MACHINE LEARNING WITH CARE

INTRODUCING A MACHINE LEARNING PROJECT METHOD

Master thesis
Management of Technology
by Steven A. Hoozemans

TU Delft

DIGICAMPUS

Logius

This page is intentionally left blank.

Machine Learning with Care - introducing a Machine Learning Project Method

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in Management of Technology

Faculty of Technology, Policy and Management

by

Steven Alexander Hoozemans

Student number: 4598954

To be defended in public on 10-08-2020

Graduation committee
Chair: Prof.dr.ir. M.F.W.H.A. Janssen, Section ICT
First supervisor: Dr.ir. N. Bharosa, Section ICT
Second supervisor: Dr. M.E. Warnier, Section SE
External supervisor: B. Groenveld, Logius

This page is intentionally left blank.

Acknowledgments

A free translation of Alice in Wonderland, “If you don’t know where to go, you won’t get lost” can be seen as a metaphor to find the right answers in exploring immense databases. As a child it gave me a defensible excuse as a young explorer to enjoy everything that crossed my path, without having a clear, defined goal and direction. As an MSc-student, it gave me the right direction to and focus for my graduation project. Without a good research question and confidence in the direction of my research and the intended end station, I would not have been able to finish my thesis successfully. I am therefore proud to present my thesis, as graduation project for the master's program Management of Technology (MoT) of the Technology, Policy, and Management Faculty of the TU-Delft.

With the defence of my thesis, a wonderful time as a student in Delft comes to an end. I can look back on a very instructive transition from BSc Mechanical Engineering to the master program of MoT. This study period created the perfect environment to develop myself: the professional atmosphere the TUD has offered me, the challenging curriculum, and the personal, constructive guidance, has been very supportive.
Also, I am grateful for the unique chance to do my graduation project for the DigiCampus, an exciting place with challenging projects. It gives a lot of satisfaction to work for this organisation and especially to be part of the team of inspiring colleagues. I hope my results will contribute to the future research and branding of the Digicampus.

I would like to thank everybody who contributed to my work last six months while completing my thesis. Because of my intrinsic motivation and interest for the subject, I was more than excited to work on this research project. The previous period has been one of the most demanding periods in my student days but I have always enjoyed working towards the final result.

I specifically want to thank my first supervisor, Nitesh Bharosa for providing guidance throughout the entire thesis period. During this process, Nitesh has been a very supporting and devoted supervisor and always willing to schedule time for me. I also want to thank my chair, Marijn Janssen and my second supervisor, Martijn Warnier, for providing their valuable feedback. They have always been very willing to reflect in the meetings and I am very grateful for that.

I also want to thank Bas Groenveld, Logius, and all the interviewees that have provided me with their valuable time and input. Their input has been an essential aspect to prove my theory in building a proper machine learning method.

Last but not least, I want to thank my family and my friends and for their patience and mental support.

Steven Hoozemans
Student Management of Technology, TU Delft.

Executive summary

Worldwide, data is an essential resource for economic growth, innovation, job creation and societal progress. Therefore, technology focusses on applications translating data into usable variables for science or organisations. Optimising data analytics is essential for improving efficiency and innovation. Artificial Intelligence, in particular its' subset Machine Learning, is increasingly used to achieve optimal analytics and outcomes in business and governmental organisations. Successful Machine Learning projects are dependent on multiple factors. These include correct algorithm selection, compatibility with organisation strategy and being in harmony with ethical standards. Combining these aspects into a method to set up machine learning projects could create great potential.

The demand for structural guidance for implementing machine learning has become evident – exampled by the recent issue regarding child allowance and Dutch Tax organisation. Lack of understanding regarding the technical aspects of machine learning, the organisational aspects of successful implementation of machine learning projects and the difficulties coming with meeting ethical standards, contribute to the suboptimal use of machine learning projects. In the governmental sector, Standard Business Reporting (SBR) is used by Dutch organisations to improve the exchange of financial data. Within the need to improve the ongoing development circle of machine learning techniques, SBR could be a promising case. Systematic scientific literature analysis identified the gap of a structured method for setting up machine learning projects that consider a combination of the technical steps, organisational aspects and ethical aspects. A demand for structured guidance is called upon.

The research question of this master thesis is to fill this existing gap and is stated as follows: **How can technical, organisational and ethical aspects be combined into a method that supports stakeholders to systematically set up machine learning projects in SBR context?** The outcome of this thesis is a method to help actors in SBR context to systematically set up machine learning projects. Furthermore, it will assess the use of machine learning in SBR context. Additionally, it will support the stakeholders to determine if the proposed machine learning project could be viable in their organisation so that the stakeholders have a better chance to develop a successful machine learning project.

The chosen research approach to develop this method is the Design Science Research Methodology (DSRM). The DSRM contains six steps: problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, and communication. The first step, problem identification and motivation, was formulated in the previous paragraphs of this summary.

In order to complete the second step of the DSRM, defining the objectives for a solution, the problems identified has been converted into design objectives to formulate what type of solution would be desirable. The design objectives were formulated on the basis of an extensive literature review towards ethical, technical and organisational aspects, and based on experiments conducted in collaboration with DUO. These experiments included one regression experiment, one classification experiments and the production of a Strategy Map.
As a result, the following six design objectives were formulated and divided into two categories: what the method should include, and what the method should provide.

Design objective focussing on the method, to include:

1. **The designed method should include an ethical framework**

2. **The designed method should include machine learning steps to create a model in SBR context**

3. **The designed method should include a machine learning algorithm selection method, including multiple machine learning techniques**

4. **The designed method should include organisational factors relevant for creating a machine learning project**

Design objectives focussing on the method, to provide:

5. **The designed method should provide an understandable process for creating a machine learning project in SBR context**

6. **The designed method should help decision-makers to understand if machine learning can create added value in their organisation**

The third step of the DSRM is focussed on the design and development of the method. The relevant data and insights for the development of the first version of the method in SBR context were derived from the literature review and experiments conducted in collaboration with DUO, an SBR stakeholder. Three important factors were extracted from the literature review and are combined into the method: the Ethical Impact Assessment for ethical aspects, the Strategy Map for organisational aspects and Knowledge Discovery in Databases for technical aspects. In addition, in collaboration with DUO, two machine learning experiments and one Strategy Map based on the strategy “using Machine Learning” were carried out. The input of the literature and the experiments resulted in a first tailor-made Machine Learning Project Method version. The designed method includes ten unique steps for setting up machine learning projects in SBR context, taking into account ethical, organisational and technical aspects.

Before the second iteration in the development of the method (Design Phase 2), a small but relevant selection of interviewees was made, and six semi-structured interviews were conducted. During these interviews, the respondents were asked to systematically evaluate the first version of the designed method and provide suggestions for improvement. The input of the interviewees was analysed, and eleven relevant suggestions were determined and implemented in the second design iteration.

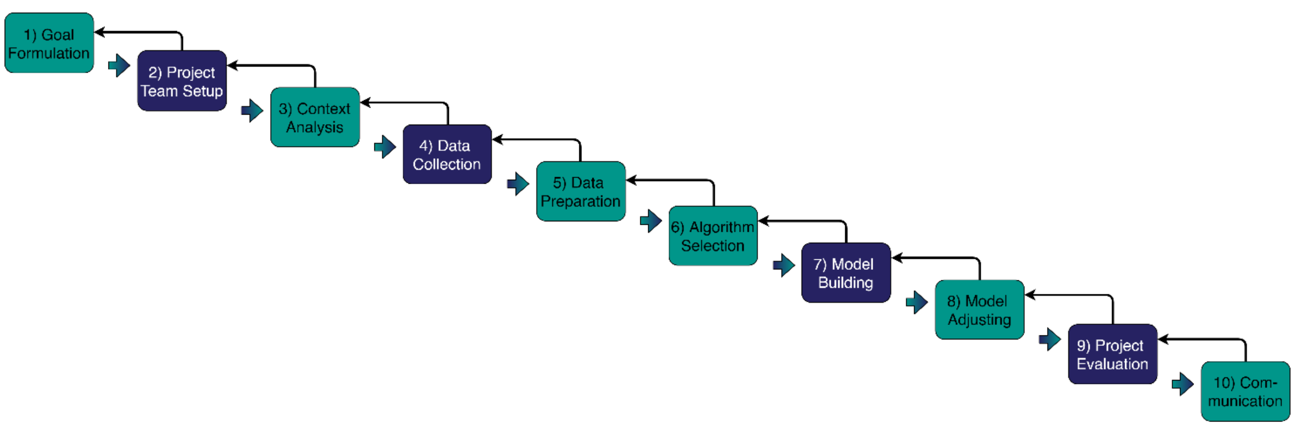

Figure 1. Concise version Machine Learning Project Method

This exercise resulted in the updated and second version of the method. The improved method is subdivided into ten unique steps: goal formulation, project team setup, context analysis, data collection, data preparation, algorithm selection, model testing, model adjusting, project evaluation, communication; all with related sub steps. A concise overview of the final research product, the Machine Learning Project Method, is visualised in Figure 1.

The method helps the user to evaluate the use of machine learning in their organisation by providing the stakeholders with a systematic process for creating a machine learning project in SBR context. Furthermore, the method includes an algorithm selection method. The designed method provides a comprehensive list of options which algorithms to use for a specific case, a framework to assess the ethical impact and various other aspects important to create a successful machine learning project in SBR context.

The fourth step of the DSRM is to demonstrate the designed artifact as a result of step three. In this research, the Demonstration of the designed method is done in collaboration with WSW. A project team was set up to use the designed method. The aim of the project was to estimate the added value of machine learning for WSW by building a machine learning model that determines the financial risk labels. After completing all steps of the method, a machine learning model was delivered: providing WSW with insight into the added value of machine learning. WSW was therefore able to expand their insights on determining the financial risk of their stakeholders. Furthermore, the demonstrate was successful in showing the methods' applicability on a real case.

Evaluation is the second to last step of the DSRM. First, the method is evaluated. In summary, the design objectives are essentially completed. Although the design objectives were ambitious, they are all incorporated in the designed method. However, in order for the designed method to become a fully operational and validated method, the design objectives should be further investigated and developed. Supported by the interviews, the designed method has demonstrated to be effective in setting up a machine learning project in a real case. Second, machine learning in SBR context is put in perspective. Although this research recognises the potential of machine learning in SBR context, the experiments show that for implementing machine learning in the SBR context, the conditions are still not optimal and therefore are not ready yet to replace the current systems. However, it should be taken into consideration that the method is developed for a specific context: the SBR context. It is not clear whether it can be applied in different context. Furthermore, two out of four machine learning techniques are included in the designed method: the supervised and unsupervised machine learning techniques, and supervised machine learning has been tested during this research. The designed method is published on GitLab and the TU-Delft Repository to facilitate further research and development (*Digicampus and Data | European Data Portal*, 2020), thereby completing the final step of the DSRM.

In conclusion, this thesis describes the successful development and testing of a method: the Machine Learning Project Method. This method includes an algorithm selection method. The Machine Learning Project Method provides a structured method that helps managers to understand the process of setting up machine learning projects and provides them with guidelines on how to setup a project. Different aspects of managerial domains are integrated: organisational and ethical aspects and guidelines on managerial implementation.

The scientific contribution of this thesis lies in the theory of the designed method. The new method enhances the understanding of machine learning projects in SBR context, a form of structured data. This method, which combines technical, ethical and organisational aspects in a systematic approach, enables its users to obtain knowledge of the added value of machine learning, and to set up machine learning projects. The integration of these three pillars into a single method was not yet available. Therefore, this method fills the gap that other methods left open. At this moment it is unique in comparison to other methods, as it combines interdisciplinary aspects into one method. The newly created method is added to the scientific field and is shared to facilitate further research and development. Furthermore, the research provides insight of machine learning in SBR context. Evaluating the results of this research, it is found that at this moment, machine learning is not yet capable of generating the desired application and outcome in SBR context. Therefore, this research contributes to the development of the use of machine learning on structured data.

The practical contribution of this research is that the method provides a structured and partly iterative process to set up machine learning projects. Following the Machine Learning Project Method, a machine learning model can be made with respect to organisational, ethical and technical aspects. This allows the user of the method to evaluate the added value of machine learning in the organisation: it guides users towards asking the right questions, also making them aware of the limitations and impossibilities of machine learning. This method prevents initiating a machine learning project without an estimation of the applicability. Furthermore, it gives managers, policy makers and engineers an overview of what it takes to start a machine learning project, including preconditions and restrictions. It provides insight into possible applications of machine learning and enables a structured process for both engineers and managers, creating alignment and understanding between management and engineers. When all steps are completed, the method provides the following deliverables: insight into whether machine learning has an added value for the organisation and an ethical, potentially cost efficient, yet simple, prototype machine learning model.

The designed method has proven to be reliable in achieving a machine learning project with usable outcome and is presented on GitLab and shared on the European Data Portal (*Digicampus and Data | European Data Portal*, 2020) for further use and development. The method developed fills a niche in the current (knowledge) gap reviewing the application of machine learning within SBR in line with the research question. Considering this demonstrable added value, it is useful to further elaborate and apply this method. Recommendations for future research can be summarized in three-fold.

1. Further focus on fine-tuning of the methodology, with the aim of a fully tested and operational model, including sufficient iteration steps. This will be an important next step to translate this prototype into a reliable model in a professional, operational organisation using SBR.
2. The promising generic part of this model should be conceptualised in order the broaden the concept for machine learning in wide scale of algorithms and applications. This could ultimately contribute to uniformity within use of machine learning.
3. Machine learning and ethical issues go hand in hand. The risks this entails is still underestimated in operational applications. Further research into the possible negative impact of machine learning on society must be conducted. Therefor it should be researched how organisations and society can be included in the process of constructing machine learning.

Keywords: Machine Learning, DSRM, Standard Business Reporting, ethics, Machine Learning Project Method, Strategy Map

Table of content

List of figures

List of tables

Part I

Problem identification and defining objectives for a solution

Chapter 1. Problem Introduction

This chapter elaborates on framing this thesis' research objective in the broad context of working with data and artificial intelligence. In Section 1.1, the background of the problem is described. Subsequently, in Section 1.2, the societal relevance and managerial knowledge gap are displayed. Following the managerial knowledge gap, the related academic relevance and knowledge gap is provided in Section 1.3. The last section, Section 1.4, focusses on defining a clear research objective, including problem statement and research questions.

1.1 Background

Data, the fuel of human society

Data is inseparable with human society. Since the start of human history, data has been collected and used. In ancient history libraries were built to store observations, the first theorems were created. Data collection is being used ever since, to gain insights in processes in almost every domain (Miyazaki, 2015). The evaluation of the form of data itself - from tally sticks to digits - and of the tools to analyse it - from calculators to spreadsheets - has ultimately laid the basis for the data processing we know today (Cohen, 2018).

Data as a resource contributes to economic growth, innovation, job creation and societal progress. Therefore, developing applications for better data use can benefit research and business. Important improvements consist of improving efficiency of processes, generating new products and service and providing a solid foundation for innovations (Volman, 2017).

The growing interactions between data, algorithms and big data analytics, connected things and people are opening huge new opportunities. Market analysis, analysing customer feedback, discover the strengths and weaknesses of their competitors, analyse valuable customers, and make smart business decisions could be mentioned as examples (Han et al., 2011, p. 27). In the same time, they are also giving rise to issues as "data governance" at the national and international levels. These include questions around the management of data availability, accessibility, usability, integrity and security, as well as concerns about ownership and implications for personal privacy and ethics.

Artificial Intelligence, computers that behave like humans

Last decades, due to an exponential increase in digital data, Artificial Intelligence has built up its base in the analytics of data. Quoting Moore who is Carnegie Mellon Dean of Computer Science, the modern definition of Artificial Intelligence can be described as "*the science and engineering of making computers behave in ways that, until recently, we thought required human intelligence*" (High, 2017). It is a broad definition as Artificial Intelligence involves a program doing an analysis that normally would rely on the intelligence of a human. Research provided a wide variety of learning techniques that have the potential to transform many scientific and industrial fields. Benefits of Artificial Intelligence compared to the traditional data quarry's, are the reduction in Human Error, the availability of 24x7, the access to daily Applications (like Google) and, the ability to show quick insight for decisions (Korinek & Stiglitz, 2017). In 2017, Deloitte emphasised that the United States government could save billions annually with the implementation of Artificial Intelligence (Viechnicki & Eggers, 2017).

Machine Learning, a subset of Artificial Intelligence

Artificial intelligence comprehends various elements, one of the main elements being machine learning. Machine learning provides systems the ability to automatically learn and improve from experience without being explicitly programmed. The process of learning begins with observations of data in order to look for patterns in data and therefore make better decisions in the future based on the provided examples (Raghani, 2019). It has found to be indispensable in many fields, including computer science, engineering, mathematics, physics, neuroscience, and cognitive science.

Machine learning, as a promising technique, is under development and improvements follow from controlled application. Implementing novel techniques is a challenge in itself and comes with the necessity to be continuously critical on the results. For instance, machine learning techniques used by thousands of scientists to analyse data are unfortunately, capable of producing results that are misleading and often completely wrong (Allan, 2017).

The recent issue of the child allowance in the Netherlands is a clear example in which machine learning has been used to detect fraud, but the consequences of errors in the algorithm have not been sufficiently taken into account. The social impact of this result is that a specific group of parents were wrongly labelled as fraudsters and did not receive any allowance at all (*SyRI-wetgeving in strijd met het Europees Verdrag voor de Rechten voor de Mens*, n.d.).

Standard Business Reporting, the new standard

For analysis, data is needed. In the Netherlands, most organisations write and receive business reports. In the past, many of these business reports were produced manually and in variable data formats. Nowadays, some of these organisations have switched to "Standard Business Reporting" (SBR) (Lim & Perrin, 2014; Madden, 2011; Robb et al., 2016). Standard business reporting is used as a set of standards for the digital exchange of business reports. SBR allows organisations and their intermediaries to reduce reporting and administration work in the exchange of business information to authorities and banks (Bharosa et al., 2015, p. XVII–XXI). SBR follows specific rules with eXtensible Business Reporting Language (XBRL) at its centre. Following these carefully chosen rules, SBR provides highly structured data (Bharosa et al., 2015, p. 110). Furthermore, SBR is getting more accepted: in 2018 42,8 million messages were sent over SBR with a growth of 11.5% compared to 2017 (SBR Feiten En Cijfers 2018_0.Pdf, n.d.). Currently, SBR in the Netherlands is mainly used by large organisations such as the Tax Authorities, the Chamber of Commerce (KvK), the Central Bureau of Statistics (CBS), the Education Executive Agency of the Ministry of Education (DUO), Culture and Science (OCW) and various banks. The vision for the future is that growth is expected in the number of sectors that work with SBR in the coming years. In addition, this amount of structured data provides a lot of opportunities.

1.2 Societal relevance, from a managerial perspective

This section further elaborates on the societal relevance of this thesis. This is done by scanning both published and unpublished literature (Sekaran & Bougie, 2010, p. 18). The section finishes with a clear managerial knowledge gap.

Machine learning, the big potential

As stated previously, Machine learning is gradually being implemented in multiple domains, such as science, business and government (Lecun et al., 2015). In 2016 companies invested more than 26 billion US dollars in artificial intelligence (Bughin et al., 2017). For instance, in medicine machine learning is being used for clinical decision making (Kohli et al., 2017), such as determining the risk of a mass being benign or malignant (Ardila et al., 2019).

"*AI COULD POTENTIALLY CREATE $3.5 TRILLION TO $5.8 TRILLION IN ANNUAL VALUE IN THE GLOBAL ECONOMY*" (Chui et al., 2018, p. 17)

Many papers and books on artificial intelligence have been published in recent years. When searching "machine learning", countless hits come up. A quick analysis of the results shows that top-cited papers write about specific algorithms, such as random forest (Breiman, 2001)or support vector machines (Chang & Lin, 2011). Furthermore, several excellent textbooks are dedicated on how machine learning works and give detailed descriptions of available algorithms (Chang & Lin, 2011). Furthermore, Machine learning has turned out to be very useful at discovering intricate structures in high-dimensional data, finding patterns and making predictions. Therefore, machine learning is applicable in many domains (Lecun et al., 2015). Therefore, it can be stated that machine learning comprehends a broad scientific field with different domains and applications.

A report published by Deloitte stated that machine learning could help companies become more efficient (Viechnicki & Eggers, 2017). This is supported by Bughin et al, (2017), who surveyed numerous organizations on the use of Artificial Intelligence and concluded that in 44% of the cases Artificial Intelligence helped reduce cost. However, the downside is that in 56% of the cases, cost reduction was not achieved.

Consequently, many companies still do not use machine learning. A recent study emphasises that only 23% of businesses have adopted machine learning (Allan, 2017). A contributing factor to the tendency of companies being reluctant to implementing machine learning is because many of these projects are unsuccessful (Bughin et al., 2017; Press, 2019).

While the implementation of machine learning in businesses is gradually developing, the use of machine learning in governments is still running behind (Nagorny et al., 2017). This discrepancy is being accentuated by Batarseh & Yang (2018), who indicate that the use of machine learning in the governmental system of the United States could be extremely profitable.

Batarseh & Yang (2018) further explain the major problem why implementation of machine learning in the United States is stalled. This is because most files are in pdf and word, which are types of unstructured data.

Klievink et al. (2017) agree with Batarseh & Yang (2018) that the lack of structured data in the public sector is what obstructs the potential use of various forms of analytics. They further emphasize the critical need to inject more Artificial Intelligence in the public sector because this will eventually lead to great benefits in multiple domains (Batarseh & Yang, 2018). Consequently, this offers an opportunity for the stakeholders of SBR, since they have a large source of highly structured data (Bharosa et al., 2015, p. 110).

Klievink et al. (2017) & Adadi et al. (2015) stipulate that data activities often fail because the goals and objectives of the organisation and stakeholders are not taken into account. Furthermore, a review by Valle-Cruz et al. (2019), which included 75 papers on AI in the government, has found that at this stage in time AI is not completely understood in the public sector. Therefore, there is a demand on how to help governments understand artificial intelligence, the potential of Artificial intelligence and how to successfully implement AI systems such as machine learning.

Machines that act like humans, and its ethics

There is no talking about Artificial Intelligence without considering the ethical perspectives. Common ethical difficulties are how to maintain transparency or where responsibility lies. Therefore, realisation towards the need for integrating ethics into building machine learning projects is growing. Improper use of algorithms is potentially fundamental for ethical errors. Engineering is focused on building the best model for creating the best solution. If for example, the goal is to build a model with the best outcome in an engineering perspective, this may be in conflict with societal ethics.

In the previous section, the example of the scandal surrounding the Dutch governmental organisation "de Belastingdienst" was given. They did not foresee the ethical impact of their machine learning outcomes: not only were people wrongfully labelled, there were also no-fail saves preventing the prosecution of the accused. More importantly, the enormous impact on the accused was not foreseen. The Supreme Court in the Netherlands confirmed that the procedure was unethical (*SyRI-wetgeving in strijd met het Europees Verdrag voor de Rechten voor de Mens*, n.d.).

"***Gaps between the design and operation of algorithms and our understanding of their ethical implications can have severe consequences affecting individuals***"
(Mittelstadt et al., 2016, p. 1)

A potential contributing factor, as earlier stated by (Mittelstadt et al., 2016), is that the core problem when developing machine learning models, lies in the difference between the goal of the project and the outcome of the algorithm. This can have unintentional consequences with a possible negative impact on stakeholders (or society). Therefore, it is advised to take into account ethics when using algorithms or when building machine learning models (Mittelstadt et al., 2016).

1.2.1 Managerial knowledge gap

In conclusion, machine learning shows great potential in many applications. However, there is a need for structural guidance for implementing machine learning projects. Support is summarized in threefold. First, although in business, machine learning is gradually being attempted, projects are often not successful. Second, governmental organisations lack acquaintance with the possibilities of machine learning and how to structure data to implement this correctly. Third, machine learning projects in general, but especially those in a governmental project, have to meet high ethical standards. Therefore, this research concludes that there is a need for a method which helps organisations set up successful machine learning projects.

1.3 Academic relevance

In the previous section, the need for structural guidance implementing machine learning projects, has been detected. This section describes the literature review (Sekaran & Bougie, 2010, p. 18). The literature review focused on answering the following search question: Is there structural guidance for setting up machine learning projects, which includes machine learning specific aspects, organizational aspects and ethical aspects combined? And if not, what supporting literature is available that provides evidence to help answer this question. As machine learning has potential in a wide range of different applications, the academic literature review is focussed on its application in SBR context.

Systematic literature review

A systematic literature review will be done to determine whether there are any methods available that guide organisations on how to implement machine learning projects, as established in the search question. Furthermore, connecting terms will be explored. To do this a set of closely related keywords is used as search terms in the prominent database Scopus. All keywords were searched using quotation marks to retrieve precise results during the search. The search terms were: “Machine learning Projects”, “Machine learning project steps”, “Machine learning steps”, “Machine learning project framework", "Machine learning project method" and "Machine learning project strategies". The results of the searched terms are displayed in Table 1.

Table 1. Literature Search

Keyword	**Total number of hits**	**After screening titles, keywords and abstracts**
“Machine learning projects”	30 results	1
“Machine learning project steps”	0 results	0
“Machine learning steps”	31 results	0
“Machine learning project framework"	0 results	0
"Machine learning project method"	0 results	0
"Machine learning project strategies"	0 results	0

In summary, the results of the systematic literature review clearly show the absence of a method which combines all the factors that are needed at the moment of the search. However, much literature does elaborate on the separate fields. These are Machine learning, Organizational aspects and ethical aspects. In order to find papers on the subject described previously in this chapter, a second search has been done. Search terms include Machine learning, operational aspects and ethical aspects.

Knowledge Discovery in Databases and Machine Learning

When searching the academic sources, it becomes evident that there are certain frameworks available. However, these present frameworks focus on guiding the process of knowledge discovery from data (Fayyad et al., 1996; Han et al., 2011, p. 6). Subsequently, Qiang & Xindong (2006) emphasise that there is a need for a theoretical framework that helps organisations use different Machine learning techniques, such as classification or clustering. Both Fayyad et al. (1996) & Han et al. (2011, pp. 6-8) provide a method on how to discover knowledge from datasets: technical steps are described that give guidance in a partial machine learning environment (Fayyad et al., 1996; Han et al., 2011, pp.6-8). However, both methods do not elaborate on the ethical and organisational context. Furthermore, no algorithm selection method is provided.

Organisational view
As previously described in the Societal Relevance section, literature shows a relevant correlation between discrepancy of machine learning project goal and strategy of the organisation, and unsuccessful implementation of machine learning projects (Adadi et al., 2015; Klievink et al., 2017). A prominent work on evaluating strategy is the Strategy Map of Kaplan & Norton (2004). In this method, the strategy is analysed from four different perspectives: mission, beneficiaries, method and learning and growth.

Klievink et al. (2017) & Adadi et al. (2015) suggest that to properly apply data analytics, one of the most important factors is that the aim must be to ensure data activities to be in line with the goals and objectives of the organisation and with stakeholders' needs.

Ethical view
Scientific literature is found that specifically focusses on the ethical properties of machine learning. Qiang & Xindong (2006) explain the demand for guidance for using data mining and keeping ethical issues intact such as privacy and data integrity. They anticipate that "data mining will become a derogatory term to the general public." Mittelstadt et al. (2016) implicate that analysing the ethical concerns produced by algorithms, are an important subject. An editorial Gibney (2020) wrote for Nature displayed that ethic controversies should be implemented in the design process of machine learning models.

A framework focussing on the ethics in information systems providing guidelines for guaranteeing ethical designs is a method called Value Sensitive Design (VSD) (Friedman et al., 2013). Umbrello & Bellis (2018) add to the work and stipulate the need for ethics when building machine learning models. However, Umbrello & Bellis (2018) do question if Value Sensitive Design is the most efficient way to ensure ethics. Another framework that assesses the ethical impact, is Ethical Impact Analysis, which any project or program involving information technologies (Wright, 2011). Raab (2020) continues on the work of (Wright, 2011) and discusses the use of the ethical impact analyses on algorithms and what the current ethical frameworks are.

Standard Business Reporting (SBR)
The literature regarding SBR is foremost about how SBR can be used and what its pros and cons are (Lim & Perrin, 2014; Madden, 2011; Robb et al., 2016). Madden (2011) expresses that when SBR was introduced in Australia, the results were significantly positive regarding cost reduction. However, implementation and use of SBR is less than the government originally expected, indicating that there is room for improvement of the SBR system (Lim & Perrin, 2014). Furthermore, Lim & Perrin (2014) conclude that the opinions of the reporting businesses were not properly taken into account. Therefore, the perspective of the reporting businesses was missing. A study by (Robb et al., 2016), shows the assessed perceptions of Australian business stakeholders in relation to the benefits of the Australian SBR instantiation for financial reporting.

Bharosa et al. (2015, p. XVII–XXI) elaborate on SBR and its use in the Netherlands. They provide insights and best practices relevant for readers that do not use SBR and are looking for cost-effective information chains. Furthermore, the book provides the reader with an overview in the turbulent world of SBR from two perspectives: On the one hand, it provides insights into the creation of an initiative such as SBR and describes the challenges that actors face when striving to redesign and improve information exchange and processing in information chains. In this respect, we can see SBR as a challenge for information chains. On the other hand, this book provides concrete descriptions of the SBR solution components (building blocks) realised in the Netherlands" (Bharosa et al., 2015, p. XIX).

eXtensible Business Reporting Language (XBRL)
Bharosa et al. (2015, p. 146) argue that XBRL will be responsible for a transformation in the business reporting sector and information chains. XBRL is expected to grow in the following year; however, extensive research towards prerequisite conditions and distinct environmental settings for correct application is desired (Bharosa et al., 2015, p. 146). The structured data of SBR provides a clear research field for obtaining this knowledge. Roohani et al. (2010) reviewed all the scientific publications on XBRL between 1998 and 2008. The reviewed publications include trade, practitioner and academic journals to identify trends and patterns, milestones, and organisations actively contributing to this development (Roohani et al., 2010). Roohani et al. (2010) emphasises that "*in the future academic publications on XBRL may include a wide variety of areas such as corporate governance, data analysis, integration of financial information with non-financial information, behavioural impacts, social responsibility, and financial reporting and data warehousing."* These prospects support further research on data analytics by machine learning on XBRL data, a main part of the SBR context.

1.3.1 Academic knowledge gap

In conclusion, a thorough literature survey clearly results in the outcome that there is no method for setting up machine learning projects that consider the technical steps, organisational aspects and ethical aspects combined.

Although methods on how to retract information from datasets or separate frameworks on how to implement ethics in designs are present, these methods do not take into consideration the technical steps that are needed when setting up a machine learning project, or the organisational structure of the entity for which the machine learning project is intended. Therefore, this research concludes that there is a need for the development of a method including technical machine learning steps, organisational strategy and ethical aspects. Furthermore, no literature can be found providing insight in the use of machine learning in SBR context. By developing and testing a method in this environment, the research provides insights in whether machine learning has potential in SBR context.

1.4 Research objective

The previous sections elaborated on the problem from both managerial and scientific perspective. These factors combined form the base for defining the problem and clarifying the research objectives of this thesis (Sekaran & Bougie, 2010, p. 18).

Problem statement
This chapter elaborates on the potential of machine learning in different areas. However, besides the fact that it has potential, there are also many problems surrounding machine learning.

For example, ethical challenges easily arise in machine learning. Furthermore, friction with organisation fit and misunderstanding c.q. lack of knowledge of Machine learning – how to use it and what it can offer – are a selection of factors complicating projects.

As stated previously, there are many factors that influence machine learning projects, and there are different ways to view these factors. It is not possible to investigate every factor, therefore this research focusses on three factors: technical factors, organisational context and ethical context. Support for choosing these three factors is as follows: technical factors are of course essential in building a machine learning model, and these will be defined via Knowledge Discovery in Databases. Second, for successful implementation of projects within an organisation, taking into account organisational factors contribute to this process. Last, the ethical factors are included because taking into account ethical considerations from the beginning of the process ensures feasible outcomes that can actually be implemented in human society. Besides the aforementioned three factors, numerous other factors influence and contribute to machine learning projects. However, these are outside the scope of this thesis, to ensure proper time-management and to define a clear framework for this thesis.

When searching scientific databases, no method that combines organisational, ethical and technical factors of machine learning is found. An explanation for this might be that focus lies on one of the factors, for example, either highly technical or organisational factors. As the three subjects all require expert knowledge, this thesis explores the balance in this field when developing a method that is understandable for both manager and technician. This will probably be at the expense of the three explored aspects since expert knowledge is not easily transferrable.

Next to the numerous factors that influence machine learning projects, there are also many different situations or contexts where machine learning projects are performed. SBR provides an interesting context as no scientific literature can be found providing the understanding of the potential of machine learning in this context. Analysing all the different situations or projects or contexts in which machine learning is applicable, is not realistic in the timeframe of this thesis. Therefore, it is decided to scope on the context of SBR, thus making the research objective to analyse the factors in an SBR context.

Consequently, this thesis does not has the illusion to create a fully-functional method that can be implemented in general. Rather the goal is to create a comprehensible working prototype method for a specific context and to provide insight into the possibilities and drawbacks of combining these three factors in one method in SBR context. Furthermore, it researches the translation of the theoretical potential of machine learning in SBR context.

Research objective
In the previous paragraphs, the argumentation why this thesis focuses on machine learning in SBR context and the rationale behind the choice of the three factors has been given. Consequently, the scope of the thesis is more comprehensible and thus pragmatic. Subsequently, the following research objective is formulated:

“to develop a method including technical machine learning steps, organisational strategy and ethical aspects in order to guide stakeholders in systematically setting up machine learning projects in SBR context”

Research Question
This thesis will research how to implement machine learning, with respect to organisational and ethical steps, into one method. To do this, the specified fields will be explored separately before a method combining these elements will be created. The created method will be analysed and tested in experimental settings.

Based on the previous paragraphs, the following research question is formulated:

Main research question: How can technical, organisational and ethical aspects be combined into a method that supports stakeholders to systematically set up machine learning projects in SBR context?

In order to answer this main research question, it is divided into four sub-questions:
1. What are the relevant machine learning, organisational and ethical factors for setting up a machine learning project in SBR context?
2. How can the identified factors be combined into a method?
3. Does the designed method provide the guidelines needed for systematically setting up machine learning projects?
4. Does the designed method provide clear insight in the value that machine learning might provide?
5. What is the potential of machine learning in SBR context?

1.5 Conclusion
This chapter elaborated upon the potential of machine learning and illustrated the problem of ineffective use of machine learning in organisations. Three factors influencing machine learning projects were identified: technical, organisational and ethical factors. From a managerial and scientific perspective, a gap of a structured method, including these factors, on how to set up machine learning projects, was identified. This defined the scope of this thesis: creating a comprehensible working prototype method for a specific context and providing insight into the possibilities and drawbacks of combining these three factors in one method in SBR context. Furthermore, it researches the translation of the theoretical potential of machine learning in SBR context. The research question is “How can technical, organisational and ethical aspects be combined into a method that supports stakeholders to systematically set up machine learning projects in SBR context?”. This research question will be subject to further detailed discussion and arrangements divided into clear steps of this thesis. Essential elements for answering this research question, while giving substance to the academic knowledge gap, are the quest of the added value of machine learning combined with the analysis of the ethical implications and the organisational factors. A further specification and elaboration are described in the chapter Research Approach.

Chapter 2. Research Approach

A decent research study always starts with the formulation of a proper research methodology (Sekaran & Bougie, 2010, p. 220). This chapter elaborates on this thesis' methodology based on standards advised by Hevner, March, Park, & Ram (2004), Sekaran & Bougie (2010) & Peffers et al (2014). Following this method, it is required to follow a proven structured research process. This chapter will elaborate on the research approach. First (2.1), the chosen research methodology is described, and each step of the research will be clarified. Second (2.2), a short description on the communication on GitLab is given. Third (2.3), several data collection methods are illustrated. Last (2.4), an overview of the study and its results will be visualised.

2.1 Design Science Research Methodology

The previous chapter elaborates on the research objective of this thesis, **"to develop a method including technical machine learning steps, organisational strategy and ethical aspects in order to guide stakeholders in systematically setting up machine learning projects in SBR context"**. This research objective, designing a method, or in other words, an artifact, perfectly fit within the vision of Hevner et al. (2004) and Peffers et al. (2014). Hevner et al. (2004) created guidelines on how to conduct proper design science in information systems. These guidelines are fundamental for a validated research methodology to build those methods in information systems. This methodology is called Design Science Research Methodology (DSRM) (Peffers et al., 2014). DSRM does not only provide a structured process but also contributes to a uniform research methodology. Consequently, outcomes or strategies created with DSRM are universally acknowledged.

This thesis aims to create a scientific method to guide SBR-stakeholders to successful machine learning projects. The creation of such a method fits with the characteristics of DSRM. Therefore, DSRM is used as a methodology to conduct this research.

The DSRM contains six steps: problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, and communication. In the following paragraphs, each phase of Design Science concerning this thesis are explained.

2.1.1 Problem identification & motivation

The main objective of this first step of the DSRM is the development of a sound method that can effectively provide a solution. Furthermore, emphasising the potential added value of the method supports the acceptance of the research results. Additionally, it helps to grasp the reasoning of the researcher. By describing the societal relevance and doing a preliminary literature review, Chapter 1 focuses on the identification of the problem and the motivation why there is a need for the proposed method.

2.1.2 Define objectives for a solution

Following problem definition in step 1, this step focusses on shaping the objectives for problem solution. Aim of this thesis is to create a new method, which requires the formulation of qualitative design objectives. These research objectives are clarified in Chapter 1. They include identifying machine learning steps based on structured data derived from SBR-stakeholders, ethical steps focused on ethics in machine learning, and organisational steps derived from machine learning as a strategy. Thus, there is a strong need for knowledge that provides a clear outline of what is possible and feasible within the context of this research. This is done by the exploration of academic databases and examining prior research, which is done in Chapter 3. This eventually provides support for the design process and help to legitimise the research. Specific design objectives for the method will be further elaborated on in Chapter 4.

2.1.3 Design & development

The third activity of the design science for information systems encompasses the design and development of the actual artifact. In this case, the methodology that guides setting up a successful machine learning project. The design of the artifact in this thesis is divided into two phases.
First, the input for the first design phase is gathered during the state-of-the-art literature review executed in Chapter 3 and the experiment in Chapter 4. Furthermore, Chapter 4 provides additional content on this thesis' specific objectives for strategy and machine learning steps. The gathered input provides the desired functionally and architecture of the artifact and therefore can be transformed into the first artifact, which is elaborated on in Chapter 5. After the first design phase, the artifact is evaluated by experts during six semi-structured interviews, as described in Chapter 6. These interviews assess the current design and provide additional input for the second design phase. This second design phase, based on the output of these interviews, is described in Chapter 7.

2.1.4 Demonstration

The fourth step in the DSRM is the demonstration of the created method. In this thesis, the demonstration is divided into three main sections. First, the demonstration explains how to implement the designed method. Second, the method created in this thesis, is demonstrated in an experiment on the relevant data of the organisation WSW. In this experiment, step by step guidance through the overall method is provided. The experiment shows the ability of the method to provide guidance and provides a real-life example giving insight to other practitioners how the method can be used. Third, the demonstration phase includes semi-structured interviews with experts. The experts are requested to evaluate the functionality of the designed method. This will provide partial content for answering sub-question 3. Chapter 8 will comprehend the demonstration.

2.1.5 Evaluation

The second to last activity of DSRM is the evaluation of the methodology. The evaluation can be found in Chapter 9. The assessment is essential to observe and to measure if and to what level, the designed method supports the problem solution that is defined in the first step of the design science methodology and answering the research questions. In this research, this step is done in twofold. First, the results of the experiment with WSW are. The results of the experiment provide insight if the method was successful in delivering a successful machine learning project. The second part of the evaluation is based on the results from the semi-structured interviews. In this part, the interviewees are requested to provide their expert opinion on if the method supports the informed decision-making and therewith setting up successful machine learning project. Furthermore, the interviewees elaborate on whether the method is understandable for the internal stakeholders (managers and technical experts). The combination of both the experiment and the evaluation of relevant experts in the field, provide a superb level of evaluation.

2.1.6 Communication

The sixth and final step of DSRM focusses on communication. It is essential to communicate the detected problem area (described in Chapter 1), to understand how much potential value a possible solution has, in this thesis, the created method. Furthermore, communication is critical for transparency. Explaining the design of the method, the contribution to existing methods and correct application of the method enhances validity and effectiveness. This is necessary to create support by other researches for the acceptance of the method. This research is communicated in twofold, through sharing this thesis on the TU-Delft repository and via GitLab. This is elaborated on in Chapter 10.

2.2 Communication via GitLab

This section elaborates further on the communication of the research on GitLab. In this research communication is an ongoing process; the iterations of the method will be uploaded. As described in the DSRM, creating a GitLab is part of the communication.

GitLab is an online platform where developers and researchers can share their work. On GitLab, users can view, check, reproduce and provide additional input to the uploaded work. This creates a community on a specific topic. In this case, the method and experiments are uploaded and described. Uploading the method not only contributes to the transparency of the research, but also to substantiate the research. The added value of GitLab is, by reading on the method, the community may be inspired to use and test the method. When users subsequently share their findings, experiments and data, this will lead to further development of the method, including the generalisation of the method.

2.3 Data Collection

As explained in the previous section, the research uses the DSRM of Peffers et al. (2014). This section elaborates specifically on how the data will be gathered and which research methods will be used. There are several data collection methods. Sekaran and Bougie (2010) explain two main forms. Primary data and secondary data. Where primary data refers to information obtained first-hand such as interviews and experiments, secondary data applies to data gathered from sources which already exist previous to the research (Sekaran & Bougie, 2010, p. 220). This research starts with collecting secondary data during a state-of-the-art literature review. Second, data is collected by experimenting, which methodology will be further explained in this section. Third, data is collected with interviews using the methodology as will be described in Chapter 6. Combining these three forms of data collection provide a proper basis for this research.

2.3.1 Experiments

As part of the research, two experiments will be completed. The first experiment, which is executed by the researcher in a controlled environment, is done with SBR data from DUO. This experiment provides data that has high internal validity, as it is done in a controlled environment. However, the difficulty and thus the risk of experimenting in a controlled environment, is that the external validity or generalisability relatively low is. Therefore, the second experiment, which is done by WSW, relates more to a field experience (Sekaran & Bougie, 2010, p. 233), as the designed method is tested by the users in a not controlled setting. This supports the independent external validity of the method. The combination of both types of experiments provides a scientifically validated method, with both internal and external validity.

2.3.2 Interviews

After the literature survey and the first experiment, this part of the research focusses on conducting interviews. In total, six semi-structured interviews will be held. There is chosen explicitly for semi-structured interviews as this type of interviews provides in-depth qualitative data. The interviews use predefined questions but additionally allow the interviewees to address its opinion. This produces additional insights during the interview (Sekaran & Bougie, 2010, p. 142). Besides the gathering of qualitative data, the interviews will also gather quantitative data. The quantitative data is gathered by asking the interviewee to evaluate each step of the designed method and the functionalities of the total method.

2.4 Research visualisation

The section provides a simplified overview of the research methodology, visualised in Figure 2

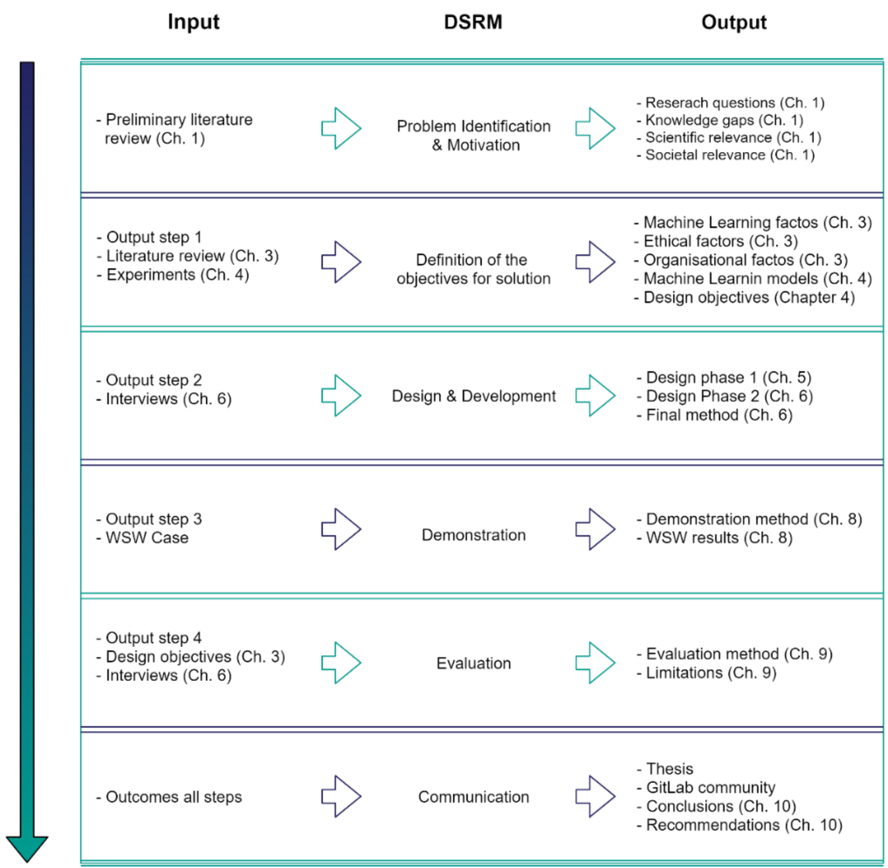

Figure 2. Research Methodology overview

2.5 Conclusion

This chapter describes the use of Design Science as a scientific research methodology. The DSRM contains six steps: problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, and communication which provide the structure in this thesis. Furthermore, this chapter clarifies the procedure collecting the relevant data, by an extensive literature review, two experiments and six interviews. This is proven to be sufficient to provide a scientific and proper research method.

Chapter 3. Literature review

Having identified the problem area and research approach in the previous chapters, this chapter elaborates on conceptualising of the theoretical base as input for setting up the method. The chapter consists of an extensive literature review on the specified concepts in order to identify the design objectives for the method. This chapter partially contributes to answering sub-question 1, by enumerating the relevant concepts related to creating machine learning projects. The ideas identified in this section will be the basis for the input of the experiment and the first design phase of the method.

This chapter starts with describing the concept of structured data in Section 3.1. This is followed by literature on Knowledge Discovery in Databases and its connection to machine learning, Section 3.2. Then, the definition of machine learning, a selection of algorithms of machine learning and a summary on how to interpret the results of the machine learning models, Section 3.3. The last part of the literature review focusses on relevant organisational (3.4) and ethical frameworks (3.5).

3.1 Data

Data offers room for many interpretations and is far too often seen as a homogenous concept (Janssen & Kuk, 2016). Data can be defined as symbols, the products of observation (Janssen & Kuk, 2016). In order to enable interpretation, data should first be placed in the context. Janssen & Kuk (2016) present four main data characteristics between the degree of structuredness and openness in the data. Bharosa et al. (2015, p. 101) continuous on one form of data characteristic, data structure and describe three types: structured, non-structured and semi-structured data. Structured data is organised according to a specific structure and usually resides in a database. In addition, structured data can be interpreted by computers and can be searched by type of data. Compared to structured data, unstructured data does not have an identifiable structure. Somewhere between structured and unstructured data resides semi-structured data. Some data is hierarchical and has recursive structures and can, therefore, be regarded as structured or semi-structured data (Bharosa et al., 2015, p. 101).

This thesis focuses on data of the Standard Business Reporting (SBR), a specific format of structured data. Following the specifications set by the SBR chain provide a high standard of data quality. The data specifications for SBR chains are described by Bharosa et al. (2015, p. 110):

- "The use of XBRL taxonomies within the SBR Programme
- Specific requirements for SBR taxonomies, including organisational needs, such as compliance with the Netherlands Taxonomy Architecture (NTA).
- The taxonomy development process applied for the Netherlands Taxonomy (NT). This process will be analysed in various stages, from the requirements analysis up through the publication stage.
- Relevant developments in XBRL that could provide new opportunities for SBR."

XBRL, as one of these specifications, stands for eXtensible Business Reporting Language (Bharosa et al., 2015, p. 110) and refers to a framework for the business reporting domain. XBRL enables publication, exchange and processing of business reports over the internet. Furthermore, XBRL is also a taxonomy for a collection of controlled context and how they connect (Bharosa et al., 2015, p. 113).

The definition of data in this thesis follows the characteristics described above. However, it does not support the exact data specifics of SBR chains as the data is transformed into CSV data. Therefore, it "loses" some of its structure but remains "highly" structured data.

3.2 Retracting knowledge from data, the Knowledge Discovery in Database process

Although the data provided by SBR is highly structured, data alone is not enough. Han et al. (2011, p. 5) describe that data can also be seen as "golden nuggets". But to be able to use data as "golden nuggets", a predefined process should be followed. This process, Knowledge Discovery in Databases (KDD), is defined as: "*discovering interesting patterns and knowledge from large amounts of data*" (Han et al., 2011, p. 8). This KDD process provides an interactive and iterative process, including the following seven steps (Han et al., 2011, pp. 6–8):

1. Data cleaning
    a. to remove noise and inconsistent data
2. Data integration
    a. where multiple data sources may be combined
3. Data selection
    a. where data relevant to the analysis task are retrieved from the database
4. Data transformation
    a. where data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operations
5. Data mining
    a. an essential process where intelligent methods are applied to extract data patterns
6. Pattern evaluation
    a. to identify the fascinating patterns representing knowledge based on interestingness measures
7. Knowledge presentation
    a. where it is essential how the results are represented and accessible.

Another prominent description of KDD is done by Fayyad et al. (1996). These authors define the nontrivial process of identifying valid, novel, potentially useful, and ultimately and understandable patterns in data. The process developed by Fayyad et al. (1996) is visualised in Figure 3.

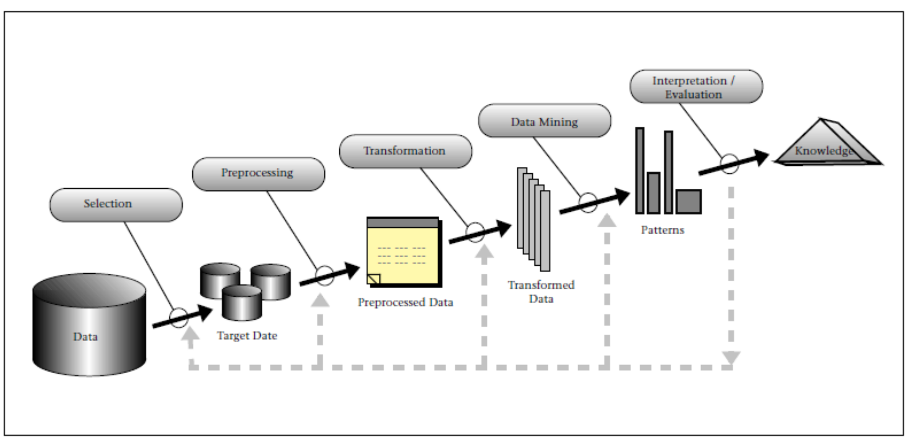

Figure 3. KDD process (Fayyad et al., 1996)

Slightly different from Han et al. (2011, pp. 6-8), this KDD process involving nine steps with decisions made by the user:

1. The first step is to develop an understanding of the application domain, the relevant prior knowledge and finally to define/identify the goal of the knowledge discovery from the stakeholder's viewpoint.
2. The second step is to select the data set on which the discovery is performed
3. The third step focuses on the data preparation in which the data needs to be cleaned and pre-processed. Actions involve, apart from cleaning the data, the removal of noise and the proposal of strategies for handling missing data.
4. The fourth step is to understand the data and to find relevant features to represent the data
5. The fifth step is to align the goal of the project to a particular data mining method, such as machine learning.
6. The sixth step of the KDD is to choose a unique set of data mining algorithms. Furthermore, one should select the relevant parameters for the proposed goal. So, it might be possible that it is more important for the user to understand the model rather than its predictive capabilities.
7. The seventh step is focused on the quest for the relevant patterns, or in other words using the algorithms to find relevant information.
8. The eight-step is to analyse and understand the results of the mined patterns
9. The last step focusses on how the additional knowledge found, can be used in practice. This includes how to use the model, communicating the model, documenting the model/knowledge and reporting the result to interested parties.

Although slightly different in the KDD process steps, both Fayyad et al. (1996) & Han et al. (2011, p. 5) agree that the possibilities of gathering knowledge from data are enormous. Furthermore, they both elaborate on the process of knowledge discovery and also highlight the importance of the data mining step, as reviewed in the following section.

3.2.1 Data-mining, an important step of the KDD

Fayyad et al. (1996, p.41) define Data mining as "*a step in the KDD process that consists of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns (or models) over the data.*". This definition enables many options for implementing data mining. Han et al. (2011, p. 33) further elaborates on the concept of data mining and defines it as "*is the process of discovering interesting patterns from massive amounts of data.*". Herewith, specifically stipulating the importance of data mining but also the variability.

Figure 4 visualises the factors related to data mining, showing datamining encompasses many different ways and techniques to analyse data.

The next paragraph provides some additional factors of importance, and therefore strictly necessary, in the process of data mining (Fayyad et al., 1996; Han et al., 2011, p. 85).

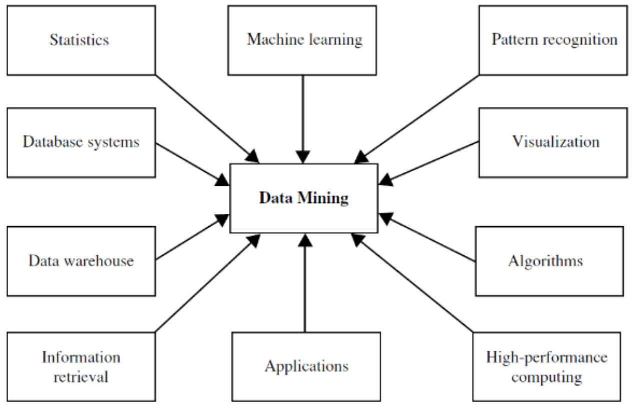

Figure 4 - Data mining (Han et al., 2011 p.28)

Overfitting: when a model is built for one particular dataset using a limited set of data, the model might not be able to detect general patterns or explore the noise in the dataset, resulting in poor performance in the test dataset and for use on other/new datasets. Therefore, testing on a test dataset also investigates the level of overfitting.

Understandability of patterns: In many applications of data mining, it is essential to make the knowledge that is discovered, understandable so that it can be analysed by and communicated to a broad audience.

3.2.2 Data-mining and Machine learning

Data mining itself strictly focusses on data patterns without predictive capacity. A next step is to make use of this dataset as a training data set to make predictions by training the program. Therefore, machine learning is discussed as a promising and potential technique to achieve this next step. When searching for what encompasses machine learning, one finds many similarities with the functions of data mining. Therefore, one tends to interpret them wrongfully as equal. Maybe the main difference between data mining and machine learning is stipulated when reading the definition of machine learning by Han et al. (2011, p. 24) who highlights that machine learning focusses on how “to automatically learn and recognise complex patterns and make intelligent decisions based on data.”. Whereas data mining focuses more on how to manually program computers, machine learning focuses on the questions of how to make computers program themselves. The next section reviews the topic of machine learning in more detail.

3.3 Machine Learning

As elaborated in the previous section, the use of Knowledge Discovery in Databases (KDD) to extract knowledge in databases is essential. Also is discussed how data mining and machine learning are connected. From now on, this thesis further uses machine learning as it focusses on the field of study that provides computers with the ability to learn without being explicitly programmed (Samuel, 1959). Therefore, in this section, only machine learning will be reviewed in more detail.

3.3.1.1 Machine learning, a definition
Literature shows that machine learning systems can improve their performance due to gaining more experience for a particular use-case (based on a sound set of data). It can analyse data to eventually result in added knowledge, new novel patterns and models using a specific set of techniques. These techniques are referred to as machine learning (Bishop, 2006). The literature review indicates that machine learning is connected both in artificial intelligence and data mining. Another more recent definition implies that machine learning tools endow a program with the ability to learn and to improve its performance over time (Shalev-Shwartz & Ben-David, 2014). In this thesis, the formulation by Han et al. (2011, p. 24) is used defining machine learning as“ *to investigates how computers can learn (or improve their performance) based on data*”.

3.3.2 Types of machine learning
Machine learning can be subdivided into three main learning techniques: supervised learning, unsupervised learning and reinforcement learning. There is also a fourth technique which is a combination of supervised and unsupervised learning, semi-supervised learning (Bishop, 2006). The next section will further elaborate on each technique. However, this thesis focusses strictly on supervised and unsupervised machine learning, and its related techniques showed in Figure 5.

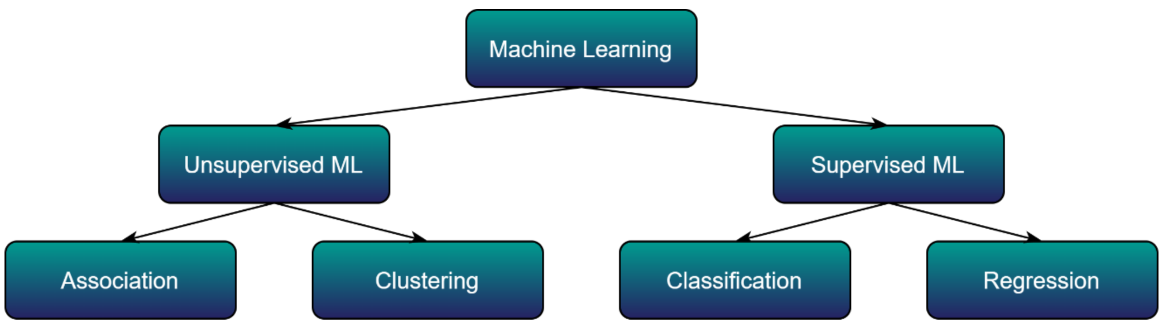

Figure 5. Machine Learning techniques

3.3.2.1 Supervised Learning
Supervised learning is a form of machine learning where both input and output data are provided (Fayyad et al., 1996; Han et al., 2011, p. 24). With supervised machine learning the algorithm is employed to determine a function from the input variable (x) and the output variable (y) (Bishop, 2006, p. 3). It is therefore creating the function y = f(x). The produced function (f) can analyse a new input variable (x) and estimate the output variable (Y) (Fayyad et al., 1996; Han et al., 2011, p. 327). It is called supervised learning because the process of an algorithm, learning from the training dataset, can be seen as a teacher supervising the learning process. Supervised learning uses classification algorithms and regression techniques to develop predictive models. Where classification predicts categorical labels, regression models predicts numerical labels. The term prediction refers to both numeric prediction and class label prediction. (Fayyad et al., 1996). Next section will review these techniques in more detail.

3.3.2.1.1 Classification
Classification can be defined as a form of data analysis that extracts models describing essential data classes (Han et al., 2011). Such models are called classifiers. Classifiers predict the categorical class labels which, as explained for supervised learning, means that a classification algorithm uses input

data (x) to learn the function (f) to estimate the categorical output value (Y) output (Bishop, 2006, p. 179). For example, one can build a classification model to categorise bank loan applications as either safe or risky. The classification has numerous applications, such as fraud detection (*SyRI-wetgeving in strijd met het Europees Verdrag voor de Rechten voor de Mens*, 2020), performance prediction, and medical diagnosis (Han et al., 2011, p. 327).

3.3.2.1.2 Regression

Shortly discussed in the section of supervised learning, regression is a method used to predict a numeric value (Han et al., 2011, p. 19). In the regression method, an algorithm is used to predict continuous or discrete quantity output. Where the process is similar as in the classification, where algorithms use input data (x) to learn the function (f) to estimate the categorical output value (y). Output, (y) is now a numeric value (Bishop, 2006, p. 137). An example is the prediction of the price of houses.

3.3.2.2 Unsupervised learning

The process of unsupervised learning is focused on data that is not class labelled. This encompasses cluster analysis or association analysis. As said before, the learning process is unsupervised since the input examples are not labelled. The goal of unsupervised learning is to draw interferences and discover hidden structures within the data (Bishop, 2006, p. 3). With this analysis, it is possible to create unseen semantics. However, the analysis needs to have a proper understanding of the data to understand relevant results.

3.3.2.2.1 Clustering

As mentioned in the previous section, unsupervised learning focusses on data that is not class labelled (Han et al., 2011, p. 19). Clustering comes in handy at the start of projects where in many cases no class labels are available. The cluster analyses the data and generates clusters or class labels for specific groups of data. The data in the groups or clusters have high similarity to each other in comparison with the data in different clusters (Bishop, 2006, p. 424).

3.3.2.2.2 Association

Association learning is a machine learning technique searching for important connections within variables or features of a data set. The association rule involves a single attribute that repeats itself (Han et al., 2011, p. 17). With this generated association rule, new searches can be conducted, and associations can be made.

3.3.2.3 Semi-supervised machine learning

Semi-supervised machine learning uses aspects from both supervised and unsupervised machine learning (Han et al., 2011, p. 432) as it uses both structured and unstructured data. Semi-supervised machine learning algorithms are typically used when there is limited structured data and a lot of unstructured data available. Furthermore, modifying the unstructured data into structured data requires valuable resources. A precondition for using semi-supervised learning is that both the structured and unstructured data is relevant for the same objective. The objective of semi-supervised learning is to understand how combining labelled and unlabelled data may change the learning behaviour (Zhu & Goldberg, 2009).

3.3.2.4 Reinforcement machine learning algorithms

Reinforcement Learning is a type of machine learning technique that uses an interactive environment where systems try-out new actions to examine if these actions are sufficient to reach the predefined goal. An example of reinforcement learning is credit assignment, where the learner explores new actions and analyses if these actions are worth taking. As the actions provide a reward, credit, the system will use the actions that are known for providing a high reward to reach the goal and earn the highest reward (Bishop, 2006, p. 3).

3.3.3 Machine learning Algorithms

There is an immense number of algorithms available (Fayyad et al., 1996). Therefore, it is hard to find out which algorithm fits were. This thesis delves into a few popular algorithms as is not possible to try each potentially relevant algorithm in the selected time-frame this thesis. These algorithms are derived from and mentioned by (Fayyad et al., 1996; Han et al., 2011; Wu et al., 2008). Table 2 shows the selected algorithms.

Table 2. Algorithms

Algorithms:	**ML-goal:**	**Classification**	**Regression**	**Clustering**	**Association**
Logistic Regression		X			
Linear Regression			X		
Naive Bayes		X			
k-Nearest Neighbors		X	X		
Decision Trees		X	X		
Support Vector Machines		X	X		
Artificial Neural Network			X		
The k-means algorithm				X	
The Apriori algorithm					X

3.3.3.1 Naïve Bayes

The Naïve Bayes classifier is a machine learning model that is one of the simpler Bayesian networks, models based on Bayes theorem (Han et al., 2011, p. 394). Bayes theorem investigates the probability of A happening, provided that B has occurred, shown in Figure 6. In the example, B is the evidence, and A is the hypothesis (Bishop, 2006, p. 380). The term Naïve is used because the classifier assumes that there are no dependencies between attributes (Bishop, 2006, p. 380). Naïve Bayes algorithms are used for classification.

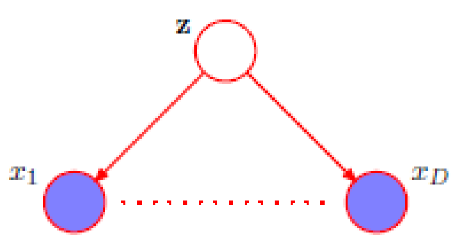

Figure 6. Naïve Bayes (Bishop, 2006, p. 394)

3.3.3.2 K-nearest-neighbor

The k-nearest-neighbor classifier is an algorithm used for pattern recognition working via learning by analogy. Data is not structured and is compiled as a tuple. The classifier functions via comparison of a test tuple with training tuples (Han et al., 2011, p. 423). The training tuples are produced by n-attributes which present a point in a n-dimensional space. When the model is presented with an unknown tuple, the model searches the pattern space for a tuple that is most similar or, in other words, is its nearest neighbour.

3.3.3.3 Decision Trees

The Decision Trees is a flowchart-like tree structure for classification and uses univariate splits, as shown in Figure 7 (Han et al., 2011, p. 330). The tree is built up with nodes. Each internal node denotes a test on an attribute; each branch represents an outcome of the test; and each leaf node holds a class label. The highest node in a tree is the root node (Han et al., 2011, p. 330). Construction of a decision tree does not require any knowledge of domains or parameter setting. Therefore, it is relatively easy to use and understand, resulting in being one of the most common learning techniques that are used (Han et al., 2011, p. 330).

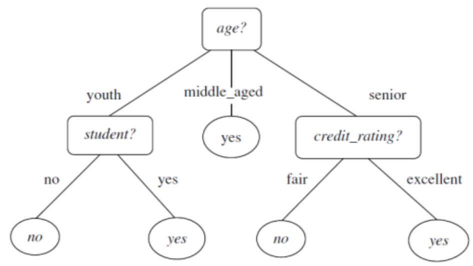

Figure 7. Decision Tree (Han et al., 2011, p. 311)

3.3.3.4 Artificial Neural Networks

Neural networks were originally researched by psychologists and neurobiologists looking for computational analogues of neurons (Han et al., 2011, p. 398). The Artificial Neural networks are algorithms that can solve complex problems via their associating and self-organising capabilities. Where Decision trees and Naïve Bayes are good to interpret, because of these capabilities, Artificial Neural Networks have poor interpretability. However, Neural Networks do provide the ability for continuous-valued inputs and outputs, where many decision tree algorithms are sensitive for. Furthermore, Neural Networks process noisy data well. A Neural network can be described as a learner that provides connections between input/output units, as shown in Figure 8. These connections have corresponding weight and learning the model adjusting the weights to find a proper solution. This provides the algorithm to predict the correct class label of the input variables.

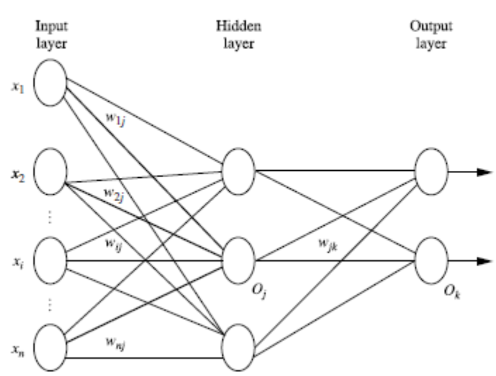

Figure 8. Neural Networks (Han et al., 2011, p. 399)

3.3.3.5 K-means Clustering

The K-Means algorithm is a clustering algorithm that divides data into groups. It is a form of unsupervised machine learning. The algorithm searches the dataset to define a centroid for each cluster. Corresponding data points are found in a multidimensional space. The clusters comprehend a group of data points whose inter-point distances are small compared with the distances to points outside of the cluster (Bishop, 2006, p. 424).

3.3.3.6 Support Vector Machine

Within the category of supervised machine learning techniques, Support Vector Machines (SVM) is a widely used classifier. The SVM is an algorithm that uses labelled training data and original data. The original data is then transformed into a higher dimension via nonlinear mapping, where the SVM algorithm searches for the linear optimal separating hyperplane to split the data points into corresponding labels. The SVM is used for both linear and nonlinear data, and an example of an SVM is shown in Figure 9. (Han et al., 2011, p. 409)

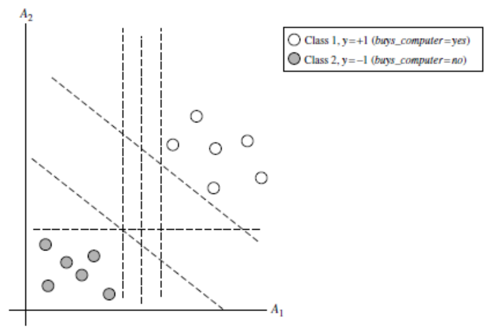

Figure 9. SVM (Han et al., 2011, p. 409)

3.3.3.7 Logistic and Linear Regression

Logistic regression is part of the field of statistical analysis. The basic principle of the logistic regression is finding the probability of a certain class. Logistic and linear regression models are not the most complex models. On the contrary, they provide a clear and understandable model. The algorithms use the input attributes and combine them linearly to predict the output value. Where the linear value is numeric (Bishop, 2006, p. 138), the logistic output value is binary (Bishop, 2006, p. 205).

3.3.3.8 Apriori Algorithm

The Apriori algorithm is an association algorithm and is used over relational databases (Wu et al., 2008). The algorithm identifies item sets by using prior knowledge of frequent itemset properties. Apriori scans through the database and searches for a frequent k-itemsets. The process is iterative until no more frequent k-itemsets are found (Han et al., 2011, p. 249).

3.3.3.9 Ensemble Machine Learning

Ensemble Machine learning is a procedure to increase the accuracy of certain machine learning techniques. An ensemble is a combination of classifiers. There are multiple forms of ensemble machine learning. Han et al. (2011, p. 377) describe three frequently used methods: Boosting, Bagging and Random Forrest. Within this thesis, the focus lies on boosting. Boosting is primarily used for converting weak learners to strong classifiers. As supported by Han et al. (2011, p. 377), Wu et al. (2008) described Boosting as one of the most promising techniques to improve models.

Specifically, Wu et al. (2008) mention ADABooster, Adaptive Boosting. The ADABooster algorithm runs many training iterations focusing on the weaker links. The training models provide votes on which links should be used. These links are ensembled into the final model (Han et al., 2011, p. 380).

3.3.4 Interpreting Machine Learning Models

When building machine learning models, one has to be able to understand how to interpret the results. Every machine learning technique provides unique results. For example, regression analysis provides Root Mean Squared Error, classification provides, among other things, the precision, and clustering provides clusters. This sections shortly describes how to interpret the results of each of the following two machine learning techniques: classification and regression. As explained earlier in this chapter, unsupervised machine learning provides clusters which need to be analysed manually (Han et al., 2011, p. 443).

3.3.4.1 Classifier Evaluation

Han et al. (2011, p. 365) describe that for a proper assessment if the classifier is excellent in predicting the class labels, one has to consider different evaluation techniques. For example, when a dataset is unbalanced, accuracy can be misleading. The evaluation techniques, therefore, differ for each model. Starting the evaluation, one first has to look at the confusion matrix (Han et al., 2011, p. 366). A confusion matrix is a tool for analysing how accurate the model is in classifying labels, shown in Figure 10.

Predicted class

		yes	*no*	Total
Actual class	*yes*	TP	FN	P
	no	FP	TN	N
	Total	P'	N'	$P + N$

Figure 10. Confusion matrix (Han et al., 2011, p. 366)

Where:

- True positives: TP: refer to the positive labels that were correctly labelled by the classifier
- True negatives: TN: are the negative labels that were correctly labelled by the classifier
- False positives: FP: are the negative labels that were incorrectly labelled as positive
- False negatives: FN: are the positive labels that were mislabelled as negative

With knowing the confusion table on can go further and use the evaluation techniques which are visualised in Figure 11.

Measure	*Formula*
accuracy, recognition rate	$\frac{TP+TN}{P+N}$
error rate, misclassification rate	$\frac{FP+FN}{P+N}$
sensitivity, true positive rate, recall	$\frac{TP}{P}$
specificity, true negative rate	$\frac{TN}{N}$
precision	$\frac{TP}{TP+FP}$
F, F_1, F-score, harmonic mean of precision and recall	$\frac{2 \times precision \times recall}{precision + recall}$
F_β, where β is a non-negative real number	$\frac{(1+\beta^2) \times precision \times recall}{\beta^2 \times precision + recall}$

Figure 11. Classification evaluation metrics (Han et al., 2011, p. 365)

3.3.4.1.1 Accuracy

The accuracy of a classifier on a given dataset is the percentage of data labels that are correctly classified by the classifier (Han et al., 2011, p. 366).

3.3.4.1.2 Sensitivity/Recall & Specificity

Sensitivity is the number of positive labels that are correctly identified, while specificity is the proportion of negative labels that are correctly identified (Han et al., 2011, p. 367).

3.3.4.1.3 Precision

Precision refers to the percentage of labels that are labelled as positive and are actually positive labels (Han et al., 2011, p. 368).

3.3.4.1.4 F-Score & Fb

An alternative way to use precision and recall is to combine these into a single measure.
The F-score is an alternative way to combine precision and recall into one measure. The F-score measures the harmonic mean of precision and recall. It provides equal weight to precision and recall. Fb, on the other hands, uses different weights to combine precision and recall in a single measure. In practice, two forms of Fb are used. F2 where the weights of recall weight twice as much as the weights of precision and F0.5 where the weights of precision weight twice as much as recall (Han et al., 2011, p. 369).

3.3.4.1.5 Regression evaluation

In order to analyse a regression model, the root mean square error (RMSE) should be analysed. The RSME is used to measure the differences between values predicted by a model or an estimator and the values observed (Bishop, 2006, p. 7). The RMSE is always non-negative, and a lower RMSE is preferred. So, the lower the RMSE, the better the regression model will be in the prediction

3.3.4.1.6 Extra evaluation metrics
Han et al. (2011, p. 369) describe the following evaluation measurements:
- **Speed:** Which are the computational costs involved in generating and using the model.
- **Robustness:** Which is the ability of the model to make correct predictions given noisy data or data with missing values.
- **Scalability:** Which is to the ability to produce a classifier that is efficiently given large amounts of data.
- **Interpretability:** This refers to the level of understanding and insight that is provided by the classifier. Decision trees and classification rules might be relatively understandable. On the other hand, neural networks are "black boxes" and therefore, difficult to interpret.

3.4 Organisational aspects
In the research objective, it is stated that there is a need for organisational aspects. A proved method is provided by Kaplan & Norton (2007). The approach of Kaplan & Norton (2007) is called a Balanced Score Card. This methodology is a strategic planning and management system that help align, prioritise and measure actions within the company and to communicate the results in the organisation. It tracks the strategic objectives the organisation is trying to accomplish in line with the mission, vision, and strategy of the organisation. Creating a Balanced Score Card investigates the "balance" between strategic measures and operational management.

The BSC analyses from four different perspectives:
- Financial / mission
- Customer/Stakeholder / beneficiaries
- Internal Process
- Organisational Capacity / Learning & Growth

3.4.1 Strategy Map
The Strategy Map is based on the same principles as the Balanced Score Card. It is a more visual representation of the organisation's objectives and how they relate to each other.

A strategic map is created during the strategic planning process and is used as the primary reference material during periodic strategy meetings. In this thesis, the Strategy Map can help to identify organisational aspects as the strategy map is focused on the strategy using machine learning in the SBR context (Kaplan & Norton, 2004). The Strategy Map uses the same perspectives as the Balanced Score Card to help strategic planning and management, including the previously described four different aspects, by implementing the structure of the four perspectives as it helps to identify problems with the proposed strategy. Furthermore, this prevents the wrong approach before starting a project.

3.5 Ethics in information systems

The need for ethical integration is stipulated in Chapter 1. This section discusses different methodologies on ethics in information systems. Exploring the relevant literature about ethics and technology systems result in many theories to choose from.

3.5.1 Value Sensitive Design

A grounded theory on establishing the ethical considerations in the design of a product is Values Sensitive Design by Friedman et al. (2013). Friedman et al. (2013) propose an approach where the values of direct and indirect stakeholders are incorporated into the design of a technology system. The procedure consists of three phases of investigation; conceptual, empirical and technological, visualised in Figure 12. The inquiries are iterative and help to modify the design continuously. Based on the work of Friedman et al. (2013) Davis & Nathan (2015) add some adaptations and also provide a strong basis for this concept

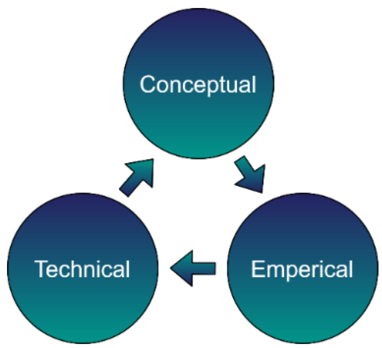

Figure 12. Value Sensitive Design

3.5.1.1 Conceptual investigations

First, VSD investigates both direct and indirect stakeholders. When analysing the stakeholders, the focus lays on understanding and articulating the stakeholders. It also presents their values considering the use of the proposed technology, in this case, machine learning. One must carefully analyse these values and clarify fundamental issues such as conflicting values.

3.5.1.2 Empirical investigations

Second, the researcher must examine the “understanding” the identified stakeholders have regarding the context in which the artifact is situated.

3.5.1.3 Technical investigations

In this phase, the investigation focusses on the technical part. It concentrates on how people use related technologies on the design of systems to support values identified in the conceptual and empirical studies.

3.5.2 Ethical Impact Assessment

Another arising ethical approach for information systems is the Ethical Impact Assessment. The Ethical Impact Assessment is a process during which an organisation, considers the ethical issues or impacts posed by a new project, technology together with stakeholders (Wright, 2011).
The Ethical Impact Assessment is developed focusing on responsible research & innovation. It facilitates decision-makers and researchers by the formulation of guidelines to ensure ethical

implications are seriously taken into account. As the increasing impact of research and innovation on society, researchers should reflect on the effects of these new technologies and incorporate the reflections during the design of new technologies. Key points, as stipulated by Wright (2011), are the precondition of transparency of the design of new technologies. Furthermore, it specifies the importance of examining the stakeholders before accepting the latest technologies. Wright (2011) provides clear guidelines and questions that need to be asked. Reijers et al. (2016) continued on the work of Wright (2011) and constructed a framework, consisting out of six steps.
1. Conduct an EIA threshold analysis
2. Prepare and EIA plan
3. Set up and execute an ethical impact identification assessment
4. Evaluate the ethical impacts
5. Formulate and implement remedial actions
6 Review and audit the EIA outcomes
Each step is subdivided into multiple sub-steps. All the steps can be found in Appendix A.1. Figure 13 shows a simplified visualisation of the framework of an EIA.

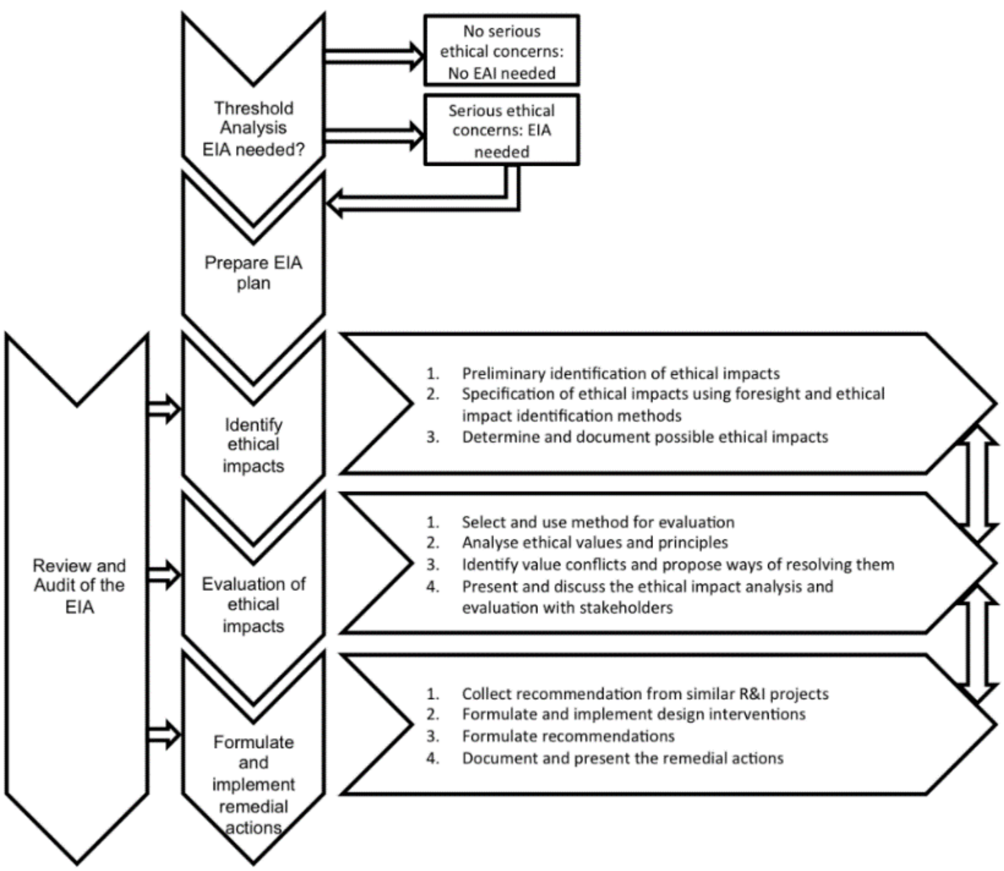

Figure 13. Ethical Impact Assessment (Reijers et al., 2016)

3.6 Conclusion

By conceptualising the theoretical base as input for setting up the model, a detailed literature survey has been carried out. The focus and emphasis of this literature review were on the selection of relevant concepts in order to identify the potential design objectives as part of the overall method. These selected concepts will be the basis for the input of the experiment, and in addition to that, the first design cycle of the method. The literature survey specifically focussed on the concept of structured data, data mining and its connection to machine learning, artificial intelligence in respect to machine learning, and a selection of algorithms of machine learning including ethical and organisational aspects. The outcome of the survey shows that for the development of a successful ML project method as part of the research question, the relevant concepts are referring to data, knowledge discovery in databases, machine learning, artificial intelligence, ethical frameworks and organisational frameworks. These selected concepts, as input for the first design cycle, are indispensable and therefore, crucial for a successful machine learning project. Therefore, this chapter partly answers sub-questions 1.

Chapter 4. Design phase 0: Exploration to design objectives

Designing a reliable method supported and or recognised by the relevant stakeholders, is an iterative process where literature, experiments and interviews together will be used to produce and refine the proposed method.

In the previous chapter, an extensive literature review was conducted. This chapter reports the experiments done in the SBR context and create guidelines (design objectives) for creating the proposed method.

This chapter is divided into three sections. First, it focuses on the machine learning experiments conducted on the DUO case, which will be discussed in Section 4.1. These are two experiments, first, a classification experiment in which the educational institution is classified. Second, a regression experiment in which governmental subsidy is estimated. Section 4.2 addresses the creation of the Strategy Map; the overall strategy is focused on the use of machine learning. The last section, Section 4.3, relies on the data found in the previous chapters and data from the experiments and Strategy Map, to create the Design Objectives. This is an essential aspect of developing the method since it lays out the guidelines for the next chapters, following Step 3 of the Design Science by Peffers et al. (2014).

4.1 Experiments

In order to conduct the machine learning experiments, the process of (Fayyad et al., 1996; Han et al., 2011, pp. 6–8) is used as described in Chapter 3. The process of knowledge discovery in databases is translated and specified only on the machine learning steps of the knowledge discovery in databases. The derived steps can be found in Table 1.

Table 3. Combining aspects Knowledge Discovery in Databases

Step name	**Han et al. (2011, pp. 6-8)**	**Fayyad et al. (1996)**	**Action**
1) Data collection	collect and combine relevant datasets	this step is to select the data set on which the discovery is performed	collect the data that is needed
2) Data preparation	to remove noise and inconsistent data	the third step is focused on the data preparation in which the data needs to be cleaned and pre-processed.	cleaning the data, the removal of noise, and the handling of missing data.
3) Algorithm selection	select the proper data mining techniques	to choose a unique set of data mining algorithms	select algorithms from the Table 2
4) Data mining	an essential process where intelligent methods are applied to extract data patterns	search for relevant trends, or in other words using the algorithms to find relevant information.	use the algorithms to create models
5) Evaluation	to identify the truly interesting patterns representing knowledge based on interestingness measures	focused on analysing and understanding the results of the mined patterns	evaluating the machine learning model, patterns and project

4.1.1 Machine learning workbench

There are different ways to build models with machine learning algorithms. For example, it is possible to build these models via programming or machine learning workbenches. In this thesis, there is specifically chosen to use a machine learning workbench. There are different kind of workbenches, such as Azure, Weka and more. A machine learning workbench uses a collection of machine learning algorithms and is relatively user friendly as the user does not have to program. In this research is chosen for Weka, as Weka enjoys a widespread acceptance internationally in both academia and business environments (Hall et al., 2009). The machine learning algorithms can be applied directly to the selected data or called from Java code. Weka is a collection of tools for regression, clustering, association, classification, data pre-processing and visualisation (Hall et al., 2009).

4.1.2 Background DUO

DUO (Dienst Uitvoerend Onderwijs) is an agency of the Dutch Ministry of Education, Culture and Science. This agency implements various education laws and regulations. For example, the DUO provides funding for most educational institutions. To do this properly, all educational institutions share their financial data with the DUO. The DUO checks this data and examines the financial state of the educational institution. Currently, the checks of these educational institutions are done using an analysis of predetermined key figures. In case this model pin-points an educational institution considered to be a risk, these results are manually checked by a group of employees. In cooperation with DUO, it is investigated the added value of machine learning algorithms in this process. However, as the data is not public, it has been decided to set up a similar experiment to develop a generic method, independent whether the used data is open for public or not.

4.1.3 Hypotheses

This experiment is initiated to explore the possibilities of machine learning in a real case and to create reliable input for the Machine Learning Project Method. To conduct a reliable experiment, one must test a hypothesis. A scientific hypothesis has two requirements, it should be testable and falsifiable (Sekaran & Bougie, 2010, p. 45). Following these criteria, the two hypotheses were derived:
Hypothesis 1: "With a classification model, it is possible to predict the correct educational institution of the DUO SBR dataset above 90% accuracy."
Hypothesis 2: "With a regression model, it is possible to predict the governmental subsidy with a Root Relative Square Error below 25%."
These hypotheses were set up in dialogue with DUO.

4.1.4 Experiment 1: Classification

This section elaborates on the steps that are conducted in the experiment. The main objective of the first experiment is to find out whether hypothesis 1: "With a classification model, it is possible to predict the correct educational institution of the DUO SBR dataset above 90% accuracy." is true or false. A detailed description of the steps is provided in Appendix B.1.

4.1.4.1 Step 1: Data collection

The experiments are done on the open dataset of DUO. This dataset is a transformed version of the XBRL dataset. This new format, CSV, is readable by Weka and is converted to ARFF. ARFF is the data format of Weka. The dataset consists of the financial balance per educational institution from the year 2014-2018. The total dataset consists of 8115 instances. The number of instances per year is visualised in Table 4.

Table 4. DUO-dataset

Year	**Instances**
2014	1844
2015	1622
2016	1582
2017	1533
2018	1513

Combining datasets

In addition to the "Balance" dataset, the "Governmental Subsidies" dataset is used. This dataset contains the attribute "Governmental Subsidies OCW", which is added to the "Balance" dataset. Adding this attribute is advised by DUO.

4.1.4.2 Step 2: Data preparation

Handling missing data & noise

From the merged dataset, the Authorised supervision (identification number), Grouping and Name are deleted, as they are conflicting with the experiment. The full description, including all attributes, can be found in Appendix B.1. Attributes that have more than 85% missing files are removed, which are:

- "Immateriele vaste active" 92% missing
- "Voorraden" 89% missing
- "Kortlopende effecten" 90% missing

To make an attribute that needs to be analysed, the attribute is changed into a class attribute in Weka. For the first experiment, it will be the attribute sector.

Splitting the dataset
The cleaned dataset is now split into two sets. A training set and a test set, which the algorithms will use to train and test the model respectively, will be generated and used.

4.1.4.3 Step 3: Algorithm selection

The objective is to predict a nominal label. This requires a classification algorithm.
For classifying the Sector attribute, the top 5 algorithms of Weka are used, which are described in section 3.3.3. The five algorithms are

- Logistic Regression
- Naive Bayes
- k-Nearest Neighbor
- Decision Trees
- Support Vector Machines

4.1.4.4 Data mining

Training the models
It is necessary to experiment with different algorithms to understand which algorithms are promising. The model will be trained on the dataset from 2014-2017. In Table 5, the results of trying all the algorithms with the standard settings of Weka. A detailed description of the settings and the results per algorithm can be found in Appendix B.1.

Table 5. Results Classification models (1)

Algorithm	Algorithm Weka	Correctly Classified Instances
Logistic Regression	functions.Logistic	84.6993%
Naive Bayes	NaiveBayesMultinomialText	70.8226 %
k-Nearest Neighbor	lazy.IBk	18.8002 %
Decision Trees	trees.REPTree	86.8353 %
Support Vector Machines	functions.LibSVM	70.8075 %

Evaluating build models
The first results show that the REPTree algorithm performs the best regarding the accuracy of the model. However, we check the built models with the unseen data from 2018. This helps conclude whether the model responds differently to new data.

Algorithm	Algorithm Weka	CCI training	CCI 2018
Logistic Regression	functions.Logistic	84.6993%	84.6001%
Naive Bayes	NaiveBayesMultinomialText	70.8226 %	68.6715%
k-Nearest Neighbor	lazy.IBk	18.8002 %	24.4547%
Decision Trees	trees.REPTree	86.8353 %	87.31%
Support Vector Machines	functions.LibSVM	70.8075 %	68.6715%

Table 6. Results Classification models test set

The results in Table 6 shows that REPTree decision tree model has the highest accuracy.

Improving the models

The REPTree algorithm performed best in terms of the accuracy of the model. This is the starting point for examining if it is possible to create a model with higher accuracy. This is done by systematically changing each of the parameters of the algorithm. The explanation of the settings can be found in Appendix B.1 (Witten & Frank, 2002). The results of changing the parameters are described in Table 7.

Table 7. Results of systematically changing the parameters of the REPTree algorithm

Setting	Standard-setting	New setting	Increase
Batch size	100	100	0%
Debug	False	False	0%
doNotCheckCapabilities	False	False	0%
initial count	0.0	1.0 / 2.0	0.0303%
MaxDepth	-1	-1	0%
minNUm	2.0	1.0	0.0606%
minVarianceProp	0.001	0.001	0%
noPruning	False	True	0.2424%
numDecimalPlaces	2	2	0%
numFolds	3	4	0.1364%
seed	1	-2	0.4242%
SpreadInitialCount	False	False	0%

The results show that changing parameters can provide higher accuracy. The next step is to combine the parameters and explore whether it reaches a higher accuracy. Multiple combinations were tried, which did not lead to a higher result. Therefore, the most optimal settings are the standard parameters with the seed setting set to -2. Changing the seed can provide overfitting. This increases the accuracy with 0.4242% to 87.2595%, which does not confirm the hypothesis.

Ensemble Machine Learning – Improving models

For the purpose of creating a "better" model, the possibilities of ensemble machine learning have endeavoured. One of the options in ensemble machine learning is boosting. Boosting is primarily used for converting weak learners to strong ones, which is explained in Section 3.3.3.9.

The next step is to examine the possibilities of ensemble machine learning. With ensemble machine learning, it might be possible to build a model with higher accuracy. This experiment uses ADAbooster explained in Section 3.3.3.9. The explanation of the settings within ADAbooster can be found in Appendix B.

Using the Adabooster with standard parameters, the accuracy is increased with 2.3633% to 89.1981%. Table 8 shows all the REPTree parameters which were systematically changed in the ADAbooster.

Table 8. Results of systematically changing the parameters of the REPTRee with standard ADABooster parameters

Setting	**Standard-setting**	**New setting**	**Increase**
Batch size	100	100	0%
Debug no change	False	False	0%
doNotCheckCapabilities	False	False	0%
initial count	0.0	0.0	0.%
MaxDepth	-1	-1	0%
minNUm	2.0	1.0	0.0758%
minVarianceProp	0.001	0.001	0%
noPruning	False	True	1.1968%
numDecimalPlaces	2	2	0%
numFolds	3	8	0.3787%
seed	1	1	0%
SpreadInitialCount	False	False	0%

With the ADAbooster and the REPtree parameters of noPruning set to True, an accuracy of 90.3952% was achieved. The next step is to systematically change the settings of the ADAbooster. The parameters can be found in Appendix B.1 The results are shown in Table 9.

Table 9. Results of systematically changing the parameters of the ADABooster with standard REPTRee parameters

Setting	**Standard-setting**	**New setting**	**Increase**
Batch size no change	100	100	0%
Debug	False	False	0%
doNotCheckCapabilities	False	False	0%
numDecimalPlaces	2.0	2.0	0%
numIterations	-1	100	1.9391%
Resume	False	False	0%
seed	1	-1	0.0879%
useResampling	False	True	0.6060%
weightthreshold	100	100	0%

Changing all the parameters of the Adabooster systematically provided three parameters that increase the accuracy. The most promising parameters is numIterations, which provided an increase of 1.9391% in accuracy, which led to the total accuracy of 91.1377%.
The next step is to try to combine multiple options. However, this process is described more as an art form rather than an exact science (Brownlee, 2016). The results of combining the different parameters can be found in Appendix B.1. The most promising combination is when the parameters of ADAbooster are: numIteration = 100 and resampling = true, and the parameters of the REPtree are the standard parameter settings. The model was tested on the test dataset, which indicates if there is overfitting. However, the results of the model were an increase of 4.4085% compared to the previous build REPtree model and led to an accuracy of 91.2438%.

4.1.4.5 Evaluation

Walking through the machine learning steps and building a model resulted in a model which can predict the correct label with an accuracy of 91.6061%, exceeding the proposed accuracy of hypothesis 1: “With a classification model it is possible to predict the correct educational institution of the DUO SBR dataset with +90% accuracy”. Therefore, this hypothesis is confirmed. The goal of building an accurate model was focused on accuracy. In this experiment, the other classification evaluation metrics were not compared as the focus was limited to the accuracy. However, these evaluation metrics should also be advised to create a proper classification model.

4.1.4.6 Conclusion classification experiment

The goal of the experiment was to implement a full machine learning experiment. The experiment was successful in two-fold. First of all, it was possible to create a model which was able to predict the educational institution with an accuracy of 91.6%. This is 1.6% higher than the minimum value of the hypothesis. Furthermore, the experiment follows a procedure of building a classification model that is documented and is useful as input for the design objectives.

4.1.5 Experiment 2: Regression

This chapter discusses the process conducted in the experiment. The main objective of the second experiment is to find out if Hypothesis 2: “With a regression model it is possible to predict the governmental subsidy with a Root Relative Square Error lower than 25%” is true or false. A detailed description of the steps is provided in Appendix B.1.

4.1.5.1 Data gathering & Data preparation

For this experiment, the same dataset is used for the first experiment. Therefore, the first step and the second step are the same as explained in the first experiment. There is only one adjustment which is changing the class that needs to be analysed to the governmental subsidy attribute. In order to train a Support Vector Machine model, the algorithm cannot handle data attributes. Therefore, for training the SVM model, the date attribute is deleted.

4.1.5.2 Algorithm selection

The objective is to predict a numeric attribute. This requires a regression algorithm. Five algorithms are selected from Table 2 that fit the regression problem the best. These five algorithms for a regression problem are:

- Linear Regression
- k-Nearest Neighbor
- Decision Trees
- Support Vector Machines
- Artificial Neural Network

4.1.5.3 Data mining

The algorithms are trained on the training dataset with the standard settings of Weka (Hall et al., 2009). A detailed description of the settings and the results per algorithm can be found in Appendix B.1. The results of the models are displayed in Table 10.

Table 10. Results Regression models (1)

Algorithm	**Algorithm Weka**	**RMSE**	**RRSE**
Logistic Regression	functions.LinearRegression	13155867	31.6686%
k-Nearest Neighbor	lazy.IBk	113668168	273.6199%
Decision Trees	Trees.REPtree	11170476	26.8911%
Support Vector Machine	functions.SMOreg	17881820	43.0448%
Artificial Neural Network	functions.MultilayerPerceptron	12218352	29.4118%

Evaluating the build models

After building the models, they are tested on unseen data from 2018. This helps to conclude if the models respond differently to new data.

Table 11. Results Regression model on test data

Algorithm	Algorithm Weka	RMSE	RRSE
Logistic Regression	functions.LinearRegression	13155867	31.6686%
k-nearest-neighbor	lazy.IBk	113668168	273.6199%
Decision Trees	Trees.REPtree	10535042	25.3598%
Support Vector Machines	functions.SMOreg	17881820	43.0448%
Artificial Neural Network	functions.MultilayerPerceptron	12218352	29.4118%

The results in Table 11 show that the REPtree algorithm produced the lowest RMSE, which is the goal.

Adjusting parameters

In order to examine if it is possible to create an improved model, the parameters are systematically changed. A description of the parameters can be found in Appendix B.1. The parameters. The results of changing the parameters are described in Table 12.

Table 12. Results Regression systematically changing parameters REPTree

Setting	Standard-setting	New setting	Increase RMSE
Batch size	100	100	0%
Debug	False	False	0%
doNotCheckCapabilities	False	False	0%
initial count	0.0	1.0 / 2.0	0 %
MaxDepth	-1	-1	0%
minNUm	2.0	1.0	1,4356%
minVarianceProp	0.001	0.001	0%
noPruning	False	True	5.1742%
numDecimalPlaces	2	2	0%
numFolds	3	5	0.2284%
seed	1	3	2.6278%
SpreadInitialCount	False	False	0%

Ensemble Machine Learning

In order to examine the option to create a better model, the possibilities of ensemble machine learning are tried. The parameters of the Bagging are described in Appendix B.1. Using Bagging with the standard Weka parameters (Hall et al., 2009) are:
RMSE = 8810534.9278 and RRSE = 21.2099 %.

Systematically changing the parameters of Bagging produced improvement when the parameter was set to noPruning. This model achieved the best results which are a Root mean squared error of 8390352 and Root relative squared error of 20.1984 %. The model was tested on the training data, and no significant change was found.

4.1.5.4 Evaluation

Following the steps for creating a machine learning project consequently, it finally produced a model which achieves an RRSE of 20.1984%. This result is lower than the RRSE from the hypothesis 2: "With a regression model it is possible to predict the governmental subsidy with a Root Relative Square Error lower than 25%". Therefore, this hypothesis is confirmed.

4.1.5.5 Conclusion regression experiment

The goal of the experiment was to walk through a full regression experiment. The experiment was successful. First, the hypothesis was achieved. Second, the experiment provides insights into the creation of a regression model. These insights provide input for the design objectives of the method.

4.2 Strategy Map

The previous sections elaborated on the experiments that were conducted with machine learning. This section focuses on the production of a strategy map. The strategy map helps to identify four different perspectives on the proposed strategy, as described in Section 3.4.1. These perspectives are mission, customers/beneficiaries, internal processes and learning and growth. The goal of making a strategy map is to create insight of the strategy, which is: using machine learning. Furthermore, it provides insights on organisational aspects. The production of the strategy map is described in this section.

4.2.1 Mission

The mission, in this case, is the request of a standard SBR stakeholder, DUO that wants to use ML. It is derived from the perspective that the stakeholders strive to operational excellence, which is visualised in Figure 14 . For example, the inspections of DUO want to be able to be successful in detecting a financial risk of their organisations. Because of the need for operational excellence, the

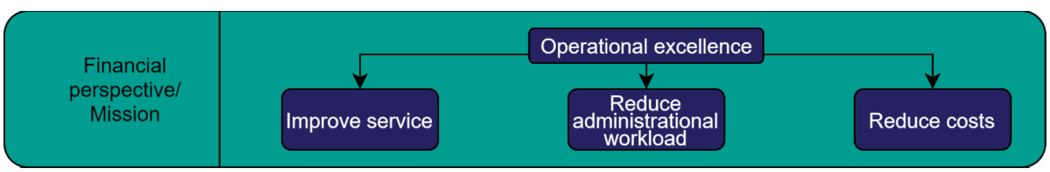

Figure 14. Mission

SBR stakeholders want to improve their service, reduce administrational workload and reduce costs. Therefore, when the goal of the strategy aligns with the mission, it has a higher chance of succeeding.

4.2.2 Beneficiaries perspective

The external stakeholder/beneficiaries value proposition is the core of the strategy, which is why it comes directly after the mission. The value propositions revolve around serving the needs of the external stakeholders/beneficiaries, which can have different values and needs. These demands need

Figure 15. Beneficiaries perspective

to be analysed. Examples could include costs, quality, functionality, service and privacy. These are visualised in Figure 15.

In the case of a financial risk estimation of the DUO, the clients of DUO (educational institutions) do not have a choice, whether to be checked or not. They do not purchase the product. They strive to be controlled in a fair process to guarantee their privacy.

4.2.3 Internal Processes

The use of machine learning is in two-fold. It can be used to acquire new insights. For example, to find new strategies. It might also be used for problem-solving, explained in Section 3.4.1. The internal processes are shown in Figure 16.

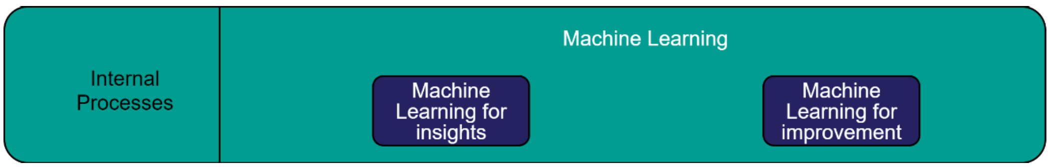

Figure 16. Internal Processes

4.2.4 Learning & Growth

Learning and growth is the foundation of the strategy. This perspective outlines the employee's skills and knowledge required to make the proposed strategy really work. For the strategy to use machine learning in an organisation, the following requirements are found. First, there needs to be an alignment in what the management team wants to achieve and what the employees can handle. It is difficult to push a (new) strategy if it is not in line with the mission. Furthermore, there has to be machine learning expertise in the organisation, or an employee willing to invest time learning the theory of machine learning. Lastly, there is a need for someone that has access to the data, because simply said, with no data, there is no machine learning. This person should also have data knowledge.

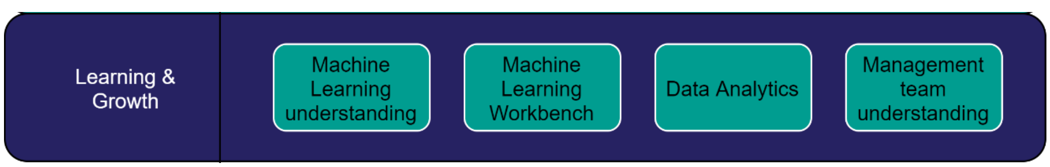

Figure 17. Learning & Growth

The reason for this is that algorithms can make predictions and show connections. These results need to be interpreted by someone that understands the predictions and connections in the context of the subject. Some examples are visualised in Figure 17

4.2.5 Strategy map

Figure 18 shows the full strategy map. However, this does not apply to each case. Therefore, the method should use concepts divined by Kaplan and Norton (2004).

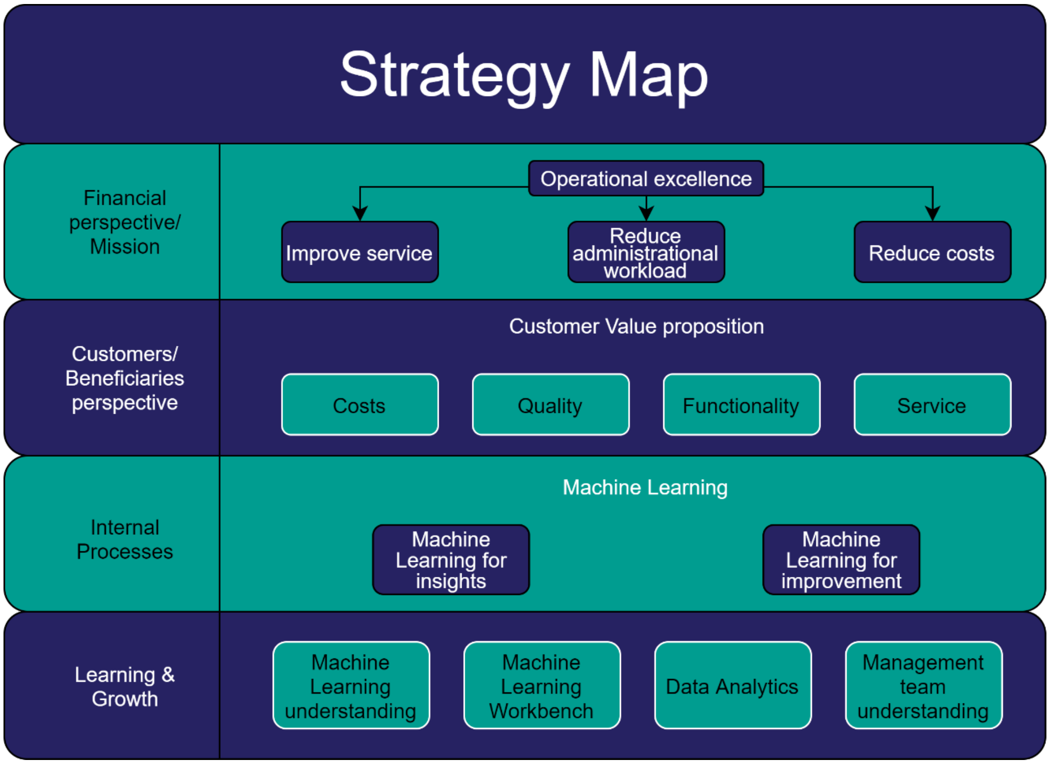

Figure 18. Strategy map

4.3 Design objectives

This section described the definition of concrete design objectives the method should meet to fulfil the research objective. These design objectives are derived from four different perspectives resulting from the actual research, 1) the research questions (Chapter 1), 2) what is found in the literature (Chapter 3), 3) the data from the experiments (Chapter 4), and 4) what is learned from producing the strategy map (Chapter 4).

The main research question as defined in Chapter 1, reads as follows: "**How can technical, organisational and ethical aspects be combined into a method that supports stakeholders to systematically set up machine learning projects in SBR context?**"

Five accompanying sub-questions have to be clarified to support the development of the method:

1. What are the relevant machine learning, organisational and ethical factors for SBR-stakeholders?
2. How can the identified factors be combined into a scientific method?
3. Does the designed method provide the guidelines needed for systematically setting up machine learning projects?
4. Does the designed method provide insight into the value that machine learning might provide?
5. What is the potential of machine learning in SBR context?

The argumentation of the combination of the machine learning, organisational and ethical factors, is elaborated on in Chapter 1.

In Chapter 3, the literature review has been discussed. In this chapter, two machine learning experiments were conducted, and a full Strategy Map on the strategy of using machine learning was built. Chapters 1, 3 and 4 together provide input on how the method should be designed. This section summarises which design objectives are selected to design a proper method which complies with the research questions.

As a result of the above described research, six design objectives are developed, subdivided into two categories; design objectives focusing on what the method includes, considering specific action steps, and design objectives focussing on what the method should provide. The following design objectives are formulated as a combination of both.

Design objective focusing on the method, to include:

1. **The designed method should include an ethical framework**

The necessity for ethical guidelines is described in chapter 1. As described in Section 0, the literature on ethical frameworks provides guidelines and frameworks that can be used in the design phase of the method.

2. **The designed method should include machine learning steps to create a model in SBR context**

The need for machine learning aspects is described in chapter 1. The guidelines are developed based on the theoretical concepts for the machine learning factors, derived from the literature survey, as well as on the results of the knowledge discovery in databases, as discussed in Section 3.2. Furthermore, the experiments in Section 4.1.1 provide additional in-depth data on the steps to be taken to conduct machine learning projects.

3. **The designed method should include a machine learning algorithm selection method, including multiple machine learning techniques**

The necessity for a machine learning algorithm selection method, including multiple machine learning techniques is specially mentioned by Qiang & Xindong (2006). The literature study on this subject in Section 3.3.3 provides clear guidelines for the proper use of these algorithms. Section 3.3.3 provides a description of how these algorithms work. Furthermore, the experiments discussed in Section 4.1.1, provide a real-life example of how the algorithms work in the SBR environment and which algorithms provide the best insights.

4. **The designed method should include organisational factors relevant for creating a machine learning project**

In chapter 1, the need for organisational guidance during machine learning projects was described. As explained by Klievink et al. (2017) & Adadi et al. (2015) which indicate that the goal or mission of the company, is often not included in the process of creating a machine learning project. Kaplan and Norton (2004) provide clear guidelines as well as the production of the Strategy map in Section 4.2, including the vision and strategy of the organisation.

Design objectives focusing on the method, to provide:

5. **The designed method should provide an understandable process for creating a machine learning project in SBR context**

As stipulated in Chapter 1, the method should provide a clear overview of what is needed for conducting a machine learning experiment.

6. **The designed method should help decision-makers to understand if machine learning can create added value in their organisation**

As discussed in Chapter 1, the potential added value of machine learning is enormous. However, organisations struggle with estimating this value that machine learning can bring to support their corporate strategy. Therefore, the method should provide decision-makers with the ability to determine if machine learning can provide added value to their organisation.

4.4 Conclusion

This chapter described the experiments that were performed on data of DUO. Two experiments were conducted: one classification experiment and one regression experiment. The experiments were done in cooperation with DUO, resulting in the definition of clear objectives and reliable data. From these objectives, two scientific hypotheses were formulated. Hypothesis 1: "*With a classification model it is possible to predict the correct educational institution of the DUO SBR dataset with +90% accuracy*" and Hypothesis 2: "*With a regression model it is possible to predict the governmental subsidy with a Root Relative Square Error lower than 25%*". These hypotheses were tested in an experimental environment, and both were found true. Furthermore, the experiments produced two machine learning models with predictive capabilities. The experiments also provided DUO with insights on the added value machine learning could provide in their organisation. The second part of this chapter describes the strategy Map used to provide insights used as guidelines for the creation of the relevant design objectives, answering sub-question 1. The last section elaborated on the initiation and justification of the design objectives, resulting in the following design objectives:

1. The designed method should include an ethical framework
2. The designed method should include machine learning steps to create a model in SBR context
3. The designed method should include a machine learning algorithm selection method, including multiple machine learning techniques
4. The designed method should include organisational factors relevant for creating a machine learning project
5. The designed method should provide an understandable process for creating a machine learning project in SBR context
6. The designed method should help decision-makers to understand if machine learning can create added value in their organisation

The six objectives have been defined in line with the research question. Although this thesis strives to fulfil these six design objectives, it is recognised that it is not realistic to fulfil all these objectives in one project, especially not in the time-frame of one thesis. The focus lays on creating a first design of the numerous design iterations into a prototype, providing a starting point and direction, and thus a framework from which to generate further research.

Part II

Design and development: building the method

Chapter 5. Design Phase 1: designing the method

In the previous chapter the design objectives were formulated. This chapter elaborates on translating the design objectives into a method for a solid set-up and implementation of a machine learning project within the goals of the thesis. In other words, this chapter elaborates on the found factors that influence a machine learning project and translates these factors into to concrete steps. In Section 5.1 the overall method is visualised. In Section 5.2, a description of the step-by-step development of the method is elaborated on.

5.1 Complete method

In order to provide the reader with a better understand of the development of the method, this chapter starts with visualising the full method, Figure 19. Hereafter, the process and each step are worked out in more detail.

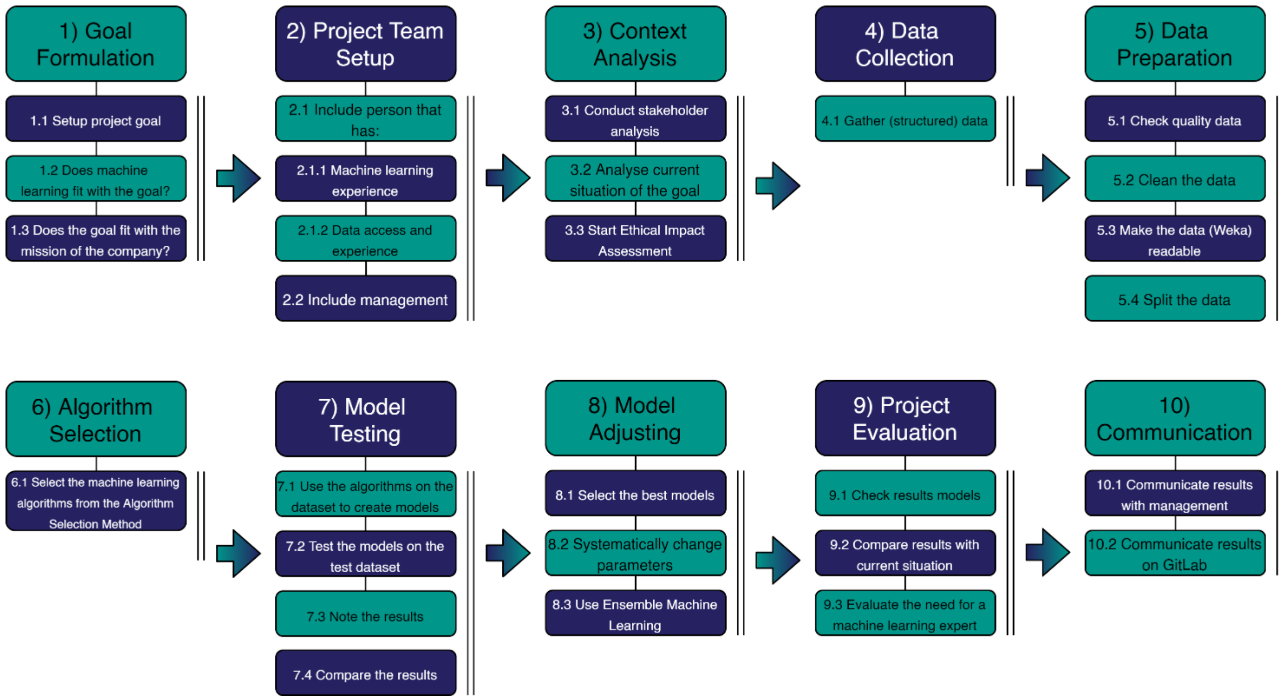

Figure 19. Design Phase 1

5.2 Method development

5.2.1 Step 1: Goal Formulation

When starting a machine learning project, it is crucial to carefully define a project goal (Fayyad et al., 1996). Furthermore, it is important to determine whether the proposed goal of the project can be supported using machine learning. If the goal of the project does not fit within the characteristics of machine learning, the original goal formulation should be redefined or machine learning might not be the best option. It may be possible that other data mining options are a better fit (Fayyad et al., 1996). Furthermore, the goal of the project is checked, whether it is in line with the mission of the company. By aligning the goal of the project with the mission of the company, the machine learning project has a higher chance of being successful (Adadi et al., 2015; Kaplan & Norton, 2004; Klievink et al., 2017). It can be concluded that if the goal of the project and the mission of the company do not align, it is recommended to redefine the original goal. When the goal of the project and the mission do not align the machine learning project can still be started but has a lower chance of succeeding.

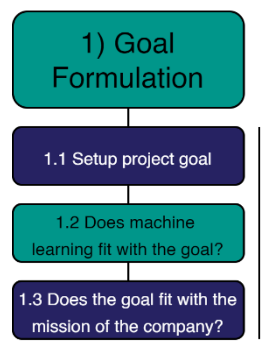

Figure 20. Step 1

Based on the above-described boundary condition, three sub-steps are identified, visualised in Figure 20:

5.2.1.1 Step 1.1 Set up project goal

Formulating and defining the goal of the project. It emphasises the importance of a clear project goal formulation.
Source: (Fayyad et al., 1996)**,** Experiment-DUO
Example/output: estimating the probability if an organisation is fraudulent.

5.2.1.2 Step 1.2 Does the goal fit with machine learning?

Analysing the goal is needed to determine whether it is compatible with machine learning. If the goal is not compatible with machine learning the goal must be reformulated or there might be other data mining options that are more appropriate. Figure 21 provides examples for each machine learning form.
Where:

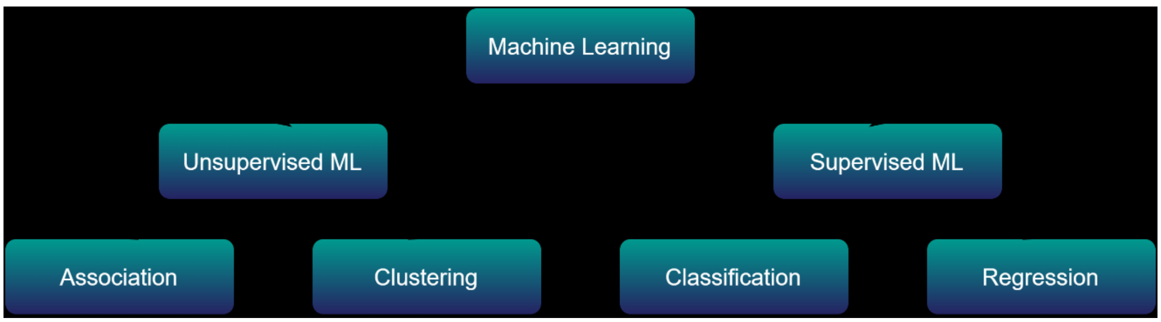

Figure 21. Machine Learning techniques

- Association & clustering are there to gain insights into the data.
  - E.g. clustering the data into groups to define new groups.
- Classification is there to predict a categorical value
  - E.g. to predict the label financial risk which can be the label yes or no
- Regression focusses on predicting a numeric value
  - E.g. to predict the amount of subsidies per organisation

Source: Section 3.3.2, Experiment-DUO
Example/Output: the estimation the of governmental subsidy which is numeric value corresponds with supervised machine learning, specifically, regression.

5.2.1.3 Step 1.3 Does the goal fit with the mission of the company?

In the Strategy Map Kaplan & Norton (2004) discussed that the mission of the company needs to be analysed. Furthermore, the authors emphasise that when the goal of the project aligns with the mission, the project receives support from the internal stakeholders. From this setting, it is derived that the project goal should clearly align with the company's mission in order to avoid a loss of support for the final result.
Source: (Kaplan & Norton, 2004), machine learning as strategy (Mission)
Example: the goal of the financial inspection of DUO is to be able to estimate the financial risk of educational organisations as good as possible

5.2.2 Step 2: Project Team Setup

In order to create a successful project, one should consider the knowledge that is needed (Fayyad et al., 1996). Kaplan & Norton (2004) explain that when applying a new strategy, certain learning and growth needs to be analysed. In other words, certain knowledge and skills are needed and, or need to be acquired. Two crucial skills that are needed for setting up a successful machine learning project are machine learning experience and data experience & access. The need for different skills does not mean that there is a need for different people. One person can have multiple skills. Lastly, Kaplan & Norton (2004) discuss the inclusion of internal stakeholders which is translated to including the management. The sub-steps are shown in Figure 22.

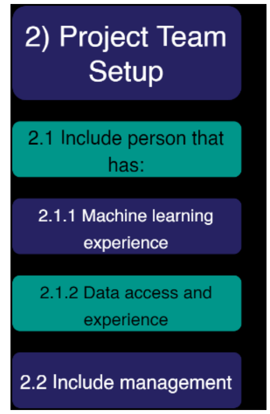

Figure 22. Step 2

Two sub-steps are therefore identified:

5.2.2.1 Step 2.1: Include person that has:

2.1.1) Machine learning experience

Experience with machine learning is necessary as preparations and construction of the algorithms are a prerequisite for successful development of the method. If knowledge of machine learning is lacking, employees should be trained to acquire these skills. One does not have to be an expert to create a preliminary project.

2.1.2) Data experience and access

Additionally, experience with and access to the data is essential in order to successfully initiate a machine learning project. Another required expertise is considerable insight in selecting and comprehending the right set of data in order to correctly use the algorithm and obtain useful outcomes.

Source: (Han et al., 2011, p. 39; Kaplan & Norton, 2004), Learning and Growth,

Example/Output: An employee that has basics of machine learning experience & an employee that works with and has access to the data

5.2.2.2 Step 2.2: Include management

Kaplan & Norton (2004) discuss the inclusion of internal stakeholders which is translated to including the management. The interaction of the several layers in the organisation, employees and management, creates support and provides acceptance of the project.

Source: (Kaplan & Norton, 2004), Strategy Map

Example/Output: Setup a project meeting including the involvement of the management

5.2.3 Step 3: Context Analysis

Figure 23. Step 3

After setting up the goal and the project team, the context needs to be analysed. First, a stakeholder analysis will be done, providing a clear understanding of what the stakeholders want and what their values are (Davis & Nathan, 2015; Kaplan & Norton, 2004). Insight of the ethical factors and impacts must be part of this analysis. Second, the current situation of project goal has been analysed in order to record a clear t=0, which in the final steps support the understanding and evaluating of the project. The last action is to conduct an Ethical Impact Assessment. Described in Chapter 1, there is a need for ethical guidance during a machine learning project. One of the actions that help to execute an ethical project is to do an Ethical Impact Assessment (Reijers et al., 2016; Wright, 2011). The framework of doing an Ethical Impact Assessment will be provided with the method. The three steps are displayed in Figure 23.

Based on the above discussion, three sub-steps are identified:

5.2.3.1 Step 3.1: Conduct a stakeholder analysis

A stakeholder analysis is done in order to identify the relevant stakeholders, to understand what is important for the stakeholders (their expectations & their values) and to analyse how the stakeholders would be affected by completing a machine learning project.
Source: (Kaplan & Norton, 2004) Beneficiary's perspective, (Davis & Nathan, 2015; Wright, 2011)
Example/output: Full stakeholder analysis on the stakeholders affected by the DUO financial inspection

5.2.3.2 Step 3.2: Analyse current situation of goal

Mapping of the current situation of the selected goal is important to compare the t=0 setting with the outcome of the machine learning experiment. This helps to evaluate the overall project on the added value of machine learning.
Source: (Kaplan & Norton, 2004), (Wright, 2011), Experiment-DUO
Example/output: At this moment fraud detection in banking cost 'X' and is done by 'Y' hours of human calculation

5.2.3.3 Step 3.3: Start Ethical Impact Assessment

The ethical impact analyses will be done in order the clarify the potential ethical effects before conducting a machine learning experiment.
Source: (Reijers et al., 2016; Wright, 2011)
Example/output: a complete Ethical Impact Analysis on the use of machine learning to detect fraud in care allowance

5.2.4 Step 4: Data Collection

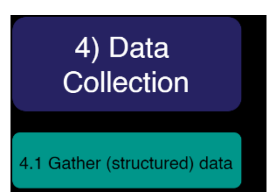

Figure 24. Step 4

The next step focused on the relevant data that needs to create a working and reliable machine learning model. This step is relatively straightforward but can be time-consuming. However, SBR data is stored properly and therefore, relatively easy to access (Bharosa et al., 2015, p. 101). Conducting the data collection has to be done by someone that has experience with data and access to the data, described in Step 2, as it is important to select the appropriate data, garbage in, garbage out (Han et al., 2011, p. 39). Figure 24 provides the visualisation of the step.

Step 4.1: gather (structured) data
Select and gather the appropriate data.
Source: (Fayyad et al., 1996; Han et al., 2011, p. 39)
Example: Retract the financial data from the "DUO" server

5.2.5 Step 5: Data Preparation

The data preparation step is done to create a dataset which can be used to train the machine learning models. First, the quality of the data is checked. This helps to indicate which actions need to be performed. Second, data preparation steps described by (Fayyad et al., 1996; Han et al., 2011, p. 83) include: cleaning the data, removing noise and handling missing data. Third, the transformation of the dataset to a format which is readable by a Machine Learning Workbench. As part of this research, Weka has been used. The last action is to split the dataset into two sets, a training set and a test set (Han et al., 2011, p. 33). This is done so that the produced model, which is built on the training data, can be evaluated on unseen data, which indicates the effectiveness of the model on unseen data and checks if the model was not overfit on the training dataset (Han et al., 2011, p. 85). The subsets are shown in Figure 25.

Figure 25. Step 5

As part of a successful data preparation phase, four sub-steps were identified:

5.2.5.1 Step 5.1: Check data quality

Analyse the quality of the data to indicate if there is a need to clean and or to restructure the data.
Source: (Fayyad et al., 1996; Han et al., 2011, p. 84)
Example/output: Analyse if there is data missing to be able to understand what needs to be done regarding data cleaning actions

5.2.5.2 Step 5.2: Clean the data

In this step, the results of the previous steps are analysed such that the data cleaning tasks can be performed. Tasks are; combining, reducing, replacing and or removing data parts.
Source: (Fayyad et al., 1996; Han et al., 2011, p. 85)
Example/output: Remove attributes in cases of too many missing instances for a clean dataset

5.2.5.3 Step 5.3: Make the data Machine Learning Workbench (Weka) readable

This step adjusts the data format as it needs to be changed into a format readable by the machine learning workbench, in this case, Weka.
Source: (Hall et al., 2009), Experiment-DUO
Example/output: Change to format from XBRL to CSV to ARFF to have a data format which can be used for building machine learning models

5.2.5.4 Step 5.4: Split the data
This step indicates the importance of splitting the dataset into a training dataset and into a test dataset. This step is indispensable as the model can be checked on unseen data and therefore checked on overfitting.
Source: (Fayyad et al., 1996; Han et al., 2011, p. 33), Experiment-DUO
Example/output: The DUO dataset is split into a dataset including data from 2015-2018 and a dataset including the data from 2019

5.2.6 Step 6: Algorithm Selection
This step requires the user of the method to select the algorithms from the algorithm selection method and is shown in Figure 26. The method is based on the most promising algorithms, explained in Section 3.3.3, and the DUO experiments. The used algorithms are explained in Section 3.3.3 and will be provided with the method. The method itself is visualised in Figure 27, showing which algorithms should be used with the corresponding machine learning techniques.

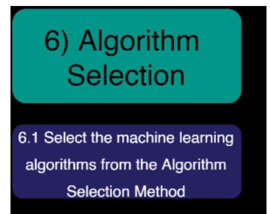

Figure 26. Step 6

The algorithm selection method provides the user with clear guidelines on which algorithm to use. In Step 1.2, the form of machine learning is already identified. Furthermore, it provides algorithms which are relatively "easy" to understand, described in Section 3.3.3. This helps projects with no expert knowledge of machine learning in understanding the considerations the model takes. Additionally, choosing an algorithm which is relatively "easy" to understand provides the ability to explain the model to stakeholders which contributes to the transparency of the project. And the transparency of the project diminished the ethical problems (Davis & Nathan, 2015).

5.2.6.1 Step 6.1: Select the machine learning algorithms from the algorithm selection method
Select the relevant algorithms that fit with the machine learning goal.
Source: Section 3.3.3, Experiment-DUO
Example/output: The machine learning goal is a classification problem. Therefore, we use the following algorithms: Logistic Regression, Naive Bayes, k-Nearest Neighbors, Decision Trees, Support Vector Machines

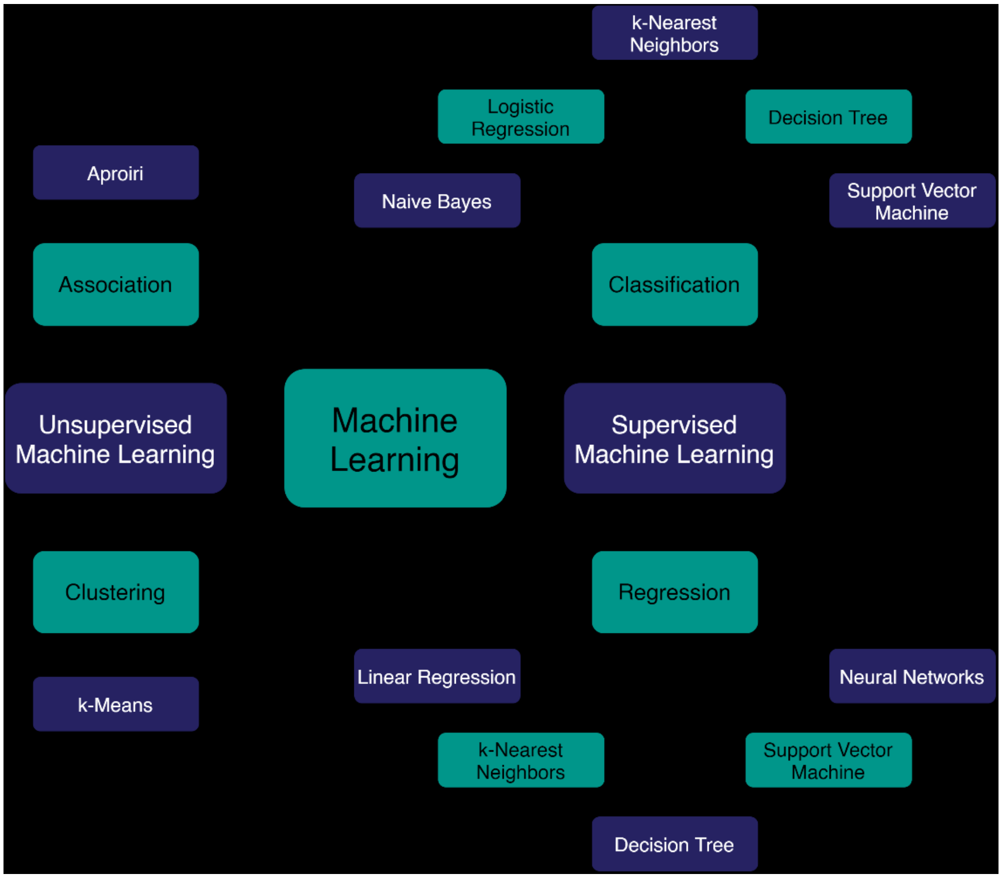

Figure 27. Algorithm Selection Method

5.2.7 Step 7: Model Building

Based on the algorithms selected in the previous step, this step uses these algorithms to build the machine learning models (Fayyad et al., 1996; Han et al., 2011, p. 8). Hereafter, the created models are tested on the unseen test data (Han et al., 2011, p. 33). Lastly, the results are noted and analysed (Fayyad et al., 1996). The steps are shown in Figure 28.

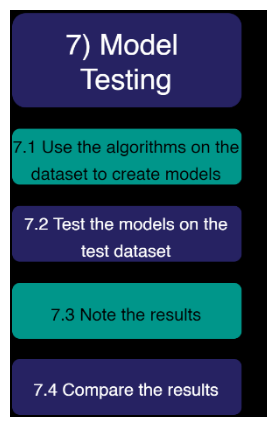

Figure 28. Step 7

As part of this process, four sub-steps are identified:

5.2.7.1 Step 7.1: Use the algorithms on the dataset to create the models

This step is to train the model, a crucial step within this process of machine learning. This is done by using the algorithms selected in the previous step and training them on the prepared dataset (step 5, data preparation). For this step, the standard settings of the algorithms are used (Hall et al., 2009).
Source: (Fayyad et al., 1996; Han et al., 2011, p. 8), Experiments-DUO
Example/output: Use the algorithms from Step 6.1 on the dataset to train the models

5.2.7.2 Step 7.2: Test the models on the test dataset

Testing the dataset on the test data helps to indicate if the model is built in a way that it is not prone to overfitting. This avoids bias and produces a “fair” model.

Source: (Fayyad et al., 1996; Han et al., 2011, p. 33), Experiment-DUO
Example/output: Test the built model on the unseen dataset of 2018

5.2.7.3 Step 7.3: Note the results
Note the results of the experiments such that they can be compared.
Source: (Fayyad et al., 1996), Experiment DUO
Example/output: See the example results in Table 13

Table 13. Example results step 7.3

Algorithm	**Algorithm Weka**	**CCI training (%)**	**CCI 2018 (%)**
Logistic Regression	functions.Logistic	84.6993%	84.6001
Naive Bayes	NaiveBayesMultinomialText	70.8226 %	68.6715
k-Nearest Neighbors	lazy.IBk	18.8002 %	24.4547
Decision Trees	trees.REPTree	86.8353 %	87.31
Support Vector Machines	functions.LibSVM	70.8075 %	68.6715

5.2.7.4 Step 7.4: Compare the results
This step is created to evaluate the results of the models built in the previous step. These results need to be interpreted and compared such that the most promising model can be selected. In order to interpret the results, Section 3.3.4 provides guidelines for interpreting the results of machine learning models.
Source: (Fayyad et al., 1996; Hall et al., 2009; Han et al., 2011, p. 8), Experiment-DUO
Example/output: Interpreting the accuracy and precision of the models and comparing them with each other

5.2.8 Step 8: Model Adjusting
The next step is to adjust the parameters of the selected model and/or use ensemble machine learning to examine if it is possible to improve the results. Nonetheless, this step is not always necessary, as the original defined goal can already be achieved. Yet, it might be possible to create a model which is superior to the defined goal. The sub-steps are visualised in Figure 29.

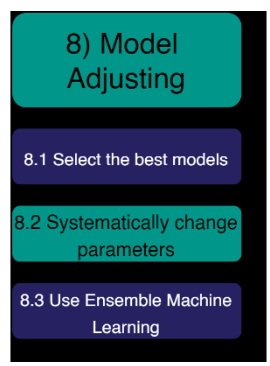

Figure 29. Step 8

Three sub-steps are identified:

5.2.8.1 Step 1: Select the best model
This step indicates the selection of the best model for further improvement of the model in line with the research question. In this step, it is important to select the model which results fit the goal of the project the best.
Source: (Fayyad et al., 1996; Han et al., 2011, p. 243)
Example/output: from the models that are built model 'X' has the highest accuracy and the best precision. Therefor model 'X' is chosen.

5.2.8.2 Step 2: Systematically change parameters
In order to improve the model, it might be possible to systematically change the parameters of the algorithm. The results can be noted and compared.
Source:(Bishop, 2006; Hall et al., 2009), Experiment-DUO

Example/output: turning the parameter "NoPruning" in a decision tree algorithm provided an increase in accuracy.

5.2.8.3 Step 3: Use Ensemble Machine Learning

Ensemble machine learning is found to be effective in increasing the results of algorithms in some specific cases (Han et al., 2011, p. 377). Therefore, this step stipulates the option to select an ensemble machine learning technique to improve the created model.
Source: (Han et al., 2011, p. 377; Wu et al., 2008), Experiment-DUO
Example/output: Use ADAbooster to discover if it is possible to improve the model

5.2.9 Step 9: Project Evaluation

After the previous step, the new model needs to be tested on the previous unseen dataset. Furthermore, it needs to be decided if the machine learning project has been successful. This will be done by comparing the final results to the original situation, which is analysed in Step 3.2. Lastly, it needs to be determined if there is a need for additional machine learning expertise to further develop and improve the model.

To fulfil this step, three sub-steps are identified, which are shown in Figure 30. Step 9:

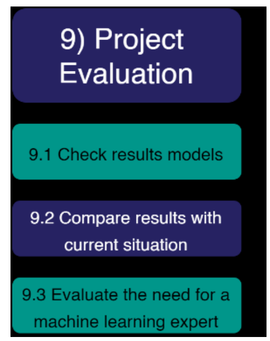

Figure 30. Step 9

5.2.9.1 Step 1: Check results model / test the results on the test dataset

This step focusses on evaluating the improvements that are made in the previous step. This is done by testing the dataset on the unseen test dataset. Furthermore, the results of the model are evaluated.
Source: (Han et al., 2011, p. 21), Experiment-DUO
Example/output: The improved model is tested on the unseen test data set of 2018

5.2.9.2 Step 2: Compare results with current situation

This step focusses on evaluating and comparing the model on the situation it was produced for.
Source: (Fayyad et al., 1996; Han et al., 2011, p. 8), Experiment-DUO
Example/output: At first, a system was able to predict fraud with an accuracy of 90%, the new model is able to predict fraud with an accuracy of 95%

5.2.9.3 Step 9.3: Evaluate the need for a machine learning expert

This step evaluates whether it is needed to include a machine learning expert on the project. The experiment can be conducted by someone with a minimal machine learning experience, but to further improve the model, there might be a need for a machine learning expert.
Source: Experiment-DUO
Example/output: Because the project was done with a minimal machine learning experience, and the model shows protentional, a machine learning expert is advised.

5.2.10 Step 10: Communication

This final step describes the communication of the project. Although the project has been implemented in closes interaction with the internal stakeholders, the management and employees, a final and transparent communication by the project members is crucial. It supports the potential impact on the work-processes by the management in line with the strategy of the organisation.

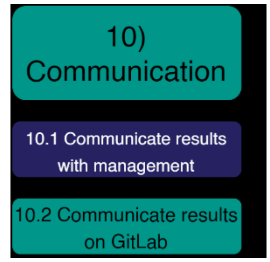

Figure 31. Step 10

As stipulated by Umbrello & Bellis (2018) & Wright (2011) it is important for projects like these, they are thoroughly communicated with the internal and external stakeholders. This contributes to the transparency and reproducibility of the project. Visualisation is provided in Figure 31.

Two sub steps are identified:

5.2.10.1 Step 10.1: Communicate results with management

Communicating results with management keep the alignment with employee and employers such that the work provides information to the decision-makers and is not redundant.
Source: (Fayyad et al., 1996; Kaplan & Norton, 2004)
Example: Set up a meeting with management to explain the final results so that informed decision making can be conducted.

5.2.10.2 Step 10.2: Communicate the results on GitLab

As stipulated by (2015) it is important in projects like these to thoroughly communicate with the stakeholders and outside world. This inspires transparency and reproducibility. Additionally, it creates open communities and inspires people to attend, think along and even experiment on their own.
Source: (Davis & Nathan, 2015)
Example: Upload the full experiment on GitLab

5.3 Conclusion

In this chapter, a method is discussed in order to set up a machine learning project in line with the research question. The method is based on the literature provided in Chapter 3, as well as the data gathered from the experiments and the created strategy map, both described in Chapter 4. The combined and final product of this chapter is the first version of the scientific method, taking into account the design objectives set in Chapter 3. This scientific method is subdivided into ten unique steps (goal formulation, project team setup, context analysis, data collection, data preparation, algorithm selection, model testing, model adjusting, project evaluation, communication) all necessary to complete a first model. This first model will be validated in the next phase, by interviews of a selection of stakeholders.

Chapter 6. Interview methodology

After having designed the first version of the method from an intensive literature review and the DUO experiments as described in the previous chapter, the next step in this research is to validate the designed method, test its comprehensiveness and determine the missing parts. To be able to do so, a tailor-made interview format is required, which will be described in detail in the next sections. In Section 6.1, the outline of the interview goal is given. In Section 6.2, the interview methodology is explained. In Section 6.3, the interview protocol is displayed. Hereafter the selection of the interviewees and justification is described (6.4). Finally, the interview analysis (6.5) provides guidelines on how the interviews will be transformed into useable data.

6.1 Interview goal

Having drafted an initial methodology as described in Chapter 4, the framework of a successful interview should give insight in and provide answers on the following four objectives: First, the results of the experiments must be analysed with the stakeholders of the experiments. Second, the interviewee must provide the usefulness and understandability of each step of the designed method, so that irrelevant and not well-explained content can be adjusted. Third, the interviews should help in identifying relevant improvements in the method that have not been included in the first version. The last goal of the interviews is to validate the overall method on its relevance, applicability, practical use and ability to provide value in their organisation. The outcome of the interviews, when successfully conducted, will bring the following results:

- Evaluation of the experiment
- Evaluation of each step of the designed method
- Evaluation of relevance and applicability of the designed method
- Improvements of the designed method as input for design cycle 2

6.2 Interview methodology

This part of the research is in place to fulfil the goals set up in Section 6.1, and a tailor-made round of individual semi-structured interviews will be done. To conduct interviews, appropriate literature has been consulted (Sekaran & Bougie, 2010). After selecting what the preferred output is resulting from the interviews, a semi-structured method is chosen to guarantee this. The selected method allows adjusting questions during the interview, based on the knowledge and experience of the interviewee. This semi-structured approach is open for adding remarks and suggestions during the interviewing phase. Next to the semi-open questions, the interviewee is requested to give input in a quantitative part of the interview. This part allows the researcher to evaluate each step of the designed method, the relevant factors of the total method and the experiment quantitatively.

6.2.1 Ethical considerations

The interviews are recorded and attached in this report to guaranty trackability and repeatability to the research (Sekaran & Bougie, 2010, p. 19). Additionally, the participants will be asked to give their written consent. Therefore, the TU-Delft has constructed a form of consent that will be provided to the participant before the interview. In the consent form is stated how the collected data is used in this thesis considering privacy. In Appendix C.1 Consent Form for Master thesis of Steven, a standard consent form is attached. From the recordings, an interview transcript is created. The drafted transcripts processed after the interviews will be sent to the participants for final validation and clearance if the data can be used in the thesis. Participants will be asked to check the transcripts thoroughly for irregularities. Irregularities will be corrected.

6.2.2 Justification evaluation criteria

In order to comply with Design Science (Peffers et al., 2014), it is vital to obtain useful data from the interviews. Therefore, relevant evaluation criteria are selected to evaluate the method and the experiments (Prat et al., 2014). First, the interviewees are asked to rank each step on its usefulness for the method and its understandability. This is done by ranking via a Likert scale (Sekaran & Bougie, 2010, p. 172). Secondly, the interviewee is asked to evaluate the overall method. Various evaluation criteria (Prat et al., 2014) have been selected to ensure that the data from the interviews meet the desired standards of Hevner et al. (2004). The evaluation criteria can be found in Table 14. These evaluation criteria have been thoroughly selected so that the data from the interviews can be used specifically to evaluate and improve the developed method.

Table 14. Artifact evaluation criteria

Criteria	Description	Statement
Effectivity	the degree to which the artifact produces its desired effect	The method is effective in achieving a machine learning project
Utility	Degree of quality of practical use	The method is practical to use
Understandability	Degree of ease of use	The method is easy to understand
Fit organisation	Degree of usefulness in the organisation(experiment)	The model suits my organisation
Completeness	Level of detail and consistency	The method is complete
Robustness	ability to respond to the fluctuations of the environment	The method has the ability to respond to environmental fluctuations
Accuracy experiment	Accuracy of the method in line with the experiment	The method is consistent with the experiment

6.3 Interview protocol

The interview protocol includes six crucial elements.

1. First, the interviewer starts a “Coffee Chitchat” to put the interviewee at ease and gain a relaxed atmosphere. The interview is not a one-directional exam, but a transfer of information essential to improve the method; a mutual benefit.
2. Explanation of the used methodology: Before asking detailed questions to the interviewee, an overall explanation of the designed method and experiments are presented. Afterwards, the interviewee is explicitly asked if there are additional questions referring to the designed method.
3. Validation experiment: This step is set-up to determine whether the experiment was successful and is useful for the organisation. This step can only be successful if the interviewees have insight into the experiment and the related procedures in their organisation.
4. Validation of identified method: This step is committed to validating if all actions that are identified in the first design cycle, are of relevance for the designed method. Furthermore, it

checks and therefore guarantees if the steps are useful for creating a machine learning project. It supports a thorough analysis of the total use of the designed method.

5. Discussion on other relevant variables: This step is open for extra input for the method by additional remarks and suggestions by the interviewees. The interviewed experts provide information for the second design phase of the method.
6. Lastly, the talking points and conclusion are recapped in order to confirm the interpretations and to minimise misinterpretation.

6.4 Interviewee selection & justification

As the methodology is specifically developed for SBR users, there are only a restricted group and/or organisations able to provide the relevant information. For this type of interviews, judgment sampling is the most suitable sampling design (Sekaran & Bougie, 2010, p. 285). Judgement sampling uses the opinions of the interviewee, which contains a certain level of bias. A vital step to reduce the possibility of an adverse selection (one-sided/too narrow) is to identify a relevant target group. This should be a reflection of representatives in the relevant knowledge field within the organisations, as shown in Table 15. The interviewees should have proper knowledge of the different areas such as machine learning, data mining, the experiment and SBR.

Table 15. Interviewee selection

Interviewee ID	Organisation	Position	Justification
IN1	DUO	Adviser OCW Taxonomy	IN1 is an innovation consultant within DUO. He seeks new business opportunities for DUO. Therefore, he made the connection between DigiCampus and DUO so that a machine learning experiment could be conducted. Lastly, he was the DUO connection during the experiment.
IN2	MD Aguilonius | Taxxor | SBR Logius/BZK | SBR Nexus | OpenSBR.org | XBRL International	Advisor/ Managing director/ XBRL-expert	IN2 has over 20 years of experience as a business consultant and has a strong background in finance and IT. IN2 support companies and governments to prepare for the transformation of manual, paper-based reporting to digital, standards-based reporting. IN2 has worked on projects within governments, financial institutions and multi-nationals in the areas of Standard Business Reporting, XBRL.
IN3	Visma B.V.	Solution Architect at Visma Connect B.V / IT expert / XBRL expert	IN3 is a seasoned expert on IT. Worked at many IT companies and has the ability to master new complex domains and implement elegant and creative solutions. IN3 has an excellent track record in implementing solutions in an agile fashion.
IN4	DigiCampus / Logius	Business Consultant Logius	IN4 is a business consultant for Logius. He seeks new business opportunities for Logius. As IN4 is creating experiments for SBR-stakeholders on a daily base, he will be useful for analysing the usefulness of the designed method.

IN5	WSW	IT exploration	IN5 is involved in risk management including which data is requested, how this data is processed and how WSW can best organise its risk management system.
IN6	WSW	Risk Manager at WSW	IN6 is involved in risk management including which data is requested, how this data is processed and how WSW can best organise its risk management system.

6.5 Interview analysis

The interviews structure is subdivided into four parts, two qualitative and two quantitative. First, it provides a qualitative reflection on the experiments. Second, it provides quantitative data on each step. Third, it provides another quantitative part of the functionality of the full method. Last, it provides qualitative improvements on the method.

As the first part of the analysis of the interviews, they have to be transcribed. Because the quantitative section of the interviews is already coded during the interview, these parts are not included in the transcription. Furthermore, the coffee chit chat is also not transcribed as it is not relevant for the research. So the qualitative parts are reproduced in the form of words (Sekaran & Bougie, 2010, p. 370), meaning that per interviewee, the results of the quantitative part are noted in a table, and the qualitative part is transcribed into words. Both can be found in Appendix C.1.

6.5.1 Qualitative analysis

To draw meaningful conclusions out of the qualitative data (Sekaran & Bougie, 2010, p. 370) suppose a specific process. First, the data needs to be reduced. This is done by coding where units of text are given labels. Hereafter the labels subdivided into categories (Sekaran & Bougie, 2010, p. 372). Third, the categories are analysed by the researcher with deductive reasoning. Lastly, the results are used to improve the designed method and create the next iteration. (Sekaran & Bougie, 2010, p. 374).

To draw meaningful conclusions out of the qualitative data Sekaran & Bougie (2010, p. 370) suppose a specific process. First, the data needs to be reduced. This is done by coding, where units of text are given labels. Hereafter the labels are subdivided into categories (Sekaran & Bougie, 2010, p. 372). Third, the categories are analysed by the researcher with deductive reasoning. Lastly, the results are used to improve the designed method and create the next iteration (Sekaran & Bougie, 2010, p. 374).

6.5.2 Quantitative analysis

In order to analyse the quantitative part of the interview, the following steps are executed to examine the input. As previously stated, the quantitative part consists of two main parts; 1) to evaluate each step for understandability and effectivity, and 2) to evaluate the designed method on its ability to create a successful machine learning experiment.

In order to start the analysis, first, the interviews need to be coded, which is done during the interview. Second, the responses need to be assigned to a number (Sekaran & Bougie, 2010, p. 316) and put into a database (Sekaran & Bougie, 2010, p. 319).

The next step is to analyse the data. Because the data is interval data, Sekaran & Bougie (2010, p. 318) explain that the standard deviation, in conjunction with the mean, is the appropriate tool because of the statistical rules, in a normal distribution (Sekaran & Bougie, 2010, p. 318). The mean or the average is a measure of central tendency that offers a general picture of the (Sekaran & Bougie, 2010, p. 316). And the standard deviation is a measure of dispersion, which offers an index of the spread of distribution or the variability in the data (Sekaran & Bougie, 2010, p. 318).

6.6 Conclusion

In this chapter, clear interview objectives were formulated. These objectives were translated into an interview protocol, containing both qualitative and quantitative sections. Based on these objectives, a small but relevant selection of interviewees were chosen, and six semi-structured interviews were conducted. The interviews provide input for Design Phase 2, which will be described in Chapter 7. Furthermore, the interviews provide the data for the evaluation of the functionality of the overall method and will be described in Chapter 8.

Chapter 7. Design Phase 2 - dotting the i

In the previous chapter, Chapter 6, the process of conducting the interviews was explained. In this chapter, the output of these interviews is used to analyse and improve the method that was built in Design Phase 1. The first section of this chapter, Section 7.1, provides an overview of the result of Design Phase 2. Hereafter, Section 7.2 expounds on the quantitative data from the interviews, and thus the first interview goal is met. Section 7.3 states the relevant suggestions for improvements made by the interviewees and describes in detail how the suggestions can be used as clear improvements. In Section 7.4, the relevant suggestions for improvements are translated into steps to be adopted in the method, resulting in an update of the designed method. The last part, Section 7.5 of this chapter, describes user guidelines on how and in which context the method is to be used.

7.1 Result Design Phase 2

Figure 32 visualises the results of the improvements made in this chapter. This is done in order to help the reader to understand the process of improving the method and placing the improvements in the overall context.

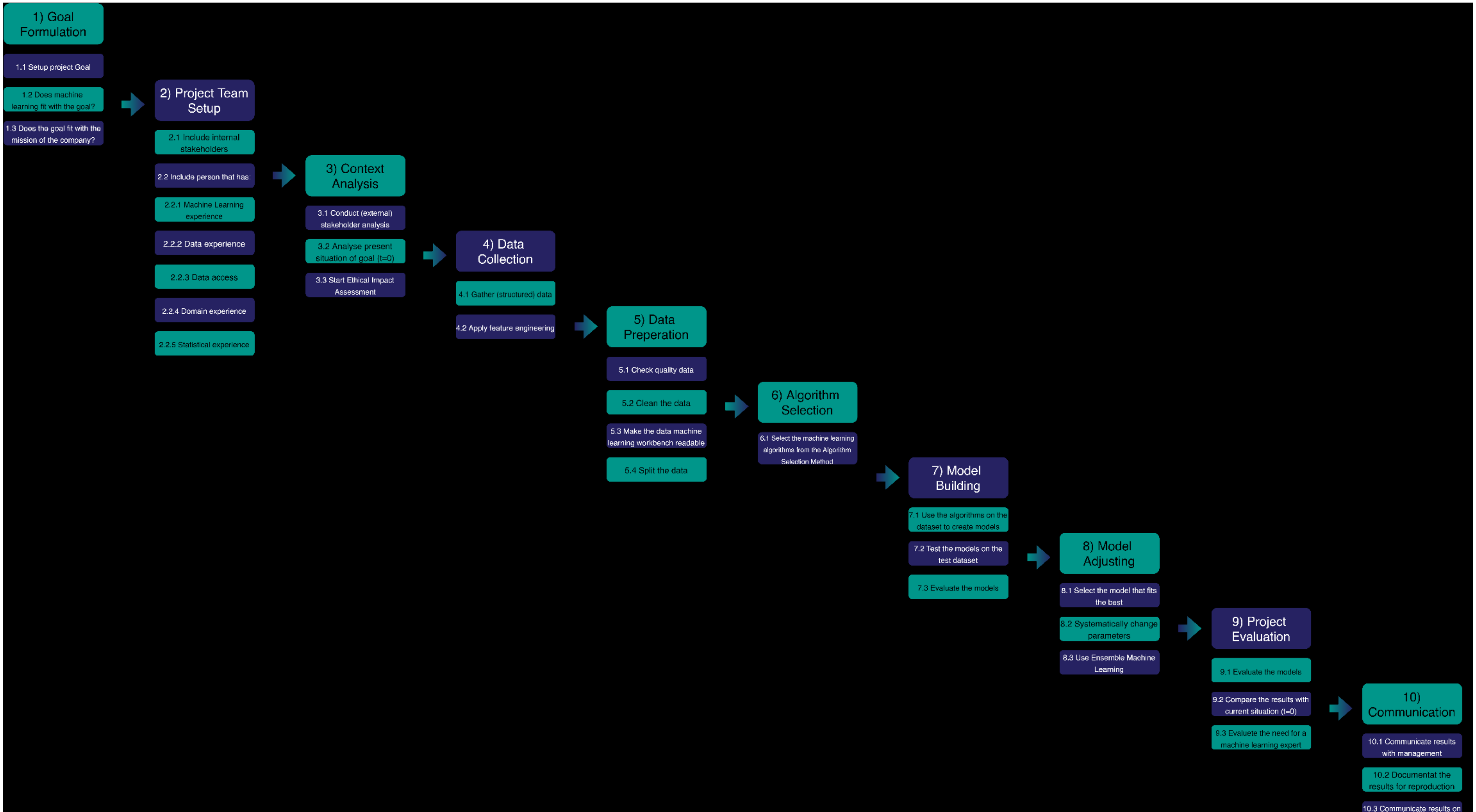

Figure 32. Machine Learning Project Step Design Phase 2

7.2 Quantitative interviews

In Section 6.5.2 is described how the quantitative data of the interviews will be used. The results of this conversion, translating the interval scale to numbers and inserting them into a database, are shown in Table 16. The shown data represent the mean and the standard deviation.

All steps were quantitatively graded on understandability and usefulness for creating a machine learning project. The results are also shown in Table 1; the numbers are representing the mean and the standard deviation.

Analysing the data in the table, there are several grades that stand out: 1.3, 3.1, and 10.2, corresponding with "Does the goal fit with the mission of the company?", "Conduct a stakeholder analysis" and "Communicate results on GitLab". These steps are discussed during the relevant suggestions for improvements in Section 1.3. As part of the qualitative interview, the interviewees provided associated suggestions to improve the method designed in phase-1. Before this information is used, irrelevant steps are deleted. The result of this exercise, a measure of importance and understandability of each step, is presented in Table 16. This transparent subdivision clearly shows that only a few of these sub-steps are classified as lower than 4 by the interviewees, a number on the interval scale corresponding with useful and understandable. These sub-steps will be adjusted based on the suggestions made during the interviews.

Step evaluation

	Understandability		**Usefulness**	
Step:	Mean	Standard deviation	Mean	Standard deviation
1	4.7	0.5	4.8	0.4
1.1	4.7	0.5	4.8	0.4
1.2	4.2	0.4	4.7	0.5
1.3	3.8	1.5	4.2	1.6
2	4.8	0.4	4.8	0.4
2.1.1	4.8	0.4	4.7	0.5
2.1.2	4.5	0.5	4.5	0.5
2.2	4.5	0.8	4.0	0.9
3	4.5	0.8	4.8	0.4
3.1	3.8	1.5	3.8	1.6
3.2	4.5	0.5	4.7	0.5
3.3	5.0	0.0	4.5	0.8
4	5.0	0.0	5.0	0.0
4.1	5.0	0.0	5.0	0.0
5	5.0	0.0	5.0	0.0
5.1	5.0	0.0	5.0	0.0
5.2	4.8	0.4	5.0	0.0
5.3	4.7	0.8	5.0	0.0
5.4	4.7	0.5	4.7	0.5
6	4.7	0.8	5.0	0.0
6.1	4.7	0.8	5.0	0.0
7	4.7	0.8	4.8	0.4
7.1	5.0	0.0	5.0	0.0
7.2	4.8	0.4	4.8	0.4
7.3	4.2	0.8	4.5	0.8
7.4	4.3	0.8	4.8	0.4
8	4.5	0.8	5.0	0.0
8.1	4.5	0.8	5.0	0.0
8.3	4.3	0.8	4.8	0.4
8.3	3.8	0.8	4.5	0.8
9	4.7	0.5	4.8	0.4
9.1	4.3	0.8	5.0	0.0
9.2	5.0	0.0	4.8	0.4
9.3	4.5	0.5	4.5	0.8
10	4.7	0.5	4.7	0.5
10.1	4.8	0.4	4.5	0.8
10.2	3.5	1.6	3.8	1.8

Table 16. Results Evaluation Steps

7.3 Relevant suggestions for improvements based on interviews

In Table 17, the results of coding and categorising of the quantitative data from the interviews, relevant suggestions for improvement, are presented. The categories provide relevant suggestions made by the interviewee. Detailed descriptions of the interviews can be found in Appendix C. In order to adopt the suggestions made by the interviewees, a strict protocol should be used to adapt the proposed
suggestions for improvements. First, the suggestions must be checked on contradictions. When there are contradictions, the reasoning and explanation behind the suggestions need to be clarified in order to clearly understand if a suggestion needs to be adopted. Second, relevant literature is advised to check whether the suggestion can be supported or rejected. Last, if there is no relevant literature that provides support for the suggestion, the suggestions will not be adopted.

Table 17. Suggestions for improvements

Step	Suggestions for improvement	Action	Interviewee
Step 1-10	Clarify in the figure that the process is partly iterative in the method	Add	IN3, IN5, IN6
Step 1	Add a context check for using the method	Add	IN3, IN5, IN6
	- Where the method can be used for and in what context		
Step 2	Include internal stakeholders instead of management only management	Add	IN2, IN3, IN5
	- And explain the importance of communication with internal stakeholders		
Step 2	Add domain experience	Add	IN1, IN2
Step 2	Add statistical experience to the ability	Add	IN6
Step 2	Split the step "Data access and knowledge" into two separate steps	Change	IN2
Step 4	Add the step "Feature engineering"	Add	IN2, IN3
Step 7	Combine the step note the results and compare the results	Change	IN1
Step 8	Explain the risk of overfitting	Add	IN3
Step 10	Include proper internal documentation for reproducibility	Add	IN2
Step 2	Explain the importance of communication with internal stakeholders	Adjust	IN1, IN5
Total	Make the visualisation more iterative	Adjust	IN3, IN5

7.4 Implementation of relevant suggestions into the method

The suggestions presented in Table 17, will be implemented in the phase-1 designed method to create the phase-2 version. The new steps are described step-by-step in this section.

7.4.1 Step 1: Goal Formulation

Step 1.1: Check context method

The method is originally designed for specific users and for a specific goal. Before using this method, it is important to check if this method really fits the proposed use. As explained, this method provides clear guidelines in setting up a machine learning project considering ethical, organisational and machine learning factors. The designed method is only tested on a specified user (SBR-stakeholders) working with a specified data structure (SBR structured data). The specific context is elaborated on in Section 7.5 . The updated Step 1 is visualised in Figure 33.

Source: IN3, IN6

Example/output: Check if the method fits with the context of the user

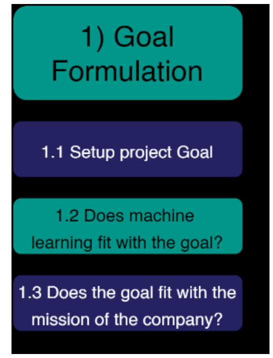

Figure 33. Step 1 updated

7.4.2 Step 2: Project Team Setup

Step 2.1: Include internal stakeholders

This step was originally described as to "include management", but changed to "include internal stakeholders". As stipulated by the interviewee, although the project has to be agreed by the management, the management and employees together are relevant internal stakeholders combining all kinds of relevant knowledge. It also supports the potential impact on the work-processes by the management in line with the strategy of the organisation. Furthermore, this contributes to the internal communication with the relevant internal stakeholders, so that all can benefit from the project and provide additional input. The update of Step 2 is shown in Figure 34

Source: IN1, IN2, IN6, (Kaplan & Norton, 2004)

Example: Include internal stakeholders and communicate intern

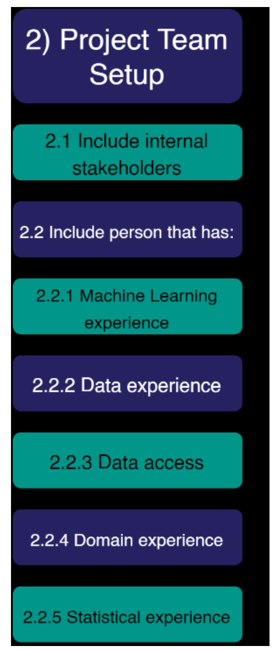

Figure 34. Step 2 updated

Step 2.2: Include person that has:

A suggestion for improvement was to split data knowledge and data access in two separate actions. Furthermore, it was suggested to include domain knowledge as which the researcher party indicated with data experience, but was not fully understood. Therefore, this step is added to make it more specific and to avoid misunderstanding. Furthermore, IN6 suggested that one of the project members requires statistical experience to validate the actions and results. Both suggestions, domain knowledge and statistical knowledge, are confirmed by Han et al. (2011, pp. 39–44). Adopting the suggestions, Step 2.2 includes:

1. Machine learning experience
2. Data experience
3. Data access
4. Domain knowledge
5. Statistical experience

Source: IN3, IN6, (Fayyad et al., 1996; Han et al., 2011, pp. 39–44)

Example: Add people who have in multiple knowledge fields (Step 2.2.1-2.2.5) relevant experience

7.4.3 Step 4: Data Collection

Step 4.2 Apply feature engineering

IN3 stipulated the importance of feature engineering. This step perfectly aligns with the need for that the team member with the domain experience is the same person that should conduct the feature engineering, as also stipulated by IN1, IN2, IN4. Furthermore, Han et al. (2011) describe the process as data transformation, confirming the relevance for this step. The new version of Step 4 is visualised in Figure 35.

Source: IN2, IN3, (Fayyad et al., 1996; Han et al., 2011, p. 83)

Example: Combine two attributes to one attribute that has a higher predicting power

Figure 35. Step 4 updated

7.4.4 Step 7: Model Adjustment

Step 8.2 Systematically change parameters

IN3 stipulated to emphasise that by changing the parameters of a model, overfitting might be the result. Overfitting is elaborated in Section 3.2, and countermeasures are already described, such as Step 4.4. However, it is useful to explicitly stipulate the risk of overfitting before starting systematically changing the parameters of the algorithm. The updated Step 7 is shown in Figure 36.

Source: IN3, (Fayyad et al., 1996; Han et al., 2011, p. 85)

Example: Changing the seed of a decision tree can lead to an overfitting model.

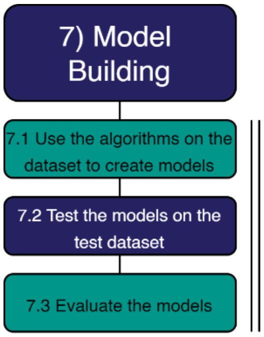

Figure 36. Step 7 updated

7.4.5 Step 10: Communication

Step 10.2: Document the project

During the interview with IN2, it became apparent that the overall project should be documented internally. Documenting the project internally helps other employees in the organisation to easily access the findings of the project. Furthermore, if the employees who participated in the project are no longer working for the company, the project is clearly documented. The new Step 10 is visualised in Figure 37.

Source: IN2, (Fayyad et al., 1996)

Example: Transcribe and store all the relevant deliverables in the cloud of the organisation

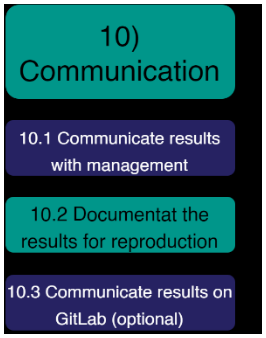

Figure 37. Step 10 updated

Step 10.3: Share on GitLab (optional)

Most interviewees mentioned this step as important; however, IN6 advised it to be optional as not all organisations' can share their data, e.g. because of privacy reasons. However, by sharing the project and its results, the organisation creates transparency, which is preferred by Friedman et al. (2013) & Wright (2011).

Source: IN6, (Davis & Nathan, 2015; Friedman et al., 2013)

Example: Share the project on GitLab to provide insight on how the model was built

7.4.6 Visualise the iterative process

The literature on Knowledge Discovery in Databases stipulates that the process is iterative (Fayyad et al., 1996; Han et al., 2011, p. 6). The fact that the process is iterative was previously discussed. Only IN3 and IN5 emphasised the need for change in visualisation. Therefore, adjustments are made to the visualisation as part of the method. The new visualisation is shown in Figure 32.

7.5 Context method

This section elaborates on the context of the designed method. This context provides clarity on when this method can be used. The user has to find out if it is appropriate to use this method so that the user does not start a dead-end process.

When to use machine learning?

It is not so easy to answer the question when to use machine learning. There are many situations where machine learning is not a sufficient technique, and it requires inside in the potential added values before a time-consuming implementation of a less successful methodology. However, it is difficult to determine in advance whether machine learning will actually work. That's why making a "simple" prototype is an excellent way to determine the potential. Before a decision can be made, there are a number of requirements in advance, described in the method. A simplified representation can be found in Appendix D.1. This consists of steps, as shown in Figure 38.

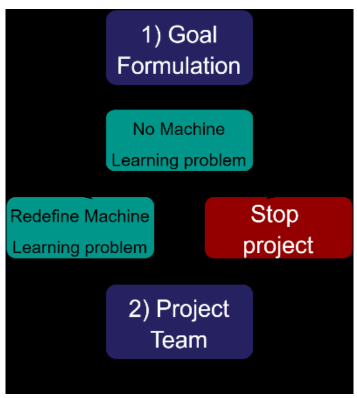

Figure 38. Simplified version "Stop project"

7.5.1 When can the method be used?

There is an endless number of factors that prevent you from using this method. It's impossible to list and discuss all of them. This section, therefore, describes for whom this method is intended and stipulates a number of key points of the method, such as the degree of knowledge and the users. In addition, this section describes what the method can provide. This is based on the experience of the experiments.

As mentioned, several kinds of knowledge are required before this method can be used. In addition, there are situations in which, when the knowledge level is higher, the method produces a better model. First, the situation of the project is being analysed. Second, the level of knowledge is evaluated. The user must comply with the following requirements in order to use this method successfully:

- Stakeholders that have a SBR data structure
- Stakeholders that want that have the following knowledge levels or ability to gather this expertise:
  - Machine learning experience
  - Statistical experience
  - Data experience
  - Domain experience

Experience

Each sub-step of Step 2.1 is described in the next section. This gives the user insight into the level of experience required to use this method. Experience is hard to quantify. The following factors are based on the experiments to approximate this.

Machine Learning Experience

In a situation where there is no machine learning knowledge at all or access to someone with machine learning experience, this method will not produce a machine learning prototype. What is possible is that the user of the method has a technical background, such as data science or ICT, and will take an (online) machine learning course. As a result, sufficient knowledge can be developed over a relatively short period of time to implement the machine learning components of the method. This experience level was the input for the experiments in this research and defining the required minimum experience level.

On the other hand, the method might be less effective for a machine learning expert, as the method is set up and designed in a format that both the technician as well as the manager are able to use it and oversee and understand the outcome and consequences. This asks additional qualifications of the specialist to abstract and to communicate his knowledge in such a way the manager (non-specialist) does understand the results and is able to use the output to support the strategy of the organisation.

In conclusion, the method can be used to create a model in a situation where no significant machine learning expertise is the minimum experience level available. “No significant” described as “by a user with a technical background who has followed at least 40 hours of Machine Learning Workbench (online) courses”. In the event that someone has more machine learning expertise, it is more likely that the model will result in a more accurate model. At the same time, if the model becomes more accurate, it might lower the transparency of the model as it is more difficult to understand. Without transparent abstraction, transfer and explanation to and acceptance by the management (or society), this model might be a sub-optimal or limiting factor in the primary process.

Domain experience

Domain knowledge refers to the knowledge necessary in the researched domain. In this case, the knowledge and insight in the used SBR data and domain. Two experiments have been carried out in this study; in both cases, employees who have worked within this domain for at least one year. These experiments show that with one-year experience, it is possible to create a working model using the method. Therefore, the minimum experience is set for one year. Less experience than one year experience has not been formally tested and falls therefore, out of the scope. To conclude, one year is the minimum domain experience needed to work with the method.

Data experience
In this study, data expertise indicates that a person has the ability to retrieve data from databases, modify data and other relevant actions. The difference between data expertise and domain expertise may seem to be limited, but in this research, it is specifically differentiated. Data expertise focuses primarily on what an employee must be able to do in order to use the data in a proper way. Domain expertise is more focused on which data and how it should be used, such as combining two different attributes to create more predictive power. This study used employees who had been working with data in the company for at least one year.

Statistical experience
Statistical expertise in this research was also self-learned. Han et al. (2011, p. 44) provided the guidelines for understanding the basics. However, more experience results in a model more fit for the problem (Han et al., 2011, p. 44).

7.5.2 What can the method provide?

What the method actually provides can be divided into two parts. First of all, the method gives the user a general insight into what is involved in setting up a machine learning project. It also helps the user to make a reliable estimate for which cases machine learning can be used. These aspects give the potential user a proper idea of what it takes to set up a model, and if machine learning has added value within this context. This process step is extremely relevant to prevent initiating a machine learning project without a sound estimate of the applicability. What the method provides, before actually starting a machine learning project, is:

- It gives policymakers and engineers an overview of what it takes to start a machine learning project (including preconditions and restrictions)
- It provides insight into and examples of, possible applications of machine learning
- It enables a structured process for both engineers and managers; creating alignment and understanding between management and engineers can lead to synergy
- It presents guidelines to implement the project ethically

When employing the developed method, the user endeavours a structured and partly iterative process to set up machine learning projects. The user is guided step by step through the process of a machine learning project, taking into account organisational and ethical aspects as part of a machine learning project.
When all steps are completed, the method provides the following deliverables:

- insight into whether machine learning has added value for the organisation
- a machine learning model/prototype (and/or insights)
    - produced ethically
    - cost-efficient
    - simple
- insights into what the future organisational need might be e.g. in human recourses, more employees with machine learning expertise, fewer employees with "old skills".

7.6 Conclusion

The interviews have a clear added value in the development of an improved method. The quantitative results clearly evaluated the method and pointed out the weak spots. The results from the qualitative interviews were translated into clear suggestions for improvements of the method. In total eleven relevant improvements, shown in, Table 18, were adopted.

Table 18. Suggestions for improvement

Step	Suggestions for improvement	Action	Interviewee
Step 1-10	Clarify in the figure that the process is partly iterative in the method	Add	IN3, IN5, IN6
Step 1	Add a context check for using the method	Add	IN3, IN5, IN6
	- Where the method can be used for and in what context		
Step 2	Include internal stakeholders instead of management only management	Add	IN2, IN3, IN5
	- And explain the importance of communication with internal stakeholders		
Step 2	Add domain experience	Add	IN1, IN2
Step 2	Add statistical experience to the ability	Add	IN6
Step 2	Split the step "Data access and knowledge" into two separate steps	Change	IN2
Step 4	Add the step "Feature engineering"	Add	IN2, IN3
Step 7	Combine the step note the results and compare the results	Change	IN1
Step 8	Explain the risk of overfitting	Add	IN3
Step 10	Include proper internal documentation for reproducibility	Add	IN2
Step 2	Explain the importance of communication with internal stakeholders	Adjust	IN1, IN5

The improvements were implemented in the first version of the method, resulting in the final method, visualised in Figure 32, and therefore answering sub-question 2.

In addition to the method being developed, this chapter also clearly outlines what the method provides to its users (additional to the present situation). Furthermore, what kind of experience the users must have? This method has been developed for a very specific group of organisations with SBR data. These are organisations willing to investigate whether machine learning can be of added value in SBR context, taking into account an interdisciplinary perspective, an important strength of the method. The next chapter elaborates demonstrating the designed method on a real-life case.

Part III

Demonstration, Evaluation

Chapter 8. Demonstration

The development of the second version of the designed method was discussed in the previous chapter. Following the DSRM the next step is to demonstrate the designed method and provide a proof-of-concept of the method by working out a real-life case. This chapter, therefore, starts with the identification of an appropriate case, Section 8.1. The method is then applied to the selected case and a step - by - step explanation is provided, Section 8.2.

8.1 Identification WSW case

For the application of the method, the “Waarborgfonds Sociale Woningbouw” (WSW) case was chosen. WSW , that translates as Social Housing Guarantee Fund, is the guarantee fund for social housing in the Netherlands. A primary activity of WSW is to assess and manage the potential risks of housing associations and the overall sector. WSW supports social housing associations to be financed at the lowest possible cost.

WSW was deliberately chosen as it is affiliated with SBR-Wonen. The problem statement, to analyse financial risks, is specifically SBR related. The data format used by WSW almost completely matches SBR. From the very beginning, WSW was enthusiastic about the case as they would like to understand whether machine learning is an application that can improve WSW's internal processes.

8.2 Step by step explanation WSW case

In this section, all performed steps on the WSW database are described in order to identify the financial risks.

8.2.1 Step 1: Goal Formulation

In a meeting with an employee of SBR-Wonen, WSW proposed a project using machine learning to identify the energy label of houses, Step 1.1. The goal was to analyse whether this subject is applicable for machine learning, which is Step 1.2. The proposed goal fits with the description of a classification problem. However, the goal ”identify the green labels” did not fully align with the mission of WSW, which is to provide guaranties to social housing associations that build and manage social housing, Step 1.3. However, without the alignment with the mission, the next step was initiated as a meeting with was set up to create alignment.

8.2.2 Step 2: Project Team Setup

After step 1 was completed, a meeting was set up with the internal stakeholders of WSW, Step 2.1. During this meeting, both management and employees attended. The project goal was refined to fit with the mission of the company, which is discussed in the following section, returning to goal formulation. Furthermore, the people with the relevant skills were identified, and step 2.2 was completed.

8.2.3 Back to Step 1: update of the original Goal Formulation

As explained in the previous section, one iteration during the first two phases was conducted to refine the project goal. The adjusted project goal is to investigate if it is possible to create a model able to predict the financial risk of the organisations that are guaranteed by WSW, and therefore using the financial data (SBR) as input. The objective is to clarify the protentional of machine learning in WSW and to build a prototype model. This project goal requires a supervised machine learning model as the nominal label, Financial risk (yes/no) has to be predicated, as shown in Figure 39. An additional benefit is that the experiment provides the examiner's insights on how the algorithm learns and what the relevant data is for the selected algorithm. If the experiment is successful, it could be an extra indicator for defining the financial risk of the organisations. Furthermore, these insights can help improve the existing method. A vital facet of the goal is that it is more important that the algorithm predicts all the risk cases (yes) than the non-risk cases (no). In other words, in the case of WSW is the goal to predict the financial risk as good as possible.

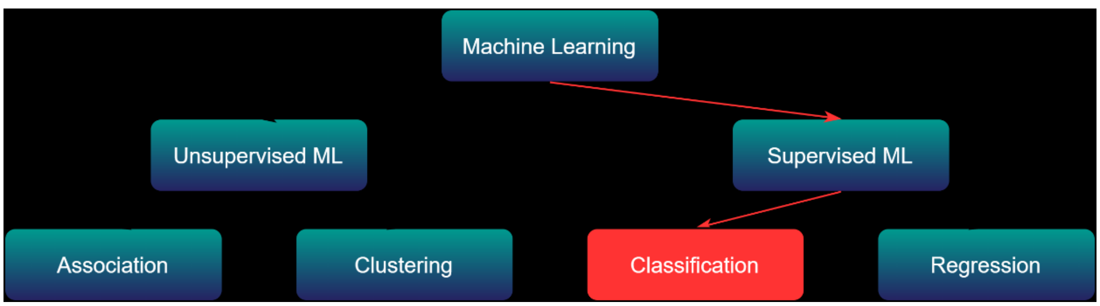

Figure 39. WSW Machine Learning technique

8.2.4 Step 3: Context Analysis

First, the current setting (t=0) is analysed, Step 3.2. Until now, WSW uses a method to analyse if an organisation is at financial risk. This method is developed by in-house employees of WSW together with an outside organisation (S&P). Every year, the method is verified, and every five years, the overall method is redeveloped. Checking and redeveloping the method is an immense and time-consuming operation at high costs.

The main objective of predicting the financial risk by WSW is that WSW decides, based on their internal analysis, if they need to introduce measures to support the organisation to find a way forward. Furthermore, it restraints WSW from providing guarantee to organisations which form a risk. By checking the organisations, WSW safeguards the sector. In case WSW endeavours into wrong propositions, all the related social housing builders have to pay the price as part of the contribution. This reflects in a negative way to the social renters and limits constructors to build new social housing as there is no finance. On the other hand, if WSW makes it too difficult to provide organisations with the necessary finance to build social housing, there will be a shortage of social housing, which affects the people that need social housing. Figure 40 shows the most important stakeholders.

Figure 40. WSW

The first step of the ethical impact analysis (EIA) is to assess the need of the EIA. However, as the aim of this project is to create a prototype, it is decided full EIA does not have to be performed. The added value is in this stage to understand the value that machine learning can bring and to learn from the model how it produces its outcome.

8.2.5 Step 4: Data Collection

In Step 4 an assessment is done of what data is needed and how to collect it. This is executed by a WSW employee with domain knowledge. This person also collected the relevant data. Furthermore, the employee continued to Step 4.2, where certain data sources were combined into data that might provide a better model. WSW collected the structured financial data making it relatively easy to acquire. However, because the data contains information of organisations, which information is not public, there is chosen to anonymise the data.

8.2.6 Step 5: Data Preparation

The Data preparation step starts with checking the quality of the data, Step 5.1. The first impression of the employees was that the overall quality of the data was good but unbalanced. Noisy data was removed by the domain expert, which coincides with Step 5.2. The third step, Step 5.3, is changing the data format that is readable by Weka. First, the structured data was transformed into CSV. The CSV formatted data is readable by Weka, which converts this format to the workable Weka format, ARFF. The last step of the data preparation, Step 5.4, is to split the dataset into two sets: one set to train the model and one to test the model.

8.2.7 Step 6: Algorithm Selection

As described in Step 1.2 the goal of the project is to predict the nominal label (yes/no), which indicates if an organisation is a financial risk for WSW. This condition is identified as a classifying problem and therefore requires classification algorithms. Following the algorithm section method, provided with the method, five algorithms were identified to train the model. The five algorithms and that part of the algorithm selection method that provides the algorithms is visualised in Figure 41. The five algorithms are: Logistic Regression, Naive Bayes, k-Nearest Neighbors, Decision Trees, Support Vector Machines.

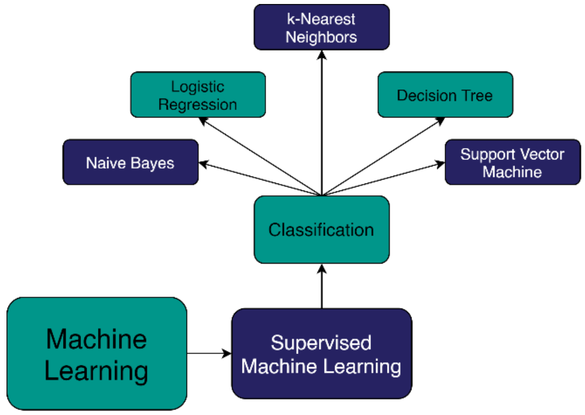

Figure 41. WSW Algorithm selection

8.2.8 Step 7: Model Building

In Step 7.1 the five selected algorithms are used with the standard settings of Weka, to train five different models. These models provided the following results where two evaluation metrics are displayed in Table 19. Accuracy, Sensitivity and Recall are the proportion of labels that are correct.

Table 19. WSW classification models

Algorithm	**Algorithm Weka**	**Correctly Classified Instances**	**Sensitivity (yes)**	**Recall (no)**
Logistic Regression	functions.Logistic	95.44 %	0.371	0.992
Naive Bayes	NaiveBayes	38.50 %	0.895	0.353
k-Nearest Neighbors	lazy.IBk	93.31%	0.308	0.973
Decision Trees	trees.REPTree	94.98%	0.329	0.989
Support Vector Machines	functions.SMO	94.02%	0.007	1.000

The following action, Step 7.2, is to test the models on the test data. This resulted in no significant changes in the results. Evaluating the results of the models, Step 7.3, is done by interpreting and comparing the evaluation metrics provided by the method. As the dataset is unbalanced, and it is important for WSW to identify the risky organisations (yes), multiple evaluation criteria must be advised. One of these important evaluation criteria is the sensitivity, displayed in Table 19, where the sensitivity identifies the proportion of (yes) labels that are correctly identified.

8.2.9 Step 8: Model Adjusting

In Step 8.1 the model is chosen that fits the goal best. As explained in Step 7.3, this depends on different metrics and interpretations. The employees of WSW chose the Naïve Bayes model as it was the model that performed best on sensitivity (yes), rather than the logistic regression, which produced the highest accuracy. The next step is to systematically change the parameters of the Naïve Bayes algorithm, which is Step 8.2. A description of the parameters can be found in Appendix B.1. The results of systematically changing the parameters of the Naïve Bayes, are displayed in Table 20.

Table 20. WSW Model Adjusting

Setting	Standard-setting	New setting	Increase accuracy	Increase Sensitivity (yes)	Increase Recall (no)
useKernelEstimator	False	True	28,18%	-0.105	0.306
numDecimalPlaces	2	2	0%	0	0
batchSize	100	100	0%	0	0
Debug	False	False	0%	0	0
displayModelInOldFormat	False	False	0%	0	0
doNotCheckCapabilities	False	False	0%	0	0
useSupervisedDiscretization	False	False	0%	0	0

Step 8.3 guides the user to apply ensemble machine learning. As elaborated in Section 3.4.3.9, in this step the method makes use of the ADABooster. However, the ADABooster was not successful in improving the model.

8.2.10 Step 9: Project evaluation

In the previous step, the final model was produced and therefore, the first action, step 9.1, is to evaluate the output of the model. It showed that changing useKernelEstimator provided a result interpreted as improving by the employees of WSW. Even with the sensitivity (yes) decreasing with 0.105, the accuracy of the model went up from 38.5% to 66.68%, which means the model is less accurate in predicting the risk cases. However, the model is now much better in predicting the (no) labels. Therefore, the total model increased in quality. The next step is to evaluate the results of the model with the present situation (t=0).

In Step 2 was stipulated that the current evaluation system of WSW is quite extensive. The capability of the produced prototype/model does not yet come close to the present evaluation model. However, it is important to realise a limited number of hours were spent on this first prototype model. In conclusion, at this moment, WSW does not foresee machine learning replacing the current method. However, the potential of machine learning at WSW is recognised, and will continue to experiment with machine learning.

8.2.11 Step 10: Communication

The last step of the method indicated the need to document Step 10.1, and communicate Steps 10.2 and 10.3. The documentation of the process is kept in-house at WSW, where all the conducted steps are noted, and data and models are stored. If necessary, the overall process can be reproduced. The second step is to communicate the results with the internal stakeholders. It has been agreed that the employees of WSW will communicate the results in the next “future of data in WSW” meeting. Therefore Step 10.2 is also conducted. The last step, uploading the process on GitLab is not viable for WSW as the data that is used is confidential. By implementing this last step, the method has been completed.

8.3 Conclusion

This chapter describes the demonstration of the in this thesis designed method. By applying the designed method to a real case study, this research follows the guidelines of DSRM. To demonstrate the designed method, the WSW-case was selected. WSW is allied to SBR-Wonen and fits perfectly within the context provided to use this method.

The demonstration of the method has been done in close collaboration with WSW. WSW employees, together with the researcher, went through all steps of the method to set up a machine learning project. In this project, a model has been developed that determines the financial risk of social housing associations. By developing the model, WSW would be able to understand the added value machine learning could provide. After completing all steps of the method, a prototype machine learning model was developed. The developed model did not yet come close to the capabilities of the current evaluation model.

The goal of WSW was to determine if machine learning has potential within WSW to support the primary business process. Where the model provided relevant insights in the decision making towards the identification of financial risk, it did not come close to the current system. Therefore, it could be concluded that at this moment machine learning will not replace the current systems. However, WSW will continue exploring the possibilities of machine learning as it was able to create relevant results in a short period and at low costs.

Summarised, the aim was to demonstrate that the method has been successfully applied to a real case, rather than just a fictional one. By demonstrating the method on a real case, it provides the research with empirical evidence and an example on how the method should be used. In the next chapter the designed method is evaluated including the demonstration of the designed method.

Chapter 9. Evaluation

First, the method will be evaluated in Section 9.1. First, the design objectives will be evaluated, followed by the evaluation of the total design. Afterwards, the application of the designed method is evaluated, taking into account the demonstration in the previous chapter. Hereafter, the evaluation of machine learning is SBR is discussed in Section 9.2. Then the theory behind building the method and machine learning in the SBR context is evaluated upon in Section 9.3. Followed by the limitations of the conducted research in Section 9.4.

9.1 Evaluation of the Method

The evaluation is an important step to observe how well the designed method supports the problem solution and is in this thesis done in three-fold. First, the evaluation of the design objectives is described. Second, the evaluation based on the demonstration on the WSW case is discussed. The final evaluation, Section 9.1, focusses on the evaluation criteria provided by the expert opinion during the semi-structured interviews.

9.1.1 Evaluation of Design Objectives

In Section 4.3, six design objectives were formulated. These design objectives were chosen to guarantee that the designed method will meet the desired functionality.

Design objective 1: The designed method should include an ethical framework

Based on the literature review as described in Section 3.5, several ethical theories were explored. The concepts defined by the literature have been partly incorporated into the method. Apart from these concepts, the Ethical Impact Assessment (EIA) framework is implemented in the method, providing a validated method for assessing the ethical impact of technical designs. Furthermore, if the user strictly follows the EIA framework, the ethical reliability of the machine learning project will be enhanced. Therefore, design objective 1 is incorporated. However, reading the literature and following the experiments, it is evident that ethical considerations are highly contextual and highly subjective. By including them in the design of the model, it may help to create a model that is relatively more ethical. However, ethics comprehend a broad domain, and it is not expected to cover the entire ethic field with one ethical framework. Furthermore, the question arises as to whether the public accepts an ethically produced machine learning model as machine learning itself might not entirely be accepted yet.

Design objective 2: The designed method should include machine learning steps to create a model in SBR context

In the literature, based on knowledge discovery in databases, several factors were found providing the steps for data analysis. These steps were specified for machine learning. Furthermore, the experiments provided detailed steps on what users of SBR context can do with machine learning. The combination of these two input criteria resulted in clear machine learning steps that contribute to the development of a successful machine learning model, incorporating design objective 2. Although the method provides clear technical steps, the method is limited to a specific group of users. The steps might be too difficult for someone with no technical experience at all and too easy for a machine learning expert. Furthermore, there are different approaches in creating machine learning models. This research focussed mainly on two different types of machine learning approaches (the

KKD process) to define the technical steps. Consequently, other potential machine learning approaches are not considered.

Design objective 3: The designed method should include a machine learning algorithm selection method, including multiple machine learning techniques

Based on the integration of the knowledge derived from relevant literature and the results from the conducted experiments and interviews, a method of machine learning algorithm selection has been developed. This method provides a clear overview of multiple machine learning algorithms based upon project-specific variables. These machine learning techniques were derived from the literature of the most frequently used machine learning algorithms and further tested in the experiments. Therefore, design objective 3 is incorporated. However, the designed algorithm selection method does not include all machine learning algorithms – for the scope of this thesis only frequently used algorithms were included. Subsequently, most of the included algorithms are transferable to non-experts. A risk is that this selection is at the expense of the model performance. This comprehensive selection of included algorithms adds to the overall transparency and was deemed appropriate for this method.

Design objective 4: The designed method should include organisational factors relevant for creating a machine learning project

The need for organisational guidelines was established in Chapter 1. The literature survey resulted in the selection of the Strategy Map as a proven method, providing academic guidelines for including organisational factors as part of the method. Second, the Strategy Map was created on the strategy "using machine learning", which provided input from the users' perspective. These two inputs were combined and provided the final organisational guidelines in the method. However, there are other organisational factors and methods that may be relevant. Due to the limited timeframe of this thesis, only the Strategy Map is investigated and implemented. It has not been possible to consider other options within the organisational context. Within the scope of this thesis, design objective 4 can be considered as incorporated.

Design objective 5: The designed method should provide an understandable process for creating a machine learning project in SBR context

The designed method provides a clear overview of the technical (Design objective 2 and 3), ethical (Design objective 1) and organisational input (Design objective 4) and thus providing the guidelines for setting up a machine learning project in SBR context. This has been tested in the experiments, evaluated by interviews and demonstrated in the WSW-case. Furthermore, describing the process clarifies possibilities of machine learning using SBR data. The interviewees rated the understandability of the method a 4.2 out of 5, this is further elaborated upon in Section 9.1.3. Therefore, design objective 5 is adequately elaborated upon. However, the combination of these aspects makes the method applicable to a very specific group of users and reduces the usability to both managers and engineers as it is more generalised. In addition, the designed method is a prototype. Additionally, the method provides only a small portion of all machine learning opportunities. The suggestions as output of the method, are mostly directed to SBR stakeholders, as they are the main stakeholders. Therefore, the method does not provide an understanding of the overall field of machine learning: it provides limited insight. Elaborating upon more domains within machine learning was not realistic within the timeframe of this thesis.

Design objective 6: The designed method should help decision-makers to understand if machine learning can create added value in their organisation
The method must provide clear insights for the user, whether machine learning, when used accordingly, can create added value. The design objectives provided a method that can be implemented systematically and efficiently within an organisation. By using the method, the organisation obtains insight whether machine learning in the SBR context can create potential c.q. added value for their organisation. However, the method provides only a small and relatively easy setup of machine learning capabilities, only the tip of the iceberg. There are many other ways to explore the potentials of machine learning in the organisation. However, providing all the options would make it more difficult to follow the method and might require a higher level of expertise. Furthermore, this does not fit within the timeframe of this thesis. Therefore, design objective 6 is completed for the scope of this thesis.

Evaluation of design objectives
The evaluation of the design objectives in the previous section can be explained in different ways. In general, the objectives are all incorporated in the designed method, and ,therefore, technically completed. However, in this thesis, "completed" does not mean that they do not require further work. It should be clear that only a few options were investigated when giving substance to the design objectives, as can be expected within a timeframe and focus of a master thesis. For example, several ethical frameworks are available, but not all frameworks have been investigated in this thesis. The same goes for organisational factors and technical factors such as machine learning algorithms.

The design objectives are quite extensive. When formulating the objectives, it was evident that they were ambitious. From the start of developing the design objectives, the goal has been to incorporate all these requirements. However, the objective was not to completely fulfil the objectives, because it was recognised that it is almost not possible to fulfil all these objectives in one project, especially within the scope of one thesis. Even for one project with multiple subprojects, it is an ambitious goal to achieve a fully proven and implementable method. Not only numerous cases and design iterations are needed to create scientific substance, but also extensive expert opinions on the different perspectives stipulated in this thesis, which correlate with design objectives 1, 2 and 3. There are different ways to analyse the ethical, organisational and technical inputs of this method. When setting these first steps towards the fulfilment of each design objective, it became evident that in order to combine these specific factors, it was necessary to prioritise within content and context. This developed prototype method has met the first stage of the design objectives.

9.1.2 Evaluation based on the demonstration
The demonstration has proven the applicability of the method by an organisation in SBR context. Policymakers of WSW obtained the insights needed to start the machine learning project. It also gave the engineers the guidelines to go through the technical aspects of the method. All steps have been found to be effective and useful, evaluated by the WSW project members. Furthermore, consistently following the steps has provided users with insights into what is the potential value of machine learning in their organisation: it triggered a WSW employee to investigate opportunities to develop machine learning knowledge and skills, in anticipation of adjusting the desired HR profile of future employees. However, as the researcher of this thesis was part of the WSW project team, research bias cannot be excluded, since it is difficult to say if the team would have been able to complete the method without the researcher. The WSW project members did, however, provide a positive evaluation of the method, including its understandability. The Demonstration shows that the method is applicable

and thereby fulfilling the expectation of this thesis. However, the prototype needs testing to further evaluate its' performance in real-life cases.

9.1.3 Evaluation based on interviews

During the semi-structured interviews, the interviewees were asked to give their expert opinion on the designed method. The evaluation criteria were carefully extracted from the literature, describing the most promising evaluation criteria, as is discussed in Section 6.2.2. The expert opinion of the interviewees was coded. Based on this input, the mean and standard deviation was determined. After analysation and aggregating the data, this resulted in the overall mean and standard deviation of the seven evaluation criteria as visualised in Table 21. The first criterium elaborates on the effectivity of the designed method on producing a machine learning project considering the organisational, ethical and machine learning factors. This aspect was graded with a 4.5, which is translated into useful/very useful. The second evaluation criterium analyses the utility the method provides and is graded with a 4. Third, all experts found the designed method understandable, relating to grade 4.2. The fourth criterium is the completeness of the method; the interviewees judge this with the grade 3.3, belonging to neutral. The reason for this qualification is as it is difficult to interpret if the method is, or ever will be, fully completed, as has been evaluated upon in the previous paragraphs and will be evaluated upon further in this Chapter. The fifth evaluation criterium refers to the robustness of the method, which indicates if the designed method provides clear guidelines for building a machine learning project in a different situation. This was graded with a 3.8. The last two evaluation criteria were only graded by the interviewees who were taking part in the experiments. These show that the accuracy of the experiment (3.7) is in accordance with the method since the method is developed based on the experiment. Furthermore, the method was evaluated to fit the organisational context of SBR (4.0). Overall, the evaluation was positive. However, some interviewees emphasised that the method is applicable only for particular situations in SBR context and that further development is desired.

Table 21. Evaluation results interviews

Criteria	Mean	Standard deviation
Effectivity	4.5	0.8
Utility	4.0	0.9
Understandability	4.2	0.8
Completeness	3.3	0.8
Robustness	3.8	1.0
Accuracy experiment (SBR)	3.7	0.6
Fit organisation (SBR)	4.0	1.0

9.1.4 Why is this method better than no method?

By developing a method that is understandable for managers as well as technicians, a synergy between these different stakeholders can be achieved. This method provides managers with tools to assess what they need for and what they can do with machine learning in their organisation in line with the strategy of the organisation. Without this method, they would not have had guidance on how to start setting up a machine learning project or hire an external expert to explore their options. Hiring external experts is not only expensive but more importantly, an employee with enough insight into machine learning might be able better to estimate what the company needs. In addition, the method also offers in-company-tools for making a prototype with minimal machine learning knowledge (minimum knowledge is defined in Section 7.5). The developing process itself is part of the learning curve of the individual employees as well as of that of the organisation.

This prototype model, as a result of implementing the designed method, can be set up by employees of the company who have some knowledge as described above. As a result, a limited pilot version can already be used to explore whether machine learning has any potential in the company. This approach gives the manager the opportunity to take further actions based on information of a relatively simple pilot. For the technician, the method provides the tools to create a "simple" machine learning prototype. As a result, the technician does not have to be a machine learning expert, and the prototype can be built with relatively little effort. Although this latter model is nothing more than a first prototype, it does, however, provide insights into whether machine learning has enough potential for a company. The main aim of the method, for now, is therefore to make prototype models, the test phase. Inherently, the outcomes are also a prototype.
From the researchers' perspective, it is important to emphasise that the outcomes are also a prototype, since the method is not finished yet. Care while using machine learning and interpreting the results of machine learning projects, is essential.

Placing the method into context with other comparable methods is difficult since no other method combining the three aspects can be found. When compared to methods only including one aspect, it should be expected that this designed method may be less accurate on one individual aspect, than the other method to which it is compared. However, this is the price paid when combining essential factors. The consideration needs to be made how results of projects derived with methods focused on one specific aspect should be interpreted, especially since the design process elucidated the complexity of setting up machine learning projects while taking into account multiple aspects.

The development of the designed method has a societal contribution since it provides a prototype of a machine learning project method combining certain technical, organisational and ethical aspects. This is summarised in that the European Data Portal (*Digicampus and Data | European Data Portal*, 2020) has already picked up on the development of this method, confirming the need for a machine learning project method including aforementioned factors, as also established in Chapter 1 of this thesis.

9.2 Evaluation of Machine Learning in SBR

One of the reasons that machine learning projects fail, is that machine learning as a method is not necessarily the right fit for solving a specific problem. A contributing factor is that machine learning in general is often not sufficiently understood and therefore not applied at the right moment or in the correct context.
In the SBR context, it becomes clear that machine learning certainly has potential. For successful application of machine learning, structured data and clear variables are contributing to creating proper machine learning models. Knowledge about the SBR domain is essential for using machine learning in the SBR context since SBR is a complex domain.

However, the experiments show that for implementing machine learning in the SBR context, the conditions are still not optimal and therefore not ready yet to replace the current systems. Contributing to this statement is the question of whether machine learning is a reliable approach to tackle the problem if results are within a desirable margin. Furthermore, the social scepticism on machine learning is still widely existent.

In the situation of SBR, one could argue that specific knowledge of the application of machine learning is not fully available yet. However, experimenting and building prototypes is currently the right way. When the techniques are further developed and socially more accepted, and they are also applied in a responsible and decisive way, machine learning can certainly contribute to a good basis of data analysis. It also forces companies to think more strategically about how and which data to collect in order to obtain the right ingredients to build a sound model. A noted bottleneck is that the current databases are relatively limited and therefore at this moment, an additional limiting factor to build a reliable model.

9.3 Evaluation on the theory behind the research

This master thesis has delivered multiple contributions to practice. The research developed a new practical method that aims to enhance the understanding of machine learning projects in a SBR context and provides the organisation a basis for action. This method, which combines technical, ethical and organisational aspects in a systematic approach, enables its users to accumulate knowledge of the added value of machine learning. The integration of these three pillars into a single method has not yet been available.

Next to the practical contribution, the research also contributes to the theory development on the use of machine learning on structured data. The scientific contribution is found in the identification and conceptualisation of ten steps (including the sub-steps) that influence the use of machine learning on structured data. These steps are derived from the three selected pillars of technology, organization and ethics. To the best of the researcher's knowledge, no literature exists considering all steps (together) in order to provide a theoretical interpretation of the method. The integration of these three pillars in a systematic approach and into a single method, is not yet available and/or has not been published.

There are different perspectives on what defines a theory. Gregor (2006) summarises the various theoretical perspectives as “abstract entities that aim to describe, explain, and enhance understanding of the world and, in some cases, to provide predictions of what will happen in the future and to give a basis for intervention and action”. Gregor further elaborates that theories “are practical because they allow knowledge to be accumulated in a systematic manner and this accumulated knowledge enlightens professional practice” (Gregor, 2006).

Five different types of IS theory distinguished by Gregor (2006, p. 611) which are: “(1) theory for analysing, (2) theory for explaining, (3) theory for predicting, (4) theory for explaining and predicting (EP theory), and (5) theory for design and action”. The different types of theory are interrelated and some comprehensive, well-developed theoretical bodies could include components from all the types of the theory discussed.

The theory developed in the research fits within the description of the design and action theory type. This theory type “gives explicit prescriptions for constructing an artifact” (Gregor, 2006, p. 620). In this research, the theory (designed method) sets out prescriptions to construct an artifact (Machine learning model).

This can be seen as an extension of previous contributions by Fayyad et al. (1996) & Han et al. (2011) on Knowledge Discovery in Databases, Reijers et al. (2016) and Wright (2011) on the Ethical Impact Assessment and Kaplan & Norton (2004) on the Strategy Map. These main contributions helped to fill the gap described in Section 1.3.1

9.4 Limitations of conducted research

The developed Machine Learning Project Method shows strengths and limitations. The limitations are elaborated upon and put into perspective in the following paragraphs.

The method is developed for a specific context, namely that of SBR. It is not clear whether it can be applied in another context. Furthermore, one of the applications is that SBR data is highly structured. It is not analysed whether this method can be used on other types of data. Pre-processing unstructured data to structured data was not necessary for these experiments. Although in theory, this method could be used on other structured data, this has not been tested in this thesis.

Out of four machine learning techniques, two are included in the designed method: the supervised and unsupervised machine learning techniques. The reinforcement and semi-supervised techniques are not part of the methodology since this would not fit in the scope and timeframe of this thesis. Both techniques have potential in the SBR domain of machine learning. Of the two types of machine learning techniques used in the method, only supervised machine learning was tested using regression and classification algorithms. Unsupervised machine learning, clustering and association are not conducted in this research due to time-management. Therefore, within this part of the methodology, further experiments with these machine learning techniques are required before generalising the use of this method. However, within the scope of this thesis, three experiments were conducted: two experiments on the classification problem and one experiment on the regression problem, creating a prototype of a method that needs more experiments to gain further support.

A complete, formal Ethical Impact Assessment is not executed in this research. However, it is part of the method, and the subject itself has been evaluated with the experts and included in the guidelines. The EIA has been chosen as a format to analyse the potential ethical impacts. However, other ethical models could be applicable as an ethical format for the method.

The prototype method is built on literature, experiments and interviews. A limitation of the literature search is that multiple other frameworks and contributing factors, for example, for successful projects could have been investigated, although not within the timeframe of this thesis. This applies to ethical aspects, organisational aspects and technical aspects. For determining the technical aspects for the method, Knowledge Discovery in Database was chosen. There might be other methods from which machine learning steps can also be derived. However, this was not explored since an adequate option was found. This also accounts for the organisational and ethical aspects. There might be other literature applicable for integration in the designed method.
A limitation of the experiments is that three experiments have been conducted. For building a prototype within the context of a master thesis, this may result in interesting outcomes. However, this is not sufficient for creating a fully operational and validated method. The findings of this prototype method can be used with care.

A limitation of the interviews from the researcher's perspective is that the positive evaluation might be biased because of professional connection to the interviewees. Furthermore, even though fitting in the scope of a thesis, six interviews were performed, and the sample size of both the number of organisations and of interviewees was relatively small. Also, no ethical expert was interviewed, and thus, this specific expert perspective is missing. Furthermore, one round of interviews was conducted in which several subjects were discussed. As a result, implementing user feedback took place only once.

The designed method is developed by one researcher only. The researcher conducted the literature review, the experiments as well as the interviews and thus could be an unknowingly potential influencing factor. However, the research was systematically produced following a validated research methodology, thus providing adequate fail-safes reducing potential research bias.

9.5 Conclusion

In this chapter, the research conducted in this thesis and the designed method is evaluated and reflected upon. The evaluation of the method is the second to last step of the DSRM and divided into three parts.

First, the designed method, including the design objectives, have been evaluated. In summary, the design objectives are essentially completed. Although the design objectives were ambitious, they are all incorporated in the designed method. However, in order for the prototype method to become a fully operational and validated method, the design objectives should be further investigated and developed.

Second, the application of the method on the hand of the demonstration phase, is evaluated. Supported by the interviews, the designed method has been effective and found usable in setting up a machine learning project in a real case. Furthermore, machine learning in SBR context is put in perspective. Although this research recognises the potential of machine learning in SBR context, the experiments show that for implementing machine learning in the SBR context, the conditions are still not optimal and therefore are not ready yet to replace the current systems. However, the method has proven to be reliable in achieving a machine learning project with the usable outcome and is presented at the European Data Portal for implementation via GitLab.

Last, six limitations of the research are found. The six limitations are: The method is developed for a specific context, namely that of SBR. It is not clear whether it can be applied in another context. Secondly, out of four machine learning techniques, two are included in the designed method: the supervised and unsupervised machine learning techniques. Only supervised machine learning has been tested during this research. Third, no complete Ethical Impact Analysis was conducted during the research. Fourth, the method is compiled out of selected literature relevant for the thesis. However, other literature might be applicable as input for the method. Fifth, only three experiments with machine learning were conducted, more experiments are needed for creating a fully operational and validated method . Sixth, no ethical expert was interviewed during the round of interviews, therefore missing an ethical experts' opinion on the method. The above-mentioned elements of the evaluation together ensure that the method has been appropriately evaluated, so that the evaluation contributes to the added value of the method, and answers sub-question 3 and sub-question 4.

Part IV
Placing the method in context

Chapter 10. Conclusion

In this chapter, the content of this thesis is unified in order to draw conclusions, place these into perspective and ultimately give recommendations. First, the main research question is answered, and the main product of this thesis, the designed method, is presented in Section 10.1. Second, in Section 10.2, the method is put in a final context, including the contribution of the research and its generalisability. Third, in Section 10.3, recommendations for further research are made. Fourth, the reflection of the research on machine learning in SBR is given in Section 10.4. Finally, the link between research and MoT is described in Section 10.5.

10.1 Answering the main research question

Recapitulating the research objective
The aim of this research was to design a method that helps organisations, in particularly SBR-stakeholders, set up machine learning projects with respect to organisational, ethical and machine learning factors. As stated previously in Chapter 1, the potential value that machine learning can provide is enormous. However, organisations struggle with creating successful machine learning projects. Furthermore, guidance towards understanding how machine learning can bring added value to support their corporate strategy that is called for. Therefore, this method should provide decision-makers with the ability to determine if machine learning can provide added value to their organisation. The main research question for this research was formulated: *"How can technical, organisational and ethical aspects be combined into a scientific method that supports stakeholders to systematically set up machine learning projects in SBR context?"*

Building the conceptual model of the Method
Based on the formulated research questions and sub-questions (Chapter 1), the chosen structure of the research was according to the Design Science Research Methodology (DSRM) (Chapter 2). Together with the extensive literature review (Chapter 3) and the baseline data experiments (Chapter 4), six design objectives were formed and evaluated in detail in Chapter 9. Evidence supported that in the development of a successful Machine Learning Project Method as part of the research question, criteria as concepts data, knowledge discovery in databases, machine learning, ethical frameworks and organisational frameworks are indispensable and therefore crucial. These criteria were implemented in the method for a machine learning project in Design phase 1 (Chapter 5).

Optimising the Method based on expert interviews
Method validation was conducted via expert interviews. Clear interview objectives were formulated and translated to an interview protocol, which contains both qualitative and quantitative sections. A selection of relevant expert interviewees was made, and six semi-structured interviews were conducted. The interviews provide sufficient input for optimising the Machine Learning Project Method in Design phase 2 (Chapter 7). Furthermore, the interviews provide the input for the evaluation of the functionality of the overall method (Chapter 9).

10.1.1 The Research outcome: Methodology of the Machine Learning Project Method: 10 steps

The three sources crucial to answer the research question – the literature survey, the two experiments and the six interviews – provide sufficient input for the design of the Machine Learning Project Method following the DSRM method. For this project method, tailor-made for an SBR environment, ten unique steps, including multiple sub-steps, were identified and merged into a method, giving guidance to the implementation of machine learning projects. A concise description of the final research product, the Machine Learning Project Method, is provided in the following section and is visualised in Figure 42.

<u>Step 1: Goal formulation</u>: A simple yet crucial first step. Proper project goal formulation is critical, especially in machine learning projects. Correct goal correspondence with machine learning outcome not only enhances technically successful machine learning projects, but it is also essential for creating optimal fit with the company's mission and thus successful implementation from a managerial perspective.

<u>Step 2: Project team setup:</u> Project teams have a standard structure. For setting up successful machine learning projects, standard prerequisites are defined: the team should include a person with (limited) experience with machine learning and one with data access and data user experience. Especially since machine learning is a technical aspect with managerial consequences, management should be included.

<u>Step 3: Context analysis</u>: Context setting is important for framing the work field and setting up clear begin- and endpoints. Stakeholder analysis and analysation of the current goal situation provides intelligence on attachments points within the organisation for an excellent fit of the project. Ethical impact analysis is conducted in this step and provides important guidelines to set up an ethically justified project.

<u>Step 4: Data collection</u>: Compiling optimal datasets of structured SBR-data, facilitates the necessary input for the machine learning project. How the data is accumulated provides extra opportunities for analysation; feature engineering optimises data sets, thus data input, and consequently, data output.

<u>Step 5: Data preparation:</u> Focus lies on the technical aspects of the data. Via checking the quality of the data, cleaning up noise and searching sub-steps, the data is prepared for better analysation. In this part the data is split into the training and the test set and facilitates a safeguard for overfitting the data.
<u>Step 6: Algorithm selection:</u> This step fills the void of an absent framework that combines different machine learning techniques. An algorithm selection method, including different machine learning

techniques, is created for this method. It provides the user with clear guidelines regarding which algorithms could be used best in the particular context: based on the project goal set up in the previous steps, it provides a custom-made advice on which algorithms to use.

Step 7: Model building: The algorithms train different machine learning models based on the dataset. All these models have different forms of outputs. The outputs are evaluated on model evaluation metrics.

Step 8: Model improvement: The model that performs best is selected. Improving the model with different improvement techniques is explored: systematically changing the parameters of the algorithm and ensemble machine learning.

Step 9: Project evaluation: The final model is evaluated and compared with the results of the current situation. Analysis whether there is a need for a machine learning expert is done. The method can be completed without an expert; however, care is advised, because this influences the quality of the model and thus the outcomes.

Step 10: Communication: Communicating the results and implications of the project with the internal stakeholders is essential for implementation and for reproducibility. External communication of the results on a platform such as GitLab is optional.

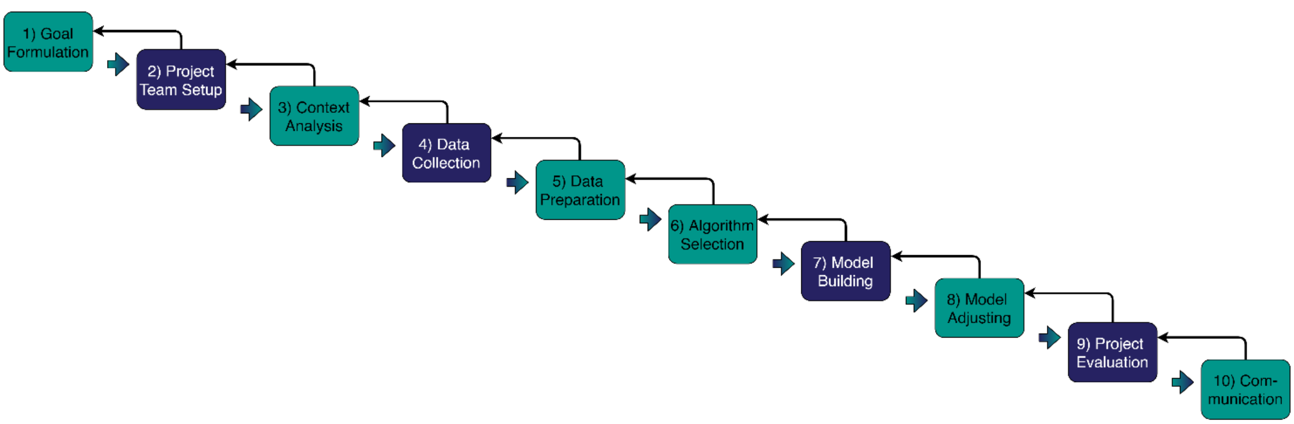

Figure 42. Concise version Machine Learning Project Method

Proof of concept

The burden of proof of the research question lies in developing a Machine Learning Project Method as the main deliverable. It can be concluded that the research question of this thesis is successfully answered and tested in an experimental environment. Furthermore, the experiments provided SBR stakeholders with insights on machine learning and implementation. The designed method has proven to be reliable in achieving a machine learning project with usable outcome and is presented at the European Data Portal for implementation via GitLab. However, the experiments show that for implementing machine learning in the SBR context, the conditions are still not optimal and therefore are not ready yet to replace the current systems. From the researchers' perspective, it is important to emphasise that the outcomes are also a prototype since the method is not finished yet. Care while using machine learning and interpreting the results of machine learning projects is essential.

10.2 The final context

10.2.1 Scientific contribution

In Chapter 9.3, Evaluation on the theory behind the research, the framework of the theoretical foundation of the research and the developed method is discussed. The formulation of this theoretical contribution (in addition to the practical contribution) goes hand in hand with the related scientific contribution and academic relevance. The scientific contribution of the method developed in this thesis comes in threefold.

The literature review resulted in the identification of a selection of influencing steps crucial for substantiating and elaboration of the first part of the research question, were the three aspects technical, ethical organisational are identified as leading factors. Exploring the literature regarding these three factors resulted in identification and conceptualisation of ten unique steps (including the sub-steps): goal formulation, project team setup, context analysis, data collection, data preparation, algorithm selection, model testing, model adjusting, project evaluation, communication. These found steps influence the use of machine learning on (SBR) structured data.

The in this thesis developed method, combining and integrating technical, ethical and organisational aspects into a single method, was not yet available. Therefore, this method fills the gap that other methods left open. At this moment, it is unique in comparison to other methods, as it combines interdisciplinary aspects into one method. The newly created method is added to the scientific field and is shared to facilitate further research and development (Digicampus and Data | European Data Portal, 2020). This can be seen as an extension of previous contributions by Fayyad et al. (1996) & Han et al. (2011) on Knowledge Discovery in Databases, Reijers et al. (2016) and Wright (2011) on the Ethical Impact Assessment and Kaplan & Norton (2004) on the Strategy Map.

Last, this research contribution comes from the experiments which enhanced the understanding of machine learning projects in SBR specifically but can be extended to research into machine learning on structured data in general. Evaluating the results of this research, it can be concluded that machine learning, at this moment, is not yet capable of generating the desired application and outcome in this specific context. However, by more experimenting, building prototypes and learning by doing, this might result in a reliable technique. When the techniques are further developed and socially more accepted, and they are also applied in a responsible and decisive way, machine learning can contribute to a sound basis of data analysis in SBR context.

10.2.2 Managerial contribution

The method provides a structured and partly iterative process to set up machine learning projects. Via following the machine learning project method, a machine learning model can be made with respect to organisational, ethical and technical aspects. This allows the user of the method to evaluate the added value of machine learning in the organisation: it guides users towards asking the right questions, also making them aware of the limitations and impossibilities of machine learning. This method prevents initiating a machine learning project without estimation of the applicability. Furthermore, it gives managers, policymakers and engineers an overview of what it takes to start a machine learning project, including preconditions and restrictions. It provides insight into possible applications of machine learning and enables a structured process for both engineers and managers, creating alignment and understanding between management and engineers. When all steps are completed, the method provides the following deliverables: insight into whether machine learning has added value for the organisation and an ethical, potentially cost-efficient, yet simple, prototype machine learning model.

Including ethical and organisational factors at the beginning of setting up a machine learning project method, results in early consideration and awareness regarding the correct fitting of organisational and ethical standards by the users of the method. This results in fitting outcomes of the machine learning project method, necessary for successful implementation of the project, and important for the stakeholders. The importance and practicality of the management involvement while setting up the machine learning project, is once again emphasised by the previously mentioned issue regarding the Dutch Tax organisation and child allowance. A clear example where machine learning is used to detect fraud, but the consequences of errors in the model were failed to be recognised and consequently gave more weight to a discussion about the ethical impact. This structured method helps managers understand the process of setting up machine learning projects and gives them guidelines on how to set up a successful project. Interdisciplinary domains are integrated: organisational and ethical aspects, and guidelines on managerial implementation.

10.2.3 Translational relevance

By creating a method that is accessible for the public, transparency is being created on how machine learning is being applied. Especially in the abstract world of Artificial Intelligence, the creation of a clear method can contribute to a better understanding of and more faith in machine learning in SBR by general society.

Implementing successful Machine learning projects is dependent on two domains: engineering and management. Therefore, it could be stated that Machine learning projects fall into a translational domain. For the development of the Machine Learning Project Method, both engineering and management aspects were integrated. The structured method creates alignment between and understanding for both management and engineers, a synergy aspired by true management of technology.

10.2.4 Generalisability

This research gave guidance to the theoretical basis for optimising data analytics in SBR context with machine learning. By the development based on scientific literature and a working prototype from two practical experiments, in close collaboration with the relevant stakeholders, the original academic question led via applied research, to a concrete working method and insights into the use of machine learning in SBR.

“A researcher can take a newly formulated and untested theory that has the property of generalisability (i.e., it is generalisable, but has not yet been generalised) and then (2) the researcher, in examining it empirically, can demonstrate its generality” (Baskerville & Lee, 1999, p. 51).

Whereas the designed method demonstrated to be able helping its user to set up a machine learning project in SBR context, it has the property of generalisability towards other cases to the wider domain of machine learning on structured data. Furthermore, the implementation ethics in the designed method contributes the field of ethical machine learning.

The same goes for the insights regarding machine learning in SBR context, as the two experiments show that machine learning in SBR is at this moment not ready to replace the current systems, this insight has the property of generalisability towards other cases where machine learning is used on structured data. However, as it has the property of generalisability, two cases do not provide enough evidence to demonstrate its generality.

The designed method provides the first step towards optimising machine learning projects in SBR context with regard to technical, organisation and ethical aspects. Furthermore, by facilitating transparency and sharing the method, experiments and data, the input is provided to optimise the method and implementation of machine learning in SBR context. This research contributes to the generality by uploading the Machine Learning Project Method onto GitLab and the TU-Delft repository. This facilitates other organisations to do experiments on their own, which is needed for further research and development. This contributes to generating more data, which is needed for translation of the method for general application.

10.3 Recommendations for further research

As indicated in the previous section, the designed method provides the first step towards optimizing machine learning projects in SBR context regarding technical, organization and ethical aspects. The method developed in this thesis therefore fills a niche in the current (knowledge) gap reviewing the application of machine learning within SBR in line with the research question. Considering this demonstrable added value, it is useful to further elaborate and apply this method. Recommendations for future research can be summarized in three-fold.

Improving the method

Further focus on fine-tuning of the methodology starts with expanding the literature research regarding the three main aspects which are, ethical, organisational and technical aspects and considering more approaches. Furthermore, with the aim on creating a fully tested and operational model, more cases must be analysed. For finding more cases, GitLab could be a good source of relevant data as the users are encouraged to experiment with the method and upload the full process of creating a machine learning experiment. Extending the literature research and testing the method on more cases, will be an important next step to translate this prototype into a reliable model for a professional, operational organisation use.

Applying the method on different machine learning contexts

To further extend the societal relevance, this method, with the potential for a more generic application, should be explored. Within this research a limitation has been imposed to limit the development of the method at SBR. It is advised to expand beyond SBR to explore if the method is applicable to a broader field of structured data. The promising generic part of this model should be conceptualised in order to broaden the concept for machine learning in a wide scale of algorithms and applications. This could ultimately contribute to uniformity in the use of machine learning.

Ethical considerations

Machine learning and ethical issues go hand in hand. The risks this entails is still underestimated in operational applications. Further research into the possible negative impact of machine learning on society must be conducted. Therefor it should be researched how organisations and society can be included in the process of constructing machine learning.

10.4 Reflection on the future of Machine Learning

At the end of the research, it makes sense to reflect, according to the researcher, on what the future of Machine learning looks like. Particular attention is paid to the technical possibilities that the researcher recognises in machine learning and what type of machine learning problems need to be solved.

It might be concluded that based on the gained insights and knowledge during the research, machine learning can be seen as a promising technique. It shows enough potential to justify further research and development, and not only in the domain of analysing structured data. Zooming in on potential, this added value restricts itself in this thesis by focussing mainly on the possibilities that machine learning offers. It can be said that in many areas, machine learning provides better results (for instance by sophisticated analytics) and is faster and cheaper compared to other data-analysis techniques. In a theoretical optimal situation, a fully automating of data research and analysis, with no change of human errors during the data analysis, sounds almost too good to be true. Therefore, a first restriction is that we are generally unable to develop machine learning models that are perfect c.q. do not make any mistakes at all. Who is to blame when a mistake is made, the developer, the owner, the stakeholder, the user? It is often thought, too easily, that the results/output no longer needs to be thoroughly checked as the output is produced by a model, instead of human action. A major fallacy which has regularly led to misunderstandings and, therefore, the question can be asked whether it is socially accepted that “decisions are made by a machine” as long as the (end)user does not take full responsibility for analysing the outcome of the model. In can be concluded that, additional to the question whether machine learning has much to offer in terms of technical potential, it is certainly not yet fully developed. Another restriction is that the accuracy of machine learning is often at the expense of explainability, the ongoing quest for finding the balance between the expert developer (best technical model) and the end user/stakeholder (applicability, broad support and a tool in line with the strategy of the organisation).

In the researcher’s opinion, apart from the quality of the output itself, the biggest problem right now is the (unforeseen) ethical aspect of the implementation of machine learning. Even if machine learning is better at certain tasks, it still needs to be guided and accepted. Sufficient support should be created and limits should be set on the parameters used for the analysis. Using machine learning but without filtering on ethical concepts, undesirable causal relations and output can be produced. Therefore, it is recommended that strict guidelines should be developed for the use of machine learning. It is also necessary to investigate what the society thinks of machine learning and how it can be accepted; insight ultimately leads to acceptance. The researcher believes the government should take the lead in developing these guidelines. They should provide an environment in which society, organisations and government can provide input into what ethical machine learning must involve.

10.5 Management of Machine Learning - Relating to MoT perspectives

The Machine Learning Project Method moves on the interface of technology and administration. The presented method in the thesis, where technical and organisational aspect and ethical issues go hand in hand, both demands and stimulates direction acquiring vision on the application of machine learning techniques in organisations. The method supports the management of an organisation to be proactive in the continuous reflection and update of knowledge processes within its' organisation in line with corporate strategy.

Recommendation for the director of MoT

During this research it has become evident that the current playing field of machine learning, including all different aspects that influence machine learning projects such as organisational settings, ethics and regulations by the government, is a perfect fit with the curriculum and goals of a MoT student. Therefore, it is advised that MoT should include machine learning in the curriculum of the MoT program as it will challenge its students with an exciting, complex and relevant field of expertise.

During the research and the attended congresses, the researcher noticed that there is a great demand for guidance on machine learning, both from the business community as from the government. As a MoT student, there is a clear added value in including machine learning in the present curriculum. It is a relevant research area asking an on-going policy development in the coming years. As MoT student you should be able to maintain a solid academic overview and give guidance to and participate in research of the crucial core pillars, such as technology development, organisational aspect s and ethics in general.

Part V

Appendices and reference

References

Adadi, A., Berrada, M., Chenouni, D., & Bounabat, B. (2015). Ontology based composition of e-Government services using AI Planning. *2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA)*, 1–8. https://doi.org/10.1109/SITA.2015.7358430

Allan, K. (2017, November 21). *How many businesses actually use machine learning? | IDG Connect*. https://www.idgconnect.com/idgconnect/analysis-review/1001644/businesses-actually-machine-learning

Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., Naidich, D. P., & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, *25*(6), 954–961. https://doi.org/10.1038/s41591-019-0447-x

Baskerville, R., & Lee, A. S. (1999). Distinctions among Different Types of Generalizing in Information Systems Research. In O. Ngwenyama, L. D. Introna, M. D. Myers, & J. I. DeGross (Eds.), *New Information Technologies in Organizational Processes* (Vol. 20, pp. 49–65). Springer US. https://doi.org/10.1007/978-0-387-35566-5_5

Batarseh, F. A., & Yang, R. (2018). Making the Case for Artificial Intelligence at Government: Guidelines to Transforming Federal Software Systems. Guidelines to Transforming Federal Software Systems. In *Federal Data Science: Transforming Government and Agricultural Policy Using Artificial Intelligence* (pp. 41–51). Scopus. https://doi.org/10.1016/B978-0-12-812443-7.00004-1

Bharosa, N., van Wijk, R., & de Winne, N. (2015). *Challenging the Chain: Governing the Automated Exchange and Processing of Business Information*. Ios Press.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. Scopus. https://doi.org/10.1023/A:1010933404324

Brownlee, J. (2016). Machine learning mastery with Weka. *Ebook. Edition: V. 1.4*.

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Henke, N., & Trench, M. (2017). Artificial intelligence: The next digital frontier. *McKinsey Global Institute*, 1–80.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A Library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3). Scopus. https://doi.org/10.1145/1961189.1961199

Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018). Notes from the AI frontier: Insights from hundreds of use cases. *McKinsey Global Institute*.

Cohen, P. (2018, March 16). How the History of Data Gathering Lead to the Age of Big Data & Personalization. *SmartData Collective*. https://www.smartdatacollective.com/history-data-gathering-lead-age-of-big-data-personalization/

Davis, J., & Nathan, L. (2015). Value Sensitive Design: Applications, Adaptations, and Critiques. *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, 11–40. https://doi.org/10.1007/978-94-007-6970-0_3

Digicampus and data | European Data Portal. (2020, July 7). https://www.europeandataportal.eu/en/news/digicampus-and-data

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, *17*(3), 37–37. https://doi.org/10.1609/aimag.v17i3.1230

Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value Sensitive Design and Information Systems. In N. Doorn, D. Schuurbiers, I. van de Poel, & M. E. Gorman (Eds.), *Early engagement and new technologies: Opening up the laboratory* (pp. 55–95). Springer Netherlands. https://doi.org/10.1007/978-94-007-7844-3_4

Gibney, E. (2020). The battle for ethical AI at the world's biggest machine-learning conference. *Nature*, *577*(7792), 609. https://doi.org/10.1038/d41586-020-00160-y

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, 611–642.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10–18. https://doi.org/10.1145/1656274.1656278

Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. https://books.google.nl/books?hl=nl&lr=&id=pQws07tdpjoC&oi=fnd&pg=PP1&dq=data+mining+concepts+and+techniques&ots=tzLtYVkCZX&sig=KZFU16u7Rzt4VL9QpbjGgQywrMk&redir_esc=y#v=onepage&q=data%20mining%20concepts%20and%20techniques&f=false

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). *Design Science in Information Systems Research*. 32.

High, P. (2017). *Carnegie Mellon Dean Of Computer Science On The Future Of AI*. Forbes. https://www.forbes.com/sites/peterhigh/2017/10/30/carnegie-mellon-dean-of-computer-science-on-the-future-of-ai/

Janssen, M., & Kuk, G. (2016). Big and Open Linked Data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 3–13. https://doi.org/10.1080/10919392.2015.1124005

Kaplan, R. S., & Norton, D. P. (2004). The strategy map: Guide to aligning intangible assets. *Strategy & Leadership*.

Kaplan, R. S., & Norton, D. P. (2007). *Using the balanced scorecard as a strategic management system*.

Klievink, B., Romijn, B.-J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers*, *19*(2), 267–283. Scopus. https://doi.org/10.1007/s10796-016-9686-2

Kohli, M., Prevedello, L. M., Filice, R. W., & Geis, J. R. (2017). Implementing machine learning in radiology practice and research. *American Journal of Roentgenology*, *208*(4), 754–760. Scopus. https://doi.org/10.2214/AJR.16.17224

Korinek, A., & Stiglitz, J. E. (2017). *Artificial Intelligence and Its Implications for Income Distribution and Unemployment* (Working Paper No. 24174; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w24174

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. Scopus. https://doi.org/10.1038/nature14539

Lim, N., & Perrin, B. (2014). Standard Business Reporting in Australia: Past, Present, and Future. *Australasian Journal of Information Systems*, *18*(3). https://doi.org/10.3127/ajis.v18i3.895

Madden, P. (2011). Greater Accountability, Less Red Tape: The Australian Standard Business Reporting Experience. *International Journal of E-Business Research (IJEBR)*, *7*(2), 1–10. https://doi.org/10.4018/jebr.2011040101

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society*, *3*(2). Scopus. https://doi.org/10.1177/2053951716679679

Miyazaki, M. (2015, March 11). *A Brief History of Data Analysis | FlyData*. FlyData | Real Time MySQL Replication to Amazon Redshift. https://flydata.com/blog/a-brief-history-of-data-analysis/

Nagorny, K., Lima-Monteiro, P., Barata, J., & Colombo, A. W. (2017). Big Data Analysis in Smart Manufacturing: A Review. *International Journal of Communications, Network and System Sciences*, *10*(3), 31–58. https://doi.org/10.4236/ijcns.2017.103003

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2014). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*. https://doi.org/10.2753/MIS0742-1222240302

Prat, N., Comyn-Wattiau, I., & Akoka, J. (2014). *ARTIFACT EVALUATION IN INFORMATION SYSTEMS DESIGN-SCIENCE RESEARCH – A HOLISTIC VIEW*. 17.

Press, G. (2019, July 19). *This Week In AI Stats: Up To 50% Failure Rate In 25% Of Enterprises Deploying AI*. Forbes. https://www.forbes.com/sites/gilpress/2019/07/19/this-week-in-ai-stats-up-to-50-failure-rate-in-25-of-enterprises-deploying-ai/

Qiang, Y., & Xindong, W. (2006). 10 Challenging problems in data mining research. *International Journal of Information Technology and Decision Making*, *5*(4), 597–604. Scopus. https://doi.org/10.1142/S0219622006002258

Raab, C. D. (2020). Information privacy, impact assessment, and the place of ethics. *Computer Law and Security Review*. Scopus. https://doi.org/10.1016/j.clsr.2020.105404

Reijers, W., Brey, P., Jansen, P., Rodriques, R., Wright, D., Koivisto, R., Tuominen, A., & Bitsch, L. (2016). *A Common Framework for Ethical Impact Assessment*. European Commission EC. https://satoriproject.eu/media/D4.1_Annex_1_EIA_Proposal.pdf

Robb, D. A., Rohde, F. H., & Green, P. F. (2016). Standard Business Reporting in Australia: Efficiency, effectiveness, or both? *Accounting and Finance*, *56*(2), 509–544. Scopus. https://doi.org/10.1111/acfi.12094

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

SBR feiten en cijfers 2018_0.pdf. (n.d.). Retrieved November 14, 2019, from https://www.sbr-nl.nl/sites/default/files/SBR%20feiten%20en%20cijfers%202018_0.pdf

Sekaran, U., & Bougie, R. (2010). *Research methods for business: A Skill-Building Approach*. John Wiley & Sons.

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press. https://doi.org/10.1017/CBO9781107298019

SyRI-wetgeving in strijd met het Europees Verdrag voor de Rechten voor de Mens. (n.d.). Retrieved May 31, 2020, from https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-wetgeving-in-strijd-met-het-Europees-Verdrag-voor-de-Rechten-voor-de-Mens.aspx

SyRI-wetgeving in strijd met het Europees Verdrag voor de Rechten voor de Mens. (2020). https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-wetgeving-in-strijd-met-het-Europees-Verdrag-voor-de-Rechten-voor-de-Mens.aspx

Umbrello, S., & Bellis, A. F. D. (2018). *A Value-Sensitive Design Approach to Intelligent Agents*. 18.

Viechnicki, P., & Eggers, W. D. (2017, April 26). *Artificial intelligence in government | Deloitte Insights*. https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/artificial-intelligence-government-analysis.html

Volman. (2017, January 10). *A European Strategy for Data* [Text]. Shaping Europe's Digital Future - European Commission. https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy

Witten, I. H., & Frank, E. (2002). Data mining: Practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record*, *31*(1), 76–77.

Wright, D. (2011). A framework for the ethical impact assessment of information technology. *Ethics and Information Technology*, *13*(3), 199–226. https://doi.org/10.1007/s10676-010-9242-6

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*(1), 1–37. https://doi.org/10.1007/s10115-007-0114-2

Appendix

Appendix A

Appendix A.1

A common framework for ethical impact assessment (Reijers et al., 2016)
The six steps of the Ethical Impact Assessment (EIA)

1. “Conduct an EIA threshold analysis
    a. Complete the EIA questionnaire
    b. Send the finished documentation to the ethics assessor or conduct a self-assessment
    c. The threshold analysis is either accepted, rejected or there will be a request for amendments
2. Prepare and EIA plan
    a. Assess the scale of the EIA
    b. Allocate a budget to the EIA
    c. Compose a team for the EIA
    d. Review and approval of the EIA plan
    e. (Optional) Repeat the threshold analysis at different stages of the project, critically when there are significant changes in the project
    f. (Optional) Consult with relevant stakeholders to raise awareness of the project taking place and gather more details about possible ethical impacts
3. Set up and execute an ethical impact identification assessment
    a. Assess the Technology Readiness Level (TRL) of the R&I project’s outcomes
    b. Review existing work in the relevant R&I field
    c. Select appropriate methods for conducting the ethical impact identification based on the TRL and the threshold analysis
    d. Gather relevant data (evidence based, by consulting experts, by interacting with stakeholders, based on creativity)
    e. Determine possible, probable and/or preferable ethical impacts
    f. Document and present the ethical impacts
4. Evaluate the ethical impacts
    a. Decide which methods should be used (desk research, expert consultation or participatory method)
    b. Conduct a contingency analysis to evaluate the likelihood of ethical impacts to occur
    c. Assess the relative importance of ethical impacts
    d. Identify potential or actual value conflicts and, if possible, aim at resolving these Formulate workable conceptualisations of the relevant ethical impacts
    e. Document and present the ethical impacts evaluation
5. Formulate and implement remedial actions
    a. Gather relevant information about remedial actions proposed by other R&I projects
    b. Formulate and implement design interventions
    c. Formulate different types of recommendations
    d. Document and present the remedial actions
6. Review and audit the EIA outcomes

a. At the beginning of the EIA: set the milestones and criteria for the review and audit process
b. During the EIA: evaluate the EIA documentation and the agreed upon criteria and milestones
c. At the end of the EIA: ensure proper documentation, follow-up and signing off of the EIA
d. Document and present the review and audit outcomes”

Appendix B

Appendix B.1

Downloaded the sets from DUO
Add “Governmental Subsidies” and “Total result” which are the cost & benefits
Changed the csv format to an ARFF, because Weka uses the ARFF format
Changed the “date” from a numerical attribute to a date attribute
Deleted the Authorised supervision, Grouping, Name to make the instances not relate to each other

Attribute original	Attributes English	Attribute format	Missing instances (%)
Jaar	Year	Date	0%
Immateriële vaste activa	Intangible assets	Numeric	92%
Materiële vaste activa	Tangible fixed assets	Numeric	6%
Financiële vaste activa	Financial fixed assets	Numeric	59%
Totaal van vaste activa	Total of fixed assets	Numeric	6%
Voorraden	Stocks	Numeric	89%
Vorderingen	Progress	Numeric	1%
Kortlopende effecten	Short-term effects	Numeric	90%
Liquide middelen	Liquid assets	Numeric	1%
Totaal van vlottende activa	Total current assets	Numeric	0%
Totaal van activa	Total of assets	Numeric	0%
Eigen vermogen	Equity	Numeric	1%
Voorzieningen	Services	Numeric	8%
Langlopende schulden	Long-term debt	Numeric	72%
Kortlopende schulden	Current liabilities	Numeric	0%
Totaal van passiva	Total of liabilities	Numeric	0%
Toename (afname) van liquide middelen	Increase (decrease) in cash and cash equivalents	Numeric	0%
Totaal resultaat	Total result (Cost & benefits)	Numeric	0%
Rijksbijdragen OCW/EZ	Governmental subsidies	Numeric	0%
Sector	Sector	Nominal	0%

REPTree parameters (Witten & Frank, 2002):

- “Seed: The seed used for randomising the data.
- minNum: The minimum total weight of the instances in a leaf.
- numFolds: Determines the amount of data used for pruning. One fold is used for pruning, the rest for growing the rules.
- numDecimalPlaces: The number of decimal places to be used for the output of numbers in the model.

- batchSize: The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.
- Debug: If set to true, classifier may output additional info to the console.
- noPruning: Whether pruning is performed.
- spreadInitialCount: Spread initial count across all values instead of using the count per value.
- doNotCheckCapabilities: If set, classifier capabilities are not checked before classifier is built (Use with caution to reduce runtime).
- maxDepth: The maximum tree depth (-1 for no restriction).
- minVarianceProp: The minimum proportion of the variance on all the data that needs to be present at a node in order for splitting to be performed in regression trees.
- initialCount: Initial class value count.

The ADAbooster parameters (Witten & Frank, 2002):

- Seed: The random number seed to be used.
- weightThreshold:Weight threshold for weight pruning.
- numDecimalPlaces: The number of decimal places to be used for the output of numbers in the model.
- batchSize: The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.
- numIterations: The number of iterations to be performed.
- Debug: If set to true, classifier may output additional info to the console.
- Resume: Set whether classifier can continue training after performing the requested number of iterations.
- Classifier: The base classifier to be used.
- doNotCheckCapabilities: If set, classifier capabilities are not checked before classifier is built
- useResampling : Whether resampling is used instead of reweighting."

Naïve Bayes parameters (Witten & Frank, 2002):

- "useKernelEstimator: Use a kernel estimator for numeric attributes rather than a normal distribution.
- numDecimalPlaces: The number of decimal places to be used for the output of numbers in the model.
- batchSize: The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.
- Debug: If set to true, classifier may output additional info to the console.
- displayModelInOldFormat: Use old format for model output. The old format is better when there are many class values. The new format is better when there are fewer classes and many attributes.
- doNotCheckCapabilities: If set, classifier capabilities are not checked before classifier is built

- useSupervisedDiscretization: Use supervised discretisation to convert numeric attributes to nominal ones.”

Appendix C

Appendix C.1: Interview IN1

Interview IN1

IN1: Ja, gaat eigenlijk prima. Ik had vorige week een leuke sessie met duo N 1 en de mensen van digicampus. Is eigenlijk best goed verlopen en ik heb daarna ik ook ongeveer twee weken terug – diverse contacten gehad met managers bij ons op om te proberen twee uur dagen in de week voor digicampus te kunnen gaan werken.

Steven: Oké.

IN1: De ene manager die trapte enorm op de rem. Er speelt zoveel, nou zeg ik , laat mij overzicht creëren. Ja, maar dat heb ik al aan iemand gevraagd. Ik denk, ooh maar dat weet ik toch niet.

Steven: oke.

IN1: Kan ik helpen of wat dan ook, allemaal moeizaam, moeizaam moeizaam... Mijn tweede manager belt mij van de week hopen dat hij rond AI en machine learning

IN1: ik had jouw nog een chatje gestuurd..

Steven: Ja...Ik zie nu dat ik daarop niet geantwoord heb. Inderdaad, via skype heb je die berichten soms door.

IN1: Maar die wil binnenkort op bezoek gaan naar TU Twente, is leuk om te kijken naar hoe het gaat met het lerarentekort? Wat kunnen wij met machine learning op het gebied van lerarentekort? Eigenlijk gaat het om besturen betere besluiten te laten nemen.

Steven: Oké, interessant, Snap ik, leuk er zijn inderdaad zoveel opties met AI.

Steven: We gaan vandaag, dit is mijn eerste interview, voor mij ook spannend, want ik heb voor mijn methode een x aantal experts nodig. Experts die met de experimenten hebben meegelopen, ga ik nu interviewen op een systematische manier, wat dan ook input is voor mijn model.

Steven: Ik heb jou een een mail gestuurd – echt net voordat we begonnen – dit hoort erbij. Dat is namelijk het consent form.

IN1: Dan kijk ik eventjes in mijn mailbox. Steven, daar is, consent form verplicht vanuit de TUD plus samen doorlopen. Ja is goed.

Steven: Kunnen we even snel doorheen, maar in het kort, het gaat erom dat dat je meedoet met een research vandaag.

Steven: Dat je mag stoppen, je mag weigeren om vragen te beantwoorden als je dat wilt. Het interview wordt opgenomen en daarna ga ik het transcriberen en dat zou ik opsturen en wordt de opname vernietigd.

Steven: Begrijp je dat dit interview gebruikt wordt voor mijn methode en dat in mijn thesis ik jou benoem, niet bij naam, maar gewoon als interview nummer 1. Maar als ik er een quote uit wil gebruiken, ik die dan wel mag gebruiken met naam.

Steven: nou in ieder geval, dat zijn de stappen. Is allemaal heel formeel, maar dit moet van te voren gezegd worden.

IN1: Ja...Natuurlijk.

Steven: Dus ga jij je van tevoren alvast, ik weet nl eigenlijk niet of zoiets van tevoren al ondertekend moet zijn, maar dat je in ieder geval nu van tevoren akkoord gaat.

IN1: Ja hoor, anders zouden we niet kunnen beginnen.

IN1: Je hebt wel bij mijn handtekening nodig op een zeker ogenblik. Dat is wel handig dat ik dit invul en even scan en jouw dan doe toekomen.

Steven: ja, dat zou heel fijn zijn.

IN1: Ga ik doen.

Steven: Maar dat heeft dus niet extreem veel haast, zeg maar dat dat. Het is nu opgenomen, en je hebt ja gezegd, dus dat is oke .

IN1: Ik kan dat niet meer terug goed ha ha. Ik zorg ervoor dat het formulier ingevuld en ondertekent wordt.

Steven: Top

Steven: Eens even kijken en nou, dan gaan we.

IN1: van start.

Steven: Dan gaan we van start. Dat is eerst met een open vraag.

Steven: Wat is jouw connectie met het project en eerst kan je jezelf kort introduceren en daarna zeggen wat de connecties zijn.

IN1: Ja.

Steven: Mijn project. De connecties met mijn project, de machine learning en het DUO.

IN1: Ja, het is wel een hele aaneenschakeling, een soort ketentje, laat ik het even wat formeel zeggen.

IN1: Ik werk bij DUO, ik hou me in die hoedanigheid als adviseur bezig met de taxonomie. Taxonomie is onderdeel van SBR/logius Een via logius ben ik in contact gekomen met mensen van digicampus. Ik vind dat waanzinnig interessant, in die zin dat ik merk dat jij niet zozeer bij logius hoort, juist helemaal niet bij logius, maar vooral bij de digicampus, waar de echte innovatie plaatsvindt, er is daar een andere vibe , vaak met jongeren mensen, maar ook het hele gedachtengoed om wetenschap en bedrijfsleven bij elkaar te trekken geeft een hele andere dynamiek . Die dynamiek is ook wel onderkend, vind ik ook wel mooi door digigcampes zelf, ja, daar moet het eigenlijk wel gebeuren, want dat dat zijn factoren die juist ontbreken in de hele formele organisaties die gewoon volgens mij, dat weet je wel, een soort big brother achtige opzet hebben, zij voeren uit en kijken of het goed gaat en ja, dat past niet meer helemaal in deze tijd. Nou, dat is de reden hoe ik daar gekomen ben en daar kom ik machine learning, artificial intelligence tegen. als één van de onderwerpen die daar hot is.

Steven: En zodoende.

IN1: Ja, wij samen tot de voldoende duidelijk, ja.

Steven: Voldoende duidelijk ja...De volgende stap is...

Steven: Ik ga even uitleggen wat ik tot nu toe gedaan hebben bij DUO.

Steven: We hebben het experiment met duo gedaan. Daarbij hebben we de financiële data gebruikt om een label te voorspellen. Helaas is dat niet gelukt om dat risicolabel te voorspellen, maar we hebben toen daar nog even over gesproken, maar ik heb er een project van gemaakt die alle stappen doorlopen die tot zo'n risicolabel komen.
Steven: Echter, is het experiment gebaseerd op door middel van een financiële data de onderwijsinstelling voorspellen, en dat had ook kunnen zijn door middel van een financiële data het risico label voorspellen.
Steven: Echter, dit kwam allemaal niet heel perfect bij elkaar, maar de stappen die doorlopen zijn die komen op hetzelfde neer. En wat ik hierbij ook wil zeggen, is dat het belangrijk is dat je dat we wel kijken alsof we dat risicolabel hebben voorspeld, want dat is eigenlijk waar we naartoe zouden willen. Maar, dan zou dan vanuit DUO zelf op een gegeven moment moeten komen.
Steven: Maar wat ik dus gedaan hebt, is dus een methode ontwikkelt die helpt om zo'n experiment op te zetten en aan welke factoren allemaal gedacht moet worden. Nu gaan we kijken naar het experiment an sich, en dan met de theorie dat we dus zo'n risicolabel zouden voorspellen.
Steven: nr 1: Ja
Steven: Want daar, daar zou ik dan het volgende over willen weten.
Steven: Denk je dat het handig is voor DUO om zo'n machine learning experiment op te stellen en te beginnen.
IN1: Ja, absoluut –
IN1: ik hoor mezelf terug, kan dat?
Steven: ja
Steven: ik hoor het zelf ook.
Steven: Nu nog?
IN1: Nu niet meer.
IN1: ik vind dat DUO zich echt zou moeten verdiepen, en dat doen we ook wel in enige mate..... waarom? Kijk, de inspectie zelf kan het ook doen, maar er is een soort waterscheiding tussen, zeg maar het voorbereiden van alles, van data dat dit ligt dan bij DUO en het uitvoeren van het toezicht en dat ligt bij de inspectie. In de praktijk zien we ook een afdeling analyse bij de inspectie, maar ja, een beetje ik vind, misschien wel omdat ik bij DUO werk, dat dat gewoon een prachtige taak is waarin wij ons zouden kunnen ontwikkelen. En dat niet alleen op dit punt, maar ik heb daar nog veel meer ijzers in vuur liggen en dat proberen we ook op te zetten in de richting van de samenwerking met de afdeling informatieproducten van DUO, maar ook samenwerken veel breder, bijvoorbeeld met mensen die zich bezighouden met studiefinanciering in Groningen, waarin wat meer geavanceerde toepassingen om al die data worden gezocht. Ook daar is hetzelfde besef wel gekomen.
IN1: Als het gaat om machine learning weet je, eigenlijk is mijn stelling we zouden alles moeten doen om zo veel mogelijk uit de data te halen wat erin zit. We hebben nauwelijks te maken, althans individu, hier met bij financiële data al helemaal niet, maar misschien ook met heel veel andere data waar we over beschikken bijvoorbeeld rendement informatie, leerlingenaantallen, vanaf een leerling niveau, individuele leerling niveau, hebben we dat beschikbaar? We weten niet wie het is, maar we kunnen in principe wel elke individuele leerling persoonlijk terugvinden.
Steven: ja
IN1: En alle aggregaties die op basis daarvan mogelijk zijn, alle gegevens van kenmerken van instellingen – ook kenmerken van de leerling, dus, overal, de regionale gebieden, samenwerking relaties enzovoort, enzovoort.
Steven: Dus eigenlijk zou je ook kunnen zeggen dat dat door middel van zo'n klein experiment gaat er een deur open voor allemaal andere experimenten.
IN1: ja ja dank je wel.
Steven: ik ga ik dit even opschrijven
Steven: oke top en dan even specifiek op dit experiment. Het maken van het machine learning model wat we nu gedaan hebben is relatief hapbaar en het kost niet extreem veel tijd, het is ook nog niet volledig af, zeg maar, er zou daarna nog wel bv een expert moeten komen om het volledig te maken, maar ten opzichte van de huidige situatie, hoe lang zijn ze bezig met het controleren van de onderwijsinstellingen? Hoeveel tijd gaat er in het maken van het aanpak model?
IN1: Ja, van het huidige model bedoel je ...want dan zou je t de meerwaarde van machine learning er naast kunnen zetten en dan kunnen kijken wat het oplevert. Je moet bedenken dat ik vertelde over het risico detectie model.
Steven: Ja.
IN1: En ik vertelde ook over, dat is er en dat levert bepaalde input, maar daarnaast kijken pakweg zo'n 30 toezichthoudende financieel de inspecteurs, met name naar kengetallen.
Steven: ja
IN1: En weet je wel, veel verder dan dat kijken ze niet. Ze kijken naar liquiditeit, rentabiliteit en solvabiliteit.
Steven: Ja.
IN1: En dan als zich een probleem in voordoet en dan kijken ze ook nog een klein beetje naar risico model.
Steven: Ja.
IN1: Ja, dan gaan ze vervolgens alles lezen wat erover te vinden is.
Steven: Oké, wel.
IN1: Ze gaan dan een meta analyse doen op het werk van de accountant.
Steven: Ja.
IN1: Dus ze controleren eigenlijk het werk wat al gedaan is door een accountant om zeker te weten dat bij die instellingen waar ze vermoeden dat daar zich iets voordoet of een beetje de gevarenzone zouden kunnen komen, of misschien wel zitten, dat dat het werk van accountant daar goed is gebeurd en daarmee valideren zijn eigenlijk de betrouwbaarheid van de data, is dit inderdaad wel goed gebeurd, heeft die accountant daar goed naar gekeken en vervolgens krijgt dan die accountant op zn falie, zal ik maar zeggen: als die het niet goed gedaan heeft. Dat is eigenlijk het allerbelangrijkste wat ze doen, dus we hebben een soort meta toezicht rol op accountants controle. Ja, je begrijpt wel dit allemaal uitvoeren en ook op deze manier een bijdrage leveren, dat dat ontzettend bewerkelijk is.
IN1: Op de manier waarop zij daar nu naar kijken en aan werken, is er nauwelijks ruimte voor iets nieuws.

Steven: Nou, terwijl eigenlijk, als je naar het experiment kijkt, zoiets opzetten en het bekijken van die organisaties, dat kost veel minder tijd. Dan kan het een hele efficiënte manier zijn om hier opnieuw naar te kijken. En misschien wel leiden tot, in plaats van 30 toezichthouders, maar vijf toezichthouders.
IN1: Ja zeker, ik denk dat dat echt wel zo is. Weet je, wel in en aantal opzichten ook omdat de aanpak van alles gaan lezen, op het gebied van het hele bestuur vlak, ook de jaarrekening wordt doorgeploegd, dat wat accountant heeft gedaan wordt overgedaan.
IN1: Ja weet je, op het moment dat je dat allemaal hebt en je hebt vertrouwen in de data, zoals die zijn, dat is dat ook wel een beetje een probleem. Ze zitten vooral met papier ook hier te werken, om kijken wel een beetje naar data, maar ze kijken vooral naar papier. Ze laten de data los en gaan naar het papier. Ja, daar kun je natuurlijk veel efficiënter in handelen, laat staan dat je dus ook nog een een algoritme hebt gemaakt, adhv waarvan je met behoorlijk wat betrouwbaarheid op een bepaald pad wordt gezet en dat pad gaat volgen en dat pad gaat volgen, blind.
Steven: Oke top.
IN1: Ik zie hier een beetje gradaties bij de huidige situatie uitermate de werkelijk ook grotendeels op papier gebaseerd. Één stap verder is om het papier los te laten en helemaal digitaal te gaan werken.
Steven: ja
IN1: En nog een stap verder is om de techniek zodanig te vertrouwen, machine learning dus, dat je je daarop durft te verlaten. Al die dingen die gaan gepaard met volgens mij met een reductie van arbeidskracht en dus ook gewoon met de fte's die je nodig hebt.
Steven: Oké, denk je dan dat het experiment dat we nu in het kort hebben laten zien, met wat er mogelijk is met zo'n financiële data en hoe je zo'n label kan voorspellen, en we zijn met voorspellen van de onderwijsinstellingen op een accuratie van één 91% procent gekomen, maar dat even terzijde, dat het doorlopen van zo'n experiment op DUO data denk je dat dat van toegevoegde waarde is voor de voor het duo? het doorlopen van van dit experiment?
IN1: Ja, absoluut. Sterker, ik zou eigenlijk ervoor pleiten dat we dit en andere experimenten op een rijtje gaan zetten en de komende jaren daar een soort, laat ik zeggen, een agenda voor ontwikkelen en die ook gaat uitvoeren, want als we dat niet doen, dan lopen op een gegeven ogenblik, we lopen eigenlijk nu al verschrikkelijk achter de feiten aan. Je moet ook bedenken die financiële data, die geven weer van datgene wat zich vorig jaar heeft voltrokken, dus als jij wil, en het gaat er natuurlijk over om te inschatten of iemand, een bestuur, er op termijn in de toekomst, een probleem mee heeft.
IN1: **Steven**: ja precies
Steven: Hoeveel verder je van tevoren kan inschatten hoe eerder je kan gaan helpen hoe minder grote problemen zijn? Ja.
IN1: Klopt. En weet je der zijn, dit is geen geheim, we hebben een aantal debacles gehad, ROC Leiden. Je kunt ook even terugzoeken er zijn heel veel nieuwsberichten over verschenen en die ROC Leiden, heeft heel veel investeringen gedaan in gebouwen, en vervolgens liepen de studentenaantallen terug en dat was best wel een probleem. En toen hadden ze een enorme financiële problemen. En toen moest de overheid wijzigingen bijkomen.
Steven: Dat heb ik meegekregen je inderdaad.
IN1: Er was iets in Groningen een poos geleden over investeringen in het buitenland, Universiteit Groningen en zo zijn er nog een paar financiële debacles geweest in het in het onderwijs. En iedere keer is dan de vraag "wist de inspectie dat?" Ja, dat wisten we wel. Waarom heeft de inspectie dan niet tijdig... nou weet je wel en dan komen ze met de verdediging dat dat niet hun taak is aan de ene kant, want zij controleren alleen maar accountants.
Steven: Ja.
IN1: Ja, nou weet je , daar kun je niet meer mee thuiskomen natuurlijk. En twee is ja, ze houden zich aan de wet, zeggen ze dan, maar dat is zo zielig om dat te doen. De samenleving heeft gewoon behoefte aan iets anders, namelijk een proactieve houding en grip op de ontwikkelingen.
Steven: Oke top dit beantwoord mijn vraag goed, want dit is dan een stukje over het experiment en wat de toegevoegde waarde
Steven: N1 zegt dan mooi
Steven: En dan gaan we nu naar de volgende stap en de gaat over de methode en daarvoor ga ik eventjes met jou delen via het scherm.
IN1: leuk gordijntje open doen.
Steven: Hallo.
IN1: Ja.
IN1: Dit is de versie van de methode die ik nu ontwikkeld heb . dit is de versie met alle stappen en hoe het er ongeveer uit gaat zien.
IN1: Het idee is dat je met het doorlopen van deze methode tot een machine learning experiment komt.
Steven: Zodat de organisaties zelf kunnen gaan experimenteren. Dus wat je zegt dat hierna ook DUO na een eerste experiment ook zelf de eerste stappen van een experiment kan gaan doen.
Steven: En voor de rest qua stappen, want dat is inderdaad wel een beetje de context, dus op het moment dat je nog niet 100 procent zeker weet of je eigenlijk wel een machine learning experiment wilt gaan doen.
Steven: Zijn er dan nog stappen die duidelijkheid geven of die je mist?
IN1: Ja, misschien wel iets tussen nog van tussenresultaten.
Steven: Ja.
IN1: En misschien wel gebruik maken van de kennis van , laat ik in ieder geval de inspecteurs hé, om als je bezig bent bij algoritme cq model testing en model adjusting dat je daar misschien ook hun ideeën daarbij nog kan toetsen. Zo van nou, we zijn nu halverwege het experiment of halverwege het project. We hebben nu een paar uitslagen we zien dit en dit en dit wat roept dat op. maar voorwaarde, is dan wel dat die input dan ook gebruikt kan worden bij het aanpassen van het model en de keuze van bepaalde parameter of whatever.
Steven: En dek je dat niet dan door iemand erbij te hebben zitten, die dus de data experience en access heeft.
IN1: Dus ja, als die er al bij zit dan wel ja.
Steven: Oké.
Steven: Dus ja dus dat wie we wat team selection als team set up, data experts en access.
Steven: Deels.

IN1: Dan kun je deels – ik denk ook echt wel bij de inspectie – in ieder geval zo dat dat je daar nadrukkelijk ook iemand met verstand van financieel toezicht bij moet hebben. En die hoeven niet zoveel te hebben van je data, maar die nemen weer andere domeinen in zijn beoordeling mee.
Steven: Oké, oké, oké, oké, nou, en dan voor de, ik ben het inderdaad wel met je eens met wat je zegt, maar dat is wel voor dat specifieke geval, dus op het moment dat we aan het begin hebben gezegd dat we iemand voor data experience. Met data experinence en access hebben, dan zou dat in theorie voor dat project genoeg zijn.
IN1: ja, dat ben ik helemaal met je eens.
Steven: Oké, en voor de rest nog bij stappen en dingen die missen dingen die je graag zou willen zien.
IN1: Nee, eigenlijk niet nee, ik ben blij dat ik überhaupt iets kon opmerken. Het is wat ik ook wel merkt is dat het wel echt gewoon een beproefd model is met hele logische stappen en waar goed over nagedacht is. Belangrijk vind ik management en communicatie erbij. Nou ja, daar is aandacht voor. Dus.
Steven: Oké, kijken ja, ja, ik vind dat ook soms moeilijk dat inderdaad dat je een model aan het maken bent wat in sommige gevallen gewoon best wel recht toe recht aan is. Maar doordat je het allemaal zo opschrijft, dat je alle stappen doorloopt, zeg maar en iemand echt bij de hand neemt door alle stappen. Dan zorg je er wel voor dat in theorie dat het machine learning project tot een succes komt.
Steven: Ook al zijn sommige dingen heel logisch.
IN1: Maar eigenlijk geef je nu zelf bijna een stap....voeg jezelf toe.
IN1: explain de model to .... en dan komt er wat , explain en neem ze even bij de hand.
Steven: Ja misschien in communication of zoiets.
IN1: Ja, zeker zeker wel.
Steven: Oké, oké, oké, hé, er komt eigenlijk. Wat moeten we ook een beetje een einde aan breien ook voor jou.
IN1: Ja, ik moet boodschappen doen.
Steven: Boodschappen doen: oké, we hadden we hebben nog een klein kort likert schaaltje. Ik moet de excel zien. Ja, oké,
Steven: dit zijn nog zes korte vragen en het is ook weer met sterk mee oneens en sterk mee eens.
Steven: De methode is effectief in het bereiken van een machine learning project.
IN1: Ik zie nu de die vraag, die jij stelt, zie ik niet voor.
Steven: Dat klopt.
IN1: Oké, nog een keer over.
IN1: de methode is effectief in het bereiken van een machine learning project ja en het bereiken van het doel van een machine learning om het succesvol he denk ik dan of überhaupt te starten kan natuurlijk ook. Ja, daar ben ik het wel heel erg mee eens. Daar ben ik sterk mee eens .
Steven: De methode is effectief in het bereiken van een succesvol .
Steven: De methode is praktisch goed te gebruiken.
IN1: Ja, dan ben je wel eens.
Steven: De methode is makkelijk te begrijpen.
IN1: eens.
Steven: Eens.
IN1: Dit model deel De ene kant het heel evident en aan de andere kant denk je nou......
Steven: Ja, ja, ja, maar ik ga op zich, dus dat is dus de vraag in. In hoeverre stel, als je een klein beetje machine learning ervaring hebt, dan zijn dus de stappen an sich zijn goed te begrijpen of niet goed te begrijpen.
IN1: Ja.
Steven: Het model past bij mijn organisatie.
IN1: Ja, vind ik wel eens, kan er makkelijker relateren roep er van alles bij
Steven: De methode is compleet.
Steven: ja
Steven: De methode heeft het vermogen om te reageren op de schommelingen van de omgeving, een moeilijke vraag, maar dus of de methode werkt bij verschillende problemen. Om het zo te zeggen: hé, het is niet een probleem waar die mee kan omgaan, zijn of denk je dat die nu echt wel specifiek is op dat ene probleem.
IN1: Ja, ik durf, ja ja ja nou ja je kan die daarop , heeft de methode het vermogen, er wordt ook met stakeholders gesproken. Er is in dit team, ik denk wel dat dit zou ik gewoon een eilandje in de organisatie kunnen zijn. Is het, laat ik hier maar neutraal zijn – zo goed top top.
Steven: De methode komt ongeveer overeen met het experiment wat we doorlopen.
IN1: Dat begrijp ik effe niet.
Steven: Ja, hoe legt ik het uit, het machine learning experiment wat ik voor DUO heb gedaan, om het zo te zeggen, de stappen die ik daarbij doorlopen, komen die overeen misschien wel een moeilijke vraag omdat voor jou de situatie natuurlijk wel is dat jij niet het hele experiment doorlopen hebt, maar misschien een beetje een kijk erop.
IN1: Nou, je bent wel bezig ja komt overeen. Zeker. Daar mag je gerust eens neerzitten, want ik merk wel dat, de vraag die je stelde wat we gedaan hebben en wat jij gedaan hebt. Dat zijn wel ongeveer stappen die ook in het model genoemd worden.
Steven: Ja, ja, want kijk, dat is natuurlijk ook nog eens een ding, want we hebben natuurlijk inderdaad met duo dat experiment gedaan en het is gebaseerd op duo data. Alleen, het is niet helemaal dat het helemaal bij DUO is geland is denk ik dat en dat maakt een aantal van deze vragen misschien wat onduidelijker
Steven: Is even kijken? Ja, we hebben, we hebben elke stap beschreven, we hebben we hebben bekeken of er dingen missen in de methode. We hebben het experiment kort behandeld. Ik moet het allemaal gaan naar de transcriberen, gaat een mooie, mooie lange tekst worden en ik maak er dan een kleine samenvatting van en die ga ik naar je opsturen goed en zijn er voor de rest nog vragen.
IN1: Nee, nee, ik hartstikke leuk. Ik leer je ook wel weer van. Wat is iedere keer een beetje moeilijk om onderscheid te maken tussen wat je doet, hè en het methodisch wat je tegelijkertijd ontwikkeld, nu wel wat meer inzicht in gekregen. Het grappige is dat wij dat ook zelf gaan doen bij de experimenten overleg. Ik heb net met x gesproken over daar zit van alles alvast, maar weet je wel x die wil je wil iets doen. We gaan een workshop houden, ja, in welke mate is die workshop nou een experiment? Ja, dat moet wel wat anders wordt het niet

bekostigd. Nou ja, we hebben de truc bedacht, gaan er een methode van maken, ongeveer wat jij ook gedaan. Dus ik zeg maar de regering: oké, we gaan hier een methode ontwikkelen om de diffuse dingen, behoefte en dergelijke bij hun doelgroep te bevragen.
***Steven**:* Ja.
***IN1**:* Is het van belang is van belang voor allerlei niet alleen voor duo waar of op de OCW In dit geval, maar dat is ook van belang voor we bijvoorbeeld bonen.
***Steven**:* Ja.
***IN1**:* Oké, we hadden truck , oke we gaan een methode maken en dat zie ik bij jou, Je test de methode.
***Steven**:* Ja, en ik en ik snap dat dat af en toe het gevoel is dit is heel logisch. Maar ja dat op het moment dat je dat niet doet, dan baken je niet af om het zo.
***IN1**:* logisch snap het helemaal.
***Steven**:* Dat dat is wel leuk, is hoe naar buiten het interview om een zou zijn en lekker bezig nu nog drie weken hard knallen, dit allemaal verwerken en hopen dat het allemaal op zn tot op zn pootje komt. Eh nee, hoor bedankt sowieso voor het interview kost toch stiekem meer tijd aan dan ik had gedacht.
***IN1**:* Het hier een uur nu en tien minuten.
***Steven**:* Ja.
***IN1**:* Maar een uur moet je er altijd voor uittrekken.
***Steven**:* Ja, dat is dat.
***IN1**:* Moet je moet, je er nog meer doen?.
***Steven**:* Ja, ik heb nog vier interviews deze zeg maar nog eentje, deze week en volgende week, drie en alles verwerken ehm. En ja, kijk.

	Cijfer
De methode is effectief in het bereiken van een machine learning project	5
De methode is praktisch goed te gebruiken	4
De methode is makkelijk te begrijpen	4
Het model past bij mijn organisatie	4
De methode is compleet	4
De methode heeft het vermogen om te reageren op de schommelingen van de omgeving	3
De methode komt overeen met het experiment	4

Step:	Usefull	Understandable
1	4	4
1.1	4	4
1.2	4	4
1.3	5	4
2	5	5
2.1	4	5
2.1.1	4	5
2.1.2	4	5
2.2	5	5
3	5	5
3.1	3	4
3.2	4	4
3.3	5	5
4	5	5
4.1	5	5
5	5	5
5.1	5	5
5.2	5	5
5.3	5	5
5.4	5	5
6	5	5
6.1	5	5
7	4	3
7.1	5	5
7.2	5	5
7.3	5	3
7.4	5	3
8	5	3
8.1	5	3
8.3	5	3
8.3	5	3
9	4	4
9.1	5	4
9.2	5	5
9.3	3	4
10	4	4
10.1	5	5
10.2	5	2

Appendix C.2: Interview IN2

***Steven**:* Lang niet gezien. Het is alweer een tijdje geleden dat we elkaar gesproken hebben.
***IN2**:* We hebben er dan ook nu de tijd voor.
***Steven**:* Dat is op zich prima en je mager de tijd voor nemen; daar wordt het alleen maar beter op.
***IN2**:* Hoe is het met jou?
***Steven**:* Goed, druk bezig, de laatst weken voor groen licht.
***Steven**:* Het is heel hard werken, maar dat is fijn op zich. Ik moet zeggen dat ik echt verreweg het meeste doe in deze laatste weken.
***IN2**:* Ja.
***Steven**:* Het is een beetje de standaard, de 70 30, 30 70 regel
***IN2**:* Daar gaat de meeste tijd in nu. Echt gewoon schrijven; typen en nog eens typen.
***Steven**:* En na de laatste interviews is het experiment met WSW afgerond.
***IN2**:* Ja.
***Steven**:* Nu moet ik goed opschrijven, en van elke stap in die methode aangegeven waarom die er in zit en waar die vandaan komt, Etcetera. Ik moet dat nog wel beter opschrijven. Ik heb al wel best veel , maar het moet allemaal wat academischer neergezet worden. Daar heb ik ook nu de tijd voor. En ik hoop dat het allemaal goed gaat lukken en daarom ben jij ook één van de mensen die ik interview. Te beginnen bij jou en IN2. Ik ga IN4 ook interviewen en nog twee mensen van WSW. En dat is dan deels om de methode, het valideren en te kijken en deels of er nog dingen missen en of er stappen uit moeten.
***IN2**:* Doen.
***Steven**:* Laten we maar gewoon het interview doen, want dan gaan we eigenlijk ook door mijn hele mijn hele methode heen en zie je wat we de laatste tijd gebeurd is.
***IN2**:* Ik snap zo wel even op de fiets, maar volgens mij moet gewoon lukken.
***Steven**:* Oké, oké, jij stapte op de fiets
***IN2**:* ja, ik moet even ergens heen, maar dat is prima.
***Steven**:* Nou, ik moet je ook wel een stuk laten zien, namelijk.
***IN2**:* Oh, oké, ja.
***Steven**:* Hoe laat ga je fietsen?
***IN2**:* Nou, eh, nu.
***Steven**:* Nu?
***IN2**:* En eh, dat duurt, dat is een kwartiertje.
***Steven**:* Een kwartiertje fietsen, oké, kijken hoe dit anders kan.
***Steven**:* Even kijken: wat is wijsheid? Er zijn wel een paar dingen die je echt moet bekijken, eigenlijk.
***IN2**:* Oké, dat lukt wel denk ik, ja of tenminste, ik ga zo uit wifi, dan schakel ik nu even over, een tel..
***Steven**:* Yes
***IN2**:* Ja, ik hoor en zie je.
***Steven**:* Over een kwartier kan je wel gewoon kijken . Dan is het misschien handiger om dan, ver te gaan. Ik moet de methode stap voor stap doorlopen. En dan is het misschien wel handig dat je een totaalbeeld ziet.
***IN2**:* Ok, zijn er dingen die we nog wel even kunnen doorspreken nu.
***Steven**:* Eigenlijk gaat het wel echt op de methode om dit te bekijken en waar we nu zijn.
***IN2**:* Oké
***Steven**:* Kan je zeggen dat je akkoord gaat met wat de onderdelen die op het consentvorm staan.
***IN2**:* Ik ga akkoord met het consent
***Steven**:* Ik ga mijn scherm nu met jou delen. Zie je nu de methode? Eventjes snel uitleggen wat ik tot nu heb toegedaan. We hebben een experiment met de DUO dataset gedaan. We hebben een experiment met WSB gedaan om het risico van de organisatie te voorspellen. Dat experiment is ook afgerond. Succesvol. Een op basis van de literatuur en de experimenten heb ik deze methode ontwikkeld. Het idee is dat je met het doorlopen van al deze stappen tot een succesvol Machine Learning experiment komt wat rekening houdt met een aantal organisatorische factoren en een aantal ethische vragen.
***IN2**:* Hartstikke goed joh.
***Steven**:* Ja, en het idee is nu dat dat we door een deel open vragen lopen en een deel systematisch vragen met check of elk onderdeel duidelijk is en of het goed is. Ik denk dat we er best wel snel doorheen kunnen gaan. Want op zich zit jij wel redelijk in de stof. We gaan gewoon de methode doorlopen. Allereerst vragen van: als je nu snel naar de methode kijkt, zijn er dan veel onduidelijkheden of ziet het er allemaal wel recht toe recht aan uit.
***IN2**:* Het ziet er heel logisch uit. Alleen wat kleine punten. Bij één, dat is de machine learning, heb je, wat daar onder ligt, neem ik aan een checklist
***Steven**:* Ja, dat is natuurlijk de vraag, inderdaad, van in hoeverre je dit uitwerkt. Nu heb ik zeg maar dit plaatje en daar onder zit nog een uitleg. Maar ja, dat is moeilijk, want anders wordt er weer een heel boekwerk qua beeld wat ik laten zien.
***IN2**:* Het hoeft hier ook niet in, maar wel was ik even benieuwd, of daar iets van een methodiek achter zit. Maar prima. Misschien is het ook handig als je iemand met domeinkennis hebt.
***Steven**:* Ja, bedoel je ..
***IN2**:* Data experience and access. Iemand die bedreven is met data, maar betekent niet altijd, dat het iemand is die de betreffende data ook goed kent.
***Steven**:* Oké, domein, kennis. Deze dingen doe ik, want ik moet dit systematisch doorlopen. Aan het einde bespreken we inderdaad de dingen die je anders zouden willen zien en die je mist.
***IN2**:* Nou ja, we zijn er bijna doorheen hoor, punt tien, nog even communicatie en het communiceren van results.
***IN2**:* Lijkt me ook heel belangrijk, vastleggen, documenteren en communicatie Ja, daar doe je daar alles in wat ik ook bedoel.

IN2: Dan ligt de nadruk heel sterk op het naar buiten brengen, terwijl het ook een heel groot belang is aan het goed documenteren en vast te leggen.
Steven: Oké, dat is een goeie aanvulling/
Steven: Ik heb dit nu opgeschreven. We gaan nu door elke stap en dat kunnen we best wel snel doen, want dat is gewoon een heel systematisch. Maar ik wil bij stap graag weten of je de stap usefull vindt, effectief, en daarmee kun je het sterk mee oneens of eens zijn. Dus zeg maar sterk mee oneens, neutraal eens, sterk mee eens. En wat we wat misschien handig is, is denk ik.
IN2: Vier of vijf punt schaal?
Steven: vijf
Steven: Waarbij drie is neutraal: vijf is sterk mee eens en één is sterk mee oneens. En dan van elke stap wil ik weten of die effectief is en of die begrijpbaar is. Zullen we dit gewoon snel doen. Dan begin ik eerst met alle kleine stapjes dus, één punt één, en dan daarna door grote de en dan daarna goal formulation, zeg maar één en daarna weer de kleine stapjes. En dan verder.
IN2: Ja.
Steven: Laten we beginnen met set up project goal. Nuttig?
IN2: Ja, ja, heel erg, dus vijf.
Steven: Vijf, Understandable?
IN2: ook vijf punt schaal, en understandable, vijf
Steven: Does the goal fit with machine learning?
IN2: Belangrijk vijf, understandable vier zou ik op eerste gezicht zeggen. Want ja, het heeft gewoon meer uitleg nodig, maar die heb jij.
Steven: Does the goal fit with the mission of the company?
IN2: Mmm.
Steven: Ik vind dat lastig de context hierbij bij te zeggen want de context hier is wanneer het doel van je machine learning project in lijn is met het doel van het bedrijf, dan is het makkelijker om een project op te zetten.
IN2: financiële, ethische,etc
Steven: Ja, bij WSW bijvoorbeeld is het het bepalen van het risico van een organisatie zodat zij kunnen kijken of ze een lening kunnen geven. Dat is één van de grotere taken van hun als bedrijf.
IN2: Ik vind het vijf
Steven: En over de algemene stappen, de goal formulation?, Nuttig?
IN2: Ja, door doen vaststellen is ja, het is gewoon een vijf
Steven: Een vijf, en understandable?
IN2: een vijf
Steven: Het zijn veel getallen; Ik moet ze uiteindelijk allemaal opschrijven, het gaat niet lang duren, maar het zijn veel getalletjes en dingetjes, en ik moet ze allemaal horen van jou. Oké, voor the team setup, include person that has machine, learning experience.
IN2: Vijf, vijf
Steven: En data experience?
IN2: Ja, vijf, vier vanwege die eerder gevraagde.
Steven: Include the management?
IN2: vier, vijf,
Steven: Ja, eh de totale stap voor de teams setup?
IN2: Vijf,vijf
Steven: Context analysis, conducted Stakeholder analysis?
IN2: Vijf vijf.
Steven: Analyze current situation of goal
IN2: Ja, dat heeft iets meer context nodig. Ga je die geven of moet ik het beoordelen zonder dat je de context geeft.
Steven: Het idee is dat je bijvoorbeeld inderdaad een risico model hebt, bijvoorbeeld een WSW, ze hebben al een risico model hebben waarmee ze kijken hoe zo een organisaties staat. Je analyseert eerst wat de situatie nu is.
IN2: vijf,vier
Steven: Je zou dus zeggen dat je dit misschien iets op een andere manier neer moet zetten. Dit doe ik aan het einde.
Steven: start ethical impact analyses?
IN2: Vijf vijf.
Steven: totale stap?
IN2: Ja, vijf vijf, Ik ben heel gul.
Steven: Ja.
IN2: Maar het is dus gewoon duidelijk.
IN2: Ja.
Steven: Derde collection, één stap?
IN2: In dit geval redelijk rechttoe-rechtaan.
IN2: vijf, vijf.
Steven: Ik zie dat jij je dingen bij wel zeggen, maar die gaan we zo doen.
Steven: Derde preparation, één, check quality delen?
IN2: Vijf, vijf
Steven: clean the data?
IN2: Vijf,vijf.
Steven: make the data weka-readable ?
IN2: vijf,vijf
Steven: Is this the data set?
IN2: Ja, het zal wel belangrijk zijn, maar dat kan ik niet zo goed beoordelen.

***Steven**:* Je splitst de dataset in een training data set en een test data set
***IN2**:* vijf vijf.
***Steven**:* Dan is de totale stap waarschijnlijk ook een vijf, vijf omdat alles tot nu toe vijf, vijf is.
***IN2**:* Ja, wat ik mis, en sorry dat ik hier even tussendoor kom, maar zeker bij structured data is het heel belangrijk om goed de metadata, dus de betekenis van de data, te doorgronden. Dus ik kan me voorstellen – dat doe je of bij deze collectie preparation – dat je de data beschrijft.
***Steven**:* De data beschrijven
***IN2**:* Ja.
***Steven**:* Oké, ik ga je die zo nog een keer vragen.
***IN2**:* Ja.
***Steven**:* Is even kijken in zes en hier komt ook weer een klein beetje meer context.
***IN2**:* Maar die is mij duidelijk uit de selectie table.
***Steven**:* Zag je hem?
***IN2**:* Ja, het was duidelijk, vijf, vijf.
***Steven**:* Vijf vijf, en daarmee ook de stap.
***IN2**:* Ja.
***Steven**:* Zeven punt één: model testing, using the algorithmn of the dataset to create models?
***IN2**:* vijf vijf.
***Steven**:* Test the model on the test data set?
***IN2**:* Vijf vijf.
***Steven**:* Note results?
***IN2**:* En dat is? Het documenteren van de resultaten, bedoel je dat ermee?
***Steven**:* Ja, ja, dit is vooral om te vergelijken. Misschien moet ik dit tot één stap brengen
***IN2**:* Ja, dat zou zou mij logische lijken, dus in die zin drie, vier.
***Steven**:* compare the results?
***IN2**:* vijf vijf.
***Steven**:* Totale stap?
***IN2**:* Vijf,vijf.
***Steven**:* acht punt één; select the best model?
***IN2**:* vijf vijf.
***Steven**:* systematic...
***IN2**:* vijf, vijf
***IN2**:* use a booster is mij niet duidelijk.
***Steven**:* ik moet het denk ik anders neerzetten. Het is een improvement algoritme waarbij je, hoe kan ik het makkelijk uitleggen, een booster maakt bijvoorbeeld 100 decisions trees, waarbij een decision tree een een decisions tree maakt, maakt die booster er 100, en die pakt dan alle onderdelen die hij het best vindt, die voegt die dan samen en dan creëer je een nog betere. Dus, het is een tuning.
***IN2**:* vijf,vier, overall, vijf, vijf.
***Steven**:* We gaan zo er snel doorheen? Het is goed. Dat heb ik liever ook zo, want dit is namelijk nu meer het kwantitatieve. Deze score komt er in en daarna is de discussie beter.
***Steven**:* Project evaluation, check the result model?
***IN2**:* Vijf,vijf
***Steven**:* compare the results with the currency situation.
***IN2**:* Vijf,vijf
***Steven**:* Evaluate the need of a machine learning expert?
***IN2**:* Vijf ,vijf, is van belang ook
***Steven**:* Ja, dat je op een gegeven moment iemand toevoegt aan het project wanneer de eerste stappen klaar zijn.
***Steven**:* Totale stap?
***IN2**:* Vijf,vijf.
***Steven**:* laatste, communicate results with the management?
***IN2**:* vijf,vijf
***Steven**:* Communicate results and git-lab?
***IN2**:* Vijf,vier.
***Steven**:* Ja, ik zit er nog over na te denken. Is het nodig om een succesvol machine learning project te hebben? Nee. Maar is het handig voor de totale community?
***IN2**:* Je kunt het natuurlijk wel iets generieker maken. Communiceer externaly.
***IN2**:* Of bepaal of je de resultaten extern wilt te communiceren.
***Steven**:* Ja, een goede. Want inderdaad, even de laatste want zijn we er namelijk bijna.
***Steven**:* Een team als geheel?
***IN2**:* Vijf,vijf.
***Steven**:* Dan even wat generieke vragen over de totale methode, ook stelling en dan sterk mee oneens, oneens, neutraal, ééns. sterk mee eens.
***Steven**:* De methode is effectief in het bereiken van een machine learning project?
***IN2**:* Effectief, vijf,
***Steven**:* de methode is praktisch goed te gebruiken?
***IN2**:* Vijf.
***Steven**:* De methode is makkelijk te begrijpen?

IN2: Ik zou zeggen vijf met de voorwaarde dat iemand wel een basiskennis heeft. Maar ja, dat spreekt voor zich.
Steven: de methode is compleet?
IN2: ik zou op zich zeggen een vijf, maar laat ik zeggen, vier, want ik mag denken dat hij compleet is, maar het zou kunnen zijn dat we dingen mis.
Steven: Ja, dat kan altijd. De methode heeft het vermogen om te reageren op schommelingen van de omgeving, oftewel op verschillende gebieden toepasbaar?
IN2: Bedoel je dan...?
Steven: In de machine learning.
IN2: Is het ook in een dynamische omgeving toepasbaar dan wat je bedoeld of is op verschillende? Is de methode zelf te kneden naar afhankelijk van wat er nodig is? Of is het de methode in te zetten in de omgeving die je zelf gedaan hebt
Steven: Aan de omstandigheden die er zijn. Dus generiek in zekere zin.
IN2: Ja, dat is een vijf omdat je een duidelijke context analyse doet.
Steven: Dat zijn alle kwantitatieve onderdelen. Even kijken of we dat hadden inderdaad. Als we nu naar de stappen gaan. Je hebt gezegd, ik zou bij een communication, het vastleggen willen zien.
IN2: Ja, of een aparte stap, maar het lijkt mij best wel daarbij passen.
Steven: Ja, dat kan een aparte stap en het doel van die stap is dan het vastleggen.
IN2: Ik bedoel met een paar stappen, het zou bijvoorbeeld nummer 11 kunnen zijn, maar ik vind 'm best wel passen bij communiceren.
IN2: Hij zou, wat mij betreft, ook tien punt drie kunnen zijn.
Steven: Dan even wat uitgebreid, het gaat er dan omdat je dus de resultaten vastlegt en het model?!
IN2: Ja, dat je gedurende het hele traject documenteert, zodat het herhaalbaar is en dat je dat vervolgens ook op een juiste manier vastlegt. Zou ik de keuzes die je hebt genomen, dus dat is niet per se iets wat je extern wilt communiceren niet per se iets wat je op script vast wilt zetten, maar wel voor de organisatie waar je dat doet. Als een leer document: waarom hebben we dat gedaan?!
Steven: Zou je dat dan misschien al eerder in het document of in de stappen willen hebben, zodat het een soort start is van documenteren.
IN2: Ja, ik vind dat best. Ik vind dat een beetje impliciet. Eigenlijk alle vragen die je stel lijkt me dat je die documenteert en de keuzes documenteert.
IN2: Ik neem ook aan dat je dat doet.
Steven: Ja, in zekere zin, maar dan moet ik het er wel expliciet bij zeggen dat had gedaan moet worden. Want stel dat je niet bij zegt en je zegt van ja, de methode is toereikend en iemand doet het niet dan en dan aan het einde zeg je dat je de documentatie moet toevoegen.
IN2: Ja dat moet jij zelf even beslissen. Het gaat erom dat als je dit bij een organisatie toepast, dat die organisatie wil weten wat er is gebeurd, stel één van de medewerkers neemt ontslag.
Steven: Ja.
IN2: Dan moet een ander dat kunnen oppakken. Stel er komen onregelmatigheden aan het licht. Dan moet de organisatie kunnen terugkijken op wat is daar gebeurd? Stel er komt een ethische commissie die vragen gesteld over het gebeuren en dan moet je dat allemaal kunnen documenteren. Het gaat eigenlijk een interne transparantie.
Steven: Ja, maar dit is dus veel uitgebreider. Heel goed.
Steven: De volgende was dat je zei dat de domeinkennis niet helemaal duidelijk naar voren komt en dat data experience niet per se altijd domeinkennis is.
Steven: Een andere is dat bij data collection het heel belangrijk is om voor jezelf de metadata dus de semantiek achter de data te begrijpen en misschien ook wel te documenteren.
Steven: En als iemand die domein kennis heeft, begrijp hij dit dan? Is dat dan niet inherent dat hij dat al heeft?
IN2: Ja, als het, als het iemand is die perfect op dat topic zit en net als bij de DUO data. Ja, als jij iemand trekt die verstand heeft van jaarverslagen van onderwijsinstellingen, die begrijpt dat. Als je iemand erbij haalt met generieke kennis van jaarverslagen, dan zal ie toch nog effe moeten kijken "Oh, ja, wacht dat is dit getalletje"
Steven: Dus op het moment dat ik in stap twee mogelijk domeinkennis toevoeg, is het dan belangrijk dat er bij data collection die meta understanding ook nog komt?
IN2: Ja, het lijkt me namelijk belangrijk dat de domein kenner dit ook op de juiste manier kan communiceren naar de machine learning kenner.
IN2: Het is een soort referentie document of zou dat moeten zijn. Net als waar wij kijken naar die data, dat we dan daarna gaan kijken in dat we wel zien van "Oh ja, daar zit een fulltime series in deze getallen opgeteld bij elkaar".
Steven: Inderdaad, dat hebben we zo met WSW ook op die manier gedaan. Soms gaat dat automatisch doordat iemand al daarmee bezig is. Maar het is nog niet een gegeven dat dit altijd zo is. Dus inderdaad, dat expliciet benoemen. Dus interne transparantie, communicatie en traceerbaarheid en domeinkennis en data collection/meta, dat zijn de drie die we hebben toegevoegd.
Steven: Zie je bij stappen nog andere onderdelen waarvan je denkt van dit is iets wat ik graag zou willen zien?
IN2: Nee, ik heb het allemaal al gespuid, denk ik.
Steven: Zijn er stappen die volledig overbodig zijn?
IN2: Nee, niet in mijn ogen.
Steven: Oké, dat is goed. We hebben in principe alle stappen doorlopen, we hebben ze gerankt of ze begrijpelijk zijn en of ze effectief zijn. We hebben we drie extra toevoegingen gevonden en er zijn geen dingen die er af zijn.
Steven: Dus ja dit was 'm in zekere zin.
IN2: Mooi.
Steven: We hebben het in een half uur exact gedaan.
IN2: Ja, we gaan we ook al weer een tijdje mee samen. Precies en bedankt.

	Cijfer
De methode is effectief in het bereiken van een machine learning project	5

De methode is praktisch goed te gebruiken	5
De methode is makkelijk te begrijpen	5
Het model past bij mijn organisatie	
De methode is compleet	4
De methode heeft het vermogen om te reageren op de schommelingen van de omgeving	5
De methode komt overeen met het experiment	

Step:	Usefull	Understandable
1	5	5
1.1	5	5
1.2	5	4
1.3	5	5
2	5	5
2.1	5	5
2.1.1	5	5
2.1.2	5	4
2.2	4	5
3	5	5
3.1	5	5
3.2	5	4
3.3	5	5
4	5	5
4.1	5	5
5	5	5
5.1	5	5
5.2	5	5
5.3	5	5
5.4	5	5
6	5	5
6.1	5	5
7	5	5
7.1	5	5
7.2	5	5
7.3	3	4
7.4	5	5
8	5	5
8.1	5	5
8.3	5	5
8.3	5	4
9	5	5
9.1	5	5
9.2	5	5
9.3	5	5
10	5	5
10.1	5	5
10.2	5	4

Appendix C.3: Interview IN3

***Steven**:* Laat ik gewoon beginnen met wat en wat ik gedaan hebt en wat het idee is. Ik ben bezig geweest met het ontwikkelen van een stappenplan methode die organisaties, en dan SBR organisaties, helpt om hun eerste machine learning experiment op te zetten. Voor jou zijn waarschijnlijk de echnische stappen heel simpel. Maar het gaat in principe om het totale plaatje za dat men de eerste machine learning experimenten kan neerzetten, rekening houdend met een stukje ethiek en een stukje organisatie. Daar ga ik zo nog even op door maar ik vergeet nl iets wat belangrijk is voordat ik dit interview mag gebruiken ik met jou doorneem dat ik het interview opnemen dat ik het daarna het transcript zal transcriberen n en samenvatten, en die transcriptie en samenvatting, aan jou zal toesturen zodat je kan kijken of je het eens ben met wat daar staat. Daarnaast wordt ook een consent formulier gestuurd waarin staat dat de informatie die ik krijg zal gebruiken in mijn thesis en wordt het wel openbaar gemaakt. Dus van tevoren: ga je daarmee akkoord nu dat het wordt opgenomen en de vervolgstappen?
***IN3**:* Ja hoor
***Steven**:* Top. Dan ga ik eerst mijn scherm delen. Dit is mijn versimpelde versie van mijn stappenplan. Wanneer dit plan gebruikt wordt, zit daar ook context bij dus iets meer informatie dan wat je nu ziet. Alleen voor de purpose of de interviews gaan we het vanaf deze insteek bekijken en ga ik je daar een x aantal vragen over stellen. Zie jij nu het stappenplan?
***IN3**:* ja hoor
***Steven**:* Wat ik hiermee wil doen is, ik heb nu tien stappen met een aantal kleine stappen en ik wil graag weten, per stap, of die begrijpelijk is en of die nuttig is voor het opzetten van een machine learning experiment. Laten we eerst door de stappen heen lopen, of er nu al vragen zijn.
***Steven**:* Als je er nu snel doorheen scant, zijn dan dingen die je je al afvraagt.
***IN3**:* Moet ik even doorlezen.
***Steven**:* En kan ik wel een klein beetje context geven bijvoorbeeld over hoe we het nu gedaan hebben? We zijn bijvoorbeeld met WSW bezig geweest van SBR wonen. Zij bepalen op basis van financiële data of een organisatie een financieel risico is of niet. Dat hebben wij met machine learning in zekere zin geprobeerd na te maken. Daar omheen hebben we deze stappen doorlopen om tot het machine learning model te komen. Nou, dat was in zekere zin een succesvol en we hebben bijvoorbeeld een decision tree algoritme gebruikt om uiteindelijk de voorspellingen te doen. Zo kwamen we tot een eerste model. En wat het WSW nu verder gaat doen, is een expert in de hand nemen om de vervolgstappen te doen. Maar het doel hebben we dus wel bereikt. Dat is dus het eerste experimenten opzetten en te kijken of het überhaupt mogelijk is om machine learning te gebruiken om te doen met wat zijn willen. Yes?
***IN3**:* ja
***Steven**:* Oké, dit is nu een heel systematisch stukje wat ik ga doen, want ik wil dus per stap weten of die een nuttig is en of je daar dan sterk mee eens of oneens bent. Eens, neutraal, eens of sterk, en dat zou je dan in de cijfers één tot en met vijf kunnen doen. En of die begrijpbaar is . J had net een aantal dingen bij stap, drie, stakeholder analyse. Je zegt dat eigenlijk de stakeholders al in het projectteam gezet moeten worden? Begrijp ik dat goed?
***IN3**:* Nou, het was voor twee, twee, including het management. Maar dat betekent dus ook include de stakeholders. Ik denk dat dat gewoon een hele diverse groep moet worden van hoog tot laag in een organisatie.
***IN3**:* Management kan dat fantastisch vinden om met A.I. aan de gang te gaan maar als een andere stakeholders het niet zien zitten, dan gaat je projecten mislukken.
***Steven**:* En daarmee bedoel je dat eigenlijk alle stakeholders van de organisatie betrokken moeten worden. Dat is iets wat ik inderdaad dan in het projectteam duidelijker neer moet zetten. En bij die context analyse bedoel ik meer dat de stakeholders buiten de organisatie geanalyseerd worden.
***IN3**:* Oké, bij WSW ik mij daar iets bij voorstellen omdat daar een hoop externe stakeholders zijn.
***Steven**:* Ja dat is natuurlijk het geval. De projecten zijn bij SBR stakeholders. En ook bij duo gaat het om organisaties die beoordeeld worden. . Het is wel vaak dat er nog buiten de eigen organisatie mensen beïnvloed worden of effected zijn door de door het doel van het project.
***IN3**:* Wat ik daar ook nog bij wilde noemen, is dat een stukje verwachtingsmanagement nodig is van wat je kunt verwachten, van AI.
***IN3**:* waarbij mensen soms de meest wilde ideeën, hebben we bij wat allemaal kan.
***Steven**:* Oké, een verwachting, en waar zou je dat dan graag zien?
***IN3**:* In het principe al in het begin van je gaat het doorzetten en is dat realistisch? Past de verwachting van een organisatie bij de mogelijkheden van a i.
***IN3**:* Zegt iets over de mogelijkheden van a i, maar ook een stukje over de verwachtingen van het management.
***Steven**:* Oké, dat is een goeie aanvulling. Zijn er voor de rest nog andere dingen waarvan je denkt, dit wil ik anders zien?
***IN3**:* Bij het stukje data mis ik nog een stukje feature engineering. Ik weet niet waar overigens een Weka tool je daar in gaat helpen.
***IN3**:* Even om een voorbeeld van duo te gebruiken als ik dat kan bedenken.
***Steven**:* financiële data uit bijvoorbeeld XBRL?
***IN3**:* Ik geef en verander voorbeeld wat gewoon wel concreet is. Stel dat je een ticket systeem hebt waar je waar je vragen kan stellen aan een helpdesk of verstoring kan melden, en daar wordt altijd vooral opgenomen wanneer een ticket aangemaakt wordt.
***IN3**:* En met de datum en tijd zo dan kun je wel iets mee doen. Maar het kan bijvoorbeeld interessant zijn om uit die datum, de dag van de week te destilleren
***IN3**:* Om te kunnen zien van nou, hoeveel procent van de data komt nou op een maandag binnen of op een vrijdag, en dat zijn soms heel bepalende dingen. En stel dat je een a i service wilt maken om de bezetting van je servicedesk mee te kunnen plannen, dan zou je dat met dat soort extra features kunnen doen.
***Steven**:* Inderdaad, bij het experiment met WSW waren dat bijvoorbeeld ook al van tevoren een aantal data attributen die bijvoorbeeld gedeeld door elkaar werden, al meer informatie gaven dan asl dit niet gebeurt.
***IN3**:* Ja, relatie inderdaad kan een hele goeie zijn.
***Steven**:* Oke, dat dus explicieter neerzetten,
***IN3**:* ja,

Steven: en dat wordt dan ook gedaan door iemand die de domein kennis heeft van de data, dus een expert op de data van het bedrijf zelf.
IN3: Ja, en in samenwerking met iemand die wat meer kennis heeft van ai.
IN3: Een puntje vijf, drie ; ik was ook wel benieuwd omdat ik op de gitlab gezien heb op wat voor een manier die duo data daar beschikbaar is gemaakt. Dat is een soort excel dat je zeg maar.
Steven: Ja.
IN3: Zit er toolbox die nu bedacht wordt? Zit daar ergens een tool in dat de XBRL-fomaat kan omvormen naar dat excel formaat.
Steven: Nee, de data van het duo is wel csv, en dat wordt dan weka readable gemaakt. Die zit daar een stap. En dan gaat het naar ARRF.
IN3: En reageer ik op omdat je in dat excelletje ook weer data mist die in de bron data weer wel zit.
Steven: ja, dus dat is het probleem en ook in een stukje van de mutatie. Ik heb nu niet alleen op die XBRL data, maar ook op de highly structured data. De stap naar XBRL heb ik niet kunnen maken en zeg maar zoals nu bijvoorbeeld, die duo data. Het is inderdaad, je verliest wel de mate van structuur en wat je zegt in in jouw blok, is dat dat natuurlijk wel zonde is.
IN3: hoe moet een organisatie te werk gaan om met dit proces vanuit de ruwe data tot een rekenmodel komen te komen?
Steven: Als die csv data beschikbaar is dan kan dat omgezet worden tot data leesbaar door weka
IN3: Ja, en ik weet dat duo is een soort eigen database ware house heeft waar vroeger ook al alle data in kwam, dus zal die csv dan de export van zijn. En ik weet dat dat WSW een soort van conversies slag heeft gemaakt om de XBRL data in een relationele database te krijgen.
Steven: Ja, dat bedoel ik met die stap. Het is nu wel verplicht om het gebruik te maken van weka, de workbench die ik gebruikt heb.
IN3: Plus dus nog moeten er wel partijen zijn die dus al die kennis en kunde hebben om de eerste stap te maken, steeds vrij om te zetten.
Steven: Ja, dat is wel wat ik probeer in die eerste stap, twee. Inderdaad dat het idee is dat je wel iemand vanuit het bedrijf hebt die de data experience en toegang heeft tot. Toen we het project met WSW, het experiment met WSW, waren er ook twee jongens die de domein kennis hadden en toegang.
IN3: En je de flow, zeg maar, als op één volgende stappen en het kan best zo zijn dat je stap zeven bemerkt dat er iets in de data zit, waardoor je slechte resultaten krijgt.
Steven: Ja, dan moet je, dan kun je weer terug.
IN3: Moet je weer terug naar stap vijf, dus dat is eigenlijk een soort van iteratie loop waar je misschien wel een paar keer doorheen moet voordat je een goed model hebt met goede scores.
Steven: Ja, dit moet ik er duidelijker inzetten .
IN3: Je agile technieken zou mooi passen, denk ik, Het is een beetje denken in iteraties.
Steven: Oké, en voor de rest.
IN3: Ja, die algoritme selectie: wat daar te binnen schoot was dat je hebt waarschijnlijk nu vooral algoritmes die goed met getalletjes om kunnen gaan.
Steven: Ja.
IN3: En er is ook best wel data waar het over tekst gaat. Zou je iets met teksten te kunnen doen.
Steven: De experimenten die ik nu gedaan heb, zijn wel gebaseerd op de financiële data.
IN3: Dat moet je dan ook duidelijk moet maken in het begin van je proces; van is deze aanpak geschikt om een bepaald probleem aan te pakken. Het is een soort van referentiekader wat je moet benoemen als van dit soort problemen kun je wel met deze aanpak doen, maar deze ook niet.
Steven: Ja inderdaad, dat is een goeie aanvulling. En voor de rest.
IN3: Waar ik bij stap 10 wel aan zat te denken maar misschien is het op 11. Op een gegeven moment wordt het in gebruik genomen en een beetje afhankelijk van wat voor data is en wat voor model, moet je ook gaan her-trainen om omdat de data misschien op den duur net iets anders wordt.
IN3: Het hele proces is in die zin dan ook weer iteratief en dat je er altijd weer mee bezig zult moeten zijn om het meest optimale model te hebben.
Steven: Constant dus. Ja dat ja, ik denk dat dat natuurlijk ook wel een beetje zo is dat ik hierbij eigenlijk ook probeer neer te zetten. Dat men het eerste experiment kan voltooien. En wat je zei net ook, qua getallen, voor of je met deze methode een effectief machine learning project kan opzetten. Daar was je neutraal en ik denk dat dit is wat je net bedoelt te zeggen: Dit is leuk misschien de eerste stap te zetten, maar het totale constante werk wat je nodig hebt om een goed project op te zetten, daarvoor mis je nog onderdelen zoals je benoemt.
IN3: Dat is wel één van aspecten. Iets anders kan zijn dat je een model kan trainen met de beschikbare data en dat het model fantastisch performed met jouw test data en op het moment dat je er nieuwe data ertegenaan houdt het dan blijkt dat toch niet zo goed doet. Dat ,zo op gefocust is op de patroontjes in de trainingsdata, je een overfit krijgt en dat het wel heel lastig wordt voor een organisatie die alleen maar dit als een soort draaiboek heeft om om daar iets mee te doen.
Steven: Ja, twee reacties inderdaad. Maar ik split wel eerst de data in een training sit en een test set om toch dat overfitten deels te voorkomen of in ieder geval controleerbaar te maken.
IN3: Want dan heb je waarschijnlijk van een jaargang, of weet ik veel hoe je de data krijgt. Dat zegt nog niet alles over hoe de data morgen eruit ziet
Steven: Dat is waar ik ben ook wel eens. Dat dit niet een methode is die een constant machine learning project, gaande te houden. Misschien is dan de naam project niet eens per se de beste benaming. Want het idee is dat je vooral die eerste stappen kan zetten, zodat je kan er überhaupt kan kijken of machine learning een optie is en of het of je iets werkends kan krijgen met de data die je hebt. Daarvoor geeft het handvatten. Dat is op zich prima.
IN3: Dan moet je dit ook heel helder maken waar je dit voor gaat gebruiken.
Steven: Ja, dat moet ik duidelijker zeggen. Oké, vind ik prima en als je dan kijkt naar het opzetten van een experiment. Is het, dan is het dan wel in z'n totaliteit effectief om een machine learning experiment op te zetten en alle stappen te doorlopen?
IN3: Ja, dan zou ik zeggen dat je ook weer rekening moet houden met wat is het resultaten wat je uiteindelijk het management wilt vertellen, welke boodschap je wilt overbrengen? Bijvoorbeeld, Ja, het is haalbaar, maar dan moet je wel dit en dit en dit doen, maar wel op zo een manier.
IN3: Het zou anders natuurlijk heel gevaarlijk, zijn als men ziet van nou, er komt een hele goeie score uit: zet maar in productie.

IN3: Je moet aan het begin al bezig zijn met aan te geven van we gaan dit doen, maar het wordt geen eindproduct.
IN3: Dus dan moet je het echt als een soort van prototyping tool positioneren om niet in die valkuil te trappen dat management denkt van nou, we zijn klaar.
Steven: Prima en bedankt. Pak ik op. Dat waren eigenlijk wel mijn vragen. Heel erg bedankt en als je nog andere onderdelen hebt of een vragen, dan kun je natuurlijk doorgeven aan mij.
IN3: Op dit moment niet, als het me nog te binnen schiet, dan weet ik je te vinden.
Steven: Dan wil ik heel erg bedankt. Ik ga dit gesprek transcriberen en zal je een samenvatting met de punten die jij net gezegd hebt, opsturen. En daarnaast ja, komt er een paar pagina's aan van de inhoud dit gespreken. Ik moet het in ieder geval opsturen. Nogmaals heel erg bedankt.

	Cijfer
De methode is effectief in het bereiken van een machine learning project	3
De methode is praktisch goed te gebruiken	3
De methode is makkelijk te begrijpen	4
Het model past bij mijn organisatie	
De methode is compleet	2
De methode heeft het vermogen om te reageren op de schommelingen van de omgeving	3
De methode komt overeen met het experiment	

Step:	Usefull	Understandable
1	5	4
1.1	5	4
1.2	5	4
1.3	5	4
2	5	4
2.1	5	5
2.1.1	5	5
2.1.2	5	5
2.2	5	5
3	5	3
3.1	5	4
3.2	5	5
3.3	5	5
4	5	5
4.1	5	5
5	5	5
5.1	5	5
5.2	5	5
5.3	5	3
5.4	5	5
6	5	3
6.1	5	3
7	5	5
7.1	5	5
7.2	5	5
7.3	5	4
7.4	5	4
8	5	4
8.1	5	4
8.3	5	4
8.3	3	3
9	5	4
9.1	5	4
9.2	5	5
9.3	5	5
10	5	5
10.1	5	5
10.2	5	4

Appendix C.4: Interview IN4

Steven: En daarna krijg je het transcript hiervan, daarna wordt de recording verwijderd en het transcript krijgt je toegestuurd zodat je kan kijken of je het eens bent met wat daarin staat. Dat is gewoon letterlijk vertaald, maar ik doe er ook een klein samenvatting bij dat je niet dat hele ding door te lezen. Voor de rest gebruik ik de antwoorden die jij geeft om mijn methode te verbeteren, evalueren. Als er een hele mooie quote in zit, dan zou ik die mogelijk willen gebruiken en zou je naam erin komen maar in principe komt jouw naam niet in mijn thesis. Ga je daar van tevoren mee akkoord.
IN4: Zeker.
Steven: Dan gaan we gewoon gelijk door. Ik ga eerst uitleggen wat de methode ongeveer is, dan daarna moeten we stap voor stap elke stap een cijfer geven op usefullness, van is dus of het usefull is dus en begrijpenbaar is, en daar kunnen we best wel snel doorheen of in ieder geval handig om dat dan snel te doen en dan gaan we daarna komt er een stuk kwalitatief, waarbij jij vertelt of je nog dingen mist in de methode in zijn totaal. Even korte uitleg van de methode. Het is een soort stappenplan en bij elke stap hoort een opdracht, een vraag die gesteld moet worden en het idee is dat je door middel van deze methode een machine learning experimenten op kan zetten. Kijk even snel door een methode en zijn er dan dingen waarvan je nu al denkt, dit snap ik nu niet.
IN4: Weet je of er een doel past bij machine learning.
Steven: Goeie vraag. Naast de plaatjes die je nu ziet, komt er wel een stukje context bij en daar worden ook voorbeelden geven van wat voor een machine learning is . In dit geval zou dan zou dan de onderliggende context, machine learning, twee varianten, unsupervised, supervised.
IN4: Inderdaad, prima.
Steven: Het is moeilijk misschien inderdaad, want sommige dingen zijn niet in z'n totaliteit uitgelegd, maar dit is vooral het actiepunt en daaronder ligt dus soms nog context. Laten we gewoon de methode doorlopen en dan kun jij zeggen of je het nuttig vindt en of je het begrijpelijk vindt. En als er dan dingen bijvoorbeeld, als je nog een klein beetje context erbij wilt, dan geef ik u dat op dat moment. Ga ik je ook vertellen hoe dit gaan doen, want ik gebruik de likert schaal oftewel, één tot en met vijf sterk mee oneens, oneens, neutraal eens, sterk mee eens.
Steven: Dus dat waren alle cijfers en vragen. Dus dat hebben we snel gedaan. Ik heb drie dingen nu opgeschreven. Je zegt compare result en select the model; dat is een beetje inherent aan elkaar.
IN4: Ja, als je resultaten gaat vergelijken, dan komt daar een conclusie uit. Dat verwacht ik dat jij in die conclusie zegt, daar gaan we mee verder. Het zou een mooi stapje zijn om het daarmee af te ronden en met het volgende blokje verder te gaan.
Steven: Oke, dat is duidelijk.
IN4: Dat vind ik trouwens altijd prettig bij maken van zo een model is dan de input van het volgende stap, wat is de output van het volgende stapje. Misschien heb je dat in je tekst daarover
Steven: Ja.
IN4: Logisch natuurlijk bij zes uitkomt dat daar op geselecteerd wordt; je hebt op meerdere modellen algoritmen getraind en dan selecteer je wat op zich natuurlijk logisch is.
Steven: maar die input en output wordt er inderdaad wel bij beschreven en kijk, dit is wel natuurlijk een versimpelde versie van de methode en het is wel dat er straks inderdaad meer lijntjes komen. Ik ben benieuwd of het dan nog moeilijker of juist minder moeilijk wordt. Maar dat gaan we meemaken.
Steven: Je zegt dat het communiceren van de tussenresultaten; dat je dat eerder zou willen zien.
IN4: Bijvoorbeeld bij twee zie ik nu ook staan nu ,stap twee include management.
IN4: Denk dat je daar met elkaar afspreekt we houden elkaar op de hoogte is, bespreken resultaten.
Steven: Maar is dan dat ik dat iets duidelijker moet neerzetten. Dat het een continue deling is van resultaten wat goed kan zijn voor de bedrijfsvoering
IN4: Ja, precies.
Steven: Leuk en IN2 zegt precies hetzelfde, dus je bent niet de enige, dus jij en IN2 denken daarin hetzelfde.
IN4: Ja, wat je vaak doet is een analyse; van hoe betrekken we stakeholder erbij , is het informatief, dan deel je gewoon wat met de stakeholder of wil die meer betrokken zijn. Dus inderdaad om de zoveel tijd gewoon een update geven, Praat je wat bij. er zijn verschillende manieren hoe je met zo een stakeholder omgaat kan je daar prima organiseren en dan kun je dat ook in het verdere proces toepassen.
Steven: Ja, een goede aanvulling: voor de rest, zijn er nog andere onderdelen die onduidelijkheden of dingen die je iets anders zou willen zien of dingen die je zou willen toevoegen.
IN4: Ja, misschien is de vraag van noem je dit project of misschien meer een andere naam. Weet ik niet, project heeft misschien weer een grote lading.
Steven: Een projecten of een experiment.
IN4: Inderdaad.
Steven: Ja, dat is ik heb gisteren met met IN gesprokenen die zei ook inderdaad van ja kijk, je zet inderdaad wel het eerste deel van experimenten op, maar als je een constant machine learning voor continue gebruik wil maken, dan komen er nog wel wat stappen bij. Is het een project, is een experiment?; dat moet ik misschien wat duidelijker neerzetten. Inderdaad, zoals als experiment, maar je zou ook kunnen zeggen dat het begin van een project is.
IN4: Ja, dat kan ook.
Steven: En voor de rest nog onderdelen?
IN4: Ja, misschien ets dat er al in zit hoor: . Kun je de resultaten zichtbaar maken, visueel voor de omgeving, maar misschien doe je dat ook wel. Ja, dat is er wel eentje.
Steven: Ja, misschien is dat toch wel in een github
IN4: Wat is nou het eindresultaat: je hebt model, dan heb je uiteindelijk check resultaten van een model, je hebt de resultaten gecheckt, je vergelijkt de resultaten, en je evalueert ze: Wat er leeft je uiteindelijk aan het einde van rit echt op.
Steven: Dat is inderdaad misschien ook wel gelijk het verschil inderdaad van een project of een experiment. Je levert de resultaten van een machine learning experiment in en daarbij kijken of het mogelijk is of we met de data die hebt, of de data die je hebt gekregen, kunt

uitvoeren wat het beoogde doel is. Dus in het geval van bijvoorbeeld de WSW was het om te kijken of we een risico model konden maken met de data die we hadden en wat de verbanden daarvan zijn en of je daaruit ook nieuwe inzichten kan halen.
IN4: Maar dat is eigenlijk communication dan de laatste stap? De laatste stap is iets met je doel: je checkt of je doel gehaald is, stap 10 dat je onderdeel maakt van communicatie over dat doel, als het gehaald is, dan worden de resultaten er bij gezet.
Steven: Is dat niet meer bij stap negen, project evaluation, en dan van hebben we het doel gehaald.
IN4: Zoiets ja.
Steven: Je doet nu iets met compare the result with the current situation, maar eigenlijk heb je in het begin al het doel gesteld. Is je doel gehaald.
Steven: Ja, dat is wel een goede.
IN4: Als het niet gehaald is, wat ga je dan? Ga je proces starten, het doel bijstellen of je ziet dat sommige modellen ook weer met elkaar verbonden zijn dan kan je blijven doorgaan, dit soort cadans op integratie.
Steven: Dat laatste is zeker goed. Inderdaad, want inderdaad, ik probeer dat wel met die project evaluation echter staat dat er inderdaad nu iets meer gericht op het model en een model is niet het enige dat je maakt gemaakt. Het doel is om een middel om een bepaald doel op te lossen.
IN4: Oh ja, evalueren heb ik altijd een beetje het idee bij van hé wat we nu allemaal gedaan hebben. Als er nog een keer zouden doen, hoe zou dat dan anders doen? Dat heb ik meestal mee met evalueren.
IN4: Evalueren is ook wel een beetje de pro's van agile: We hebben nu een sprint gehad. Wat ging er goed. Wat ging er minder goed. Communicatie ging niet goed dus hoe kunnen we zorgen dat we het de volgende keer beter kunnen communiceren? Dat vind ik altijd evalueren.
Steven: Zou je die dan anders noemen?
IN4: Zou ik het inderdaad wat jij net gezegd hebt evalueren of je je doel gehaald hebt, ze zeggen dat check results with project goal
Steven: Top, dan is hierbij het interview afgelopen

	Cijfer:
De methode is effectief in het bereiken van een machine learning project	5
De methode is praktisch goed te gebruiken	4
De methode is makkelijk te begrijpen	5
Het model past bij mijn organisatie	
De methode is compleet	4
De methode heeft het vermogen om te reageren op de schommelingen van de omgeving	5
De methode komt overeen met het experiment	

Step:	Usefull	Understandable
1	5	5
1.1	5	5
1.2	4	4
1.3	4	4
2	5	5
2.1		
2.1.1	5	5
2.1.2	4	4
2.2	3	3
3	5	5
3.1	5	5
3.2	5	5
3.3	5	5
4	5	5
4.1	5	5
5	5	5
5.1	5	5
5.2	4	4
5.3	5	5
5.4	4	4
6	5	5
6.1	5	5
7	5	5
7.1	5	5
7.2	5	5
7.3	5	5
7.4	5	5
8	5	5
8.1	5	5
8.3	5	5
8.3	4	4
9	5	5
9.1	3	3
9.2	5	5
9.3	4	4
10	5	5
10.1	5	5
10.2	5	5

Appendix C.5: Interview IN5

Interview IN5

***Steven**:* Bij een interview hoort ook een consent formulier. We gaan hier kort doorheen en dit krijg je meegestuurd. Ik ga het interview transfigureren. De recording worden verwijderd. Ik maak er ook een samenvatting van deze transfiguratie, zo dat je er ook in het kort even snel doorheen kan kijken welke punten ik gebruik. Ik zal je naam niet noemen in het onderzoek, alleen als er een hele mooie quote uitkomt, dan zou ik wel graag je naam noemen. Het kan zijn dat je zegt van nou, dan hoef ik niet. Aan het einde word mijn thesis wordt gedeeld op de repository van de TU Delft, dus daardoor is die ook openbaar. Daar moeten wij sowieso misschien even naar kijken, want ik heb natuurlijk met WSW ook het contract getekend dat ik niks mag delen voordat er consent is. vanuit jullie.

***IN5**:* Het is goed het daar nog even over te hebben van hoe gaan we dat dan doen.

***Steven**:* Ik heb natuurlijk in principe niet de data nu gehad. dus dat scheelt al. Zie jij nog problemen eigenlijk?

***IN5**:* Nee. Ik weet niet hoe je je scriptie precies het ingedeeld, of je bijvoorbeeld een hoofdstukje of paragraaf wijdt aan een WSW.

***Steven**:* Ja, dat wel

***IN5**:* Dan is het misschien goed om te zeggen deel dat stukje even om te kijken of er geen gekke dingen in staan.

***Steven**:* Wat ik nu nu gedaan heb; ik heb hier gewoon beschreven wat WSW is, want het doel was, financiële risico's van organisaties hoe het project is gedaan, context analyse. Wat voor een effect dat zou kunnen hebben. Ik heb hier wel de data die we gebruiken, hebben we? Dit hoeft niet per se. Je zou bijvoorbeeld ook kunnen zeggen van dat lijkt me niks. Daarna programma data preparatie. Welke modellen we hebben gebruikt, en dan welke accuraties we hebben gehad met modellen en dan focus op de ? en daarna een stukje over de model tuning, dat we met dit op een verbeterd resultaat kwamen. Daarna hebben we nog wat andere dingen geprobeerd, maar die laat ik buiten de methode. Ik zet nu alleen die adabooster erin. Dat is de enige die ik beschrijf en daarna is evaluatie. Dus het zal niet veel meer dan dit worden.

***IN5**:* Het lijkt niet heel schokkend en als je het geen probleem vindt, dan zou het fijn vinden om dit nog één keer toegestuurd te krijgen.

***Steven**:* Tuurlijk, Ik zou het gewoon doorsturen, maar zoals je ziet zit dan alleen nog de interviews die ik nu geef zou ook nog datakunnen zijn voor mijn methode. Maar dan gaat het vooral om de methode en niet meer om de gevoelige informatie als het goed is.

***Steven**:* We gaan vandaag mijn methode bespreken.

***Steven**:* Wat ik gedaan heb is een methode gemaakt om een machine learning experiment op te zetten, rekening houdend met de organisatie en context. Het is nu een versimpelde versie, dit soort stappenplan enna morgen heb ik alle interviews gedaan en dan ga ik een verbeterde versie maken. Wat ik nu met jou zou willen doen is in driefout. Ik wil eerst drie vragen stellen over het experiment. Daarna gaan we stap voor stap door die methode waarbij ik je zal vragen een cijfer te geven voor elke stap. Daar kunnen we gewoon even snel doorheen knallen en daarna aan het einde, vraag ik jou of jij dingen anders zou willen zien en of je nog dingen zou willen toevoegen. Dat is het doel van vandaag. Dan ga ik nu inderdaad de eerste vraag stellen. We hebben een experiment met WSW gedaan, een machine. Learning experiment, denk jij dat dit nuttig was?

***IN5**:* Ja.

***Steven**:* En waarom.

***IN5**:* Dus wij zijn dan tijdje bezig om te kijken wat er nou precies doen met onze data, wat meer uit zouden kunnen halen, want we gebruiken al best lang heel veel data. Alleen zijn we nog een beetje zoeken of nog meer mogelijkheden zijn, maar één van de mogelijkheden die we hadden bedacht, is machine learning. Het is een techniek die hadden tot nu toe iemand niet gebruikt, misschien niet het mogelijkheden die iemand niet kennen, waar we wel eens ver dat we dachten dat we een keer mee zouden willen experimenteren. En nog voordat we dat dan goed en wel op papier hadden geschreven, kwam het aanbod vanuit jouw hoek voorbij om een experimenten te gaan voeren. Dus dat kwam eigenlijk op een heel mooi moment.

***Steven**:* Denk je dat het met machine learning mogelijk is om de huidige situatie te verbeteren?

***IN5**:* Ja, daar heb ik zo mijn twijfels bij. Wij voor wij het willen gebruiken is om risico's door te rekenen. Waar we ons steeds beter bewust van worden, is dat risico 's misschien niet zozeer kunnen worden gevonden door te kijken naar wat er in het verleden is gebeurd, maar dat je meer aan scenario's zou moeten denken. Er komt een keer iets voorbij dat nog niet kennen; hoe groot zijn dan de klappen die we kunnen opvangen. Als je steeds meer naar het verleden blijft kijken, dan zie je dat je alleen maar gaat kijken of die risico's die zich al een keer hebben voorgedaan, of je die nog een keert verwacht.

***Steven**:* Ja.

***IN5**:* Jou eigenlijk heel vaak hebben we gezien dat als er een keer een probleem voordoet, dat dat een ander probleem is dan dat je hiervoor hebt gezien.

***IN5**:* En je zag ook een beetje in lijn daarmee dat het model waar we nu mee hebben gewerkt, verklaard wel voor een deel de problemen die we in het verleden hebben we gezien, dus je ziet wel een link tussen de bron Data, maar die link is niet heel sterk. Het is ook een beetje afhankelijk van welke methode je kiest, zie ???? je of degene die wel bijzonder meer ziet. Er zitten heel goed en degene die ze niet in zitten, niet, of andersom het al om een beperkt. Dat blijkt heel lastig te zijn om op basis van mijn data die we hebben, echt te verklaren, of er wel of niet probleem komen. En je kunt wel gebruiken in aanvulling op methode die wij hebben om eigenlijk op wat geavanceerdere manier een bepaalde correlatie s eruit te halen. Dat je bijvoorbeeld wel zag dat je een link heb tussen s hoe hoog de schuld is die een corporatie heef en of die al dan niet in de n problemen m kom. Dat zit op zich ok onze huidige relatie maar komt nu e op een andere manier naar voren. Dus da soort s correlatie kun je misschien op n een andere manier je eruit halen. ? Ik zie het niet echt als een vervanging van bestaande methode om dat je toch te veel moet steunen om historische data.

***Steven**:* Ja, en maar de keerzijde is natuurlijk inderdaad wel dat je bijvoorbeeld het huidige systeem blijft gebruiken, maar dat je door middel van machine learning mogelijk nieuwe inzichten krijgt om het huidige systeem te verbeteren.

***IN5**:* Ja, dit is met name voor dat soort dingen kan het interessant zijn, Houdt het systeem nu hebt en kijken of je met wat eigenlijk geavanceerdere wiskunde andere dingen kunt zien waarmee we een bestaand systeem kunt aanscherpen, en dat zou dan het doel zijn waarvoor je het kunt gebruiken.

***Steven**:* Oké, goed om te horen. Inderdaad, het is niet alleen maar positief, maar het biedt ook wel weer een andere kijk op het gebruik van machine learning. Maar als we dan even naar het experiment wat wij hebben gedaan, ben je tevreden met hoe het experiment verlopen is en waar dat in heeft geresulteerd.

IN5: Ja, ik vond eigenlijk we dat die applicatie best wel makkelijk te gebruiken is. Ik had mij voor die tijd wel wat verdiept in machine learning en ik zag het gebeuren dat je allerlei algoritmes zelf bij elkaar moet gaan zoeken en items in moest gaan gooien. Wat we hebben gebruikt is eigenlijk best wel makkelijk te gebruiken. Wel de beperking dat als dat als een relatief leek gebruikt, dat je dan lukraak algoritmes aan het aanklikken bent. En ik niet weet hoe het eruit komt bij toeval of omdat het gewoon een goed algoritme is, maar het is wel heel praktisch toepasbaar. En ja, wat je ziet, net als bij heel veel andere data analyses. De grote ellende zit bij het van verzamelen van de data. We hebben geprobeerd om vanaf 2012 alle data te ontsluiten. En nou ja, dat is eigenlijk weken aan tijd in gaan zitten. En vervolgens komt dit experiment en hebben we het in drie sessies, van elke ontving uur gedaan. En dan heb je al een analyse gedaan, en zie je zelf eigenlijk doordat de applicatie best wel makkelijk is, een klein stuk deel van het werk, dan moet je wel goede data hebben en snappen wat je doet.
Steven: Ja, maar dat is dus ook weer mooie van het hebben van gestructureerde data; dat je inderdaad al snel die analyses kan uitvoeren. En inderdaad jij zegt van, ik wil eigenlijk niet naar een systeem gaan wat ook situaties analyseert; dat op het moment dat dan die situatie die wij beschreven hebben komt, dat wij daar snel op kunnen reageren. Misschien zou het op een andere manier ingezet kunnen worden, waarbij je bijvoorbeeld machine learning gebruikt om meer de huidige situatie te beschrijven. Zo dat je weet op het moment dat die situatie komt, dat je weet ik nu moet om op een andere manier te gaan handelen.
IN5: Om een voorbeeld te geven; er zal een algoritme zijn waarschijnlijk dat voorspelt ???? dat eigenlijk gebeuren te komen. Dan kun je nog zoveel machine learning op loslaten, maar dat ga je niet vinden. Dan wil je wel eens weten dat als zich een keer z iets voordoet, en het kan ook een keer iets anders zijn we ook nog niet kennen dat eraan komt, wat kun je bijvoorbeeld opvangen aan daling van de waarde van je bezit of een verslechtering van je kas stromen.
IN5: En dat zijn allemaal dingen die je heel simpel kunnen doorberekenen, en dat geeft misschien al veel meer inzicht dan dat je op basis van historische data gaat proberen te extrapoleren wie er misschien in de problemen zou kunnen. Je hebt het allebei wel nodig,. Om in de toekomst te kijken, machine learning zou je dan kunnen gebruiken om in je historische data te kijken of je nog aanvullende dingen ziet die het huidige model nog niet heeft gezien.
Steven: Maar als je bijvoorbeeld, corona is natuurlijk een moeilijk voorbeeld, maar neem SARS; Wimbledon is de enige die hier bijvoorbeeld goed op gereageerd heeft, en gezegd, wij willen ook een verzekering hebben waarbij onbekende pandemieën ook verzekerd zijn. Dus er zijn organisaties die al wel de situatie hebben voorspeld dat het zou kunnen gebeuren dan ben je wel aan het baseren op het verleden. Ann de andere kant is natuurlijk zo dat je zou kunnen zeggen dat dat jullie 100 hypothetische situaties zouden kunnen schetsen
IN5: Het is goed want die dingen gebruiken we wel daadwerkelijk. Wij hebben bijvoorbeeld een scenario model: er zitten 1000 macro-economische scenario's in; allerlei variaties van inflatie, bouwkosten, renteontwikkeling dat soort zaken. En dan gooi je gewoon al je bron data door dat scenario model heen, je data aanpassen en wat je daarmee kunt doen is als je er 1000 doorberekent en je pakt dan een paar slechtste resultaten, het 999 slechte scenario, kun je zekerheid op 99,9 procent zeggen dat jij ???? en alle maakt. Het economisch scenario is dat je doorgerekend een bepaalde voorwaarden.
IN5: Het blijft natuurlijk ook maar statistiek voor 99,9 is ook maar gebaseerd op de input die je er in stopt.
Steven: Ja.
IN5: Maar zo zie je wel, voor de toekomst kun je dus heel veel doen. Maar scenario's en wat ik voor ons dus vanuit het risico perspectief vooral zie is dat je machine learning kunt gebruiken om het verleden wat beter te verklaren, maar ik vind het nog lastig om te zeggen van we gaan een machine learning gebruiken als voorspelling model.
Steven: of als real time
IN5: Ja
IN5: Je krijgt de data maar één keer per jaar.
Steven: Dat is natuurlijk inderdaad een probleem, maar je kan misschien op het moment dat je het binnenkrijgt, binnen een dag analyseren, maar dat doe je in principe met model nu ook al.
IN5: Het hangt ook een beetje af van welke data. Met de data die we hebben en op de manier waarop wij werken, krijg je een paar andere dingen die je moet machine learning wel of niet kunnen doen en stel je zit in een hele andere sector, of je hebt een hele andere dataset, een hele andere dingen.
Steven: Oké, nou, dat is voor mij genoeg over het experiment. Wat we nu gaan doen is inderdaad naar de methode. Die zie jij nu ook op je scherm . Het is een soort stappenplan met de verschillende iteraties; denk dat je eerst even moet kijken, of je alles een beetje begrijpt. Wat je normaal hierbij krijgt is ook de context, maar op het moment dat ik bij de methode ook alle context er in gaan zetten, dan wordt het wel heel uitgebreid, dus dit zijn de essentiële stappen die doorlopen worden en bij elke stap hoort nog een beschrijving. En als je er dus even snel doorheen kijkt, zie je dan dingen die niet begrijpt?
IN5: Ik heb nog niet gekeken naar de zwarte tekst eronder, maar wat er in die blauwe pilot staat, dat ziet er op zich wel redelijk logisch uit.
Steven: Dit is een combinatie inderdaad van hoe je over het algemeen een knowledge discovery doet uit data en dan rekening houdend met ethiek en organisaties.
IN5: Ja, je zegt het zelf als, het is iteratief dus het kan best zijn als je bij stap vijf bent dat je denkt ; Hé gaat iets niet goed, moet een paar stappen terug.
Steven: Dan moet ik er nog iets duidelijker inzetten. Inderdaad, maar ik heb het nu ook versimpeld, maar zoals ik het nu voor ogen hebt, is inderdaad één tot en met drie iteratief en dan vier een cirkel en vier tot het acht is een cirkel.
IN5: Ja, ik denk dat je in elke stap wel terug zou kunnen.
IN5: Wat je vaak ziet in data analyse is elke keer dat je weer een stapje verder bent, loop je tegen bepaalde problemen aan. Je verwacht bijvoorbeeld dat bepaalde data er zijn, maar ze zijn of ze er niet. Of ze zijn er wel maar het zegt niet zo veel of je zo nog moet bewerken. En heb je bewerkt en zie je de eerste uitkomsten, dat kan toch niet kloppen en moet je nog een stap terug. Het opschonen of ik moet toch andere data pakken, dus eigenlijk vanuit bijna elke stap dat je je data moet aanpassen, of misschien moet je wel zeggen: nou, we hebben probeert dat te doen met de data, we hebben eruit gehaald wat we willen, dus in stap één of drie is iets misgegaan en moet je iemand anders erbij betrekken, dus stap twee gaan passen.
IN5: In theorie kunnen zou je dit misschien zou clusterbommen, maar in de praktijk zie je toch dat je vanuit zo'n beetje elke stap wel een keer terug zou kunnen gaan.

Steven: Ja, elke stap na iedere stap
IN5: Ja van stap één tot en met tien, het geld op zich wel, maar vanuit welke stappen je een keer terug kan proberen te visualiseren, dan wordt het echt een band.
Steven: Da ga nu kwantitatieve data verzamelen voor jou, oftewel ik wil weten wat voor cijfer je geeft en daarna bespreken we de de verbeterpunten en de aanpassingen voor zover nodig.
Steven: Met alle getallen zijn we nu klaar en dan kunnen we inderdaad weer even een stukje terug. Want inderdaad er zijn eigenlijk best wel wat dingen over verteld tijdens het gesprek en ik ga ze dus nog wel herhalen. Wat jij hebt gezegd, het project goal en het machine learning goal, dat zou je graag anders willen zien, als twee aparte onderdelen.
IN5: Dat zit met name dan in stap drie, want als het goed is heb je als organisatie, dan een strategie en als je een project wilt beginnen en dan toets je dat project eerst aan die strategie. Pas wanneer je dat hebt gedaan, ga je kijken; Zou voor dit project machine learning relevant zijn . als je eerst een projecten hebt bedacht je toetst of machine learning daar bij past, is dat wel geschikt met deze strategie, dan is het de omgekeerde volgorde.
Steven: Conclusie: we hebben inderdaad aan het begin bekeken
IN5: je komt met 1 -3 onderhoud ??? aardig in de buurt
IN5: Kijk op het moment dat een project bedenkt dat niet aansluit op de doelen van het bedrijf, dat moet je überhaupt niet aan dat project beginnen en kom je aan mijn vraag of machine learning geschikt is.
Steven: In stap één wordt natuurlijk wel gezegd, does the goal fit with the mission of the company. Dus laten we zeggen dat de strategie en het project dan overeenkomen en daarna wil je pas kijken of het past bij machine learning en dan current situation of the goal zou dan aar stap één moet gaan.
Steven: Want stakeholder analyse is meer iets in twee.
IN5: Dat je deels interne en deels externe stakeholders onderscheid
IN5: Ja, want als je al een heel projectteam samengesteld, maar je komt daarna erachter dat ze jouw project niet wil uitvoeren. Dan gaat er ook iets niet helemaal lekker.
IN5: In stap één heb je dat heel aardig uitgewerkt en dan is het ook weer een beetje iteratief, want je hebt dus een project bedacht op de missie van je bedrijf en uit allemaal analyse methodes die er zijn heb je vastgesteld dat machine learning geschikt is. Je haalt de juiste mensen erbij.
IN5: En zal je daarna inderdaad verder uitdiepen van wat kunnen we nou precies doen? Wat voor data hebben we nou; zijn er misschien nog andere methodes die we kunnen toepassen? Is dit ethisch verantwoord, etcetera. Dus dan ga je bij Drie twee staat verder uitwerken. Hoofdstap een heb je als het goed is dan wel gedaan, maar drie kun je nog wat verder doen. Dat steekt in elkaar.
Steven: Maar kan je dit dan zomaar doen zonder de expertise van de verschillende mensen?
IN5: Nee, meestal is waarschijnlijk één persoon die een project start en of het idee heeft; uiteindelijk moet iemand een initiatief nemen. Dus ik begin mij een beetje als vragen of één en twee niet samen lopen. Want één iemand bedenkt we gaan doen.
IN5: Kan ook zijn je bijvoorbeeld een MT die zegt: we gaan doen of het kan gewoon een individu zijn die dit heeft bedacht, dat project wat we gaan doen heeft een bepaald doel en sluit aan bij de missie van het bedrijf. En machine learning zou misschien zou misschien wel kunnen, maar dan begint het als een heel klein ideetje of projectplan en dan ga je meer mensen erbij halen. Als je die meer mensen hebt, dan ga je projectplan verder uitwerken.
IN5: Zou je kunnen zien als drie, werk je projectplan verder uit en betrek dan alle aspecten erbij en kijken naar wat willen we nou precies doen? Hoe willen dat precies doen? Hebben we al die data? Is het ethisch verantwoord? Dat doe je. Dit is stap Één, twee, drie en oogt misschien ook een beetje door elkaar. En hoe werkt projectmanagement inderdaad, het begint met een klein project idee, Je haalt er een team bij, werkt het uit en heel klein projectplan.
Steven: Het loopt een beetje door elkaar eigenlijk.
IN5: Het begint met iets kleins en wordt steeds groter. Een, twee, drie is iets wat je in meerdere rondes steeds verder werkt
Steven: Mijn idee erachter is ook niet dat je dit in je eentje dat van tevoren allemaal al gedaan hebt.
IN5: Ik denk dat als je daar rekening mee houdt, dat het op zich gewoon allemaal wel klopt, en als je dan in hoofdstuk een, stap twee en drie omdraait dan kom je een heel eind.
Steven: En nog eventjes concreet, we moeten het eigenlijk een 2 stappen doen, het project formuleren en dan vervolgens kijken of machine learning daar bij past, of het doel fit met machine learning
Steven: en dus een tweede doel moeten maken voor machine learning
IN5: misschien komt dat überhaupt pas bij stap drie en ik vind het lastig. Ik denk pas dat als het het probleem verder hebt uitgewerkt , pas weet of machine learning erbij past. Maar als je eenmaal een machine learning aan de slag gaat, dan heb je het was wel weer de mensen nodig die daar verstand van hebben. Dus als je stap één, twee en drie in iteraties doet, misschien pas bij drie bepalen of machine learning geschikt is en begin je daarna het riedeltje van voren af aan.
Steven: Ja eens even kijken hoe ik dit ga verwerken.
IN5: Kijk als je twee ziet als je gaat extra mensen toevoegen en de eerste keer dat je bij één moment ben je nog meer alleen, dan kun je de eerste keer dat je bij één bent waarschijnlijk nog niet bepalen dat machine learning geschikt is. En pas als je je project wat verder gaat uitwerken, en dat zie ik wat meer bij drie, de context analyse, dan ga je erachter komen, als we dit allemaal weten dan blijkt inderdaad van alle analyse methodes te zijn, laat machine learning zien geschikt zijn.
Steven: Het is een deel van uiteindelijk kunnen zeggen en bij de context al geregeld moet zijn of machine learning geschikt is
IN5: Ja, dan ga je verder uitwerken wat je nu precies moet doen en hoe je dat gaat doen en kan het wel aan? Mag dat wel.
Steven: Nu heb ik in principe in stap één inderdaad set the project goal, does the goal fit in machine learning; dat zou je in theorie best wel snel kunnen zien,
IN5: maar misschien zie je die dat pas bij stap drie.
Steven: Ja, maar dat is op zich niet erg toch?
IN5: Nee.

***Steven**:* Want stel, je hebt een idee, je weet niet zeker of het bij een machine learning past. Jij hebt het er wel over. Je kijkt in globaal door middel van een beslisboom of het bij machine learning past met wat voorbeelden. Daarna, ga je het overleggen met de interne stakeholders; kijken wat van expertise ze in huis hebben. En dan ga je het eigenlijk wat beter uitwerken.
***Steven**:* Dan analyseren dus inderdaad echt de uitgebreide situatie, de ethische effecten en dat is eigenlijk ook weer een punt om te kijken of het überhaupt in zijn totaliteit wel gaat passen, en dan begin je meer en het aan de data knowledge extraction.
IN5: Ja, stel dat je bij stap een hebt of machine learning geschikt is. Dan zou je drie kunnen zeggen; Kijk nog steeds denkt dat het geschikt is het uitgewerkt.
***Steven**:* Ja, precies
IN5: Zou je het bijvoorbeeld ook kunnen doen; want het kan best de uitkomst van bijvoorbeeld ethiek of beschikbaarheid van data of wat dan ook; dat je in het begin nog dacht van ja kan wel en dan heb je projectgroep je samengesteld, context uitgewerkt, denk je toch niet.
***Steven**:* Toch ethisch niet verantwoord of zoiets? Dat is een mooie, concrete. Mbt het management, zei je ook dat including het management is per organisatie apart maar je wilt eigenlijk ook interne stakeholders erbij betrekken.
IN5: Ja, de juiste mensen bij hebben en dan kijken. Sommige bedrijven zijn zo georganiseerd dat management heel inhoudelijk is en erbij betrokken wil zijn. Weer andere organisaties kiezen voor het model: Wij hebben management en gaan alleen over planningen, voortgang en de algehele strategie; en het personeel zoekt de rest maar uit.
***Steven**:* Inderdaad, intern interne stakeholders samen met het management.
***Steven**:* jij zegt, ethiek hangt af van het doel; vind je daardoor dat je daar dan niet naar hoeft te kijken?
IN5: Machine learning is eigenlijk maar gewoon en analysetechniek; niet anders dan dat je ook andere manieren van rekening kunt toepassen. En er zijn organisaties die aan de lopende banden allerlei onderzoeken uitvoeren en als je daarin gewoon doet, wat je altijd doet, en er zitten geen moeilijke ethische dingen achter, dan ga je niet bij elk project opnieuw allerlei ethische vragen beantwoorden. Dus het hangt heel erg af van wat voor data heb je het over en wat doe je mee? Soms komt dat gewoon niet aan de orde. Is wel goed om eens een keertje te bedenken van is het überhaupt niet relevant.
***Steven**:* Om je een je voorbeeld re geven. Belastingdienst analyseert wel een jaren of kindertoeslag fraude gevoelig is of niet.
IN5: Dan ligt het gevoelig, want daar hebben w het over n persone. Maar op het moment dat jij analyse doet die helemaal niks met personen te maken hebben, we bijvoorbeeld je bent en machinebouwer en je gaat met een algoritme voorspellen wanneer je machines en onderhoud toe zij, dan denk ik dat je weinig ethische vragen tegen gaat gekomen.
IN5: Dus het hangt heel erg af van wat voor onderzoek je het over hebt of ethiek aan de orde is; en als je gewoon doet wat je altijd dat doet en ethiek is daarbij nooit aan de orde geweest en je nieuw onderzoek doet, dan hoeft dat op zich niet meteen aan de orde te zijn, tenzij er is iets heel anders gaan doen
***Steven**:* het model is in zekere zin wel gericht op SBR stakeholders en organisaties die daarbij horen, het gaat vaak om gegevens van personen en organisaties. Maar ik ben er mee eens op de manier die je zegt. Ik heb eventjes een klein beetje stakeholder analyse gedaan, maar dan denk je dat het voor bijvoorbeeld het project wat wij doen, dat nodig is.
IN5: Met waar we uiteindelijk op uit zijn gekomen hebben we gezegd; we willen, eerst kijken of het überhaupt voorspellende waarde heeft.
IN5: Als je experimenteel onderzoek een voorspellende waarde heeft, maar je doet verder nog helemaal niks mee, dan ben je alleen maar data intern aan het gebruiken, data die niet over personen gaan maar over organisaties en die organisaties die hebben ons de data beschikbaar gesteld om dit soort dingen te kunnen doen; niks anders dan voorheen. Dus daarmee komen niet echt nieuwe ethische vragen aan de orde. Misschien zou je kunnen zeggen, aangezien je een generiek model wil hebben, vraag je af of je een ethische analyse moet uitvoeren!
***Steven**:* wie bepaalt dat of het ethisch erbij hoort,
IN5: Dan kun je het gewoon makkelijker houden en zeggen: dit is een generiek model. Zo hebben we het gedaan
IN5: Die analyse kan soms heel kort zijn. We hebben geen probleem op dit punt, klaar.
***Steven**:* Maar het gaat er hier dat er kort aan wordt gedacht, want je kan inderdaad heel snel doorgaan en een model maken; En kijk zijn inderdaad goed in het voor 95 % in voorspellen: kom laten we het gebruiken.
***Steven**:* Eens even kijken; data verder uitsplitsen. Nou, daar hebben we natuurlijk beschreven. Op het moment dat de context erbij wordt gegeven en dan is dat. Data samenvoegen, zeg jij nog bij data preparation.
IN5: Soms kan zijn dat je allerlei verschillende bronnen bij elkaar haalt.
IN5: Als je data uit vier verschillende bronnen haalt, dan moet je op de één of andere manier op elkaar afstemmen, samenvoegen, structureren, weet ik veel wat. Dus het is een beetje de vraag: zie je dat dan als stap vier of vijf.
***Steven**:* Dat vind ik een goede dat samenvoegen: want in principe hadden wij dat ook.
IN5: De verschillende setjes, die samengevoegd moesten worden.
IN5: En omdat je vaak bij machine learning met een bepaalde input een bepaalde output wilt voorspellen, je de kans hebt dat het wel verschillende datasets zijn
***Steven**:* Je zegt dat het communiceren in de organisatie , dat dat eigenlijk een constant proces moet zijn.
IN5: Het hangt ook helemaal van je doel af. Wat je wel of niet wilt communiceren, en ik denk dat het uiteindelijk wel netjes is, dat je op de één of andere manier resultaten communiceert met je organisatie. Dat hoeft niet alleen maar ??? te zijn. Dat kunnen ook andere zijn. In ons geval betekent dat bijvoorbeeld: we waren bezig met het uitwerken van een visie op datamanagement bij onderwerp toekomst van data analyseren van de thema's was.
IN5: In dat thema is dan machine learning één van de technieken die worden benoemd.
IN5: Dus wat wij daar mee gaan doen op het moment dat we die visie afronden. Want anders dan gewoon de eerste ervaringen daarmee opschrijven. En dan is dat onze belangrijkste manier van communiceren, en iedereen die dan meer van af wil weten, die kan dat natuurlijk weten.
***Steven**:* Dat is wel leuk om te doen.
***Steven**:* Maar dat geeft ook wel het belang aan in de organisatie.
***Steven**:* De randvoorwaarden, die heb je natuurlijk nu niet gezien; Die moeten duidelijker opgeschreven worden. Van tevoren.
IN5: Het is ook maar hoe je het wilt visualiseren, kijk jij, je hebt waarschijnlijk verschillende plaatjes in verschillende overzichten.
***Steven**:* Nee, maar het kan natuurlijk ook om aan het begin een check te doen van voldoen aan de randvoorwaarden.

IN5*:* Ja, dat is misschien wel een goeie.
Steven*:* Dus dan is het ook in het totale model duidelijk, en dat niet iemand dit model pakt en dan zegt oh, ik wist niet dat er ook nog andere dingen waren.
Steven*:* Het iteratief maken van alle onderdelen, en eigenlijk beschrijven dat bij elke stap het mogelijk is om terug te gaan.
IN5*:* Ja, en dan is niet per se één stap terug, maar je kan dan terug naar alle stappen die daarvoor zijn geweest; ook om iets te kunnen aanscherpen. Overal kun je iets aanscherpen.
Steven*:* Zijn er voor de rest nog dingen die jij hebt gezien of die je anders zou willen zien. Eigenlijk hebben we de meeste nu denk ik al benoemd van tijdens het proces, maar het kan zijn dat je nog iets te binnen schiet.
IN5*:* Nee, ik denk niet. Ik kan niks anders bedenken.
Steven*:* Dat is ook goed, Top.
Steven*:* Dan wil ik jou bedanken voor het interview. Ik ga nu alles in het transcript samenvatten: de verbeterpunten verwerken in de methode en dan hebben wij sowieso nog eventjes contact over wat ik wil delen qua thesis. Ik zou wel gewoon mijn hele thesis doorsturen .

	Cijfer
De methode is effectief in het bereiken van een machine learning project	5
De methode is praktisch goed te gebruiken	5
De methode is makkelijk te begrijpen	4
Het model past bij mijn organisatie	5
De methode is compleet	3
De methode heeft het vermogen om te reageren op de schommelingen van de omgeving	3
De methode komt overeen met het experiment	4

Step:	Usefull	Understandable
1	5	5
1.1	5	5
1.2	5	5
1.3	1	1
2	5	5
2.1		
2.1.1	5	5
2.1.2	5	5
2.2	3	5
3	5	5
3.1	1	1
3.2	5	5
3.3	3	5
4	5	5
4.1	5	5
5	5	5
5.1	5	5
5.2	5	5
5.3	5	5
5.4	5	5
6	5	5
6.1	5	5
7	5	5
7.1	5	5
7.2	5	5
7.3	5	5
7.4	5	5
8	5	5
8.1	5	5
8.3	5	5
8.3	5	5
9	5	5
9.1	5	5
9.2	5	5
9.3	5	5
10	5	5
10.1	3	5
10.2	2	5

Appendix C.6: Interview IN6

Steven: Vandaag, omdat ik dit interview ga gebruiken voor mijn methode, moet ik ook allemaal officiële dingen doen, ik ga mijn scherm met jou delen eens even kijken toe, zie jij mijn scherm?
IN6: En ik zie jouw scherm moet even .. ja nu stuk beter
Steven: Oke ik stuur dit straks nog mee, maar waar we het even van tevoren behandelen. Het interview wordt opgenomen, wordt getranscribeerd, de recording wordt verwijderd, jij krijgt transcript doorgestuurd, die moet je ondertekenen, samen met dit formulier. Als je het er mee eens bent. Het formulier zegt dat ik je naam niet gebruikt, tenzij je een hele mooie quote heb en dan mag je mag zeggen of je dus inderdaad die quote met of die met echte name gebruiken of niet. En dat was het wel. Een het wordt en mijn thesis wordt op de TU Delft repository gezet.
IN6: En die study information ....was vandaag en is dit gewoon.
Steven: Dus dat stuur ik nog wel door, maar het is omdat het interview nu is en ik dacht van, omdat ik ook nog die transcribatie door moet sturen dan doe ik gewoon alles in één keer, is dat oké? Even kijken hoor. Ik heb ook een schema voor mezelf gemaakt even bij pakken, dan kan u het interview beginnen.
Steven: Mijn eerste vragen gaan over het experiment en dit zijn gewoon open vragen. Laten we beginnen. Denk je dat het handig was om dit machine learning experiment uit te voeren en waarom.
IN6: Ja
IN6: Het is wel iets wat bij meerdere mensen binnen het team al een beetje sudderde van, is. dit iets waar mee iets mee kunnen doen Zo ja, en dan hoe. En het is dan best wel fijn om dan iets te doen.
Steven: Iets te doen.
IN6: En het is dan best wel fijn dat er iemand is die er iets van af weet, tenminste meer dan wij, als volledige leken en die je kan zeggen: nou, ja, dit, het programma kun je gebruiken dit is de manier waarop je de data kan bewerken, dus het is wel fijn om een beetje bij de hand genomen te worden . Wij gingen er natuurlijk niet in wij hebben het plan om met **Steven** dit probleem op te lossen, we denken dat het al best goed werkt maar we zijn nieuwsgierig.
Steven: Ja, ja, precies we hebben met z'n drieën al best wat geprobeerd en hebben we ook best nog wel wat leuke resultaten gekregen.
Steven: Oké, kijken, dat is positief.
Steven: Denk je dat machine learning de huidige situatie kan verbeteren in jouw organisatie.
IN6: Ja, het kan allicht geen kwaad, tenminste ik heb nog niet zoiets van wat wij hebben gedaan wordt kan een concreet bijdragen aan XY . En het is wel zo dat we nu kunnen zeggen, nou ja, het is wel iets waar we naar kunnen want het is om te kijken.
Steven: ja ja
IN6: Je kunt nog een extra laag om maar in de zekerheden inbouwen om een andere manier naar de data te kijken.
Steven: En we hebben natuurlijk ook inderdaad in snelle in snelle vorm even bekeken welke punten het algoritme belangrijk vond om het voor het voorspellen van de resultaten.
Steven: Dus als laatste ben je tevreden met de resultaten van het Machine Learning experiment.
IN6: Ja, nou ja, op zich wel is, ik denk niet dat we iets bruikbaars hebben, wat ik je al zei, dat we op zoek waren naar .
Steven: Nee, nee.
IN6: echt een oplossing of een antwoord van iets, 't was meer we wilden kijken hoe het werkte.
Steven: Ja, en dat maar dat is inderdaad ook qua doel, het was, ik pretendeer ook niet dat ik dat mijn methode ervoor zorgt dat je een volledig werkend machine learning algoritme hebt, maar wel die eerste stappen waardoor je weet of het überhaupt wel nuttig zou zijn of niet.
IN6: nee want ook wat we de laatste keer dat wij zaten bespraken, voordat we zoiets zouden implementeren of het verder kunnen gebruiken, dat we zoiets zouden laten toetsen door iemand die echt meer van statistiek afweet dan jij en ik.
Steven: Maar dat is op zich een mooie conclusie
IN6: nog niks geprobeerd maar dit is onzin of iemand die zegt ja ja, ja,
IN6: maar ik kan het oordeel nog niet maken.
Steven: Nee, maar de volgende keer weer inderdaad als je weer een dataset krijgt, dan weet je nu ook welke stappen je moet doorlopen, om eventjes in ieder geval te kijken. Nou, dat waren eigenlijk de vragen over het experiment. Daarnaast heb ik een methode ontwikkeld. Dit is een versimpelde versie ervan. Zie je hem nu in beeld, oké, ja, De e verdere loop van het interview is in tweevoud.
Steven: Allereerst gaan we door de methode lopen en dan wil ik van jou weten, per stap of je het nuttig en begrijpelijk vindt, en daar moeten we denken is.
IN6: Stap zelf.
Steven: Ja, en ook de kleine stapjes, dus daar moeten we doorheen en daarna komen de vragen als: wat zou je willen verbeteren? Zou je willen toevoegen? Zou je iets willen aanpassen? Et cetera, ET cetera? Dus het is een eerste stuk is kwantitatief en dan daarna een stukje open vragen.
Steven: Nou, eerst naar de methode. We hebben tien stappen. Elke stap heeft een x aantal acties die gedaan moeten worden. Sommige acties zijn wel recht toe recht aan, het idee van de methode is dus dat je in eerste instantie een machine learning experiment op kan zetten, maar wel rekening houden met een stukje organisatie en een stukje ethiek. Nu gaan we dan kijken naar de hele methode of er dingen zijn die jij erbij zou willen zien, of dingen zijn die je anders zou willen zien. En ik heb even kijken, twee dingen opgeschreven, dus de machine learning experience en zou je een methode experience willen hebben?
IN6: Ja, gewoon misschien voor de personalized statistical method experience ik denk wel dat het belangrijk is. Ja, het kan, dat kan wel een hele grote overlap hebben met iemand met machine learning experience, ja maar ik vind het wel een aparte categorie.
Steven: Ja, oké.
IN6: Ja want de manier waarop je data is gestructureerd zou wel leidend moeten zijn met welke methode gebruikt.
Steven: Oké, en zeg maar op het moment dat jij dat is ook wat andere mensen zeiden dat het iets duidelijker, ook een verschil moet gemaakt moet worden met de domein kennis. En hoe zou jij jezelf zien als iemand die deze statistiek kennis al heeft?
IN6: Ik zou zeggen dat ik het ooit heb gehad, ik heb het nu niet zo paraat.

IN6: Ik heb het niet zo dat ik het nu niet kan zeggen: nee, je moet dit wel op die manier bootstrappen. Het is meer een beetje aan is volgens mij moeten iets doen. Dus ik nee, ik zou het wel kunnen worden. Dat ik heb dit ook met mijn management besproken dat het misschien wel een beetje op ga pakken en gaan leren, maar het is wel gewoon. Ik ga niet zeggen dat ik dit nu paraat als kennis heb.
Steven: Laten we zeggen dat het misschien ook afhankelijk is hoeveel statistiek kennis de machine learning expert heeft. En anders moet je die kennis aanvullen? Dan is het prima?
IN6: Ja, deze drie categorieën kunnen allemaal in principe dezelfde persoon zijn, ja, maar het is wel. Kijk, ik ken ook gewoon mensen die wel die werken bij de universiteit hele leven werken met statistieken die zoeken alsnog een message expert erbij, kan ik dit nou gebruiken of moet misschien net een andere manier, gewoon iemand die er veel meer van afweet.
Steven: En zou, zou je deze stap dan bij die eigenlijk ook bij die, aan het eind dat bij stap negen, dat je toch die is statistical method experience ook controleert?
IN6: ja
Steven: want wij hebben in principe ook een een klein experiment opgezet, zonder dan echt een echte expertise te hebben qua statistiek.
IN6: Ja
Steven: Dus dan zou het.
IN6: Ik denk inderdaad dat het wel voordat je iets natuurlijk zijn voordat je iets implementeert voordat je het gaat publiceren of gebruiken in verder analyse is. Moet je wel een check dat wat je hebt gedaan ook juist ja.
Steven: Oké, dan ik denk dat ik dit er dan bij stap negen bij zet, want dan als ik deze er bij stap twee erbij zet, dat is denk ik niet een verplichting om alles door te lopen en die is de eerste experiment op te zetten. Echter om door te gaan met het project is dat, dan is het wel nodig.
IN6: In ben ik niet helemaal met je eens. Ik denk dat je heel erg de mist in kan gaan door de data, op een bepaalde manier voor te bereiden die uiteindelijk niet handig blijkt omdat je het toch op een andere manier moet doet. En als je in het begin duidelijk had gehad welke methode je had moeten gebruiken had je je veel koppijn bespaard.
Steven: Dan zet ik 'm bij stap 2
IN6: ik zou zeggen Stap twee en stap 9. Je moet nooit gewoon één keer een beslissing maken en dan denken dat het voor eeuwig goed is.
Steven: Ik denk ik ga ze gewoon loskoppelen : machine learning experience, data experience, domein experience en dan zeg ik ook een stukje: statistical experience. Oké, dat is prima, ehm input, impact analyses en dan inderdaad op data gebruik. Dus je wil, of je wel de data mag gebruiken.
IN6: Nou, ja, GPR zorgt ervoor dat je heel erg, je moet echt goed documenteren voordat je zoiets doet dat je de toestemming hebt als dataprocessor om de data te gebruiken voor het doel waarvoor je het nu voor gebruikt. Dus het is echt een stap die zeker bedrijven die persoonsgegevens hebben, die weg te lopen.
Steven: Oké.
IN6: Of academici, die persoonsgegevens gebruiken
Steven: Nee, maar daarom, ethisch moet het ook handvaten gaan geven. Dus oké, is even kijken en voor de rest nog dingen die je anders zou willen zien.
IN6: nee niet zo 123.
Steven: Oké, dat is eigenlijk goed, beter weinig opmerkingen, weinig veranderingen, extreem veel. Nee, ik heb nu op zich best een mooi lijst van dingen die inderdaad nog met anders opgeschreven moeten worden of net toegevoegd moeten worden. Wat duidelijker beschreven. Dus nou dat als jij verder niets ziet, dan zijn we denk ik klaar qua interview.
IN6: Oke .
Steven: Heel erg bedankt vond je het leuk om hiermee bezig te zijn om deze.

	Cijfer
De methode is effectief in het bereiken van een machine learning project	4
De methode is praktisch goed te gebruiken	3
De methode is makkelijk te begrijpen	3
Het model past bij mijn organisatie	3
De methode is compleet	3
De methode heeft het vermogen om te reageren op de schommelingen van de omgeving	4
De methode komt overeen met het experiment	3

Step:	Usefull	Understandable
1	5	5
1.1	5	5
1.2	4	4
1.3	4	5
2	4	5
2.1		
2.1.1	4	4
2.1.2	4	4
2.2	3	4
3	4	4
3.1	4	4
3.2	4	4
3.3	4	5
4	5	5
4.1	5	5
5	5	5
5.1	5	5
5.2	5	5
5.3	5	5
5.4	4	4
6	5	5
6.1	5	5
7	5	5
7.1	5	5
7.2	4	4
7.3	4	4
7.4	4	4
8	5	5
8.1	5	5
8.3	4	4
8.3	4	4
9	5	5
9.1	5	5
9.2	4	5
9.3	4	4
10	4	4
10.1	4	4
10.2	1	1

Appendix C.7

			Usefull					Understandable		
Step:	Sterk mee oneens	Oneens	neutraal	Eens	Sterk mee eens	Sterk mee oneens	Oneens	neutraal	Eens	Sterk mee eens
1										
1.1										
1.2										
1.3										
2										
2.1										
2.1.1										
2.1.2										
2.2										
3										
3.1										
3.2										
3.3										
4										
4.1										
5										
5.1										
5.2										
6										
6.1										
7										
7.1										
7.2										
7.3										
7.4										
8										
8.1										
8.3										
8.3										
9										
9.1										
9.2										
9.3										
10										
10.1										
10.2										

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
The method is effective in achieving a machine learning project					
The method is practical to use					
The method is easy to understand					
The model suits my organization					
The method is complete					
The method has the ability to respond to environmental fluctuations					
The method is consistent with the experiment					

Appendix C.8: Consent Form for Master thesis Steven Hoozemans

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information dated [DD/MM/YYYY], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	◻	◻
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	◻	◻
I understand that taking part in the study involves an audio recorded interview which will be transcribed and the recording will be destroyed.	◻	◻
Use of the information in the study		
I understand that information I provide will be used to evaluate the designed method and I understand that the interview results will be processed according scientific qualitative data processing methodologies.	◻	◻
I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will not be shared beyond the study team.	◻	◻
I agree that my information can be quoted in research outputs	◻	◻
I agree that my real name can be used for quotes	◻	◻
Future use and reuse of the information by others		
I give permission that the summary of the interview (not including: the interview transcript and corresponding audio files) that will be based on the interview that I provided to be archived in *TU Delft thesis repository*2 so it can be used for future research and learning.	◻	◻

Signatures

__________________________ __________________________ __________

Name of participant [printed] Signature Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

___________________________ ____________________ _________

Researcher name [printed] Signature Date

10.6 Appendix D

10.6.1 Appendix D.1