction industry and how machine learning can make the design process more efficient, it was decided to look at sectors

where machine learning have been adopted in some areas and see how machine learning has affected the processes and what the consequences of its implementation were. By looking at other industries, it creates a better understanding on how machine learning can be applied, difficulties with finding sufficient and correct data as well as other problem that may arise in the process.

2.7.1 The financial sector

The first industry used as a reference is the financial industry, where some banks have started to use machine learning on algorithms to improve their efficiency on the market [77]. There are numerous applications for machine learning in the finance sector [15]

- Risk management in financial services
- Customer service
- Digital assistants
- Marketing
- Algorithmic trading
- Money-laundering prevention

JP Morgan Chase and CO is one of the oldest and largest bank in the USA [78] and are using machine learning in their business [79]. The ambition is to make better decision based on data analysis as this will give the investment managers an edge [79]. The algorithm will have access to real-time information and can supply with organized data to the investors [79]. Accordingly to JP Morgan, this is now possibly because of 3 things [79]

- Exponential increase in available data
- Increase in computer power and data storage capacity
- Advancement in machine learning methods to analyze large data sets

Machine learning and big data are believe to change how investments are done as the ability to get new feeds and other data from the market analysts fast and to make trade from these analysis instantaneously which will

increase revenues [79]. Other advantages are reduced operational cost, improved user experiences for the customer and improved security [75]. This have lead to many large financial institutions believing that machine learning is one of the more important areas to invest in [75]. In a survey conducted the by Bank of England, a majority of the respondents (companies in the financial service industry in UK) that they are using machine learning and that it has had huge benefits for their businesses [80]. The report concluded that machine learning will only increase in the sector and follow up surveys are needed to better understand how machine learning will change the sector and how regulators can get an insight into how machine learning executes its performance [80].

In Sweden, banks like Nordea, Swedbank and SEB are in progress with using machine learning [81][82]. Nordea's ambition is that machine learning will effective and quickly analyze document from different central banks around the world, which can be long and difficult to process if done by humans [81]. This will help predict interest-change and will be available for Nordea's market customer [81]. SEB believes that machine learning can help them increase the customer services, sales [82]. In 2017 they deployed a virtual co-worker in order to increase availability [83]. Swedbank is using machine learning to help their customer to plan their private economy better and to shorten the time it takes to evaluate the credit rating for companies [84].

2.7.2 The medical sector

The second area where machine learning have started to get attention, and has been used for some time, is the medical sector[63]. Some of the applications for the medical world are [14]

- Diagnosis
- Personalized treatment
- Drug discovery
- Clinical trial research
- Radiology
- Epidemic outbreak prediction

In a paper by Greek researchers, they discuss the use of machine learning to understand cancer progression and to model the treatment of cancerous conditions. With new technology and and large amount of data for the research community, machine learning can help with predicting different cancer types and improve the accuracy of the cancer susceptibility and the survival prediction. By including variables as family history, diet, weight, lifestyle etc which is important, the algorithm increases its predicting for the risk of cancer [63]. In an other article on nature, a group of American researchers discuss how deep learning can be used for early detecting of breast cancer [85]. Accordingly to the authors, the development of deep learning have lead to an interest for it and the application it can have on medical imaging problems and their model that they built could achieve a sensitivity of 86,7% and specificity of 96,1%. which is considerable [85]. The study showed that training the model could have a high accuracy and will only be better as more data is collected [85].

Beyond detecting cancer can machine learning be applied to treatment of different sorts of diseases. In January 2020 did a group of British research use machine learning to find an effective antibiotic [86]. By using machine learning, they manage to create an algorithm that have been processing through 2500 drugs and natural substance and how they affected the growth of E coli. After the algorithm was familiar with the substances, it went through a library of more than 6000 substance under investigating for treating human diseases and were task with looking for a substance that was not similar to existing antibiotics but still effective, doing so reduce the risk of the bacteria having develop a resistance against the new substance. One of the authors of the study, Jonathan Stokes said that this would not have been possible if done by humans, as going through all compounds and testing would have taken far too much time to be reasonable [86].

There are one big similarity for every industry that have enable them the use of machine learning and that is data and in large quantities as data is essential for training an algorithm [87].

2.8 Machine learning in the design phase in the construction sector

The design phase of the construction is the phase where the preparation for the construction takes place. Calculations for the buildings structure are made, planes and routines are drawn up, discussions with suppliers are being done etc. Resources are required in the design process, but more importantly, it is in the design stage where the resources are plan to be used in the construction. The construction industry in general includes a high risk for projects and a low return in profits [88]. This may change to a degree with machine learning entering the construction sector [13]. Machine learning has yet made its entrance into to the wider market in the construction industry as it is still one of the most under-digitized industries but has great potential for the industry [13]. Machine learning is however expected to increase its market shares in the construction industry as demonstrated in figure 16.

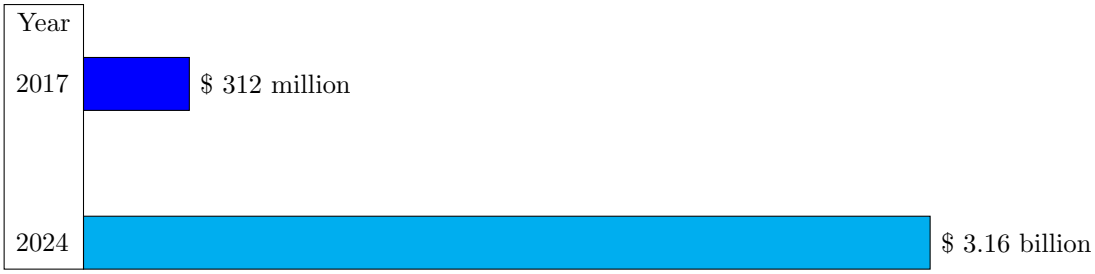

Figure 16: How the global market for AI is expected to grow from 2017 till 2024. Source: Global AI-In-Construction Market, 2019

There are a number of areas in which machine learning can be applied and can in many ways be seen as one of the next technological frontiers for the construction industry to explored [89]. Examples of areas are [45]

- **Risk mitigation**: Using machine learning to identify potential quality risks, making the project team more alert [90]. Doing this can reduce costs and increase profit for the company [45]. The algorithm would also be able to predict every sub-projects consequence if failure occur to the project and give them a risk score, furthering helping the project team [91].
- **Project planning**: Scanning the construction sites and feed the data to a neural network would allow the algorithm to understand when and

how the different sub-projects would execute and how to manage any problems along the way [92]. This would quickly help the project team to prevent problems of different natures before they even occur [90]. Besides that, the ability to use algorithms for repetitive tasks would be more efficient than if performed by humans [92]. Therefore using algorithms to monitor human productivity to provide guidance would help to increase productivity [92]. Other possibilities could be algorithm assisting with BIM to improve processes in the work, eliminating some inefficiencies [93]

- **Safety on the workplace**: The construction industry has the most casualties among industries in Sweden [94]. Reducing the accident frequency for personal would make the construction sites a safer place for people and reduce costs for society [95]. This can for example be done by letting the algorithms look on data from previously projects with data about where and when accidents happen and create a list of probabilities of where and when accidents will happen. This is helping the project team to identify these risk and to take actions against it. The ability to identify risks will be even greater with real-time footage on the sites, as visibility and transparency are important for the algorithms, that is however done on the construction sites and not in the project phase [96]

- **More realistic costs for projects and prevent cost overruns**: Many larger project goes over budget and delays which can be prevented with machine learning [90]. Utilizing machine learning can result in more realistic time frames and cost analysis, giving a more accurate picture of the project [90].

- **Better and safer designs**: Generative design is a process to find the most optimized solution by using an evolutionary approach and as well fulfilling the the parameters set up by the project team [3]. This is done through a digital platform as building information modeling for example or any other digital platform [3]. Through lots of iterations, the algorithm will find a 3D model that gives engineering, architecture, electrician, plumber etc access to plans, the sequence of activities and time of execution [92].

The requirements for the used of machine learning in the construction industry are the same as for the other cases, and one important thing besides the

access to data is the ability to share data with other [91]. Creating a platform for a integration between parties would allow construction companies to analysis large amounts of different data on a common ground [91].

It is not secret that the construction industry has a lower productivity increase compared to other industries, with around 1% vs 2.8% for the hole economy [13]. It lags in innovation, projects often takes longer time than predicted and many times goes over budget [97]. As the population and urbanization in Sweden increases, there is a need to stream line the industry to become more efficient [97]. Streamlining the industry increase the financial returns as well as preparing the industry to face future challenges [97][98]. Today the building sector in Sweden has a market share of around 11-12% of the GDP on average [2]. For 2018, this was 11,3% and haven relatively steady been rising from the record low level of 6,2% in 1997 [2]. Given the industry's size to the economy, it is important that the industry stream line itself to make it more compatible [1]. Much of this improvement can be done in the design phase. In an article written by Mckinsey& Company [1], the authors discuss several options do execute this. One promising option is to build up libraries of data from previously projects and to standardize the design elements. Doing this improves the design process [1].

The construction industry have been slow to adopt new technologies and that is one of the reason that productivity has not grown near as much as the rest of the economy since WW2 [13]. It appears that it is a lot of room for improvement in the construction industry and that machine learning can help increase the productivity, innovated and introduce new techniques [89]. Globally, construction companies only invest around 1% of the their resources into technology, significantly less than other industries, such as financial or manufacturing [89]. A global survey executed by KPMG reveals that many in the construction industry recognise the promise of utilizing new technology to increase value for the company but few have a strategy to implement these new technology [99].

There is also a different between age in the industry. Younger people and especially millennials are finding it important that being updated with the technology is important for the company [99]. This can be compared to the older generation whom have not grown up with the internet and are not equally custom to quickly have the availability to access information [99].

Besides keeping graduates interested in the construction industry, it is of a great importance to satisfy the customers and to be competitive.

2.8.1 Safety

Accidents is a problem that is plaguing the construction industry [100]. Even with decades of experience and statistic to know how and where accidents are happening, it is difficult to predicate them in the design phase as their are too many variables. Humans has a limit to the amount of information a person is able to process and handling hundreds of details, small issues etc can be difficult under time pressure. Even a person with experience can under stressful circumstances miss details. Machine learning on the other hand has the possibility to assist the project team in predictions on safety risks, done through processing data of previous projects and accidents [88]. Given the quantity of data generate, it is close to impossibly for humans to process that amount of data as good as algorithms.

According to Arbetsmiljöverket, the Swedish ministry of work environment, the construction industry has the most work accidents with fatal outcome [100]. This creates a financial burden for both the companies and the society as well as personal sorrow for friends and families [95]. The companies are gathering data about all accidents to better prevent them from happening at all, however machine learning has a much greater ability to understand the data and to predict where the accidents might happen [100]. In a project where machine learning will be tested, Mälardalens högskola, NCC Svergie AB and Chalmers tekniska högskola are hoping to create an algorithm that will help NCC to reduce the numbers of accidents in the industry by looking through data over previously accidents. This will also increase the production (for the company) and value (for the customer) in the long term accordingly to one of the participants [100].

Amongst companies working to utilize machine learning is Autodesk, which is a company currently working on a project, project IQ, where they are trying algorithms in real projects to help project teams detected and mitigate risk [88]. Project IQ has access to millions of construction observations, building information models, project outcome etc to learn from and the purpose is to creates an engine to make predictive analyses [88]. With this, the risks will be detected in an early stage and classified into actionable categories

according to Autodesk [88].

2.8.2 Time

Many projects fails to keep up with the time frame that have been projected in the design process [97]. This leads to economical loses and reduced trust from the buyer, which is damaging the company's reputation. The problem is especially effecting the larger construction projects (skyscrapers, bridges, tunnels etc), where a majority fails to kept the original time table and are delayed [97]. As mention before regarding the proposal to build up a digital library, the potential for training algorithms would be enormous [1]. However it is not just designing element that machine learning will be able to do. Machine learning would have the ability to improve the project planning and optimize the the schedules with more realistic schedules [101]. Having more reliable schedules will decrease the uncertainty for both the customer and the company. Another important factor is to optimizing the supply chain, which would help reduce the time needed for manufacturing and increase the predictability [101].

2.8.3 Cost

A concern for the industry is the low margin of return for the companies [88]. Together with high risks, this can be seen as something that is hampering the development of the sector [102]. Even though a number of unpredictable events can occur, most of all actions can be planned in the design phase in the early days of a project but the ability to influence cost decreases as the projects moves forward. Figure 17 shows the connection between cost and the ability to change cost.

Training the algorithm to understand inputs and outputs can help the project team to understand cash flow through a project and decrease the risk of cost overruns. This is important as many projects have cost overruns and/or delays [88]. The project team would have the ability to optimizing routes for transportation and improving traffic navigation, reducing the time for the building material and equipment on the road [89]. In the end, reducing accidents, optimizing the manufacturing process and working with more realistic time frames will help projects to become more innovative and profitable.

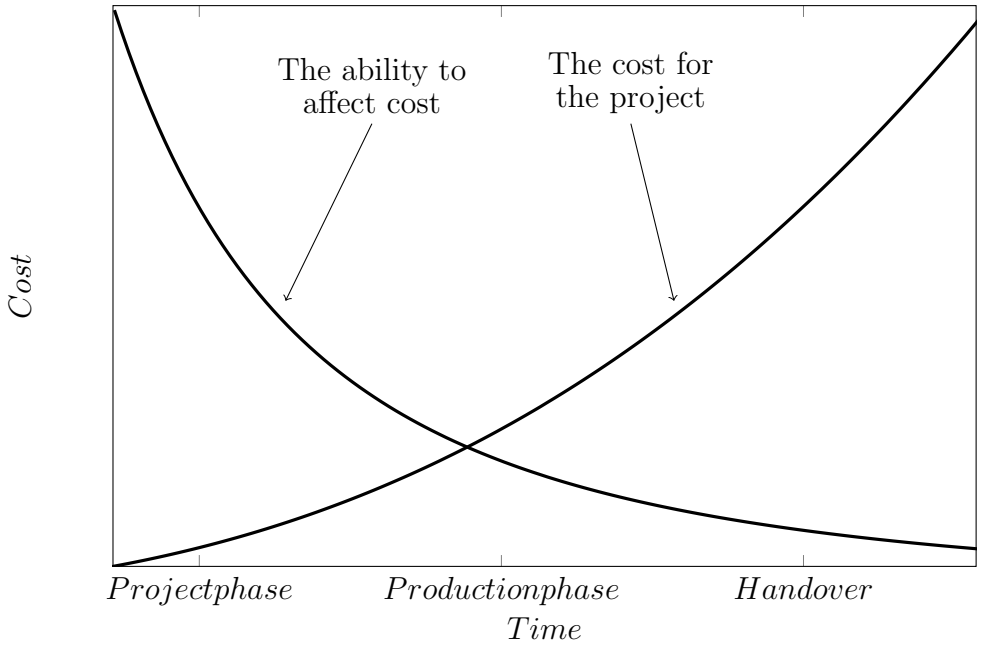

Figure 17: The corresponding lines between the ability to affect cost in a project and the cost of the project.

2.8.4 Limitations

Despite the benefits of machine learning, it is important to remember that machine learning is not a flawless technology without limitations and implementing the technology will be challenging [103]. Their are a number of things which should be known about machine learning that may limit its success [103]

- Ethics: How will machine learning affect algorithms ethically and who is to be blamed if the algorithms fail to achieve its task. Failure that can mean both human and financial losses.
- Data: A algorithm is only as good as the data that it has been processing. Lack of data will strongly hamper the progress of machine learning and is currently one of the limiting factors for the use of machine learning.
- Interpretability: How to interpret and evaluate the output from an algorithm can be challenging, but to achieve success it is important to

for the result to be intelligible. This may be easier said than done but in order to be able to be integrated into the business sector, focusing on both the algorithm's training behind to learn the it to be understood by people is of upmost importance. As stated in "Business Data Mining — a machine learning perspective":

A business manager is more likely to accept the [machine learning method] recommendations if the results are explained in business terms

2.8.5 Implementing machine learning

The use of machine learning in the construction industry is still in its infancy but starting from the beginning may not be necessary as their are a few early-stage examples that can be tried [89]. These first test with machine learning will likely be optimization of schedules or analytic process of previously project to see how margins can be increased [89]. As a report from Mckinsey points out, machine learning is based on learning and needs a certain amount of data to start be of used, which many construction companies are lacking [89].

However algorithms do not have to be created by construction companies and only have to be used for designing structures. Algorithms can bought as a product and be integrated into software, used as calculation tools by constructions companies. It is therefore important to note that there is a difference in how machine learning is implemented and to what extent.

2.9 Companies working to digitize the construction industry

2.9.1 Autodesk

Autodesk is an American company working with software services for architecture, engineering, construction etc. The company is incorporating new emerging technologies such as 3D printing, artificial intelligence, generative design and robotics. It was founded in 1982 and has its headquarters in San Rafael in California.

2.9.2 BIMobject

BIMobject is a company based in Malmö, Sweden, and they are working to digitize products from the construction industry through a cloudbased platform. Digitizing the construction industry will accordingly to them enable lower costs, more precise design phases, more sustainable solutions and more efficient property management.

3 Method

3.1 Litterateur

As the interest in machine learning and AI has risen in the last years, so have the number of written articles, books and documentaries about the subject. In order to find information about machine learning, google, LTH's library and LTH's search engines had to be used. Usual search key words were

- Machine learning
- Machine learning applications
- Machine learning in the construction industry
- Machine learning algorithms
- Machine learning in the industry
- Digitizing

These were some of the more common used key words on both google and LTH's search engine to find information about machine learning. Further more, a book about machine learning written by Peter Flach, describing and explaining the math and science behind algorithms. Searches for other master thesis about machine learning in the construction industry were unsuccessful.

3.2 Methodology

To understand how machine learning can be implemented and what challenges are laying ahead, this master thesis will include two sectors where machine learning already have been utilized. The two reference industries/areas that have been chosen is the financial and medical sector. Showing how machine learning have been implemented into these sector on varies level can help construction companies on how machine learning can improve certain processes in the design phase in terms of time, safety and capital.

The method for this master thesis will be carried out in several steps

- Preparing the hypotheses that machine learning can be used to optimize the design phase of a construction project

- Conducting literature study about machine learning and how it can be applied in the design stage of a project. The focus will be on this step as machine learning in the construction sector is a new frontier that has not really been explored on any depth
- An interview with a party in the financial sector
- An interview with a party in the medical sector
- An investigation with seven parties in the construction industry
- Analysing the data
- Drawing conclusions from the data
- Present the process and conclusions made of this master thesis

This process over a semester or 20 weeks.

3.3 Collection of data

The data needed will be collected through an investigation. The investigation with the parties in the financial and medical sector aim to understand how machine learning has work so far and to understand how demanding it was to set up the algorithms in the first place and acquire the competence to do so. It was first decided to conduct a interview with both the medical and financial sector but the outbreak of Covid 19 change those plans to a virtual interview, where the questions were sent and they replied when they had the time for it. The investigation with the companies in the construction sector will have the purpose to investigate the attitude towards new technology such as machine learning and if there are a desire for an acquisition of this technology. The questions that will be answered are of similar types as the ones for the banks and medical sector. The questions will be sent to people, working with BIM or has a higher position in the companies, giving the replies some legitimacy. All representative will be anonymous.

The two parties in the medical and financial sector had been selected after some investigation revealing their progress in utilizing machine learning. There was not a lot of information about how far they had progressed, only that they had implemented machine learning to some degree. Given the

hardship with finding organizations working with machine learning and be willingly to participate in a investigation of this types, they were chosen. Regarding the construction companies, they were somewhat randomly chosen, as there were little to no information about machine learning connected to them. Many of them have publish information about their work to innovate the sector but it was unclear if machine learning was included into their plans. NCC had be mention in an article regarding their cooperation with a project about machine learning and safety [100]. The only criteria was that they were some of the largest construction companies in Sweden.

The data (the replies from the parties) will be gathered together in their respective field. The medical and the financial sector in their respective section and the construction companies in another section. As the respondents from the medical and financial sector are adding in their own comments and thought, they idea is to get an overall picture of how their organizations have benefited from using machine learning. If someone points out something of value, it will be noted in the result. The data is to be analyzed, by looking at what they wrote and discussed. The result will be present in a mix of the questions together with the replies and floating in a text.

3.4 The reference industries

The questions are showed in Appendix but the heads lines are shown below. The first questions are general in nature on machine learning and gives a bit background on the person answering the questions. Knowing how much effort the organizations put in utilizing machine learning will help predict cost and time for a construction company to spent should the company be interesting in acquire the technology.

General questions

If the acquirement was internal or external

Data and models

Here the purpose is to understand how the technology was acquired, what kind of model was used, how is was implemented and why. This is because it is not always clear what type of model that should be applied to what sit-

uations and/or how it will performs. Therefore knowledge can be extracted from previously experiences.

Implement of machine learning

Threats against machine learning

The last questions are on how well the model have functioned while operating. The ambitions are that these questions will give answers to how different parties have implemented machine learning successfully into their organizations. Understanding how machine learning have been utilized will give this master thesis guidance in how to help to implement machine learning in the construction industry.

3.4.1 The medicine sector

For the medical sector has Vårdgivare Skåne be chosen and has implemented machine learning in form of voice recognition. Vårdgivare Skåne or Caregivare Scania is a organization responsible for the medical situation in region Scania for both private and public use. They started with voice recognition to make it simpler to write (or speak) journals. Since writing journals takes a lot of time, the idea was to implement a system that could reduced the time required for it. Other positive effects would be that the journal is digital updated instantaneously, which lets the patient have access digitally to the journal after a meeting and increases the patient security.

3.4.2 The financial sector

Two banks to be questioned for this master thesis, Nordea and Handelsbanken. Both those banks uses machine learning for a numerous activities, including customer service and internal service. The questions were sent at the same time as the questions for Vårdgivare Skåne. Both the banks are using machine learning for external service. The public information for both reveals that the ambitions are to increase customer service and satisfaction.

3.5 The construction sector

Seven companies have been chosen to evaluated. Most of these company have not public expressed a desire to acquire the technology but given the resources and need, these companies are some of the more likely to look into the acquisition of machine learning algorithms.

- Wästbygg
- Peab
- Skanska
- Serneke
- JM
- NCC
- BIMobject

These first sex companies are some of the largest construction companies in Sweden and of that reason, the more likely to implement machine learning into the core business as they have the resources required for it and most to gain from it [104]. The seventh company, BIMobject has been included because of its work to digitize the construction industry, but itself is not a construction company. The questions for the construction company will be a little different compared to the questions asked to the financial and medical sector. The reason it here that to understand how the construction companies views around digitizing and machine learning. The head lines are shown below.

General questions

If BIM is used

If BIM is not used

If machine learning is being used/tested

If machine learning is not being used/tested

The near future

The questions have been divided into section where one section focus on questions for those whom are using machine learning for those how are not. The questions have been selected in a way to investigate to understand the companies mindset, if they are interested in using machine learning, if they have any ambitions to utilize it (in the near future) etc. A second section of questions regarding BIM has been added as well as. The purpose is to cover the possibilities and combinations of the use of BIM, machine learning or non of them. This helps to understand if the company has any ambition of digitizing its business, where in the process of going digital are the company (if it is interested) and if the company has an interest or a long term goal to integrated machine learning into the business core.

3.6 Validity and reliability

The credibility of any investigation builds on validity and reliability This investigation can be seen as a survey on a much smaller scale with on a few participants. This method has be chosen to be as reliable as possible but it is its execution which will determine how reliable the conclusions will be. How many replies which will be received do also affect the reliability of the investigation. It is expected with some lapse in the replies as this happens parallel to the current situation with the 2020-present Covid 19 pandemic and the financial stresses it has put on the economy. With a total of maximum of seven replies from all sectors and only three from the construction sector, how valid should these replies then be seen? There are multiple factors which affects the validity and reliability for the investigation.

- How the questions were formed: Forming wrong questions or focusing on less important areas can lead to inaccurate data and in extent, wrong conclusions. Therefore it is crucial to have a clear picture of the situation together with a clear aim. From the aim can the questions be derived.

- Commitment of the representative: How serious were the representative when they replied to the investigation and what position did they held in the organization they represented.

- Representation: How representative is it for the replies to represent the mindset of a broader portion of companies in the construction sector. The replies comes from some of the largest construction corporation in Sweden and have a lot of influence in the construction sector but that does not mean an unique voice for the entire sector.

In the report, a numerous sources are used to gather information, such as articles, research reports, literature, web-pages etc. It is therefore of upmost importance to critically review all sources used, and the writers behind them. To understand how they are relevant for the master thesis, if the information is reasonable and if the information can be back up by other sources. When the sources were chosen, it was regard to the year of published, articles more than 10-15 years seen too old and not up to date. This was done because of the intensity of change which is happening to the technology, as the progress is rapid for any digital technology. This made it important to have new published sources which are up to date and currently of interest. A lot of focus was made to find multiple sources on information on the theory of machine learning to increase the reliability. The technology has picked up a lot of speed in the last years and there were much information available which help to increase the reliability.

Some information were pick from corporations. As they have their own agenda and not always may be reliable, the information were carefully inspected. However that did not appear to be a problem as several companies were researching on the topic and their information were easily verified. Information from articles could be more difficulty to verify, however scientific articles in general are often peer-review which helps them to become more reliable.

Besides to consider the factors about the investigation, so is it important to consider the factors involved in the method used. There are pros and cons with using a method of this type. Advantage with method are multiple

- It can be done through email, meaning the questions can be sent a any time and be replied when the respondents have time to answer properly without an immediate time pressure
- It can gather a lot of data without the risk of any loss of information during the transition back from the respondents, which can happen

during a conversation where a lot of information is shared

- It is easy to create a form and takes little effort to change it should their be a mistake or if the respondents are asking to clarify one or more questions

The cons on the other hand are multiple as well with using this method

- If the questions are focusing on wrong areas or missing questions on a specific subject, can result in inaccurate data and in extent, wrong conclusions
- The respondents can decide to skip some questions, which results in missing data
- The questions may not be able to reflect the respondents possible answer to previously or later questions

These factors as the previously mention affects For the medical and financial sector, the replies will be even fewer than from the construction sector but they will highlighted how machine learning can be used. With that said, it is crucial to point out they replies reflected their experience which may or may not reflect others struggle to implement machine learning with their success. Since the investigation for the construction companies only received replies from a maximum of five respondents, the reliability of any drawn conclusions can be disputed but not dismissed. Chapter 2.8 talked about how the construction industry in general are behind from a technological perspective [13]. How the replies will correspond with the data presented and if it will strengthen both the reliability of the investigation and the validity of the conclusions are yet to be seen.

3.7 Method criticism

As with all thing, there are pros and cons with choosing a method to evaluate a number of questions. It was decided that a investigation would be way to gather information, as the were numerous representatives who were of interest for this paper. A investigation is a practical way to collect information from numerous companies/other parties and gives the representative the same prerequisites. It was considered to interview all parties but that suggestion was eventual dismissed. The time required for such an operation

was simple too much. A factor which may draw criticism is the formulation of the questions. As the the investigation was divided into two forms, one for the medical and financial sector and on for the construction sector, the questions were not individual customized. This problem was somewhat diminished when the questions were formed in such a way that they cover all aspects of the use of machine learning (for the medical and financial sector). However with only two participants, the data gather from them will be limited. For the construction companies, the questions cover both the cases of using/not using BIM and using/not using machine learning.

A negative aspect of a investigation is no physical contact occurs, preventing a judgement of the representative. This however happen to become a positive aspect as the coronavirus Covid 19 started to spread around the global, becoming a pandemic. This would have prohibit any physical interaction with the representatives. Despite the cons with using investigation, it is still considered to be the best option here as it allows to gather lots of information from the participants given the circumstances and is on the same conditions for everyone. The result can be presented on equally terms and with same measures, from which conclusions easier can be drawn.

4 Results

4.1 The medical sector

Vårdgivare Skåne was buying the algorithm for voice recognition from a company and therefore has no knowledge in the creating and training of the algorithm. However the consequences of implementing it was a success and Vårdgivare Skåne have a lot to say about the new system.

1. Can you tell a little about your background and how you came to work for this company/organization?

   The respondent has a long history of working for Region Skåne and started as a medical Secretary.

2. Can you tell how long you have been working for the company/organization and what you position is?

   Almost 30 years and currently has the position as system administrator.

3. How long have the company been working with utilizing machine learning?

   Vårdgivare Skåne has recently bought the algorithm from a supplier, named Nuance.

Given the fact that the algorithm was bough makes further questions on the list harder to answer. In the reply, the respondent points out the advantage with machine learning is many. After implementing voice recognition, the patient security has increased, the journal can be updated almost instantly, giving the patient the ability to read the journal after the visit. Secretary time is released from journal writing and dictates lines becomes shorter. The algorithm has been of great use and Vårdgivare Skåne has as an organization been pleased with the results so far.

4.2 The financial sector

Nordea able to respond to the questions, no response was given from Handelsbanken. However, given the position of the person (and the work/responsibility

that comes with that) and the current Covid 19 outbreak, the person responsible for the reply for Nordea was somewhat unavailable. Through conversation however, some questions were able to be answered and some general information retrieved.

1. What have the advantage been with using machine learning?

   What the bank is hoping for and working towards is to let the customers or the end users be able to get answers from chatbots at any time with one single point of contact, instead of multiple support channels as of today. The advantage will be better customer service.

2. Was an estimation done before on the profitability of using machine learning?

   Estimations were done and showed that the bank would save resources and increase customer satisfaction

3. How much time and money (roughly speaking) were require before the algorithms could be used (put into service)?

   Costs were taken centrally and not by units, and it took three months before the algorithm could be launch with initial functionality for customer.

4. Are the infrastructure used for machine learning and competence an internal or external service?

   It was an external service and the bank did not program the algorithms

5. How hard was it to collect data (both finding the right data and in enough amount)?

   It varies depending on the data but overall, not too difficult. However it was time consuming to collect the data.

6. Have machine learning resulted in a more efficient process?

   That is too early accordingly to the bank but the ambition is for the algorithms to be so

Even though lot of information is missing, a general picture of the benefits of using machine learning could be understood. The bank is determined to continue to work to improve the use of machine learning and believes it will play a bigger role in the future. Nordea thinks a substitute to machine learning does not exist and with so many applications for it, utilizing it to continue to be competitive is necessary.

4.3 The construction sector

Three companies replied, Serneke (R1), JM (R2) and Peab (R3). The information given by these representatives, shows that they recognise the promise of digitization but only JM is looking to improve the design process with machine learning, though the process is still in it's infancy. The representatives from the companies had different roles within the three companies but all work in one way or another with BIM, digitizing or had a higher position in the company, giving them legitimacy.

1. What is the company's view around digitization?

   All representatives responded that digitizing is important to the company, although it varies between companies and within a company on how the work to digitize proceeds.

2. Is the company working to become more digital?

   All representatives said their respective company is working towards to becoming digitized. However, the work is not always consistence accordingly to the representatives.

3. Do you know about machine learning?

   R2 and R3 respondents that they were familiar with machine learning

   If BIM is used

4. How long has BIM been used and how has it work so far?

All representatives had used BIM for some time now and it has help to improve the business for all.

5. Has the use of BIM optimized the business core more efficient?

   They all believed BIM had help to optimize the business core and that BIM will be a part of the future for them, though R3 points out that optimize is a strong formulation.

6. Is BIM used to collect data from previously projects to improve the process and routines?

   Yes accordingly to the representatives, although in varies degrees.

7. Is the company collaboration with other to build up a digital library of projects and elements?

   The answers were divided here, R2 and R3 did try to build up a library, R3 with another company, named BIMobject. The respondent for R1 was not informed if such efforts were taking place.

If machine learning is being used/tested

8. Has it been hard to collect data to use for training?

   R2 one has started to look into machine learning and it was still to early to say anything. The company was trying to build up its own library which would help them to collect data, mitigating somewhat the struggle to find sufficient data.

If machine learning is not being used/tested

9. Does the company have ambitions to utilize machine learning in the near future?

   R2 said that they have ambitions to utilize machine learning since they believe machine learning will increase productivity, flexibility and help the company to be competitive on the market. R3 said they have

machine learning in the software they have bought. They believe that machine learning will come naturally as things start to digitize, however it was currently not plans to incorporate machine learning. R1 had no information regarding to incorporate machine learning.

10. Is machine learning too costly to utilized?

    No accordingly to the representative who is working with it. It costs but they expect a return in profit when the algorithms will be commissioned. The other two did not know.

11. Is it difficult to acquire the technique and data for machine learning?

    R2 which is working with machine learning already collects data for it and is will continue to collecting. It is however time consuming to find the right data. The other two did not know.

12. What is the main obstacles that is preventing the company from utilizing machine learning?

    Data and knowledge accordingly to R2. The answer continues with saying for the construction sector, it is unawareness about the technology or the over believes for the technology and then, the disappointment when it is not working properly. R3 was focus more on digitizing than machine learning and currently did not working towards utilizing machine learning. R1 company did not know.

The future

13. Do you believe that the construction industry will become more digitized in the near future?

    All representatives agreed that digitizing the business will happen to both their company and with time, the industry. Given the importance, the demands from more clients and the competitive edge, it will be hard for those how do not incorporate this.

14. Do you believe that machine learning will play a bigger roll in the design phase in the construction industry in the near future in order to

be competitive?

R2 and R3 firmly believed machine learning is the future and incorporating it will give a strategic edge over the competitors in the construction sector. However both points out that it depends on how long in the future and the expectations on what machine learning can deliver. R1 did not have a opinion on this.

5 Analysis and Discussion

5.1 General discussion

The construction industry has to a very little extend been digitized and is a factor which complicates productivity increase for the construction industry, compared to the rest of the economy. Machine learning may not be the solution to all problems the construction industry will face in the next few decades but it have the ability to help on multiple level on the design phase (as well in the production, though the production phase is an areas which has not been investigated or analysed in this paper). Machine learning has the potential to help companies to increase productivity, become more sustainability by reducing waste and increased safety, if they can acquire enough data for it [13]. Data may be in short supplies today but will grow with time in today's society, which is only becoming more digitized, and digitized means more stored data [13]. The construction sector has been behind compared to other sectors here but as many in the construction industry realize the promise and benefits in becoming more digital, the use will increase.

All replies recognise the fact that the future will become more digital for the construction industry and platforms like BIM are likely to increase in use as well as in complexity. The investigation showed that many are aware of the benefits of digitize its processes and collecting data for future use. However it also enlighten that the awareness of the concept of machine learning was low to medium, more about that in chapter 6.2. As the global survey executed by KPMG revealed, that many in the construction industry recognise the promise of utilizing new technology but few have a strategy to implement shows that many are not ready be on the forefront [99]. This is unfortunate as it does not appears to exist a strong will to innovate the industry. The replies from the survey did also reveal the same mindset. It can be speculated that the people in change are in not millennials and therefore have not been custom with digitization. After all, the internet is a relative new technology in the society and its abilities are not known to the older generation, the generation that now is in position of running the companies.

So does this mean that machine learning will be absents in the construction industry for the next 10-20 years? The answer is with a high certainty no. Despite the low awareness about machine learning, it seems that as the

technology matures and the world becomes more digitized, more companies will find benefits with the way machine learning can be used to handle large quantities of data set. Companies that fails to adapted new technology will likely not be able to compete with other in the long run as technology is an important part of being innovative and to increase productivity for a company [105]. The information received from the reference sectors, the financial and medical sector gave an insight into how the process went (though incomplete). However to what extent machine learning will be used is highly uncertain and open for debate. As one company in the survey noted that they are already using algorithms but has no ambitions today to fully incorporate it into every section of their business. So even though machine learning will be use to train algorithms in a company, it will perhaps not emboss each section of it.

The answers from the medical and financial contributes shows the struggle as well as the benefits of machine learning. Though the the were a lot of information missing, preventing from building up a full picture of the process, it was clear what their position were regarding towards the utilizing of machine learning. Accordingly to them, machine learning is a technology which is expected to have a profound impact in the way data is being processed. For an organization or business being capable of using data to improve will always have an edge amongst competitors. Understanding and using data will help corporations become more profitable and competitive. With time it will likely be hard to compete without such an edge. It is worth noting that not every organization will use machine learning for profit. For an organization as Vårdgivare Skåne, the purpose is not to generate profit, but to help as many patient as possible with the tools at hand.

What is interesting with Nordea and Vårdgivare Skåne is non of the parties created the algorithms themselves, instead it was an external service. In Nordea's case, they trained the algorithms internally, much thanks to their supply of data available. This results in lost information regarding the creating and training of the algorithms them self, which was unfortunately. What type of algorithm, which learning model were used etc are all unknown variables. However given the nature of the assignments for both the medical and financial sector, it seems probably that in both cases reinforcement machine learning have been used. The fact that neither party had written their own codes put an interesting perspective on the complexity with writing codes

for an algorithm. To write codes for an algorithm can be something the construction companies may not have the necessary knowledge to do. If construction companies should writes their own codes or use an external service will depend on multiple factors, which will not be discussed here (given the magnitude of information, assumptions and situation awareness for every company required). One of the construction companies did only use machine learning which was include in the software but as with the case with the other parties, it was an external service and the company did not reveal information about its software.

5.2 Knowledge about machine learning

The construction sector has a high level of awareness of the benefits with going digital but are still lagging behind compared to other industries when it comes to implementing [106]. Most of the construction companies showed in the survey that they had little to medium knowledge about machine learning and how it can be used to increase productivity. This should not be too much of a shocking discovery as machine learning is still a relatively new technology on the market [31]. This is likely to change with time as machine learning is integrated into different sectors in the economy at a higher rate and the benefits are starting to becomes well known for the society as a whole.

Given how few companies this master thesis included in the survey, they can not be representative for the entire sector. However it is not unreasonable to think that machine learning is absents in the construction sector as the sector is slower to acquire new technologies. Another thing to point out is that some of the representative who replied in the survey for the construction companies were not the decisions-makers and therefore may not know the course which the board have decided for their respective companies to take for the next few years.

5.3 Data

The lack of data is something that will hamper the progress of machine learning and the companies will need to start to collect data on digital platforms before they can start to use machine learning algorithms. All the larger companies in the survey do however use CAD or BIM to increase the efficiency for their projects, and so many are already in the process of collecting digital

data. This data can later be combined with other data sets if the platform can be integrated with other platforms [107]. Many companies besides the large ones may not sufficient data to utilize machine learning which makes it hard to implement the technology [108]. It it therefore not unreasonable to think that a or several common digital platform(s) would help companies to collect data and help each other. One problem appears before that, multiple platforms that are not integrated or cannot be integrated [108]. BIM is the only platform to be discussed in this master thesis but there are numerous options. Trying to integrate them can be a bit tricky as every platform have its own codes and language.

The willingness to share data is another factor to consider. Many of the companies are competitors on the same market and may therefore not want to integrated their platform and share data with others. As data rise in value, it is not unlikely that it will be seen by many as an asset, which leads to the question on how companies will manage it [24]. Sharing it for free with competitors is probably not likely, but cooperation can occur. This was pointed out by one of the companies in the survey.

5.4 The future for machine learning in the construction sector

Machine learning is likely only to increase in use for organizations, companies, institutions etc in other sectors and become more advance as time goes on [109]. Today can algorithms use computer vision to recognize and analyze pictures much better and faster than humans [110]. Machine learning is already today being used in many aspects of a persons life, e-mail filtering, recommending on streaming sites, searches on the web etc. In the financial sector is the use of machine learning started to become widely adopted and is only likely to grow as the technology becomes better [111]. Given its abilities to handle and process data, machine learning will have a edge over humans and continue to further increase productivity for companies and organizations whom manage to incorporate the technology, all around the world [112].

One only have to look at BIM and how it has developed over time [113]. BIM has become widely used in the construction industry and with such benefits as it has provided to the sector [42]. It appears therefore not com-

pletely unreasonable to think that machine learning will have a future in both the society and in the industrial sector. It will be an increasingly important tool for companies as more data becomes available [111]. That leads to the realization that machine learning will sooner or later be adapted by the construction industry and integrated into its activities as it becomes more available on the market. The question is how long it will take before it is widely adopted. To put a time frame on when that will happen is very difficult as their are many factor to consider. The construction industry is neither a single entity, nor starts at the same technological level and companies will most likely start at different time to acquire the technology, not simultaneously. However when the technology is incorporated into a company's business on wider and more advance front, the company will have a competitive edge over other and put pressure on more construction companies to become innovative [92]. This will likely include the acquisition of the machine learning to keep up [92].

This raises another question, why the construction industry is so slow with adapting new technology. That question is however out of the scope of this master thesis but that question needs to be addressed in order to know how and when companies in Sweden will start to become more innovative. Innovative means more competitive on the market and higher profit margins for a company. To acquire the technique for machine learning is one important way to become more competitive [92].

6 Conclusions

This master thesis tried to understand what were required in order to make the design phase in the construction industry more efficient by using machine learning. Throughout the process of gathering data and analyzing the information at hand, conclusions can be made. Machine learning has an efficient way to handle large quantities of data compared to any means before. It has an enormous potential to increase productivity and value to the construction sector, at least on a theoretical level. The construction companies in Sweden (as well as global) are slower to adapt to new technology and have a lower productivity increase compared to the rest of the economy. Here machine learning can be a solution but despite the advantage with machine learning, many companies do not have the interest to acquire this technology yet, or are unaware of the technology. Many are however focusing on digitizing their core business, or at least have ambitions to do so. Companies will with time start to become more digital and will with that have the ability to gather large quantities of data that can be needed for machine learning.

As mention in chapter 2.7.1, three factors have made the progress of machine learning possible, exponential increase in available data, increase in computer power and data storage capacity and advancement in machine learning methods to analyze large data set. When these factors starts to affect the construction sector and lay the foundation for the utilization of the technology, construction companies will likely integrate machine learning on a large scale into their businesses (to what extent it will be integrated and used depends on the people in charge), as times goes on and as more younger people enter the sector. The pressure from them to acquire modern technology will increase with time. It is difficult to put a time frame to when construction companies will start to initiate projects for utilizing machine learning, as there are a lot of factors and considerations for each and every construction company. However companies will probably start to look into the technology in the next decades or so, and this is a process that likely will proceed over several decades as more younger (millennials) are entering the market and as the technology becomes more widely available.

6.1 Further research

This master thesis had a narrow scope and with only 20 weeks, not much time to evaluate all factors. For future studies, including more companies and to go deeper down on why the companies in the construction industry are more slow to adapt new technique and technology would be interesting. Given that many of the people working in the construction sector are engineers, shows how surprising it is for the sector to so slowly innovate. This would be useful as knowing what is stopping the construction companies from utilizing new technology would help the industry to act against it and therefore become more innovative. How machine learning can be applied in the production is an another area which would be captivating to learn about. it is after all just so much that can be planed in the design process, and thing might go differently in the production than planed. Understanding how machine learning can be utilized to handle these deviations, accidents, delays etc would be of great use.

This master thesis did only focus on one digital platform (BIM) but as their are many different digital platforms out their on the market, looking into other platforms and how they can be integrated would also be important. As mention before, the inability to integrate different platforms may be one of the reasons for not be able to collect large data set and in extension, preventing machine learning to be utilized. Looking at how to integrated platforms together would therefore be of interest. Besides bringing data into a common platform, including more focus on data, and how more can be generate for use would also give some insight into how the construction industry can become more digitized. Other questions are how would the companies know what type of data is needed and when to use it. Evaluate what type of algorithms and learning models should be used for different situation. Finally a master thesis on how machine learning more concretely will be implemented, how a company best should collect and share data, how the algorithms are created and trained. How the evaluation is executed etc. Overall are their a lot of studies to look at, as this subject is a relative new research area for the construction industry.

References

[1] F. Barbosam, J. Mishke, and M. Parsons. (Jul. 2017). Improving construction productivity, [Online]. Available: https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/improving-construction-productivity. accessed: 23.02.2020.

[2] B. Organisation. (Jan. 2020). Total construction investments, [Online]. Available: %7Bhttps://byggforetagen.se/statistik/bygginvesteringar/%7D. accessed: 19.04.2020.

[3] B. Kour. (Jan. 2018). The rise of machine learning and ai is improving lives in 2018, [Online]. Available: https://www.smartdatacollective.com/rise-of-machine-learning-ai-improving-lives/. accessed: 25.03.2020.

[4] J. H. Snabe. (Nov. 2015). What will digitalization do to the future?, [Online]. Available: https://www.weforum.org/agenda/2015/11/what-will-digitalization-do-to-the-future/. accessed: 10.03.2020.

[5] M. Bayern. (Sep. 2018). Why machine learning will see explosive growth over the next 2 years, [Online]. Available: https://www.techrepublic.com/article/why-machine-learning-will-see-explosive-growth-over-the-next-2-years/. accessed: 19.03.2020.

[6] N. Routley. (Nov. 2017). Visualizing the trillion-fold increase in computing power, [Online]. Available: https://www.visualcapitalist.com/visualizing-trillion-fold-increase-computing-power/. accessed: 18.03.2020.

[7] J. Zulick. (Aug. 2019). How machine learning is transforming industrial production, [Online]. Available: https://www.machinedesign.com/automation-iiot/article/21838038/how-machine-learning-is-transforming-industrial-production. accessed: 04.03.2020.

[8] I. D. Marco. (Sep. 2019). How machine learning is boosting business growth, [Online]. Available: https://www.business.com/articles/machine-learning-boosts-business-growth/. accessed: 26.03.2020.

[9] S. Rose. (Aug. 2019). Machine learning applications across different industries, [Online]. Available: https://hackernoon.com/application-of-machine-learning-across-different-industries-v27yc30wp. accessed: 23.02.2020.

[10] O. Kharkovyna. (Jul. 2019). 7 reasons why machine learning is a game changer for agriculture, [Online]. Available: https://towardsdatascience.com/7-reasons-why-machine-learning-is-a-game-changer-for-agriculture-1753dc56e310. accessed: 25.02.2020.

[11] A. Gonfalonieri. (Nov. 2019). Deep learning algorithms and brain-computer interfaces, [Online]. Available: https://towardsdatascience.com/deep-learning-algorithms-and-brain-computer-interfaces-7608d0a6f01. accessed: 24.02.2020.

[12] J. Morgan. (May 2014). A simple explanation of 'the internet of things', [Online]. Available: https://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#43fd64a11d09. accessed: 23.04.2020.

[13] Mckinsey&Company, "Reinventing construction: A route to higher productivity", p. 20, 2017.

[14] D. Faggella. (Mar. 2020). 7 applications of machine learning in pharma and medicine, [Online]. Available: https://emerj.com/ai-sector-overviews/machine-learning-in-pharma-medicine/. accessed: 17.02.2020.

[15] A. Eisenberg. (Nov. 2018). 7 ways fintechs use machine learning to outsmart the competition, [Online]. Available: https://igniteoutsourcing.com/fintech/machine-learning-in-finance/. accessed: 15.02.2020.

[16] A. Saxena. (Mar. 2019). What is data value and should it be viewed as a corporate asset?, [Online]. Available: https://www.dataversity.net/what-is-data-value-and-should-it-be-viewed-as-a-corporate-asset/#. accessed: 23.02.2020.

[17] J. F. Puget. (2016). What is machine learning, [Online]. Available: https://www.ibm.com/developerworks/community/blogs/jfp/entry/What_Is_Machine_Learning?lang=en. accessed: 04.02.2020.

[18] P. Flach, *The LaTeX Machine learning. The art and science of algorithms that make sense of data*. Grafos S.A. Arte Sobre Papel, 2012.

[19] Y. Guo, *The 7 steps of machine learning (ai adventures)*, 2017. [Online]. Available: https://www.youtube.com/watch?v=nKW8Ndu7Mjw, accessed: 28.02.2020.

[20] M. Heller. (May 2019). What is machine learning? intelligence derived from data, [Online]. Available: https://www.infoworld.com/article/3214424/what-is-machine-learning-intelligence-derived-from-data.html. accessed: 22.03.2020.

[21] J. Hurwitz and D. Kirsch, *Machine learning for dummies*. John Wiley& Sons, inc., 2018.

[22] (2018). What is machine learning, [Online]. Available: https://www.ibm.com/topics/machine-learning?mhsrc=ibmsearch_a&mhq=machine%20learning%20history. accessed: 19.03.2020.

[23] Y. Guo, *What is machine learning? (ai adventures)*, 2017. [Online]. Available: https://www.youtube.com/watch?v=HcqpanDadyQ, accessed: 28.02.2020.

[24] I. Larru. (Apr. 2018). The rising value of data, [Online]. Available: https://www.ie.edu/insights/articles/the-rising-value-of-data/. accessed: 05.02.2020.

[25] IBM. (2018). Data science and machine learning, [Online]. Available: https://www.ibm.com/analytics/machine-learning. accessed: 27.01.2020.

[26] M. T. Jones. (2017). Models for machine learning, [Online]. Available: https://developer.ibm.com/articles/cc-models-machine-learning/#reinforcement-learning. accessed: 02.02.2020.

[27] P. Madan. (Mar. 2020). An introduction to deep learning, [Online]. Available: https://developer.ibm.com/technologies/deep-learning/articles/an-introduction-to-deep-learning. accessed: 21.04.2020.

[28] S. Sattel. (2016). How the integrated circuit works: Everything you need to know (and moore), [Online]. Available: https://www.autodesk.com/products/eagle/blog/integrated-circuit-moores-law/. accessed: 19.02.2020.

[29] P. Chojecki. (Feb. 2019). Moore's law is dead, [Online]. Available: https://towardsdatascience.com/moores-law-is-dead-678119754571. accessed: 27.03.2020.

[30] B. Marr. (May 2018). How much data do we create every day? the mind-blowing stats everyone should read, [Online]. Available: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#6b7d7bbd60ba. accessed: 30.03.2020.

[31] L. Columbus. (Jan. 2020). Roundup of machine learning forecasts and market estimates, 2020, [Online]. Available: https://www.forbes.com/sites/louiscolumbus/2020/01/19/roundup-of-machine-learning-forecasts-and-market-estimates-2020/#2a57f9665c02. accessed: 11.04.2020.

[32] R. Desai. (Jan. 2020). Top 9 algorithms for a machine learning beginner, [Online]. Available: https://towardsdatascience.com/top-10-algorithms-for-machine-learning-beginners-149374935f3c. accessed: 13.04.2020.

[33] S. Ray. (Sep. 2017). Commonly used machine learning algorithms (with python and r codes), [Online]. Available: https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/. accessed: 05.03.2020.

[34] S. Asiri. (2018). Machine learning classifiers, [Online]. Available: https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623. accessed: 21.02.2020.

[35] R. S. (Nov. 2019). Understanding dimensionality reduction for machine learning, [Online]. Available: https://towardsdatascience.com/understanding-dimensionality-reduction-for-machine-learning-ad9a3811bd89. accessed: 19.03.2020.

[36] N. S. Chauhan. (Oct. 2019). Introduction to artificial neural networks(ann), [Online]. Available: https://towardsdatascience.com/introduction-to-artificial-neural-networks-ann-1aea15775ef9. accessed: 16.03.2020.

[37] G. Saporito. (Sep. 2019). What is a perceptron?, [Online]. Available: https://towardsdatascience.com/what-is-a-perceptron-210a50190c3b. accessed: 04.02.2020.

[38] M. Chatterjee. (Feb. 2019). Top 20 applications of deep learning in 2020 across industries, [Online]. Available: https://www.mygreatlearning.com/blog/top-15-applications-of-deep-learning/. accessed: 22.03.2020.

[39] M. Heller. (Aug. 2019). Unsupervised machine learning explained, [Online]. Available: https://www.infoworld.com/article/3429017/unsupervised-learning-explained.html. accessed: 03.03.2020.

[40] R. Shaikh. (Oct. 2018). Feature selection techniques in machine learning with python, [Online]. Available: https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e. accessed: 23.03.2020.

[41] J. Hall. (Jul. 2018). Top 10 benefits of bim in construction, [Online]. Available: https://connect.bim360.autodesk.com/benefits-of-bim-in-construction. accessed: 26.03.2020.

[42] S. Lorek. (Jul. 2018). What is bim (building information modeling), [Online]. Available: https://constructible.trimble.com/construction-industry/what-is-bim-building-information-modeling. accessed: 27.03.2020.

[43] H. Cherkaoui. (Dec. 2016). What is bim? what are its benefits to the construction industry?, [Online]. Available: https://www.letsbuild.com/blog/what-is-bim-what-are-its-benefits-to-the-construction-industry. accessed: 23.03.2020.

[44] S. Paul. (Feb. 2018). What is bim, [Online]. Available: %7Bhttps://www.geospatialworld.net/blogs/what-is-bim/%7D. accessed: 27.02.2020.

[45] M. Venugopal. (May 2018). What is ai and machine learning in construction? our definitive answer, [Online]. Available: https://connect.bim360.autodesk.com/what-is-machine-learning-in-construction. accessed: 16.04.2020.

[46] M. Galay. (Oct. 2019). How ai is improving bim and helping the construction industry, [Online]. Available: https://www.kreo.net/blog/how-ai-is-improving-bim-and-helping-the-construction-industry. accessed: 26.03.2020.

[47] NBS. (Aug. 2016). What is building information modelling (bim)?, [Online]. Available: https://www.thenbs.com/knowledge/what-is-building-information-modelling-bim. accessed: 13.04.2020.

[48] K. Johnson. (Feb. 2019). Autodesk launches construction iq to analyze building project safety, [Online]. Available: https://venturebeat.com/2019/02/26/autodesk-launches-construction-iq-to-analyze-building-project-safety/. accessed: 16.04.2020.

[49] Autodesk. (Mar. 2020). Construction iq (preview), [Online]. Available: https://knowledge.autodesk.com/support/bim-360/learn-explore/caas/CloudHelp/cloudhelp/ENU/BIM360D-Insight/files/BIM360D-Insight-About-Construction-IQ-html-html.html. accessed: 13.04.2020.

[50] M. Alba. (Jun. 2019). How machine learning is improving construction, [Online]. Available: https://www.engineering.com/BIM/ArticleID/19317/How-Machine-Learning-is-Improving-Construction.aspx. accessed: 24.02.2020.

[51] (2017). What is machine learning? a definition, [Online]. Available: https://expertsystem.com/machine-learning-definition/. accessed: 02.02.2020.

[52] D. Fumo. (Jun. 2017). Types of machine learning algorithms you should know, [Online]. Available: https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861. accessed: 15.03.2020.

[53] M. T. Jones. (Feb. 2018). Supervised learning models, [Online]. Available: https://developer.ibm.com/articles/cc-supervised-learning-models/. accessed: 27.01.2020.

[54] M. Heller. (Jun. 2019). Supervised machine learning explained, [Online]. Available: https://www.infoworld.com/article/3403403/supervised-learning-explained.html. accessed: 03.03.2020.

[55] J. Brownlee. (Mar. 2016). Supervised and unsupervised machine learning algorithms, [Online]. Available: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/. accessed: 30.01.2020.

[56] (2016). Overfitting in machine learning: What it is and how to prevent it, [Online]. Available: https://elitedatascience.com/overfitting-in-machine-learning. accessed: 14.02.2020.

[57] A. Al-Masri. (2019). What are overfitting and underfitting in machine learning?, [Online]. Available: https://towardsdatascience.com/what-are-overfitting-and-underfitting-in-machine-learning-a96b30864690. accessed: 01.03.2020.

[58] R. Gandhi. (2018). Introduction to machine learning algorithms: Linear regression, [Online]. Available: https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a. accessed: 27.02.2020.

[59] P. Gupta. (2017). Decision trees in machine learning, [Online]. Available: https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052. accessed: 05.03.2020.

[60] S. Mishra. (May 2017). Unsupervised learning and data clustering, [Online]. Available: https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a. accessed: 27.02.2020.

[61] V. Roman. (Mar. 2019). Unsupervised machine learning: Clustering analysis, [Online]. Available: https://towardsdatascience.com/unsupervised-machine-learning-clustering-analysis-d40f2b34ae7e. accessed: 10.03.2020.

[62] S. Li. (Jul. 2019). Anomaly detection for dummies, [Online]. Available: %7Bhttps://towardsdatascience.com/anomaly-detection-for-dummies-15f148e559c1%7D. accessed: 17.02.2020.

[63] K. Kourou, T. Exarchos, K. Exarchos, M. Karamouzis, and D. Fotiadis, "Machine learning applications in cancer prognosis and prediction", *Computational and Structural Biotechnology Journal*, vol. 13, Nov. 2014. DOI: 10.1016/j.csbj.2014.11.005.

[64] S. Ullman, T. Poggio, D. Harari, D. Zysman, and D. Seibert. (2014). Unsupervised learning clustering, [Online]. Available: http://www.mit.edu/~9.54/fall14/slides/Class13.pdf. accessed: 18.03.2020.

[65] M. Heller. (Aug. 2019). Semi-supervised learning explained, [Online]. Available: https://www.infoworld.com/article/3434618/semi-supervised-learning-explained.html. accessed: 07.02.2020.

[66] Anastasia. (Jan. 2020). Semi-supervised machine learning algorithms, [Online]. Available: https://hackernoon.com/semi-supervised-machine-learning-algorithms-fnm32cw. accessed: 04.03.2020.

[67] M. Heller. (Jun. 2019). Reinforcement learning explained, [Online]. Available: https://www.infoworld.com/article/3400876/reinforcement-learning-explained.html. accessed: 02.03.2020.

[68] M. Esposito, T. Tse, and K. Bheemaiah. (May 2017). You've heard about it, but do you understand? everything you need to know about machine learning, [Online]. Available: https://www.weforum.org/agenda/2017/05/what-is-machine-learning. accessed: 013.04.2020.

[69] A. Chakure. (Aug. 2019). Introduction to machine learning and different types of machine learning algorithms, [Online]. Available: https://hackernoon.com/introduction-to-machine-learning-and-different-types-of-machine-learning-algorithms-3z57i242u. accessed: 14.03.2020.

[70] Garychl. (). Applications of reinforcement learning in real world, [Online]. Available: https://towardsdatascience.com/applications-of-reinforcement-learning-in-real-world-1a94955bcd12. accessed: 13.03.2020.

[71] B. Baker, I. Kanitscheider, T. Markov, Y. Wu, G. Powell, B. McGrew, and I. Mordatch. (Sep. 2019). Emergent tool use from multi-agent interaction, [Online]. Available: https://openai.com/blog/emergent-tool-use/. accessed: 19.04.2020.

[72] A. Violante. (Mar. 2019). Simple reinforcement learning: Q-learning, [Online]. Available: https://towardsdatascience.com/simple-reinforcement-learning-q-learning-fcddc4b6fe56. accessed: 25.02.2020.

[73] K. Keshari. (Mar. 2020). Top 10 applications of machine learning : Machine learning applications in daily life, [Online]. Available: https://www.edureka.co/blog/machine-learning-applications/. accessed: 28.03.2020.

[74] Q. Bi, K. E Goodman, J. Kaminsky, and J. Lessler, "What is machine learning? a primer for the epidemiologist", *American Journal of Epidemiology*, vol. 188, pp. 2222–223, Oct. 2019. DOI: 10.1093/aje/kwz189.

[75] K. Didur. (Jul. 2018). Machine learning in finance: Why, what how, [Online]. Available: https://towardsdatascience.com/machine-learning-in-finance-why-what-how-d524a2357b56. accessed: 06.04.2020.

[76] D. Faggella. (Mar. 2020). 7 applications of machine learning in pharma and medicine, [Online]. Available: https://emerj.com/ai-sector-overviews/machine-learning-in-pharma-medicine/. accessed: 08.04.2020.

[77] ——, (Nov. 2017). Machine learning in finance – present and future applications, [Online]. Available: https://www.visualcapitalist.com/visualizing-trillion-fold-increase-computing-power/. accessed: 16.03.2020.

[78] R. Wille. (Mar. 2012). Duels, bombings and apple: The incredible story behind the creation of jpmorgan chase, [Online]. Available: https://www.businessinsider.com/the-history-of-jpmorgan-chase-2012-3?r=US&IR=T. accessed: 18.04.2020.

[79] M. Kolanovic. (2017). Informing investment decisions using machine learning and artificial intelligence, [Online]. Available: https://www.jpmorgan.com/global/cib/research/investment-decisions-using-machine-learning-ai. accessed: 25.02.2020.

[80] C. Jung, H. Mueller, S. Pedemonte, S. Plances, and O. Thew. (Oct. 2019). Machine learning in uk financial services, [Online]. Available: %7Bhttps://www.bankofengland.co.uk/-/media/boe/files/report/2019/machine-learning-in-uk-financial-services.pdf?la=en&hash=F8CA6EE7A5A9E0CB182F5D568E033F0EB2D21246%7D. accessed: 021.04.2020.

[81] (May 2019). Nordea markets börjar med ai för snabbare analyser, [Online]. Available: https://www.nordea.com/sv/press-och-nyheter/nyheter-och-pressmeddelanden/news-sv/2019/nordea-markets-borjar-med-ai.html. accessed: 17.03.2020.

[82] (Nov. 2016). Avancerad maskininlärning knyter ihop alla kanaler, [Online]. Available: https://sebgroup.com/sv/press/nyheter/avancerad_maskininlarning_knyter_ihop_alla_kanaler. accessed: 23.03.2020.

[83] K. Lindström. (Oct. 2016). Storbanken seb utforskar ai – anställer "digital medarbetare", [Online]. Available: https://computersweden.idg.se/2.2683/1.666950/seb-ai-digital-medarbetare. accessed: 09.03.2020.

[84] (Dec. 2019). Swedbank använder ai för att förbättra kundupplevelsen, [Online]. Available: https://www.di.se/brandstudio/swedbank/swedbank-anvander-ai-for-att-forbattra-kundupplevelsen/?loggedin=true. accessed: 16.04.2020.

[85] L. Shen, L. R Margolies, J. H. Rothstein, E. Fluder, R. NcBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography", *A natureresearch journal*, 2019. DOI: 10.1038/s41598-019-48995-4.

[86] T. Guardian. (Feb. 2020). Powerful antibiotic discovered using machine learning for first time, [Online]. Available: https://www.theguardian.com/society/2020/feb/20/antibiotic-that-

kills-drug-resistant-bacteria-discovered-through-ai. accessed: 05.03.2020.

[87] M. Reiner. (Jan. 2019). Data essentials for effective machine learning, [Online]. Available: https://medium.com/cityai/data-essentials-for-effective-machine-learning-5cbd375c3ac2. accessed: 19.02.2020.

[88] M. Venugopal. (Nov. 2018). How ai and machine learning are helping construction reduce risk and improve margins, [Online]. Available: https://constructionexec.com/article/how-ai-and-machine-learning-are-helping-construction-reduce-risk-and-improve-margins. accessed: 26.02.2020.

[89] J. L Blanco, S. Fuchs, M. Parsons, and M. J Ribeirinho. (Apr. 2018). Artificial intelligence: Construction technology's next frontier, [Online]. Available: %7Bhttps://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/artificial-intelligence-construction-technologys-next-frontier%7D. accessed: 24.03.2020.

[90] D. Burrus. (May 2019). Smart construction: How ai and machine learning will change the construction industry, [Online]. Available: https://medium.com/@DanielBurrus/smart-construction-how-ai-and-machine-learning-will-change-the-construction-industry-406b83a30140. accessed: 25.03.2020.

[91] A. Rajagopal. (Dec. 2017). The rise of ai and machine learning in construction, [Online]. Available: https://medium.com/autodesk-university/the-rise-of-ai-and-machine-learning-in-construction-219f95342f5c. accessed: 12.02.2020.

[92] S. Rao. (Jan. 2019). The benefits of ai in construction, [Online]. Available: https://constructible.trimble.com/construction-industry/the-benefits-of-ai-in-construction. accessed: 06.02.2020.

[93] K. Myers. (May 2020). How artificial intelligence is improving the efficiency of bim, [Online]. Available: https://www.pbctoday.co.uk/news/bim-news/ai-bim-systems/76032/. accessed: 28.05.2020.

[94] (Mar. 2019). Sammanställning arbetsolyckor med dödlig utgång 2018, [Online]. Available: https://www.av.se/globalassets/filer/statistik/dodsolyckor/arbetsolyckor-med-dodlig-utgang-2018.pdf. accessed: 21.04.2020.

[95] (Dec. 2011). Samhällets kostnader för olyckor, [Online]. Available: https://rib.msb.se/filer/pdf/26114.pdf. accessed: 23.04.2020.

[96] M. Man. (Apr. 2019). 3 ways ai and machine learning are keeping construction workers safer, [Online]. Available: https://www.roboticsbus%20inessreview.com/construction/3-ways-ai-and-machine-learning-are-keeping-construction-workers-safer/. accessed: 28.05.2020.

[97] S. Chatterjee. (Sep. 2017). Re-imagining future of construction with ai, [Online]. Available: %7Bhttps://towardsdatascience.com/re-imagining-construction-with-ai-54d8c8077939%7D. accessed: 24.03.2020.

[98] J. Björklund, G. Hägglund, A. Lööf, V. Johnson, K. Eklund, H. Jonsson, E. Nordmark, M. Saatchi, J. Rockström, L. Trägårdh, P. Mellergård, and S. Westerberg, "Future challenges for sweden", p. 325, Mar. 2013. [Online]. Available: %7Bhttps://www.regeringen.se/contentassets3%2089793d478de411fbc83d8f51cb5013/future-challenges-for-sweden--final-report-of-%20the-commission-on-the-future-of-sweden%7D, accessed: 10.04.2020.

[99] L. Brookes. (Apr. 2018). Why construction companies should embrace technology, [Online]. Available: https://www.constructiondive.com/news/why-construction-companies-should-embrace-technology/520070/. accessed: 29.03.2020.

[100] Byggnorden, "Mdh-research shall prevent work accidents on the construction", Jan. 2020. [Online]. Available: http://www.byggnorden.se/mdh-forskning-ska-fa-rebygga-arbetsplatsolyckor-pa-byggen, accessed: 22.02.2020.

[101] Admin. (Feb. 2019). Construction industry – the future of artificial intelligence, [Online]. Available: https://decondia.com/construction-industry-ai-future/. accessed: 17.02.2020.

[102] M. Galaev. (Apr. 2018). Why are construction profit margins so low?, [Online]. Available: https://www.kreo.net/blog/why-are-construction-profit-margins-so-low. accessed: 19.03.2020.

[103] M. Stewart. (Jul. 2019). The limitations of machine learning, [Online]. Available: https://towardsdatascience.com/the-limitations-of-machine-learning-a00e0c3040c6. accessed: 14.04.2020.

[104] J. Åfreds. (Nov. 2019). Här är sveriges 30 största byggare 2019, [Online]. Available: https://byggindustrin.se/artikel/nyhet/har-ar-sveriges-30-storsta-byggare-2019-28909. accessed: 14.04.2020.

[105] K. Jones. (Dec. 2018). How technology is reshaping the construction industry, [Online]. Available: https://www.constructconnect.com/blog/technology-reshaping-construction-industry. accessed: 06.04.2020.

[106] B. Arena. (May 2017). Låg digital nivå i byggsektorn – men nu tar utvecklingen fart, [Online]. Available: https://www.nyteknik.se/sponsrad/lag-digital-niva-i-byggsektorn-men-nu-tar-utvecklingen-fart-6844058. accessed: 29.03.2020.

[107] B. Dave, A. Buda, A. Nurminen, and K. Främling, "A framework for integrating bim and iot through open standards", *Department of Computer Science, Aalto University, Espoo, Finland*, vol. 95, pp. 35–45, Nov. 2018. DOI: 10.1016/j.csbj.2014.11.005.

[108] G. Ellis. (Jan. 2020). How machine learning is making construction more human, [Online]. Available: https://blog.plangrid.com/2020/01/machine-learning-in-construction/. accessed: 22.03.2020.

[109] M. Garbade. (Jul. 2018). 5 forecasts about the future of machine learning, [Online]. Available: https://readwrite.com/2018/07/19/5-forecasts-about-the-future-of-machine-learning/. accessed: 13.03.2020.

[110] B. Neo. (Dec. 2019). The future of machine learning, [Online]. Available: https://towardsdatascience.com/the-future-of-machine-learning-ce0a9dc18cb8. accessed: 08.03.2020.

[111] J. Vigo. (Jun. 2019). Why machine learning is the future of business culture, [Online]. Available: https://www.forbes.com/sites/julianvigo/2019/06/06/why-machine-learning-is-the-future-of-business-culture/#fe14b6a2bab7. accessed: 07.03.2020.

[112] D. Wadhwa. (Mar. 2019). How to improve your productivity with ai and machine learning?, [Online]. Available: https://bigdata-madesimple.com/how-to-improve-your-productivity-with-ai-and-machine-learning/. accessed: 01.03.2020.

[113] H. Cherkaoui. (Dec. 2016). A simple explanation of 'the internet of things', [Online]. Available: https://www.letsbuild.com/blog/what-is-bim-what-are-its-benefits-to-the-construction-industry. accessed: 29.03.2020.

7 Appendix

7.1 Questions regarding the acquisition and use of machine learning

1. Can you tell a little about your background and how you came to work for this company/organization?
2. Can you tell how long you have been working for the company/organization and what you position is?
3. How long have the company been working with utilizing machine learning?
4. How much time and money (roughly speaking) were require before the algorithms could be used (put into service)?
5. What type of infrastructure for using machine learning were required and how was it required?
6. Were the personal working with machine learning an internal or external service?

If internal

7. How was the computer engineers for this venture acquired?

If external

8. How did the coordination between the company and the external party work regarding the collecting of data for the training?

Data and models

9. How hard was it to collect data (both finding the right data and in enough amount)?
10. What type of model was selected and why?
11. How did you evaluate the performance of the model?

12. How did the model perform when deployed?

Implement of machine learning

13. Opportunities for machine learning
    - Have machine learning result in reducing cost for the company/organization?
    - Has machine learning resulted in a more efficient process?
14. Threats against machine learning
    - Was it difficult to acquire the technology for machine learning?
    - Has the technique matured enough for machine learning to be utilized in a meaningful way when the company/organization decide to work it?
    - Does the algorithm(s) often needs to be updated in order to be optimized?
    - What was the main obstacles that hampered the progress of utilizing machine learning?
15. Was an estimation done before on the profitability of using machine learning?
16. Has it been economical sustainable?
17. Is it likely that machine learning will play a bigger role in the organization in the future, given its performance?

7.2 Questions for construction companies regarding machine learning and BIM

1. Can you tell a little about your background and how you came to work for this company/organization?
2. Can you tell how long you have been working for the company/organization and what you position is?
3. What is the company's view around digitization?
4. Is the company working to become more digital?

5. Do you know about machine learning?

If BIM is used

6. How long has BIM been used and how has it work so far?
7. Has the use of BIM optimized the business core more efficient?
8. Is BIM used to collect data from previously projects to improve the process and routines?
9. Is the company collaboration with other to build up a digital library of projects and elements?

If BIM is not used

10. Does the company consider to use BIM?
11. Is the company working to digitized its business?

If machine learning is being used/tested

12. Has it been hard to collect data to use for training?
13. Has machine learning resulted in a more efficient process in the design phase?
14. Was an estimation done before on the profitability of using machine learning?
15. Has it been economical sustainable?
16. Is it likely that machine learning will play a bigger role in the organization in the future, given its performance?
17. Threats against machine learning
18. Was it difficult to acquire the technology for machine learning?
19. Had the technique matured enough for machine learning to be utilized in a meaningful way when the company decided to use it?
20. Does the algorithm often needs to be updated in order to be optimized?

21. What was the main obstacles that hampered the progress of utilizing machine learning?

If machine learning is not being used/tested

22. Does the company have ambitions to utilize machine learning in the near future?
23. Is machine learning too costly to utilized?
24. Is it difficult to acquire the technique and data for machine learning?
25. What is the main obstacles that is preventing the company from utilizing machine learning?

The near future

26. Do you believe that the construction industry will become more digitized in the near future?
27. Do you believe that machine learning will play a bigger roll in the design phase in the construction industry in the near future in order to be competitive?