Aalto University
School of Science
Degree Programme in Industrial Engineering and Management

Michael Wallin

Developing an analytics-driven sales process: a case study in the field of corporate banking

Master's Thesis

Espoo, 24.6.2015

Supervisor: Professor Henri Schildt

Instructor: M.Sc. Pekka Vimpari

AALTO UNIVERSITY SCHOOL OF SCIENCE Degree Programme in Industrial Engineering and Management	ABSTRACT OF THE MASTER'S THESIS

Author: Michael Wallin

Subject of the thesis: Developing an analytics-driven sales process: a case study in the field of corporate banking

Number of pages: 63 + 25	Date: 24.6.2015	Library location: TU

Professorship: Strategic management	Major subject code: TU-91

Supervisor: Professor Henri Schildt

Instructor: M.Sc. Pekka Vimpari

To deal with an increase in the quantity and availability of data, companies are coming up with new ways to support their decisions and processes with data. This master's thesis describes the design of a data-driven sales process for the acquisition of new corporate customers for Aktia, a medium-sized Finnish bank. The purpose of this redesign is to move from an ad hoc, relationship-based method of customer acquisition to a more systematic one, where data is utilized throughout the process.

The sales process design is approached from both an empirical and a qualitative perspective. From the empirical perspective, the main deliverable is a statistical scoring model for the selection of new corporate customers. From a qualitative perspective, the thesis presents a process for deploying the customer selection suggested by the developed scoring model. Additionally, the thesis outlines a plan for establishing an iteratively improved process for the acquisition of new corporate customers.

In terms of empirical findings, our modelling work brought to light several statistically significant predictors of customer quality. Among the most important predictors were variables relating to financial health, such as the relative indebtedness, and variables relating to profitability, such as the return on assets and the return on investment. In terms of methodological results, we found decision tree-based models to be the strongest-performing class of models.

A successful implementation of the proposed analytics-based sales process has the potential to generate significant business gains for Aktia. Compared to the current, limited customer selection process, a formal customer selection model can lead to a significant improvement in the quality and profit potential of acquired customers. Additionally, by introducing techniques for analyzing and improving sales efforts, Aktia can improve its customer acquisition and retention rates. Concretely, these retention gains lead to more productive sales efforts and an increased rate of new customer acquisition, both of which match Aktia's strategic growth targets.

Keywords: Credit scoring, customer selection, corporate banking, sales process development	Publishing language: English

AALTO-UNIVERSITETETS Högskolan för Teknikvetenskaper Utbildningsprogrammet för produktionsekonomi	REFERATET AV DIPLOMARBETET	
Författare: Michael Wallin		
Diplomarbetets ämne: Utvecklandet av en data-driven försäljningsprocess		
Sidoantal: 63+25	Datum: 24.6.2015	Placering i biblioteket: TU
Professur: Strategisk företagsledning		Professurskod: TU-91
Arbetets övervakare: Professor Henri Schildt		
Arbetets handledare: M.Sc. Pekka Vimpari		

För att handskas med den alltjämt växande mängden data som samlas och lagras, utvecklar företag fortsättningsvis nya sätt att stöda sina beslut och processer med data. Detta diplomarbete beskriver planerandet och implementeringen av en ny data-driven försäljningsprocess för anskaffandet av nya företagskunder för den medelstora finländska banken Aktia. Syftet med omorganiseringen av försäljningsprocessen är att övergå från en ostrukturerad metod för nykundsanskaffningen till en mer systematisk process där data används i processens varje steg.

Konstruktionen av försäljningsprocessen behandlas från både ett empiriskt och kvalitativt perspektiv. Ur ett empiriskt perspektiv utvecklar diplomarbetet en statistisk kund-scoring-modell som kan användas för att välja kunder med hög kvalité och lönsamhet. Kvalitativt presenterar diplomarbetet en process för att genomföra kundvalet baserat på den statistiska modellen. Ytterligare presenteras en plan för etableringen av en iterativt uppdaterad process för anskaffning av nya företagskunder.

Modellen identifierade flera statistiskt signifikanta prediktorer av kvalitén av en företagskund. Faktorer som beskriver företagets finansiella hälsa var bland de mest betydande. Också faktorer som relaterar till företagets lönsamhet kom fram som statistiskt betydande. Av de testade modellklasserna, presterade beslutsträdsmodeller det bästa resultatet.

En lyckad implementering av den föreslagna data-baserade försäljningsprocessen kan leda till betydande fördelar för Aktia. Jämfört med den nuvarande, något begränsade kundanskaffningsmetoden, kan en mer systematisk kundvalsprocess leda till en avsevärd ökning i kvalitén och lönsamheten av nya kunder. Genom att introducera metoder för analyserandet och förbättrandet av försäljningsinsatser, kan Aktia också effektivera sin nykundsanskaffning. Dessa effektiveringar leder till mer produktiva försäljningsinsatser och en ökad träffsäkerhet inom nykundsanskaffningen, som är i linje med Aktias strategiska tillväxtmål.

Nyckelord: Kreditscoring, kundval, företagsfinansiering, försäljningsprocesser	Publiceringsspråk: Engelska

Acknowledgements

Contents

1 Introduction

With the recent increase in affordable computational power, and the quicker dissemination of data across a number of different devices and sensors, we are seeing rapid development in data storage and analysis techniques. To deal with this development, companies are coming up with new ways to support their decisions and processes with data. This master's thesis describes the design of a data-driven sales process for the acquisition of new corporate customers for Aktia, a medium-sized Finnish bank. The purpose of this redesign is to move from an ad hoc, relationship-based method of customer acquisition to a more systematic one, where data is utilized throughout the process, from the initial stages of customer selection and sales resource allocation to the later stages of evaluating the retention of contacted firms.

1.1 The underlying business problem

In the current environment of persistently low interest rates, the profit margins from retail lending are slim. In comparison, corporate lending offers higher loan margins at the expense of higher risk. At the same time, some of the larger Finnish lenders seem to be focusing their corporate banking efforts on larger firms, and seem to be willing to let go of some of the small- and medium-sized firms (SME) in their portfolio. These SMEs fit well into Aktia's offering to corporate customers, and they are the types of firms that Aktia is looking to attract in the future.

The key challenge in expanding Aktia's corporate customer portfolio is the current lack of a data-based process for the acquisition of new customers. As of now, new corporate customers are mostly acquired through marketing efforts, through the existing networks of sellers or by offering corporate banking services to existing retail customers of the bank. A relatively small sales force is also a limiting factor in further developing the corporate portfolio of the bank, as most of the time of the personnel of the corporate bank is spent on managing the existing customer base instead of acquiring new customers. This further enhances the need for a data-driven approach to sharpen the focus of customer acquisition activities.

1.2 Research questions

To deal with the business problem, this thesis proposes a data-driven approach to customer selection, sales resource allocation and the continuous monitoring of sales performance. This leads to a natural division of the research problem into three research questions:

1. How can Aktia use data to select the corporate customers with the highest profit potential?

2. How should Aktia allocate its sales resources to match the selections proposed by the customer selection model?

3. How can Aktia transform the data-driven customer selection routine into an iterative, continuously monitored sales process?

Of the three research questions, the first defines the empirical focus of the thesis. The main deliverable of the project is a statistical customer selection model, and a significant proportion of the thesis is dedicated to discussing the development and validation of this model. The second and third research questions focus on how the theoretical modelling efforts should best be utilized in a business context. Concretely, the second research question investigates the practical implementation and deployment of the customer selection proposed by the constructed model. The third research question moves beyond customer selection to the more ambitious objective of a comprehensive data-driven sales process. Such a process requires efforts to monitor the success of sales activities, and to continuously update the customer selection model and sales methods accordingly.

1.3   Structure of the thesis

The thesis is structured into three main parts. First, a description of the underlying business problem establishes the background of the thesis (section 1), a general overview of Aktia and its operating environment (section 2), and a review of the relevant academic literature (section 3). The second part of the thesis covers the empirical parts of the work, through a discussion of the development, validation and interpretation of the customer scoring model. The empirical section begins with a discussion of the methods used (section 4) and ends with an inspection of results (section 5). The final part of the thesis discusses the implementation of the analytics-driven sales process, and evaluates the findings of the thesis work (sections 6 and 7).

In terms of the three research questions presented in 1.2, the first, i.e. the customer selection problem, is mainly discussed in sections 4 and 5. Based on the proposed model for customer selection, the deployment of the proposed sales process is discussed in section 6.2. The final research question, which pertains to the continuous use of a full-fledged analytics-based sales process, is discussed in section 6.3.

2   Aktia's operating environment

In this section, we present a general overview of Aktia's operating environment. First, we look at the general composition of the Finnish banking sector. Then, we investigate the state of corporate banking and financing in Finland. Finally, we present an overview of Aktia's background, its activities and its competitive position in the market.

2.1   Banking in Finland - a general overview

In the following table, the market shares of the largest Finnish banks are presented.

Table 1: Market shares of Finnish banks as of 31.12.2014

Institution	Loans, market share (%)	Deposits, market share (%)
OP	34.2	36.4
Nordea	28.8	28.7
Danske Bank	9.8	11.9
Handelsbanken	5.8	3.4
Aktia	3.1	3.1
Säästöpankkiryhmä	2.7	4.4
POP	1.8	3.1
Ålandsbanken	1.1	1.1
Hypo	0.6	0.3
Others	12.0	7.7

Source: Finanssialan Keskusliitto (2014)

As can be seen in Table 1, the Finnish banking industry is an oligopoly, with OP and Nordea being the two clearly largest banks. In third and fourth place are the Nordic financial conglomerates Danske Bank and Handelsbanken. The remaining banks are smaller, local players, with Aktia being the largest of the small banks by loan market share.

A recent newcomer to the Finnish banking sector is S-Pankki, which offers banking services at the supermarkets of the Finnish retailing co-operative S-Ryhmä. In Table 1, S-Pankki is covered under the "Others" category, but exact estimates of S-Pankki's market share are difficult to find. As all S-Ryhmä loyalty customers automatically become customers of S-Pankki, the bank has several million customers. Of these, an estimated 1.3 million are active customers, and 200 000 have directed their monthly salary to an S-Pankki account. Recently, S-Pankki has also acquired the banking arm of the Finnish insurance company Lähi-Tapiola and the entirety of

the investment firm FIM. (Niemeläinen, 2014) All in all, these observations suggest that S-Pankki's market share should be somewhere between those of Aktia and Danske Bank.

After examining the firms in the Finnish banking sector, we can investigate the distribution of the customers of Finnish banks. In the following chart, the composition of the loan portfolio of Finnish lending institutions is depicted.
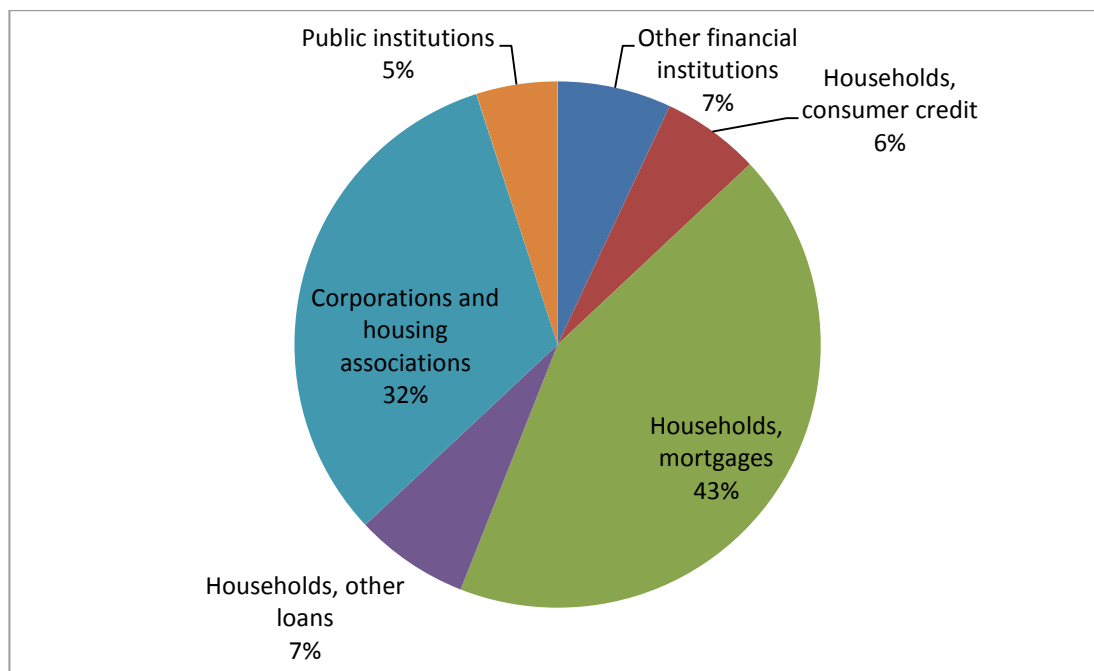
Source: Finanssialan Keskusliitto (2015a)

Figure 1: Distribution of the loans of Finnish financial institutions, as of 31.10.2013

As can be seen in Figure 1, households are the most significant lending segment and account for 56% of all lending. Corporations hold 32% of loans and other financial and public institutions stand for the remaining 12% of lending.

On an international level, the Finnish banking sector performs quite well, particularly in terms of solvency. The Tier 1 solvency of Finnish banks in 2013 was among the best in the Eurozone, at roughly 15%. This was almost at the same level as Germany (15.5%), and clearly above Austria (13.5%), the Netherlands (12.5%), Italy (11%) and Spain (10.8%). The share of non-performing assets in the Finnish loan portfolio has also remained at close to 0% throughout the post-millennial period, which is a sign of healthy balance sheets. (Finanssialan Keskusliitto, 2015b) Additionally, all of the three Finnish banks (Nordea, OP and Danske) included in ECB's stress test passed the test comfortably (Turtola, 2014). In terms of profitability, Finnish banks also perform quite well by European standards. In 2013, the Return on Equity of Finnish banks was among the highest in Europe. At 10%, the RoE exceeded that of both France (8%) and

Germany (7.5%). Still, Swedish banks clearly outperformed Finnish banks with a RoE of roughly 16%. (Finanssialan Keskusliitto, 2015b)

2.2 Corporate banking and financing in Finland

There are two main sources of debt-based financing for corporations: bank loans and bonds. In Finland, and in Europe in general, the emphasis on bank loans has been relatively significant. In 2005, the split between loans and bonds for the Eurozone was 89%-11% (the Finnish split was 86%-14%). This contrasts with the United States, where the split was much more even, at 61%-39%. (Mattila, 2013) In general, capital markets are much more developed in the United States than in Europe. According to a report by the European Commission (2015), medium-sized U.S. firms get five times more capital markets financing than EU firms(European Commission, 2015).

For a glimpse into the distribution of Finnish loans, the following chart displays the industry composition of Finnish corporate bank loans.
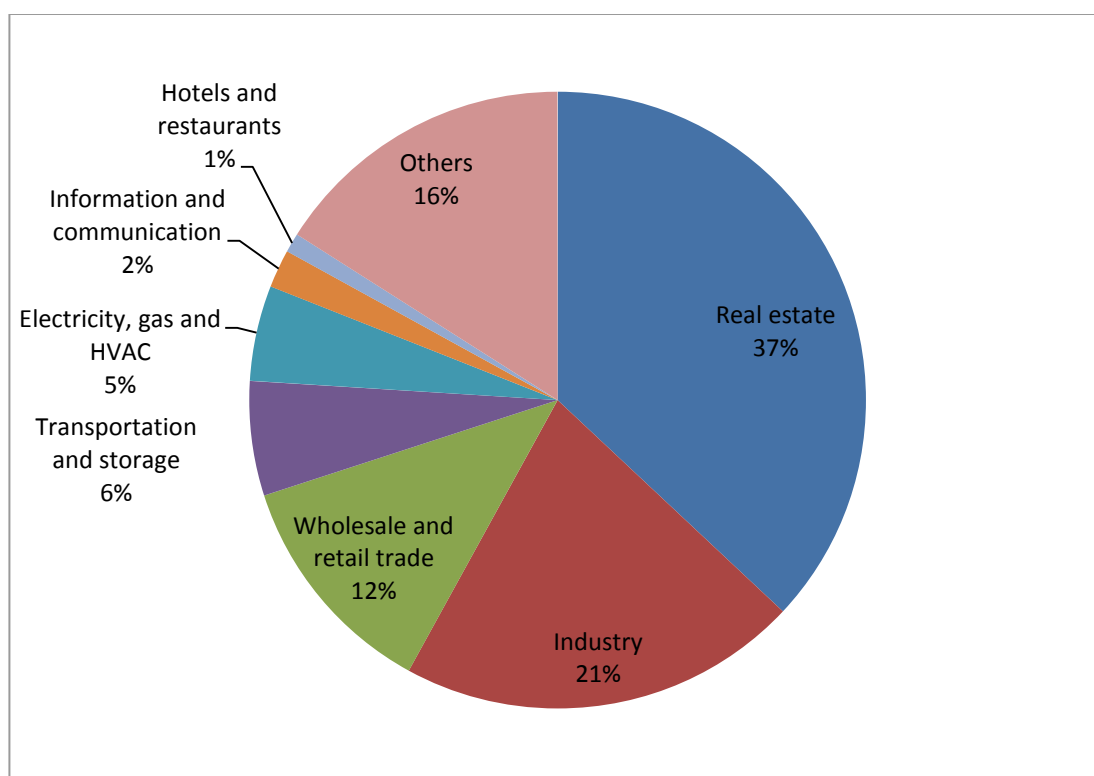
Source: Finanssialan Keskusliitto (2015c)

Figure 2: Finnish corporate loans by industry sector (as of 30.9.2012)

As can be seen in Figure 2, real estate is the single most important sector of industry for corporate lending, with 37% of all loans. Loans to industrial firms and to wholesale and retail firms also exceed 10% of the total loan portfolio. For other sectors, the share is less than 10%.

As previously mentioned, a defining trait of both Finnish and European corporate financing is the relatively strong emphasis on bank loans. Recently, this has begun to change. After the global financial crisis of 2008, a stricter regulatory environment and a more challenging financing environment has hurt the availability of bank loans. To adjust to this change, corporate financing through bonds has increased at the expense of bank loans. In the first quarter of 2013, the number of newly issued bonds in Europe exceeded the number of new bank loans, and Finland's financing mix is also shifting in this direction. For SMEs, bank loans remain the only realistic source of new financing, as issuing bonds is rarely an option for these firms. As larger firms exceedingly move towards bond-based financing, banks can commit more of their balance sheets to funding SMEs (Pylkkönen and Savolainen, 2013). (Mattila, 2013)

While the focus of corporate debt financing has moved towards bonds, the growth in Finnish corporate loans has been strong throughout the last decade. The yearly growth in Finnish corporate loans peaked at roughly 10% halfway through 2011, but has remained clearly positive since then. This contrasts with other European nations, where corporate loan growth has generally been stagnant or negative. In terms of loan margins, Finland is on the lower end of the spectrum for Eurozone nations. In January 2015, the average interest rate on corporate loan contracts in the Eurozone was 2.41%. For Finland, the interest rate was 2.05%. Of other Eurozone countries, Germany, Netherlands, Austria and France had lower interest rates (2.03%, 1.87%, 1.83% and 1.79%, respectively). (Finanssialan Keskusliitto, 2015b)

2.3 Aktia Bank plc

Aktia Bank plc is a Finnish financial services company that offers a wide variety of services in banking, asset management, insurance and real estate. In its current form, the company was formed in 1993, when Helsingfors Sparbank and several other savings banks on the Finnish coast merged to form Aktia Sparbank. Today, Aktia Group serves some 350 000 customers through 50 local branches, as well as by phone and through digital channels. Aktia's key operating areas are the Finnish coastal regions from Loviisa to Oulu and the mainland growth centers of Helsinki, Tampere and Turku. Aktia's vision is to be the best financial adviser for families and their companies. (Aktia, 2015a)

In our general overview of the Finnish banking sector, we saw that Aktia is the fifth-largest bank by loan market share. In the following table, we have gathered some other key performance ratios of the largest Finnish banks to further investigate the competitive landscape.

Table 2: Key performance ratios of Finnish banks for 2014

Bank	RoE	Cost/profit-index	Tier 1 solvency
Nordea	9.40%	48.00%	14.20%
OP	8.10%	57.00%	15.50%
Danske	6.90%	62.00%	14.50%
Aktia	8.30%	71.00%	14.60%
Säästöpankki	5.80%	64.00%	16.90%
POP Pankki	n/a	72.00%	n/a
Ålandsbanken	8.70%	80.00%	10.90%
S-Pankki	4.20%	87.00%	14.30%
Hypo	6.70%	56.00%	n/a
Evli	15.20%	80.00%	n/a

Source: Finanssialan Keskusliitto (2015d)

As Table 2 shows, Aktia performs quite well in terms of profitability, as Aktia has the fourth-highest RoE after Nordea, Ålandsbanken and Evli. In terms of operative efficiency, Aktia doesn't perform quite as well, falling in the lower half of firms in terms of the cost/profit-index, with an index that is more than 20 percentage points higher than that of the strongest performer Nordea. In terms of solvency, Aktia's performance was average according to 2014 figures. In early 2015, the Finnish Financial Supervisory Authority granted Aktia and its subsidiaries permission to use the IRBA-method for computing capital requirements. With this method, Aktia's Tier 1 capital solvency ratio increases from 14.6% to 22.7%, which moves Aktia to the very top of Finnish banks in terms of solvency(Aktia, 2015b). One key factor that has made Aktia an attractive investment is the strong and stable dividend that the company has paid in recent years. In an investigation of stocks with high dividend yields, Jaakko Tyrväinen from Evli found Nordea and Aktia to be the two stocks on the Finnish stock exchange with the highest dividend yields (Tyrväinen, 2015).

For a more operations-oriented perspective into the competitive landscape of Finnish banks, the following chart displays the composition of profits of Finnish banks.
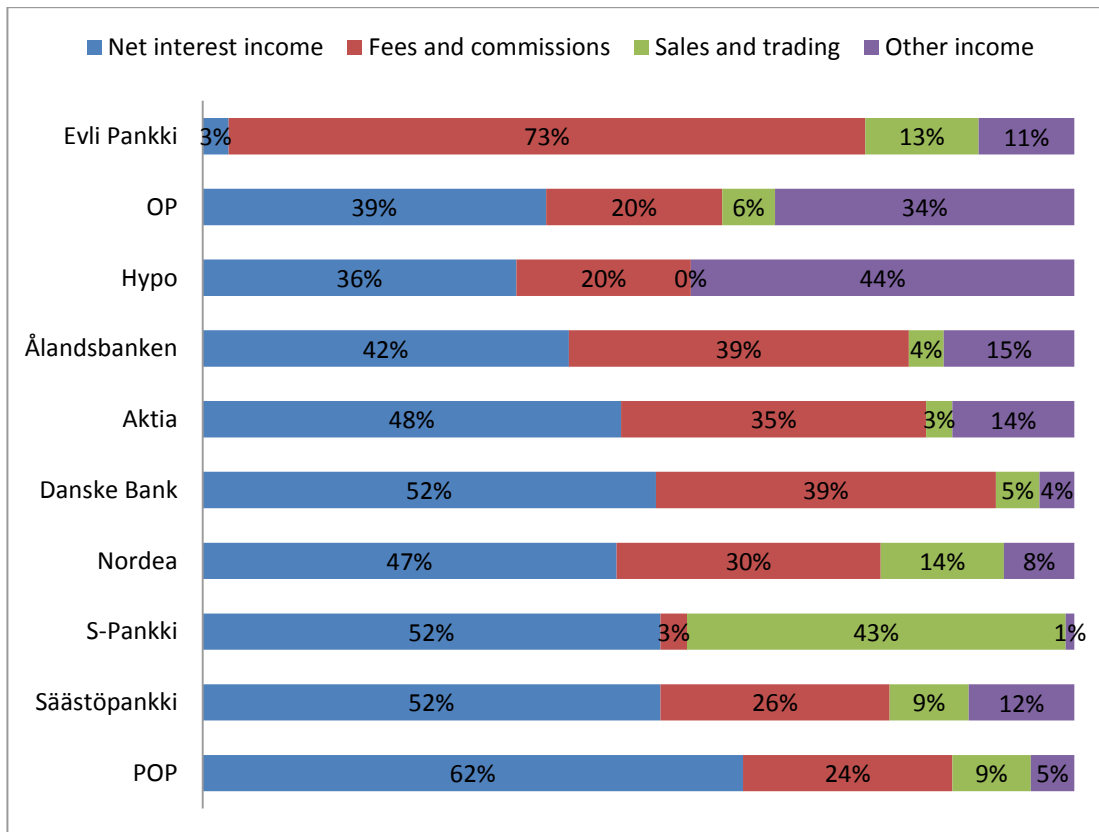
Source: Finanssialan Keskusliitto (2015a)

Figure 3: Distribution of profits of Finnish banks, as of 31.12.2014

As can be seen in Figure 3, Aktia's profit distribution is comparable to some of the similar smaller banks (e.g. Säästöpankki, Ålandsbanken), and also quite similar to the profit distribution of Nordea. The vast majority of Aktia's income comes from net interest income and fees, while sales and trading only accounts for 3% of the firm's income.

In terms of strategic aspirations, Aktia is currently looking to transition from a cost-cutting phase to a growth phase. After restructurings in 2012 and 2013, and a soon-to-be-finished revamp of its core banking IT infrastructure, the bank will aim for higher growth in the coming years, as well as lowered IT costs from its new core banking system. With the significant increase in Aktia's solvency ratio after the new IRBA capital requirements, Aktia can use more of its balance sheet to finance these growth efforts. A significant part of the growth needs to come from new customer acquisition, in both retail and corporate sectors.

3 Literature

In this overview, the literature pertaining to each of the three research questions is treated separately. For the first question, regarding customer selection, the overview is further divided into a quantitative and qualitative part. In the quantitative part, we discuss the academic literature that relates to quantitative credit scoring methods. In the second part, the relationship between banks and their customer firms is considered from a qualitative perspective through concepts such as relationship banking and bank switching. For the second research question, the literature overview covers research conducted in the field of analytics-based resource allocation. For the third research question, the literature overview first presents some key findings on how businesses can use data to improve their processes. After this, the review inspects the literature related to the development of data-supported sales processes and the emergence of new types of sales organizations.

3.1 Customer selection in corporate banking

3.1.1 Quantitative approaches

The act of quantitatively evaluating the quality of a bank customer, i.e. credit scoring, is a process where a financial institution tries to classify customers into 'good' and 'bad' borrowers on the basis of some data and some type of decision model. Commonly, the resulting credit score is a probability of the borrower in question defaulting on the loan. While the origins of formal credit scoring can be traced back to the early 1940s (Durand, 1941; Finlay, 2010, p. 528), the adoption of formalized credit scoring processes did not increase dramatically until the mid-1990s when consumer credit started growing dramatically (Hand and Henley, 1997). Recently, the development of credit scoring techniques has been further bolstered by the advance of statistical classification techniques and the increased availability of consumer data.

Various statistical methods have been used to score consumer credit. In earlier credit scoring models, linear models, such as logistic regression and linear discriminant analysis (LDA), were primarily used. More recently, non-linear models such as neural networks (NN), decision trees and support vector machines (SVM) have been investigated as potential replacements for linear models. In a comprehensive benchmarking of classification algorithms, Baesens et al (2003) compared the predictive performance of a wide variety of methods. They found that the non-linear SVM and NN models performed very well. Importantly, however, the previously favored LDA and logistic regression models also performed well, which suggests that the tested credit scoring data sets were only weakly non-linear (Baesens et al., 2003).

While linear models have been shown to be competitive with non-linear models in some experiments, the credit scoring research is increasingly shifting towards more advanced and complex non-linear models. As even incremental improvements in default prediction accuracy can lead to noteworthy monetary gains, slight improvements in accuracy from complex and advanced credit scoring models quickly become significant (Huang et al., 2007). The main drawback of complex, non-linear models is the lack of comprehensibility. For linear models, the coefficients of the predictor variables allow for the model to be easily understood, but for non-linear models, such as support vector machines and neural networks, there are no such comprehensible decision rules. In the financial industry, where regulations often tend to require transparency in credit decisions, the opaqueness of non-linear models is problematic. In the U.S., for example, the Equal Credit Opportunity Act requires a financial institution to provide specific reasons why a loan application was rejected (Martens et al., 2007, p. 2). To improve on the transparency without sacrificing the increased accuracy of advanced models, work has been done to derive explicit decision rules from non-linear models (Martens et al., 2007).

Compared to the scoring of consumer credit, research on credit scoring in the corporate sector is a bit more limited due to the lack of publicly available data, for privately owned firms in particular(Fernandes, 2005, p. 2). Nonetheless, significant work has been done in the field of corporate bankruptcy prediction. The earliest methods for bankruptcy prediction were univariate methods, where selected financial ratios were used as predictors of business failure (Beaver, 1966). Unsatisfied with using just a single financial ratio in bankruptcy modelling, Edward Altman introduced his own multiple discriminant analysis model, the Z-Score Model, in which the Z-Score is computed as a weighted sum of five different financial ratios (Altman, 1968). The original Z-Score model has, since its adoption, been updated several times, e.g. for the UK market (Taffler, 1984), and for private manufacturing and non-manufacturing firms (Altman, 1983). Additionally, simpler models with the same core idea as the Z-Score Model, such as Laitinen's Z-Score, have emerged after Altman's original model (Balance Consulting, n.d.). While there has been vast research on failure prediction, the Z-Score Model, in its original form or as one of the updated variants, remains the most widely used bankruptcy prediction model today. (Altman et al., 2014)

While the Z-Score model remains popular, some more complex bankruptcy prediction models have also been developed. Similar to methods for the scoring of consumer credit, logistic regression and neural networks have had some success in corporate bankruptcy prediction. In addition to accounting-based models (such as the Z-Score models), market-based models that

utilize option pricing theory have seen some popularity. In a comparison of the predictive accuracy of market-based models and accounting-based models, there was, however, little difference between the two model categories (Agarwal and Taffler, 2008). Most recently, hazard models, which utilize both accounting and market information, were found to be slightly superior to comparable models utilizing only accounting or market information (Bauer and Agarwal, 2014). (Altman et al., 2014)

The academic approaches to credit scoring are quite varied, but on the industry side, the FICO score has emerged as the most widely used credit score in lending decisions, with around 95% of the largest U.S. financial institutions being FICO clients, and a total of 100 Billion FICO scores having been sold since FICO's inception in 1956 (FICO, 2015a). The exact model behind the score is a trade secret, but according to FICO, the score is a combination of five, weighted, importance factors: Payment history (35%), amounts owed (30%), length of credit history (15%), new credit (10%) and types of credit used (10%). The resulting score ranges between 300 and 850, with a higher number indicating a higher level of creditworthiness, and a score of roughly 700 being a 'good' score. (myFICO, 2011)

It is noteworthy that a majority of the credit scoring approaches covered above treat the output as a binary variable where all evaluated cases are classified into either good or bad payers. As a result, any two instances that exceed a given threshold of distress will be treated as equally unwanted loan takers by the classification model. Consequently, the typical credit scoring approaches have not taken profitability into account. This is unfortunate, as the lender would optimally not only provide loans to reliable individuals and firms, but to profitable ones as well. A key challenge has been the difficulty of accessing profitability figures on the level of individual accounts and loans (Finlay, 2008, p. 922). To alleviate this, Finlay has introduced models which attempt to introduce profitability considerations into the loss function in a continuous fashion (Finlay, 2010, 2009, 2008). Recently, similar profit-based classification measures have been developed by Verbraken et al. (2014).

3.1.2 Qualitative considerations

Traditionally, one of the fundamental tasks of banks has been to mitigate informational asymmetries. Banks develop close relationships with borrowers, which over time allows them to monitor the business of the borrower, and helps in eliminating the informational asymmetry that originally existed between the bank and the borrower. In this process, the task of collecting information about the creditworthiness of potential borrowers is delegated from private investors to banks (Farinha and Santos, 2002, p. 1). This aspect of banking, where the primary purpose of a bank is to establish and cultivate relationships with its customers, is

commonly referred to as relationship banking. This contrasts with transaction-oriented banking, or trading, where the focus is on executing transactions rather than on forming information-intensive relationships with customers (Boot, 2000, p. 10).

For relationship banking to be useful for borrowers, they need to exhibit a certain level of informational opaqueness. Otherwise, if their creditworthiness was apparent to all market participants, they could gather the necessary funding through transactional lending, or other sources, such as bond issuance, or private equity funding (Berger and Udell, 2002). Hence, relationship banking largely revolves around lending to small businesses.

While small businesses benefit from relationship banking through the 'soft' information that banks gain in the relationship, prolonged relationship lending may eventually incur unwanted costs. If the relationship lender is small, it may not be able to meet the increasing credit needs of a growing company. Additionally, a relationship lender may 'informationally capture' the borrower and charge a higher interest rate (Gopalan et al., 2007, p. 1). These restrictions are the basis for the graduation hypotheses, which is a common research topic in the literature on bank switching behavior. According to this hypothesis, firms tend to 'graduate' from smaller banks to larger ones as their business grows.

In a study of roughly 30 000 U.S. commercial bank loans during 1990-2005, Gopalan et al. (2007) found fairly strong support for the graduation hypothesis. It was found that informationally opaque firms were less likely to switch banks. Surprisingly, the most transparent firms were also less likely to switch banks. For the firms that switched banks, the switch was directed from smaller banks to larger banks and from smaller bank markets to larger bank markets. It was also found that switching banks allowed for firms to obtain higher loan amounts, undertake higher capital expenditures, and increase their leverage after the switch. (Gopalan et al., 2007).

Ioannidou and Ongena (2010) came to a similar conclusion on the consequences of bank switching in their investigation of Bolivian loans during 1999-2003. It was found that firms were able to substantially lower their loan rate by switching. About a year and a half into the switch, however, banks started increasing the loan rate, to the point where the rate reached parity with the loan rates at the previous firm, about four years after the switch. (Ioannidou and Ongena, 2010)

In another study, Howorth et al. (2003) investigated factors associated with bank switching in the U.K. small firm sector. The main drivers in bank switching were found to be dissatisfaction with the provided service, and difficulties in obtaining finance, the latter of which supports the

graduation hypothesis. As a motivation for their study, Howorth et al. note the paradoxically large number of firms that have indicated a willingness to switch banks, compared to the small number of banks that actually end up making the switch, quoting a 1998 study by the Federation of Small Businesses which found that 4% of small businesses had switched banks while as many as 34% had considered switching (FSB, 1998).(Howorth et al., 2003)

In addition to bank switching, there are also firms that graduate from single- to multiple-bank lending relationships. In a study on Portuguese lending relationships during 1980-1996, Farinha and Santos (2002) found that firms with more growth opportunities, more bank debt and less liquidity were more likely to switch from a single bank to multiple banks (Farinha and Santos, 2002, p. 18). On the other hand, they also found that firms with low profitability and challenges in paying bank loans on time were more likely to switch to multiple-bank relationships. Consequences of the move to a multiple-bank relationship were an increase in the firm's reliance on trade credit and a reduction in the importance of the incumbent bank as a funding provider, but no significant improvements in firm performance were observed (Farinha and Santos, 2002, p. 18). Importantly, Farinha and Santos (2002) found the original relationship between the firm and the incumbent bank to be valuable. For one, well-performing firms were found to have longer exclusive relationships with their incumbent banks before switching compared to poorly performing firms. Even in the case of poorly performing firms, the initial exclusive relationship was found to be valuable, as borrowing from the incumbent bank usually continued even after the switch to multiple banks. Farinha and Santos (2002) concluded that the substitutability and value of bank-firm relationships depends on the duration of the relationship. (Farinha and Santos, 2002)

With the increased focus on quantitative, scoring-based approaches to lending (such as FICO scores), there is a worry that relationship lending to small businesses will get crowded out. As loan officers get replaced by credit scoring algorithms, the risk is that loans will mostly be provided to firms that are already successful and financially secure. For small businesses, credit scores may not be good enough, or difficult to calculate due to lacking publicly available information, which may leave small businesses largely without loan funding. (Smith, 2014) To measure the value of loan officers and their relationship lending, Wang (2014) performed a study based on Chinese lending data. He found that the added value from the soft information provided by loan officers clearly exceeded the average pay of the loan officers, and hence argued that loan officers are a valuable addition to the hard information provided by credit scoring algorithms (Wang, 2014).

3.2 Allocation of sales resources

In their investigation of the sales analytics process, Kawas et al. (2013) divided the process into two sub-problems: the *predictive* problem of evaluating potential selling opportunities, and the *prescriptive* problem of assigning sellers to selling opportunities (Kawas et al., 2013, p. 2). The previous section on customer selection and credit scoring methods covered the predictive problem; this section focuses on the prescriptive problem related to resource allocation.

Given a set of profit estimates for a selection of potential customers, the corresponding resource allocation problem can be expressed as an optimization problem where the objective function is the profit, and the constraints are determined by the practical limitations of the business. This optimization problem is usually solved through some method of mathematical programming, and has been a fairly active area of research in management science.

Zoltners and Sinha (1980) reviewed some of the proposed solution methods for the sales resource allocation problem. They formulated the mathematical problem around three concepts: sales resources, sales entities, and sales response functions. Sales resources are the decision variables that are being allocated, for example sales representatives, sales time, sales effort or sales budget shares. Sales entities are the groups to which the sales resources are assigned, such as sales districts, products or markets. The sales response function represents the tradeoffs that follow from different choices of resource allocations. A wide variety of different sales response functions have been suggested in the literature, including linear, S-shaped, logit, exponential, piece-wise and discrete functions. Additionally, a wide variety of solution approaches have been proposed, such as control theoretical approaches, probabilistic methods, dynamic programming and even heuristic methods. As their own contribution to the existing body of work, Zoltners and Sinha proposed several integer programming-based models with a discrete revenue response function. (Zoltners and Sinha, 1980) In a recent, practically deployed solution to the sales resource allocation problem, Kawas et al. (2013) used a linear programming approach where teams of sellers were assigned to sales opportunities. The constraints of the linear program included a cost constraint and a constraint on the extent of headcount variation (Kawas et al., 2013).

In addition to the problem of sales resource allocation, approaches have been proposed for optimizing the entirety of the sales deployment process. Drexl and Haase (1999) suggested a mixed integer programming approach for simultaneously solving four, interrelated sales force deployment problems: sales force sizing, salesman location, sales territory alignment and sales resource allocation. They also developed a fast approximate solution method, which they

showed to be no further than 3% from the actual optimum (Drexl and Haase, 1999). Skiera and Albers (2008) suggested another, equally holistic approach to the sales force decision problem by estimating a core sales response function that allows managers to detect where there is most room for profitability improvements. The sales response function estimates the sales potential on both the individual level and the organizational level, and allows for managers to evaluate where the organization is lagging behind its potential. This then allows the organization to adapt by adjusting the sales force size or by providing more motivating compensation schemes for individual sellers. (Skiera and Albers, 2008)

3.3 A data-driven sales process

3.3.1 The state of the art of business analytics

There is considerable evidence that analytics is becoming an increasingly important function of modern businesses. In an international, cross-industry survey of 3000 executives, managers and analysts conducted by MIT Sloan and the IBM Institute for Business Value, it was found that half of the respondents agreed that the 'improvement of information and analytics' was a top priority in their organization. More than one fifth of respondents also said that they were under significant pressure to 'adopt advanced information and analytics approaches'. Moreover, top-performing organizations had a much higher tendency to use analytics than intuition compared to low-performing organizations. (LaValle et al., 2011, p. 22) In another study conducted by Bloomberg Businessweek in 2011, it was found that 97% of companies with revenues exceeding $100 million used some form of business analytics, compared to a figure of 90% from two years earlier (Businessweek, 2011). As a third data point, a study performed by SAS Institute and MIT Sloan, which covered around 2500 respondents from roughly 25 industries, found that 67% of respondents felt that they are gaining a competitive edge from their use of analytics (Kiron et al., 2013, p. 2).

While the potential benefits of business analytics are generally accepted, there exists a large gap between strong and weak performers in the utilization of analytics. Kiron et al. (2013) found that 11% of surveyed firms could be classified as exceptional performers in analytics (coined analytical innovators), 60% could be classified as moderately strong performers (analytical practitioners), while 29% of surveyed firms were still struggling to utilize data beyond basic reporting and marketing applications (Kiron et al., 2013, p. 7). LaValle et al. (2011) performed a similar classification of surveyed firms into three 'levels of capabilities' (Aspirational, Experienced and Transformed users of analytics), and found that Transformed

organizations were three times more likely to substantially outperform their industry peers than Aspirational firms (LaValle et al., 2011, p. 23).

Much work has been done to understand what differentiates exceptional analytics performers from average and poor ones. One characteristic that is frequently mentioned as a distinguishing feature of strong analytics performers is the ability and willingness to utilize real-time data in everyday decisions. Good practitioners use data to evaluate previous decisions, to automate operations and to report on performance, but excellent practitioners use real-time data to drive their day-to-day decision making and even their innovation activities. As Kiron et al. (2013, p.7) put it, exceptional performers view "data as a core asset" and for these companies "analytical insights are part of the culture of the organization and are utilized in strategic decisions, both large and small". (Kiron et al., 2013; LaValle et al., 2011)

Exceptional analytics performers are also able to extract more insights from their data. While regular performers monitor past performance through monitoring realized KPIs, excellent performers use predictive statistical techniques to forecast future performance (Kiron et al., 2013, p. 16). In addition to predictive techniques, exceptional analytics performers gain mileage from their data by using data visualization techniques (such as interactive dashboards), as well as simulations and scenario development (LaValle et al., 2011, pp. 26–27). Importantly, exceptional analytics performers also tend to identify the business challenges that can lead to the most significant gains for the company and focus on those challenges. Additionally, exceptional performers start the analytics process from questions, and then figure out what data and processing to use to answer those questions. Here, weaker performers do worse due to a lack of focus in choosing analysis tasks and a bottom-up tendency to build their analytics processes to fit their data without having clear business objectives in mind (LaValle et al., 2011, p. 25).

Decisions related to the organization of analytics activities are another important differentiator between exceptional and average performers. LaValle et al. (2011) found that a centralized analytics unit offers superior performance by allowing advanced skills to come together and by allowing companywide standards and best practices to form (LaValle et al., 2011, p. 28). Kiron et al. (2013) also found that a fragmented analytics ecosystem was a key factor in holding average analytics practitioners back from greatness. Despite problems with fragmentation, Kiron et al. (2013) urged companies to bring access to analytics to all levels of the business, but they emphasized the importance of an integrated approach over a fragmented one (Kiron et al., 2013, p. 15).

As a final differentiating factor between average and exceptional analytics performers, Kiron et al. (2013) mention the importance of revision and innovation in maintaining a competitive edge. After a brief period of success from an innovative analytics approach, other competitors will observe the success and start developing competing approaches. As their approaches improve, the company's competitive edge will diminish. As an example, Kiron et al. (2013) presents the story of the Oakland Athletics baseball team. In 2002, the team's manager Billy Beane was able to build a playoff-caliber team using advanced analytics despite facing the most significant salary constraints in the league. After this, other teams adopted similar data-driven approaches and eliminated the team's competitive edge. Only in 2012, after the Athletics managers invented new and effective analytics metrics, were the Athletics able to return to the playoffs (Kiron et al., 2013, p. 6).

While the success factors of strong analytics performers are fairly well established, there are several roadblocks which prevent average performers from becoming great. According to a 930-firm study conducted by Businessweek (2011), the number one analytics challenge for companies is the ability to collect and store reliable and timely data. Data accuracy, data consistency and even data access still challenge many companies (Businessweek, 2011, p. 2). Many firms are stuck with batch processing of sales information, and moving to real-time processing would require sweeping changes to existing systems. These changes would require significant costs in terms of money, time and potential disturbances to ongoing operations (Ferguson, 2013). In addition to data collection and storage, a significant challenge was found to be a lack of analytics talent. Without proper talent, companies struggle to process their data into results. (Businessweek, 2011, p. 2).

Perhaps the most crucial challenge for aspiring users of business analytics lies in the company culture. Intuition is still valued highly in the decision processes of many firms and data-driven approaches may struggle to catch on (Businessweek, 2011, p. 2). For older firms, it is difficult to compete with firms that have engrained analytics into their business from their very inception, such as most modern web companies. Still, there are encouraging examples of older, non-digital firms that have transformed into strong analytics performers. Kiron et al. (2013) present the case of Oberweis Dairy, which has its background in door-to-door-delivery of milk bottles to customers. By adding analytics talent from outside, the company was able to develop a competitive edge through data-based customer segmentation and targeting. (Kiron et al., 2013, pp. 4–5).

3.3.2 Towards a data-driven sales process

Cespedes (2014) identifies three factors that drive the productivity of a sales model: customer call capacity, close rate, and profit per sale (Cespedes, 2014, p. 2). Data has been used to improve all three of these factors with varying degrees of success. Of the three factors, the first is seemingly the most problematic source for productivity gains, as immediate improvements will require overworking the current sales force or hiring new personnel. Regardless, one of the proposed benefits of salesforce automation systems has been to increase call capacity by automating administrative activities. However, in a study on the impact of salesforce automation systems in the pharmaceutical industry, Eggert and Serdaroglu (2011) found that the cost-cutting qualities of salesforce automation systems do not have a direct effect on performance. The administrative qualities only improved efficiency when sellers were able to use the time gains for relationship-buildings tasks, which was not always the case. Instead, the main performance gains of salesforce automation systems came from an improved customer understanding. (Eggert and Serdaroglu, 2011, p. 182)

An approach that can benefit all three productivity factors is customer selection. As a motivation for studying the customer selection approach, Cespedes et al (2013) mention the observation that profitable sales are generally attributable to relatively few customers (Cespedes et al., 2013, p. 54). They also found that the growth of entrepreneurial firms, in particular, is held back by an overreliance on "heroic" efforts by individual sellers and a lack of customer selection criteria. As an example of the implementation of a customer selection process, Cespedes et al present the case of BusinessProcessingCo. (actual name disguised), a growth company that provides web-based payroll services to small and medium-sized businesses. Through investigating its internal data on profitability, selling cycles and lifetime values, BusinessProcessingCo was able to identify the ideal customer group to target: professional services firms with more than 15 employees. These companies were small enough to not have internal IT staff and thereby required the types of outsourced services that BusinessProcessingCo offered, while being big enough to offer a stable revenue stream. Within a year of the completed implementation of the customer targeting, BusinessProcessingCo was able to increase its bookings by 25% with fewer sales representatives. Additionally, the selected customers churned at half the previous rate, which further supported profitability and revenues. In this case, customer selection directly improved the close rate (increased bookings) and the profit per sale (less churn), but it also improved the first productivity factor (the sales capacity), as salespeople could focus on sales tasks with a higher impact. (Cespedes et al., 2013)

Another way of improving the third productivity factor (the profit per sale) is to sell more to each customer. To this end, real-time recommendation systems have been proposed. As an example, a desk lamp could be recommended to a purchaser of economical bed sheets in August because he may be buying supplies for his college dorm room (Cameron and Brunette, 2006, p. 14). For digitally distributed services such as Netflix and Amazon, such systems have been in place for a fairly long time due to the availability of real-time data, and the unobtrusive nature of the recommendations. For other types of services, implementing such systems is quite challenging. Most firms have batch processing of data, which is not conducive to real-time recommendation systems. These companies would have to undergo significant and costly changes to establish real-time data pipelines. On top of that, building real-time decision systems is another technically challenging initiative. Additionally, there are distribution systems which do not fit well with real-time recommendation systems. For example call center interactions are rarely suitable for product recommendations, as customers do not want to be 'interrogated', which severely limits the interactivity of the process. (Cameron and Brunette, 2006)

While individual components of a larger sales process have been developed in many variations and contexts, there are few cases in the literature that describe an implementation of data-driven sales process that would cover all the aspects, such as customer selection, resource allocation, sales monitoring and the interactions between them. Van der Linden and Jain (2012) offer a possible explanation in their presentation of Accenture's seven principles of sales analytics. They acknowledge that analytics as a part of the sales division is still in its infancy, and in one of their seven principles they urge practitioners to start slowly by only introducing analytics into one or two functions (Van der Linden and Jain, 2012). Efforts to develop analytics-based sales processes seem to be mainly driven by the largest business software providers. To this end, IBM has published several papers of process implementations which cover multiple stages of the sales process. Lawrence et al (2007) developed two analytics-based models to support sales: a probabilistic model that identifies new sales opportunities in existing client accounts and non-customer companies (coined OnTarget), and another model to drive the sales allocation process on the basis of analytical estimates of future revenue opportunities (coined the Market Alignment Program) (Lawrence et al., 2007). In a similar vein, Baier et al (2012) present a sales methodology implemented through three analytical models, a Growth and Performance program (GAP), a Territory Optimization Program (TOP), and a Coverage Optimization with Profitability (COP) program. Of the three models, GAP optimizes sales capacity and profitable sales growth, TOP optimizes the assignment of customers to sellers and sales channels and COP provides recommendations on

adjustments to the sales coverage on the basis of customer profit estimates (Baier et al., 2012, p. 1).

As a contrasting view to the perspectives presented up to this point, Adamson et al. (2013) question the effectiveness of a rigid sales process. As customer buying behavior has changed, sales performance has grown erratic with lower conversion rates and less reliable forecasts. Customers are increasingly knowledgeable about available options and successful sales now require unexpected insights and solutions from sellers. Adamson et al. argue that the traditional and formulaic sales process will no longer thrive in this new environment and suggest a new type of selling, coined insight selling, which relies on the insight and judgment of the sales representatives. (Adamson et al., 2013)

4 Research setting & methods

In this section, we present the approach and methodology used to solve the business problem at hand. The first subsection describes the general research philosophy of the thesis. The second section presents the source data, and the software used for the data analysis. After this, we move on to the concrete methodology used to solve the business problem. Here, the focus is on the first research question (i.e. the customer selection problem), as it forms the main empirical part of this thesis. For the second and third research questions, the presented solutions are more qualitative in nature and will be discussed in sections 6.2 and 6.3.

4.1 Research setting

The focus of this thesis is on solving a given business problem and describing the solution process. Academic research of this type falls under the term *design science*, and focuses on "tackling ill-structured problems in a systematic manner" (Holmström et al., 2009, p. 67). Throughout the solution process, we encounter decisions where solutions are not directly available in the academic literature. In these cases, we attempt to explain the alternatives and reasons for choosing a particular direction. Still, the customer selection solution is heavily anchored in quantitative methods through the use of statistical data processing techniques. Our benchmarking of different statistical models in the customer selection section can also be seen as an addition to the literature on statistical customer scoring models.

As the intended end product of this thesis is a new sales process, our ambition is to create a process innovation. To get a theoretical framework to guide this innovation, we borrow the concept of "lean innovation" from the startup world. A lean startup abandons detailed business plans and fully functional prototypes and instead focuses on developing a "minimum viable product", which focuses on presenting the main business idea with the bare minimum of features and functionalities. This product is then iteratively improved based on customer feedback. In this thesis, the customer selection model and accompanying sales process is a somewhat rough first attempt at establishing an analytics-driven sales process in corporate banking. To get the most out of this proposed sales process, it needs to be iteratively improved and adjusted in collaboration with the relevant stakeholders, such as the sellers and regional managers. (Blank, 2013)

4.2 Data

4.2.1 External data

The primary data source for the customer scoring model is the Voitto+ application developed by Suomen Asiakastieto. The application contains financial information on some 90 000 Finnish firms, reported at regular time intervals. The following financial ratios are included in the application in precomputed form:

- Revenue
- Revenue per employee
- Revenue change-%
- Gross profit
- Gross profit per employee
- Gross profit change-%
- Earnings before interest, taxes, depreciation, and amortization-% (EBITDA-%)
- Profit-%
- Return on Investment-%
- Current ratio
- Quick ratio
- Equity ratio
- Return on Assets-%
- Gearing
- Relative indebtedness-%
- Working capital-%
- Inventories per revenue-%
- Turnover of receivables in days
- Turnover of payables in days

In addition to these precomputed financial ratios, the application also includes complete financial statements for all the reported companies. Hence, we can also compute other financial ratios, if necessary. (Suomen Asiakastieto, n.d.)

The Voitto + application also contains the official industry classification code for all reported firms, as defined by Statistics Finland. For our analyses, we use the highest level of classification, which covers the following categories:

- Agriculture, forestry and fishing

- Mining and quarrying

- Manufacturing

- Electricity, gas, steam and air conditioning supply

- Water supply; sewerage, waste management and remediation activities

- Construction

- Wholesale and retail trade; repair of motor vehicles and motorcycles

- Transportation and storage

- Accommodation and food service activities

- Information and communication

- Financial and insurance activities

- Real estate activities

- Professional, scientific and technical activities

- Administrative and support service activities

- Public administration and defence; compulsory social security

- Education

- Human health and social work activities

- Arts, entertainment and recreation

- Other service activities

- Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use

- Activities of extraterritorial organisations and bodies (Statistics Finland, n.d.)

Prior to starting this project, the Voitto + data for selected companies had been uploaded to Aktia's data warehouse for years 2009-2012 Hence, the financial data used in this thesis is limited to those years. There were some observations for other years as well, but these observations were too few to perform a successful imputation of missing values for the dataset (see section 4.4.3.2). Hence, we decided to limit the predictor variable dataset to years 2009-2012.

Conveniently, the uploaded dataset also contains information on the location of the companies. The classification has been made according to Aktia's operative regions into the following (anonymized) regions:

- Region 1

- Region 2

- Region 3

- Region 4

- Region 5

- Region 6

To summarize, the external data contains the following information:

- Financial ratios and financial statements for years 2009-2012

- The industrial classification of the firm

- The geographical region of the firm

4.2.2 Internal data

In addition to the external data available from the Voitto + application, Aktia's own data warehouse provides a wide selection of data that can be used in the scoring. Most importantly, Aktia's database contains rolling profitability measures and forecasts for all of Aktia's corporate customers on a monthly basis. Additionally, the database contains other information that can be used to evaluate the quality of a customer for Aktia, such as the extent of the firm's customer relationship with Aktia.

4.3 Software

All of the internal data used in this thesis was stored in Aktia's Oracle databases. To process and query this data we used an Oracle client and the PL/SQL query language (Oracle, 2015). For the modelling work and for plotting, we used R, which is a programming language and environment for statistical computing(The R Project, 2015). The R language can be extended via packages which add new functionalities to the environment. In the following, we list the most important packages that were used for this thesis project, and briefly describe how they were used.

- For connecting to the Oracle database, we used the ROracle package (Mukhin and Luciani, 2014).
    - ROracle offers functions for querying and updating existing tables, and for writing new tables to an Oracle database.
- For our statistical modelling experiments, we used the mlr-package (Bischl et al., 2015)
    - The mlr-package provides a standardized interface to a variety of popular machine learning algorithms along with other supporting features that allow for the benchmarking, cross-validation and tuning of different models. The library allows for machine learning experiments to be performed in a modular fashion without too much extra coding.

- For plotting, we used the ggplot2-package (Wickham and Chang, 2015)
    - The ggplot2-library is an implementation of the grammar of graphics in R. It allows for a wide variety of plots to be constructed in a step by step fashion.
- For filling in missing values in our data, we used the mice-package (Buuren et al., 2014)
    - The mice-package allows for missing data to be imputed using a variety of academically well founded imputation algorithms.
- For selecting statistically significant predictors, we used the Boruta package (Kursa and Rudnicki, 2014)
    - The Boruta package implements the Boruta feature selection procedure which determines important predictors by adding completely random "shadow" variables to the data and comparing the cross-validated feature importance of each of these shadow variables to the feature importance of each of the predictors. Predictors that have a significantly higher impact on the response variable than the shadow variables are labeled as "confirmed", statistically significant predictors.

The main deliverable of this thesis is a first version of an analytics pipeline for developing a customer selection model and using this model for scoring new, non-Aktia customers. The technical composition of this pipeline is described in detail in Appendix 7.

4.4 Customer selection model

The aim of the customer selection process is to define a model that finds the most fitting non-Aktia customers to approach for new customer acquisition. In a more formal sense, we are looking for a function $f(x) = y$ that most optimally maps a selection of predictor variables x to a response variable y. Here, the predictor variables function as a "footprint" of a company and its activities, while the response variable functions as a quantified measure of the fit of the firm for Aktia.

Concretely, the customer selection approach combines external data on the financial quality of companies from the Voitto + database and internal data on Aktia's existing corporate customers. The goal is to determine the response variable based on Aktia's internal data, and to use the external data as predictor variables. With this approach, the model is fit to Aktia's internal data (i.e. for companies for which the response variable is known) and predictions are made for external data (i.e. non-Aktia companies in the commercial database). This allows for external, non-customer companies to be scored.

In the following examination, the modelling problem is divided into three components:

- Determining the predictor variables x
- Determining the response variable y
- Constructing a model that maps x to y

Each component of the model is discussed in its own subsection.

4.4.1 Predictor variables

The predictor variables are the variables that form the input to our customer selection model. Importantly, the predictor variables need to be chosen so that they can be determined for both non-Aktia customers and Aktia customers. Without this property, the model cannot be generalized to the population of non-Aktia customers. Fortunately, the Voitto + database offers this property and can be used as the main data source for the predictor variables.

To further bolster the financial information found in the Voitto + dataset, Laitinen's and Altman's Z-scores were computed and included in the selection of predictor variables. As mentioned in the literature review, these scores are accounting-based measures of the financial health of a company. The definitions of the two types of Z-scores can be found in Appendix 1. The inclusion of the Z-scores in the predictor variables is intended to strengthen the explanatory power of the statistical scoring model by adding some academically well founded measures of financial health to the dataset. Additionally, the Z-scores can function as baselines that other scoring models can be compared to.

To summarize, the dataset of predictor variables contains the following information:

- Yearly financial ratios for each company including Laitinen's and Altman's Z-scores for years 2009-2012 (ratios listed in 4.2.1).
- Laitinen's and Altman's Z-scores for years 2009-2012, including the following component financial ratios required to compute the Z-scores:
    - Altman's component ratios T_1, T_2, T_3, T_4 and T_5
    - Current assets
    - Current liabilities
    - Book value of equity
- The geographical region of the company's operations (see 4.2.1). Companies from outside Aktia's operative regions were excluded from the dataset.
- The industry class of the company (see 4.2.1).

With four years of data, this selection of variables leads to roughly 150 variables in total. With a dataset of approximately 2000 scored customers, there are only about a dozen of

observations for each variable. Hence, this selection of predictor variables was deemed sufficient and no other variables were computed from the available financial statements.

A key question in the modeling process is how to deal with the existence of several yearly data points of accounting information for each company. As will be described in the following subsection, we decided to formulate the response variable in such a way that the information on the profitability and quality of each company was aggregated into one single response variable. Hence, we also needed to aggregate the yearly predictor data of each company into one data point. There are several ways of performing this aggregation. One option would be to take the mean of all the yearly predictor variable measures and use this as an aggregated predictor variable. Another option is to weigh the predictor variables using some type of heuristic, e.g. by placing a higher emphasis on more recent predictor variables. In the end, we decided against these types of aggregation methods. Instead, we included all the available variables by giving every yearly financial ratio its own column in the matrix of predictor variables. We then used a statistical feature selection procedure to extract the significant features from the predictor variable matrix. This feature selection is discussed in sections 4.4.3.2 and 5.1.

4.4.2   Response variable

Another central decision in the modelling process is defining the response variable. In the academic literature, the most common choice of a response variable in corporate scoring is a binary variable indicating defaulters and non-defaulters. For our case, using this type of binary variable is not quite optimal due to the fairly low number of Aktia customers firms that have defaulted, and challenges relating to choosing a suitable default horizon. Hence, we chose an approach which aims to model customer profitability and quality in a more general sense. This is a somewhat rarer approach than the typical classification of customers into good and bad payers, but is more in line with actual business objectives than a binary classification approach. Additionally, Aktia maintains rolling measures of realized and predicted profits for all its customers, which means that profitability information is readily available.

Initially, the idea was to simply use the income generated by the firm as the response variable, but this was found to be excessively noisy. Corporate customers with very similar financial performance were found to have large differences in generated incomes e.g. due to different lending terms. Additionally, it was not entirely clear as to which year's profitability data to use; should we only focus on the latest profitability number or take into account the trend or stability of profitability measures? To circumvent these problems, we chose a more holistic approach to quantifying the customer value of Aktia's corporate customers by using a

27

"balanced scorecard"-type of evaluation system. A deciding factor in pursuing this type of approach was the fact that the international market leader FICO seems to use scorecards in forming its credit scoring models (FICO, 2015b). Additionally, balanced scorecards are frequently used in the management literature when dealing with decision problems with multiple criteria (Kaplan and Norton, 1995).

After discussions with my thesis supervisor at Aktia, the decision criteria found in Table 3 were included in the scorecard.

Table 3: Decision criteria for customer scoring

Criterion	Measure
Profitability I - Latest profitability	Customer generated income for the year 2014 (rolling 12 months measure)
Profitability II - Stability of profitability	The average customer generated income for years 2012-2014
Profitability III - Profitability trend	The average yearly absolute change in customer generated income for years 2013 and 2014
Strength of customer relation I – Primary bank status	A variable indicating whether the firm regularly handles its payments through Aktia
Strength of customer relation II – Cross selling	The number of different product categories that the firm has purchased from Aktia

Importantly, Aktia's profitability measure also includes expected losses and capital costs. Hence, the profitability measure also contains default information, and we do not need to add a separate criterion for loan defaults.

For a first attempt at a scorecard, we used a very simple binning approach, where the decision criteria are divided into bins, and then companies are distributed into bins and scored accordingly. This approach was a bit challenging, as it led to a somewhat narrow distribution of scores and a high number of ties between firms. A few scores were much more common than others, with e.g. roughly 5%-10% of firms each having a score of 5, 10 or 13.

To counteract this, another scoring approach was tested. Instead of binning profitability measures, we took the logarithm of the profitability measures. This resolved the majority of ties and made for a much less narrow distribution of scores. Still, this approach was not quite satisfactory, as the logging produced a very steep increase in scores around the lower end of

the profitability spectrum. At the higher end of the spectrum, an increase in the generated profit led to only a small increase in score. E.g. an increase in generated profit from 3000€ to 3500€ increased the score only by $\log(3500) - \log(3000) \approx 0.067$. We would like to reward corporations for increases in generated profits more evenly throughout the entire range of profits.

As the simpler approaches did not quite work out, we performed a more thorough process of determining a suitable scoring system for Aktia's corporate customers.

4.4.2.1 *Utility function for customer profitability*

As a theoretical foundation for our scoring process, we used utility theory. In utility theory, a common task is to map different outcomes x to achieved utilities with a utility function $U(x)$. The utility function encodes preferences so that larger utility values correspond to more favorable outcomes. Using utility functions of different shapes allows for different types of preferences to be encoded. A concave utility function corresponds to risk-averse preferences, a convex utility function corresponds to risk-seeking preferences, and a linear utility function corresponds to risk-neutral preferences.(Garvey, 2008, p. 65)

The first significant hurdle in determining suitable utility functions for our profitability measures is the somewhat skewed distribution of profitability measures. While more than 95% of profitability observations fall within an interval of roughly $[-x€, x€]$ (profitability measures omitted), the remaining outlier observations deviate significantly from this central mass. If we fit utility functions to all of the observations (including the outliers), our utility function ends up dominated by the outliers, and the differences in utilities for measurements in the central mass of the distribution will be miniscule. Still, we would rather not remove the outliers altogether, as e.g. the firms with highly negative profitability values tell us which firms to avoid, and can add significant explanatory power to our model. After some consideration, we decided to solve the outlier problem by "cutting off" the utility functions at certain fixed thresholds. If the profitability reaches a certain threshold, we no longer care for further profitability shifts in that direction. An interval of -y€ to z€ was found to be suitable, as it contains more than 96% of all observations for all three of our profitability measures. If the profitability of a company is less than -y€, the firm will be given the minimum utility score. At the other end of the spectrum, companies with a profitability of more than z€ will be given a full utility score.

After dealing with the problem of outliers, we moved on to determining the utility function for the central mass of the profitability distribution. We approached the problem by plotting a few

of the most common shapes of utility function and then asking Aktia to choose the function that most accurately resembles its profitability preferences. The plot of utility functions is displayed in Figure 4.
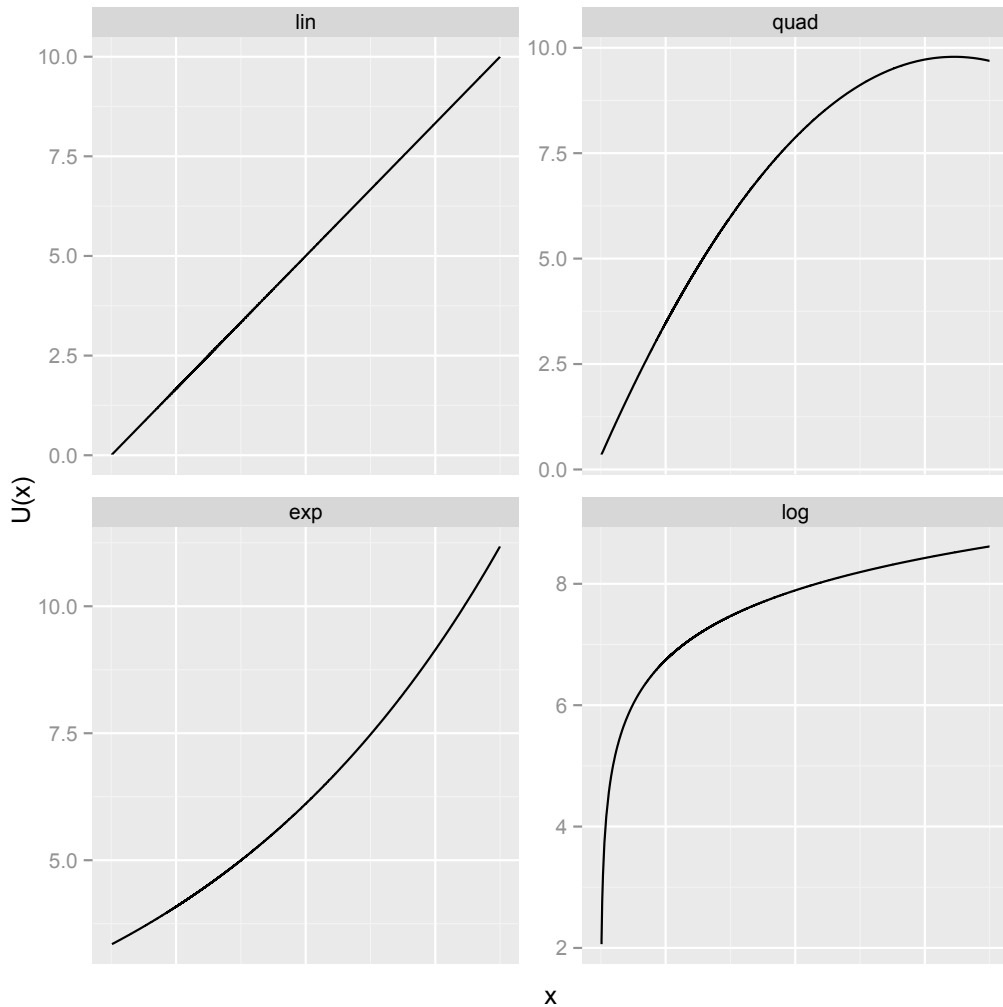
Figure 4: Alternative utility functions for profitability scoring

In Figure 4, we have displayed four types of utility functions:

- A linear utility function for risk-neutral preferences (labeled lin in the graph)
- A quadratic utility function for risk-averse preferences (quad)
- An exponential utility function for risk-seeking preferences (exp)
- A logarithmic utility function for highly risk-averse preferences (log)

The functions were fit in a fairly approximate fashion by defining a handful of points with risk averse preferences and utility values between 0 (our intended minimum utility score) and 10 (our intended maximum utility score), and then fitting curves to these points. Due to the small

number of points, the curves are not exactly bounded between 0 and 10, but the general shapes are more important for the utility function choice than the exact values.

Of the four different types of utility functions, the linear utility function was found to be the most suitable, as Aktia had neither particularly risk averse nor risk loving preferences. Additionally, the linear utility function is robust, understandable and easy to interpret.

4.4.2.2 Final score weights

To filter out firms that have recently experienced significant losses, we added a slight adjustment to the scoring procedure. If the latest profitability measure of a firm is less than -y€ (i.e. the minimum cut-off for criterion one) and its average profitability for the last three years is less than -y€ (i.e. the minimum cut-off for criterion two), all the other three scoring criteria will also be scored as zeroes. This was mainly done to exclude firms that have made significant losses throughout the last three years, but have recovered significantly in the preceeding year. Without the adjustment, the third decision criterion would inflate the scores of these types of firms significantly.

For scoring the two non-profitability measures, we use a binning approach. A customer with a primary bank relationship gets extra points. For the cross-selling index, we encoded preferences in a non-linear fashion.

Using this scoring philosophy, we get the following weights for the different criteria:
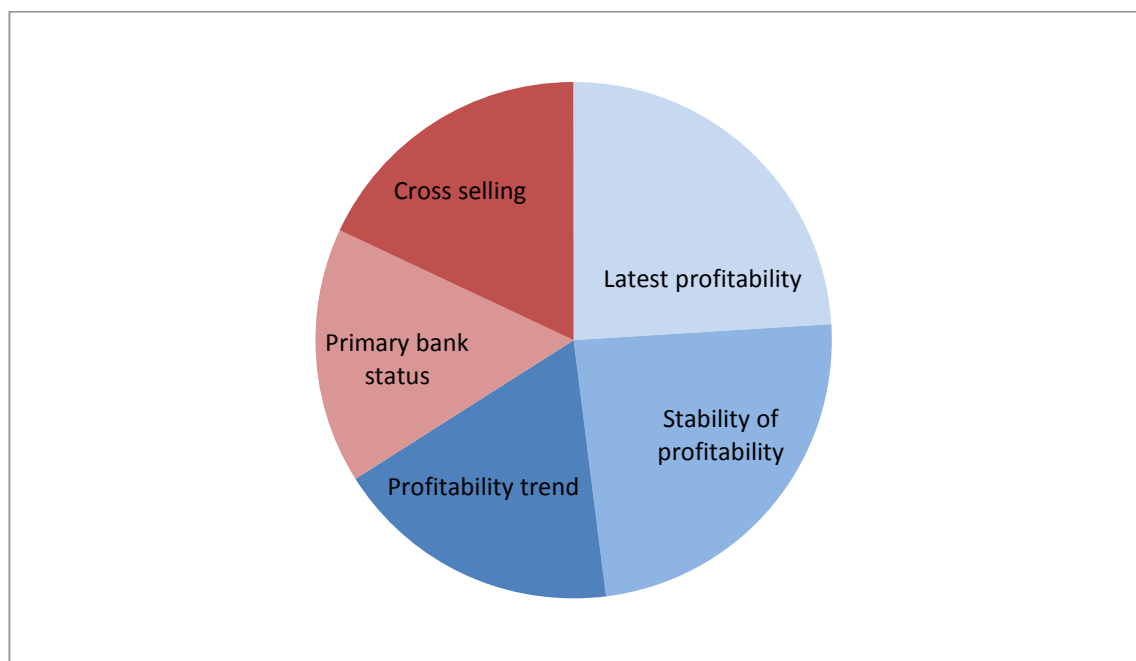
Figure 5: Component weights in corporate score

31

As can be seen in Figure 5, the different profitability measures stand for roughly two thirds of the total customer score. The criteria related to the strength of the customer relation compose the remaining third of the score. The score weights were estimated by computing the average contribution of each component criterion to the total firm score for all the scored companies.

4.4.3 Scoring model

In this sub-section, we describe the procedure of building and comparing our firm scoring models. Each part of the procedure is discussed in turn, starting with the problem formulation and pre-processing and then moving on the model fitting and evaluation.

4.4.3.1 Problem formulation

The first question in building the scoring model pertains to the formulation of the modelling problem. Perhaps the most obvious problem representation is to treat the scoring problem as a regression problem, where we use the predictor variables x to predict the customer score as a continuous variable y. This problem formulation is quite easily interpretable, with higher scores implying a higher level of customer quality.

As the ultimate objective of the scoring process is to find a way of selecting the best firms from the dataset, we can also approach the problem from a classification perspective. In this formulation, the response variable y is a categorical variable with two states, good and bad firms. For our modelling, we decided to include the 30% of firms with the highest scores in the category of good firms, and the rest of the firms in the bad category. This binary formulation is quite similar to the typical loan default classification problem. The hope is that separating good firms from bad firms may be an easier problem than directly quantifying the quality of a firm as a continuous numeric measure.

For the two different problem formulations (regression & classification), we compare models both within the formulation class and across the two formulations. The objective is two-fold: to find well-performing scoring models and to compare the performances of the two problem formulations.

4.4.3.2 Pre-processing

The first pre-processing challenge addresses missing data. Companies can have data gaps for several reasons. In some cases, the company has only existed for a few years, which means that it is missing some data points for earlier years. In the case of some small firms, it is also possible that the financial statements are not complete enough to compute all financial ratios. To fill in missing data, we used R's mice-package for missing value imputation. We tried out two methods: predictive mean matching (the library's suggested default method) and a

method utilizing Bayesian linear regression. (Buuren et al., 2014) The imputation methods were evaluated by comparing the distributions of the imputed predictor variables to the distributions of the non-missing predictor variables using histograms and quartiles. Based on this inspection, the predictive mean matching-method was found to lead to more similar distributions for the imputed and non-missing datasets, and was chosen as the selected imputation method.

Our second pre-processing decision relates to the categorical variables in our dataset (the geographical region and the industry classification). The R programming language offers support for categorical variables through factors (Spector, 2007), but not all statistical models offer support for the factor data type. Hence, we decided to handle categorical variables by encoding them into binary dummy variables. With this approach, a categorical variable is split into as many binary dummy variables as there are unique categories, with each binary dummy variable taking a value of one only when an observation belongs to the corresponding category. This type of approach is frequently used in the literature to encode categorical variables (Hardy and Reynolds, 2004).

Another challenge relates to the large number of features in the dataset. Without pre-processing, there are 142 predictors and roughly 1700 customers for the dataset for which we have response variable information. To alleviate this, we performed an initial feature selection using R's Boruta package. The Boruta procedure finds relevant features by comparing the importance achievable by the predictor variables to the importance achievable at random. The underlying statistical model used by Boruta to determine the feature importance of each is a random forest. Using the random forest base learners, the predictor variables are then classified into confirmed, tentative and rejected predictors. This feature selection was performed separately for both the classification and regression formulations of the problem. (Kursa and Rudnicki, 2014)

As the final step in pre-processing, we performed feature scaling. Some of the statistical models we used are distance-based, and most commonly used distance measures (such as the Euclidean distance) assign a greater weight to features with wider ranges than features with narrow ranges (Aksoy and Haralick, 2001). The financial ratios we use as predictor variables have very different ranges: some ratios vary between 0 and 1 while others have ranges that span thousands. Hence, we need to scale the features to have similar ranges. We chose to scale our features to have zero mean and unit variance, which is a fairly common scaling strategy (Aksoy and Haralick, 2001; Geladi and Kowalski, 1986). For performing this scaling, we used the scale-function of R's base library.

4.4.3.3 Statistical models

Due to our decision to use R's mlr-package for our statistical experiments, we limited our investigation of different statistical models to models supported by the library. Fortunately, the selection of available models is sufficiently comprehensive, and all of the most common types of machine learning models (linear models, decision trees and ensembles of decision trees, support vector machines, nearest neighbors-methods, neural networks) are represented. Importantly, the package seems to support the implementation of all the model types encountered in the literature review of credit scoring models. Brief descriptions of all the models included in our experiments can be found in Appendix 2.

4.4.3.4 Model performance

A key issue in the modelling process is evaluating the quality of a particular statistical model. The theory of comparing different classification and regression models is quite well established, but comparing models across two model formulations is somewhat more unconventional. In this section, we discuss each of these three model evaluation categories separately.

4.4.3.4.1 Regression metrics

In an overview of model performance metrics for air quality models, Wilmott (1982) divides regression metrics into two categories: correlation measures and difference measures. Correlation measures, such as Pearson's correlation coefficient and its square (i.e. the coefficient of determination), are based on computing a correlation index between the observed and predicted values. Difference measures, such as the mean squared error and the mean absolute error, are based on computing the difference between the predicted and observed values. Both types of metrics have their advantages and disadvantages. On one hand, the coefficient of determination (a correlation measure) seems to be slightly more popular in the literature as a measure of model performance; on the other hand, difference measures are easier to understand and explain. In the end, we chose to use two difference measures (the root mean squared error and the mean absolute error) to compare our regression models. Of the two metrics, the root mean squared error penalizes highly inaccurate predictions more harshly than the mean absolute error. The mean absolute error is slightly easier to interpret, as it tells us how far our prediction is from the actual target on average. (Willmott, 1982)

4.4.3.4.2 Classification metrics

For classification models, a very widely used and interpretable metric is classification accuracy. Given a validation set, the accuracy is the proportion of items in the validation set that are

properly classified (Alpaydin, 2014). In cases where labels are distributed unevenly, the accuracy can be a problematic metric. For example in cases where the proportion of positive samples in the dataset is low, a trivial classifier that predicts all cases as negatives will perform very well in terms of accuracy.

A more robust measure of classification performance is the F1-score, which is computed as the harmonic mean of the precision and the recall. The recall is equal to the true positive rate (i.e. the number of correctly predicted positives divided by the total number of positives in the sample), where the precision is the number of true positives divided by the number of all predicted positives (including possible false positives). The F1-metric puts a stronger emphasis on returning the actual true positives than the accuracy metric. (Forman, 2003)

The aforementioned metrics evaluate the quality of a classification or grouping of a classification model, but there are also metrics that evaluate the quality of the predicted probabilities of a model. For the binary classification problem, one popular metric of this type is the receiver operating characteristic (ROC). In a ROC curve, the true positive rate of a binary classifier (the number of true positives divided by the total number of positives in the sample) is plotted against its false positive rate (the number of false positives divided by the total number of negatives in the sample) for various prediction thresholds. Here, the prediction threshold refers to the cutoff in predicted probabilities above which an observation is predicted to be positive. The information provided by the ROC curve can be conveniently summarized in one metric, the area under the ROC curve (sometimes referred to as the AUC), which allows for models to be compared. A perfect classifier has an AUC of 1, and a completely random classifier has an AUC of 0.5. (Alpaydin, 2014)

For our comparisons of classification models, we decided to use the F1-score and the AUC. Of the two metrics, the AUC is of particular interest, as it directly evaluates the quality of the predicted probabilities. The F1-score is affected by the choice of the prediction threshold, as it evaluates the quality of a classification (with a threshold of 0.5 being a commonly used default value). In the final business application of the scoring model, the intention is to rank firms in a continuous fashion, which means that there is no need to limit model diagnostics to only comparing class predictions.

4.4.3.4.3 General metrics

Comparing the scoring performance of a regression model with that of a classification model is a somewhat rare occurrence, and is not really covered in the academic literature. Hence, we need to approach the question from the perspective of this particular business problem. The

final application case of the scoring model is to use the predicted customer score to select a subset of high quality firms to contact for new customer acquisition. Hence, the order of the scores is more important than the absolute scores. This is quite useful, as we don't have to worry about the fact that the predicted regression scores take on different values than the classification probabilities.

In addition to ordering firms, the model will also be used for the classification of firms into good and bad classes. This classification will, however, be preceded by a decision on the number of firms to select from the population of scored firms. In this sense, the scoring classification task differs from the standard classification task, where the classification model is expected to also accurately predict the correct number of positive samples in the prediction set.

To help solve these two decision problems, we decided to use two general metrics: Kendall's Tau and the true positive rate (i.e. the recall). Kendall's Tau is a rank correlation measure that can be used to assess the quality of a ranking. In essence, Kendall's Tau divides the two compared sets of objects into pairs, and counts whether each pair is discordant or concordant. The Tau metric is then computed as

$$\tau = \frac{c - d}{c + d}$$

where c is the number of concordant pairs, and d is the number of discordant pairs (Nelsen, 2011). In the case of ties, a normal approximation adjustment is added to the formula (Kendall, 1948). We use R's Kendall package for computing the Tau, and the package implements the adjustment for ties (McLeod, 2011).

The Tau coefficient has a fairly intuitive interpretation. If we take any pair of objects from our comparison sets, the Tau is equal to the probability of the objects being in the same order in both sets minus the probability of the objects being in different order in the two sets. A perfect ranking has a Tau value of exactly one, while a reversed ranking has a Tau of minus one. For two independent rankings, the Tau value is around zero. (Abdi, 2007) We chose Kendall's Tau as our measure of ranking performance for three reasons. Firstly, it seemed to be the most interpretable of the encountered rank correlation coefficients. Secondly, it is quite robust to different types of inputs, as it only depends on the relative ordering of the inputs and not their absolute values. Thirdly, it can also handle ties, which are likely to occur in our experiments, as some of the simpler statistical models will learn very simple decision rules with only a few distinct prediction outcomes.

To evaluate the quality of a classification proposed by a scoring model, we used the true positive rate, i.e. the number of true positives divided by the total number of positives in the sample. This metric fits well with the business objective: we want our selection of good firms to contain as many truly good firms as possible. To allow for the classification and regression approaches to be compared fairly, we needed to introduce a common classification rule. The classification predictions, the regression predictions and the true firm scores were sorted in descending order with the best firms coming first and the worst firms coming last. For the classification approach, this meant sorting predictions in descending order of predicted probabilities; for the regression approach, this meant sorting predictions in descending order of predicted scores; and for the ground truth values, the observations were sorted in descending order of company scores. In the case of ties, the tied elements were shuffled randomly. For both types of predictions and for the ground truth observations, the top-30% observations were classified as positives and the rest as negatives. With these classifications, we can compute the true positive rates for the two approaches by comparing their proposed classifications to the ground truth.

4.4.3.4.4 Validation

The final point in assessing model performance is to define a cross-validation method for evaluating the models. As is fairly well acknowledged, statistical models should not be tested on the same dataset that they are trained on. This type of behavior leads to overfitting and overestimation of the model's generalization performance. Instead, different sets should be used for model fitting and testing. For our model comparisons, we used the popular K-Fold cross-validation method. In this method, the dataset is split into K different folds, and each one of the K folds is in turn used as the test set while the rest of the folds are used as training data. To get the final cross-validated metrics, we take the mean of the metrics of the different folds. For the case of the classification problem (and for the general investigation across problem formulations), we used a stratified version of K-Fold cross-validation, where the data is divided into folds so that all folds should contain roughly equal proportions of positive and negative samples. (Alpaydin, 2014) We chose to set the number of cross-validation folds at K = 6.

4.4.4 Combining statistical models

In the discussions so far, we have only considered single statistical models. According to machine learning theory, it is possible to improve model performance by combining multiple base learners into an ensemble of models. This approach corrects for the errors made by different statistical models on different observations by utilizing other models to classify some

of the more difficult observations. Assembling a successful ensemble of models requires establishing a diverse base of learners that complement each other. This diversity can be reached in many ways, by using different algorithms, by training the same algorithm with different hyperparameters, by using different input features or by using different subsamples of the same training set. The individual base learners need to be diverse and should be reasonably accurate, but need not be very accurate individually.(Alpaydin, 2014)

Just as there are many ways of choosing suitable base learners, there are also many ways of combining base learners. The simplest method of model combination is voting, which entails taking a linear combination of base learners. Other model combination methods include bagging, where base learners are evaluated on different parts of the training set, and boosting, where base learners are trained sequentially on samples that were misclassified by other base learners. In terms of more advanced approaches, there is stacked generalization, where the voting weights are not linear but learned by another classifier, and cascaded classifiers, where the misclassified or uncertain training samples are passed from one classifier to the next. (Alpaydin, 2014)

For our machine learning experiments, we used a fairly simple ensembling strategy. For base learners, we choose a handful of the best performing models for the two problem formulations, according to Kendall's Tau and the true positive rate. For model ensembling, we use a voting approach. Every predicted score (regression) or predicted probability (classification) is mapped to a vote so that the smallest prediction gets one vote, the second smallest prediction gets two votes, and the largest prediction gets as many votes as there are observations in the validation set. In the case of equal predictions, each tied element is given points according to the maximum sorted index of the tied predictions. For example, if both our second-smallest and third-smallest predictions have a prediction value of 0.2, both of the predictions will be given three points. The final ensemble prediction is then achieved by computing the mean of the vote totals of the individual base learners. To determine the suitability of ensembling for our modelling problem, we try out different combinations of base learners to see if the ensembling improves model performance.

4.4.5   Summary of modelling procedure

The chart in Figure 6 summarizes the different stages of the scoring procedure, starting from data preparation, then moving on to model benchmarking and finally to model interpretation.

1. Data preparation

Generate predictor variables x

Define company scores y

Regression formulation
Predictors: x
Response variable: y

Classification formulation
Predictors: x
Response variable:
- Positive class: best 30% of scores
- Negative class: remaining 70%

Preprocessing
Missing value imputation
Encoding of categorical variables
Feature selection
Feature scaling

2. Model evaluation

Regression model benchmarking
Performance metrics: RMSE, MAE

Classification model benchmarking
Performance metrics: F1-score, AUC

General model benchmarking
Performance metrics: Kendall's Tau, True positive rate

Model ensemble benchmarking
Performance metrics: Kendall's Tau, True positive rate

Choice of final model

3. Model interpretation

Model interpretation

Figure 6: Summary of modelling procedure

5 Results

In this section, we go through the results of the initial customer scoring process. The discussion proceeds in roughly the same order as the presentation of the scoring model in section 4. First, we go through the results of the feature selection procedure. After this, the results of the modelling experiments are presented, with the regression results coming first, followed by the classification results and finally the general comparison of the two types of models. Based on the general comparison, we choose the models with the strongest performance and evaluate how an ensemble of models performs on this dataset. After performing the model comparisons, we choose a final model (or ensemble of models) to use for the customer selection. In the final part of the results section, we interpret the proposed decision models of a few well-performing scoring models.

5.1 Feature selection

As described in section four, the Boruta procedure divides features into three categories: accepted, rejected and tentative features. By increasing the number of iterations, the tentative features are resolved into either accepted or rejected features. We chose to run the procedure for 200 iterations to minimize the number of tentative variables. For the regression formulation of the problem, 200 iterations of the Boruta procedure led to 144 rejected predictors, 5 tentative predictors and 16 confirmed predictors. For the classification formulation, the procedure also confirmed 16 predictors, rejected 145 and left 4 as tentative. In Figure 7, the confirmed predictors and the importance of their random forest features are displayed for both the regression and classification formulations of the problem, with the regression results on top and the classification results on the bottom.
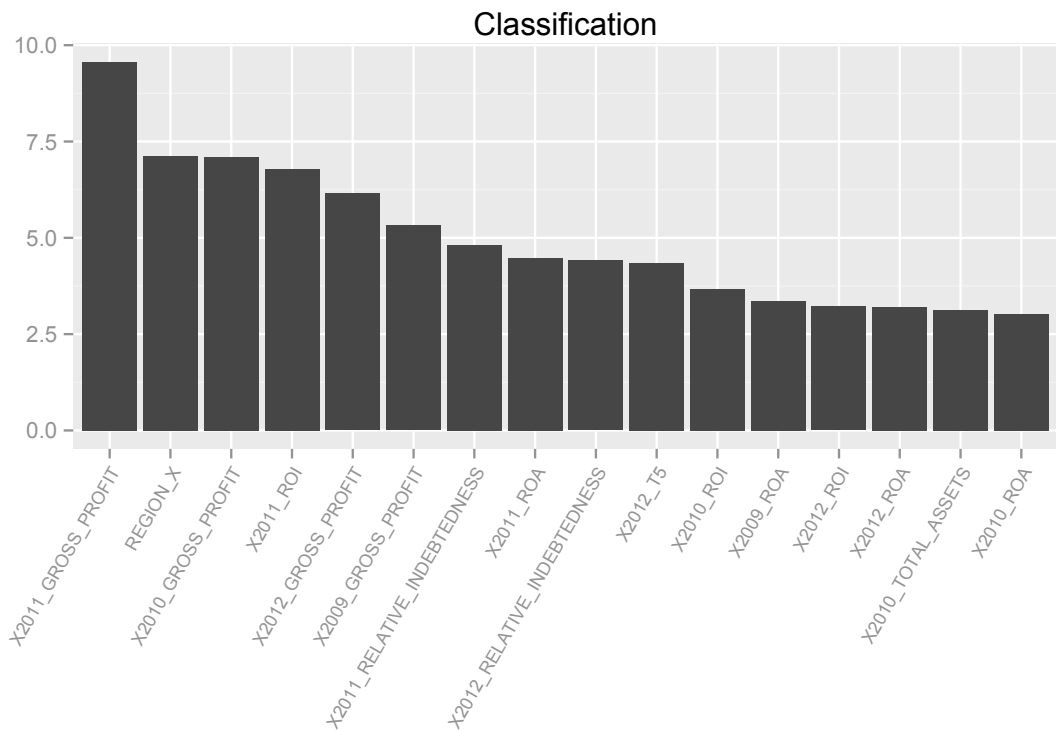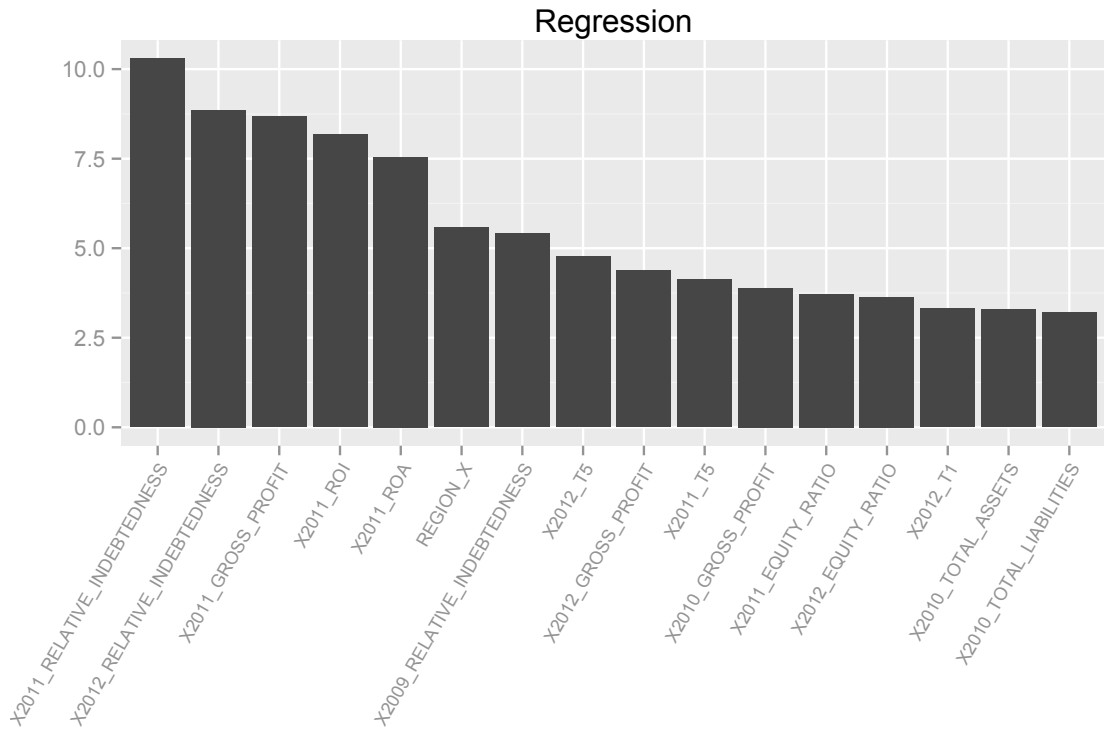
Figure 7: Confirmed, statistically significant predictors for the regression and classification problems

As can be seen in Figure 7, the two formulations lead to roughly the same confirmed features. For both formulations, the gross profit and relative indebtedness-measures emerge as significant. Both formulations also consider the dummy variable representing one of the

geographical regions to be important (Region X), which is somewhat unexpected. The regression formulation seems to have a slightly higher focus on measures of financial health, e.g. when it comes to the high importance of the relative indebtedness-measures, and the inclusion of the equity ratio-predictors for the years 2011 and 2012. The classification formulation seems to place an increased focus on profitability measures through the inclusion of as many as seven different ROI or ROA measures and the high importance assigned to the gross profit measures for years 2010 and 2011. Of the engineered variables (i.e. Altman's and Laitinen's Z-scores and their component variables), only the T1 and T5 component variables emerge as significant. The T5 variable is the Sales/Total assets-ratio and the T1 expresses the working capital divided by the total assets. None of the predictors relating to the industry classification of firms emerged as significant.

5.2 Regression models

Using the feature selections proposed by the Boruta procedure, we generated the training dataset by keeping only the confirmed predictors in the dataset. Using this dataset, we benchmarked all of the available and functioning regression models supported by the mlr-library. The benchmarking results are reported in Appendix 3. In the table of Appendix 3, the model evaluation metrics for the regression models are displayed. The table reports the model name, the cross-validated root mean squared error (RMSE) and the cross-validated mean absolute error (MAE) in descending order of RMSE values.

The best models reach a mean absolute error of roughly 8.7 and a root mean squared error of about 10.5. Of the different model classes, decision tree models clearly have the strongest performance. The bartMachine, cforest and extraTrees models, as well as the random forests are all collections of decision trees. The models consisting of single decision trees (ctree and rpart) also have above average performance. The non-decision tree models with the strongest performance are the earth and mars models, which are both implementations of the multivariate adaptive regression splines-algorithm (MARS). Of the other model classes, support vector machines perform well in terms of the MAE, but not in terms of the RMSE. For example the svm model has the lowest MAE of all compared models while having only average performance in terms of the RMSE. The linear models (glmnet, lm, and the lasso and ridge-models) all have almost identical performance with RMSE and MAE values of roughly 10.70507 and 9.017976, respectively. Neural networks (brnn and nnet) have somewhat average performance while nearest neighbors methods (fnn, kknn) perform poorly.

Here, it should be noted that the hyperparameters of the models were not tuned to this case, and instead we used the defaults proposed by the mlr-package (except for the xgboost-model, for which the number of trees in the ensemble was increased from the default value of 1 to 100). This can make the results somewhat skewed in favor of the decision tree ensembles, which are known to work well without much tuning. Additionally, the base learner used in the Boruta procedure is a random forest, which can further skew the results in favor of the decision tree models(Kursa and Rudnicki, 2014). For support vector machines and, in particular, for neural networks, a more careful tuning of model parameters could lead to some improvements. Additionally, the dataset used is quite small, which can be problematic e.g. for neural networks, which tend to excel in problems with larger amounts of data.

5.3 Classification models

In Appendix 4, the comparison of classification models is displayed. For each model, the area under the ROC curve (AUC) and the F1 score (F1) are reported. The table is sorted in descending order of the AUC.

A completely random classification corresponds to an AUC value of 0.5. It is encouraging to see that most of the tested models clearly beat the trivial random benchmark. The exception to this is the single decision tree model rpart, which failed to find anything and has an AUC of 0.5 and an F1 score of 0. Still, even the highest AUC values are only around 0.64, and about half of the models have an AUC of less than 0.6, which is quite weak.

In terms of the relative strength of the models, decision tree models, again, have the strongest performance, as the top four models according to AUC are decision tree models (bartMachine, cforest, ada and gbm). Linear models also performed reasonably, as glmboost, lqa and glmnet all performed well. In general, boosted models seem to perform slightly better for the classification formulation than the regression formulation, as shown by the success of the glmboost, ada and gbm-models. While discriminant analysis models had roughly average performance, nearest neighbors models, support vector machines and neural networks, on the other hand, had below average performance.

As can be seen in Appendix 4, the F1 scores for this comparison vary quite wildly. The models are so uncertain about their predictions that they do not predict many positive samples (and in some cases none at all) with the default prediction threshold of 0.5. Hence, the F1 scores are probably not as informative for this comparison as the AUC values are.

5.4 General comparison

The table in Appendix 5, presents the general comparison of models. For each model, Kendall's Tau and the true positive rate (TPR) is reported. The table is sorted in descending order of the Tau-value.

In this general comparison, we have included a few baselines as comparison points for our models. For baselines, we use a uniformly random prediction, the equity ratio for 2012 and Laitinen's and Altman's Z scores for 2012. We see that just about all of the tested models beat the three baselines by a clear margin.

There are quite a few jumps in performance from the earlier comparisons. According to the RMSE and MAE metrics, the rpart and blackboost-regressors only ranked 8th and 10th respectively, but for this general comparison, they are the two best models in terms of Kendall's Tau. The blackboost-regressor also does very well in terms of the true positive rate. Another significant improvement in performance is shown by the nnet classifier, which had below average performance for the AUC, but now ends up being the best classifier in terms of the Tau. On the other hand, a several findings from the classification and regression comparisons also hold true for the general comparison. For one, decision tree models remain the best model class. Additionally, the glm-based classifiers do quite well.

As for the comparison between regression and classification models, regression models tend to do better in terms of Kendall's Tau, while classification models do better in terms of the TPR. Nonetheless, regression models are slightly better overall, with the best models having TPRs that are quite close to the TPRs of the best classification models. The svm-regressor is somewhat of an exception, as it clearly outperforms the svm-classifier in terms of both the Tau and the TRP.

5.5 Model ensembling

For the ensemble models, we chose the 6 best regression models and the 6 best classification models according to Kendall's Tau. Using these base learners, we constructed the following ensembles using uniform voting (as described in the methods section)

- An ensemble with all the 12 best regression and classification models (ensemble_all)
- An ensemble with the single best regression model and the single best classification model (ensemble_two)
- An ensemble with the 6 best regression models (ensemble_reg)
- An ensemble with the 6 best classification models (ensemble_clf)

The model performance metrics for these ensembles (and the base learners) can be found in Appendix 6.

From the table in Appendix 6, we see that ensembling seems to work quite well in this case. All of the four ensembles outperform all other models except for the rpart-regressor in terms of the Tau, with the ensemble consisting of all 12 base learners having the highest Tau-value at 0.193. In terms of the TPR, the three best ensembles also outperform all base learners, with the ensemble of the 6 best classifiers having the highest TPR.

5.6   Final model choice

Based on the performance comparisons, we chose the ensemble with the 12 best regression and classification models as our final model. It had the best performance in terms of the Tau and the second best performance in terms of the TPR. Additionally, we presumed that having as many as 12 base learners could bring some stability to the predictions, as a few extreme predictions from a few of the base learners will not have a colossal effect on the ensemble predictions.

5.7   Model interpretation

In this subsection, we investigate the decision rules behind the predictions some of our better models. First, we take a look at the decision tree of our best-performing single regression model, rpart. In the second section, we investigate the importance of the individual features of another fairly well-performing regression model, the cforest.

5.7.1   Decision tree

A somewhat surprising finding was the strong performance of the rpart decision tree model, particularly in terms of Kendall's Tau. It outperformed a number of complex models, such as gradient boosting machines and random forests, which are formed by averaging and combining the predictions of several individual decision trees in different ways. In the plot in Figure 8, we display the decision tree of an rpart-model that was fit to the entire training set.
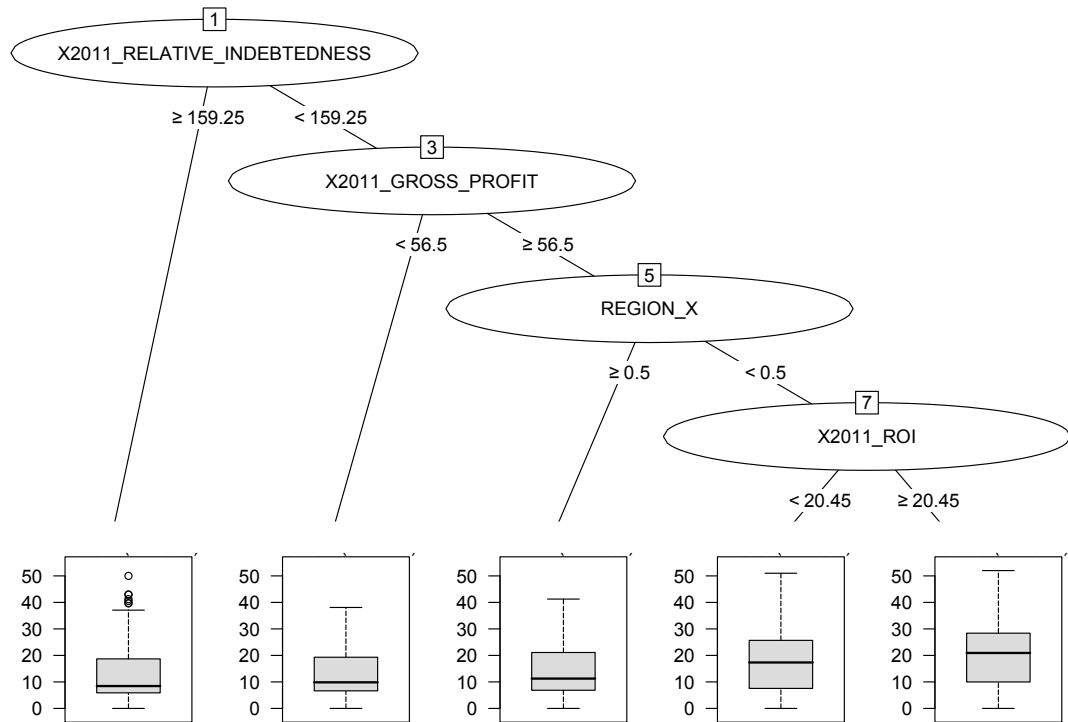
Figure 8: Decision tree plot

The decision tree is fairly simple, with four decision nodes and five outcome nodes. For each of the outcome nodes at the bottom, the model predicts the median of the displayed box-and-whisker plot. As there are only five outcome nodes, the model only predicts five different scores. The decision rules can be deduced by reading the tree from top to bottom. The lowest score prediction occurs in the leftmost node, for cases where the 2011 relative indebtedness exceeds 159.25%. The highest score prediction occurs in the rightmost node for cases where the 2011 relative indebtedness is below 159.25%, the 2011 gross profit exceeds 56 500€, the firm is not located in the X region, and the 2011 ROI exceeds 20.45%.

5.7.2 Random forest interpretation

In this sub-section, we investigate the decision rules of the cforest-regressor, which was the second-best single regression model in terms of the RMSE. The cforest-model is a random forest, so it is formed by averaging a large number of decision trees of the type presented in the previous sub-section.

A brief interpretation of the importance of the individual features of a random forest was conducted with the earlier investigation of the importance of the individual features using the Boruta procedure. A key shortcoming of that investigation was that only the magnitude of the influence of a variable was considered and not the direction of the relationship. Admittedly,

interpreting the direction of influence is not always straightforward in a collection of decision trees, as it is possible that both very large positive and very large negative values of a particular variable are associated with the same predictive outcome. Nonetheless, we made an attempt at estimating the direction of the relationship between the variables and the response. We fit the cforest-regressor to all of the training data and used this model to predict customer scores for the test set of non-Aktia companies. The firms of this test set were divided into good and bad firms by classifying the top 5% of the highest scoring non-Aktia customers as good firms. The means for all the predictor variables were computed for the two classes of firms, and for each predictor variable, we computed the percentage difference between the good group and the bad group. This percentage difference in the group-wise predictor means was then used as a measure of the direction of the relationship between the predictors and the response. Much as for the plots of the Boruta procedure, the magnitude of the relationship was quantified by using the feature importance of the random forest. This information was then plotted on a bar graph, where the lengths of the bars represent the importances of individual features (the magnitude), and the colors of the bars represent the group-wise percentage differences in predictor means (the direction). For the cforest-regressor, this particular plot can be found in Figure 9.
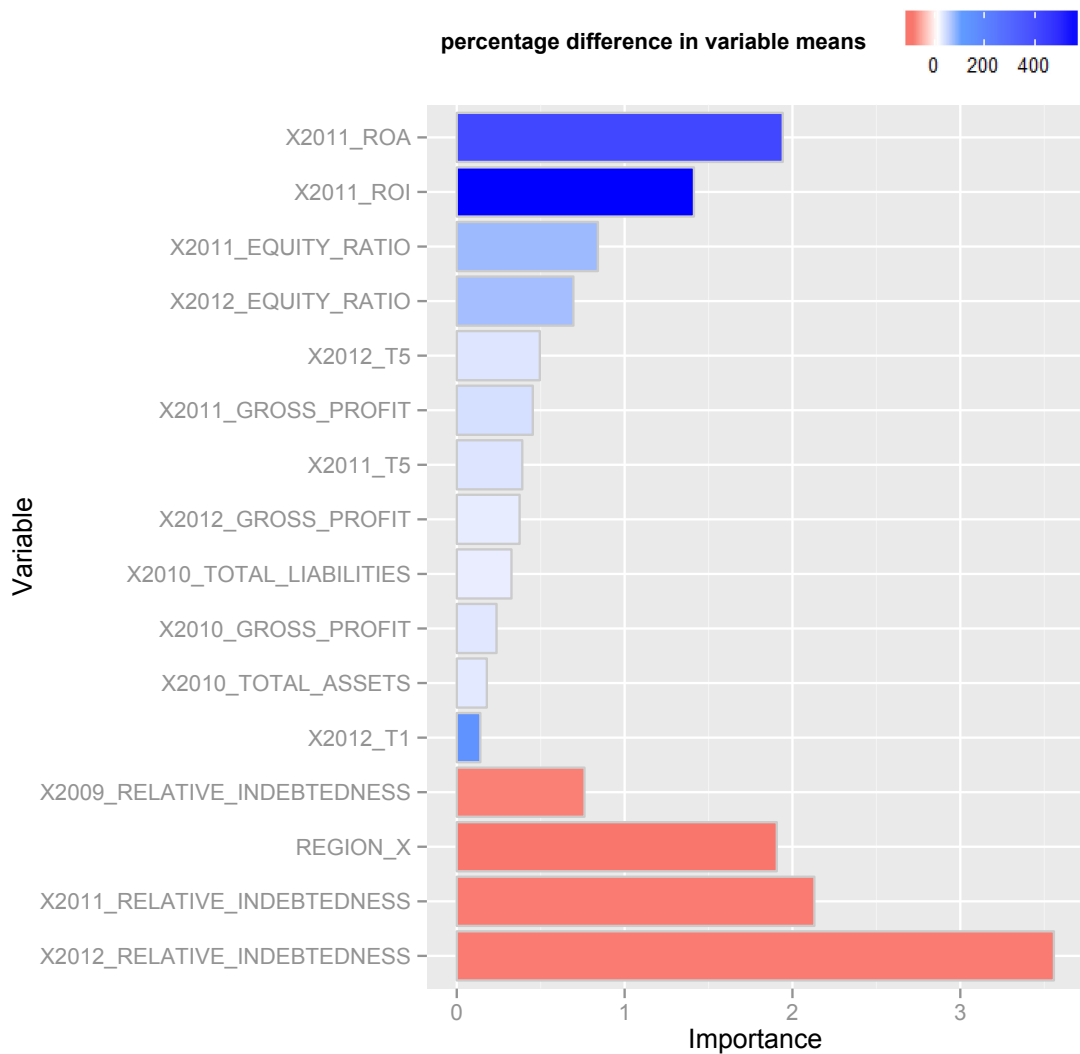
Figure 9: Variable importance and direction

In Figure 9, the solid red and blue bars have the largest directions of influence numerically. The variables with the clearest positive influence on the response are the 2011 ROA and the ROI for the same year. For these variables, the percentage difference between the good and bad groups is several hundred percent. In other words, the 2011 ROI and ROA values are several times larger for the group of (predicted) good firms than for the group of bad firms. The variables with the largest negative influence on the response are the relative indebtedness-variables and the dummy variable for region X. For these variables, the values in the group of good firms are only a small fraction of the values of the bad firms. For the rest of the variables, the percentage differences between the variable means in the two groups are positive. For a few variables (the T5 values), the percentage differences are quite small, while e.g. the T1 variable for the year 2012 has a fairly large percentage difference between the good and bad groups.

As a final target of investigation, we grouped the importance of individual features by year in Figure 10.
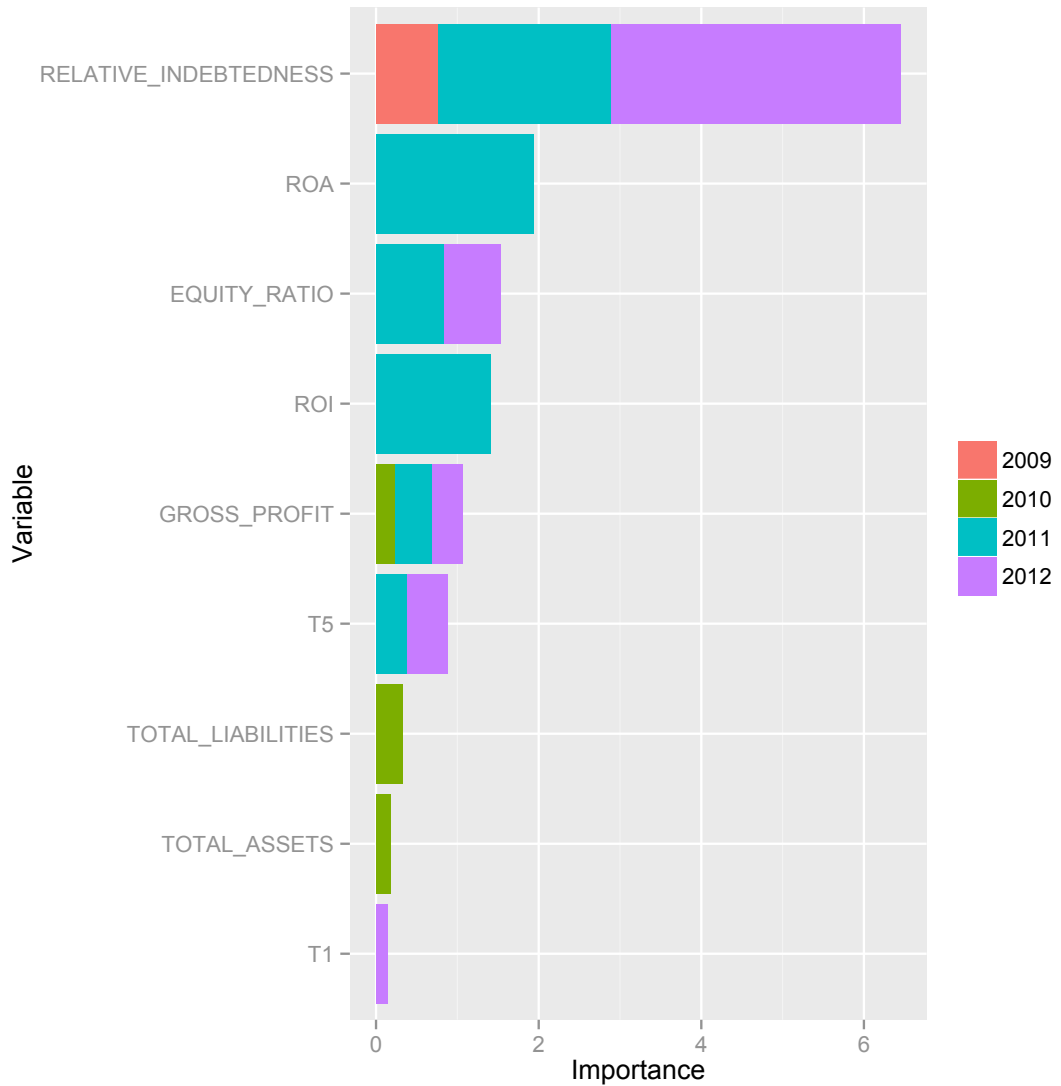
Figure 10: Variable importance by year

As can be seen from Figure 10, the emphasis is on more recent financial ratios. Only one financial ratio from 2009 is included (the relative indebtedness); for the year 2010, three ratios are included (the gross profit, the total assets and the total liabilities). Whenever a financial ratio occurs many times, the latest figure tends to be the most significant, except for the gross profit, where the 2011 gross profit has a slightly higher feature importance than the other years. Figure 10 also shows the strong emphasis that is placed on the relative indebtedness-variable.

6   Discussion and evaluation

In the first part of this section, we evaluate the merits and the shortcomings of our customer scoring model. After this, we move on to our second and third research questions. We approach the sales resource allocation problem by proposing an action plan for carrying out the customer selection suggested by our scoring model. In the third subsection, we present some ideas on how Aktia could move towards an analytics-driven sales process, and how this scoring procedure could inform the design of such a process. In the fourth and final subsection, we relate our findings to the academic literature.

6.1   Scoring model – Key findings, shortcomings and ideas for further research

Perhaps the most interesting discovery of the corporate scoring process is the proposed feature selection. The significant features found by the Boruta procedure were in line with expectations, and the feature importances of the final random forest were reasonable both in terms of the magnitude of the effect and the direction. A higher level of indebtedness led to less favorable scores, while strong performance in e.g. the ROI and ROA metrics led to more favorable scores.

In a somewhat unexpected turn, our model ended up penalizing companies from one particular region (Region X) quite heavily. The reason for this should be investigated further. One possible explanation is that Aktia's market share in this region is quite low and a few poorly performing or excessively competitively priced loans may stand for a large share of Aktia's business in the region. As a result, the models will associate the region with customers of poor quality, and flag all companies from that region as unwanted customers. One needs to consider whether future customers should be penalized for the region's past challenges. One possibility is to completely remove the region variable from future versions of the model to avoid this problem.

Another aspect of the feature weightings that could warrant further research is the relative importance of the same ratio for different years. For the regression formulation, the relative indebtedness ratio was found to be statistically significant for three years (2009, 2011 and 2012). On the other hand, the return on assets was found to be significant only for 2011. This variable was, however, the most important single financial ratio according to feature importance. It is not entirely clear as to why certain variables are confirmed only for one year,

while other variables are included for several years. Perhaps the RoA does not vary much over the years and it is sufficient to only include the feature for one year.

Another component of the model that could be useful for Aktia in the future is the proposed scoring of Aktia's corporate customers, and the scorecard approach underlying the scoring process. It offers a new perspective into customer segmentation: instead of discrete clusters of thousands of customers, the customer score distributes customers into "segments of one" in a more continuous fashion. As we noticed when struggling with outliers during the score definition process, a few select companies stand for surprisingly large shares of the bank's losses and profits. Our scoring approach brings special attention to these outliers, and thereby encourages action to either correct the worst losses or to nurture the most profitable customer relationships of the bank.

Additionally, our scoring process led to a few theoretical insights. Of the tested model classes, decision tree models had the best general performance across the board. Of the two proposed problem representations, the regression formulation ended up performing slightly better than the classification formulation, particularly in terms of Kendall's Tau. For the true positive rate, however, the best classifiers ended up slightly outperforming the best regressors. Another interesting finding was our success in improving model performance through ensembling. With a simple linear voting ensemble involving the strongest-performing base learners, we were able to improve performance in terms of both the Tau and the true positive rate. The fairly large differences in relative performance for different scoring metrics could warrant some further investigation. We chose Kendall's Tau as our ranking metric due to its robustness and ability to handle ties, but it is possible that some other metric could be a better choice. In this case, selecting the metric was challenging due to the need to compare regression and classification models.

The most significant shortcoming of the model was its lack of explanatory power. While both our regression and classification models outperformed random predictions, the absolute performance of our models remained quite weak. For classifiers, our best models only reached a cross-validated AUC of roughly 0.65, which is somewhat mediocre. In terms of the Tau, even our best ensembles stayed below a value of 0.2, which means that the probability of ranking a pair of objects correctly is only just below 0.2 higher than the probability of an incorrect ranking. The weakness of the model led to a general lack of robustness for the experiments. For two different cross-validation strategies or even for different random seeds, we could end up with fairly significant mix-ups in the best-performing models. Additionally, we found that

different random seeds resulted in notably different decision trees. For some random seeds, we were unable to build decision trees due to the lack of explanatory power in the variables.

One possible solution to the lack of explanatory power is to acquire more training data. The dataset used for this thesis was intentionally quite small, and also slightly outdated, as the scoring model developed for this thesis aims to function as a pilot. With roughly 100 000 companies in the Voitto + database, and Aktia's market share of a few percentage points, the upper limit for the maximum size of the training set is somewhere around 5 000 companies, implying a tripling of the current training set. In order to further increase the size of the dataset, one could consider abandoning the current approach of aggregating all yearly data into one observation and instead adding several annual observations for each company. Another possible approach to improve the explanatory power could be to tune the hyperparameters further or to try other statistical models, but these fine tuning efforts are unlikely to help much.

While we may experiment with several potential means of improving the explanatory power of the model, it is likely that there is only a weak relationship between financial statement information and the quality of a customer. In addition to being in good financial health, a company needs to select Aktia as its primary bank to become a highly scored customer. The process of selecting a bank is difficult to model with using only information about the financial statements of a firm. Hence, it seems reasonable to assume that the explanatory power of financial information with regards to customer quality will be somewhat limited.

6.2 Implementing the customer selection proposed by the scoring model

For the customer scoring model to be of any practical use for Aktia, it needs to inform some type of action. Hence, this subsection is dedicated to describing a process for deploying the customer selection recommended by the scoring model. The suggested process is outlined in the flowchart in Figure 11.

Step 0: Updating data + model

↓

Step 1: Gathering contact information

↓

Step 2a: Regional pilot - cold calling

↓

Step 2b: Regional pilot - follow-up appointments

↓

Step 3a: Large-scale cold calling

↓

Step 3b: Follow-up appointments

Figure 11: Initiating the sales process

As can be seen in Figure 11, we suggest a two-tiered customer acquisition process, where the first layer of cold calls is handled by phone sellers, with corporate sellers being responsible mainly for handling the follow-up appointments arranged by the phone sellers. The idea is that the first stage of cold calling is somewhat standardized and does not require the domain expertise and experience of the corporate sellers. Freeing corporate sellers from making cold calls allows them to keep their workload at reasonable levels. Importantly, corporate sellers should still be included in designing the cold calling process, e.g. when it comes to the sales script, and the customer selection. In the following discussion, we go through each of the stages of initiating the sales process.

6.2.1 Step 0: Updating the data and the model

As discussed in the previous section, the limited dataset could be a possible explanation for the fairly weak explanatory power of the model. Hence, we recommend for more training data to be used for the model that will be used in the actual customer selection.

In addition to increasing the amount of training data, the data should be updated to reflect a more recent period. Currently, the most recent financial statements are from 2012. If we start contacting customers based on this data, there is a risk of contacting firms that have gone bankrupt or undergone a severe deterioration in financial health over the past few years.

To ensure the best possible fit for the selected firms, some firms could also be filtered out from the training data. For our first version, we included all firms regardless of their size, as

our training data was not particularly large to start with. For the next version, we could filter out very large firms (e.g. firms with more than €100m in revenue), as very large firms may require somewhat more tailored approaches for customer acquisition.

Following the update of the data, we should also update the model by repeating the model comparisons, and making sure that our chosen models still perform relatively well with the new dataset. Also, if a more complicated statistical approach is deemed too opaque, it is possible to replace the model with some type of simple decision model, such as e.g. ranking companies in descending order of ROI. If necessary, the firm scorecard could also be updated.

6.2.2 Step 1: Gathering contact information

Our external dataset with financial information on companies contains the business ID of the companies (Y-tunnus). This ID also provides a connection to further, publicly available, contact information, such as the registration address of the company, but the phone numbers for companies are rarely included in open data sources. In order to efficiently contact selected firms on a larger scale, some type of automated process for mapping business IDs to phone numbers should be developed. Most likely, this requires the assistance of an external provider of contact information, such as Fonecta. The first step would be to identify what types of sources for contact information Aktia currently uses, and the applicability of these sources for the current context. Whereas the acquisition of contact information is currently conducted in a somewhat ad-hoc fashion, we should explore the feasibility of creating a more systematic process for gathering contact information for corporate customers.

6.2.3 Step 2: Regional pilot

The cold calling hit rate is one key parameter that influences the implementation of the customer contacting procedure, as it determines the number of calls that needs to be done to achieve a certain number of appointments with potential customers. Correspondingly, it also determines the number of sales resources that needs to be involved in the cold calling step. Due to the importance of the hit rate, we propose arranging a regional pilot for determining a reasonable first guess for the hit rate. For example, Aktia could call 500 or 1000 Helsinki companies with a high predicted customer quality, and then compute the hit rate as the proportion of customers that agree to a follow-up appointment at one of Aktia's branch offices. For these follow-up appointments, we could further investigate what proportion of appointments lead to sales, and what products drive successful sales.

In addition to providing us with a first estimate for the hit rate for cold calls, the pilot would provide an opportunity to test out different sales pitches and approaches to new customer

acquisition. For instance, we could write down two different sales scripts for the cold calls and compare the relative performance of the two scripts. Additionally, we could apply the modelling ideas used for the customer scoring in order to determine whether certain financial ratios help in predicting the success of cold calls.

The information gathered from the follow-up appointments is also very valuable. In addition to investigating the close rate for the selling appointments, we can interview the corporate sellers to gather their opinions on the selected companies: Do the selected companies seem to have potential or does the customer selection do more harm than good?

After the regional pilot, one option is to completely abandon the customer selection procedure, and not proceed with the analytics-driven sales process, e.g. if the hit rate in the cold calling phase is too low, if the selected customers do not align with Aktia's desired customer profile, or if the appointments do not generate enough sales.

6.2.4 Step 3: Execution phase

If the customer selection procedure is deemed worthy of continuing after the pilot, we can move on to performing customer acquisition on a larger scale, across all of Aktia's operative regions. In the resource planning of this execution step, there are three key parameters: the hit rate of the cold calling phase, the close rate of the follow-up appointments, and the desired number of new corporate customers to acquire. The former two will be determined by the pilot, while the latter should be chosen by Aktia to match the company's growth objectives. In the short run, we assume the sales resourcing will remain fixed at current levels, as it significantly simplifies the problem of resource planning.

In the chart in Figure 12, we present a rough assignment of sales resources, where one 40-hour work week of time from each corporate sales resource is assigned to holding follow-up appointments set up by cold calls. In the assignment, we use a cold call hit rate of 5%, a follow-up appointment close rate of 25%, a cold call duration of ten minutes, and a follow-up appointment duration of two hours. In the name of corporate secrecy, the regions have been anonymized and the sales resource quantities altered.

Regions	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6
New customers, total			270			
Customers, regions	25	25	15	75	100	30
Appointment close rate			25%			
Appointments	100	100	60	300	400	120
Hours à 2h	200	200	120	600	800	240
Available resources	5	5	3	15	20	6
Hours/resource	40	40	40	40	40	40
Cold call hit rate			5%			
Cold calls	2000	2000	1200	6000	8000	2400
Total cold calls			21600			
Cold call hours à 10 mins			3600			
Total workweeks à 40 h			90			

Figure 12: One proposed sales resource assignment (for illustrative purposes only)

With the aforementioned assumptions, the sales resource assignment of Figure 12 leads to the acquisition of 270 new customers. With a low cold call hit rate of 5%, we notice that the workload of new customer acquisition falls quite heavily on the cold callers. In total, the workload of the corporate sales resources stands at 54 workweeks while the cold calling requires as many as 90 workweeks. In our simple model, the required cold calling work depends linearly on the hit rate. Hence, a doubling of the hit rate to 10% would halve the required cold calling work to 45 workweeks, and an increase of the hit rate to 25% would shrink the load to 18 workweeks. The sensitivity of the workload to assumptions about the hit and close rates further emphasizes the usefulness of an initial pilot round of cold calls and follow-up appointments.

In the execution phase, it would be highly beneficial to collect as much information about the attempted customer acquisitions as possible. In addition to the financial, geographical and industry classification information in the Voitto + database, information about the types of products or sales pitches used in approaching the customer would be very useful. Additionally, different corporate sellers can perform better in approaching different kinds of companies.

Hence, documenting information about the seller of each new customer acquisition attempt could be useful.

6.3 Moving towards an analytics-driven sales process

In this thesis, the focus has mainly been on detecting and acquiring customers that resemble Aktia's current high-quality customers. In the longer term, the objective is not only to detect high-quality customers, but also to identify customers that have a high probability of becoming Aktia's customers. At the time this master's thesis project began, there was very little data about successful and failed attempts at new customer acquisition. Hence, we focused on a slightly easier problem for which data was available. The methods and models used in the customer scoring problem are, however, also applicable for the more ambitious problem of predicting customer acquisition probabilities. The main challenge lies in establishing a data pipeline that stores relevant information about the sale as well as the outcome of the sale, for both cold calls and follow-up appointments.

The first question in designing the data pipeline relates to data collection. Currently, we have the financial data of the Voitto + database, and the corresponding industry classification and geographical data. To further bolster this data, we suggest collecting some additional data for both the cold calls and the follow-up appointments. For the cold calls, e.g. the following data could be useful:

- The type of sales pitch used (e.g. based on a categorization of sales scripts)
- The duration of the call
- The caller
- The outcome (appointment/no appointment, favorable/unfavorable responses)

For the follow-up appointments, we suggest the following information be gathered:

- The products and offers suggested to the firm
- The seller
- The outcome (sale/no sale)
- Information about why the firm was willing to agree to the follow-up appointment
- Information about why the firm did/did not become a customer

These are initial suggestions for what variables should be documented in customer acquisition attempts. For the follow-up appointment phase, in particular, there may be good reason to include more information, as the appointments are longer and less formulaic than the cold calls. This data could even be documented in slightly less structured form, e.g. by allowing

sellers to write down short descriptions of the appointments. A key challenge is making sure that the data collection process is not obtrusive for corporate sellers or phone sellers. Hence, a lot of effort should be put into choosing data points that are informative and quickly documentable.

By gathering and analyzing information about the cold calls and the follow-up appointments, we can hopefully improve the hit rates and close rates by updating the sales propositions to potential new customers. However, we will not be able to use this information for training our customer selection model, as none of the newly collected data will be available for non-encountered firms. Hence, for our customer selection model, we are limited to the Voitto + data that was used for the initial selection model. Nonetheless, the newly gathered information will provide us with two new response variables: one variable denoting the success of the cold calls and another denoting the success of follow-up appointments. One possible way of utilizing these new variables would be to represent the problem as a three-class classification problem, where one class contains firms that fall out in the cold calling phase, another class contains firms that pass the cold calling phase but fall out in the appointment phase, and a third class contains firms that pass both phases and end up as new customers. This type of model would allow us to choose firms that are likely to become Aktia's customers, or to focus on firms that are likely to pass the first cold calling phase. In addition to collecting information on the hit and close rates, we are also interested in eventually computing the realized customer scores for the customers acquired with the new sales process. This will give us valuable validation data for evaluating our model. For our current score, we use profitability data from a three-year period, which causes a significant delay when scoring new customers. If we want to avoid a three-year validation cycle, we may have to make some adjustments to the scorecard for newly acquired corporate customers.

In the diagram in Figure 13, the envisioned sales process is depicted.
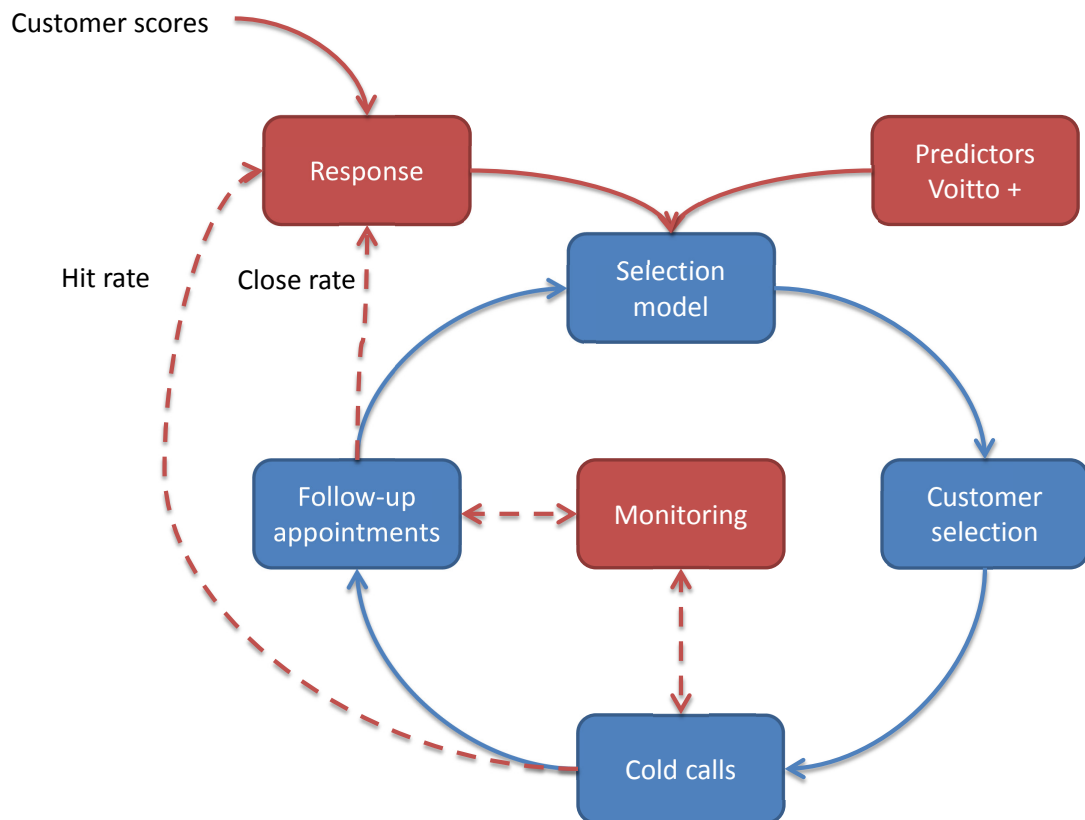
Figure 13: Flowchart of data-driven sales process

The main operative process is drawn in blue. At certain parts of the sequence, data is used to improve the process. These "feedback" points in the process are indicated in red. Furthermore, data that is currently available is denoted by solid arrows while new proposed data inputs are denoted by dashed arrows.

The sales process is executed in batches. For each batch, we start with our selection model, and perform a customer selection according to the model. For the very first batch, the model will be based on only the Voitto + predictors and the initial customer score information. After the selection, phone sellers contact the selected customers. Follow-up appointments are arranged for interested customers. This marks the end of the batch. After each batch, the selection model is updated. We investigate the collected information on the successes and shortcomings of the sales efforts, and use this information to improve future sales activities. Additionally, data on the quantities of successful and failed sales efforts are used as response variables for the updated selection model. The initial scoring model only uses the corporate scores as response variable information, while future batches also take into account the hit and close rates of the two sales phases. After updating the model, we select a new batch of non-contacted customers, and the process starts anew.

6.4 Implications for future research

This thesis can be seen as an addition to the academic literature on credit scoring models. Our perspective is slightly different to most existing approaches, as we model customer quality on a more holistic level than models which only focus on default probabilities. Additionally, we considered both regression and classification formulations of the scoring problem. For our dataset, decision tree models were found to perform the best. In the literature review that preceded our modelling work, decision tree models did not stand out while e.g. neural networks and support vector machines did. This could warrant some further research. If decision tree models can match or even exceed the performance of complex non-linear models, their interpretability could make them a very attractive option for future credit scoring models. This thesis is also somewhat more transparent than traditional academic credit scoring papers in terms of explaining and documenting the underlying predictor variables and interpreting the fitted statistical models.

Another important finding was the long-tailed distribution of the profitability values of the modeled corporate banking customers. The vast majority of firms are concentrated in a fairly narrow band of modest profitability values, but a few outliers stand for significant portions of the total profits and losses. The type of profitability-based scoring that was proposed in this thesis could be useful in identifying these outliers, and thereby drawing attention to the most profitable customers. This continuous scoring method also supports a "segment-of-one"-type of approach to customer portfolio management, which has been suggested as a fitting segmentation method for banking by Winger and Edelman (1989). The relationship-based nature of corporate banking requires customized services and products, and a continuous customer score could support this by facilitating more fine-grained customer segmentation.

For our second and third research questions, i.e. how Aktia should allocate its sales resources and redesign the sales process, this thesis presents a roadmap for future action. Here, some of our choices in the process design could lead to valuable insights during the implementation process. In our sales resource assignment, we propose a two-tiered sales process, where customers are first approached by phone for appointments, where the business of the customer is hopefully won. Here, the intention is to gather data on successes and failures of the cold calls and follow-up appointments, and to hopefully use these insights to improve the hit and close rates of the two stages of the process. If deployed, this procedure can offer many valuable insights about systematic sales processes:

- Does a two-tier sales approach work, or is one tier of sellers (only cold calls/only appointments) better?
- Can we gather useful and clean data without distracting the efforts of sellers?
- Can we use the gathered data to improve the hit rates of the cold calls and the close rates of the follow-up appointments?
- Does a more careful selection of new customers lead to any noticeable bottom-line gains?

Just as the deployed sales process can inform future research, existing research should also inform the implementation of the sales process. In designing the analytics-based sales process in section 6.3, we considered some of the best practices of strong analytics performers. By performing the process in batches, we can move towards a self-correcting process, where data is not used for reporting and evaluating performance after the fact, but rather for guiding sales efforts in, close to, real time. There is also further room to utilize findings from the literature in other aspects of the sales process. The literature importantly identified the value of good loan officers. When only considering quantitative factors, corporate banks will overwhelmingly end up financing similar, financially stable companies at the expense of some more opaque firms with significant upside but higher risk. As much of the real value of corporate lending comes from resolving information asymmetries between financers and SMEs, the incorporation of the experience and expertise of loan officers in the sales process should also be considered. This could be realized by e.g. redesigning the corporate scorecard with the help of the corporate sellers, or by performing an initial filtering of the dataset on the basis of their input.

7 Conclusion

In this master's thesis, we have laid the foundation for a comprehensive, analytics-driven sales process for corporate banking. From an empirical perspective, we have constructed, compared and validated a variety of different scoring models for the selection of corporate customers. From a qualitative perspective, we have proposed a process for implementing the customer selection suggested by our scoring model. Additionally, we have outlined an initial plan for establishing a continuously updated and improved process for the acquisition of new corporate customers.

In terms of empirical results, our modelling work brought to light several statistically significant predictors of customer quality. Among the most important predictors were variables relating to financial health, such as the relative indebtedness, and variables relating to profitability, such as the return on assets and the return on investment. In terms of methodological results, we found decision tree-based models to be the strongest-performing class of models. Additionally, we tried out two different formulations of the customer scoring problem: a regression formulation, where customer quality is modelled as a continuous numerical score; and a classification formulation, where customer quality is modelled as a binary variable that divides firms into good firms and bad firms. Of these two formulations, the regression formulation performed better in terms of Kendall's Tau, while the classification formulation performed better in terms of the true positive rate. When considering both metrics, the regression formulation performed slightly better overall. By combining our best models using a linear voting approach, we were able to exceed the performance of our base learners in terms of both Kendall's Tau and the true positive rate.

In terms of the results of our qualitative work, a key finding was the importance of a pilot implementation of the customer selection procedure. In our initial quantitative estimations, we found the workload required for new customer acquisition to be highly dependent on the hit rate of initial cold calls and the close rate of follow-up appointments. A pilot would give us some initial estimates for these retention rates, and would help us generate a realistic understanding of the feasibility of a large-scale customer acquisition process. Additionally, our qualitative work brings attention to key data points that should be collected from sales activities, and how this data could be used to improve future sales efforts and customer selections. In the lean innovation framework, our initial scoring model is the "minimum viable product" that should be iteratively improved according to input from users, and the results of using the model.

In terms of concrete benefits for Aktia, this thesis project adds several new tools to Aktia's analytics toolbox. Through the data cleaning and processing of the Voitto + data, Aktia now has four years' worth of financial data for thousands of Finnish firms in its data warehouse. The scoring of Aktia's corporate firms allows Aktia to rank the firms in its corporate portfolio and to identify its most and least profitable corporate customers. Additionally, the programming work related to feature selection, statistical modelling and model validation is directly applicable in other contexts, e.g. for loan default prediction, customer retention prediction or customer revenue prediction.

In addition to the technical contributions of this thesis, a successful implementation of the suggested analytics-based sales process has the potential to generate significant business gains for Aktia. Currently, Aktia's customer selection is limited to flagging risky firms with insufficient credit ratings. With a fairly rudimentary baseline, even a weak customer selection model could lead to a significant improvement in the quality and profit potential of acquired customers. In addition to this, setting up a system to collect more detailed information about successful and failed cold calls and appointments could generate remarkable business gains. By introducing techniques for analyzing sales efforts and using these techniques for honing sales activities, Aktia could improve its customer acquisition and retention rates in many different sales channels. Concretely, these retention improvements would lead to a higher productivity of sales activities and an increased rate of new customer acquisition, which are both in line with Aktia's strategic growth targets.

8 Bibliography

Abdi, H., 2007. The Kendall rank correlation coefficient. Encycl. Meas. Stat. Sage Thousand Oaks CA 508–510.

Adamson, B., Dixon, M., Toman, N., 2013. Dismantling the Sales Machine. Harv. Bus. Rev. 91, 102–109.

Agarwal, V., Taffler, R., 2008. Comparing the performance of market-based and accounting-based bankruptcy prediction models. J. Bank. Finance 32, 1541–1551. doi:10.1016/j.jbankfin.2007.07.014

Aksoy, S., Haralick, R.M., 2001. Feature normalization and likelihood-based similarity measures for image retrieval. Pattern Recognit. Lett. 22, 563–582.

Aktia, 2015a. Om Aktia - Aktia [WWW Document]. URL http://www.aktia.com/sv/tietoa-aktiasta (accessed 4.5.15).

Aktia, 2015b. Capital adequacy - Aktia [WWW Document]. URL http://www.aktia.com/en/velkasijoittajat/vakavaraisuus (accessed 5.17.15).

Alpaydin, E., 2014. Introduction to machine learning. MIT press.

Altman, E.I., 1983. Corporate Financial Distress: A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy, 1 edition. ed. Wiley, New York.

Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J. Finance 23, 589–609.

Altman, E.I., Iwanicz-Drozdowska, M., Laitinen, E.K., Suvas, A., 2014. Distressed Firm and Bankruptcy Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. Available SSRN 2536340.

Altman, E.I., others, 2000. Predicting financial distress of companies: revisiting the Z-score and ZETA models. Stern Sch. Bus. N. Y. Univ. 9–12.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., Vanthienen, J., 2003. Benchmarking state-of-the-art classification algorithms for credit scoring. J. Oper. Res. Soc. 54, 627–635.

Baier, M., Carballo, J.E., Chang, A.J., Lu, Y., Mojsilovic, A., Richard, M.J., Singh, M., Squillante, M.S., Varshney, K.R., 2012. Sales-force performance analytics and optimization. IBM J. Res. Dev. 56, 8–1.

Balance Consulting, n.d. Tunnusluvut - Kolmen muuttujan Z-luku [WWW Document]. URL http://www.balanceconsulting.fi/tunnusluvut/kolmen_muuttujan_z_luku (accessed 12.27.14).

Bauer, J., Agarwal, V., 2014. Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. J. Bank. Finance 40, 432–442. doi:10.1016/j.jbankfin.2013.12.013

Beaver, W.H., 1966. Financial ratios as predictors of failure. J. Account. Res. 71–111.

Berger, A.N., Udell, G.F., 2002. Small business credit availability and relationship lending: The importance of bank organisational structure. Econ. J. 112, F32–F53.

Bischl, B., 2015. Mlr Tutorial - Integrated Learners [WWW Document]. URL http://berndbischl.github.io/mlr/tutorial/html/integrated_learners/index.html (accessed 5.4.15).

Bischl, B., Lang, M., Richter, J., Bossek, J., Judt, L., Kuehn, T., Studerus, E., Kotthoff, L., 2015. mlr: Machine Learning in R.

Blank, S., 2013. Why the lean start-up changes everything. Harv. Bus. Rev. 91, 63–72.

Boot, A.W., 2000. Relationship banking: What do we know? J. Financ. Intermediation 9, 7–25.

Businessweek, B., 2011. The Current State of Business Analytics: Where Do We Go from Here? Bloom. Businessweek Res. Serv. Httpwww Sas Comresourcesassetbusanalyticsstudywp08232011 Pdf.

Buuren, S. van, Groothuis-Oudshoorn, K., Robitzsch, A., Vink, G., Doove, L., Jolani, S., 2014. mice: Multivariate Imputation by Chained Equations.

Cameron, M., Brunette, A., 2006. Real-Time Analytics: Leveraging Your Sales Process. DM Rev. 16, 12–16.

Cespedes, F., 2014. Putting Sales at the Center of Strategy. Harv. Bus. Rev. 92, 23–25.

Cespedes, F.V., Dougherty, J.P., Skinner III, B.S., 2013. How to Identify the Best Customers for Your Business. MIT Sloan Manag. Rev. 54, 53–59.

Drexl, A., Haase, K., 1999. Fast approximation methods for sales force deployment. Manag. Sci. 45, 1307–1323.

Durand, D., 1941. Risk elements in consumer instalment financing. NBER Books.

Eggert, A., Serdaroglu, M., 2011. Exploring the Impact of Sales Technology on Salesperson Performance: A Task-Based Approach. J. Mark. Theory Pract. 19, 169–185.

European Commission, 2015. Green Paper - Building a Capital Markets Union [WWW Document]. URL http://ec.europa.eu/finance/consultations/2015/capital-markets-union/docs/green-paper_en.pdf (accessed 5.18.15).

Farinha, L.A., Santos, J.A., 2002. Switching from single to multiple bank lending relationships: Determinants and implications. J. Financ. Intermediation 11, 124–151.

Ferguson, R.B., 2013. Big Data and Big Change Management: A Path Forward. MIT Sloan Manag. Rev.

Fernandes, J., 2005. Corporate credit risk modeling: Quantitative rating system and probability of default estimation. Available SSRN 722941.

FICO, 2015a. ABOUT US [WWW Document]. FICO.com. URL http://www.fico.com/en/about-us (accessed 1.1.15).

FICO, 2015b. FICO Score [WWW Document]. URL http://www.fico.com/en/products/fico-score#overview (accessed 4.26.15).

Finanssialan Keskusliitto, 2015a. Pankit Suomessa - Toukokuu 2015 [WWW Document]. URL https://www.fkl.fi/materiaalipankki/esitysaineistot/ppt/Pankit_Suomessa.ppt (accessed 5.16.15).

Finanssialan Keskusliitto, 2015b. Kansainvälisiä Vertailutietoja Finanssimarkkinoilta - Huhtikuu 2015 [WWW Document]. URL https://www.fkl.fi/materiaalipankki/esitysaineistot/ppt/Kansainvalisia_vertailutietoja_Suomen_finanssimarkkinoilta.ppt (accessed 5.16.15).

Finanssialan Keskusliitto, 2015c. Yritysten Rahoitustilanne Suomessa - Huhtikuu 2015 [WWW Document]. URL https://www.fkl.fi/materiaalipankki/esitysaineistot/ppt/Yritysten_rahoitustilanne.ppt (accessed 5.16.15).

Finanssialan Keskusliitto, 2015d. Tulos- ja tasetiedot 2013-2014 [WWW Document]. URL https://www.fkl.fi/materiaalipankki/julkaisut/Julkaisut/Pankkikonsernien_tulokset_2014.pdf (accessed 5.16.15).

Finanssialan Keskusliitto, 2014. Suomen rahalaitosten taseissa olevat lainat ja talletukset pankeittain/pankkiryhmittäin, 31.12.2014 [WWW Document]. URL https://www.fkl.fi/materiaalipankki/julkaisut/Julkaisut/Pankkien_markkinaosuudet_2014.pdf (accessed 5.16.15).

Finlay, S., 2010. Credit scoring for profitability objectives. Eur. J. Oper. Res. 202, 528–537.

Finlay, S., 2009. Are we modelling the right thing? The impact of incorrect problem specification in credit scoring. Expert Syst. Appl. 36, 9065–9071. doi:10.1016/j.eswa.2008.12.016

Finlay, S.M., 2008. Towards profitability: a utility approach to the credit scoring problem. J. Oper. Res. Soc. 59, 921–931. doi:http://dx.doi.org/10.1057/palgrave.jors.2602394

Forman, G., 2003. An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3, 1289–1305.

FSB, 1998. Small Businesses' Finance and the Economy. Federation of Small Businesses, U.K.

Garvey, P.R., 2008. Analytical methods for risk management: A systems engineering perspective. CRC Press.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185, 1–17.

Gopalan, R., Udell, G.F., Yerramilli, V., 2007. Why do firms switch banks?, in: EFA 2007 Ljubljana Meetings Paper.

Hand, D.J., Henley, W.E., 1997. Statistical classification methods in consumer credit scoring: a review. J. R. Stat. Soc. Ser. A Stat. Soc. 160, 523–541.

Hardy, M., Reynolds, J., 2004. Incorporating categorical information into regression models: The utility of dummy variables. Handb. Data Anal. 229–255.

Holmström, J., Ketokivi, M., Hameri, A.-P., 2009. Bridging practice and theory: a design science approach. Decis. Sci. 40, 65–87.

Howorth, C., Peel, M.J., Wilson, N., 2003. An examination of the factors associated with bank switching in the UK small firm sector. Small Bus. Econ. 20, 305–317.

Huang, C.-L., Chen, M.-C., Wang, C.-J., 2007. Credit scoring with a data mining approach based on support vector machines. Expert Syst. Appl. 33, 847–856.

Ioannidou, V., Ongena, S., 2010. "Time for a change": loan conditions and bank behavior when firms switch banks. J. Finance 65, 1847–1877.

Kaplan, R.S., Norton, D.P., 1995. Putting the balanced scorecard to work. Perform. Meas. Manag. Apprais. Sourceb. 66.

Kawas, B., Squillante, M.S., Subramanian, D., Varshney, K.R., 2013. Prescriptive Analytics for Allocating Sales Teams to Opportunities, in: Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on. IEEE, pp. 211–218.

Kendall, M.G., 1948. Rank correlation methods.

Kiron, D., Perguson, R.B., Prentice, P.K., 2013. From Value to Vision: Reimagining the Possible with Data Analytics. MIT Sloan Manag. Rev.

Kursa, M.B., Rudnicki, W.R., 2014. Boruta: Wrapper Algorithm for All-Relevant Feature Selection.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N., 2011. Big Data, Analytics and the Path From Insights to Value. MIT Sloan Manag. Rev. 52, 21–32.

Lawrence, R., Perlich, C., Rosset, S., Arroyo, J., Callahan, M., Collins, J.M., Ershov, A., Feinzig, S., Khabibrakhmanov, I., Mahatma, S., others, 2007. Analytics-driven solutions for customer targeting and sales-force allocation. IBM Syst. J. 46, 797–816.

Martens, D., Baesens, B., Van Gestel, T., Vanthienen, J., 2007. Comprehensible credit scoring models using rule extraction from support vector machines. Eur. J. Oper. Res. 183, 1466–1476.

Mattila, V.-M., 2013. Pankit ja Yritysrahoitus Finanssikriisissä [WWW Document]. URL http://www.hyol.fi/assets/files/Talous%20tutuksi/Mattila_Taloustutuksi_2013_Turku.pdf (accessed 5.16.15).

McLeod, A.I., 2011. Kendall: Kendall rank correlation and Mann-Kendall trend test.

Mukhin, D., Luciani, D.A.J. and J., 2014. ROracle: OCI based Oracle database interface for R.

Howorth, C., Peel, M.J., Wilson, N., 2003. An examination of the factors associated with bank switching in the UK small firm sector. Small Bus. Econ. 20, 305–317.

Huang, C.-L., Chen, M.-C., Wang, C.-J., 2007. Credit scoring with a data mining approach based on support vector machines. Expert Syst. Appl. 33, 847–856.

Ioannidou, V., Ongena, S., 2010. "Time for a change": loan conditions and bank behavior when firms switch banks. J. Finance 65, 1847–1877.

Kaplan, R.S., Norton, D.P., 1995. Putting the balanced scorecard to work. Perform. Meas. Manag. Apprais. Sourceb. 66.

Kawas, B., Squillante, M.S., Subramanian, D., Varshney, K.R., 2013. Prescriptive Analytics for Allocating Sales Teams to Opportunities, in: Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on. IEEE, pp. 211–218.

Kendall, M.G., 1948. Rank correlation methods.

Kiron, D., Perguson, R.B., Prentice, P.K., 2013. From Value to Vision: Reimagining the Possible with Data Analytics. MIT Sloan Manag. Rev.

Kursa, M.B., Rudnicki, W.R., 2014. Boruta: Wrapper Algorithm for All-Relevant Feature Selection.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N., 2011. Big Data, Analytics and the Path From Insights to Value. MIT Sloan Manag. Rev. 52, 21–32.

Lawrence, R., Perlich, C., Rosset, S., Arroyo, J., Callahan, M., Collins, J.M., Ershov, A., Feinzig, S., Khabibrakhmanov, I., Mahatma, S., others, 2007. Analytics-driven solutions for customer targeting and sales-force allocation. IBM Syst. J. 46, 797–816.

Martens, D., Baesens, B., Van Gestel, T., Vanthienen, J., 2007. Comprehensible credit scoring models using rule extraction from support vector machines. Eur. J. Oper. Res. 183, 1466–1476.

Mattila, V.-M., 2013. Pankit ja Yritysrahoitus Finanssikriisissä [WWW Document]. URL http://www.hyol.fi/assets/files/Talous%20tutuksi/Mattila_Taloustutuksi_2013_Turku.pdf (accessed 5.16.15).

McLeod, A.I., 2011. Kendall: Kendall rank correlation and Mann-Kendall trend test.

Mukhin, D., Luciani, D.A.J. and J., 2014. ROracle: OCI based Oracle database interface for R.

myFICO, 2011. Understanding Your FICO Score. Fair Isaac Corporation.

Nelsen, R.B., 2011. Kendall tau metric - Encyclopedia of Mathematics [WWW Document]. URL http://www.encyclopediaofmath.org/index.php/Kendall_tau_metric (accessed 4.27.15).

Niemeläinen, J., 2014. S-Pankin Ylihurula varoittaa aliarvioimasta pankkia [WWW Document]. Taloussanomat. URL http://www.taloussanomat.fi/rahoitus/2014/06/04/s-pankki-lupaa-lattialle-saa-tuoda-hiekkaa/20147902/12 (accessed 5.16.15).

Oracle, 2015. Oracle PL/SQL [WWW Document]. URL http://www.oracle.com/technetwork/database/features/plsql/index.html (accessed 5.4.15).

Pylkkönen, P., Savolainen, E., 2013. Suomen Pankki - PK-yritysten rahoituksen tila Suomessa [WWW Document]. URL http://www.suomenpankki.fi/fi/julkaisut/euro_ja_talous/rahoitusjarjestelman_vakaus/Docum ents/ET213_Pylkkonen-Savolainen.pdf (accessed 5.16.15).

Skiera, B., Albers, S., 2008. Prioritizing sales force decision areas for productivity improvements using a core sales response function. J. Pers. Sell. Sales Manag. 28, 145–154.

Smith, Y., 2014. Why Getting Rid of Loan Officers Hurt Banks and the Economy. Naked Capital.

Spector, P., 2007. Factors in R [WWW Document]. URL http://www.stat.berkeley.edu/~s133/factors.html (accessed 4.26.15).

Statistics Finland, n.d. Tilastokeskus - Luokitukset - Standard Industrial Classification TOL 2008 - [WWW Document]. URL http://www.stat.fi/meta/luokitukset/toimiala/001-2008/index_en.html (accessed 6.15.15).

Suomen Asiakastieto, n.d. Voitto+ Käyttöohje [WWW Document]. URL http://www.asiakastieto.fi/voitto/ohje/voitto1.htm (accessed 4.25.15).

Taffler, R.J., 1984. Empirical models for the monitoring of UK corporations. J. Bank. Finance 8, 199–227.

The R Project, 2015. R: What is R? [WWW Document]. URL http://www.r-project.org/about.html (accessed 5.4.15).

Turtola, I., 2014. Suomalaispankit selvisivät EKP:n testeistä kuivin jaloin [WWW Document]. Yle Uut. URL

http://yle.fi/uutiset/suomalaispankit_selvisivat_ekpn_testeista_kuivin_jaloin/7552277 (accessed 5.17.15).

Tyrväinen, J., 2015. Houkuttelevia tuottoja perinteisistä osinko-osakkeista - Evli Pankki Oyj [WWW Document]. URL https://www.evli.com/net/evli-visio/evli-visio/01-2015/houkuttelevia-tuottoja-perinteisist--osinko-osakkeista (accessed 5.17.15).

Van der Linden, J., Jain, N., 2012. Bringing Science to the Art of Sales. CRM Mag. 7–7.

Verbraken, T., Bravo, C., Weber, R., Baesens, B., 2014. Development and application of consumer credit scoring models using profit-based classification measures. Eur. J. Oper. Res. 238, 505–513. doi:10.1016/j.ejor.2014.04.001

Wang, J., 2014. Why hire loan officers? VoxEU.org.

Wickham, H., Chang, W., 2015. ggplot2: An Implementation of the Grammar of Graphics.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bull. Am. Meteorol. Soc. 63, 1309–1313.

Winger, R., Edelman, D., 1989. Segment- of- One Marketing. Perspect. Boston Consult. Group.

Zoltners, A.A., Sinha, P., 1980. Integer programming models for sales resource allocation. Manag. Sci. 26, 242–260.

9 Appendices

9.1 Appendix 1: Z-score definitions

Laitinen's Z-score (Balance Consulting, n.d.) is computed as follows:

$$Z = 1.77 * Financial\ result\ \% + 14.14 * Quick\ ratio + 0.54 * Equity\ ratio$$

The financial result-percentage is computed as:

$$Financial\ result\ \% = 100 * Financial\ result/Revenue$$

$$Financial\ result = Net\ result + Depreciation, amortization\ and\ write - downs$$

A larger Z-score implies a higher creditworthiness, with the following thresholds:

- Over 40: Excellent
- 28-40: Good
- 18-28 Satisfactory
- 5-18: Weak
- Below 5: Awful

Altman's Z-score for private firms is computed as

$$Z = 0.718 * T_1 + 0.847 * T_2 + 3.107 * T_3 + 0.420 * T_4 + 0.998 * T_5$$

where

$$T_1 = \frac{Current\ Assets - Current\ Liabilities}{Total\ Assets}$$

$$T_2 = \frac{Retained\ Earnings}{Total\ Assets}$$

$$T_3 = \frac{EBIT}{Total\ Assets}$$

$$T_4 = \frac{Book\ Value\ of\ Equity}{Total\ Liabilities}$$

and

$$T_5 = \frac{Sales}{Total\ Assets}$$

Just as Laitinen's Z-score, a larger Altman's Z-score implies a higher level of creditworthiness. The Z-score has the following "Zones of Discrimination":

- Z > 2.9: "Safe" Zone
- 1.23 < Z < 2.9: "Grey" Zone
- Z < 1.23: "Distress Zone" (Altman and others, 2000, pp. 12, 20–21)

9.2 Appendix 2: Overview of compared statistical models

This appendix offers an overview of all the compared classification and regression models that we used in this thesis. Most of the information has been gathered from the online documentation of the Mlr package (Bischl, 2015). In some cases, additional information was sought from the documentation of the packages hosted on the CRAN repository for R packages.

For each model we document the following information

- The model identifier
- A brief description of the model
- The model class
- The supported modelling task (classification and/or regression)

The overview of statistical models is structured around the following model classes:

- Boosting models
- Decision tree models
- Discriminant analysis models
- Generalized linear models
- Nearest neighbors models
- Neural network models
- Support vector machine models
- Other models

Some models fall under multiple model classes. These models are documented for the model class which comes first in the alphabetical order.

9.2.1 Boosting models

Model identifier	Description	Type	Task
ada	Decision trees boosted using the Ada meta-algorithm	Boosting, Decision trees	Classification
blackboost	Gradient boosting with regression trees	Boosting, Decision trees	Classification, Regression
gbm	Gradient boosted decision trees	Boosting, Decision trees	Classification, Regression
xgboost	Gradient boosting (as implemented in the xgboost-package)	Boosting, Decision trees	Classification, Regression
glmboost	Boosted generalized linear models	Boosting, Generalized linear model	Classification

9.2.2 Decision tree models

Model identifier	Description	Type	Task
bartMachine	An implementation of the Bayesian Additive Regression Trees algorithm	Decision trees	Classification, Regression
cforest	Random forest of conditional inference trees	Decision trees	Classification, Regression
ctree	Conditional inference tree model	Decision trees	Classification, Regression
extraTrees	Ensemble of extremely randomized decision trees	Decision trees	Classification, Regression
J48	J48 decision trees	Decision trees	Classification
randomForest	Random forest of decision trees	Decision trees	Classification, Regression
randomForestSRC	Random forest of decision trees (as implemented in the randomForestSRC-package)	Decision trees	Classification, Regression
rpart	Decision tree model	Decision trees	Classification, Regression

9.2.3 Discriminant analysis models

Model identifier	Description	Type	Task
lda	Linear discriminant analysis	Discriminant analysis	Classification
mda	Mixture discriminant analysis	Discriminant analysis	Classification
plsdaCaret	Partial least squares discriminant analysis	Discriminant analysis	Classification
qda	Quadratic discriminant analysis	Discriminant analysis	Classification
rda	Regularized discriminant analysis	Discriminant analysis	Classification
sda	Shrinkage discriminant analysis	Discriminant analysis	Classification

9.2.4 Generalized linear models

Model identifier	Description	Type	Task
binomial	Binomial regression	Generalized linear model	Classification
glmnet	Generalized linear models with Lasso or Elasticnet regularization	Generalized linear model	Classification, Regression
LiblineaRLogReg	Logistic regression (as implemented in the LiblinearR-package)	Generalized linear model	Classification
lm	Simple linear regression	Generalized linear model	Regression
logreg	Logistic regression (as implemented in base R)	Generalized linear model	Classification
lqa	Penalized generalized linear models with the LQA algorithm	Generalized linear model	Classification
multinom	Multinomial regression	Generalized linear model	Classification
penalized.lasso	Lasso-regularized linear regression	Generalized linear model	Regression
penalized.ridge	Ridge-regularized linear	Generalized	Regression

	regression	linear model	
plr	Logistic regression with a L2 penalty	Generalized linear model	Classification
probit	Probit regression	Generalized linear model	Classification

9.2.5 Nearest neighbors models

Model identifier	Description	Type	Task
fnn	Fast k-nearest neighbors model	Nearest neighbors	Classification, Regression
Ibk	K-nearest neighbors (as implemented in the Rweka-package)	Nearest neighbors	Classification, Regression
kknn	K-nearest neighbors	Nearest neighbors	Classification, Regression

9.2.6 Neural network models

Model identifier	Description	Type	Task
brnn	Feed-forward neural network with Bayesian regularization	Neural network	Regression
elmNN	Extreme learning machine for single hidden layer feedforward neural networks	Neural network	Regression
nnet	Single-hidden layer neural network	Neural network	Classification, Regression

9.2.7 Support vector machine models

Model identifier	Description	Type	Task
ksvm	Support vector machine (as implemented in the kernlab-package)	Support vector machine	Classification, Regression
rvm	Relevance vector machine	Support vector machine	Regression
svm	Libsvm-based support vector machines (as implemented in the	Support vector	Classification, Regression

| | e1071-package) | machine | |

9.2.8 Other models

Model identifier	Description	Type	Task
bdk	Supervised version of Kohonen's self-organising map	Other	Classification, Regression
cubist	Rule- and instance-based regression modeling	Other	Regression
earth	Regression model based on Friedman's Multivariate Adaptive Regression Splines-procedure (as implemented in the Earth-package)	Other	Regression
Jrip	Propositional rule learner	Other	Classification
mars	Regression model based on Friedman's Multivariate Adaptive Regression Splines-procedure (as implemented in the mda-package)	Other	Regression
naiveBayes	Naive Bayes	Other	Classification
OneR	Rule-based OneR-classifier	Other	Classification
pcr	Principal component regression	Other	Classification
plsr	Partial least squares regression	Other	Regression
rsm	Response surface regression	Other	Regression
xyf	X-Y-fused self-organising maps	Other	Classification, Regression

9.3   Appendix 3: Comparison of regression models

Model identifier	Model class	RMSE	MAE
bartMachine	Decision trees	10.47276	8.725286
cforest	Decision trees	10.4931	8.761483
randomForestSRC	Decision trees	10.50747	8.797584
randomForest	Decision trees	10.53103	8.829922
mars	Other	10.53293	8.765768
earth	Other	10.54626	8.797766
extraTrees	Decision trees	10.58358	8.841442
rpart	Decision trees	10.60242	8.882099
ctree	Decision trees	10.61077	8.87554
blackboost	Boosting, Decision trees	10.62642	9.009433
brnn	Neural network	10.6554	8.945144
svm	Support vector machine	10.6865	8.676228
glmnet	Generalized linear model	10.69469	9.011372
lm	Generalized linear model	10.70507	9.017976
pcr	Other	10.70507	9.017976
penalized.lasso	Generalized linear model	10.70507	9.017976
penalized.ridge	Generalized linear model	10.70507	9.017976
rsm	Other	10.70507	9.017976
plsr	Other	10.70507	9.017976
cubist	Other	10.75295	8.824273
gbm	Boosting, Decision trees	10.80042	9.202922
ksvm	Support vector machine	10.9807	8.888162
nnet	Neural network	11.09808	9.15241
rvm	Support vector machine	11.24974	9.168726
xgboost	Boosting, Decision trees	11.3975	9.241743
kknn	Nearest neighbors	11.59112	9.430867
fnn	Nearest neighbors	12.05799	9.744438
elmNN	Neural network	13.31599	10.64525
xyf	Other	13.53085	10.48837
lbk	Nearest neighbors	14.27701	11.13187
bdk	Other	14.28295	10.94394

9.4 Appendix 4: Comparison of classification models

Model	Model class	AUC	F1
bartMachine	Decision trees	0.645944	0.148382
cforest	Decision trees	0.641541	0.050029
ada	Boosting, Decision trees	0.628138	0.214354
gbm	Boosting, Decision trees	0.622422	0
glmboost	Boosting, Generalized linear model	0.622063	0.028447
randomForestSRC	Decision trees	0.621416	0.147259
lqa	Generalized linear model	0.621375	0.031587
extraTrees	Decision trees	0.62086	0.190735
glmnet	Generalized linear model	0.620377	0.003584
LiblineaRLogReg	Generalized linear model	0.619445	0.041137
plsdaCaret	Discriminant analysis	0.619124	0.021282
sda	Discriminant analysis	0.618551	0.090294
lda	Discriminant analysis	0.61802	0.04446
randomForest	Decision trees	0.617878	0.139401
rda	Discriminant analysis	0.61742	0.07406
binomial	Generalized linear model	0.617071	0.044323
logreg	Generalized linear model	0.617071	0.044323
multinom	Generalized linear model	0.617071	0.044323
plr	Other	0.617071	0.044323
mda	Discriminant analysis	0.616641	0.050928
probit	Generalized linear model	0.616078	0.034424
naiveBayes	Other	0.595799	0.463711
blackboost	Boosting, Decision trees	0.595557	0
xgboost	Boosting, Decision trees	0.582825	0.269461
nnet	Neural network	0.579265	0.107965
xyf	Other	0.560204	0.287946
ctree	Decision trees	0.559838	0
qda	Discriminant analysis	0.555001	0.456675
ksvm	Support vector machine	0.553753	0.010914
J48	Decision trees	0.549857	0.076334
kknn	Nearest neighbors	0.548904	0.300508
svm	Support vector machine	0.541358	0

Ibk	Nearest neighbors	0.53576	0.34486
bdk	Other	0.530486	0.240179
OneR	Other	0.51079	0.242459
Jrip	Other	0.509499	0.089676
rpart	Decision trees	0.5	0

9.5 Appendix 5: General comparison of scoring models

Model	Task	Model class	Tau	True Positive Rate
rpart	Regression	Decision trees	0.186354	0.38576779
blackboost	Regression	Boosting, Decision trees	0.176081	0.411985019
cforest	Regression	Decision trees	0.171155	0.402621723
earth	Regression	Other	0.170652	0.411985019
nnet	Classification	Neural network	0.169458	0.391385768
gbm	Regression	Boosting, Decision trees	0.168886	0.402621723
bartMachine	Regression	Decision trees	0.168832	0.402621723
mars	Regression	Other	0.1675	0.41011236
cforest	Classification	Decision trees	0.165969	0.419475655
svm	Regression	Support vector machine	0.16539	0.402621723
cubist	Regression	Other	0.164147	0.383895131
bartMachine	Classification	Decision trees	0.163018	0.417602996
randomForestSRC	Regression	Decision trees	0.162753	0.380149813
ctree	Regression	Decision trees	0.160068	0.368913858
gbm	Classification	Boosting, Decision trees	0.157742	0.393258427
randomForest	Regression	Decision trees	0.157041	0.393258427
extraTrees	Regression	Decision trees	0.150007	0.391385768
brnn	Regression	Neural network	0.149592	0.393258427
plsdaCaret	Classification	Discriminant analysis	0.147851	0.387640449
glmboost	Classification	Boosting, Generalized linear model	0.145727	0.400749064
glmnet	Classification	Generalized linear model	0.145566	0.389513109
lqa	Classification	Generalized linear model	0.145338	0.400749064
sda	Classification	Discriminant analysis	0.14291	0.400749064

ada	Classification	Boosting, Decision trees	0.142595	0.398876404
glmnet	Regression	Generalized linear model	0.142113	0.38576779
lm	Regression	Generalized linear model	0.14161	0.383895131
pcr	Regression	Other	0.14161	0.383895131
penalized.lasso	Regression	Generalized linear model	0.14161	0.383895131
penalized.ridge	Regression	Generalized linear model	0.14161	0.383895131
plsr	Regression	Other	0.14161	0.383895131
rsm	Regression	Other	0.14161	0.383895131
rda	Classification	Discriminant analysis	0.141099	0.393258427
LiblineaRLogReg	Classification	Generalized linear model	0.140477	0.397003745
extraTrees	Classification	Decision trees	0.140175	0.38576779
lda	Classification	Discriminant analysis	0.140101	0.400749064
binomial	Classification	Generalized linear model	0.138924	0.393258427
logreg	Classification	Generalized linear model	0.138924	0.393258427
plr	Classification	Other	0.138924	0.393258427
multinom	Classification	Generalized linear model	0.138917	0.393258427
probit	Classification	Generalized linear model	0.137459	0.393258427
randomForestSRC	Classification	Decision trees	0.136367	0.400749064
mda	Classification	Discriminant analysis	0.133426	0.387640449
nnet	Regression	Neural network	0.132476	0.391385768
blackboost	Classification	Boosting, Decision trees	0.132293	0.389513109
randomForest	Classification	Decision trees	0.13219	0.387640449

ksvm	Regression	Support vector machine	0.125314	0.383895131
rvm	Regression	Support vector machine	0.123869	0.383895131
naiveBayes	Classification	Other	0.118688	0.397003745
ctree	Classification	Decision trees	0.110147	0.367041199
xgboost	Regression	Boosting, Decision trees	0.108525	0.36329588
bdk	Regression	Other	0.10088	0.352059925
xgboost	Classification	Boosting, Decision trees	0.094747	0.359550562
qda	Classification	Discriminant analysis	0.093285	0.342696629
kknn	Regression	Nearest neighbors	0.091166	0.36329588
kknn	Classification	Nearest neighbors	0.082241	0.337078652
fnn	Regression	Nearest neighbors	0.070011	0.31835206
J48	Classification	Decision trees	0.06582	0.209737828
Ibk	Classification	Nearest neighbors	0.065395	0.348314607
bdk	Classification	Other	0.061036	0.312734082
ksvm	Classification	Support vector machine	0.0603	0.335205993
svm	Classification	Support vector machine	0.056184	0.323970037
Ibk	Regression	Nearest neighbors	0.056173	0.333333333
xyf	Regression	Other	0.050992	0.331460674
xyf	Classification	Other	0.047689	0.335205993
Jrip	Classification	Other	0.044212	0.234082397
Altman's Z-score 2012	Baseline	Baseline	0.040708	0.288389513
Equity ratio 2012	Baseline	Baseline	0.037779	0.279026217
Laitinen's Z-score 2012	Baseline	Baseline	0.022465	0.299625468
OneR	Classification	Other	0.01671	0.342696629
Random prediction	Baseline	Baseline	0.014636	0.303370787

rpart	Classification	Decision trees	0	0
elmNN	Regression	Neural network	-0.07531	0.213483146

9.6 Appendix 6: Comparison of ensembles and the best base learners

Model	Task	Model class	Tau	True Positive Rate
ensemble_all	Ensemble	Ensemble	0.193389	0.43071161
ensemble_two	Ensemble	Ensemble	0.191388	0.41011236
rpart	Regression	Decision trees	0.186354	0.38576779
ensemble_reg	Ensemble	Ensemble	0.186166	0.423220974
ensemble_clf	Ensemble	Ensemble	0.185622	0.434456929
blackboost	Regression	Boosting, Decision trees	0.176081	0.411985019
cforest	Regression	Decision trees	0.171155	0.402621723
earth	Regression	Other	0.170652	0.411985019
nnet	Classification	Neural network	0.169458	0.391385768
gbm	Regression	Boosting, Decision trees	0.168886	0.402621723
bartMachine	Regression	Decision trees	0.168722	0.402621723
cforest	Classification	Decision trees	0.165969	0.419475655
bartMachine	Classification	Decision trees	0.164433	0.413857678
gbm	Classification	Boosting, Decision trees	0.157742	0.393258427
plsdaCaret	Classification	Discriminant analysis	0.147851	0.387640449
glmboost	Classification	Boosting, Generalized linear model	0.145727	0.400749064

9.7 Appendix 7: Technical description of analytics pipeline

Based on the work done for this thesis, we established a first proposal for an analytics pipeline for customer selection. In the diagram below, the pipeline is represented as a sequence of R-scripts.

```
                 Predictor variables
                                                            Response variable
    ┌──────────────────────────────┐
    │  corporate_data_tables.R     │          ┌──────────────────────────────┐
    └──────────────────────────────┘          │      corporate_scoring.R     │
                    │                          └──────────────────────────────┘
                    ▼
    ┌──────────────────────────────┐
    │   data_towide_impute.R       │
    └──────────────────────────────┘

                              Scoring model

              ┌────────────────────────────────────────┐
      ───────▶│                Dataset                 │◀───────
              └────────────────────────────────────────┘
                                 │
                                 ▼
                ┌──────────────────────────────────┐
                │       feature_selection.R        │
                └──────────────────────────────────┘
                                 │
                                 ▼
                ┌──────────────────────────────────┐
                │       model_comparisons.R        │
                └──────────────────────────────────┘
                                 │
                                 ▼
                ┌──────────────────────────────────┐
                │      model_interpretation.R      │
                └──────────────────────────────────┘
                                 │
                                 ▼
                ┌──────────────────────────────────┐
                │       write_predictions.R        │
                └──────────────────────────────────┘
```
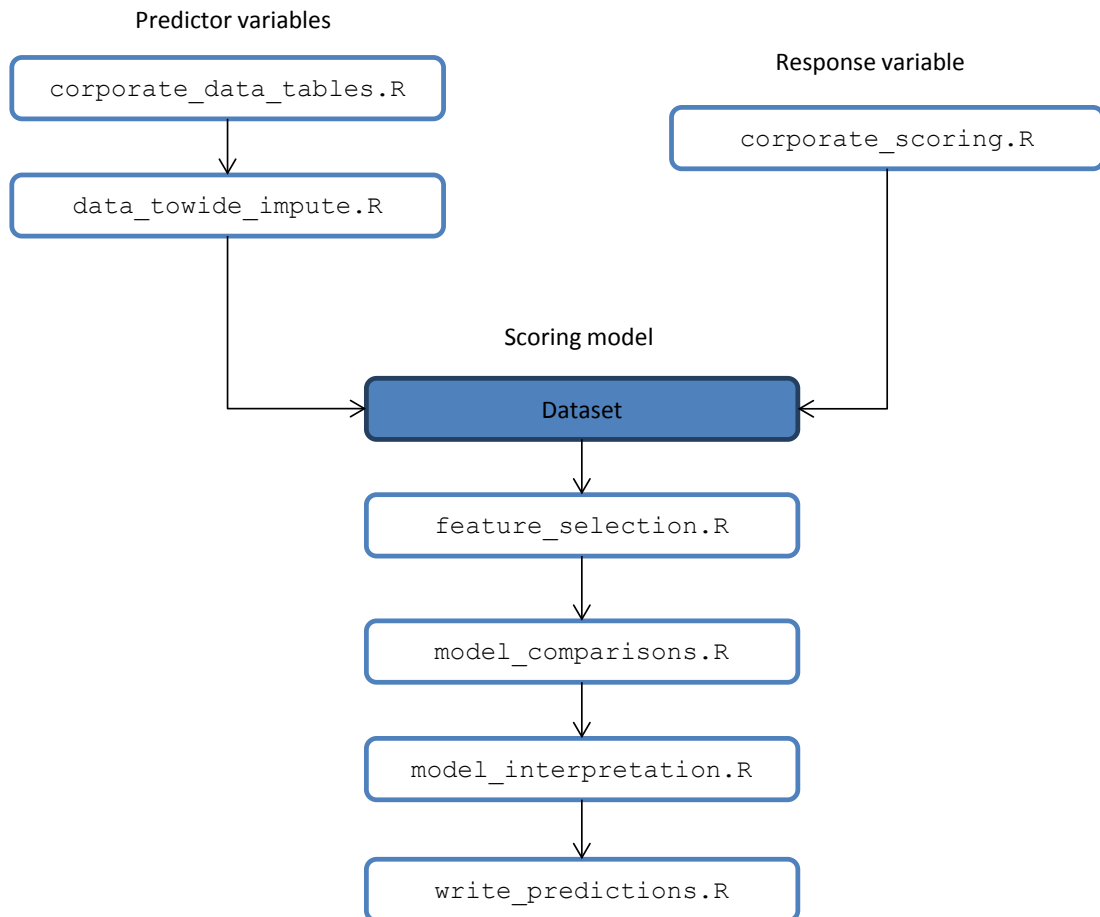
In the following we briefly describe the tasks performed by each of the seven scripts. The discussion is divided into two parts: building the dataset and building the scoring model.

9.7.1 Building the dataset

9.7.1.1 corporate_data_tables.R

The script queries and aggregates the data in the Voitto + tables in Aktia's data warehouse. There are two Voitto + tables: one table with complete financial statements and another with financial ratios. For the table with financial statement information, we pick the desired rows from the financial statements for the relevant years. For the table with financial ratios, the information is in columnar format, and we pick the desired columns. In the end, we generate a table where each company has one row of observations for each year. We use the ROracle-library to connect to Aktia's databases.

9.7.1.2 *data_towide_impute.R*

After running corporate_data_tables.R, our data is in long format, where each company has one row of observations for each year. We use R's reshape2 package to turn the data into fully wide format, where each company only has one row of observations and the year observed is indicated in the name of the column. In addition to transforming the data to wide format, the script also performs an imputation of missing values using the mice-package and its predictive mean matching-method of mean imputation. Here, we noticed that the imputation only successfully completes for years 2009-2012. Hence, we chose to limit our dataset to these years.

9.7.1.3 *corporate_scoring.R*

With the corporate_scoring.R-script, we read the relevant internal data on Aktia's own customers to determine the customer scores. Concretely, this entails querying profitability data, information on the primary bank status of customers, and information on the number of product categories bought by a customer. Based on this information, we compute the customer scores, as explained in 4.4.2. In this script, we also use the ggplot2-library to plot the utility curves displayed in section 4.4.2 (Figure 4)

9.7.2 Building the scoring model

We generate our dataset by joining the tables generated by the scripts in section 9.7.1. Based on this dataset, we fit our scoring models, and eventually perform the intended scoring of potential new customers.

9.7.2.1 *feature_selection.R*

Using R's Boruta package, we perform a random forest-based feature selection for our dataset, for both a classification and a regression formulastion of our customer selection problem. We store the results of the feature selection for later use.

9.7.2.2 *model_comparisons.R*

We select the features considered relevant by the Boruta-procedure, and fit a wide variety of classification and regression models to our data using the mlr-package. We output and store diagnostic information for both the classification and regression formulations, and the more general problem of ranking customers. We choose a handful of the best models according to Kendall's Tau-metric and ensemble these models using a simple linear voting ensemble. Again, the diagnostic information is saved on disk.

9.7.2.3   model_interpretation.R

We choose two well-performing and highly interpretable models (the cforest random forest-model, and the rpart decision tree-model), and interpret their results. For the rpart-model we plot its suggested decision tree. For the cforest-model, we plot the feature importances of the predictor variables. Additionally, we plot the results of the Boruta feature selection procedure.

9.7.2.4   write_predictions.R

For the first run of our scoring model, the ensemble of the 12 best-performing models was found to have the best performance. Hence, it was chosen as the final decision model for this initial run. In the write_predictions.R-script, we fit these 12 models on all of the available training data, compute predictions for all the non-Aktia firms in the test data, and ensemble the predictions using the simple linear voting model. The predictions are written to Aktia's database.