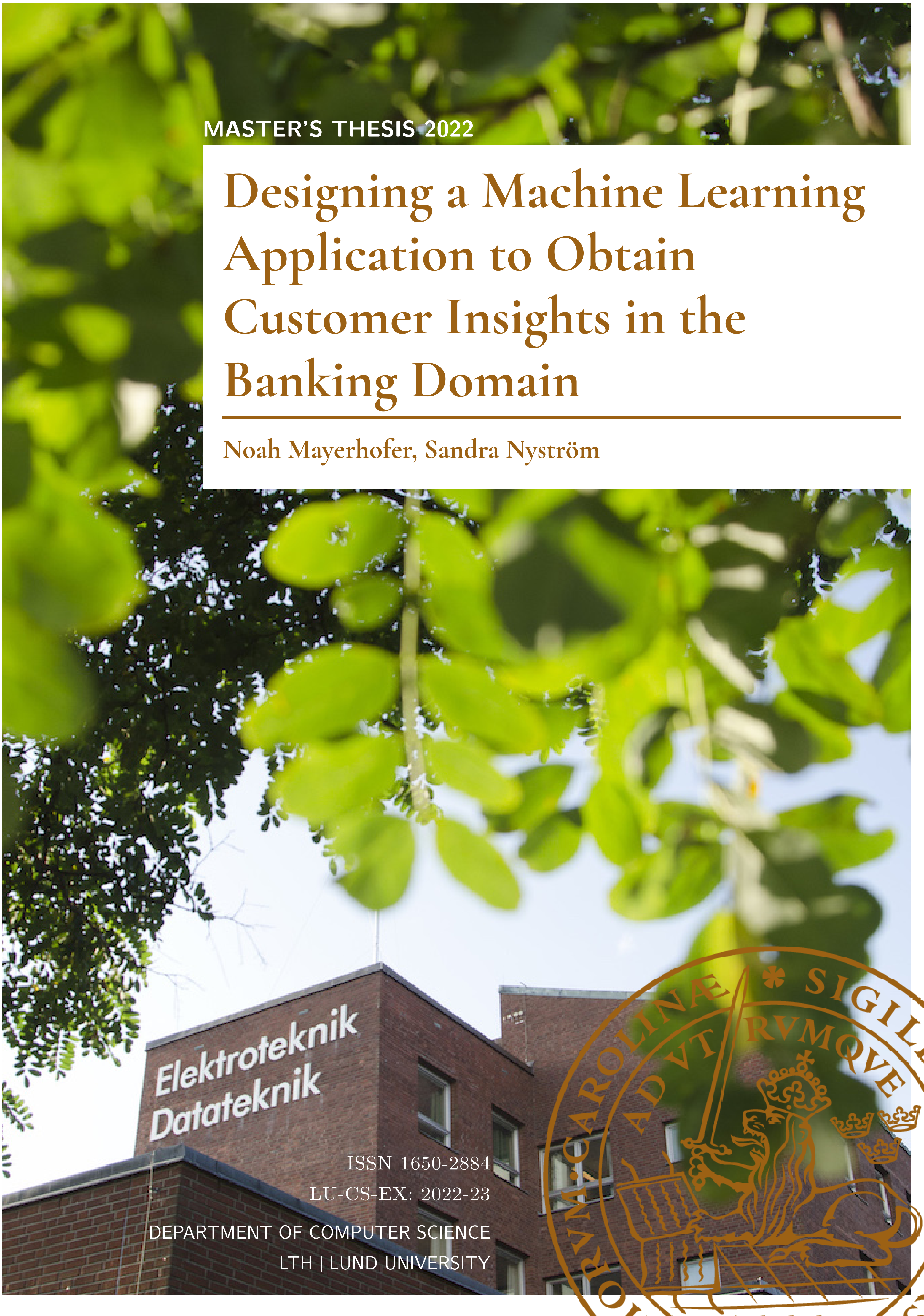
MASTER'S THESIS 2022

Designing a Machine Learning Application to Obtain Customer Insights in the Banking Domain

Noah Mayerhofer, Sandra Nyström



ISSN 1650-2884
LU-CS-EX: 2022-23

DEPARTMENT OF COMPUTER SCIENCE
LTH | LUND UNIVERSITY

EXAMENSARBETE
Datavetenskap

LU-CS-EX: 2022-23

Designing a Machine Learning Application to Obtain Customer Insights in the Banking Domain

Design av en maskininlärningsapplikation för att erhålla kundförståelse inom banksektorn

Noah Mayerhofer, Sandra Nyström

Designing a Machine Learning Application to Obtain Customer Insights in the Banking Domain

Noah Mayerhofer
no1774ma-s@student.lu.se

Sandra Nyström
sa1741ny-s@student.lu.se

June 7, 2022

Master's thesis work carried out at
the Department of Computer Science, Lund University.

Supervisor: Elizabeth Bjarnason, elizabeth.bjarnason@cs.lth.se

Examiner: Emelie Engström, emelie.engstrom@cs.lth.se

Abstract

A good customer understanding is an important foundation for successful businesses that enables providing of products and services in-line with customer needs. Information intense businesses, such as banks, have access to large amounts of user data that may be used to support customer relationship management through mining this data for insights into the customers' needs. This thesis aimed to investigate how user data can be explored to improve customer advising and possible challenges of implementing a machine learning based application within the finance sector. We used design science research at a case company in the financial domain to design and develop a machine learning based application. We also followed the CRISP-DM, the industry standard for data mining projects, to concretise the design science steps. We collected data by two sets of interviews and two iterations of focus groups which had three participants.

The final machine learning model performed better than the benchmark at predicting the next 26 weeks of seasonal patterns. When demonstrating the generated information, the users did receive new customer insights. We identified a number of challenges when implementing the ML-based application, such as long running times for almost every step, due to our large data set, unknown performance of model in deployment, and verbal misunderstandings related to tacit knowledge. This thesis contributes with an example of how CRISP-DM can be used in the finance sector and how machine learning can be used to create insights. However, further work is needed to improve the application as well as test the process for other banks.

Keywords: Machine Learning, CRISP-DM, Finance, CRM

Acknowledgements

We would like to thank the case company for working with us on this Master's thesis. The company has given us the opportunity to work with real customer data and a challenging problem. It has been a pleasure to pick your brains!

We would also like to thank our supervisor Elizabeth Bjarnason for helping us every step of the way. She has given us great feedback and ideas along the way and her dedication to helping us with the structure of the report has been invaluable.

Contents

Chapter 1
Introduction

It is more important than ever for companies, in all sectors, to keep up with the ever-changing markets and to utilise all the technological tools there are to keep themselves relevant to their customers and banks are not an exception. The finance sector is information-intensive and information technology can be seen as a business enabler, analysing data for better customer relationships is key [1]. Understanding your customers' data will give you insights into how to run your business better and this can be done by data mining. Data mining, the science of extracting useful knowledge from large data sets, is an interdisciplinary field in computer science and includes various techniques, such as statistics and machine learning [2]. Machine learning (ML) is a branch in the field of artificial intelligence (AI) and in finance, the use of AI has increased in recent years in tandem with the increase of machine-readable data [3].

The process of going from large volumes of data to having something that generates true value for a business is not a straightforward path. There is a great need for business and data understanding to be able to create an artefact that suits that particular domain and problem. In this thesis, we explored this process in the banking domain at a case company, more specifically in the banking area of foreign exchange (FX). We wanted to investigate the creation of an ML-based application to find customer insights by using user data. For this, we used data of transfers between customers' accounts of different currencies where conversions happen passively in the background without the user requesting a conversion, henceforth called passive FX transactions.

The thesis contributes with an example of how to conduct such a process and some of the challenges that one might face. It specifically addresses this in the banking domain, but we hope that some of our findings will apply to others as well.

This thesis starts by giving the reader an introduction to the topic and aims. The first chapter, Chapter 1, will also explain some necessary background knowledge, the general approach we have followed throughout, a description of our case as well as a short statement of how we have divided the work between us. The following chapter, Chapter 3, describes how we executed the different methods in detail. Subsequently, Chapter 4 presents the results, followed by a chapter of discussion, Chapter 5, putting the results in context and exploring what conclusions can be drawn. Finally, the thesis is concluded in Chapter 6 with a summary of the thesis and the conclusions of the work as well as what future work in this area should be done.

1.1 Aims and Research Questions

Customers doing business with foreign companies often have to exchange currencies. Since the FX market is always changing there is a risk, which should be managed by formulating a strategy. One way our case company works with discovering customers in need of guidance is by finding, so-called *leads* by analysing data. A *lead* is contact information and other relevant information about a customer that might be interested in a product or service. The *lead* can thereafter be used to contact the specific customer to offer them guidance of how to better manage their FX trading or to buy a product or service of use for that customer.

This Master's thesis aims to investigate how user data can be used to improve customer understanding. A bank may be able to provide better service to its corporate customers based on knowledge acquired through their user data. We have investigated this topic by creating an ML-based application generating leads using passive FX transaction data at a case company.

The specific questions we want to answer through this thesis are the following:

RQ1: What can a bank learn about its small and medium sized business customers from their passive FX transactions data?

RQ2: How can machine learning be used to gain knowledge about passive FX transactions?

RQ3: What are the main challenges when developing an ML-based application for the banking domain and how can these be handled?
RQ3a: related to data?
RQ3b: related to achieving problem-solution fit?

In the banking domain, there are very large volumes of customer data and it is, therefore, a suitable domain for data mining, to gather information by different techniques for large data sets. We intend to explore this on different levels of abstraction. RQ1 is quite general and broad, higher level, and explores what possibilities there are for our specific data. This is required to know what objectives the application has to fulfil for the bank to get insights. Whereas RQ2, on the other hand, is more specific, lower level, and addresses the use of machine learning in our application. When designing and developing an application of any sort there will be challenges along the way. But, to be able to avoid some of them in future work, we compiled the challenges we faced and how we handled them in RQ3.

1.2 General Approach

This thesis investigates the RQs in Section 1.1 by designing and creating an ML-based application. The investigation was performed at a case company, creating a concrete example of how an ML-based application can help a bank with its customer guidance.

We addressed RQ1 by exploring the problem at our case company and designing and evaluating a solution for and in that context. RQ2 was answered by investigating the customers' passive FX transaction data and building a machine learning model based on that data. RQ3 was addressed by creating an ML-based application in a new domain and noting challenges along the design and development process.

The investigation is based on the Design Science Research Methodology [4], we also incorporated the cross-industry standard process for data mining (CRISP-DM)[5] into the design science steps. The purpose of this approach is to look at the whole process of designing a machine learning-based application. The main data gathering methods used were literature review, interviews, focus groups and document studies [6]. A literature review was used to gain domain understanding where term as finance, banking, forex, and foreign exchange as well as machine learning and time series analysis was used in different combinations. Two sets of interviews were performed. The first set consisted of two group interviews and the second set of three individual interviews. Two sets of focus groups with 3 participants each were used to evaluate the final application. Document studies were used throughout the project to review current manuals and other documents provided by the case company to work with their systems.

1.3 Case Description

We conducted this thesis at a Nordic bank at their headquarters. The bank provides a wide range of economic services, everything from loan offers to insurances, to all their personal and business customers.

At the case company, there is an FX department where both developers and sales teams work to help the bank's customers by either developing and upgrading digital tools or contacting the customers to give them individual advice. The sales team we worked with for this thesis, henceforth called the sales team, contacts and guides customers about their FX strategy.

Due to a large number of customers, the sales team needs to know which customers are in the most need of guidance. Therefore, they work closely with the FX data analytics team, which helps the sales team by creating recommendations for leads.

The sales team had expressed an interest in working more closely with leads in general and specifically with leads based on the passive FX transaction data since that data had not been investigated previously. We, therefore, created an application that generates leads using the passive FX transaction data and machine learning. The stakeholders of this application are mainly:

- The sales team members (consisting of currency advisors) since they are users of the application.
- The FX data analytics team members, which included a team lead and future product owner of the created application, and developers because they will continue the development and maintenance of the application after this thesis.
- The corporate customers of the bank will be influenced by the recommendation of the currency advisors. We have not worked directly with this stakeholder, but it is still an important stakeholder to keep in mind.

1.4 Division of Work

Both authors participated in all parts of the thesis and most work was conducted as a pair. However, Sandra Nyström focused more on the implementation of the machine learning models and Noah Mayerhofer conducted the individual interviews (I2).

The writing of the report was done in parallel, but both authors proof read every part of the report and have contributed more or less to all chapters. However, a few sections had more focus from one author than the other. Sandra Nyström focused more on Aims and Reasearch Questions (Section 1.1), CRISP-DM (Section 2.1), and Define Objectives of a Solution (Section 4.2). Noah Mayerhofer focused more on the Case Description (Section 1.3), Design Decisions (Section 3.3.4), and Identify Problem and Motivate (Section 4.1).

Chapter 2
Background and Related Work

We have, for this thesis, used previous research from different fields to explore methods, solutions and ideas from similar research and to ensure our research questions have not yet been answered. The CRISP-DM methodology has been applied throughout the thesis and is therefore shortly described and put in to context in the beginning of the chapter. There are two main areas, see below, related to our thesis where machine learning has been applied, namely within Finance and within Customer Relationship Management (CRM). Machine learning is already used within the finance sector, and we have identified work on similar data and tasks as in our case. Machine learning has also been used within CRM to improve the relationship between business and customers. Finally, we have included an introduction to the chosen machine learning architectures, LSTM and N-BEATS, to provide non-ML experts with a basic understanding.

2.1 CRISP-DM

CRISP-DM, or the Cross-Industry Standard Process for Data Mining, is an industry-, tool- and application neutral methodology for data mining projects [5], used as a standard in the industry. It has been developed by a consortium with experience from "real" data mining projects and practical experiments and with funding from the European Commission [5]. Since CRISP-DM 1.0 Step by step data mining guide was released in 2000, many projects have been adopting the methodology. Meta S. Brown, a data miner and tech journalist, stated in Forbes in 2015 that it was "...by far the most widely-used analytics process standard" [7]. CRISP-DM is also widely used in research and academia, when using the search term CRISP-DM, in the database LUBsearch, 866 peer-reviewed results were obtained.

CRISP-DM is divided into five or six phases [5], business understanding, data understanding, data preparation, modelling, evaluation and sometimes deployment. CRISP-DM is described as a hierarchical process model, making it easy to go in the range of general to specific. For every phase, there are three more levels of abstraction: generic tasks, specialised tasks and process instances. For example, in the data preparation phase, there can be a generic task of cleaning data that has specialized tasks of how to clean it depending on what kind of data it is (numerical/categorical). The action, decision or result of a specialized task is recorded and this record is the process instance.

CRISP-DM also encourages the user of the methodology to change it and adapt it to the specific problem, pressing that going back and forth between the phases and tasks or skipping some tasks is a natural part of the process. We have, for example, been going back and forth between data preparation and modelling to filter our data further for the training of the models to take a feasible amount of time. Another thesis [8] has looked at this process of adapting and expanding the methodology in the domain of data-driven financial support tools by interviewing different stakeholders.

2.2 Machine Learning in Finance

A literature study from 2021 reviewed 24 studies using CRISP-DM as their methodology. Only one of the studies reviewed was from the finance sector, showing a need for further research in this area [9]. The same study concluded that CRISP-DM is used in different ways in different domains but some common denominators exist, for example, it is common to measure the performance of the model with some metrics. It is also common to not perform the deployment step because it will be done later or since the results of the model were not good enough to justify deployment [9].

In finance, machine learning is widespread which several literature studies have shown [10, 11]. The usage of machine learning is often towards prediction [11] but it has also been used for portfolio construction and fraud detection [10]. Common challenges when using machine learning is to find clean data and filter out data which would make the model worse [11]. We also found that pattern prediction is being used in finance and even more specifically in the field of foreign exchange [12], but for pricing data instead of transaction data as we have. In the study, they are using a genetic algorithm method instead of machine learning, but the paper still shows that working with patterns in time series data can generate beneficial results.

Since CRISP-DM has not been used frequently in finance but sees widespread use overall, we, therefore, thought it would be interesting to use CRISP-DM to develop an ML-based application and see what challenges we encounter. It has been shown, in the literature, that machine learning in the field of finance and pattern prediction has been successful and we, therefore, saw an opportunity for a good result on our data as well.

2.3 Machine Learning for CRM

For a business to grow it is important to have a good relationship with customers, both current and future, and with the exponential growth of data, more insights have become possible, Customer Relationship Management (CRM) being a widely used tool [13]. Data analysis on customer data both strengthens the customer relationship performance and increases the sales growth [14]. A literature study from 2020 shows that the use of machine learning within CRM is increasing and is already used in many sectors, including banking [15]. The study reviewed 35 papers presenting a variety of machine learning algorithms applied to the CRM domain and found that about half of the papers used supervised learning techniques and just over 15% used unsupervised. An increase in deep learning techniques could also be seen, indicating a shift in the field [15]. These studies show that CRM is important for businesses to strengthen their relations with customers and that machine learning can be used successfully to improve the relationship between businesses and customers.

2.4 Machine Learning Models

We compared two machine learning models, namely LSTM and N-BEATS, which both are deep learning models that can be used for time series forecasting. Deep learning is a subcategory of machine learning, which just means that the machine learning models are very large (deep) and usually results in better performance [16]. The idea of the LSTM network was published in 1997, whereas N-BEATS is much younger with three years since publication. In this section, we will provide some background on these models.

2.4.1 LSTM (Long Short-Term Memory)

Long Short-Term Memory, or LSTM for short, is part of the recurrent neural network, RNN, family [17]. An RNN is a type of deep learning neural network which is specialised to work with sequence data such as time series [16]. A common problem for deep learning methods when working with time series is that data far back in time either diminishes to something insignificant or blows up ruining the model's results [16]. The LSTM model was created to solve this problem [16] and because of this, LSTMs have shown good performance on time series forecasting [18].

2.4.2 N-BEATS

Neural Basis Expansion Analysis for Interpretable Time Series forecasting or N-BEATS for short is a fairly new deep learning-based forecasting method. It was first presented in 2020 as an answer to modern forecasting models which had become more hybrid models between deep learning and statistics. With the creation of N-BEATS, the goal was to go back to a new pure machine learning model and so a new architecture was created. The architecture is built on a simple generic building block but the models become extensive by using many more of these blocks than other machine learning models [19]. The new architecture has shown to be useful in several domains without being changed while outperforming other machine learning models [19, 20].

Chapter 3
Method

We applied the steps of design science [4], i.e. problem identification, design, demonstration, and evaluation, and detailed these using the more concrete steps of CRISP-DM. An overview of our research process is provided in Figure 3.1. First, we identified and motivated the problem by gaining insights into the business domain and related data. This knowledge was used to define objectives that the solution to be constructed had to fulfil. For the solution we compared two different machine learning models, LSTM and N-BEATS, see Section 2.4 for details. We constructed an artefact, in the form of an application displaying interesting customers, from the best model. We demonstrated the artefact to the stakeholders and evaluated how well it satisfied the business requirements. During the process, we performed two iterations of the design, demonstration and evaluation steps to improve upon the initial artefact. After the design and development process, we performed an analysis of the challenges we faced throughout the thesis.

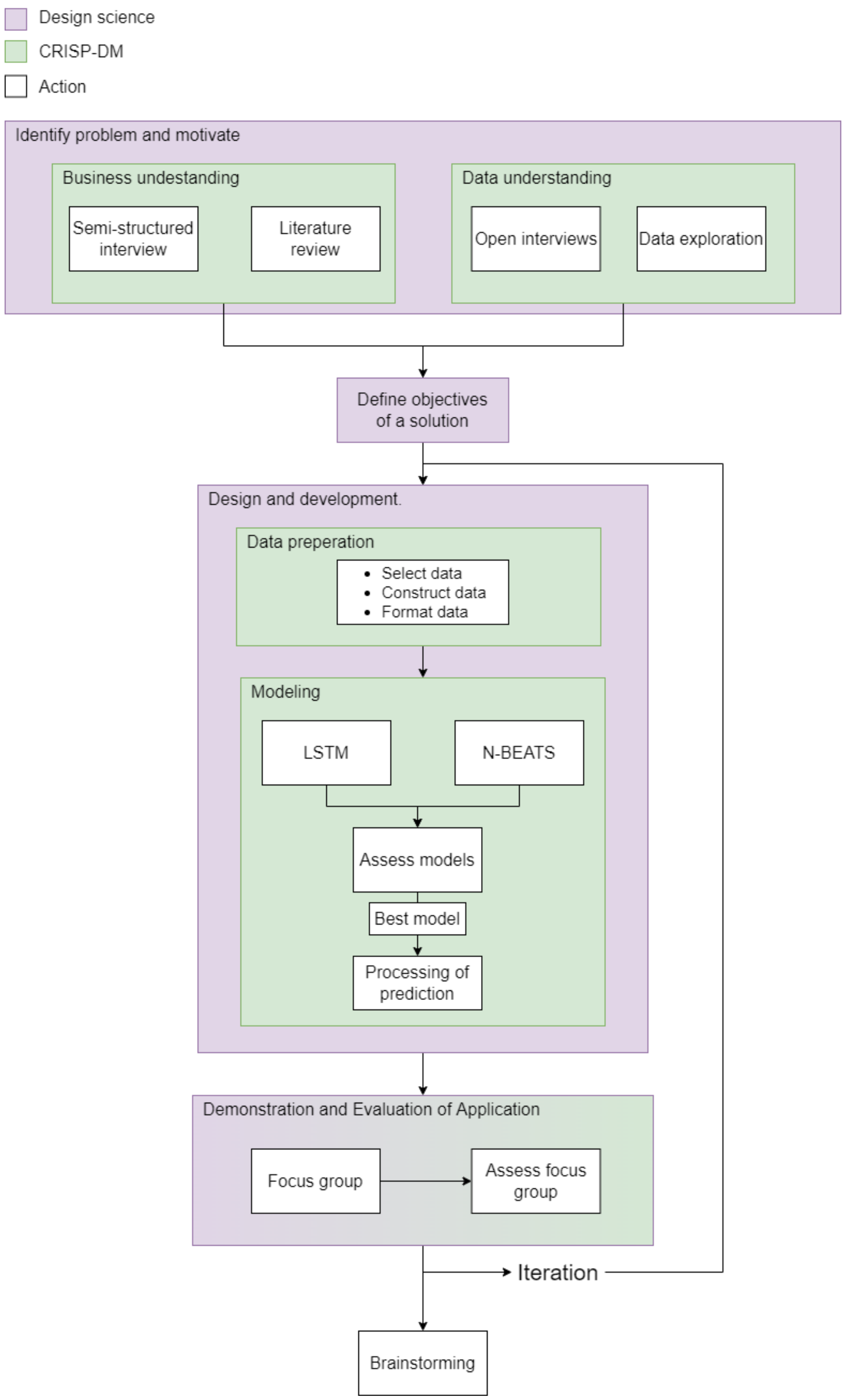


Figure 3.1: Overview of the thesis process

3.1 Identify Problem and Motivate

In the initial step of our process, we required an understanding of the business domain and of the data on which our ML-based application was to be based. We used this knowledge to define the objectives of our solution, which were then used to guide the solution design.

3.1.1 Business Understanding

To define relevant objectives of a solution and design an artefact meeting those objectives, we needed to understand the business domain, in general, and specifically for our case company. For this reason, we performed a literature review of the use of machine learning in the financial and banking domain. We also performed interviews with stakeholders and subsequent thematic analysis to gain insight into the case company.

Literature Review

We performed a literature review to increase our understanding of the banking domain and how machine learning is being used in finance. We thought that literature could explain concepts and common problems people working within the business consider obvious, since they work within the domain daily, as well as give us deeper technical knowledge in the never-ending research of machine learning. The literature review, even though most of it was performed in the initial stage of the thesis, was carried out continuously throughout as support for our analysis and to be able to motivate decisions with research.

For the gathering of literature, we have mainly used LUBsearch and Google Scholar, with some additional help from our supervisor and Google for finding websites of different Python libraries and machine learning blogs. We found a huge amount of related literature and could only considered a small part of it. We ended the literature review when we felt we had enough understanding. We have also sometimes followed the trail of references used in other papers and articles to find the original source.

To position ourselves and deepen our knowledge in the domain of machine learning in finance, search terms as **machine learning** and **time series analysis** in combinations with **finance**, **banking**, **forex**, and **foreign exchange** were used. In this phase, methodologies and going from idea to application e.g. CRISP-DM, design science in information technology/software engineering and prototyping, were also reviewed. Thereafter, we read about different interviewing methods and analyses for interviews, mainly in a book [21] about this topic referenced by the master thesis guidebook [22].

Later on, we made further research on different machine learning methods for time series analysis and on methods for data pre-processig. We re-used some of our previous search terms, such as **time series analysis**, **machine learning**, and sometimes **finance**, together with **pattern recognition**, **pattern anomaly**, **pattern detection**, **pre-processing** and **filtering**. As we gained more knowledge the search got more precise and we added terms, such as **LSTM**, Long Short Term Memory, and **multiple univariate time series**.

Semi-Structured Interviews

We conducted two sets of interviews consisting of two group interviews (I1) and three individual interviews (I2). The goal of the group interviews (I1) was to gain business understanding to define the objectives of a solution, the individual interviews (I2) were performed to obtain more concrete suggestions and wishes of what information the final application should display.

Group Interviews (I1)
The purpose of I1 was to get an understanding of our case company's business. The questions were constructed after a brainstorming session based on what information we did not get from the literature review: what problems they have, what is needed to solve the problem, and how they work at the bank. The questions were also revised twice with the help of our supervisor. These questions can be found in Appendix A.

During the interview, we used the questions as a checklist to make sure all of our questions got answered. The group interview was a one-hour video meeting which was recorded. We selected participants from all stakeholder groups to ensure all important opinions and insights were raised, the participants can be found in Table 3.1. The currency advisors listed were the only people in the sales team that could participate at the time. The senior consultant, data scientist, and team lead are the only people with the required technical knowledge at the case company.

After the first interview, a follow-up interview was performed with the same participants to clarify some answers realised after the analysis of the first interview. The follow-up interview was a recorded video interview and about half an hour in length.

Table 3.1: Participant of interviews. Column I1 and I2 denotes which interview set each person participated in

Current role at case company	Time worked at case company	Time worked within the financial domain	I1	I2
Currency advisor 1	Half a year	Half a year	X	X
Currency advisor 2	4 years	4 years	X	X
Currency advisor 3	8 years	10 years		X
Senior currency advisor	5 years	8 years	X	
Senior consultant (developer)	3 years	3 years	X	
Data scientist (developer)	Half a year	Half a year	X	
Team lead	2 years	14 years	X	

Individual Interviews (I2)
After the interviews in I1, we had some further questions about how the sales team wanted the application to be presented and what information they need in the application. Therefore, we performed an additional set of interviews with only the sales team, the interviews were individual since we did not want to let the interviewees colour each other's opinions [23].

The interviews were semi-structured and the questions were used as a checklist, just as in I1. The questions focused on how the interviewees work with leads today and how they would like our application to present leads. The questions in full can be found in Appendix B.

The interviews were performed via video meetings with all three participants, the participants were selected since they were the people that could participate at the time, some information about the participants is provided in Table 3.1. All interviews were about 20 minutes in length.

Thematic Analysis of Interviews

The interviews were analysed to locate answers to our questions and gain more insights into our subsequent design. For this, the widely used and recognised method *Thematic Analysis* was used [24]. The method suited our needs since we have qualitative data that we wanted to concretise.

Both I1 and I2 were analysed in the same way: The recordings of the interviews were transcribed and themes were created using thematic coding. The process followed Braun and Clarke's six-step guide [25]: Become familiar with the data, Generate initial codes, Search for themes, Review themes, Define themes, and Write-up. For I1, we focused on identifying themes for the creation of a problem definition. For I2, we focused on themes related to the sales team's requirements for the solution.

We familiarised ourselves with the data by one of us re-listening to the recordings and loosely transcribing them. By loosely transcribing, irrelevant parts were skipped when the conversation drifted from the scope of this thesis, and words with no meaning or rephrased sentences were left out. Afterwards, the other person read the transcript while re-listening and made notes when disagreeing with the text. Finally, the differences were discussed and a final transcript was agreed upon.

The initial codes were generated by both of us highlighting important parts of sentences, we then looked through all the highlighted parts of the text and grouped all highlighted parts with other similar parts. From the groupings, we searched for themes by naming all the groups to something that describes the underlying meaning of the contents of the group. We then reviewed and discussed the grouping and themes to see if we needed to add a subtheme, if a piece of text could fit into multiple themes or if we should move one piece of text. We updated the grouping accordingly and then produced a table with the themes and the highlighted pieces of text. The final themes can be found in Chapter 4 in Table 4.1 and 4.2.

3.1.2 Customer Data Understanding

We familiarized ourselves with the user data available at the case company as part of understanding the business domain and in preparation for our design. The data consisted of a wide range of information, in several different databases with many different tables, regarding passive FX transactions, customer information and currency rates from the beginning of 2018 until early 2022 when the data understanding was conducted.

We started by looking at the data we wanted to predict, this was the passive FX transactions. The first part of the data understanding was to decode the different fields in the table to understand what the different abbreviations meant. The decoding was done with the help of employees of the case company that worked with its databases daily. The data consisted of a few million transactions per year and consisted of all passive FX transactions of the case company's corporate customers.

After we understood each field, we plotted the passive FX transactions for a few randomly selected customers for the year 2020 to get a general feel for the data. We also investigated the seasonality of the data using an autocorrelation function [26] for a few different seasonality lengths.

The customer information data and the currency rates data were in the same database. The database had a look-up system for the fields and tables within it, making it easier to decode the fields of the tables ourselves. This, for example, made it clear how we should convert a currency into another one and that the case company has divided their corporate customers into small, medium, large and institutional, making it easy for us to later filter on this type.

3.2 Define Objectives of a Solution

After both the business and data understanding phases were conducted we developed and defined the objectives of the solution. The definition was done by brainstorming ideas, together with the Team Lead, that we thought would both help the sales team and would be possible with our prerequisites: available data, time constraints, our skill sets and hardware constraints. The purpose of defining such objectives was to provide a direction for our solution design. We got inspiration for the layout of the objectives definition from an example in a study of Design Science [4]. The objectives can be found in Chapter 4.

3.3 Design and Development

With the help of the objectives, we started the designing and development of the artefact. We prepared the data by both cleaning and filtering it to meet our defined objectives. Afterwards, we decided what models to test and made the necessary changes to the data to fit them. After comparing the models' performances, we continued with the best and trained it one last time. We then performed post-processing of the output to make the result easy to interpret. This included creating a prototype, a mock-up, of the presentation of the findings.

3.3.1 Data Preparation

When designing a good machine learning model, data preparation is a very important step [27]. The preparation was divided into several parts [5], filtering and aggregating the data in different ways, with the aim to more easily reach the objectives of a solution.

We started with a very large and broad set of user data that turned out to be too big for the purpose. This initial data set consisted of the case company's corporate customers' data for their passive FX transactions, a few million of them[1]. We started by filtering this data set by excluding large and institutional customers, transactions between DKK and EUR, the data from 2022 and customers not having data for all four years (2018-2021). The sales team work only with small and medium-sized businesses and therefore the larger corporations and institutions were irrelevant and thus not considered. EUR to DKK transactions were removed since the Danish krona has a fixed exchange rate policy with the Euro [28] and the

[1]The exact amount have been left out since the case company wish to not disclose it

value of the Danish krona can therefore not deviate that much from the Euro, making it not worth planning those transactions. The data for 2022 was not considered since there might be errors in the data, which was not sure to be corrected before two months after the transaction took place. Since we started the data collection in February 2022 there was a possibility of errors in the data from 2022 and it was therefore discarded. Only customers with data for all four years were kept for the customers to have an equal amount of data to not favour any particular pattern [29]. This filtered set resulted in approximately half of the original transactions.

We aggregated the data into weeks, to reduce the data volume and to showcase weekly patterns. To be able to add up the transactions per week for each customer the currency had to be converted to one common currency. The conversion was done by taking the average rate of the day of the specific transaction. We now had approximately tens of thousands of [2] customers with 52 fields each.

Later on, it was discovered that this data set still was too big, our computational power was not enough, and too varied, some customers had a very different pattern in their transactions. The data was therefore filtered further, both by the customers' passive FX transaction turnover, because of relevance to the sales team, and a 26-week seasonality, to find some similarity. This ended up being a data set with roughly 100 customers.

Finally, the data was formatted as a time series, to suit the chosen machine learning models, see Section 2.4 and3.3.2 for more details of the models. Every row represented a unique customer where the first column was the customer id and the other columns were a list of weeks, (year)-W(week number 1-52), with the value as the sum of the transactions for that customer that week. An example of this format can be seen in Table 3.2

Table 3.2: Example of final format of data. (The data in the table is mock data.)

Identifier	2018-W1	2018-W2	2018-W3	...	2021-W51	2021-W52
00318419	0	0	500	...	1000	0
4189481	40936	0	0	...	0	64870

3.3.2 Modelling

We designed the solution by assessing two existing machine learning models, LSTM and N-BEATS which are described in Section 2.4. The models were fed a subset of user data from the case company and evaluated using the root mean square error (RMSE).

In machine learning, it is important to split the data into three sets: training, validation, and test. The training and validation sets are used during the development of the machine learning model. The test set is used to check that the model is general since it is a risk to create a model with biases to the training set (overfitting), The test set should be excluded until the final evaluation of the model and used to test its performance on completely new data [30].

We have one time series for every customer, multiple univariate time series, and we want to predict one time series for each, to get an individual prediction for every customer. Because

[2] The exact amount have been left out since the case company wishes to not disclose it

of this particular data we have split the data both in time and for different customers, an illustration of this can be seen in Figure 3.2. The split in time is to get a test set to use to finally test the best models for all customers on unused data, the best models are then trained on both training and validation sets. Before then, to find the best models the rest of the data is split between the customers into a training and validation set, to make sure the best model is chosen for generalizability.

We split the data into these sets by excluding the year 2021 which were to be used as a test set and the rest of the data was divided into train and validation set, where 70% is training and 30% is validation, which is an often-used ratio [31]. At the time of the split, we did not know how long the test set would need to be, but we figured that a year would be enough. In the end, only the first half of the test set was used and the rest was discarded because we ended up with a model which predicts half a year, the reasoning for this approach is explained in Section 3.3.1.

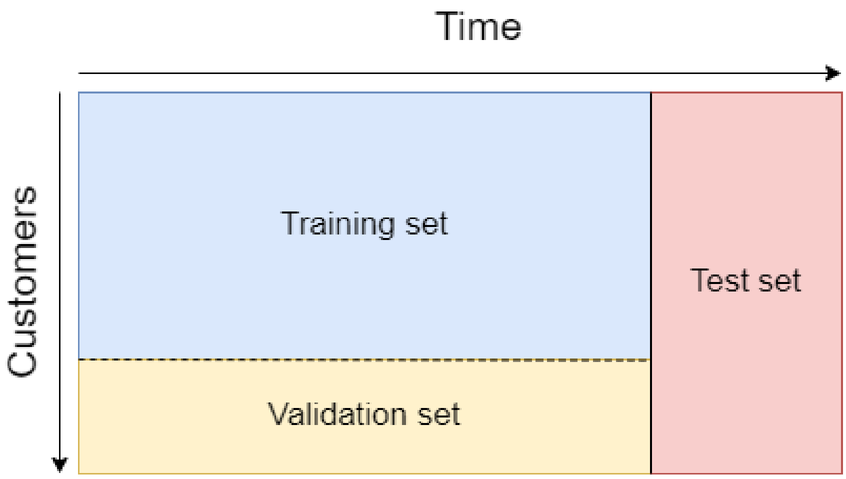


Figure 3.2: Overview of data split

During training the *Mean Squared Error* (MSE) was used as loss function, defined by: $MSE = \frac{1}{n}\sum(y_i - \hat{y}_i)^2$. where n is number of points, y_i is the true value and $\hat{y}_i$ is the predicted value. Small errors have less meaning for us than big ones and by squaring the errors their importance are also squared.

Grid search was used on the train and validation set to find the best hyperparameters for the models, the best model was chosen with respect to the lowest *Root Mean Squared Error* (RMSE) which is the root of MSE, $RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2}$. RMSE was used to compare the models since it is a good general purpose error and as stated in other research is a frequently used measurement. [32].

After grid searching, the models were trained using the combined training and validation set with the chosen hyperparameters for LSTM and N-BEATS respectively. The model performing best on the prediction of the test set was chosen as the best model, the one to keep.

We validated the models' performances by comparing their RMSE scores to a benchmark. We defined our benchmark as a simple model of taking the correct values of the prediction period with the values of the same period last year. To not rely blindly on the quantitative validation, we also performed a visual, qualitative, validation by looking at some of the predictions in the same graph as their correct values with the historic data as well.

Post-Processing

The predicted data was post-processed to find customers with predicted volumes over a certain threshold. Customers passing the threshold were considered interesting, meaning a lead should be generated for that customer.

On the predicted period a simple peak-finding algorithm was run to points in the time series with a high volume of passive FX transactions. The algorithm simply looks at the neighbouring values to see if the current value is a local maximum (peak) or not. The algorithm uses two parameters, minimum height and prominence. The minimum height is the lowest value that would be considered a peak, values below this number are not considered a peak even if they are a local maximum. Prominence is a minimum distance a peak must be above its neighbouring values to be considered a peak. We selected 0.3 for minimum height and 0.1 for prominence. These thresholds were selected by plotting a few predictions and trying to find a value where peaks are more of a peak than a plateau. For prominence, we selected a small value of 0.1 to find only the highest peak in a series of peaks.

We also gathered other information about the customer to help the user understand the machine learning model's result and the context. We generated a PDF per lead where the prediction for the customer and the additional information was displayed.

For the demonstration, explained in Section 3.4, we made three mock-ups of the PDF. Mock-ups are a type of prototype made for qualitative product testing [33]. Prototyping is a good way to get quick feedback from the end-users [34]. Mock-ups convey design details such as colour, fonts and in our case graphs and at least approximately the position and size of the elements [33]. This particular way of presentation was chosen because our artefact displayed information, which suits the properties of a mock-up and we wanted to demonstrate the design as well.

3.3.3 Implementation

The data preparation and modelling were implemented with Python and MS SQL using the code editor Visual Studio Code. We chose Python because it is widely used for machine learning purposes [35] and therefore has a lot of resources online such as guides and Stack Overflow posts. Python also provides many machine learning frameworks, tools, and libraries [35]. For training the machine learning models we used a computer with a GPU, Nvidia Quadro P2000, because GPUs can dramatically speed up the training process for deep learning [36].

We used an open-source Python tool designed for data analysis of time series [37], called Darts, to implement our machine learning models as well as filtering by seasonality in the pre-processing of the data. We mainly chose this tool because it could easily manage multiple univariate time series, meta-learning, which we found difficult to implement, and therefore time-consuming, when trying other tools such as the TensorFlow [3] library. Darts itself is

[3]TensorFlow is an open-source library to develop and train machine learning models, `https://www.tensorflow.org/`

using the PyTorch framework [4] for implementation of the neural networks. We also found Darts' selection of forecasting models to be sufficient and appreciated that it also included pre-processing methods [37] to avoid having to familiarise ourselves with one more tool or library.

To decide upon the best hyperparameters for the machine learning models, we performed a grid search of the parameters in Table 3.3. We implemented the grid search ourselves in Python since the method in Darts, at the time of writing, does not support multiple time series. Our grid search loops through every combination of hyperparameters and for every combination it trains a model on the training set and performs a prediction using the validation set and saves the RMSE, explained in Section 3.3.2.

Table 3.3: Parameters for grid search

LSTM	
# Layers	[1, 2, 4, 8, 12]
# Hidden dim	[100, 150, 200, 250]
Dropout rate	[0.2, 0.3, 0.4]
Learning rate	[0.1, 0.01, 0.001]

N-BEATS	
# Stacks	[10, 20, 30]
# Blocks	[3, 6]
# Layers	[8, 12]
Layer widths	[128, 256]
Expansion coefficient dim	[3, 5]
Batch size	[16, 32]
Learning rate	[0.01, 0.001]

To create the display of the application we generated a PDF with the information from the model as well as information gathered from the database. For every interesting customer, we generated a graph with the history of the passive transaction turnover, the predicted turnover and gathered the information on when the predicted peaks occurred. From the database, we gathered general information about the customer and summarised their passive FX transactions. With the gathered information we generated a PDF for every interesting customer which is the part of the application to be displayed to the users. The PDF was generated using the open-source python library FPDF2 [5].

3.3.4 Design Decisions

When designing the application we took several major design decisions, especially the choice of: *seasonality*, *machine learning models*, *creating one model for all customers*, *hyperparameters in the grid search* and *generation of PDFs*. These major design decisions are denoted in italics throughout this section. The reasoning behind these decisions is presented in this section, and we discuss them further, as well as other possible choices we could have made instead in Chapter 5.

During interview set I1, the interviewees emphasized that customers could, in their experience, have a seasonality of a year or more. Therefore, we initially tested to filter on seasonality of one year, but since no customer had this property no customers remained after filtering. After some exploration, we filtered on a seasonality of 26 weeks since a reasonable

[4] PyTorch is an open-source machine learning framework for going from research prototyping to production deployment `https://pytorch.org/`

[5] https://pyfpdf.github.io/fpdf2/index.html

number of customers remained. When selecting what seasonality to investigate, we did not yet know that we would only have time to test one seasonality.

The *machine learning models*, LSTM and N-BEATS, were selected because, among other things, they are from two different families of deep learning networks and therefore learn in different ways, see Section 2.4. Another reason for us to pick LSTM was because of its long history of good performance for time series forecasting [38, 39, 40]. N-BEATS were chosen for being a fairly new model, from 2020, that has shown promising results in a forecasting competition [19].

We chose to *create one model for all customers* instead of one model for each customer, thereby using a more extensive set of user data for training our (one) model. From our experience, it is possible to create a model general enough to be used on new customers without training on their previous data. New customers require less data before a lead can be generated if using a general model. For example, if we only need half a year of data to make a prediction, we can predict new customers after just half a year instead of waiting for the data to become large enough to construct a proper training set.

The choices of *hyperparameters* to test for LSTM were decided by taking inspiration from a paper investigating hyperparameters for an LSTM network [41]. However, we had to reduce the number of parameters we searched since the training time would have been heavily increased. For the N-BEATS model, we chose to use the generic architecture and changed the hyperparameters we could since we did not find a paper testing different hyperparameters for the network.

We had three options to choose between when displaying the results in our application, namely pure text, an excel sheet, or *PDFs*. From the individual interviews presented in full in Section 4.1.2 we found that the interviewees wished to have graphs and tables in the presentation. Therefore, we prioritised a display method that could display graphs, and we finally landed on the choice of PDF since we found a tool for Python, which we could easily create PDFs.

3.4 Demonstration and Evaluation of Application

In the demonstration step [42], the application is tested and evaluated, to see if it satisfies the defined objectives, see Chapter 4. For this, three mock-ups, see Section 3.3.2, were made and presented for a focus group. The group for the focus groups existed of three people from the sales team to get insights and answers from the end-users. The focus groups were conducted in a conference room at the case company. For the first iteration, the meeting lasted for about 45 minutes and in the second iteration, it lasted 25 minutes.

The purpose of the focus group, the same for both iterations, was to investigate improvements of our artefact, but also to find out what was good. Five questions, see Appendix C, were made by and for us as support to not let the focus group drift off topic too much and were only used if the conversation did not reach it by itself. The focus group approach served us well since we wanted to collect qualitative data, our group had something in common and we had a clear topic, which is all part of most definitions of a focus group [43].

By analysing the demonstration, keeping in mind the set objectives, we could get a good picture of how well the artefact performed. The analysis included our notes from the demon-

Table 3.4: Participants of the focus groups

Current role at the case company	Time worked at the case company	Time worked within the financial domain
Currency advisor 1	Half a year	Half a year
Currency advisor 2	4 years	4 years
Senior currency advisor	5 years	8 years

stration and listening to the recording, making more notes of things missed during the focus group. Later, these notes were internally discussed as we tried to answer three evaluating questions, helping us answer the ultimate business objective questions:

- Does the artefact bring new and relevant customer information to the user/sales team?
- Does the artefact provide the right information?
- Is the information easy to comprehend?

3.5 Identifying Challenges

After the final design science step, we performed a data collection of the challenges faced throughout the thesis, to answer RQ3. We performed a brainstorming session to identify the challenges experienced during our thesis work in a systematic and structured way. The participants of this session were the authors of this thesis. The brainstorming session was divided into two parts, an individual and a group part. In the individual part, we brainstormed challenges by ourselves, to not influence each other, and wrote them down. The individual part was about 20 minutes. Afterwards, we compared each other's notes and brainstormed together to end up with a final set of challenges.

Chapter 4
Results

We gained insights into the problem domain through designing and developing an ML-based application and validating this artefact in the industrial context of our case company. In this chapter, we describe our findings from each of the four parts of our study, see Chapter 3, namely *Identify problem and motivate*, *Define Objectives of a Solution*, *Design and Development*, and *Demonstration and Evaluation*. Finally, at the end of the chapter, we present the challenges we encountered during the project.

4.1 Identify Problem and Motivate

We explored the problem domain through interviews (group I1 and individual I2) and performed thematic analysis on the transcripts of both interview sets. We have summarised the most important findings and provided the final themes from the thematic analysis in Table 4.1 and Table 4.2. The themes are denoted using *italics* in this section.

4.1.1 Group Interviews (I1)

We got a wide array of information about the case company's business, data, and previous work with time series from I1. The themes from the analysis of I1 are: *Reasons to reach out to customers*, *Customers*, *Wish list*, *Challenges in implementation*, *Model building for time series*, *Segmenting*, and *The sales team*. The themes and the list of codes can be found in Table 4.1.

Our interviewees believed *Reasons to reach out to customers* mainly were related to behaviour in their FX transactions, which can be either changes or patterns in behaviour. An example of a change in behaviour our interviewees gave us is when a customer starts doing transactions in a new currency. Our interviewees also said that some customers do unnecessary transactions, either going back and forth or to an intermediate currency instead of directly to the final currency. We got informed that our interviewees have limited knowledge about behaviour changes and patterns right now, so any information they can get is desired.

Our interviewees described the *Customers* as inexperienced in hedging. One interviewee said "...they don't want to hedge because if they do a hedge they might lose money". The group clarified the meaning of hedging, describing it as a tool used to decrease or remove a customer's currency risk by buying currency before a customer needs it. Otherwise, the customer risks buying currency during a price spike, which can increase costs. However, the customer might lose money if the price decreases. They also expressed that some customers are in more need of guidance than others, for example, growth customers or customers not hedging their transactions.

We found that the sales team had a *Wish list* for new leads. The interviewees expressed a desire for alerts and recommendations to reach out to customers who needed guidance proactively and wanted to know why they got the lead. When getting an alert, our interviewees wanted to have as much information as possible to understand how the customer deals with currency and if it would be good to reach out to them.

Our interviewees mentioned possible *Challenges in implementation* of the models, especially with data which can have errors such as faulty transactions. The errors are corrected, but this can take months and can impact the accuracy of any model. The data scientist informed us that when doing *Model building for time series*, previous work at the case company has shown that it is hard to get a good forecast, that RNN is a common model for forecasting and that most customers only have one bank, so we have data for all FX transaction for those customers. It was also pointed out that customers are unique but that the size of a customer is usually a good metric for *Segmenting* them.

Finally, *the sales team* expressed a need to describe the lead in a way they understand. The interviewees pointed out that they are not data scientists and need an explanation that is not too technical. Our interviewees highlighted this by informing us that the sales team does not know what ML models are or how they work, and this can not be expected.

4.1.2 Individual Interviews (I2)

Through the interviews, we found that the sales team thought there was both needed and wanted *Information* to include in the lead. The *Needed* information was mostly background information about the customer, such as the customer's name or organisation number. Additional background information, for example, the short name or the branch of the customer, was information the sales team *Wanted* to include. This would make the look-up of the customer faster and would increase their knowledge and therefore help to decide on an action or not. The complete list of information needed and wanted can be found in Table 4.2.

When it comes to the *Delivery of the lead*, our interviewees wished that the lead would be included in the current CRM system and an email would be sent out to the people that can work with the lead, telling them a new lead is available. Our interviewees also expressed some wishes regarding the display of the lead, especially that graphs and tables would be appreciated since the comprehension would increase. All interviewees expressed a limit of a couple of times a week to notify a sales team member of a new lead, but this does not concern us since we are not doing the deployment of the lead.

Lastly, our interviewees described how the sales team *Work with leads today*, we found out that most of the leads should not require much work to see if it is relevant since all interviews

Table 4.1: Themes and codes from the analysis of I1. The themes are denoted in bold with gray background.

Reasons to reach out to customers	
Flows in new currency	Increase or decrease in turnover
Peaks or drops in hedging ratio	Change in direction of currency flows
Unnecessary transactions	Patterns in FX transactions
Hedging less than their need, risk of losing money	
Customers	
Inexperienced in hedging	Afraid of doing bad trades, losing money
Hedging less than their need	Some have flows that can change over time
Growth customers have a need for guidance	Many smaller trades indicate continuous business
Wish list	
Finding thresholds	Finding explanations for trends and pattern
Proactively reach out to customers	Being able to find the end of hedges
Alerts	Graphs
Show customers how it would have gone if they made different choices	
Challenges in implementation	
Filter out incorrect trades	One-off trades can create skewed statistics
Model building for time series	
Forecasting is difficult	RNN most used method
Active transactions data can be important to not miss the bigger picture	We see all transactions
Segmenting	
Customers are different	Size is a good measurement
The sales team	
Not computer scientists	Need to be in front of customer
Need easy to understand and good explanation	

described a lack of time.

Table 4.2: Themes and codes from the analysis of I2. The themes are denoted in bold with gray background and the sub themes in italics

Information	
Needed:	*Wanted:*
Name of the customer	Short name
Id of the customer	Relationship Manager
Organisation number	Currencies traded and the volume
The actual lead	What branch the customer is part of
Historical passive FX transactions turnover	Historical hedging
Turnover compared to company size	
Delivery of the lead	
Might be different depending on the person receiving the lead	By email first with link to CRM
Always in CRM	Reminder of the lead
Graphs would be nice	A pivot table would be nice
Couple of times a week	
Working with leads today	
Talking with the RM	Determining if the lead is relevant
Not too much work to check if relevant	Does not check CRM system daily
Several people might get the same lead	

4.2 Define Objectives of a Solution

From the findings of the business and data understanding, we defined objectives together with the Team Lead. Thereafter, we, without the Team Lead, phrased these objectives and added the motivation, ending up with the following objectives for the solution (the definition is in bold and the themes from Section 4.1.1 and 4.1.2 are in *italics*):

To develop an ML-based application that finds customers that have a seasonal pattern in their passive FX transactions and that are therefore potentially interesting to reach out to. Patterns in the customer's FX transactions were seen by the sales team as a *Reasons to reach out to customers*. Finding these patterns would also fulfil an item on the sales team's *Wish list*, giving them enough resources to proactively reach out to customers.

The customer should be a small or medium corporation with a turnover for their passive FX transactions over a certain threshold[1]. *Segmenting* of the customers was mainly necessary to decrease the volume of data. From previous experience of the case company the customer's size was often a good and easy measure to use for this.

[1] The exact threshold has been left out since the case company wish to not disclose it.

The ML-based application has to present the output in a comprehensible manner that provides the user with enough information to be able to decide whether to act upon it or not. When *Working with leads today* the sales team find it important that they quickly can determine if the lead is indeed interesting and that they do not have the time nor patience to lookup much more information themselves. The terminology and level of the technical language also have to suit *The sales team*.

Possible challenges are the quality of the data which might contain errors and that forecasting is a hard task. These two challenges were taken directly from *Challenges in implementation* and *Model building for time series*.

The ML-based application is supposed to provide the user with deeper insight into the customer's transaction patterns and information needed to quickly reach out to the customer to provide guidance.

4.3 Design and Development

We designed and developed the ML-based application keeping in mind the problem domain knowledge and our defined objectives for our solution. We will present our quantitative and qualitative results of the LSTM and N-BEATS models as well as describe the design of the mock-ups made for the demonstration of the application.

4.3.1 Model Evaluation

The grid search for the best hyperparameters for the LSTM and N-BEATS models has been provided in Table 4.3. Both models had similar performance with an RMSE that just differed by about 0.01, both RMSEs are provided in Table 4.4.

Table 4.3: Best parameters for grid search

LSTM	
# Layers	1
# Hidden dim	200
Dropout rate	0.4
Learning rate	0.1

N-BEATS	
# Stacks	30
# Blocks	6
# Layers	12
Layer widths	128
Expansion coefficient dim	3
Batch size	16
Learning rate	0.001

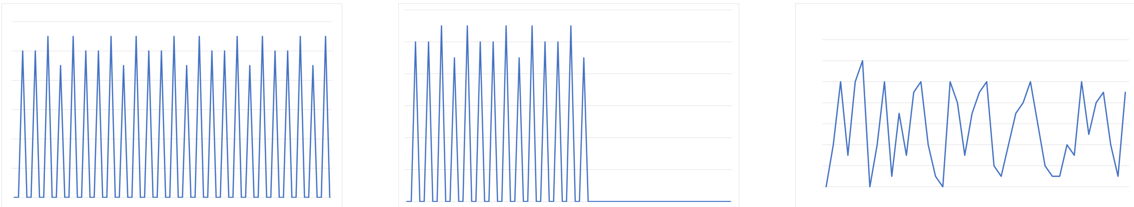
The two best performing models from the grid search were retrained using the training and validation sets. Their resulting RMSE showed that both models performed better than the benchmark and the LSTM model performed the best with an RMSE approximately 0.006 units lower than the N-BEATS, the full errors are provided in Table 4.4.

Our visual inspection of the models, see Section 3.3.2, found that there are three common patterns in the test data, see Figure 4.1, patterns that are predictable (a), patterns that have flat lines (b), and patterns that have more variation (c). For predictable patterns, both models successfully predict the pattern. For the flat-lining and varied patterns, we saw some differences. The N-BEATS model manages patterns which flat line better by almost matching the

Table 4.4: Performance on the validation and test set for every model with normalised time series.

Model	RMSE validation set	RMSE test set
LSTM	0.155872	0.144751
N-BEATS	0.165637	0.150903
Benchmark		0.202429

flat line while the LSTM model instead keeps following the previous pattern. Both models have a harder time predicting varied patterns but we saw some success in the predictions. The LSTM model was not better at any type of pattern but matched patterns with generally lower amplitude better. From the final data set of about 100 customers, our application found roughly 10 customers which it generated leads for.



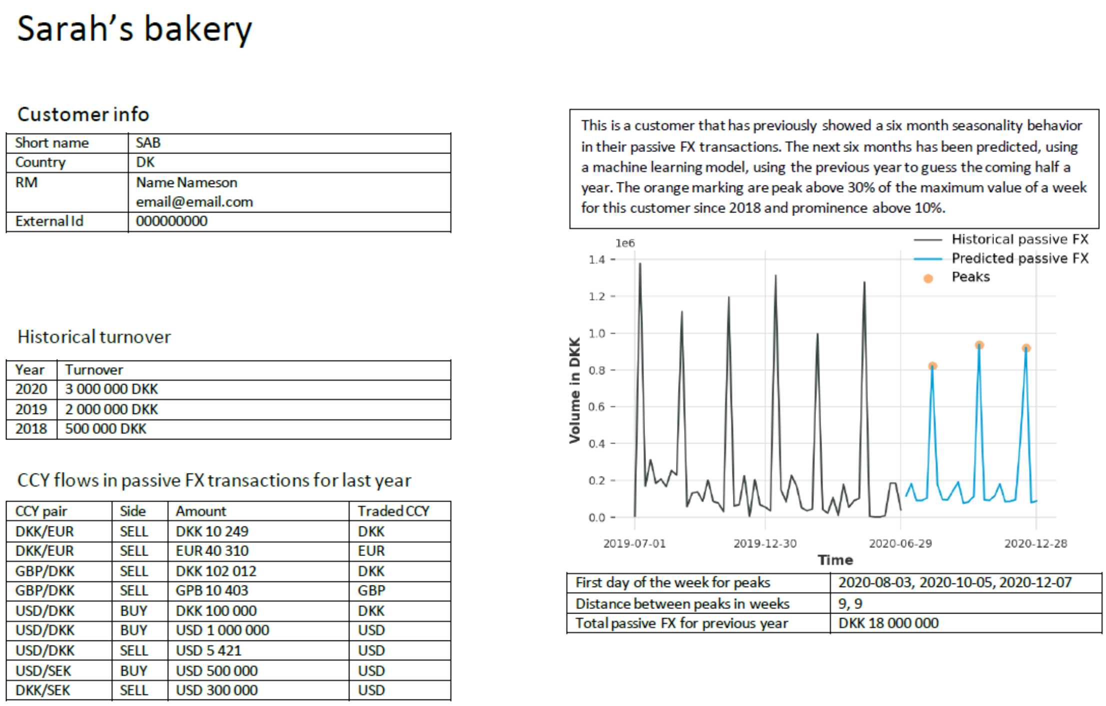
(a) **(b)** **(c)**

Figure 4.1: Common patterns of data found in the qualitative evaluation of the models

4.3.2 Mock-ups

The first iteration's mock-ups, see example in Figure 4.2, consist of a front page with two columns, the left one contains background information about the customer and the right contains information about why this lead has been generated. The left column has three tables. The first table has some background information about the customer such as their short name and country, the second table is a list of their turnover for the last three years and the last table contains their passive FX transactions aggregated by currency pair for the last year. The right column contains a text about the generation of the lead, a graph with the customer's historical and predicted passive FX turnover and a marking on each interesting point for the predicted turnover. After the graph, a table follows with information about the distance of the peaks in weeks, at what week the peaks occur and the customer's passive FX turnover for the last year. The front page of the mock-up is provided in Figure 4.2. After the front page, a variable amount of pages follows with a table of all the customer's passive transactions from the last year. The transactions include the amount of traded currency, which currencies are traded and when the trade occurred.

The mock-ups were updated in the second design iteration based on the feedback from the focus group, see Section 4.4.2. The updated version of the mock-up is provided in Figure 4.3.



Sarah's bakery

Customer info

Short name	SAB
Country	DK
RM	Name Nameson email@email.com
External Id	000000000

This is a customer that has previously showed a six month seasonality behavior in their passive FX transactions. The next six months has been predicted, using a machine learning model, using the previous year to guess the coming half a year. The orange marking are peak above 30% of the maximum value of a week for this customer since 2018 and prominence above 10%.

Historical turnover

Year	Turnover
2020	3 000 000 DKK
2019	2 000 000 DKK
2018	500 000 DKK

First day of the week for peaks	2020-08-03, 2020-10-05, 2020-12-07
Distance between peaks in weeks	9, 9
Total passive FX for previous year	DKK 18 000 000

CCY flows in passive FX transactions for last year

CCY pair	Side	Amount	Traded CCY
DKK/EUR	SELL	DKK 10 249	DKK
DKK/EUR	SELL	EUR 40 310	EUR
GBP/DKK	SELL	DKK 102 012	DKK
GBP/DKK	SELL	GPB 10 403	GBP
USD/DKK	BUY	DKK 100 000	DKK
USD/DKK	BUY	USD 1 000 000	USD
USD/DKK	SELL	USD 5 421	USD
USD/SEK	BUY	USD 500 000	USD
DKK/SEK	SELL	USD 300 000	USD

Figure 4.2: Mock-up from first iteration, all data is mock data to not disclose data about the case company's customers

4.4 Demonstration and Evaluation of Application

We created three PDF mock-ups for the first iteration, which we later updated to the second iteration, and evaluated them using a focus group. We will present the result of the evaluation for both iterations as well as changes made between iterations in this section.

4.4.1 First Iteration

The mock-up presented to the focus group is described in Section 4.3.2 and provided in Figure 4.2. The focus group's initial reaction to the mock-up was confusion as to what the lead description was and what they were supposed to get from it. After some discussion, their reaction was instead positive and they analysed one of the customers and thought the lead was interesting and told us that it was intriguing to see a customer with regular patterns. It was, however, obvious that the mock-up had to be clarified for the focus group to be able to understand and work with the PDFs generated by the application. We have summarised the analysis of the first focus group in the following list of comments:

- The quoting of currencies in the aggregated transactions table should be written from the customers perspective. The current table is ambiguous.

- The traded currency column in the aggregated transactions table is unnecessary since the first currency is usually the traded currency, having a column to explain it instead of doing it in the usual way can cause confusion.
- The intro text to the lead is not what the focus group was looking for instead they want a text explaining why the customer is interesting.
- The PDF should have a header explaining what kind of lead it is.
- The number of transactions per currency pair in the aggregated transactions table would be nice to have.
- Instead of just a table of the currencies traded a chart showing how large every currency pair is of their customers total transactions would be nice.
- The raw data after the front page is unnecessary and nothing the focus group would look at.
- Including a demonstration of their active FX transactions as well as their passive transactions would be useful to assess the customer.
- What branch the customer belongs to would be good background information.

4.4.2 Second Iteration

A change was made for all the comments in the summation, Section 4.4.1, except the addition of active FX transaction data because of time constraints. The updated mock-up is provided in Figure 4.3. The second mock-up has the same two-column layout as the first iteration although the information is spread over two pages instead. The table with currency pairs has been updated based on our notes and the number of transactions for each pair has been added. The intro text and heading have been updated to reflect the seasonality of the customer, for example, bi-weekly or monthly seasonality. What branch the customer belongs to has been added as well as a bar chart over how much every currency pair contributes to the passive FX transactions' total value.

The second focus group was positive about the updated mock-up and said it was easier to understand. There was no new information they wanted to add and was overall happy with the state of the PDF. They did, however, have a few comments:

- The bar chart, showing currency pairs, should not show entities that are less than 5%.
- The quoting of the currency pairs are wrong since there are conventions of which currency should appear first. For example euro, EUR, should always be stated first in a pair.
- The dates. in the intro text, are unnecessary and it is enough to have them below the table.

Passive FX seasonality lead

Customer Info

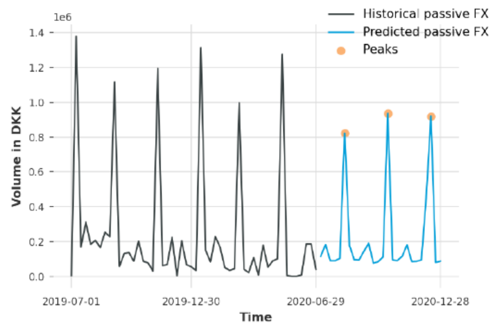
Name	Sarah's Bakery
Short name	SAB
Country	SE
RM	Name Nameson 0000 email@email.com
Branch	Branch name
External ID	00000000

Historical passive FX turnover

2020	DKK 5 000 000
2019	DKK 3 000 000
2018	DKK 500 000

Based on this customer's previous autogbveks transactions, it has been predicted to have a bimonthly seasonality pattern. In the graph below you can see the last year's weekly autogbveks turnover and the coming predicted 6 months. The peaks will likely happen bimonthly, with peaks in the week starting at date: 2020-08-03, 2020-10-05, 2020-12-07.

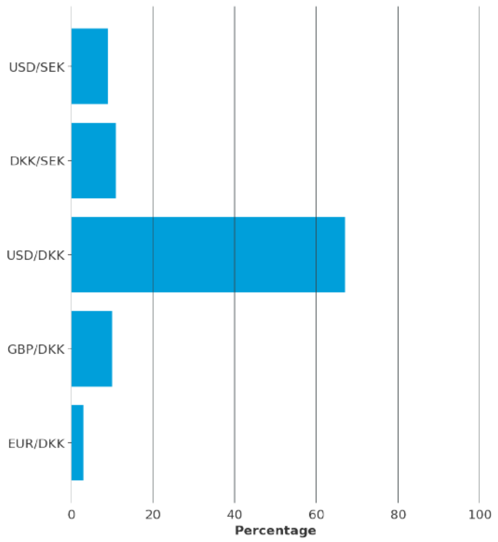
Timing of bimonthly peaks



Monday of the week per peak	2020-08-03, 2020-10-05, 2020-12-07
Distance between peaks (in weeks)	9, 9

(a) First page

Passive FX seasonality lead



CCY flows for 2020

CCY pair	CustSide	Amount	Nbr
DKK/EUR	SELL	DKK 10 249	77
DKK/EUR	SELL	DKK 40 310	47
GBP/DKK	BUY	DKK 102 012	89
GBP/DKK	BUY	GBP 10 403	24
USD/DKK	SELl	DKK 100 000	56
USD/DKK	SELL	USD 1 000 000	61
USD/DKK	BUY	USD 5 421	22
USD/SEK	SELL	USD 500 000	78
USD/SEK	BUY	USD 300 000	1

(b) Second page

Figure 4.3: Mock-up from second iteration

4.5 Challenges

The brainstorming session concluded with four different areas of challenges: Machine Learning, Database, Business understanding and organisation. Here we will present the challenges briefly, in Section 5.3 we will dive deeper and discuss the reasons behind the challenges as well as the problems they caused.

We faced a few challenges regarding machine learning, most of them related to the vast amount of data. We found that all parts of the machine learning took a very long time, especially the grid search, taking several weeks of 24/7 training to complete. Because of this, we could not perform as wide a grid search as we would have liked and it was also hard to test new ideas since the time investment would have been large and the thesis have a set end date. When we had a finished model it was challenging to know how it would perform when deployed since it generates leads over time when new passive transaction data gets generated.

The challenges associated with the database mostly concerned the data gathering for the application. We found the structure of the database confusing, the database was also partially undocumented. Therefore, we could not find data on our own and we had to rely on members of the FX data analytics team to find data. We also found that data resided in multiple places and it was impossible to be sure what data source was the correct one to use. There was also a huge amount of data and it, therefore, took a long time to run certain queries (some up to six hours), especially if we wanted to know how many rows a query returned.

During the thesis, we faced a few misunderstandings based on a lack of business understanding on our part. Even though we thought we understood the domain well enough, we misunderstood a few key points during the thesis. For example, in the first iteration during the focus group, we got feedback on our use of turnover. We had wrongly assumed that the turnover referred to the customer's turnover while the sales team meant the customer's passive FX transaction turnover. Another example during the first focus group was the quoting of currencies. We thought we understood but missed an arbitrary rule of which currency should be first as described in Section 4.4.2.

Another challenge we faced was that gaining access to data or systems could take a long time. The reason for this is that we needed to apply for access and sometimes the approval process takes some time.

Chapter 5
Discussion

The work done throughout the thesis has given us a lot of thoughts on potential improvements when designing an ML-based application and working with passive FX data. We will discuss our results per research questions and the validity of our findings.

5.1 Customer Insights (RQ1)

From the work with the passive FX data, we learned that many types of lead generating applications can be done using the data. For example, *seasonal patterns* (as we used), *changes in behaviour*, and *unnecessary transactions*. We will discuss our thoughts regarding these applications of customer data.

We saw that the *seasonal patterns* exist as described in Section 4.3.1 but a large number of customers means that there is also a large variation in behaviour patterns which we saw before filtering the customers. We found that the future can be predicted well for some customers. This was mostly the case for customers with patterns that are repetitive or seasonal.

Another idea presented in the group interviews I1 was to look at *changes in behaviour*. For example sudden changes in the turnover of their passive FX transactions or transactions in new currencies. Future work is to investigate the customer data to see how often this occurs, but we think a reasonably simple model could find additional interesting customers by analysing the data from this angle.

Something we saw during our focus groups was that our interviewees inspected the table over the total amount of currency bought and sold per currency and directly saw that some customers seem to do *unnecessary transactions* back and forth between three currencies, which they found interesting. This is something they mentioned in interview set I1 and it was exciting to see the focus group find these kinds of transactions in our application. Something to look at as future work is a model that simply summarises the data per customer and per currency which most likely could find the customer doing these kinds of unnecessary transactions.

5.2 Machine Learning (RQ2)

We developed an ML-based application to gain knowledge about the seasonal trend of a customer's transaction patterns. During the thesis, we discovered other possible uses for machine learning and possible improvements to the model we created. We will discuss the result of our models, choices regarding the model, and possible improvements to the model.

We created an application, which summarises customers' passive FX transaction data into weeks and predicts their future turnover. We filtered customers based on a 26-week seasonality and analysed the predicted turnover to find weeks where the prediction peaks. Since the focus groups found the patterns of the customers intriguing we have shown that new insights can be generated with an ML-based application using the passive transaction data. We also noted that our focus groups thought that the summary of the passive FX transaction per currency pair was interesting and could be used to find unnecessary transactions.

In the early stages of the thesis, we discussed possible approaches of what we could use machine learning for based on the data collected from the group interviews. Some of the ideas were to classify customers using the passive transaction data or detect anomalies in the behaviour patterns. Further work can be done to develop ML-based applications based on these concepts.

5.2.1 Possible Improvements to Implementation

After the modelling phase, we realised several possible improvements that could have been made but were not due to time constraints. These improvements mainly related to the *grid search* which were very small, how to test the models using *backtesting*, and how we could have handled the *confidence* of the predictions.

For the *grid search*, there are two different things to discuss. First of all, we would have liked to have the resulting best hyperparameters not at the edges of their intervals. This is to show that we have searched the space wide enough to, with some certainty, be able to say that the result would not have been better even if we expanded the interval for that specific hyperparameter. This is not the case, especially not for N-BEATS where it was mostly impossible to not be at an edge since most intervals only included two candidates. Here, we could have decided to decrease the number of searched hyperparameters to be able to increase the intervals, but we did not have an indication of what hyperparameters were more important than the others. Therefore, we decided that, although not ideal, to run the search with the initial set of hyperparameters with smaller intervals than we would have liked.

Secondly, LSTM performed best while having only one layer and therefore quite a small model overall, this surprised us. In our previous experience bigger networks, up to a certain point, usually perform better.

From the test of our models, we saw that the models performed better than the benchmark and our qualitative analysis found there are some customers the prediction work well for. However, our model only found that ten of our about 100 customers had peaks large enough to generate leads. Because we did not have the time to do *backtesting* and check other seasonalities we do not know how good this is. Our application generated only ten leads, but we expect other sesonalites to generate additional leads. Likewise, more leads will hopefully be generated over time when more data is being created. Otherwise, some tuning of the filtering might be required to include more customers in the data set.

Something that can also be done to evaluate the *confidence* in the predictions is to use the mean absolute percentage error (MAPE) which is an error used to measure the accuracy of predictions. The error could be used on historical data and indicate how good a model has been to predict the customers' previous data. If the MAPE is below some threshold for a customer we can remove it as a potential lead even though the model otherwise would classify it as interesting. By doing this we can increase the confidence in the leads we do generate to not send false leads to the sales team.

We choose to create one general model for all customers since we wanted one more general model, for example not have to create and train a new machine learning model for a new customer. A new customer here could either be a completely new customer for the bank or a customer that has changed behaviour and is showing another seasonality in their passive FX transactions. Another option would be to instead create a model per customer this could have the benefit of making the predictions more accurate but is something that would need to be investigated. However, it would increase the complexity and most likely make the performance for customers with data for a short time period worse.

Another use of the time series data would have been to try to cluster the customer into new segments instead of trying to predict the future data. Since machine learning can find patterns we humans could not. The clustering method could lead to the same result by for example clustering customers that have a monthly seasonal behaviour based on their past transaction data. Customers in groups could then be picked one by one to receive guidance and new leads could be generated if a new customer is clustered into a group.

5.3 Identifying Challenges (RQ3)

During our work with designing and implementing an ML-based application for generating leads, we identified several challenges in four areas Machine Learning, Database, Business understanding and organisation (see Section 4.5). We will herein discuss these challenges, how they impacted us and how they were or could be addressed.

5.3.1 Customer Data (RQ3a)

We encountered several challenges when working with this thesis, the largest challenge was related to *long training times* but we also faced challenges regarding *gaining access to data*, *performance of the model in deployment*, and the *unorganised database*

The most time consuming and the biggest challenge we faced was the *long training times* of machine learning models because of the large amount of data. We alleviated this challenge by aggregating the data into weeks to reduce the number of data points in each time series. If we had not done this the use of machine learning would have been infeasible with the hardware we had access to. This might be a sub-optimal solution since there might be patterns in the data related to exact days or times of the day which were lost during aggregation. Another solution would be to acquire high-end machine learning hardware to speed up training enough to make it feasible to use the data set without aggregation. The long training time also impacted our ability to do a wide grid search since every added parameter exponentially increases the total amount of models to train. At later stages in the thesis, we discovered a possible solution to this problem, a method called random search which instead of trying all possible combinations of parameters randomly tests different parameters and has even shown better performance than grid search [44]. Unfortunately, we discovered this when the grid search was almost complete and we did not have the time to test random search, but is something we will bear in mind in future projects.

A smaller challenge we faced was the time it took to *gain access to data*, after an application had been sent it took a varying amount of time for the application to be accepted. This challenge was mostly a problem before we knew about this procedure and did not plan for the delay. To be noted is that the procedure taking some time is an indication of high security and that the data request is being handled with respect. Therefore, this challenge should remain for user data integrity but not be forgotten.

Another challenge was to know the *performance of the model in deployment*, since the model can generate leads as new data is being generated by customers performing new transactions. A possible solution to this would be back-testing, a method of simulating time passing by feeding the model with old data over time and recording how the model responds. We did not have time to test this during the thesis since after the training of models was completed not much time remained until the end of the thesis.

The problem with an *unorganised database* did not impact us that much since we could easily ask the members of the data analytics team for help and got quick responses. However, without the help we got, it would have been a massive challenge. We believe the most straightforward solution would be to keep an updated overview of the whole database to which everyone has access.

5.3.2 Achieving problem-solution fit (RQ3b)

The main challenge to achieve problem-solution fit was *misunderstandings* caused by a lack of business understanding or a language barrier. We also think our problem-solution fit would have increased with additional *iterations* and time to explore more of our ideas.

In Section 4.5, we bring up two examples of *misunderstandings* that happened during the thesis. We think the mismatch of understanding happened partly because of the sales team's tacit knowledge about the economic field that we do not possess. During the thesis, we also mostly communicated with the people we worked with in English which as far as we know was everyone's second language and not the language everyone used daily. To communicate in a business setting with a second language is of course a challenge if not done regularly and we think this might be an alternative or a second cause of misunderstandings. We also realised that misunderstandings can be hard to detect since we did not detect the misunderstanding related to turnover described in Section 4.5 for a few weeks. A possible solution to this problem is regular check-ins so the party performing the work can give status updates, by doing this we hope that misunderstandings will surface earlier in the process.

Another challenge we faced regarding the problem-solution fit was the number of *iterations*. We had the time to make two iterations, but we still had improvements to be made. For example, the feedback we got from the second focus group could have been fixed to achieve a better problem-solution fit. This challenge would be solved simply by having more time to be able to increase the number of iterations. More time would also allow testing more ideas, for the machine learning, discussed in section 5.2. For example, we would have been able to test several different seasonalities of the data with more time which would increase our understanding of the problem-solution fit for our application.

5.4 Validity Discussion

There are always multiple threats to the validity of a study and it is important to address and discuss these for transparency and credibility. There are numerous ways to divide validity threats into categories and we have decided to separate them into two categories: internal threats and external threats. Internal threats are everything that influences the results without the researchers being aware of it, whereas, external threats risk the generalisability of the results [45].

During the data collection, we analysed the answers of the interviewees and constructed themes and codes. We then decided what to include in the report based on what we thought was important, which introduced our bias. We tried to mitigate this by both being part of the transcription process, one of us created a first draft of the transcription and the other person re-listened to the interview and checked the draft at the same time. We also independently highlighted what we thought was important from the transcription and then created themes and codes based on our separate highlights.

As described in Section 4.5 misunderstandings occurred when we interpreted and analysed the interviewees' answers. We tried to mitigate this threat by contacting the interviewees if something they said was unclear or if we had questions after the interview. However, as shown by our challenges some misunderstandings were caught later in the process and there is a possibility we did not catch all misunderstandings.

The results from the study are influenced by the case company and its employees that we used for our different interviews and we can therefore not claim that someone else would receive the same result at a similar project elsewhere. To mitigate this external threat, we have returned to the literature throughout the thesis to compare with and get input from other research. This has been done, for example, to decide what methods to use for our data collection, such as focus groups, and more technical decisions, such as the selection of machine learning models.

Our customer data is unique, received from our case company, and we can of course only speculate about how the results of the machine learning methods would have been on other data. We can guess that customers of other banks have similar patterns and passive FX transaction data, but there would have to be done several other similar studies in the future using another case company to confirm this.

We had relatively few people that we interviewed and that was in the focus groups for the demonstration. This could pose a threat to the generalisability, because of the small sample of opinions. However, the people all had expertise within the area and had experience in similar work. We also had the same participants in both focus groups, due to a limit in possibilities. This could skew the results of the second focus group since they had already seen the mock-ups before and might have remembered some of our explanations from the first time. We tried to avoid this by having a long time as possible in between the focus groups.

Chapter 6
Conclusion

Our goal with this thesis was to investigate, in the finance domain, how customer data can be used for CRM, how machine learning can be used to generate customer insights, and what challenges we identified throughout the process. We investigated the case company's customers' passive FX transaction data to design and develop an ML-based application to predict the customers' future transactions and find patterns in the predicted data.

We found that seasonal patterns exist for some customers, and our machine learning model predicts them reasonably well. However, the customer's data varies, and some patterns are more complicated to predict than others. We also theorised that there are other potential uses for the data we did not investigate, such as changes in behaviour or trading in new currencies. The application created predictions based on a 26-week seasonality and presented the predictions and other relevant data about the customers in a PDF format. The PDF and especially the prediction displayed in the PDF created new insights for the sales team. The sales team found the predictions intriguing during our demonstration of the application.

We identified challenges in four areas Machine Learning, Database, Business understanding and organisation. The machine learning challenges were mainly caused by the long training times of the machine learning models, which led to fewer ideas being tested, a smaller grid search, and not being able to evaluate how well the model would work after deployment. We concluded that modern and correct hardware for the task is crucial for working with machine learning to reduce the training time, which leaves more time to make changes and improvements to the application. An unorganised and unstructured database made it harder for us to find the data we wanted to incorporate into the ML models. The business understanding challenges were linked to tacit knowledge causing misunderstandings leading to errors in the design of the application. Finally, gaining access to data sources took more time than anticipated because of the organisation's security. Practitioners and researchers alike can use the challenges and knowledge collected and learned by us to not repeat the same mistakes. The case company will hopefully deploy the application, which will create value for the sales team and the customers of the case company.

Our work adds an example of how CRISP-DM can be used in the finance sector and how machine learning can be used to create insights into a new type of data. For our specific case CRISP-DM worked well. It forced us to gain the knowledge required to complete the project and if you have never worked with machine learning before it explains all necessary phases that are needed to create a well performing model.

Future work would be to test more machine learning models on this kind of data and do an extensive grid search for the machine learning models we tested. Furthermore, we do not know how our application works when deployed. Will it create new additional leads as time goes on or not? Future research has to be conducted in this area to answer the question of how the solution will work in practice. This could be done by either deploying the application and evaluate it continuously or by backtesting.

References

[1] F. Rajola, *Customer relationship management in the financial industry: Organizational processes and technology innovation.* Springer Science & Business Media, 2014.

[2] S. Chakrabarti, M. Ester, U. Fayyad, J. Gehrke, J. Han, S. Morishita, G. Piatetsky-Shapiro, and W. Wang, "Data mining curriculum: A proposal (version 1.0)," *Intensive Working Group of ACM SIGKDD Curriculum Committee*, 2006.

[3] M. F. Dixon, I. Halperin, and P. Bilokon, *Machine Learning in Finance. From Theory to Practice.* Springer International Publishing, 2020.

[4] K. Peffers, T. Tuunanen, M. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of Management Information Systems*, vol. 24, pp. 45–77, 2007.

[5] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. P. Reinartz, C. Shearer, and R. Wirth, "Crisp-dm 1.0: Step-by-step data mining guide," 2000.

[6] G. Bowen, "Document analysis as a qualitative research method," *Qualitative Research Journal*, vol. 9, pp. 27–40, 2009.

[7] M. S. Brown, "What it needs to know about the data mining process," 2015. Available: `https://www.forbes.com/sites/metabrown/2015/07/29/what-it-needs-to-know-about-the-data-mining-process/?sh=45fcbed6515f`, Accessed: 2022-05-07.

[8] S. Grimheden and J. Järlesäter, "Concretizing CRISP-DM for Data-Driven Financial Decision Support Tools," 2022. Student Paper.

[9] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying crisp-dm process model," *Procedia Computer Science*, vol. 181, pp. 526–534, 2021.

[10] S. Emerson, R. Kennedy, L. O'Shea, and J. O'Brien, "Trends and applications of machine learning in quantitative finance," in *8th international conference on economics and finance research (ICEFR 2019)*, 2019.

[11] J. Huang, J. Chai, and S. Cho, "Deep learning in finance and banking: A literature review and classification.," *Frontiers of Business Research in China*, vol. 14, no. 1, pp. 1 – 24, 2020.

[12] N. Goumatianos, I. T. Christou, P. Lindgren, and R. Prasad, "An algorithmic framework for frequent intraday pattern recognition and exploitation in forex market.," *Knowledge and Information Systems*, vol. 53, no. 3, pp. 767 – 804, 2017.

[13] Z. Soltani, B. Zareie, F. S. Milani, and N. J. Navimipour, "The impact of the customer relationship management on the organization performance," *Journal of High Technology Management Research*, vol. 29, no. 2, pp. 237–246, 2018.

[14] H. Hallikainen, E. Savimäki, and T. Laukkanen, "Fostering b2b sales with customer big data analytics," *Industrial Marketing Management*, vol. 86, pp. 90–98, 2020.

[15] N. Singh, P. Sing, and M. Gupta, "An inclusive survey on machine learning for crm: a paradigm shift," *Decision (0304-0941)*, vol. 47, no. 4, pp. 447–457, 2020.

[16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[17] A. Sherstinsky, "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020.

[18] H. Hewamalage, C. Bergmeir, and K. Bandara, "Recurrent neural networks for time series forecasting: Current status and future directions," *International Journal of Forecasting*, vol. 37, no. 1, pp. 388–427, 2021.

[19] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-beats: Neural basis expansion analysis for interpretable time series forecasting," *arXiv:1905.10437*, 2019.

[20] B. N. Oreshkin, G. Dudek, P. Pełka, and E. Turkina, "N-beats neural network for mid-term electricity load forecasting," *Applied Energy*, vol. 293, p. 116918, 2021.

[21] A. Lantz, *Intervjumetodik*. Studentlitteratur, 2013.

[22] M. Höst, B. Regnell, and P. Runeson, *Att genomföra examensarbete*. Studentlitteratur, 2006.

[23] D. L. Morgan, *Focus groups as qualitative research*, vol. 16. Sage publications, 1996.

[24] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[25] V. Clarke and V. Braun, "Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning," *The Psychologist*, vol. 26, pp. 120–123, 2013.

[26] C. Chatfield, *The analysis of time series: an introduction*. Chapman and hall/CRC, 2003.

[27] J. Brownlee, *Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python*. Machine Learning Mastery, 2020.

[28] Danmarks nationalbank, "Denmarks's fixed exchange rate policy." Available: `https://www.nationalbanken.dk/en/about_danmarks_nationalbank/frequently_asked_questions/Pages/Denmarks-fixed-exchange-rate-policy.aspx` Accessed: 2022-04-04.

[29] H. Kaur, H. S. Pannu, and A. K. Malhi, "A systematic review on imbalanced data challenges in machine learning: Applications and solutions," *ACM Comput. Surv.*, vol. 52, no. 4, 2019.

[30] C. M. Bishop *et al.*, *Neural networks for pattern recognition.* Oxford university press, 1995.

[31] Y. Xu and R. Goodacre, "On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning," *Journal of Analysis and Testing*, vol. 2, 2018.

[32] D. Christie and S. P. Neill, "8.09 - measuring and observing the ocean renewable energy resource," in *Comprehensive Renewable Energy (Second Edition)* (T. M. Letcher, ed.), pp. 149–175, Oxford: Elsevier, second edition ed., 2022.

[33] D. Olsen, *The lean product playbook: How to innovate with minimum viable products and rapid customer feedback.* John Wiley & Sons, 2015.

[34] S. Lauesen, *User interface design : a software engineering perspective.* Pearson/Addison-Wesley, 2005.

[35] R. Sebastian and M. Vahid, *Python Machine Learning : Machine Learning and Deep Learning with Python, Scikit-learn, and TensorFlow 2, 3rd Edition.* Packt Publishing, 2019.

[36] Run:AI, "Deep learning with gpus." Available: `https://www.run.ai/guides/gpu-deep-learning` Accessed: 2022-04-27.

[37] J. Herzen, F. Lässig, S. G. Piazzetta, T. Neuer, L. Tafti, G. Raille, T. V. Pottelbergh, M. Pasieka, A. Skrodzki, N. Huguenin, M. Dumonal, J. Kościsz, D. Bader, F. Gusset, M. Benheddi, C. Williamson, M. Kosinski, M. Petrik, and G. Grosch, "Darts: User-friendly modern machine learning for time series," *arXiv:2110.03224*, 2021.

[38] R. Kumar, P. Kumar, and Y. Kumar, "Analysis of financial time series forecasting using deep learning model," in *2021 11th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pp. 877–881, 2021.

[39] S. Kaushik, A. Choudhury, N. Dasgupta, S. Natarajan, L. A. Pickett, and V. Dutt, "Using lstms for predicting patient's expenditure on medications," in *2017 International Conference on Machine Learning and Data Science (MLDS)*, pp. 120–127, 2017.

[40] S. Ribeiro and R. Alquézar, "Local maximum ozone concentration prediction using lstm recurrent neural networks," 2002.

[41] H. Abbasimehr, M. Shabani, and M. Yousefi, "An optimized model using lstm network for demand forecasting," *Computers & Industrial Engineering*, vol. 143, p. 106435, 2020.

[42] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004.

[43] I. Hylander, *Fokusgrupper som kvalitativ datainsamlingsmetod.* Linköping University Electronic Press, 1998.

[44] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization.," *Journal of Machine Learning Research*, vol. 13, no. 2, pp. 281 – 305, 2012.

[45] M. Höst, "Edaa35 utvärdering av programvarusystem," 2022.

Appendices

Appendix A
Questions for I1

The following questions were used as a checklist of questions to be answered during the interview set I1. The goal with the questions were to get more understating of the business domain and how the sales team work.

- What's your name?
- How long have you worked in your current position at the bank?
- What are your work tasks?
- How are the sales team working with customers?
  - Is there a difference in your approach towards old and new customers?
  - What kind of recommendation and/or guidance are you giving the customers today?
- We know previously that your team is interested in passive FX transactions. What would you say is the goal of the analysis of this data?
  - Is there any particular patterns or information you think would be interesting to look at?
    * Size of trades? Number of trades?
    * Size of trades in relation to company size?
    * Does it matter what system they are using to trade?
    * Does product type matter?
  - Are you using this specific data for something today?

Appendix B
Questions for I2

The following questions were used as a checklist of questions to be answered during the interview set I2. The goal with the questions were to get to know how the sales team want a lead to be presented and what information they need in the presentation to work with the lead.

- What's your name?
- How do you work with leads today?
  - Which systems do you use?
  - When do you act on leads?
- How do you get the leads presented today?
  - Does it differ depending on the lead?
- How would you like the passive FX transactions lead to be presented?
  - In what system?
  - Does that change depending on how many leads per month?
- What information would you like to have on the lead?
  - What should describe the lead?

Appendix C
Questions for focus group

The following questions were used for the focus groups to evaluate the mock-ups of the lead. The questions were used as a checklist of points we wanted to get answered during the discussions.

- What are your initial thoughts?
- Would you act on any of the three leads? (Why or why not?)
- Is there any information that is missing? (is it crucial or just 'nice to have') Or any unnecessary information?
- Do you think the information is easy to find and read on the sheet? Anything you don't understand?
- Any improvements we should make that haven't already been discussed?

EXAMENSARBETE Designing a Machine Learning Application to Obtain Customer Insights in the Banking Domain
STUDENTER Noah Mayerhofer, Sandra Nyström
HANDLEDARE Elizabeth Bjarnason (LTH)
EXAMINATOR Emelie Engström (LTH)

Maskininlärning för att förbättra kundrelationer

POPULÄRVETENSKAPLIG SAMMANFATTNING **Noah Mayerhofer, Sandra Nyström**

Kundrelationshantering är superviktigt för att företag ska öka sin förståelse om sina kunder och således öka sina vinster. Detta arbete undersökte hur maskininlärning kan användas för att skapa nya insikter med hjälp av en banks stora mängder kunddata.

Alla företag behöver förstå sina kunder och ha en bra strategi för kundrelationshantering för att kunna erbjuda bra produkter och tjänster. I dagens samhälle är det väldigt vanligt att företag lagrar stora mängder data om sina kunder och användare. Datan som lagras berör ofta användarens interaktion med digitala system, så som hur länge en person tittade på en annons och vilka köp en person gjort. Ett företag kan välja att analysera denna data för att skaffa sig en djupare förståelse om sina kunder. Att analysera kunddatan är dock inget jobb för människor utan görs med hjälp av olika modeller som beräknas av datorer. En kategori av dessa modeller är maskininlärningsmodeller vilka lär sig av tidigare data för att sedan kunna göra någon form av gissning på ny data.

I vårt examensarbete undersökte vi processen av att utveckla en applikation baserad på maskininlärning som skapar insikter från kunddata. Vi gjorde vårt arbete tillsammans med en storbank och använde företagskundernas valutatransaktioner som data. Detta resulterade i en applikation som kollade på säsongsvariationer, skillnader över året, i datan och förutspådde hur kunden skulle agera under de kommande halvåret. Maskininlärningsmodellerna visade sig kunna förutspå vissa av kundernas beteende, särskilt kunder som hade ett väldigt periodiskt mönster. Kunder som hade mindre förutsägbart beteende fick vi blandande resultat för, några gick att förutspå bra och för en del gick det sämre. Även om modellerna inte kunde förutspå perfekt tyckte banken att de fick mer kundförståelse än tidigare och att de skulle kunna erbjuda bättre tjänster för sina kunder.

Vi följde CRISP-DM metodologin genom vårt arbete eftersom den är industristandard för projekt inom informationsutvinning. CRISP-DM är dock sparsamt använd inom forskning i finansvärlden, så vi bidrar med ett praktiskt exempel till vetenskapen genom att testa denna metodologi. Vi använde oss av olika datainsamlingsmetoder, såsom intervjuer och fokusgrupper.

Vi gjorde även en analys på utmaningar vi stött på under arbetets gång för att framtida arbeten inom samma sektor ska kunna dra lärdom av vårt arbete och på så sätt underlätta framtida arbeten. Den största utmaningen vi stötte på var att arbeta med stora mängder data vilket tog väldigt lång tid. Det tog exempelvis tre veckor för en dator som jobbade 24-timmar om dygnet att skapa maskininlärningsmodellen.

Vi har sett att det finns goda möjligheter för maskininlärning i praktiska projekt för CRM inom finansvärlden, men att det fortsatt finns ett behov av forskning på området. Vi hoppas framtida arbeten kan dra nytta av kunskapen i vår rapport!