Identifying Pitfalls in Machine Learning Implementation Projects

A Case Study of Four Technology-Intensive Organizations

ELIN CLEMMEDSSON
Identifying Pitfalls in Machine Learning Implementation Projects - A Case Study of Four Technology-Intensive Organizations

by

Elin Clemmedsson
Identiferings av fallgropar i implementtionsprojekt inom maskininlärning - en fallstudie av fyra teknologi-intensiva organisationer

av

Elin Clemmedsson

Examensarbete TRITA-ITM-EX 2018:224
KTH Industriell teknik och management
Industriell ekonomi och organisation
SE-100 44 STOCKHOLM
Abstract

This thesis undertook the investigation of finding often occurring mistakes and problems that organizations face when conducting machine learning implementation projects. Machine learning is a technology with the strength of providing insights from large amounts of data. This business value generating technology has been defined to be in a stage of inflated expectations which potentially will cause organizations problems when doing implementation projects without previous knowledge. By a literature review and hypothesis formation followed by interviews with a sample group of companies, three conclusions are drawn from the results. First, indications show there is a correlation between an overestimation of the opportunities of machine learning and how much experience an organization has within the area. Second, it is concluded that data related pitfalls, such as not having enough data, low quality of the data, or biased data, are the most severe. Last, it is shown that realizing the value of long-term solutions regarding machine learning projects is difficult, although the ability increases with experience.

Keywords

artificial intelligence, machine learning, pitfalls, implementation, project management
Sammanfattning


Nyckelord

artificiell intelligens, AI, maskininlärning, fallgropar, implementation, projektledning
Acknowledgements

I would like to thank my supervisor, Henrik Blomgren, for great guidance through all stages of the process as well as interesting conversations. Also, I want to thank Findwise and foremost my supervisor at Findwise, Monica Basu, for the advice and new perspectives I have been given. Last but not least are the interviewees who have agreed to take part in this study. Thank you for giving me your time and insights.

Elin Clemmedsson
Stockholm, 2018-05-28
## Contents

1 **Introduction**  
   1.1 Background ................................................. 9  
   1.2 Purpose ................................................. 10  
   1.3 Research Questions ........................................ 11  
   1.4 Scope, Limitations, and Sustainability ...................... 11  
   1.5 Expected Contribution .................................... 11  

2 **Literature Review**  
   2.1 Machine Learning ......................................... 12  
      2.1.1 Stages of Machine Learning Implementations .......... 12  
   2.2 Artificial Intelligence From a Management Perspective - an Analysis of the Future ......................... 14  
   2.3 Software Implementation Project Pitfalls .................. 15  

3 **Method**  
   3.1 Research Design and Approach .............................. 19  
   3.2 Literature Review ........................................... 19  
   3.3 Research Reasoning ......................................... 20  
   3.4 Ethics ..................................................... 20  
   3.5 Data Collection .............................................. 21  
      3.5.1 Primary - Interviews .................................. 21  
      3.5.2 Secondary - Literature Review ....................... 22  

4 **Case Study Companies**  
   4.1 Cinnober Financial Technology .............................. 23  
   4.2 iZettle .................................................... 24
5 Results

5.1 Machine Learning Pitfall Hypotheses

5.2 Similarities Through Identified Software Development Project Pitfalls

5.3 Interview Insights

5.3.1 Main Pitfalls

5.3.2 Stages of Machine Learning Projects

5.3.3 Machine Learning Hypotheses

5.3.4 Project Management Hypotheses

5.3.5 Differences in how Machine Learning Projects and Other Software Development Projects are Treated

5.3.6 Additional Comments From the Interviewees

6 Discussion

6.1 Findings

6.2 Validity, Reliability, and Generalizability

6.3 Issues and Limitations

7 Conclusion

8 Author’s Thoughts

9 Future Research

A Interview Questions
Chapter 1

Introduction

This chapter serves as an introduction to the area of research. It presents the purpose of the study along with research questions to be answered. The chapter is concluded with a description of the scope and limitations to the research.

1.1 Background

Artificial intelligence, short AI, is described by Jordan and Mitchell (2015) to address the question of how to build computers that can learn and improve automatically through experience. AI has as well been defined as a science with the goal of making machines do things that would require intelligence if done by humans (Negnevitsky, 2005). Machine learning, a technology used for achieving AI, is explained by Marsland (2015) as “adaptive mechanisms that enable computers to learn from experience, learn by adaption, and learn by analogy”. With machine learning progressing rapidly over the past two decades, it is now used for computer vision, speech recognition, natural language processing, robot control and more (Jordan & Mitchell, 2015). The last decade the term big data has emerged. It references the ability to manage large amounts of data used for machine learning. The emerge of the term is due to the technology’s ability to iterate through and gain useful insights about the data faster than humans can. With the use of Graphics Processing Unit (GPU) power for this, these computations are made even faster which is a reason for the rapid development of machine learning over the past years. (Jordan & Mitchell, 2015).

The concept of artificial intelligence was first suggested by Alan Turing (1950). Turing described the concept of having machines act as humans. In the recent years, the increased CPU and GPU speed along with increased availability of data has made the concept useful for many companies in building business value. The AI trend has quickly grown from being a vast concept to an emerging technology that many companies are looking to integrate into their businesses (Gartner, 2017).

Every year Gartner provides a visual representation of “the maturity and adoption of technologies and applications”, shown in fig 1.1 (Gartner, 2017). Machine learning is placed in the category “Peak of inflated expectations”. With the recent rise of AI, the subject is heavily discussed but might not be living up to the expectations of some
organizations. With machine learning on the peak of the hype cycle, organizations might not fully understand the technology and its drawbacks before starting the implementation projects.

Figure 1.1: Gartner Hype Cycle for Emerging Technologies, 2017

Before entering the area of artificial intelligence and machine learning, organizations need to know which resources are required and which resources are beneficial for successful implementation and usage. Gartner (2017) pointing to inflated expectations within organizations implicates the risk of companies neither possessing this knowledge nor realizing the lack of it. Also, an understanding of the type of problems which machine learning can be useful for solving is of importance. Inflated expectations may cause a belief in machine learning solving unrealistic problems. With the rise of machine learning technologies because of aforementioned factors, machine learning implementation projects are and will be realized within companies and organizations without previous experience in the specific area. Without this competence within the organizations, they will make mistakes and find themselves in pitfalls that could have been avoided. Identifying these problems can make future implementations more likely to be successful.

Although machine learning falls under the category of software implementation, the risks, pitfalls, and problems might not be the same or even a subset of other software implementation projects. Therefore, there is a need for investigating how machine learning project pitfalls might differ from software implementation pitfalls in general.

1.2 Purpose

The purpose of this report is to identify pitfalls existing, or pitfalls that have been identified and avoided, in machine learning implementation projects in four technology-intensive companies with different levels of machine learning experience. By identifying pitfalls, future implementation projects can run more efficiently and avoid non-company specific mistakes. By identifying pitfalls in different stages of the implementation projects, future project managers can know which stages that require more focus than others, which necessarily not are the same as other software
implementation projects and therefore not correspond to previous experiences.

1.3 Research Questions

Based on the problem background and purpose above, the following research questions have been defined:

- Which pitfalls exist in machine learning implementation projects within four technology-intensive organizations?
- How do pitfalls experienced vary with machine learning experience?

1.4 Scope, Limitations, and Sustainability

Four companies are included in the study. The companies are all technology intensive and based in Stockholm. To further delimit the work, no quantitative studies will be conducted and data is collected through qualitative studies. This includes interviews, articles, and literature. The interviewees are employees within machine learning implementation projects within the companies.

The research is limited to the functional level regarding project planning, optimization, and risk mitigation. In regard to organizational sustainability, this study aims at making projects waste fewer resources by identifying common mistakes and will through that increase sustainability. This means organizations’ initiatives within machine learning can continue and the area of project management within machine learning can explore new pitfalls instead of fall into old ones.

A pitfall is in this thesis defined as something affecting a project negatively but which could have been avoided. A pitfall can be a mistake due to lack of attention, a hindrance encountered due to lack of knowledge or research before starting, or a difficulty that in hindsight could have been avoided. Pitfalls are considered unnecessary and will interfere with effective use of resources. Something hindering the project that is outside of a company’s power is not considered a pitfall.

1.5 Expected Contribution

For academic and theoretical purposes, this project is expected to contribute to the research area of project management regarding machine learning projects. By conducting interviews, an early indication of problem areas is expected to be found. This can provide inspiration for further research within this area, with pointers on which direction to go retrieved from the conclusions of this project.

The results from this project are also expected to be useful for companies, and specifically project managers or C-level executives, in the early stages of implementing machine learning to optimize or automate their businesses.
Chapter 2

Literature Review

This section will guide the reader through previously done research within the area as well as provide deeper knowledge of the subject. First, artificial intelligence and machine learning as concepts are explained. Following that is a more technology intensive section about machine learning and later a perspective from a managerial point of view. The chapter is concluded with a summary of software development pitfalls developed by four different authors.

2.1 Machine Learning

Machine learning is about making computers modify or adapt actions to make them more accurate. The accuracy is measured by how well the computer’s decisions reflect the correct decisions (Marsland, 2015). Jordan and Mitchell (2015) say machine learning addresses the question of how to build computers that improve automatically through experience. Many developers of machine learning solutions recognize that some problems are more easily solved by showing a machine examples of desired input-output behavior than programming it manually to respond to every input correctly (Jordan & Mitchell, 2015).

An example of a basic machine learning problem used for education by the data science platform Kaggle (2018) is predicting who would survive the Titanic catastrophe in 1912, and who would not. Given a set of data about a subset of the passengers, including their age, sex, if they were a survivor or not, ticket class, port of embarkation, number of family members on the boat, etc., a machine can be shown this data and learn which of these traits, often called features (Marsland, 2015), that increases the chances of survival. When the machine is then presented with another subset of data about other passengers, where the “survivor or not” indication is removed, it can to a certain confidence level predict who survived and who did not. (Kaggle, 2018)

2.1.1 Stages of Machine Learning Implementations

Data Collection and Preparation is the process of collecting the data needed as well as making sure the data sets are clean, meaning they do not have significant errors or missing data. This data is to be used as training and test data. If the problem is completely new, the data needs to be collected. If the data to be used already exists, it needs to be cleaned from errors and missing values. The data could also need to be merged from different sources hosting the data on different forms. Part of this process is thereby ensuring the data is represented in the same place in the same format which could be a challenging task. Finally, the quantity of the data needs to be considered. Machine learning is a computationally expensive process and with more data comes increased computational costs. Simultaneously, not having enough data will make the model less correct in its predictions. There is a point in which there is enough data to train the model until satisfaction while using an amount of data that is not adding too much computational overhead to the process. (Marsland, 2015)

The second stage, potentially to be merged with the first stage if the problem is completely new and data needs to be collected, is Feature Selection. Features are the variables used for the algorithm, eg. age, sex, or location. Feature selection is the process of selecting which features that are (the most) useful for solving the problem and which are not useful. (Marsland, 2015)

Algorithm Choice is, as the name indicates, choosing an algorithm that is appropriate for the problem. There are different algorithms that might suit different types of problems. For many algorithms, there are parameters that have to be set manually, which happens in the stage Parameter and Model Selection. (Marsland, 2015)

When the solution is implemented, the next stage is Training. This is “the use of computational resources in order to build a model of the data in order to predict the output of new data” (Marsland, 2015). Marsland also discusses a problem known as overfitting. Overfitting happens when the machine is over-trained on the training data and is not identifying a general generating function, but rather adapts the model to perfectly fit the training data. This will reduce the generalization capacities of the model. A visual representation of overfitting is shown in figure 2.1. Overfitting means the model will not perform as good on previously unseen data as it does on the test data and might make the model appear better than it is. (Marsland, 2015)

The last step is Evaluation; testing the system on data it has not been trained on. This can include comparison with experts in the field’s results. (Marsland, 2015)

Another two steps can be added to the process. There is a stage before the start of the implementation and a stage after finishing the implementation. This is proposed by Duncan (1993) who suggests five stages of project management. The stages are initiating, planning, executing, controlling, and closing (Duncan, 1993). Marsland’s stages of machine learning implementation projects do not cover the initiating and closing phase. These phases can be seen as initiation and releasing the solution to production.
2.2 Artificial Intelligence From a Management Perspective - an Analysis of the Future

Three studies have been made by Accenture, MIT in collaboration with Boston Consulting Group, and SAS (Ransbotham et al., 2017; Kolbjørnsrud et al., 2016; SAS, 2017) regarding AI from a management perspective. The studies all focus on the status of artificial intelligence, future development and executives thoughts on this area of technology developing within their industries.

It is clear that the expectation of AI solutions is higher than the actions taken to build or implement them. 85% of a global survey of over 3000 executives believe AI solutions will benefit their company in getting a competitive advantage while only 23% have incorporated AI in any part of their business (Ransbotham et al., 2017). SAS (2017) show that around two-thirds of the participants in the study claim to believe that “AI would have a wide-ranging effect in the next five to ten years”. The study also shows that while optimism regarding AI was high among organizations, fewer thought their organization was ready for reaping the benefits they were positive about, to which SAS suggest possible problems with execution. The MIT study shows there are large gaps between today’s leaders who have adopted AI, who thereby have competence and understanding of the technology, and laggards who have not. One of the biggest gaps is the companies’ approach to data. The research showed misconceptions about resources needed to train AI. It is shown that companies sometimes erroneously believe they have access to the data needed for training.

Challenges raised by the participants in the surveys include handling of changes in jobs because of AI, both creation and loss, developing trust in AI both internally and externally, and few own-industry use cases to learn from, and lack of competence (Ransbotham et al., 2017; Kolbjørnsrud et al., 2016; SAS, 2017). A lack of trust in the output given by AI as a cultural challenge was the biggest challenge for the respondents, with 49% of the respondents’ answers (SAS, 2017). Kolbjørnsrud et al. conclude there are both readiness and resistance within companies, with resistance being more common among lower-level managers and readiness more common in
the C-suite. The findings also include the conclusion that executives need to start experimenting with AI and that the “wait and see” method is not affordable. Both SAS and MIT also bring up the issue of lack of ROI (Return on Investment) or unclear business case.

Ransbotham et al. (2017) show that the cultural resistance to AI approaches is among the top barriers to AI adoption, but on the lower half after survey respondents ranked them; around 30% believes the cultural challenge is among the top three challenges. Though, 49% of the survey respondents in the SAS study responds the cultural challenge is the biggest challenge.

2.3 Software Implementation Project Pitfalls

Below follows five authors’ research regarding problems encountered in software development projects.

Stewart

Stewart (2001) presents 25 most common mistakes with real-time software. One mistake is “Tools choice is driven by marketing hype, not by evaluation of technical needs”. This is relevant to machine learning’s stage on Gartner’s hype cycle. Stewart means there are three categories in which people are misled when choosing tools. A good looking graphical user interface does not promise a better tool than one with a not as appealing graphical user interface and argues technical capabilities need to be considered first hand. Second, the number of users of a tool has, could bias one to choose the specific tool. Once again, technical capabilities should be in focus. Promises of compatibility is the last category. Stewart argues managers being influenced by promises of compatibility even in cases where the compatibility of the tool with other software is without relevance to the project. (Stewart, 2001)

Stewart also lists documentation being written after implementation and incomplete testing as two other reasons. Remaining items on Stewart’s list are not brought up in this thesis because of them being specific to embedded real-time software and are not considered relevant for this thesis.

Boehm

Boehm, in 1991, developed a list consisting of ”Top 10 Software Risk Items” published in his article Software Risk Management: Principles and Practices in IEEE Software. The list contains the primary sources of risk items in software projects according to Boehm. The list could be used as a checklist for managers and system engineers to help with identification and risk management of the most serious risk items. Most of the items on the list concern domain knowledge or scoping, two categories often overlooked in computer science and literature and education (Boehm, 1991). The list follows below.
Keil et al.

An explanation for the high failure rate in software projects is because of the risk management and assessment skills among managers. Managers and not taking sufficient measures for assessing and managing risks existing in software projects. (Keil et al, 1998)

Keil et al. mean that risk management strategies build on first identifying risks, then assessing the relative importance of the risks and focus on the areas that contain the biggest threats to success. Last, the identified risks must be “classified in a way that suggests meaningful risk mitigation strategies” (Keil et al., 1998; Keil et al., 2001). Keil et al. identify Boehm’s work as work that has had great influence on the community. Though, Boehm’s work is in need of an update because of the considerable change in both the organizational and technical landscape. Keil et al. (1998) also discuss shortcomings in Boehm’s work by arguing it was built upon Boehm’s experiences within the defense industry in 1980 and might not be representative and generalizable in more typical business environments. Keil et al. also discuss how some risk items discovered are missing from the list developed by Boehm. A suggestion for the explanation is because Boehm’s work focuses on risk factors in the execution stage where managers have more control.

Keil et al. present a new list of 11 risk factors developed by conducting a Delphi study with managers in Finland, the US, and Hong Kong. The list of eleven risk factors are the ones ranked collectively highest among the managers. (Keil et al., 2001)
It is concluded from the survey that project managers tend to assess risks that are outside of their direct control as bigger than those they can control. With this finding, Keil et al. continue to develop a risk categorization framework presented below. (Keil et al., 1998)

![Risk categorization framework by Keil et al.](image)

The first quadrant (numbering only used for referencing), Customer Mandate, represent risks that can not be controlled by the project manager but can be influenced. Many of the study participants’ top risk fall into this quadrant. The name refers to the need for mandate among the end users, the customers, but the quadrant also cover the need for having commitment from senior managers. Risks placed in this quadrant are for example lack of top management commitment, failing to gain user commitment, or no/inadequate user involvement.

In the second quadrant fall risks that concern uncertainties that arise when scoping and developing requirements for the projects. Examples of risk factors placing in this quadrant are misunderstandings of requirements and managing change improperly. The risks in the Scope and Requirements quadrant are risks to a large extent controllable by the project manager but which requires skillful interfacing with the end users.

Execution, the third quadrant, concerns the execution of the project. Risks tied to execution are risks highly controllable by the project manager and perceived as less important than other risks. Many items on the list developed by Boehm are placed in this quadrant. This includes risks associated with poor project management skills, for example staffing issues, development process methodology problems, and improper definition of roles.

Quadrant four contains risks tied to the project environment. This includes risks both internally and outside of the organization, such as changes within the organization that could threaten the project and changes in the competitive landscape.
Risks in this quadrant have a low likelihood of occurrence and are therefore considered moderate on the scale of perceived relative importance. Though, they are all risks of which the project manager has little or no control. For example, natural disasters place in this quadrant together with changes in top management causing hindrances. (Keil et al., 1998)

Oz and Sosik

Oz and Sosik develop in Why Information Systems Projects are Abandoned: A Leadership and Communication Theory and Exploratory Study (2000) five reasons information systems projects are abandoned. The list is developed through information gathered by Oz and Sosik from a sample of chief information officers and their immediate subordinates. The data collection was made through a questionnaire with by the authors formed hypotheses which the respondents classified from 1 (Not important) to 5 (Very important). All respondents used for data collection had experience of at least one failed project. The list is provided below.

| 1. Poorly communicated goals/deliverables |
| 2. Lack of corporate leadership           |
| 3. Inadequate skills and means            |
| 4. Poor project management                |
| 5. Deviation from timetable/budget        |

Fairley and Willshire

Through comparing software development projects to the Vasa’s story of failure, Fairley and Willshire (2003) deduce nine reasons why projects fail. These reasons are deduced from an analysis of problems surfaced during the Vasa project which is also applicable to project management. The nine reasons are presented below.

| 1. Excessive schedule pressure         |
| 2. Changing needs                      |
| 3. Lack of technical specifications    |
| 4. Lack of a documented project plan   |
| 5. Excessive and secondary innovations |
| 6. Requirements creep                  |
| 7. Lack of scientific methods          |
| 8. Ignoring the obvious                |
| 9. Unethical behavior                  |
Chapter 3

Method

In this chapter, the methods used in this thesis are described along with processes and choices regarding the methodology. The primary methods that have been used are a literature review and semi-structured interviews.

3.1 Research Design and Approach

To answer the research questions, a qualitative method will be used. Interviews and a literature review are used for reaching conclusions. First, the literature review is done to cover previous research in the area and to form hypotheses from, and later interviews are held. This is believed to be an appropriate method for answering the research question because of the broad understanding the literature review provides along with interviews with people who have been a part of the implementation projects.

Walliman (2011) describes how qualitative data is data that cannot be measured and counted, and is generally expressed in words rather than numbers. Therefore a qualitative method is suitable for this study and has therefore been used. The qualitative method is chosen because of the nature of the problem.

3.2 Literature Review

The purpose of the literature review is to gain knowledge about the current state of research within the subject and assess it for relevance, quality, controversy, and gaps (Walliman, 2011). The literature review will summarize the body of knowledge within the area of artificial intelligence, machine learning, and software development projects as well as pitfalls, problems, and risks associated with mentioned areas. The literature to be reviewed was primarily found through libraries and academic databases, where Google Scholar was the most utilized.

Evaluation of web sources is done by considering several questions regarding each source. The areas considered are accuracy, bias, level of detail, date published, and what authority they are based on according to guidelines described by Walliman.
The literature review is also used for finding out how previous research in the area has been conducted. This is to gain experience and knowledge that will be used for choosing appropriate and relevant methods for solving the problem stated (Walliman, 2011).

The literature review itself was continuously reviewed during the writing process. This was done to maintain a relevant connection between the literature review and remaining parts of the thesis.

3.3 Research Reasoning

This thesis uses inductive reasoning as a method for reaching conclusions. Inductive reasoning starts from observations or experiences and from them develops general conclusions. The observations or experiences are in this case the interviews that are used to develop conclusions.

The inductive approach can be used for evaluating qualitative data and information (Thomas, 2006). The purpose of an inductive research approach is to collect qualitative data and information, analyze and establish links between the research objectives and the data collected, and finally to draw new conclusions and develop frameworks (Thomas, 2006). The approach is relevant to use in this project because of the relatively new and continuously developing field of artificial intelligence and machine learning. The new and continuous development leads to a limited amount of theory within the area to base the study on. Therefore, interviews with industry professionals are conducted to gather information and data which will serve as the basis of the analysis and developed framework in this paper.

There are though problems with inductive reasoning. One question is how many observations that need to be done before conclusions can be drawn reasonably. There need to be a certain number of observations before generalizations can be made. Another question is under which circumstances the observations are made to reveal true information and therefore be able to draw relevant conclusions. To ensure relevance of conclusions, interviews are conducted with people from different companies, where the companies vary in size and machine learning experience. It is encouraged to develop this work further by adding more observations under new circumstances to further increase reliability.

3.4 Ethics

By using an inductive approach to this study, it helps to continuously review the problematization, purpose and question formulations (Blomqvist & Hallin, 2015). The benefit of such approach is the ability to present a general critical attitude towards the methods and the gathered type of data. In a general sense concerning the purpose and meaning of ethics, the goal of this report is to remain impartial as much as possible. This is done by complying with praxis regarding scientific work and adhering to the norms of scientific work.
Being free from bias is something some people say is impossible, and many agree it is at the very least difficult. Walliman (2011) states there is danger tied to simplifying transcripts. When cleaning up or organizing data, own interpretations and bias can be introduced. This can lead to losing meaning in the collected data. Measures have been taken to minimize this risk through this thesis by sending written material from the interviews to the interviewees, asking them to confirm the concepts have been properly understood, written down, and translated from Swedish to English without losing or distorting information, and that no bias has emerged from the interpretation of what was said during the interview.

Another ethical aspect is the interviewee’s anonymity. All interviewees have been offered to be anonymized if they wished to. They are informed that the offer stands and that they can change their mind until this thesis is finished.

The study is and has been presented as a thesis project from KTH Royal Institute of Technology, conducted at a search consultancy company, Findwise.

### 3.5 Data Collection

This thesis uses both primary and secondary data to draw conclusions from. Primary data is data that is the first recording of a situation and can provide information about any observation. Secondary data is data that has been interpreted and recorded. (Walliman, 2011)

#### 3.5.1 Primary - Interviews

A semi-structured interview method was used to maximize the information gained from the interviewee. A set of interview questions were developed with the purpose of gaining as much information as possible about the company’s business within the time span provided.

To gain the most out of the interview, a discussion was prompted to ask further questions regarding the topic at hand. This type of structure will give the interview a less strict scope and also provide the ability to ask further questions when necessity or curiosity arises (Collis & Hussey, 2014). According to Collis and Hussey (2014), this type of interview structure is well suited when the goal of the interview is to create an understanding of the interviewees perspective, when the subject is confidential or commercially sensitive, or when the logic of the situation is unclear and the interviewees personal ideas and/or concepts are of importance. It is also stated by Blomkvist and Hallin (2015) that semi-structured interviews are better suited for qualitative studies, which is the primary information type in this study.

The method used for selecting the interviewees was sampling. Sampling is a method used when wanting to know information about a large group but when there are no resources to gathering information from every subject in the group (Walliman, 2011). Sampling means retrieving information from a subset of the group with the intention that the retrieved information is representative for the whole group (Walliman, 2011). Due to time constraints, this method is considered to be appropriate for answering the research questions.
The strategy was a combination of natural sampling and the comparative approach. Natural sampling involves selecting the most accessible subjects which can be useful when the researcher has little influence on the sample (Collis & Hussey, 2014). The comparative approach is a way of choosing a sample group when all cases may be unique (Walliman, 2011). Since no company is identical to another, this method was considered suitable. The comparative approach involves selecting very different subjects from both ends of a spectrum as well as in the middle (Walliman, 2011). This was done by reaching out to companies within the financial technology sector that varies in size as well as machine learning knowledge. By a referral from one of the interviewees, one interview was conducted with an interviewee at a company that was not a financial technology firm. This interview was included in the study because of its difference in size from the other companies. Many financial technology companies are relatively young and small in number of employees. Therefore it is believed that the comparative approach was strengthened by including the bigger company in the study.

The interviewees are:

- Jonas Nordström, Software developer at Cinnober Financial Technology
- Eric Hansander, Data scientist at iZettle
- Sara Väljamets, Data scientist at Klarna
- Jorrit Peters, Data scientist at Telia

### 3.5.2 Secondary - Literature Review

The secondary data is collected from:

- Reports from consultancy firms and other companies. Mainly from Accenture, Massachusetts Institute of Technology in collaborations with Boston Consulting Group, and the US company SAS.
- Academic articles and publications found through Google Scholar and the KTH Royal Institute of Technology library database service, Primo.
- Books on the topic.
Chapter 4

Case Study Companies

The companies in which pitfalls have been studied are presented in this chapter to provide context to the reader regarding the companies’ current status on machine learning and how it is utilized within the organization.

Below is a summary of numbers extracted from the following sections in this chapter. Figure 4.1 shows a visual representation of the companies’ machine learning competence and number of employees relative to each other.

![Figure 4.1: The companies’ relative machine learning competence and size](image)

4.1 Cinnober Financial Technology

Cinnober Financial Technology, short Cinnober, is a software company within the financial technology sector. Cinnober builds solutions for trading, risk management
and clearing\(^1\) among others. Their customers include stock, derivative, and commodity exchanges around the world, as well as clearinghouses\(^2\), banks, and brokers. Cinnober was founded in 1998 and has around 350 employees. (Bloomberg L.P, 2018)

**Machine Learning at Cinnober Financial Technology**

The interview at Cinnober was conducted with Jonas Nordström. Nordström has worked at Cinnober since 2016 and is a software developer with an interest in machine learning. According to Nordström, the machine learning competence within the company is very small. There are no employees hired to work solely with machine learning solutions. Today, no machine learning initiatives have been started. This is because of the customer-driven project strategy the company has adapted. The projects at Cinnober are customer driven in the meaning that the customer defines what solutions they want delivered and pay Cinnober to deliver them. To this date, no projects within machine learning have been requested by any customer. Machine learning projects are, with the limited competence and experience existing in the company, considered riskier and might not pay off. What is generating revenue at Cinnober is delivering the solutions to the customer and Nordström believes that unless a customer asks for machine learning solutions to a problem, there will be no substantial time or money investments in that area. Nordström describes that there is always something more urgent to focus on.

Despite the customer-driven project strategy and limited technical knowledge of machine learning, there are initiatives that have been discussed and there exist some understanding of the benefits of machine learning within the company. One initiative is risk calculations used by clearing houses. The problems encountered here are the ability to reproduce and the ability to account for the predictions which are challenges in machine learning. Due to regulations in the markets from authorities in different countries and confidential data, this is often not applicable. Another example, that does not face the previously mentioned problem, is predicting volatility in marketplaces. That is predicting the amounts of trades happening on different exchanges for example. The data needed for this is not confidential and the customers benefiting from this is banks and not clearinghouses, who can be more aggressive in their calculations and risk-taking due to not having as much financial responsibility.

### 4.2 iZettle

iZettle was founded in 2010. In 2011 their most known product, a credit card reader, was launched. The philosophy of iZettle is to help small businesses and small business owners compete in the market on the same terms as larger businesses. iZettle does this in several ways, and one is by issuing the credit card reader with a pricing plan that is fitted for smaller businesses that otherwise would not afford to allow payments with credit cards. Another way is through issuing credit to the

---

\(^1\) The settlement of accounts or exchange of financial instruments especially between banks (Merriam-Webster, 2018)

\(^2\) An establishment maintained by banks for settling mutual claims and accounts (Merriam-Webster, 2018)
businesses which larger banks are often not willing to do, or do at a high cost. Due to iZettle owning data from a customer through the transactions of the credit card reader, they can do reliable credit evaluations automatically and issue credit to a cheaper price. (Hansander, 2018)

**Machine Learning at iZettle**

At iZettle, the interview was conducted with Eric Hansander. Hansander started working at iZettle almost 3 years ago. During his time at iZettle, he has worked with machine learning for risk management which involves automatically detecting undesired behaviour in the form of fraud, risky behaviour, violation of terms and conditions, or plain illegal acts. Using machine learning models for analysis of behavior patterns helps identify what is deviating from normal behavior or looks suspicious.

Hansander leads a team of data scientists created from a merge of different teams that were more spread out in the organization prior to the merge. The team works with machine learning projects in all parts of the company.

The data science team at iZettle consists of seven people, an analytics department of around ten, and continuous education efforts to educate more people within the company in data science and its possibilities. The goal is to help more people identify machine learning applicable problems or business opportunities.

iZettle is using machine learning in many different areas across the company. Some of the main internal use cases are risk management, credit scoring, and predicting lifetime value for customers. Predicting customer lifetime value means predicting all future profits of individual customers and predicting the risk of the customer leaving, i.e. churn prediction. There is a continuous effort within the company to identify additional areas where machine learning can be used to optimize processes. This includes making use of customer data to help small businesses operate more effectively.

**4.3 Klarna**

Klarna is a Stockholm based company, founded in 2005 with the mission to make it easier for customers to shop online. The goal is to make payments as easy and safe as possible. Klarna is today one of the biggest banks in Europe and offers payment solutions to over 60 million customers in 14 countries. Klarna offers different payment solutions, including direct payments, installments, and invoices. Today, Klarna has 1700 employees. (Klarna, 2018)

**Machine Learning at Klarna**

The third interview was held with Sara Väljamets. Väljamets works as a data scientist at Klarna within the issuing domain, which is where credit decisions are made. The credit evaluations are real-time and instant.
At Klarna, the data science and machine learning competence is spread out and it is hard to evaluate how much competence there is, but the number is above 30 people and increasing.

Credit scoring is one of the areas where Klarna have applied machine learning. A lot of initiatives are currently worked on with their app, which Väljamets can not disclose much detail about since it has not been released yet. It is according to Väljamets “bigger” things, among other content related, meaning the use of machine learning to decide how to give the users a more personalized experience in the app and how to handle different kinds of customers. Machine learning is also used for fraud detection and ranking of the order that payment methods are shown to the user when using their service, trying to rank the most likely one at the top for every specific user for convenience reasons. The payments methods ranking can differ between different customers and different countries, because of the customers’ differences in buying behavior and relationship to credit products.

Many applications are already in production but are continuously updated with new initiatives.

4.4 Telia Sverige

Telia Sverige is the Swedish part of Telia Company and is a company within IT and telecommunication. Their core business is network and they strive to give their customers the best connection. With 6,700 employees Telia aim to be part of building the digital society. (Telia Sverige, 2018)

Machine Learning at Telia Sverige

The last interview was conducted at Telia Sverige, held with Jorrit Peters. Peters has a background in econometrics and applied mathematics. He has worked as a data analyst and now as a data scientist at Telia. He is working in a team of data scientists who act as consultants within Telia. Their focus varies between consumer, network, business, and other areas. They find areas within the company that can be improved by using machine learning solutions, implement the solution, and then move on to another project.

The team consists of three to four data scientists, working for Telia Sverige, the Swedish division of Telia. Besides that, there are also five data scientists working for Telia Group, the organization above Telia Sverige.

There are many different machine learning initiatives within Telia. One application is customer relations management, CRM, which consists of predicting which customers that will be interested in which products. Machine learning is also used within network to predict geographically where there next will be a need for a network expansion due to for example the summer tourism or technology development effects. Also within logistics machine learning technology has proven to be useful. The application is to predict how and when products should be moved from central storage to the stores and between stores to optimize storage handling. Having too many products in store storage is a cost, and so is having too few counted in lost
revenue. Mathematically, the problem is not hard to solve if all parameters are known, but as they are not, machine learning is used to best predict solutions.
Chapter 5

Results

In this section, results from the categorization and selection of software development pitfalls are presented followed by the results from the interviews in which the pitfalls have been discussed. The interview results are presented thematically for pedagogic reasons, giving the reader the full picture of each pitfall hypothesis.

From the literature review, it is concluded there are no published academic studies regarding pitfalls within machine learning implementation projects or similar. There are studies within other software implementation projects and software implementation project in general. Because of this, the pitfall hypotheses developed will be partly potential pitfalls identified through the literature review regarding machine learning, partly previously identified pitfalls within software development projects.

5.1 Machine Learning Pitfall Hypotheses

From Marsland’s work, nine technical pitfalls are identified by capturing problems raised and prerequisites stated:

- Overfitting
- Not testing properly
- Wrong choice of algorithm
- Not enough data
- Data on wrong form
- Wrong features selected
- Wrong parameter and model selection
- Training data biased
- Incomplete testing

From the studies conducted by SAS, MIT, and Accenture, three additional pitfalls are identified:

- Bad handling of changes in jobs
• Not developing trust for AI within the company, cultural challenge
• High deployment cost/lack of return on investment

5.2 Similarities Through Identified Software Development Project Pitfalls

To not get into too much detail or cause lengthy detail-oriented interviews where it is difficult to build an understanding of the bigger picture, the software development project pitfalls identified in the literature review are condensed and grouped to find patterns. The common themes among the lists are identified and were presented to the interviewees during the interviews.

By grouping the list provided by the authors in chapter 2, similarities are found. Three areas are brought up by three or more authors in different wording but with similar intent. One additional area is brought up by two authors while also mentioned in the consultancy reports provided by SAS and MIT. The four areas are presented below.

Lack of Competence

<table>
<thead>
<tr>
<th>Author</th>
<th>Pitfall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boehm</td>
<td>Personnel shortfalls</td>
</tr>
<tr>
<td>Boehm</td>
<td>Straining computer-science capabilities</td>
</tr>
<tr>
<td>Keil et al.</td>
<td>Lack of required knowledge/skills in the project personnel</td>
</tr>
<tr>
<td>Keil et al.</td>
<td>Insufficient/inappropriate staffing</td>
</tr>
<tr>
<td>Oz &amp; Sosik</td>
<td>Inadequate skills and means</td>
</tr>
</tbody>
</table>

The first item brought up by three of the authors is lack of competence, meaning organizations having difficulties acquiring talent with the competence needed for machine learning implementation projects. Items 7 and 10 on the list provided by Keil et al., *Lack of required knowledge/skills in the project personnel* and *Insufficient/inappropriate staffing* both suggest finding a competent workforce for the task is difficult. This is also brought up in Boehm’s list as *Personnel shortfalls*. Also Oz and Sosik have a staffing-related item on their list, *Inadequate skills and means*, aiming at shortcomings in competence.

Poorly Estimated Schedules and/or Budgets

| Author         | Pitfall                     |
|----------------|                            |
| Boehm          | Unrealistic schedules and budgets                         |
| Oz & Sosik     | Deviation from timetable/budget                           |
| Fairley & Willshire | Excessive schedule pressure             |

Boehm, Oz and Sosik, and Fairley and Willshire all bring up the issue of planning and budgeting. Schedules and budgets not being estimated correctly is causing problems
in the projects according to the authors. It is found on Boehm’s list as *Unrealistic schedules and budgets*, on Oz and Sosik’s list as *Deviation from timetable/budget* and on Fairley and Willshire’s list as *Excessive schedule pressure*. Fairley and Willshire do not, as opposed to the others, bring up the budgeting problem. Though, as budgeting and scheduling are dependent on each other they are chosen to be coupled with the other two.

### Requirements Changes

<table>
<thead>
<tr>
<th>Author</th>
<th>Pitfall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boehm</td>
<td>Continuing stream of requirement changes</td>
</tr>
<tr>
<td>Keil et al.</td>
<td>Changing scope/objections</td>
</tr>
<tr>
<td>Keil et al.</td>
<td>Lack of frozen requirements</td>
</tr>
<tr>
<td>Fairley &amp; Willshire</td>
<td>Lack of technical specifications</td>
</tr>
<tr>
<td>Fairley &amp; Willshire</td>
<td>Requirements creep</td>
</tr>
</tbody>
</table>

Issues regarding requirements are brought up twice by two of the authors as similar items on their lists. Keil et al. state that both *Changing scope or objections* and *Lack of frozen requirements* are potential pitfalls, and Fairley and Willshire claim that *Lack of technical specifications* and *Requirements creep* might cause projects to fail. The issue with requirements is also touched upon by Boehm who claims *Continuing stream of requirement changes* is a common problem.

### Lack of Management Support or Stakeholder Buy-In

<table>
<thead>
<tr>
<th>Author</th>
<th>Pitfall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keil et al.</td>
<td>Lack of top management commitment to the project</td>
</tr>
<tr>
<td>Oz &amp; Sosik</td>
<td>Lack of corporate leadership</td>
</tr>
</tbody>
</table>

Only by two of the four authors, lack of management support is brought up as a common problem. Oz and Sosik call it *Lack of corporate leadership* and Keil et al. mention it as *Lack of top management commitment to the project*. Although, through reports by the consultancy firm BCG and MIT as well as SAS, a lack of management support for projects are brought up as a challenge which is the reason this item is included in this section.

### 5.3 Interview Insights

#### 5.3.1 Main Pitfalls

At Cinnober, it is considered a risk to put a small group of people aside for a machine learning project that is not directly revenue creating. This prevents these types of projects from being realized. It might be because of there being a lot of older employees at Cinnober according to Nordström. The age may not be a direct factor, but the lack of a recent education within the area might. Nordström believes the
projects have to be initiated by younger people with a recent education within data science or machine learning. While he believes management will welcome possible initiatives, he also says they probably will not initiate them themselves until it is a direct request from a customer.

With many different kinds of customers running different businesses and in different countries, iZettle has lots of data, but the data varies vastly. Regarding certain types of customers there is a lot of data and regarding some others, there is not. A challenge iZettle faces are to use the large amounts of data to draw conclusions and to utilize the data in an efficient way. Hansander describes the concept of never having a satisfying amount of data because as soon as the amount of data is large enough, it is slice-able into smaller segments until each segment can benefit from having more data.

Väljamets brings up a general problem, the belief that many problems are solvable with models and machine learning, which causes people not to properly consider the problem and what kind of solution that is desired. Jumping to the conclusion of using AI over other tools or methods to solve problems is not uncommon and is considered a pitfall. Furthermore, there are often difficulties in the transition from having an algorithm or a trained machine learning model in one system, moved to the production system. At Klarna, this has been done several times, making the company build experience and knowledge of this process and its challenges, but it still takes a lot of time.

Another problem faced is the problem of evaluating the models after being deployed. For example, with the payments methods ranking, it is difficult to measure the effect of the model. This is because the customer might choose the first presented option just because of it being presented first and not because it would have been the preferred option in a neutral situation. Because of this, the model might show a change in consumer behavior, but knowing if it is for the better is the challenging part.

Evaluating which improvement projects to take on is one challenge at Telia. Estimating how much revenue a solution will generate is often unreliable because both the business value and the model’s performance need to be estimated. When considering new projects, data quality is assessed which has a big impact on whether to continue the work on the project or not. When deciding which projects to take on, both increases in business value in case of successful implementation and chances of successful implementation are considered.

One big pitfall is to not have relevant data accessible before starting the implementation. Besides accessibility, the quality of the data is also of high importance and is one of the biggest problems. At Telia, which is an old and big company, there are many sources of data of varying quality. All sources are not connected and the data management solutions look different from case to case which causes another layer of complexity to the solution.

Another challenge is people’s ability to believe that machine learning and artificial intelligence is magic. Some people view the process as sending data to the data science team and receiving money.
5.3.2 Stages of Machine Learning Projects

Considering Cinnober have not implemented any machine learning solutions yet, the process question was limited to guesses from experience of the company’s way of working and preconditions. Nordström identified data collection and preparation as the most pitfall prone area. This is because of legal issues and confidential data that are two challenges present across the financial world. Because of financial data sometimes being confidential for an amount of time after being created, it can be a challenge to acquire training data that is not too old and still relevant.

When presented with the different stages of machine learning implementations, Hansander at iZettle describes pitfalls when handling data and when deploying models to production making problems that were not previously visible arises. After deployment, there is also a need for monitoring the performance of the model. The more models deployed to production, the bigger is the work of monitoring and maintaining them. Machine learning models will not break, they will just begin to perform poorly which requires more sophisticated monitoring and maintenance work. Often there is a change in customer behavior that causes a model to perform poorly. A change in pricing in some product can cause the customers’ behavior to change, resulting in a decreased performance of a model since it is trained on old data where the customer behavior was different. There is a challenge in discovering these scenarios as fast as possible.

At Klarna, data collection and preparation is the hardest and most problematic part. When preparing data there is also a need to decide what data to evaluate the model on. If that data is biased it might look like the model is performing well during testing, but when it is released to handle unseen data it might perform worse. Väljamets also raises the issues that may arise when releasing the model to production from the internal test systems.

At Telia, besides data management feature selection, or creating features from data, is identified as a step that takes a lot of time. This is because it is hard to ever be done due to the many opportunities to improve the features.

5.3.3 Machine Learning Hypotheses

Wrong choice of algorithm

Neither at iZettle, Klarna, nor Telia, choosing the wrong algorithms is considered a big problem. The interviewees all highlight it is not easy to do right, but doing it wrong will not cause a problem. Testing different algorithms before choosing one is not a very time-consuming task compared to for example managing the data. Dedicating more time to this stage will not considerably change the time plan for the project. Väljamets raises a general problem of too advanced algorithms chosen for not as advanced problems, which results in an overfitted model that will not perform as good in production as it does during testing.
Not enough data

Not having enough data is considered a problem at all interviewed companies. At iZettle, Hansander explains how retroactive collection of data is hard or impossible, and at the very least expensive, with the cost being for example user experience or lowered conversion. Väljamets explains how models easily will be overfitted when there is not enough data to match an advanced algorithm. At Telia, the problem of owning the data is only present when implementing machine learning solutions in new areas. When doing improvements in areas where the company previously have worked with data, just not yet through machine learning, there is often a lot of data stored regarding the specific area. The problem Telia is facing in that position is instead not having the data stored properly and lack of quality in the data.

Data on wrong form or not structured properly

All interviewees responded that data stored in wrong form or data not being structured properly is definitely a pitfall. Both at Telia and iZettle, it is not seen as an unsolvable problem, but they both indicate it takes time and that this is a big part of any project. The upside is that it is an internal problem in which the customers or end user are not involved. At Klarna, the data is considered structured in the sense that it is not stored in videos, pictures etc, which are hard to translate into more concrete data. Although, it still needs structuring and handling of null values etc. Also, there is a lot of work building features from raw transaction data. A correct structure of the data is important when testing, but even more so in production, when for example a loan should be approved or not within a second. Then there is no time for a person to for example parse .json files or handling null values.

At Telia, there are people working on structuring their large amounts of data and make it accessible to everyone within the company. Because of the size and stability of the company, there are enough resources within IT to be able to make this happen. Telia have implemented a common data lake, a way of storing all data in one place. When the data exist in the data lake, there is no longer a need to access a certain system to get a certain kind of data. Everything is in one place. But, as with the problem of not having enough data, when Telia works with machine learning in a previously unexplored area, the data has not been processed and integrated into the data lake which adds complexity to the project. This will result in vast differences between projects depending on the data availability.

Wrong Features Selected and/or Wrong Parameter Selection

Selecting the wrong features or wrong parameters and model are two situations that are similarly handled at iZettle and Telia. The case is not that it always is done correctly, it is just not a relatively big problem when it is not. Neither at Klarna is this seen as problematic.
Overfitting

All interviewees agree that overfitting happens but is usually handled without much interference with the time plan or spending too much resources. There are several techniques to prevent this from happening, including testing on different sets of data and do out of time testing meaning using a test set that is from a later time period than the training set.

Biased training data

Having biased training data is defined as a more difficult problem at iZettle, because of the usual late discoveries of the problem when it happens. Väljamets at Klarna believes this pitfall is very common if there is not enough machine learning competence within the organization and that most people do not see this problem. With the ease of acquiring data and the availability of good open source software, it is not hard to build and train a model. But if the training data is not carefully chosen it might not be representative. This will result in a model that is not performing very well. Väljamets highlights the importance of understanding how much the data will affect the outcome.

Incomplete testing

At Telia, incomplete testing is not really a problem, because there is a correlation between the value of a project and the resources spent on testing the implementation. At Klarna, it is a challenge both within regular software development and machine learning projects. Since the customer data is used to give customers decisions about loans in real time, test is even more important to make sure that no customer will accidentally get a very high credit score making them eligible to get unlimited loans.

Bad handling of changes in jobs

At iZettle, problems with handling changes in jobs are not yet encountered because no jobs have been replaced due to implementations of machine learning taking their place. They aim to build processes that do not scale with people. In practice, this means that if iZettle get twice the amount of customers, they do not need twice the amount of employees. At Klarna people often move between projects, and there is a fast-paced working environment within machine learning. The issue could be a higher risk of bugs. Telia has a similar problem and Peters highlights the importance of documentation, which is not always prioritized. Regarding replacing workforce with machine learning solutions, Peters says it just changes people’s jobs, not replacing them.
Not developing trust for AI solutions within the company, cultural challenge

Neither at Klarna nor at iZettle there is a problem of not trusting machine learning solutions. Employees understand the basics of machine learning and value the outcomes without questioning it, on a developer level as well as a management level. At iZettle, employees sometimes want an explanation for the decisions or predictions made. This is not because of a lack of trust, but more because of the ease of discussing decisions and understanding them. At Telia, the trust of machine learning predictions can vary greatly between different parts of the company, possibly correlating with age and number of years within the company. There is an ongoing effort to make Telia more data-driven.

High deployment cost/lack of ROI

The understanding of machine learning projects not always succeeding is high at iZettle, meaning the lack of ROI is not a problem. The successful projects pay off. Klarna has high deployment costs but it is not something hindering the projects from being started. At Telia, a lack of ROI was a problem earlier. When entering a new area, it is hard to estimate ROI, but when doing similar projects the estimations become more accurate with time solely because of experience.

5.3.4 Project Management Hypotheses

Lack of competence

All companies believe it is hard to find machine learning and data science competence. Klarna specifically notice a shortfall in more senior competence. At Telia, there is enough of work waiting to be done to hire more people. Peters also believes many projects would benefit from having people with machine learning competence in the workforce, but this is not the situation yet.

Poorly estimated schedules/budgets

Again, at iZettle, there is a flexibility and understanding of machine learning projects not being executed as planned. There is a continuous iteration over the estimations to keep them up to date. Time estimation at Klarna is difficult, however not more than other software projects. At Telia, the estimation in the beginning is often naive and when the project has started it is often realized that it will take more time than initially planned. The estimation is especially hard in new areas. There is also an issue with communication to other employees or stakeholders with less understanding of machine learning that more resources are needed.
Requirements changes

As with schedules and budget changes, iZettle has adapted the agile mindset and the acceptance is high for requirements changes making it a lesser problem. Klarna’s business model is changing rapidly which leads to rapid changes in requirements and prioritizations as well. Väljamets describes it as being simultaneously a problem and something that is expected and handled. At Telia, the requirements changes happen mostly because there is a realization from within the machine learning team that their project is not going to be successful. Then communicating this to stakeholders is not always easy.

Lack of management support

The prioritization of long-term investments in machine learning and smaller short-term changes in order not to lose money in current projects is sometimes difficult at Klarna. Often, management wants both solutions at the same time which is not possible.

There is an effort in becoming more data-driven through the whole organization at Telia, meaning top management being supportive of machine learning even though there may be little understanding of its effects and value. Though, it is necessary to be able to motivate in numbers what can be done with data to have full management support.

5.3.5 Differences in how Machine Learning Projects and Other Software Development Projects are Treated

The difference between machine learning projects and other software development projects is that the machine learning projects are allowed to be more exploratory in the early stages at iZettle. This is because of the difficulty of knowing exactly what can be done before starting the project, which is easier in other software development projects at iZettle. At Telia, the situation is different. There is a lack of understanding of machine learning and therefore an expectation of making magic from data and not really realizing the amount of work behind the solution. Peters describes the data science team’s efforts as just solving a problem within a department and then leaving without having the department in question involved very much. At Cinnober the discussed projects are seen as high-risk projects and harder. So far, machine learning projects are seen as proof of concept projects at Cinnober.

5.3.6 Additional Comments From the Interviewees

Both at iZettle and Telia, also mentioned by Nordström at Cinnober, regulations are a big part of machine learning projects which takes up time. Sometimes this stops ideas from becoming realized at Telia. Hansander at iZettle also brings up the important question of continuous education within the company. This area of technology is a fast moving area making a plan of educating employees more important. Choice of tools and architecture design always need to be up to date.
to build solutions that can evolve. Hansander also highlights the importance of not catching on to every trend, just because something is hyped.
Chapter 6

Discussion

6.1 Findings

The companies can be divided into four groups according to their progress within incorporating machine learning into the business. The first group encapsulates companies which have yet to implement any machine learning solutions (Cinnober). The group is called Unproven path. Companies that have realized some initiatives places in the Scattered machine learning initiatives group (Telia). The last two groups are called Machine learning supporting product (Klarna) and Machine learning defines product (iZettle).

Through the insights gained from the interviews, three main conclusions are derived:

- Overestimation of machine learning capabilities varies as experience increase
- Data related pitfalls are the most severe
- Correlation between the ability to appreciate long-term solutions and machine learning experience

Below follows a discussion of their importance and impact.

Overestimation of machine learning capabilities varies as experience increase

Machine learning is categorized as being “on the peak of inflated expectations” according to Gartner. This is also indicated by the companies. This is visible from two perspectives; the developer perspective and the management perspective. The developer perspective is visible at Klarna through the willingness to use too advanced models on too small amounts of data, resulting in overfitting. The overconfidence in the model is thereby indicated to be increased. The management perspective is visible at Telia where there are difficulties experienced with accurate scheduling or acceptance when projects are not successful and an overall overestimation of what this type of technology can do. Both these perspectives can be translated to a lack of understanding of the resources which a machine learning project needs in order
to be successful. A more realistic view is beneficial, rather than believing in a black box solution that works like magic.

Based on the results of this study, indications point to overestimation varying with experience in machine learning. The overestimation seems not to be present at Cinnober, due to their unwillingness to so far starting machine learning projects. At Telia, possessing some machine learning knowledge, the overestimation is visible in the management perspective discussed above. At Klarna, with even more knowledge within the area, overestimation is visible in the developer perspective. There are no indications of overestimation of machine learning capabilities at iZettle, who are the most experienced in machine learning. This can be visualized as a bell curve, with overestimation increasing as an organization is in the early stages of learning machine learning, and after a certain point decreasing again. This could be because of the lack of knowledge from start, making it hard to overestimate something we do not know anything about. As knowledge increases and some implementations are done, an organization might start to realize the benefits of the technology and exaggerate them in excitement and hype. When becoming even more advanced within the area, a more realistic picture starts to form and will by more knowledge acquired look more like reality.

![Figure 6.1: Degree of overestimation relative experience](image)

At Telia, this problem could be increased by their way of working with machine learning. With a team working as consultants implementing machine learning solutions in different departments, they may work isolated from the departments daily work and does therefore not transfer knowledge about the resources a machine learning project needs. This would increase other employees view of machine learning working like magic. At iZettle, the opposite is happening, where continuous efforts to educate employees are made. This will spread the knowledge through the organization instead of isolating it within a group.

The belief that machine learning can solve complex problems with little resources and without much risk causes unrealistic expectations and need to be managed to prevent this. This could affect requirements, scope, time planning etc in a negative way because of the unrealistic view.
Data related pitfalls are the most severe

The results from the interviews clearly indicate that data management being the largest and most demanding part of machine learning projects. This includes collection, structuring, and cleaning of data. The problems encountered varies between storing the data on a wrong form or spread out over several sources, not having enough data, low quality of the data, or biased data. Structuring the data is indicated to be a bigger problem for Telia and Cinnober than Klarna and iZettle. This could be because the data already collected has not been collected with machine learning in mind. If that is the case, it is not surprising since machine learning places on the top of Gartner’s hype cycle now, and has not been before. With Klarna and iZettle being younger companies than Cinnober and Telia, with a large relative growth during the years of which machine learning has been increasingly discussed, their growth is likely machine learning influenced to a larger degree.

With more machine learning experience, data structuring seems to not be seen as a big problem, but there is a broad understanding of it being a time-consuming task. At more experienced companies the more complex problems and problems that are difficult to detect like biased training data is given more attention.

Ben Hamner, co-founder, and CTO at the data science platform Kaggle discusses in his speech at Strata conference in 2014 four problems in machine learning projects. They are data leakage, overfitting, data quality, and data sampling and splitting. As an experienced machine learning engineer, Hamner’s choice of four data related pitfalls strengthens this claim. If the data related pitfalls are not the most severe, Hamner’s choice of problems to present would probably have looked different. By the end of Hamner’s speech, he concludes there are lots of pitfalls in machine learning projects, although they can be avoided.

Correlation between the ability to value long-term solutions and machine learning experience

A problem shown in several of the interviews is the lack of long-term investments in machine learning. Often, problems are solved with the quickest and most short-term solutions. This is indicated by the lack of commenting code at Telia, suggesting it is higher prioritized to deliver a solution and then move on. The same thing is happening at Klarna where often the short-term economical aspect is treated more important than long-term gains. If a problem is making the company lose money because of a poorly performing model, it is more common to use a quick fix for the specific case in question, than actually correcting the underlying problem. At Cinnober, the strategy of not building machine learning solutions for problems because of the lack of concrete revenue is also an indication of not prioritizing long-term effects regarding machine learning. iZettle, on the other hand, seem to be more aware of long-term effects by driving education efforts within machine learning and continuously hiring new competence. The acceptance of projects not always being successful also indicates a long-term view on learning processes. Also Klarna indicating their team is growing. The more machine learning experienced companies also seem to be more aware of the benefits of long-term efforts shown in figure 6.2.
In an article from SAS about machine learning mistakes that beginners do, four out of five mistakes discussed are in the area of not doing sufficient planning and preparation before starting the project (SAS, 2018). A lack of planning short-term may also indicate a lack of planning long-term. This, in turn, indicates a lack of valuing long-term solutions. Otherwise, more planning would have been done.

6.2 Validity, Reliability, and Generalizability

Validity is defined by Collis and Hussey (2014) as the extent to which a test measures what the researcher wants it to measure and how well it describes the phenomenon under investigation. The validity of this thesis is high. The interviews conducted are with industry experts with experience in the area of machine learning implementation projects at their companies. They are experienced enough to discuss the topic during the interviews and provide new insights and experiences. With a literature review and hypotheses prepared for the interviews, the risk is minimized that they are spent on unnecessary discussions. Two of the interviewees wanted to add pitfalls that were not on the list of hypotheses which increases the validity since more of the topic is covered.

The reliability of a measure refers to its consistency. If the research were to be repeated, the reliability is high if there are no differences between the projects (Collis & Hussey, 2014). Reliability when doing interviews is lower, because responses and information retrieved during the interviews may differ if other interview subjects are chosen if the study should be repeated. The responses are personal and depending on background and experiences of the interviewed individual.

Generalizability is the extent to which the research findings can be extended to other cases or research findings (Collis & Hussey, 2014). As the companies are chosen as a sample group to be more generalizable and extended to other cases, the ability to extend the work to other cases exist.
6.3 Issues and Limitations

As previously discussed, additional interviews would have strengthened the thesis, making the conducted interviews a limitation but not an issue. Although, an issue with the interviews may be interviewees potential unwillingness to disclose all mistakes and pitfalls they have experienced in order not to look less professional to readers of this thesis, who might be present or future customers, or potential future employees. Due to the companies willingness to take part in this study and openly share what was said in the interviews, there is not much reason to believe they want to withhold information, although it can not be disregarded as a potential issue.

Sampling as a method is, as stated earlier, useful when resources are limited and there is no possibility to collect information from a whole group. However, it is difficult to know the confidence level of the results. Conducting interviews globally should also increase the confidence level, but would need substantially more resources.
Chapter 7

Conclusion

The questions to be answered in this study are which pitfalls organizations experience when conducting machine learning implementation projects, and how the pitfalls may vary with the level of machine learning experience that exist within the organization. To answer the research questions posed, hypotheses forming through a literature review followed by interviews and an analysis of the results have been performed. Three main findings are presented: The interviews indicate a correlation between machine learning experience and the extent to which the organization value long-term solutions, having the value increasing with experience. Additionally, data related pitfalls are by the interviewees considered more severe, taking up more time than other problems experienced. Lastly, indications show that how much a company overestimates the opportunities of machine learning varies with experience increase, starting out low, thereafter increasing to some point and then start to decrease.

With many executives believing their company will benefit from AI but not yet have implemented any solutions, the future is likely to hold many new AI projects and discoveries of new pitfalls.
Chapter 8

Author’s Thoughts

This chapter is solely based on my own thoughts and reflections after finishing this thesis and being invested in this area of research for five months. The following discussion is based on my knowledge and thought patterns gained from my master’s education within Computer Science and Industrial Management at KTH Royal Institute of Technology.

I would like to address the questions whether there could there be a general pattern here, as well as if we can generalize the findings of the thesis and to which extent.

The first thing to address is the question of what impact the size of a company has on the findings. Either size in itself is a factor for how companies handle the AI wave happening right now, or the ”size to machine learning knowledgeable employee ratio” is a more relevant factor. For example, in a very large company with lots of other people running the daily operations, a machine learning team of 50 can accomplish a lot. Although the size to competence ratio is low, a team that big can still build business changing solutions. On the other hand, the lower the competence/size ratio is, the more the non-machine learning affected parts of the company are left in the dark about the applications of the technology.

A more specific question to ask is the relevance of size regarding companies who have yet to try their hand at machine learning solutions. As company size increases, it could possibly become harder for employees to reach the actual decision makers at the top to start a new machine learning initiative. If this is true, the ratio of competence/size is more relevant than just the number of machine learning knowledgeable employees. This implies a bigger team will have a bigger impact on the decision makers.

Further, a high ratio of competence to size may start a positive spiral in companies. When there are many employees knowledgeable in machine learning, more initiatives and ideas will likely surface. This leads to more work within machine learning which will, in turn, lead to more competence creating learning opportunities. This would imply an exponential increase of machine learning competence growing faster than in a company with a low ratio.

Regarding valuing long-term solutions, very small companies without much competence might not fit the pattern. The reason behind this could be that there are higher prioritized issues to solve regarding the survival of the company, that machine
learning investments can not be considered even though the long-term benefits are understood and valued. This would imply the suggested model proposed is a better fit for companies with stability and financial margins.

Data pitfalls being the most severe is likely a general insight since all of the interviewees agreed clearly on this. Although it is not fully covered by all reports on the topic since some are more focused on a management perspective rather than a technical, it makes sense for it to be part of a general pattern. There are lots of open source software to help companies build a model that will be useful for its purpose. This part, therefore, can become general and almost identical in many companies. But, the data is in almost all cases company specific and only existing within the company. This means that without proper knowledge, all steps beside data management can be completed with good results but the data management is too company specific and sensitive to generalize. Also, due to the large amount of data used, small errors are magnified and will impact the overall solution.

All in all, with the use of the companies in this study as a sample group, I believe the indications are showing a general direction, although further research should be done to cover edge cases and increase validity or level of detail. The general pattern suggested is possibly a better fit for stable companies with financial flexibility.
Chapter 9

Future Research

The companies included in the study vary in size and competence but do for obvious reasons not cover all cases. For future research and to generalize or contradict the findings of this report, the blank areas in figure 9.1 could be covered through a similar study. This would, at first, include relatively big companies with moderate machine learning competence and even bigger companies with a bit more competence within the area. Examples of two type of companies to study are a 3,000 employee company with a machine learning team of 10 and a 7000 employee company with a machine learning team of 25. To add to this, also companies with 10,000+ employees and companies with a machine learning team of 40+ would be interesting to study. With the results from this thesis, future research can strengthen and confirm the conclusions drawn in this thesis, as well as find evidence opposing the conclusions in all or certain cases. Either way, interesting and useful conclusions can be drawn.

![Figure 9.1: Blank areas in the graph which future research could cover](image)

Further, the conclusions drawn by this case study of four companies could be investigated further. By using this study as an initial step, the conclusions can be studied
in depth to reach additional insights of their effect on organizations. By focusing on the three main conclusions, an investigation that does not study the companies as much in depth in favor of studying a larger amount of companies can be done. A suggestion is to use another method and instead of interviewing use surveying to collect data from a larger group of companies, focusing solely on the problems mentioned in the conclusions.
Bibliography


Väljamets, Sara (2018). Personal Interview.

Appendix A

Interview Questions

1. Which Artificial Intelligence/Machine Learning initiatives have been discussed within the company?

2. Which AI/ML initiatives have been started within the company?

3. Which AI/ML initiatives have been finished within the company?

4. How much AI/ML competence exist within the company?

5. Which challenges/pitfalls have you experienced in the projects?

6. What is stopping the company to continue working on started or discussed projects?

7. Division of project stages regarding machine learning. Pitfalls experiences in any of these?
   - Data Collection and Preparation
   - Feature Selection
   - Algorithm Choice
   - Parameter and Model Selection
   - Training
   - Evaluation

8. Potential machine learning hypotheses. Confirm or deny?
   (a) Wrong choice of algorithm
   (b) Not enough data
   (c) Data on wrong form/not structured properly
   (d) Wrong features selected
   (e) Wrong parameter and model selection
   (f) Overfitting
   (g) Biased training data
   (h) Incomplete testing
   (i) Bad handling in changes in jobs
   (j) Not developing trust for AI solutions within the company, cultural challenge
   (k) High deployment cost/lack of return on investment

9. Project management hypotheses. Confirm or deny?
   (a) Lack of competence
(b) Poorly estimated schedules/budgets
(c) Requirements changes
(d) Lack of management support

10. Are AI/ML projects treated differently than other software development projects? How?

11. Additional comments?