Real-time classification of SMEs credit and risk ratings and the impact of financial indicators and payment behaviour

Teemu Samuli Tuovinen

Bachelor's Thesis
Degree Programme in
International Business
2020

Abstract

23.5.2020

Author(s)
Teemu Samuli Tuovinen

Degree programme
International Business and Business Administration

Report/thesis title	**Number of pages and appendix pages**
Real-time classification of SMEs credit and risk ratings and the impact of financial indicators and payment behaviour	**59 + 10**

This thesis studies the applicability and the benefits of real-time data in the forecasting of financial risk and credit ratings of Finnish SMEs and the impact of payment behaviour to ratings. The research aims to analyse if the data of an accounting firm could improve reliability and accuracy when evaluating the probability of defaulting and the financial risk of a business. In order to achieve the objective, the development of the process, methods, and algorithm are needed. The development of an automated credit rating process is time-consuming and requires resources as accuracy and reliability are fundamentals of classification. Credit and risk ratings explain the overall financial status and risk of the entity, which makes it complicated to determine the accurate indicators of the financial status of a business. This thesis explains the theory behind the phenomenon of credit and risk rating methods when analytics and accounting are combined.

Both quantitative and qualitative elements have a significant role in the rating process and methods. Qualitative variables measure the subjectivity of the business operations as quantitative variables measure the subjectivity of the financials. Hypotheses were set according to the results of the past studies related to the matter, and data analysis authenticated the operability with the multinomial Pearson correlation coefficient and supervised logistic regression classification model. The relation of the results was analysed against the ratings of rating agencies and past research in order to evaluate the operability of the classification model in a business case scenario.

The results confirmed that real-time data increases the quality and accuracy of financial information, but also that the data can be applied in the rating processes. The impact of the additional information implemented in the rating methods was partly beneficial as an increase in payments made late was seen as an explanatory variable when evaluating the risk. The result indicates that debit entries of clearing account in relation to all transactions do not have an impact on the riskiness or creditworthiness of the entity. The results differ from the past research and guidelines of credit rating agencies in the sense that not all the financial variables examined are significant when evaluating the risk and creditworthiness of an SME. The rating process and algorithms built according to the results are a basis that will be improved into a Neural Network rating model in the future.

Keywords
Credit rating, Financial statement analysis, Real-time data, Real-time credit rating, Financial risk, Defaulting.

Table of contents

1 Introduction

The following chapter explains the purpose of this thesis, the reason that led to the research and key concepts covered in this thesis. The chapter will include all the relevant information about the thesis, such as research problems, research questions, demarcation, and information about the company that will implement the results of the research in their processes, and if the results are seen beneficial in business case evaluation.

1.1 Background

This thesis is done upon the request of Company X, which is an accounting firm operating in Country A and Country B. Accounting firms can access financial information that no other party can, at any point of the fiscal year. Company X ordered this research, which aims to study the possible benefits for the company, its customers, and stakeholders that real-time data encase when implemented in the rating processes. Possible benefits researched are the increase in forecasting accuracy of financial risk indicators and an improved rating model.

Credit ratings are often criticised in media. There is no certainty of the trustworthiness of the data that is used in credit rating. For example, there has been news of firms that have gotten A+/AAA credit rating, and there have not been any operating activities, and even the stakeholders have stated that the company is not creditworthy. As credit ratings are rated with data that consists of publicly available information, it is argued that there is a lack of information when evaluating the risk. The situation provides an opportunity for accounting firms to the benefit of the real-time data that they have. The problem with the implementation of real-time data is that the benefits have not been researched yet.

Time sensitivity in the delivery of the data used in a credit rating is problematic as financial information does not provide an adequate understanding of the current financial status of the entity as if it is outdated. The pandemic is an excellent example as it has caused a sudden stop of cash flow for specific industries and businesses. Costs have had to been cut with layoffs and other efforts, and some businesses have already gone bankrupt within the first three months. As cashflow has reduced wholly or significantly, the need for external financing is more prominent than anyone would have predicted because of the fixed costs than cannot be avoided. The basic rule for financing is that money that is arranged has a price on it. The price is commonly set according to a given credit rating, but it can be argued if the credit rating that is based on historical data tells about the company's current financial situation and risk, for example, in the times of pandemic and financial crisis.

Credit ratings have a vital role in the economy. Credit ratings measure the relative credit-worthiness of companies and are an objective way to separate risky and safe businesses, in terms of financing the operations of the company. (Gonzalez & al. 2004, 4.) Credit rating is also a way to determine the company's financial strength from a long-term point of view, as the process of the rating tends to focus on the long-term solvency of a debtor (Horrigan 1966, 45). Low credit rating correlates with the company's reduced ability to settle their payments, according to the statistics in Creighton's research (Creighton 2014, 120).

Researches about credit ratings have a long history, and it has been a highly studied subject. Although many of the researches and studies have been successful, Shon and Kim claim that it is not bright or determined which financial indicators and financial information should be used when evaluating the company's credit risk (Shon & Kim 2012, 932). Accounting firms hold a massive amount of real-time, exclusive data, which can be used as a quantitative indicator when determining the credit rating, in order to explain the real-time financial status of businesses. This thesis will examine the possible benefits of financial information and real-time data when applied to the credit and risk rating processes and the impact of payment behaviour on the real-time rating.

1.2 Research question

The thesis aims to research if real-time data, enriched with additional data of the commissioning company, could improve automated forecasting of financial risk and credit ratings. In order to seek improvement in the credit rating process, it is vital to understand how the riskiness of businesses is currently evaluated and which are the relevant indicators of financial risk. It is essential to inspect the data used in ratings properly and understand the overall status of Finnish SMEs on different matters. Creditworthiness evaluates the company's ability to settle their payments and forecasting of risk evaluates the probability of bankruptcy. In this research, the evaluation and forecasting are done with ratings in order to avoid possible credit losses and evaluate the risk correctly to improve internal processes and services provider for the customers of commissioning company X. Research question and investigative questions are set as:

- RQ: Could the implementation of real-time data improve the automated credit and risk rating methods?
- IQ1: What are the problems in current credit and risk rating processes?
- IQ2: How would it be possible for the commissioning company to determine credit and risk ratings?
- IQ3: Could real-time financial data and payment behaviour benefit the evaluation of the risk and creditworthiness of SMEs?

The aim is approached to achieve with examining if an accounting firm can develop a simple indicator that explains the financial status and risk as simple ratings. Financial ratios that have been chosen for this study are chosen accordingly to information that Finnish credit rating agencies Asiakastieto (2020) and Alma Talent (2020) have stated. The aim is to understand the financial health of the entity trough financial indicators and payment behaviour through results of statistical analysis. The statistical analysis estimates the extent of dependency of variables of financial indicators and payment behaviour to ratings. Both the influence of financial ratios and payment behaviour to ratings are analysed to evaluate and validate the operability. The hypotheses guiding the empirical analysis are set out in Chapter Two, with the examination of past studies and researches. Hypotheses are set as follows:

- H1: The business industry influences the credit and risk ratings
- H2: Profitability, solvency, and liquidity all have a significant role in credit and risk ratings
- H3: Company size impacts credit and risk ratings
- H4: Increase in the balance of clearing account has a detrimental influence on credit and risk ratings
- H5: Defaults and late payments influence credit and risk ratings

1.3 Demarcation

The impact is purely studied from the statistical point of view. The explanation is evaluated according to the results of analysis and past studies. The research will examine the relevant factors in order to provide an understanding of the possible solution for the problem. This thesis concentrates on providing the suggested solution for matter in question to an accounting firm. Financial data on this research is from SMEs that operate in Finland. All the companies are chosen from construction (Industry code 41-43) and accommodation and food service activities (industry code 55-56) industry. Accounting of the SMEs in this research is done on an accrual basis. In this research, the size of the company is determined according to revenue and number of employees. The data is formed and presented in a way that ensures the anonymity of the companies. All the businesses analysed in this thesis have operated for a full 12 months on fiscal from 1.1.201x to 31.12.201x.

The sample data have been divided into two categories. The sample that is used on examination related to credit rating includes 2790 SMEs from industries 41-43 and 55-56 that are customers of Commissioning company X. The data used in statistics of payment behaviour consist of 6113 SMEs from many different industries and includes 14.3 million purchase invoices. Ratings compared to real-time ratings in this thesis are provided by the market leaders on the rating industry in Finland. All the SMEs have been chosen randomly, even though the financial information is not publicly available. Payment behaviour is studied with a large dataset of purchase invoices and debit entries of clearing accounts. Examination of payment behaviour is research with analysis of late payments in relation to all purchase invoices and with debit entries of clearing account in relation to all transactions. This thesis is studying the financial indicators according to the information that Finnish credit rating agencies have given, and the indicators are validated with past researches and studies. Comparison is made for the datasets and ratings of the commissioning company and the chosen rating agencies that provided the data for researching the matter.

1.4 Key concepts

Credit rating in this research estimates the ability of the settlement of payments.
Kagan has stated credit rating in a simple term as a quantified assessment of the creditworthiness that is assigned by an external party (Kagan, 2020).

Forecasting can be done in several ways. In this thesis, the forecasting will be done statistically, and the data will be collected from the financial statements of SMEs. Forecasting is done with both horizontal and vertical analyses from financial data, and the aim is to get the basis to understand what kind of scenarios would be likely to happen in the future. The forecasting is usually done ahead for a certain period; in this research, the period is a fiscal year (12 months) or other if needed. (Investopedia 2019.)

Financial risk indicators and ratios – There are many ways to calculate and analyse the financial health of the company and evaluate the risk. In this thesis, financial ratios are chosen according to given information of large credit rating agencies to ensure comparability. Chosen financial indicators are Revenue Growth %, Operating Margin %, ROCE, ROA, Current Ratio, Quick Ratio, Turnover of Receivables, Equity Ratio, Net Gearing, and Relative Indebtedness. (Asiakastieto 2020; Alma Talent 2020.)

Defaulting is a failure in meeting the obligations or terms that are agreed on related to the payment of a debt, bill, or goods (Laitinen 2004, 110-114).

The clearing account is a temporary account used in accounting. Clearing account is used for the entries of the transactions that relate to the documents that do not meet the regulations stated in Accounting Act 4.6.2001/1653. Debit entries made on clearing account increases and credit entries decrease the balance of the account. (KILA 4.6.2001/1653.)

Real-time data (RTD) in this research is Big Data with velocity. It covers a variety of data related to financial information and payment information. The data is collected from databases and formed into the financial ratios. Real-time data is data that is delivered immediately after the collection. (Ylijoki & Porras 2016, 70-79.) RDR is defined as the data that is collected immediately after the delivery. RDR includes as current information as possible, which is required at the point of inspection.

"**Big data** refers to the growth in the volume of structured and unstructured data, the speed at which it is created and collected, and the scope of how many data points are covered. Big data often comes from multiple sources and arrives in multiple formats." – (Investopedia 2017.)

Pearson correlation coefficient (PCC) represents the continuous linear relationship between two variables that can be measured on a standard scale. This research measures the relation between explanatory factors in relation to credit and risk ratings. (Kenston 2020.)

A negative correlation is a negative linear relationship between two variables, which means that when the value of other variable increases, the other variables value decreases in relation. (Investopedia 2020.)

Real-time credit and risk ratings are concepts that will be formed in this research. Real-time rating is formed with real-time data and classified according to industry-specific algorithms formed with a supervised logistic regression classification model. Real-time credit rating measures the capability of the settlement of payments. Real-time risk rating measures the risk related to bankruptcy. (Gropelli, Angelico & Nikbakht 1990, 445-451.)

1.5 Commissioning company

Commissioning company X is a rapidly growing publicly listed company operating in the field of accounting in countries A and B. They offer a large variety of services, which include financial-, HR, legal- and digital services. The company has roughly a 5% market share in the industry in country X. They have over 700 employees, and their revenue in the last fiscal year was 60 million euros. The profit of the company was 8,5 million, which is 17.5% of its revenue. Roughly 90% of its revenue consists of accounting services that are the primary services of the company. The commissioning company is an innovative firm that continually seeks new opportunities in order to serve its customers on an exceptional level in the quality of the service.

2 Theoretical framework

In the thesis, there are three different theoretical concepts, credit rating, financial risk indicators, and real-time data. This chapter will introduce the main concepts and theories related to the research problem. The purpose of the theoretical framework is to guide the quantitative and qualitative analysis to be aimed towards the explanation of fundamentals behind the phenomenon (Uusitalo 1991, 63-64). The relation of critical theories is presented in figure 1.

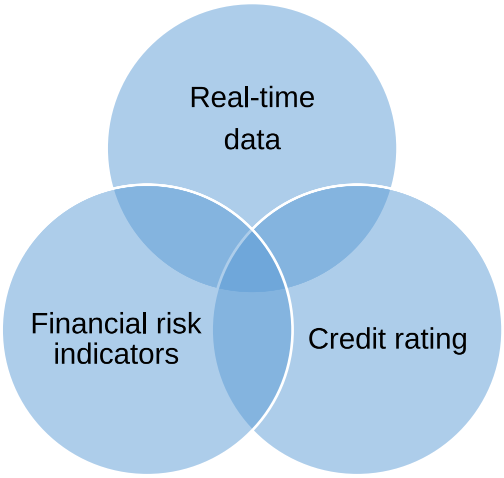


Figure 1. Illustration of critical theories of this thesis

As the thesis will examine the financial risks of Finnish SMEs, it is to be defined what a Finnish SME is. There are many definitions of SME. Finnish Accounting Act and Statistics Finland has stated clearly the definitions and indicators for an SME. The commissioning company has more than 10,000 customers that fulfil the requirements of an SME. This means that the company has every single accounting document that their customers deliver and additional, non-financial information, which together forms real-time data with massive volume, velocity, and variety. As data is carefully analysed with different kinds of indicators and parameters, it is possible to categorise customers due to their financial risk and field of industry and compare the results with set credit and risk ratings, rated by leading credit rating companies that operate in Finland.

2.1 What is an SME?

According to the Finnish Accounting Act Section 1:1a (30.12.2015/1620.) regulates the obligations of a natural person to keep accounting records as follows:

- A natural person, except for one pursuing farming or fishery, shall keep accounting records for the business and profession carried on. The obligation shall also apply to the death or bankruptcy estate continuing the activities of a person carrying on a profession or a business.

- Entries shall be made of transactions in a manner that the reporting entity can establish the number of trade creditors and debtors at all times, and information necessary to comply with tax obligations can be derived from the accounts. Notwithstanding the provisions of section 2 of this chapter, a person carrying on a profession or a business shall not be obliged to apply double-entry bookkeeping, where at the maximum one the following three thresholds were exceeded in the previous financial year and the one immediately preceding it:
- total assets in excess of 100,000 euros;
- turnover or comparable income in excess of 200,000 euros;
- average personnel amounting to more than three.
- There are many ways to define an SME. One of the definitions of an SME follows the regulations of small undertakings in the Finnish Accounting Act Section 1:4a (30.12.2015/16209).
- In this Act, a small undertaking refers to a reporting entity exceeding at the maximum one the following three thresholds at the balance sheet date of the previous financial year and the one immediately preceding it.
- total assets 6,000,000 euros;
- net turnover 12,000,000 euros;
- the average number of employees during the financial year 50.
  (Finnish Accounting Act 1:1a 30.12.2015/1620.)

Statistics Finland (2003) states that small and medium-sized enterprises (SMEs) are defined as enterprises which have fewer than 250 employees and have either an annual turnover not exceeding EUR 50 million or an annual balance sheet total not exceeding EUR 43 million and which conform to the criterion of independence as defined below.

Independent enterprises are those who are not owned as to 25 per cent or more of the capital or the voting rights by one enterprise, or jointly by several enterprises, falling outside the definition of an SME or a small enterprise, whichever may apply. The definition is valid until 31.12.2078. (Tilastokeskus 2003.) Statistics Finland (2003) states that small and medium-sized enterprises (SMEs) are defined as enterprises which have fewer than 250 employees and have either an annual turnover not exceeding EUR 50 million or an annual balance sheet total not exceeding EUR 43 million and which conform to the criterion of independence as defined below.

2.2 Credit rating and evaluation of the risk

2.2.1 Rating and credit risk

There are many definitions for credit rating, from different authors, but the central concept can be the same in the literature that defines it. Credit rating is typically needed to be determined when a company is applying for financing. The creditor will assess a price for the loan. The lending process, according to Altman (1980, 813-814), consists of four steps:

- Application for a loan
- Credit rating analysis
- Evaluation of a loan
- Payback of a loan

When discussing the credit rating, we must look at step 2, which evaluates the company's credit risk. Credit risk is illustrated with an investment-grade table that separates companies with different risks in different grades. (see table 1) The division is generally done by banks or rating agencies (Treacy & Carey 2000, 167-169).

Table 1. Gradings of Moody's and S&P as grades (Treacy & Carey 2000, 167)

Moody's	S&P	
Aaa	AAA	Investment Grade
Aa1	AA+	
Aa2	AA	
Aa3	AA-	
A1	A+	
A2	A	
A3	A-	
Baa1	BBB+	
Baa2	BBB	
Baa3	BBB-	
Ba1	BB+	Non-Investment Grade
Ba2	BB	
Ba3	BB-	
B1	B+	
B2	B	
B3	B-	

A credit rating is an evaluation completed by an external party. The purpose of the evaluation is to give the reader an understanding of the financial performance, but also about the risks involved in the financial status of the entity. Usually, the evaluation is a financial statement analysis, especially when considering SMEs, but it can also include analysing future expectations and probabilities. (Kronwald & Christian 2009, 40.) Another definition of credit rating, according to Laitinen (2005, 76-79) is to evaluate the probability of the likelihood of default, which in this case means that the creditor will not be able to meet the legal obligations of the whole balance of the debt or the payment is delivered late. Another measure is to evaluate the possible credit losses that might occur.

It is vital to notice that credit rating and financial statement analysis are defined differently and cannot be discussed as synonyms. A credit rating includes a combination of quantitative and qualitative information. Proper credit rating is determined according to all the relevant information related to the financial status and risk of the entity. (Servigny & Renault

2004, 25.) Information is gathered from multiple sources, including the financial statement and auditor's comments, but also more information that cannot be found from the financial statement. Credit rating cannot replace auditing or auditing to determine the credit rating. (Feldmann & Read 2013, 320.)

2.2.2 Insolvency

When it comes to determining the parameters for the rating, according to Krahnen & Weber (2001, 21-23), it is necessary to define the probability of default. A credit rating is an evaluation process that sets probabilities for solvency and insolvency. Solvency is defined as the degree to which the total current assets of an entity exceed the total current liabilities of an entity. In other words, the entity can meet their legal obligations in-full. (Ross, Westerfield & Jordan 2020, 67-69.) Researches related to accounting, financial statement analysis, and insolvency has introduced many definitions for insolvency.

Altman's research about insolvency concentrates on the worst form of financial difficulties, the bankruptcy. Researches have shown that delayed payments and defaults of debt, which are the indicators of financial difficulties, can be used as a measure when forecasting the insolvency. (Altman 1968, 601.)

2.2.3 Short-term and long-term ratings

The basis of the credit rating is to give a clear image of the comprehensive financial state of the business. A trustworthy credit rating should provide a solid understanding of financial health and the risk of the company on an overall level. Although this thesis examines the changes and impact of real-time data in the short-term ratings, it is essential to consider and understand if the application of long-term credit rating methods is possible. The ratings are divided into short term rating and long-term rating:

Short-term rating

Short-term rating is commonly done automatically, and it provides the rating by evaluating and comparing the data from different sources. Primary data for this kind of analysis are payment data, defaults, financial statements, necessary information about the company, such as when the company is established and who are the persons that are running the company and to which other companies the persons are related. The main reason for the short-term rating is to provide overall information about the company's current financial status and give a brief understanding of the probabilities of the risk for the future. (Kronwald & Chirstian 2009, 39-41.)

Long term rating

Short-term rating plays a role as a basis for the long-term rating, and the additional information and examination is to be done to provide a better understanding of the future of the entity that is the subject in the rating. As short-term rating is done mostly with automatisation in a short period of time, long-term rating commonly takes weeks and gathers additional information that is very hard or not possible to be collected and analysed with automated processes. Methods that are commonly used are company visits, interviews, and thorough analysis of the future of the entity. (Kronwald & Chirstian 2009, 44-45.)

When determining the long-term credit rating, the combination of an analyst's opinions and financial analysis is needed to provide the rating. The rating can be determined if there is enough information to cover the current financial status of the entity and provide a clear understanding of the risk of future operations. (Servigny & Renault 2004, 25.) Credit rating agency assigns a task for analyst or group of analysts that are assigned to provide an evaluation according to combined information from reports, interviews, and discussion with the top-level management of the entity (Moody's 2018). The criteria for the rating must be set, and every rating must strictly follow the rating process to provide a trustworthy and comparative rating. The comparison of the trustworthiness of the ratings can be debatable as the rating agencies do not provide the parameters or criteria used in the rating process, as those remain to be agencies' business secrets. In the worst case, the rating can mislead and provide false information for the reader as the reputation and opinions are the main elements when comparing the creditability of the ratings. (Servigny & Renault 2004, 28.)

2.2.4 Credit rating agencies

Standard and Poor's, Moody's, and Fitch Ratings are the most familiar internationally operating credit rating agencies, also referred to as the "Big Three" (Alessi 2013, 13). The market share of the "Big Three" is as enormous that it leaves only roughly 5.9% of the market share for the other operators in the industry on average in years 2014- 2017. (see table 2) The main reason for the non-fractured markets are the reputation and that there are only four other rating agencies in the US that have been recognised as nationally recognised statistical rating organisation (NRSRO), according to Credit Rating Agency Reform Act of 2006. (USSEC 2018) The large credit rating agencies such as "Big Three" provide ratings for large companies, but also nations, and their primary purpose is to provide long-term ratings.

Table 2. NRSRO Revenue Information: Percentage of Total Reported Revenue (USSEC 2018)

	2014	2015	2016	2017
S&P, Fitch, and Moody's	94.3%	93.7%	94.4%	94.1%
All Other NRSROs	5.7%	6.3%	5.6%	5.9%
Total	100.0%	100.0%	100.0%	100.0%

Other providers in the industry are stated to operate on niche markets in the industry. As the demand for the credit ratings has grown, the supply for certain kinds of credit ratings has grown as well. The providers have a small market share, and commonly the credit rating is not the main product that the company provides, and they operate locally or in some geographical regions rather than internationally. (Cifuentes 2013.) Finland's markets on the industry of credit rating are a great example of niche players (Junkkari 2012).

Asiakastieto, Bisnode (former Soliditet), and Intrum are all data providers that have successfully found their niche market in Finland as they produce the credit ratings for basically every single company in Finland, even for the SMEs. All the companies provide short term ratings mostly done by automatisation, but Fennorating, which is Asiakastieto's auxiliary business name, also provides long-term credit ratings. The rapid growth of the industry drives competition, and other parties than data providers are starting to provide credit rating services. For example, Alma Talent, which is commonly known to operate in the field of media in Finland, is offering short-term ratings. It is common for businesses to purchase data from other providers in the industry and determine credit rating with combined data from many sources to provide a better evaluation of the risk and financial status of the entity. (Junkkari 2012.)

2.2.5 Needs for ratings

As there are more and more credit rating providers, the demand has also grown. The needs for the credit ratings have changed in the past decade. The external need for the rating does not limit only to applying for the loan or to traditional examination when investing. The credit rating examination is a part of a simple risk management process. A credit rating is a fast indicator that provides the overall view of the risk of the entity. It is crucial to evaluate the trustworthiness of the rating as the ratings are becoming to be a significant direction pointer for the investment flow. (S&P 2019.)

A credit rating is used as a part of the communication in terms of investing. A credit rating can be considered to determine the price for the loan, investment, or financing. The capital that has been invested commonly bears an interest that is determined the risk of the investment accordingly. A common rule for the interest rate is that the riskier the investment is, the interest rate is higher as well. (S&P 2019.)

It has been studied that private investors are not interested in paying for the credit ratings when it comes to capital investments, obligations, trust funds, or stocks. External rating is seen as basic information for the investor, and the primary purpose of the rating is to speed up the evaluation process. Hyvärinen (1995, 32) states that credit rating does not have an impact on the process of evaluation of the profitability of the entity.

As well as the private investors, the financial institutions as banks or insurance providers do not have a necessary need for the purchasing of the externally provided credit rating, as it is seen that the creditor should be the one providing the credit rating and all the information that is needed in the financing process (Hyvärinen, 1995, 34-38). A trustworthy credit rating could potentially speed up the financing process and save the costs related to the process. Applying an external credit rating to the financing process is seen as a problematic matter as it does not line up with the internal process of financial institutions and regulations in concern. (Treacy & Carey 2000, 192).

Invoice funding is an increasingly more popular source for financing business operations. The creditor will get the receivables immediately or in a short time, rather than waiting for the receivables to become due. When it comes to the invoice funding, the credit rating must be accurate because of the rapid process of financing. Credit rating has a significant impact on the price of financing, which means that the sensitivity in the accuracy of the credit rating plays a significant role in both the creditors and the debtors' business. (Talenom 2020.)

2.2.6 Credit rating process

A credit rating can be issued by an external party or from the entity itself, and the process will be similar in both cases. The permission and the purpose are to be made clear, and the reason for evaluation must be clarified when issued by an external party. The credit rating evaluation process is commonly divided into different parts, which are the basis of the business operations, markets and the industry, competition advantages and capability, business operations, financial statement analysis, and cost structure. The findings from all the parts will be taken into consideration, and it is essential to bear in mind that the impact of the findings varies on when assessing entities operating in different industries. (Servigny & Renault 2004, 39.)

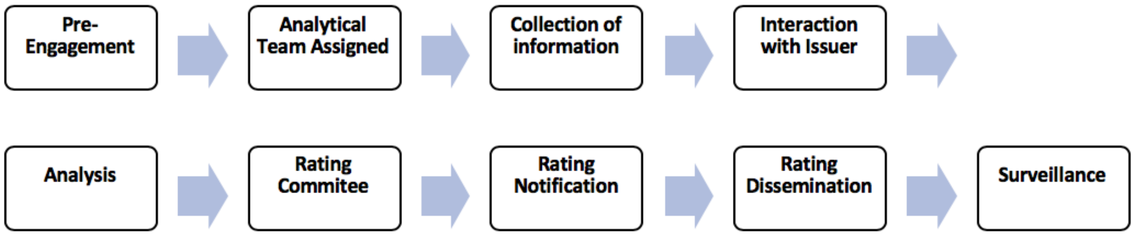


Figure 2. Moody's credit rating process (Moody's 2018)

Moody's rating process, the rating consists of 9 steps, as presented in figure 2. The pre-engagement stage involves the building of the relationship with the entity that is under the assessment. In practice, this commonly means an introduction call or a meeting. The rating process will be explained, and the practical factors are being agreed with the entity. The second step is to assign an analytical team according to the needs of individual evaluation. The team typically consists of a "Lead Analyst," who is responsible for the rating and other analysts that are needed for the case. (Moody's 2018.)

The team will start collecting the relevant data from the public sources and assessing the need for additional information that needs to be provided. The process will proceed with the collection of the precise list of financial and non-financial information that the analyst has determined on the earlier stage of the process. Information will be assessed again, and additional information will be collected until the analysts are satisfied with the amount of the information that they need for the evaluation. The interaction phase consists of team discussions with top-level management or their representative. The discussion aims to find information about matters such as credit strengths and weaknesses, but also about the trends and future of the industry. Interaction phase is an important matter when evaluating. (Moody's 2018.)

Once the required information has been gathered, the analysis is done according to MIS credit rating methodologies. The analysis includes both quantitative and qualitative factors. After the analysis is done, the rating committee will evaluate and examine the analysis by a critical consideration of quality, consistency, and integrity. The rating committee will determine the rating by the majority vote in the committee. When the decision is made, the issuer is informed about the rating. After discussing with the issuer, the rating under the evaluation published and distributed for the relevant parties. The final stage of the rating process is continuous surveillance, which aim is to monitor if there are changes that might impact the creditworthiness of the entity. If the surveillance is not done correctly, it is vital to make clear that the rating is identified as a point-in-time rating. (Moody's 2018.)

2.2.7 Inadequacies in the rating process

As the demand and supply growth increases in the credit rating industry, it is essential to evaluate the ratings and the methods that are used in the rating process. Although most of the agencies have rating processes that are broad and well-executed; there have been some cases where the rating has been a failure. Failure in the credit rating means that the agencies have done a faulty rating, and the highly-rated company has been riskier than evaluation states. Some of the agencies have been criticised for not lowering the credit rating, although there are clear indicators that increase the riskiness. So-called "issuer payer" model is to be taken into consideration when evaluating the trustworthiness of credit ratings, as the issuer is willing to pay to get a rating for their company and it is controversial in the sense that the agency has to be an impartial party in the rating process. However, the agency aims to generate profit for their stakeholder, which is the primary aim of the businesses. It can be argued that when the payment is involved, it might influence the rating. (Jollineua, Tanlu & Winn 2014, 1399-1402.)

Angus Duff stated that the entities are more likely to issue the rating for the party that evaluates the rating higher and has not caused them any problems related to the rating. The lack of transparency in the rating is also stated as a problem related to the credit rating. Credit rating agencies do not provide detailed information about the credit rating, and the basis for the rating might be unclear and partly unreliable because of the lack of information. As the competition increases, the performance of the businesses typically also does. It is stated that if there is not enough competition, it impacts the reliability of the credit rating. The lack of competition plays a significant role in the niche markets. (Duff 2015, 552-554.)

At the time of the financial crisis, usually, the criticism is aimed towards the credit rating agencies as the global financial crisis of 2007, and the financial crisis in Asia in 1997 has shown. The criticism is relevant as the credit rating agencies have been involved as a part of the causes of the financial crisis. Agencies have provided higher ratings in comparison to the risk and have failed to evaluate the riskiness of the entities. (Duff 2015, 553-554.)

Energy company Enron's case in 2001 is a famous example of the failure in the field of credit rating. Enron, the was the 7th largest company in the world at the time of occurrence, was included in the top ten performing companies in the Fortune-500 list. The size of the company, a short timeline of the crash, and the influences of the case raised a massive amount of interest in the case. Enron had been a profitable company in the past year, and even the investment-grade rating was given five days before the bankruptcy. Rating agencies have reasoned the faulty rating to be caused by a concentration on the long-term rating that did not include the short-term changes in the industry and markets. (Foss 2003.) Frost stated that the lack of information gathered on the interviews was the main driver for the faulty rating as the amount of the confidential information did not fulfil requirements to provide a reliable evaluation of the company's financial status. It is vital to notice that the past performance or size of the company is not enough information to evaluate the risk of the company. (Frost 2007, 469-472.)

2.3 Financial statement analysis

Financial statement analysis is an old method, and it has been used as early as from the end of the 19th century when determining the rating of creditworthiness and risk. There have been and still are many correct ways to conduct financial statement analysis, and the methods have developed over time. Financial statement analysis gives understanding about the financial status of the company according to information from the financial statement, which generally consists of an income statement, balance sheet, cash flow statement, changes in owner's equity, and attachments. Financial statement analysis regularly measures the profitability, liquidity, and solvency of the entity. Findings and ratios from the analysis are compared to a defined reference value or other businesses operating in the same industry. Financial statement analysis is stated to provide information about the entity's financial health. (Yritystutkimusneuvottelukunta 2011, 23-30.)

There are many purposes for financial statement analysis. Yritystutkimus Ry (2011, 23) states that financial statement analysis varies according to the needs for the analysis. The correct methods that are used in the analysis must follow the needs. The financial statement provides information for financial statements, and ratios can be compared and evaluated according to guidelines. Financial statement analysis is seen as an opinion about an entity's profitability, solvency, and overall performance and financial status. (Yritystutkimusneuvottelukunta 2011, 23.)

Leppiniemi states that a financial statement is a tool for decision making and conclusions related to the entity's finance. The analysis is seen as a method to predict the viability and future operations of the company. The financial statement must be comparable or corrected to analyse appropriately. According to Leppiniemi & al. (2017, 32), any businesses or groups of businesses cannot be compared to each other, as it might give erroneous information. Different industries commonly have utterly different reference values as the operations are also different. For example, a construction company cannot be compared to a company that sells cars or goods. (Leppiniemi & al. 2017, 32.)

On these grounds, 'H1: The business industry influences the credit and risk ratings' is set.

Laitinen states that the financial health of the entity consists of a relationship between profitability, solvency, and liquidity. All three indicators have long-term influences on financial health. For example, if the company is profitable, but the cashflow cycle drives the solvency down, the company's liquidity suffers. If a company is liquid but not profitable, they are not able to follow the payments and solvency suffers the financial health of the company. The overall financial health of the entity determined by different variables (see figure 3) (Laitinen 1992, 82.)

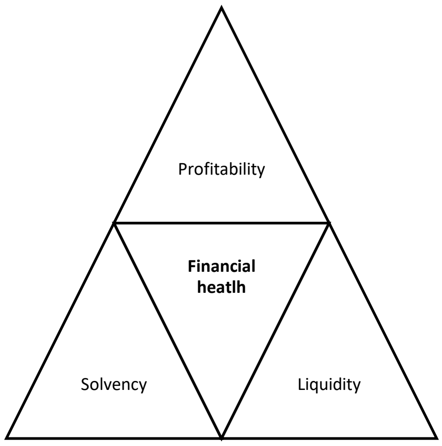


Figure 3. Indicators of financial health in financial statement

On these grounds, 'H2: Profitability, solvency, and liquidity all have a significant role in the credit and risk ratings' is set.

2.3.1 Financial ratios

A financial ratio is a relative magnitude of selected numerical values. A financial ratio is a simple way to analyse the business from a financial statement. Financial ratios are used in financial statement analysis as the ratios are objective and concrete measurement for comparison. The analysis includes calculation of ratios, and the ratios will be compared to determined values or guideline values. The weight of the ratios is to be set before an examiner can determine or conclude the operational and financial status of the company. (Gropelli & al. 1990, 433-436.)

As the financial ratios and calculations needed are simple, standardised financial ratio analysis has been criticised. It is crucial to consider the chosen ratios carefully and examine if the ratios can be comparable to any kind of business, and if averages or medians are the right value to in comparison. Financial ratios provide a general and straightforward overview of the profitability, liquidity, and solvency of a business. (Laitinen 1992, 80-83.) Laitinen states that financial ratios do not explain how well the business is operating, but rather how the past operations reflect on the financial statement (Laitinen & Laitinen 2004, 221).

Businesses are easily able to manipulate financial ratios with specific actions; for example, if the company pays off their current liabilities, their quick ratio improves. Financial ratios do not consider the payment behaviour or willingness, but rather just explain the capability of solvency. Financial ratios do not provide information outside the financial statement. Conclusions done with only a certain kind of information might not give a complete answer to a matter. With manipulation, a credit rating that is done automatically with financial ratios, might give false information and mislead financiers. (Gropelli & al. 1990, 445-451.)

2.3.2 Size of the company

The size of the business is seen as an impacting indicator when analysing the risk of the business with financial statement analysis. Amato & Furfine (2014, 2641) state that larger companies have less risk due to the variety of revenue streams and concrete products or services. It is also seen that larger companies have more information on their financial statements, and financial ratios give more information than smaller companies. (Amato & Furfine 2014, 2641-2644.)

On these grounds, 'H3: Company size impact credit and risk ratings' is set.

2.3.3 Profitability

Profitability is seen as a business's long-term capability to generate profits (Laitinen 2004, 245). The relation between net profit and capital indicates the profitability of the entity. In other words, profitability is a measure of how well company operations generate the amount of capital tied to the entity. Financial ratios that indicate profitability, according to Yritystutkimusneuvottelukunta, are sales margin, gross margin, operating profit, net profit, and financial profit. Ratios can be calculated with information from an income statement. (Yritystutkimusneuvottelukunta 2011, 54-56.)

When considering the business's efficiency to generate profits from its capital, information on the income statement and balance sheet are both needed. Laitinen states return on capital investment as the most critical measurement of profitability. ROCE gives an understanding of profits that are made in comparison to equity and long-term liabilities. (Laitinen 2004, 245.)

From the profitability point of view, the business's main priority is to maximise the profits for its owners or stakeholders. Return on equity, which indicates net result in the relation of equity is seen as the most crucial profitability ratio for owners and stakeholders. Businesses can increase their ROE with liabilities, which increases the leverage, but it is seen an increase in the risk as liabilities bear expenses. (Yritystutkimusneuvottelukunta 2011, 55-57.)

2.3.4 Solvency

Solvency gives an understanding of the capital structure of the entity. It measures the relationship between capital and liabilities, which tells about the health of the capital structure. Capital structure plays a significant role in the operational side of the business as if company operations are entirely financed with short- and long-term liabilities; the company has high leverage. However, the financing of the operations can end shortly due to the interest-bearing liabilities. In the opposite situation, if the company's capital consists entirely of equity, the entity does not have enough leverage. (Laitinen 2004, 255-256.)

Equity ratio, which measures the portion of the equity in relation to business assets is a simple way of inspecting the capital structure. The ratio is commonly used in credit rating because it tells the reader how much will be left if the operations are shut down. Equity ratio, net gearing, and return on assets are all seen as static measurements. The static ratios will measure the capital structure of the business. (Laitinen 2004, 256-259.) Dynamic measures of solvency are ratios that will indicate the sufficiency of finance. Dynamic measurements are used, for example, when determining how probable it is for business to be able to cover the payments of liabilities with the relation between debt bearing liabilities and financial results. Equity ratio to be a risky indicator, as it does not cover the terms or securities that have been agreed related to financing. (Ikäheimo, Laitinen & Laitinen 2014, 102.)

2.3.5 Liquidity

Liquidity is also divided into static and dynamic liquidity. Liquidity indicates the sufficiency of financing business operations in the short-term. Static liquidity measures the relations in the balance sheet with standard measures such as quick ratio and current ratio. The quick ratio, also known as the acid test, measures businesses' ability to cover their current liabilities with their current assets. The current ratio is like the quick ratio, but it also covers stocks and the value of work in process goods. Static ratios work as a quick tool to measure the portion of current liabilities that could be paid-off as if the business would shut down their operations. High liquidity is seen as leverage for settlement discounts, and weak liquidity will commonly drive business to defaults and late payment fees. (Laitinen 2004, 248-250.)

Unlike static liquidity measures, dynamic measures also take running expenses into a count. Dynamic liquidity measurements such as financial result percentage and operating cash flow measure the amount of finance that purely comes from the business's business operations, which is crucial, especially for growth companies. (Yritystutkimusneuvottelukunta 2011, 67-69.) Liquidity is essential as if the company cannot finance its business operations, purchases, and wages, operations will be shut, and cash flow stops, and eventually, the business will go bankrupt.

2.3.6 Financial issues at an early stage

At the early stages of the financial problems of an SME, it is vital to look at the operational processes as the decrease in financial ratios takes longer to become visible when analysing. (Frost 2007, 473.) It has been studied that the businesses that have been windup to poor and risky financial health have been facing problems on operational factors such as investments, financing, development, costs, marketing, operational management, but also in accounting. (Laitinen 2004, 233.)

Even though operational risk factors have a high impact on the risk of business, external risk factors also might influence businesses financial failure. The difference in analysing the risk is that external risk factors are less commonly visible or predictable; for example, CODVID-19 that could not be forecasted. Operational risk factors are often detectable. There are also some clear indicators of financial issues visible on a financial statement, such as changes in inventory, decreasing receivables, and increasing payables. Also, if fixed assets are being sold, it causes an increase in other operating income, or if the number of employees is decreasing, it indicates of business not being able to survive without layoffs. The clearest indicator of financial difficulties is late payments of invoices, as cash balance does not cover all the payables. (Laitinen 2004, 195-197.)

2.3.7 Clearing account

As the real-time credit rating can be done in the middle of the fiscal year, it is essential to take the clearing account into a count when considering the financial status of the company. KILA states recording of transactions to be based on a dated and numbered document certifying the transaction. The document is stated to include at least; the name of the issuer, the content of the transaction, and the amount of the transaction. The proof of expenditure must indicate the factor of production received and the performance delivered from the proof of income. It must be possible to prove the time of receipt of the factor of production and the date of delivery of the work by means of a document or an annexe thereto or otherwise. Documents related to the document, for example, contracts may also be kept separately, provided that their connection can be easily linked to business transactions, if necessary. Transactions classified as unclear may be recorded in the clearing account, provided that the account is cleared and liquidated at the latest when the financial statements are prepared. (KILA 4.6.2001/1653.)

A clearing account is a temporary account containing transactions that are to be cleared and transferred into another account. Cash transactions, including payments made with debit- or credit card, are the most common example of clearing account entry. Cash payments, ATM withdraws, and card payments are to be recorded with no delays. As the entry on clearing account is to be cleared before closing the books, the entry is not a final solution. If the document is not delivered, the transaction is seen as a personal expense. Therefore, entries can be included neither in the income statement nor balance sheet at the point of books being closed. (Section 2, Accounting Act, Paragraph 4.) It is common for entrepreneurs to pay personal expenses with company assets, which is not allowed, and the faulty transactions will be recorded in the clearing account. If the proofs of business relationships and other obligations stated are not delivered, the entry is seen as a personal expense. (KILA 5.6.2000/1615.)

On these grounds, 'H4: Increase in the balance of clearing account has a negative influence on the credit and risk ratings' is set.

2.3.8 Bankruptcy

Although insolvency and bankruptcy are topics that are often discussed in the same researches and articles; it is vital to notice that insolvency and bankruptcy are defined entirely differently.

According to the Bankruptcy Act 20.2.2004/120 §1, bankruptcy is defined as follows:

- A debtor who cannot repay his or her debts can be declared bankrupt in accordance with the provision of this Act. The court shall make the order of bankruptcy on the petition of the debtor or a creditor.
- Bankruptcy is a form of insolvency proceedings covering all the liabilities of the debtor, where the assets of the debtor are used in payment of the claims in bankruptcy. To achieve the objective of the bankruptcy, the assets of the debtor shall at the beginning of bankruptcy become subject to the authority of the creditors. An estate administrator appointed by the court shall see to the management and liquidation of the assets of the debtor and the other administration of the bankruptcy estate.
(Bankruptcy Act 20.2.2004/120 §1.)

2.3.9 Forecasting of bankruptcy

Forecasting of bankruptcy has been one of the most researched subjects in the field of financial forecasting. The central concept in past researches has been searching for common factors in financial ratios of companies that have bankrupted. Many have tried to search for the perfect model for predicting the bankruptcy, and even high accuracy rates have been reached. (Koulu 2009, 102.)

Aziz & Dar (2006, 18) divided bankruptcy forecasting models into three categories:

- Statistical model
- Theoretical model
- AIES-model

As the name of the model refers, statistical models are based on the statistics. Statistics analysed consists of the data on financial statements. The object is to find the reasoning trough financial ratios in order to understand the symptoms that are common in bankruptcy. Statistical models study the probabilities, typically with regression analysis or variate models. (Aziz & Dar 2006, 19-22.)

Artificial intelligence expert system models (AIES) are more sophisticated analytical models that examine the common factors and reasons for bankruptcy with neural network models, which replicates the problem-solving process. In simple terms, AIES-model works in a way that narrows down the options and calculates the probabilities from relations of remaining options. The object is to train and test the model to learn the relations and common factors of bankruptcy with sample data. (Laitinen 2004, 151-153.)

Theoretical models examine the causes and failures in processes. Theoretical model researches the reasons behind the analytical findings, which operates more as a help to find the causes, rather than as the causes itself. The models are commonly built to match the theoretical models or assumptions. (Aziz & Dar 2006, 19.)

Altman approached the forecasting of bankruptcy with a model that is based on 22 financial ratios that are chosen from past researches, with no other theoretical reasons or basis for the ratios (Altman 1968, 594). This model reached the accuracy as high as 95% when forecasting the bankruptcy, a year before it happens (Altman 1968, 604).

Erkki E. Laitinen has stated five best indicators for forecasting of bankruptcy in his study in 1993. Research-based on prime factors of bankruptcy and the financial information was gathered from financial statements 7-8 years before the bankruptcy. (Laitinen 1993, 95.) This method excluded the most common companies that face bankruptcy as it is studied that companies that have operated 0-4 years are most likely to go bankrupt. Laitinen (1993, 95) saw the following indicators as the best ones on forecasting the bankruptcy at an early stage:

- Net Finance Profit %
- Quick Ratio
- Turnover of payables (in days)
- Equity Ratio %
- Change in revenue % (average of the past three years)

2.3.10 Defaulting

It is seen that poor profitability and low capability to finance business operations with cash flow drives businesses to delayed payments and defaulting. Profitability will impact liquidity and solvency as the business is forced to seek external financing to run their operations. Eventually, if profitability cannot be increased, the business cannot settle the payments or follow the terms that have been agreed to relate to the financing. (Laitinen 2005, 76-79.)

If the payment is not made, a debtor will issue a reminder of payment, following by another one, and if the payment is still not made, a default of payment is recorded in to register. The most common default of payment recording is a publicly visible unaccepted bill of exchange and other records and more related to when business is filing bankruptcy. Both recordings influence the credit rating, and the impact will be determined according to the seriousness of the default. The problem with the default of payment in the eyes of credit rating agency and financier is the delay, and lack of information as the prediction of defaulting has stated to be inaccurate. (Laitinen 2005, 76-79.)

An additional factor that should be taken into consideration when analysing the influence of defaulting is extent and frequency. The default of payment can occur because of technical error or negligence. On the other hand, the frequency of defaulting might be high, and defaulting could be seen more like a habit than a nonrecurring occurrence. It can be criticised if scenarios should have an equally valued impact on credit rating. Even though forecasting of bankruptcy, according to Erkki Laitinen (2005, 78), can be done on a higher accuracy rate, it often causes complete losses for the financier, which increases the need for improvement due to the risk of losing all the remaining balance. (Laitinen 2005, 76-79.)

On these grounds, 'H5: Defaults and late payments influence credit and risk ratings' is set.

2.4 Real-time data

The term "big data" is currently becoming an often-used term in business language. There have been many purposes for this term in the past, and the meaning is understood in many ways. In 1997 the term was used to visualise large data sets (Cox & Ellsworth 1997, 87). J. R. Mashey (1998, 86) referred to term big data in hardware presentation. Also, in 1998 term big data was referred to in the context of data mining (Weiss & Indurkhay 1998, 33). F. X. Diebold (2003, 601) combined data mining with statistics.

It is considered that in 2001, big definition data reached a significant milestone as Laney (2001, 55) introduced three essential dimensions of big data: volume, velocity, and variety, which are referred to as 3 V's. "Operating with a swarm of autonomous quadcopters requires the management of high-volume, high-velocity (real-time) data that have many types (variety)." – (Ylijoki & Porras 2016, 70-73).

The purpose of practical big data solutions is to provide additional value for the company. Pioneers on the field, such as Google and Amazon, managed to apply real-time, big data in practical use and benefit of it. After 2010, it has been providing benefits and opportunities in other fields of industries, while raising popularity amongst the industries. (Ylijoki & Porras 2016, 70-79.)

3 Empirical analysis

The empirical analysis studies the differences in quality of the public data and real-time data, the correlation of explanatory factors to ratings rated by agencies, rating method for real-time data and coefficient of correlation of late payments and clearing account to real-time credit and risk ratings. The aim is to understand the complete picture of real-time credit and risk ratings and the factors impacting the ratings, which cannot be executed without the data analysis and the evaluation of the results.

The thesis examines the possible improvements for current credit rating methods and examines the possibility of increased accuracy with real-time data and additional information. The suggested credit rating model is based on real-time financial statement analysis, which is enriched with additional information on payment behaviour is not publicly available. The improvements in the credit and risk rating models are readily applicable to the process of automatically formed financial statement analysis. The model provides a quick and comparable credit or risk rating that gives the financier an understanding of the risk related to the financing and understanding of the overall riskiness related to the business. The model may increase the reliability and accuracy when evaluating the risks as the data analysis is formed with high-quality data with no manual work. The model is wholly impartial but takes the industry of the business in a count in the analysis. The model concentrates on risk related to financing by analysing the probability of defaulting and insolvency.

Research methods consist of following phases; setting the hypotheses according to the theoretical framework, data analysis and evaluation of reliability and validation. The research is explanatory research with quantitative research methods that include qualitative attributes. Explanatory research aims to provide an answer to the primary research problem, evaluate and explain the methods and process behind the credit ratings, but also seek the impact of different explanatory factors. Explanatory research tends to include an answer to casual connections of the phenomenon. (Uusitalo 1991, 63-64.)

3.1 Hypotheses

Hypotheses in this research are set for examining the variables that have an impact on credit and risk ratings. The aim is to build the model for the classification of credit ratings that are based on real-time data. Ratings that are rated with a real-time classification model are referred to as real-time ratings in this research. Hypotheses evaluate the financial ratios that are commonly used in the classification of ratings. Influence of explanatory factors of payment information and clearing account is researched in the examination of improved rating models.

The hypotheses are set according to the theoretical framework in chapter two. The theoretical framework consists of past studies and researches. The theory is built around studying the primary research question and investigative questions. The questions aim to explain if real-time financial data and factors of payment behaviour could improve the rating methods and process.

Hypotheses are either approved or disapproved from an empirical viewpoint. Hypotheses are set as:

- H1: The business industry influences the credit and risk ratings
- H2: Profitability, solvency, and liquidity all have a significant role in credit and risk ratings
- H3: Company size impacts credit and risk ratings
- H4: Increase in the balance of clearing account has a detrimental influence on credit and risk ratings
- H5: Defaults and late payments influence on to credit and risk ratings

According to the evaluation based on theory, financial ratios are a simple and easy way to understand the brief and quick status of an aspect of businesses finance and even compare in relation to other business. Financial ratios do have an essential purpose when analysing the financial health and risk, but the ratios are dependable on delivery and accuracy of the financial statement of the business. Horizontal and vertical financial statement analysis provides valuable information about relations but also changes in ratios in different periods of time. It is crucial to notice that there are also weaknesses related to the financial ratios, especially when considering the use of the ratios in automatic rating processes. It is essential to notice that not all the industries are comparable with financial ratios on the same scale, as the financials of businesses are different in different industries. Financial ratios are not to be studied as independent targets on the results as the correlation coefficient is analysed as the aim is to seek the explanation from multiple factors.

As the ratios are a relative magnitude of selected numerical values, company size does not matter as percentage or portion remains the same no matter which size of a business is under the evaluation. The model built for evaluating the financial health and risk is applicable and comparable for each of the businesses in the sample. The financial information of SMEs that is used in the classification of credit and risk ratings is seen as providing the same information as it consists of material that is delivered for accounting. The material that is delivered to accounting includes every single detail of businesses finance in terms of financial data that can be collected. Small companies in the rating process are stated to seen problematic because of the lack of financial information when data is from public sources. The basis that set the hypothesis of company size impacting on ratings is invalid according to the qualitative evaluation. Evaluation guides the quantitative analysis according to the results, and because of the invalidity, 'H3: Company size impacts credit and risk ratings' is disapproved and will not be included in the data analysis. According to qualitative evaluation, the creditworthiness and risk of the business should be measured with different indicators in order to determine reliable ratings.

The aim of studying the payment behaviour is to seek for variables that could improve the rating model. Improvement is analysed according to the impact of variables. Payment behaviour is currently studied in the form of defaults. This research studies the payment behaviour with explanatory variables that are not commonly used in rating methods as the data is not available publicly. Examination of payment behaviour seeks to understand if late payments and transactions that cannot be recorded on permanent account impact the ratings in an explanatory way.

3.2 Data analysis

Data analysis consists of 7 phases; the data mining and data qualification, comparison of differences in public data and real-time data, multinomial logistic regression analysis of financial variables, the building of classification model, comparison of the real-time ratings and public ratings, multinomial logistical regression analysis of late payments to real-time ratings, and logistical regression analysis of late clearing account to real-time ratings.

The data for this research consists of real-time data and public data. Data is either mined from commissioning company's databases or provided by Finnish rating agencies. The rating agencies use the data as a product to sell, evaluate the companies and rate the companies. Real-time data is used in providing the accounting services of the commissioning company. The number of businesses that are used for analysing the financials and payment behaviour is 6113 SMEs. Stated 6113 businesses are chosen from a variety of business industries and are commission company's customers. The comparison of data, training of the classification model and real-time ratings are executed with 2790 SMEs from construction and accommodation, food and beverages activities industries. The 2790 SMEs in question are also customers of commissioning company. The margin of error of five per cent or less is set to be acceptable in data validation in this research.

Public data is referred to as data that is provided by rating agencies. The data consists of basic information, financial data, and ratings of Finnish SMEs. Data have been chosen from the SMEs that are customers of commissioning company and operate on the construction industry or accommodation, food and beverages activities industry. The dataset is compared to real-time data. Public financial data is used in the comparison, but also for testing and training the supervised multinomial logistic regression classification model.

Real-time data referred to as the data that is consisted of the material delivered to accounting for commissioning company. The data is mined from SQL databases of commissioning company in Microsoft SQL database editor. Databases are made separately for financial data, purchase invoices and entries. Databases are connected to Python 3, and the data is sorted and formatted. The modified dataset will include the basic information of SMEs financial data that is needed, purchase invoices, and entries of Finnish SMEs that are operating on construction and accommodation, food and beverages activities industries. The analytics of an overall view of the matters is done separately. Real-time data is used for determining the differences between the datasets and evaluation of possible benefits that real-time data could bear within.

The explanatory variable in this thesis is either credit or risk ratings and the primary industries of the SMEs. The explanatory factors are financial ratios, late payments in relation to all purchase invoices, and debit entries of clearing account in relation to all transactions.

The industries that are researched separately as hypothesis H1: The business industry has an influence on the credit and risk ratings is set. Industries in this research are the construction industry with industry codes 41-43 and accommodation and food and beverages activities with industry code 55-56 according to guidelines of Statistics Finland (2013). Industry-specific factors and correlations are examined in the analysis.

Credit and risk ratings are measured separately and are entirely different variables. Credit rating measures the ability to settle the payments, and risk rating measures the probability of bankruptcy and financial difficulties. The ratings have been grouped into four categories in credit rating and risk rating to ensure the anonymity of the businesses. After the grouping, the groups are given numerical value as points for statistical analysis. Points are given from 0-3. Low values indicate low ratings and high values indicates high ratings. All the variables in this research are either financial ratios or numerical relations.

The credit ratings have been grouped according to the primary credit grade and risk grade, according to table 3. Grouping is done in order to ensure the anonymity of the companies in sample data besides executing the statistical analysis.

Table 3. Credit ratings by groups and points

Rating	Group	Points
A+,A and A-	A (Very good)	3
B+,B and B-	B (Good)	2
C+,C and C-	C (Fair)	1
D	D (Poor)	0

The grouping has also been done for the risk ratings according to table 4. Grouping is done in order to ensure the anonymity of the companies in sample data besides executing the statistical analysis.

Table 4. Risk ratings by groups and points

Rating	Points
Very low risk	3
Low risk	2
Significant risk	1
High risk	0

Explanatory factors of financial ratios are chosen accordingly to information that credit rating agencies have provided. Financial ratios are formed in Python 3 with the connection to SQL databases of commissioning company or have been provided by the rating agencies. All the real-time financial data that is used in this research are available from the fiscal year of 2019. Financial ratios are formed according to the formulas stated in Alma Talent's financial ratio guide (2020). Formulas are translated account by account in order to guarantee the correctness. Ratios are not modified in order to evaluate the real-time data in the state as it is collected. Financial ratios are used for evaluating the profitability, liquidity and solvency of the businesses. Formulas of the ratios by accounts can be found from attachments. The measures are set as follows:

- Profitability:
  - Revenue growth %, operating Profit %, ROCE, and ROA
- Liquidity:
  - Current ratio, quick ratio, and equity ratio
- Solvency:
  - Net Gearing and Relative indebtedness

Explanatory factors related to payment behaviour are chosen according to the wishes of the commissioning company. Explanatory factors of late payments are divided into five categories according to the number of days of payment made after the due date in relation to all purchase invoices; all late payments, 1-6 days late, 7-13 days late, 14-30 days late, and over 30 days late to inspect the different kinds of payment behaviour. If all the payments are made in time, the value is 0. An explanatory factor of clearing account is evaluated with debit entries of clearing account in relation to all the transactions that have been made in the same period, in order to exclude the size of the company from the evaluation. If there are no debit entries, the value is 0.

Analytics are executed in Python 3. The coding language used in Python 3 is commonly known as Py. Libraries used in statistical analysis are pandas, pyodbc, numpy, matplotlib, seaborn, sklearn, sklearn.tree, scipy.stats, sklearn.metrics, pydot, pydotplus, graphviz, and StringIO. The possibility of implementation of the Neural Network model was examined and excluded because of the lack of computing ability devices used in the research.

Logistic regression analysis aims to examine the linear dependency between two variables. Multinomial logistic regression analysis examines the impact of multiple factors on the variable. The correlation coefficient measures the dependency of variables. The values of the correlation coefficient are in the range from -1 to 1. The values imply the linear relation between the factors and variables. A positive correlation indicates an increase in both variables in relation, as negative correlation indicates the decrease in other variables when the value of the other variable is increasing. The value 0 indicates that there is no correlation between two variables. Correlation is measured in this research in the evaluation of explanatory factors to ratings. The closer the value is in the high or low range of correlation coefficient, the higher impact it has in relation to ratings. Correlation is commonly analysed when analysing the impact of variables. (Metsämuuronen 2009, 743.)

The statistical significance is determined according to the p-value. Correlation coefficient analysed in understanding the impact through odds. The p-value is a measure for the statistical significance. Holopainen and Pulkkinen (1999, 9) stated the value to be statistical significance to be reached when the p-value is lower than 0.05. The scale of statistical significance in this research only measures if there is a statistical significance or not.

$p < 0,05$ ($\alpha = 5$ %) is statistically significant

Multinomial logistic regression analysis measures the impact of each explanatory factor to explanatory variables researched in this analysis. The aim is to evaluate and compare the impact of different factors and provide an explanation for seeking statistically significant factors in relation to ratings. Logistic regression analysis and multinomial logistic regression analysis are commonly used analysis methods in evaluation of the impact of explanatory factors on ratings. Results of logistic regression analysis are the coefficient of correlation and p-value of different factors. The aim is to seek the factors that will be used in the classification model, on determining the real-time ratings. Logistic regression analysis is done for all the statistical categories measured in this research. Logistic regression analysis provides the information for comparing the correlation of different variables against the findings of past researches and the information that rating agencies have stated. (Metsämuuronen 2009, 745.)

Logistic regression analysis is sensitive in a way that if there are multiple outliers in the results, the correlation coefficient variables decreases. In the worst-case, variables correlation could be high when measured separately, but the outliers decrease the coefficient of correlation, and the result is not seen as a significant factor. All the variables are examined separately in the case of abnormality. Abnormality is analysed in comparison to past researches and information that credit rating agencies have provided. (Metsämuuronen 2009, 748.)

Logistic regression is a classification method, which is not to be mixed up with an ordinary measure of regression. The method aims to provide the relation of odds of SME with particular values of factors to be classified on a specific class of credit and risk grades according to the results of the analysis. The odds referred in terms of logistic regression classification model measures the relation of odds to an explanatory variable, instead of odd itself. Supervised logistic regression classification model in this research is used in a process that is commonly known as reverse-engineering. The model is trained and tested with public data and used in the classification of real-time data. (O’Brien 2007, 673-688.)

The accuracy of the classification algorithm involves an essential role in the analysis as the accuracy is to be as high as possible to model to be reliable. The algorithm aims to classify the ratings consisted of real-time data according to reverse-engineered classification with as high accuracy as possible. The classification accuracy of 85% has been set to the lowest accuracy rate acceptable in the model. The accuracy rate is set upon the wishes of the commissioning company. Ratings in the model have been converted into the numbers in order to execute the statistical analysis. (O’Brien 2007, 673-688.)

Improving the accuracy of the model can be done with following elements; adding the data, excluding of the missing data and outliers, narrowing the scale measured, selection of factors and variables, differentiation of algorithms, modifying the test-sample and train sample, and validation. The train-sample is used for training the model, and test-sample is used for testing the model. The seed value is used for randomising the pipeline of the data in the model. The seed is guaranteeing the sequence of numbers to be exact and anonymity of the SMEs in the sampling. Multicollinearity could provide high accuracy in the model, but also faulty conclusions. (O'Brien 2007, 673-688.)

Multicollinearity is an occurrence when the coefficient of correlation of two variables is very high. Logistic regression reacts to multicollinearity in a way that one of the variables in collinearity is not significant as the measure is equal or close to equal. When evaluating the financial ratios, this is common if ratios are alike each other. Multicollinearity can cause models that give less reliable and provide imperfect information about correlations. (O'Brien 2007, 673-688.)

3.3 The results and suggestion of the rating model

Only the relevant results will be presented. The results are presented in an illustrative way for the reader to understand the overall concept of the particular matter and the differences when inspecting certain industries. The research consists of multiple phases and covers the whole process from mining the data to the analysis of the impact of factors to variables. The results are illustrated in tables and figures in chapter 4, and the evaluation and discussion are written in chapter 5.

The suggested model of credit and risk ratings are built after evaluating the results of data analysis. The statistical significance determines the factors that are chosen to the model, and correlation provides the weighted relation of each factor used in the model. The results of data analysis are compared to past studies and information that rating agencies have provided as the evaluation of a business case is in order. Additional factors of payment behaviour will be examined according to the influence of factors to ratings. The evaluation of the benefits related to factors are examined, and suggestion is made according to the conclusions made of evaluation of additional factors in the rating process. The phases of data analysis are built in commissioning company's servers, and rating models are ready to be implemented for the rating process.

3.4 Reliability and validation

The research must be reliable and repeatable. The research has formed in a way that will be as reliable as possible. Reliability will be possible only in a case when the research phases and goals are clear and understandable. The data will have to be examined multiple times as the mistakes will decrease the reliability of research. The validity of data is to be taken into a count on every phase of the research. (Metsämuuronen 2009, 779.)

Modifying and improves made on the model cannot compromise reliability. The data includes risk as if it is not homogeneous and multicollinearity is to be avoided. Financial information will be compared between businesses that operate on the same industry, and real-time data consists of financial data of fiscal years that start 1.1.201x and ends on 31.12.201x. Different forms of companies do not have an impact on the comparability as the relevant regulations and standards related to the accounting in terms of data validation are the same for every business that has qualified in sample data of this research, which makes the data comparable and trustworthy. (Metsämuuronen 2009, 780.)

4 Results

Financial statement analysis and information on financial statements have been fundamentals for short term credit ratings and when predicting the defaulting or bankruptcy. It is seen that there are many problems related to the current credit rating methods and processes, and it can be argued if credit ratings formed with publicly available information give enough information for setting the price for financing or when evaluating the financial status and risk. The current short-term credit rating process does not cover the information outside the financial statement, other than defaults, but it is still seen as a significant factor when analysing the real financial status and risk of the company. The relevant results explaining the matter in question will be presented in this chapter. Accurate correlation coefficients and p-values of all factors on ratings analysed with multinomial logistic regression analysis or logistic regression analysis can be found in attachments.

The following chapter will present the results of the data analysis. The results and explanations are evaluated. As commonly seen in statistical models that evaluate the explanatory factors and variables behind the ratings, the constant or the correct method for rating cannot be stated according to the results. Evaluation of results discusses the possible causes behind the phenomenon when analysing the sampling used in this research.

4.1 Sampling

When evaluating the size SMEs in the sampling, it is seen that most of the randomly chosen businesses employ 0 to 4 persons. The distribution in size of the business according to the number of employees of SMEs is presented in Figure 4.

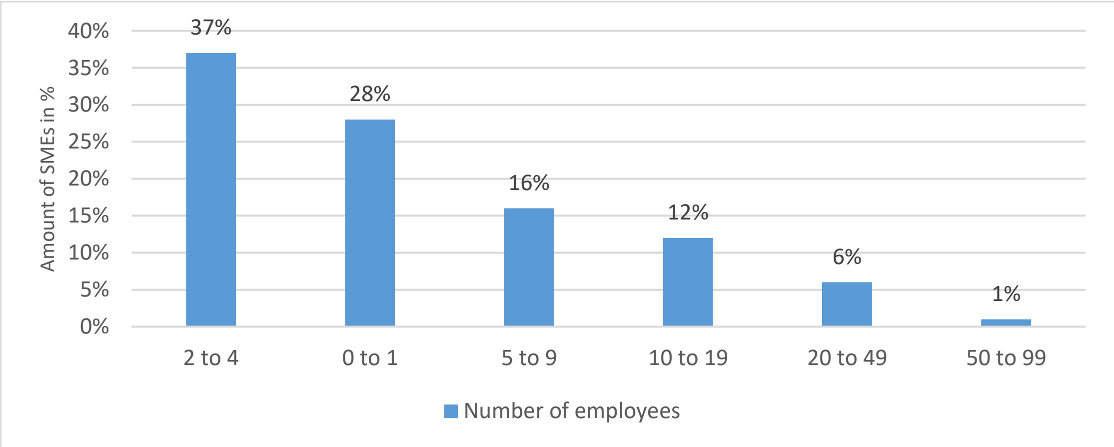


Figure 4. Distribution of the number of employees in sample data in percentages

The distribution of locations is analysed in order to ensure the variety. The variety in sampling is essential as the location might impact on the results. As presented in Figure 5, it is seen that most of the SMEs of the sample are in major cities.

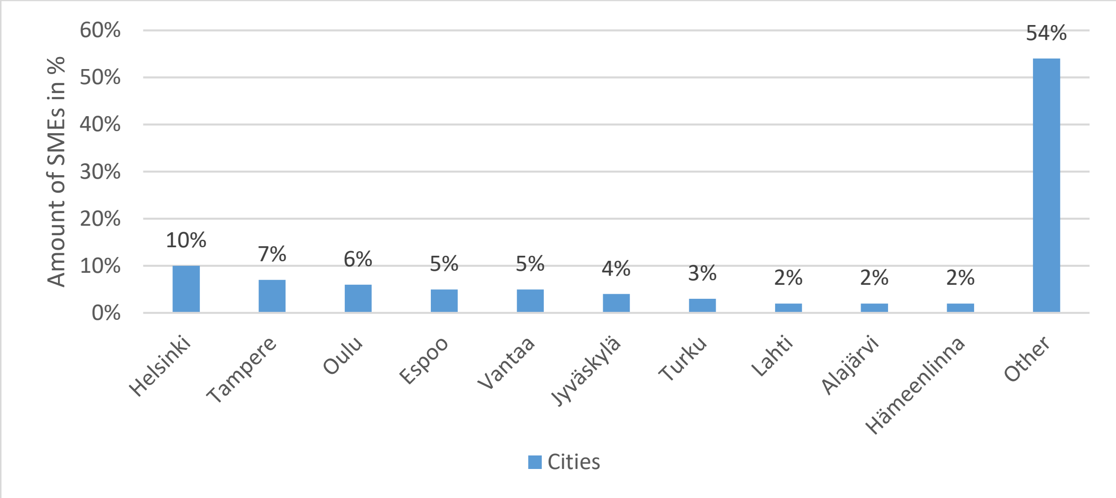


Figure 5. Distribution in the locations of the SMEs in the sampling in percentages

Financial data consists of real-time data of SMEs. This thesis concentrates on studying the possible benefits of real-time data. The real-time financial data must be correctly formed and examined. After thorough data validation that includes close inspection of errors, 1000 SMEs on industry code (55-56) 1790 SMEs on industry code (41-43). (see figure 6)

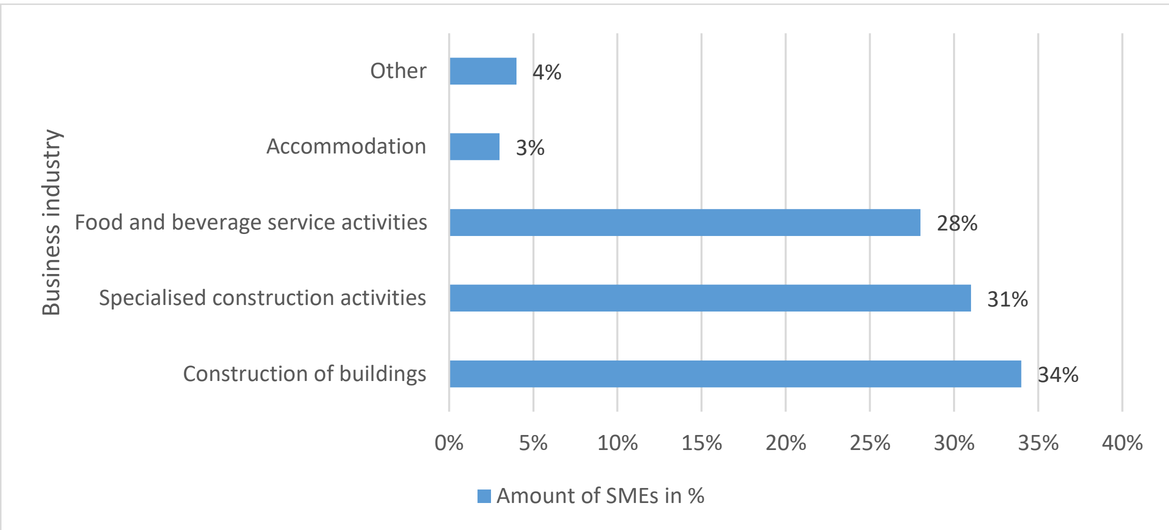


Figure 6. Industry distribution of sampling in percentages

Data in the examination of the payment behaviour in this research consist of 14.3 million purchase invoices, from 6113 SMEs. Data of clearing account consist of debit entries that are made during the fiscal year of 2019 in relation to all transactions in the stated period.

As the sizes of the companies vary, the ratio between debit entries to clearing accounts and total purchase invoices is calculated for the statistical analysis, as that could be a comparable relation that concentrates on total transactions, instead just the number of the debit entries. The amount of entry does not impact examination of clearing accounts in this research. The ratio is attached according to the unique id of 2790 SMEs in sample data that are qualified for financial analysis. If there are not debit entries on specific SMEs, the value in the analysis is 0.

4.2 Differences in real-time data and public data

As this thesis examines the implementation of real-time data into the rating process of credit rating and risk, it is essential to analyse and compare the validity of the data in order to understand the differences between two kinds of financial information. When analysing the differences, the accuracy and quality of the data are to be taken into consideration and evaluated carefully.

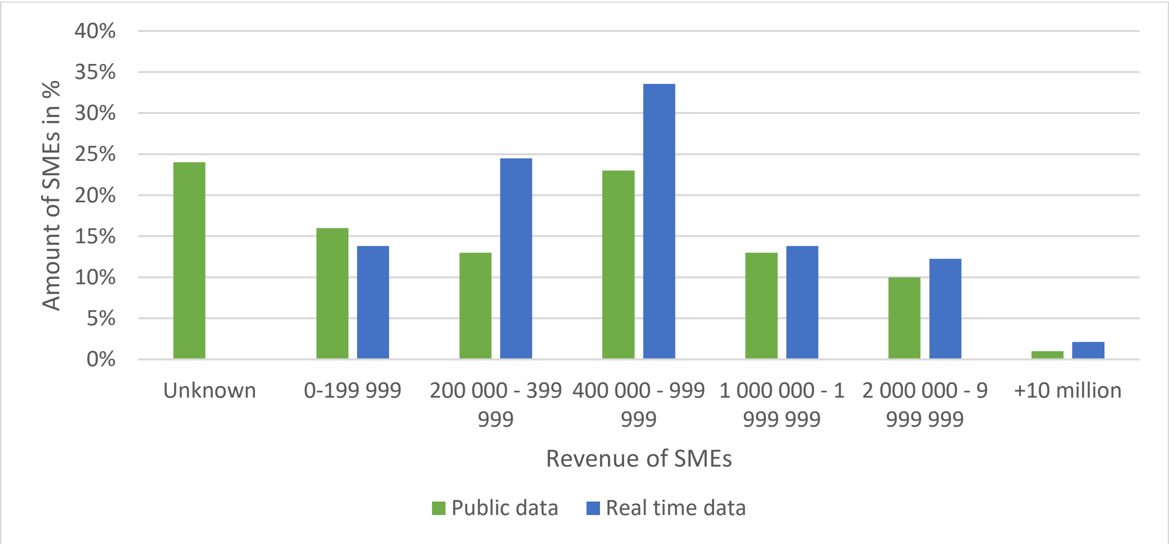


Figure 7. Differences in revenue distribution of public data and real-time data

The quality and availability of financial data cause automated credit rating to be challenging. When comparing simple financial indicators such as the revenue distribution (see figure 7), it is seen that on 24% of the businesses in the sample data revenue is unknown which causes the ratings of agencies to be invalid. Invalid data is caused because all the entities do not have to register their financial statements in-full, according to Finnish Patent and Registration Office (2019) and Accounting Act 3:9 §. In the examination of differences in revenue distribution high variability is seen when analysing the differences in revenue distribution, it is seen than the cause is found from the date of the registration of financial statements.

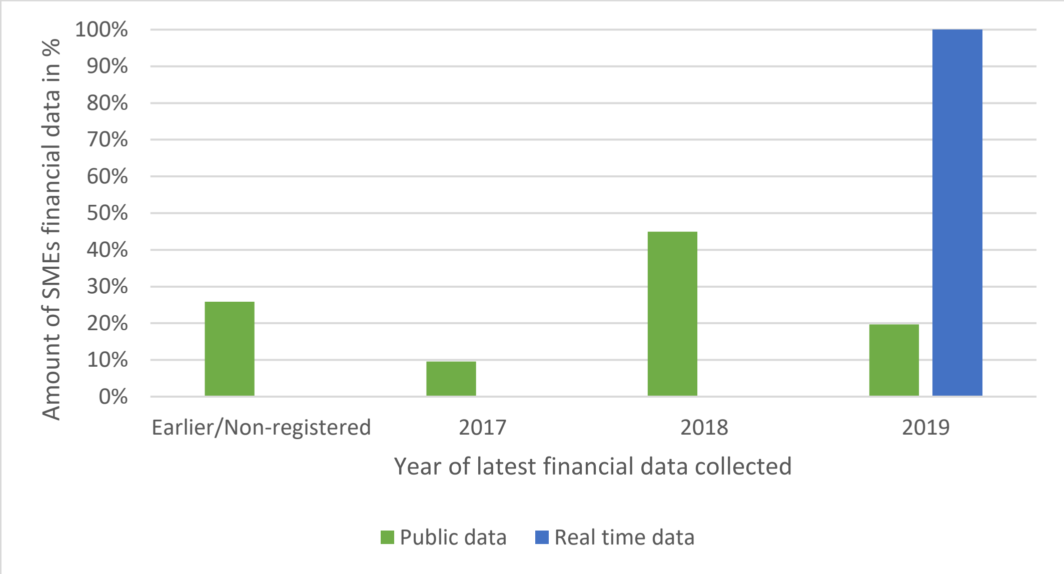


Figure 8. Distribution of fiscal year of latest financial data collected

Figure 8 presents that only 20% of businesses have registered their financial statement of the fiscal year of 2019, which means that 55% of publicly available financial statements, provide financial information from the fiscal year of 2018 or earlier. 26% of the financial statements are considered as not applicable to financial statement analysis as the last financial statement registered is older than from the year 2016 or does not exist. In order to guarantee the comparability of future researches, it is essential to notice that the data is collected on the 1st of May, which is two months before the last day of registration date set in Accounting act 3:9§.

Table 5. The error of margin grouped by years of financial data in comparison to real-time data

Latest financial information collected	The portion that has a significant margin of error (≥ 5%)
Earlier or non-registered	100%
2017	80.0%
2018	80.2%
2019	3.8%

When analysing the influences caused by non-current financial information, it is seen in table 5 that the margin of error that is set to five per cent difference on average of financial information chosen to this research is significantly high among the different years evaluated. None of the financial statements of the year 2016 or earlier could be qualified in financial statement analysis when comparing to set data validation. It is evident that if a financial statement is not registered, financial information cannot be used in the rating of agencies, as it does not publicly exist. Also, in 2017 and 2018, it is seen that roughly 80% of publicly available information on financial statements is invalid due to the registration. The margin of error on financial statements of the year 2019 is 3.8%, which is caused by an error in data collection and faulty forming of databases according to the closer examination. It is also vital to notice that the books are formed by a bookkeeper, and there might be misleading information included in both of the datasets.

4.3 Multinomial logistic regression analysis and supervised regression model

The results of multinomial logistic regression analyses are presented in tables where ‘C’ is referred to as the coefficient of correlation, and ‘p’ is referred to as the p-value. Statistical categories are presented, as stated in chapter 3.

The multinomial logistic regression analysis is executed on the whole sample to see if the same rating model can be used in the classification of both industries together and seek if there is a need for separate ratings. Investigation of the factors examines the accuracy of the classification of the ratings of risk and credit, the maximum classification accuracy of 0.669 on credit rating and 0.701 on risk rating is reached when the model is trained and tested with the whole sampling. In order to increase accuracy, the industry-specific correlation coefficients are examined, and other factors increasing the accuracy are tested.

When analysing the correlation coefficient of explanatory factors to ratings of the whole sample, it is seen that operating margin %, ROCE, ROA, equity ratio, and net gearing are the common factors on statistical categories when evaluating the statistically significant factors in relation to credit ratings. Correlation coefficient and p-values of different factors are presented in table 6. When evaluating the differences on correlations and statistical significance, it is seen that impact of different factors varies on different industries and cannot be classified with the same model, as the factors have different weighted odds in relation to credit ratings. Net gearing seems to have the highest negative correlation and equity ratio the highest correlation in relation to credit ratings. Industry-specific p-values of explanatory factors determine the chosen factors in credit rating models.

Table 6. Correlation coefficient and p-value of different factors in relation to credit rating

Industry	All		Construction		Accommodation and food service activities	
Variables	C	p	C	p	C	p
Revenue growth %	- 0,010	7,78E-01	- 0,013	7,58E-01	0,042	5,21E-01
Operating margin %	0,280	4,45E-15	0,394	6,69E-21	0,241	2,15E-04
ROCE	0,202	2,20E-08	0,178	4,09E-05	0,258	7,05E-05
ROA	0,163	7,01E-06	0,167	1,21E-04	0,162	1,34E-02
Current ratio	- 0,048	1,87E-01	- 0,069	1,16E-01	0,135	4,67E-02
Quick ratio	0,013	7,17E-01	- 0,013	7,71E-01	0,129	4,94E-02
Turnover of receivables	0,062	8,95E-02	0,061	1,64E-01	0,082	2,13E-01
Equity ratio	0,442	1,43E-37	0,458	1,43E-28	0,454	3,39E-13
Net Gearing	- 0,330	1,04E-20	- 0,372	1,10E-18	- 0,258	7,10E-05
Relative Indebtedness	- 0,115	1,53E-03	- 0,146	7,95E-04	- 0,059	3,72E-01

When evaluating the results of multinomial logistic regression analysis of explanatory factors in relation to risk ratings, it can be stated that the different factors are to be used in the classification of risk rating in comparison to credit rating. The impact of different ro credit rating varies, which indicates that the constant factors of SMEs cannot be stated. Net gearing and equity ratio does also have a high impact on risk ratings, but other statistically significant factors according to p-value vary when comparing to results of analysis executed in relation to credit ratings. The analysis confirms that the models are to be built separately for different industries and different ratings. (see table 7)

Table 7. Correlation coefficient and p-value of different factors in relation to the risk rating

Industry	All		Construction		Accommodation and food service activities	
Variables	C	p	C	p	C	p
Revenue growth %	- 0,046	2,09E-01	0,056	2,02E-01	0,021	7,49E-01
Operating margin %	0,137	1,64E-04	0,211	1,06E-06	0,102	1,21E-01
ROCE	0,073	4,48E-02	0,045	3,05E-01	0,138	3,51E-02
ROA	0,128	4,03E-04	0,126	3,94E-03	0,138	3,59E-02
Current ratio	- 0,062	8,90E-02	0,084	5,47E-02	0,105	1,11E-01
Quick ratio	- 0,002	9,49E-01	0,027	5,32E-01	0,106	1,08E-01
Turnover of receivables	0,039	2,79E-01	0,033	4,57E-01	0,070	2,91E-01
Equity ratio	0,254	1,18E-12	0,248	7,90E-09	0,281	1,38E-05
Net Gearing	- 0,185	3,16E-07	- 0,182	2,67E-05	- 0,191	3,43E-03
Relative indebtedness	- 0,040	2,67E-01	- 0,065	1,35E-01	0,006	9,32E-01

It is seen that statistically significant variables vary in different industries. Variables that are significant on credit rating in the examination of the construction industry are operating margin %, ROCE, ROA, equity ratio, net gearing, and relative indebtedness. Statistically significant variables of accommodation and food service activities industry differ as to the current ratio, and quick ratio are significant in addition to significant factors of other industry. It is also seen that Relative Indebtedness is not a significant factor in relation to credit ratings according to the p-values. (see table 6 and table 7)

Therefore, ‘H1: The business industry influences the credit and risk ratings’ is confirmed according to the results of quantitative analysis. It is seen that the constant factors cannot be found from the results as the correlation varies when measuring different variables. A thorough examination results that ratings are to be classified with models that include different explanatory variables of ratings and industries.

When inspecting the p-values and correlation, it is seen that the equity ratio has a very high correlation in measured credit and risk ratings on both industries. Equity measures the solvency and the capability of tolerance for losses, which is seen as a primary indicator when it comes to the solvency and risk. Net gearing has a high negative correlation, which means that as high value tends to correlate on other variables, low value. High net gearing is seen as a factor that increases the risk, which explains the negative correlation to both of the ratings. When examining the correlation coefficient, it is seen that all the variables are not significant, but it can be concluded that profitability and solvency have a statistically significant impact on credit and risk ratings. Liquidity does not have statistical significance when evaluating the risk rating. However, liquidity measures of quick ratio and current ratio are significant factors in relation to credit ratings, according to the p-values in results of the analysis of accommodation and food service activities industry. (see table 8) The construction industry is referred to as industry codes 41-43, and accommodation and food service activities industry is referred to as industry codes 55-56 in table 8.

Therefore, ‘H2: Profitability, solvency, and liquidity all have a significant role in credit and risk ratings’ cannot be fully confirmed as liquidity does not have significance in other ratings than credit rating of accommodation and food service activities industry. The model will be built according to the statistically significant explanatory factors.

Table 8. Statistically significant variables in relation to each rating

	Credit rating			Risk Rating		
Industry / factor	Both	41-43	55-56	Both	41-43	55-56
Revenue Growth %						
Operating Margin %	x	x	x	x	x	
ROCE	x	x	x	x		x
ROA	x	x	x	x	x	x
Current Ratio			x			
Quick Ratio			x			
Turnover of rec.						
Equity Ratio	x	x	x	x	x	x
Net Gearing	x	x	x	x	x	x
Rel. Indebtedness	x	x				

The supervised logistic regression classification algorithm is trained and tested with the variables that are statistically significant in each industry and separately for both of the ratings. Classification accuracies as high as between 93.14% - 97.24% are reached when setting the portion of training data to 0.48 and testing data to 0.52. Classification accuracy of risk in the construction industry is the lowest value, 85.14%, which indicates that the ratios are not as applicable variables as on the accommodation and food service activities industry with an accuracy of 95.52%. The training data and testing data are both halves of the sampling on industry codes 55-56. Accuracies have been reached with a seed value of four on the model of both industries. As the accuracies that are exceeding the value of 85% is accomplished, the algorithm can be built. The algorithm is used in the classification of real-time data on each rating, on both industries separately. (see table 9) The algorithms are a tool for classification, but it is vital to understand the correlation coefficient of different factors in determines the ratings. Algorithms are visualized in the form of a decision tree found in attachments due to the lack of space.

Table 9. supervised classification accuracies of logistic regression algorithms

Industry codes	Classification accuracy of credit grade	Classification accuracy of risk grade
41-43 (Construction)	93.14%	85.14%
55-56 (Accommodation and food service activities)	97.24%	95.52%

4.4 Differences in ratings between two datasets

Before comparing the differences in the datasets, it is vital to understand the distribution of ratings on the overall level to evaluate the results. When examining the overall distribution of ratings that have been rated for Finnish SMEs, it is seen that almost 70% of the businesses are rated in grades A or B, almost one third as C. Only two per cent are rated as grade D. (see table 10) Most of the SMEs that are missing the financial information in public data are rated with lower category of real-time ratings. As small companies in question commonly bear higher risk, it means that the distribution of ratings tends to be weighted down on the scale when implementing the model on real-time rating process.

Table 10. Distribution in ratings of Finnish SMEs (Alma Talent, 2020)

Credit rating Group	% of businesses
A (inc. A+, A, A-)	20.7%
B (inc. B+, B, B-)	47.0%
C (inc. C+, C, C-)	30.3%
D	2.0%

As the ratings have been rated with real-time financial data, the distribution in ratings is changed in each category. The changes are explained mostly because of the improvement in the quality of the data. It is also essential to bear in mind that the accuracy of the classification model has a margin of error that varies roughly from 3% to 15%. The results are presented in separate industries differently. ‘Pcredit and Prisk’ stand for public ratings and ‘Rcredit or Rrisk’ for real-time ratings in illustrations of results.

The differences in credit ratings on the construction industry vary when comparing the ratings. It is seen that the influences of applying the real-time data concentrate mostly as a decreasing factor in ratings. (see figure 9) The influence could be explained partly because of investments and the growth in the industry that drives the solvency down. Distribution of credit rating that is rated with real-time data on the construction industry is quite the opposite when comparing to overall distribution.

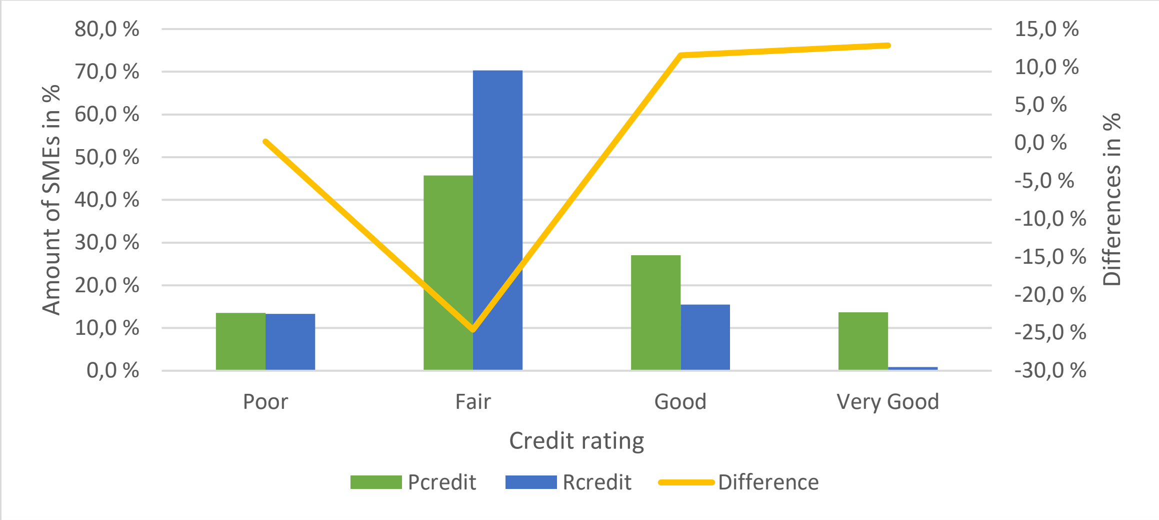


Figure 9. Distribution and differences in credit ratings between real-time data and public data on the construction industry

As the ratings rated with real-time data have a negative impact on the creditworthiness of the SMEs that operates on the construction industry, it also has an increasing impact on the risk ratings. Results are expected as the variables that determine the risk, are impacting the credit rating. One of the variables for explaining the changes in the risk rating with real-time data is the decrease in profitability. It is common for construction companies to have long-term projects, and while the costs are running, the revenue is not found in the data if the revenue related to the project is not continuously deferred on the accrual basis. (see figure 10)

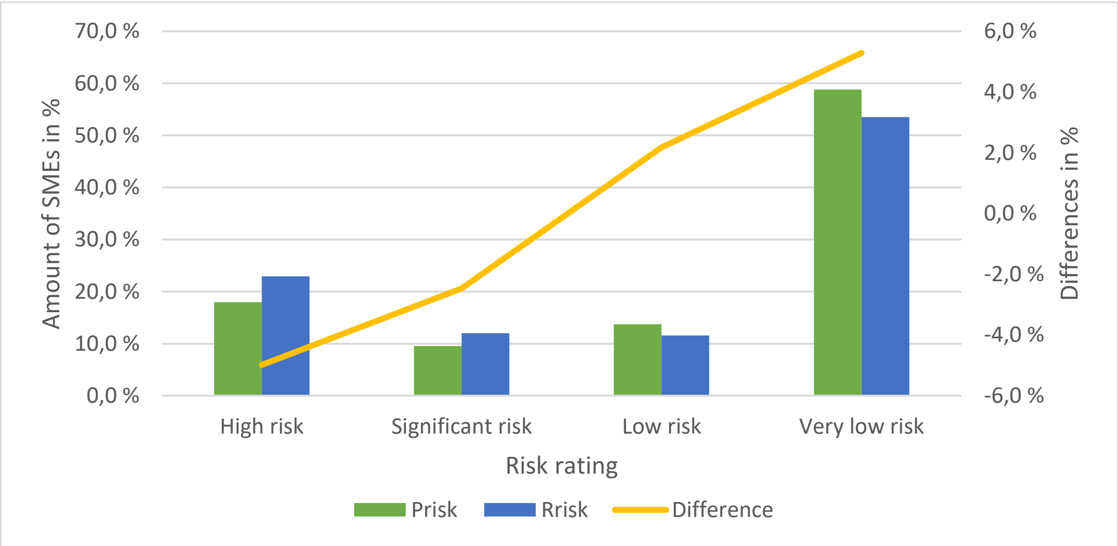


Figure 10. Distribution and differences in risk ratings between real-time data and public data on the construction industry

When evaluating the differences in accommodation and food activities industry in the research, it is seen that the differences are smaller in comparison. Small differences could be caused by the better accuracy of classification in the algorithm, but also caused by differences in variables that are significant according to the correlation coefficient.

The impact of real-time data in terms of differences in credit ratings is not seen as high on the overall view of the industry. Although, the credit ratings are commonly needed for an individual business, which means that the changes in a single rating do have a significant impact as it is concluded that real-time data provides a piece of better information for the credit rating and risk rating. (see figure 11 and figure 12)

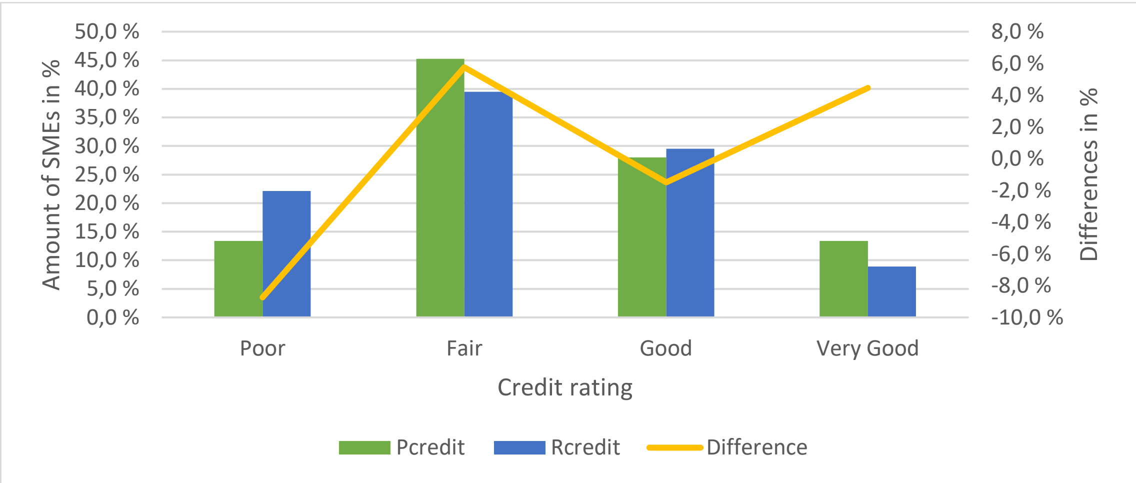


Figure 11. Distribution and differences in credit ratings between real-time data and public data accommodation and food service activities industry

The risk rating of SMEs operating on the accommodation and food service activities industry tends to be rated as a very low-risk grade. The ratings do not differ significantly, which could be an indicator that there might be similar problems related to public data and real-time data on industry in question. It is essential to bear in mind that some of the businesses in the industry might have a large amount of inventory, which must be evaluated correctly. The valuation of inventory has a significant role when evaluating real-time data, but also that it is valuated correctly on the financial statement. Deductions made in the value inventory have an impact on operating margin %, but also on the changes in inventory, which means that faulty valuation will reduce profitability. (see figure 12)

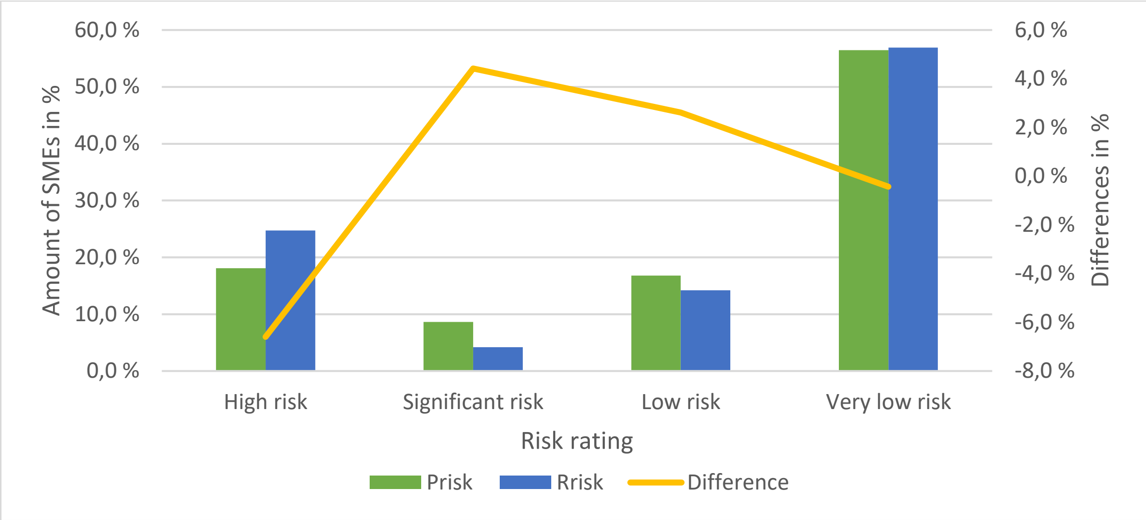


Figure 12. Distribution and differences in risk ratings between real-time data and public data accommodation and food service activities industry

4.5 Influence of payment behaviour to real-time ratings

Before concluding the results, the overall understanding of the payment behaviour of Finnish SMEs is needed. When analysing the purchase invoices of Finnish SMEs, 22.8% of payments have been made on average for the past five years. When comparing different years, the slight increase is seen as the percentage increases to 26.0% in 2018 and 25.6% in 2019. Most of the invoices are paid on the due date, but some are also paid before the due date. (see figure 13)

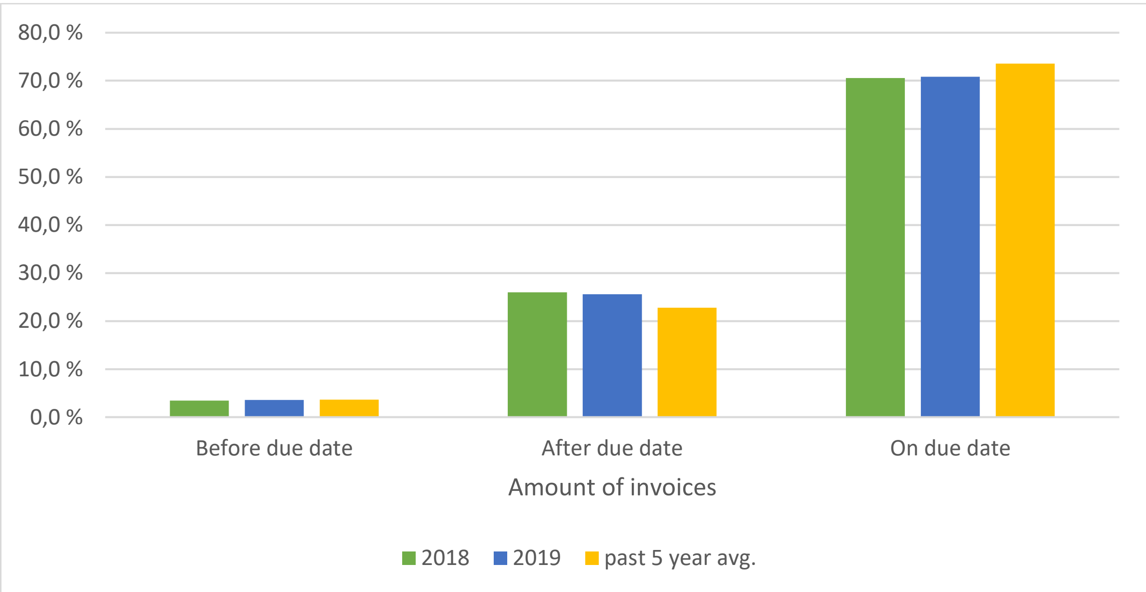


Figure 13. Distribution of payments made before, on or after the due date.

Payments made after the due date seems to be a common phenomenon among the Finnish SMEs, which means that the closer inspection of distribution is needed in order to understand the impact. As late payments are categorized into four classes, it is seen that most of the payments are paid from seven to thirteen after the due date. Analysis of payment behaviour concentrates on the willingness to settle the payment, and its' influence on the creditworthiness and risk, the correlation coefficient of four classes to ratings are analysed. Distribution of late payments is presented in figure 14. Stated four classes are grouped according to the days of payment made after the due date. Classes are 1-6 days late, 7-13 days late, 14-30 days late, and over 30 days late.

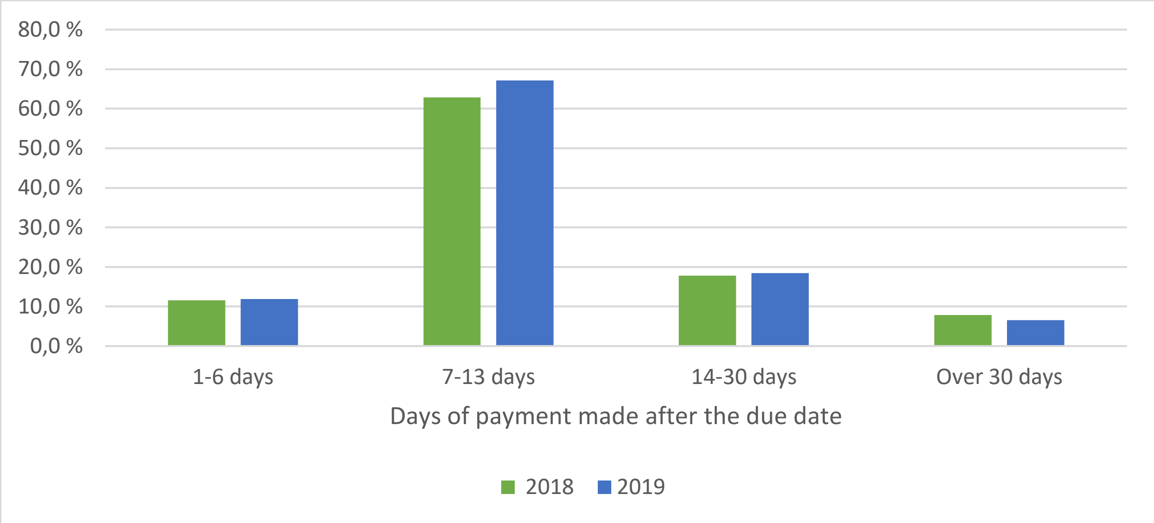


Figure 14. Distribution of late payments in days

When examining the closest comparable information that is used in current credit rating processes with public data, it is seen that as the defaults increase, the risk follows. Payment made over 30 days late is considered as an equivalent for default in this research. The overall view of the relation of defaults to the riskiness of the SMEs operating in Finland is presented in figure 15.

Figure 15. Relation of defaults to the riskiness

The results of multinomial logistic regression are presented in five classes; All late payments (LATEALL), 1-6 days late (LATE1-6), 7-13 days late (LATE7-13), 14-30 days late (LATE14-30), and over 30 days late (LAEO30). In the multinomial logistic regression analysis of the correlation between the classes of late payments to real-time credit rating on the construction industry, it is seen that the influence of late payments to credit rating is not statistically significant according to the p-value and low correlation. (see table 11)

Table 11. Correlation coefficient and p-values of classes of late payments in relation to real-time credit ratings

Credit rating	Correlation coefficient	p-value
LATEALL	-0,015691	4,53E-01
LATE1-6	-0,005447	7,94E-01
LATE7-13	-0,027251	1,92E-01
LATE14-30	-0,033486	1,09E-01
LATEO30	-0,026574	2,04E-01

The late payments tend to have a negatively linear relation to real-time risk rating if the payment exceeds six days from the due date. In other words, as the number of late payments increases and delays in payments in relation to all purchase invoices increases, the riskiness also increases. (see table 12)

Table 12. Correlation coefficient and p-values of classes of late payments in relation to real-time risk ratings

Risk rating	Correlation coefficient	p-value
LATEALL	-0,11475	3,66E-08
LATE1-6	-0,024972	2,32E-01
LATE7-13	-0,1451	3,02E-12
LATE14-30	-0,148242	1,01E-12
LATEO30	-0,1571	4,01E-14

According to the statistical analysis that examines the correlation between classes of late payments and credit ratings on the accommodation and food service activities industry, the results vary when comparing to the construction industry. The negative correlation between the late payments and credit rating in the amount of the late payments and if payments are made one to thirteen days late. Unlike one could assume, the payments made 14 to 30 days and over 30 days late, has a slight positive correlation, which indicates that creditworthiness increases as the amount of the said late payments do. The finding does not mean that SMEs should not pay their bills, but rather that amount of the payments in question compared to the total amount of the purchase invoice could not be the correct measure when analysing the late payments as the relation is higher than on other industry measured. (see table 13)

Table 13. Correlation coefficient and p-values of classes of late payments in relation to real-time credit ratings

Credit rating	Correlation coefficient	p-value
LATEALL	-0,103021	1,10E-03
LATE1-6	-0,112952	3,45E-04
LATE7-13	-0,126321	6,18E-05
LATE14-30	0,01739	5,83E-01
LATEO30	0,061791	5,08E-02

Findings in the relation between the late payments and risk ratings of SMEs of accommodation and food activities industry are almost as surprising as with credit rating. A positive correlation is not detected when analysing the risk, but it is seen that as well as on the construction industry, payments made one to six days late do not have an impact on the risk of the business. Although the payments made late after six days after the due date, the negative correlation to risk decreases as the delay goes further. (see table 14)

Table 14. Correlation coefficient and p-values of classes of late payments in relation to real-time risk ratings

Risk rating	Correlation coefficient	p-value
LATEALL	0,125996	6,46E-05
LATE1-6	0,002362	9,41E-01
LATE7-13	0,230953	1,42E-13
LATE14-30	0,193506	6,83E-10
LATEO30	0,072096	2,26E-02

As the payments made after the due date does have statistical significance on some explanatory variables, but the results are mixed in a way that would ensure the correct method of weighting the influence of late payments, it can be stated that evaluation in the relation of late payments needs improvement. Therefore, ‘H5: Defaults and late payments influence credit and risk ratings’ is partially confirmed according to the results of quantitative analysis, on the condition that the measure can be improved. When evaluating the impact of late payments to credit and risk ratings on different industries, it can be stated that late payments do have a significant impact to other ratings than credit rating on the construction industry, according to the results of multinomial logistic regression analysis. Distribution in late payments of each industry can be found on industry-specific charts in attachments.

4.6 Influence of the clearing account to real-time ratings

As the clearing account is chosen as an indicator of payment behaviour, the correlation coefficient of a factor in relation to ratings is evaluated. After an examination of logistic regression analysis, it can be stated that the clearing account does not have impaction on credit or risk ratings. A correlation coefficient of -0.22% to 0.5% can be found in the construction industry and higher values of 2.1% to 2.7% on accommodation and food activities industry. Debit entries on the clearing account cannot be used as an explanatory indicator in the credit rating process in relation to all the transactions as the correlation coefficient is not significant enough. (see table 15)

Table 15. Correlation coefficient and p-values of clearing accounts in relation to credit and risk ratings

Industry: Construction	Correlation coefficient	p-value	Industry: Accommodation and food service activities	Correlation coefficient	p-value
Credit	0,005778	7,82E-01	Credit	0,027247	3,89E-01
Risk	0,002236	0,914833	Risk	0,020567	0,515928
Clearing	1	0,00E+00	Clearing	1	0,00E+00

When inspecting the explanation for low correlations, the oblivious reason cannot be concluded as there are not past researches to compare the results as the examination of the influence of clearing account to ratings of Finnish SMEs could be something wholly new or past researches are not found publicly. The indicator of debit entries to all transactions might be a completely incorrect factor to measure the influence of clearing account as payment behaviour, or the number of entries might be too low to measure with logistic regression analysis of the variables. Therefore, ‘H4: Increase in the balance of clearing account has detrimental influences on credit and risk ratings’ is disapproved according to the results of quantitative analysis.

5 Conclusion and discussion

5.1 Suggested real-time credit and risk ratings process and model

This thesis aimed to research if real-time data is beneficial in rating methods and process and examine the impact of additional explanatory variables that are not commonly used in a credit rating or risk rating. The theoretical framework that was built for studying the matter according to past studies and researches. The framework worked as a guide for setting the hypotheses. Hypotheses were set for data analysis to separate the industries, cover the variables of financial health and risk, but also to study the factors of payment behaviour. As often seen in the researches that study the ratings, the theory does not provide the correct or complete answer for choosing the variables to examine. This thesis has also shown that not all the factors from past studies have significance in rating and forecasting when evaluating the real-time ratings.

An automated process of rating creditworthiness and risk with real-time data is seen beneficial as the quality of the data is seen to give a better understanding of the current financial status of the company. The data analysis methods of two forms of logistic regression analysis are just commonly used method when evaluating the relation of different factors to ratings, but it cannot be stated that it is the correct or only method for the evaluation. Reverse engineering of the credit rating method with supervised logistic regression classification model has been successfully executed as the aim was to exceed the classification accuracy of 85% on each rating model in different industries. The algorithm built according to the result of this research is implemented in the rating process. It must be pointed out that this model works as a basis and examines the overall benefits of real-time data when implemented on the rating process, and it is improved. The model could be used as a base model for the development of unsupervised logistic regression classification model and possible AEIS-model that concentrates on neural network classification.

The purpose of the hypotheses was to guide the data-analysis in the sense of narrowing down the factors in researching of statistically significant indicators to credit and risk ratings. As assumed, not all the financial indicators have statistical significance in rating or equal correlation coefficient. After a complete evaluation, it can be concluded that H1 is confirmed, but H2 and H5 are partially confirmed in the condition that the measures do not all have significance or measures can be improved. (see table 16) It was seen throughout the research that industry-specific rating models are needed as the factors that impact on ratings is different. Profitability, solvency, and liquidity all do have a significant role in both credit rating and risk rating. Credit rating examines the possibility of default, and the risk rating evaluates the risk of losing it all with the probability of bankruptcy.

Table 16. Final evaluation of hypotheses according to evaluation

Hypothesis	**Approved**	**Disapproved**
H1: The business industry influences the credit and risk ratings	x	
H2: Profitability, solvency, and liquidity all have a significant role in credit and risk ratings	x*	
H3: Company size impacts credit and risk ratings		x
H4: Increase in the balance of clearing account has detrimental influences on credit and risk ratings		x
H5: Defaults and late payments influence credit and risk ratings	x*	

*partially confirmed

The suggested model for real-time rating process is built according to the final evaluation of the researched matter. The research is completed in commissioning company's servers, and all the attributed researched in this thesis are readily built for commissioning company's use. Process chart found in figure 16 is a concluded process of built model, excluding the phases that were executed when researching. The process is built for the explanation of the workflow of the model built. Evaluation indicates that the ratings cannot be done entirely automatically. The suggested model will provide the ratings automatically, but the ratings must be approved by the bookkeeper of the specific SME in order to be validated, as the bookkeeper knows the financial of the customer well. The bookkeeper is also aware if the real-time data is missing relevant information, which causes rating to provide incomplete information.

Suggested automated credit rating and risk rating process for commissioning company consist of 7 steps;

- Delivery of the data from built SQL databases.
- Determining the industry code of the business that is rated.
- Determining the rating that is under classification.
- Classification in cloud servers with the correct algorithm that is built for the industry and rating in question.
- The rating is classified according to the correlation coefficient.
- The bookkeeper that knows the operations and financials of the entity either approves or rejects the rating.
- Rating is added into the train data and test data randomly, which develops the algorithm to give more accurate and reliable ratings continuously.

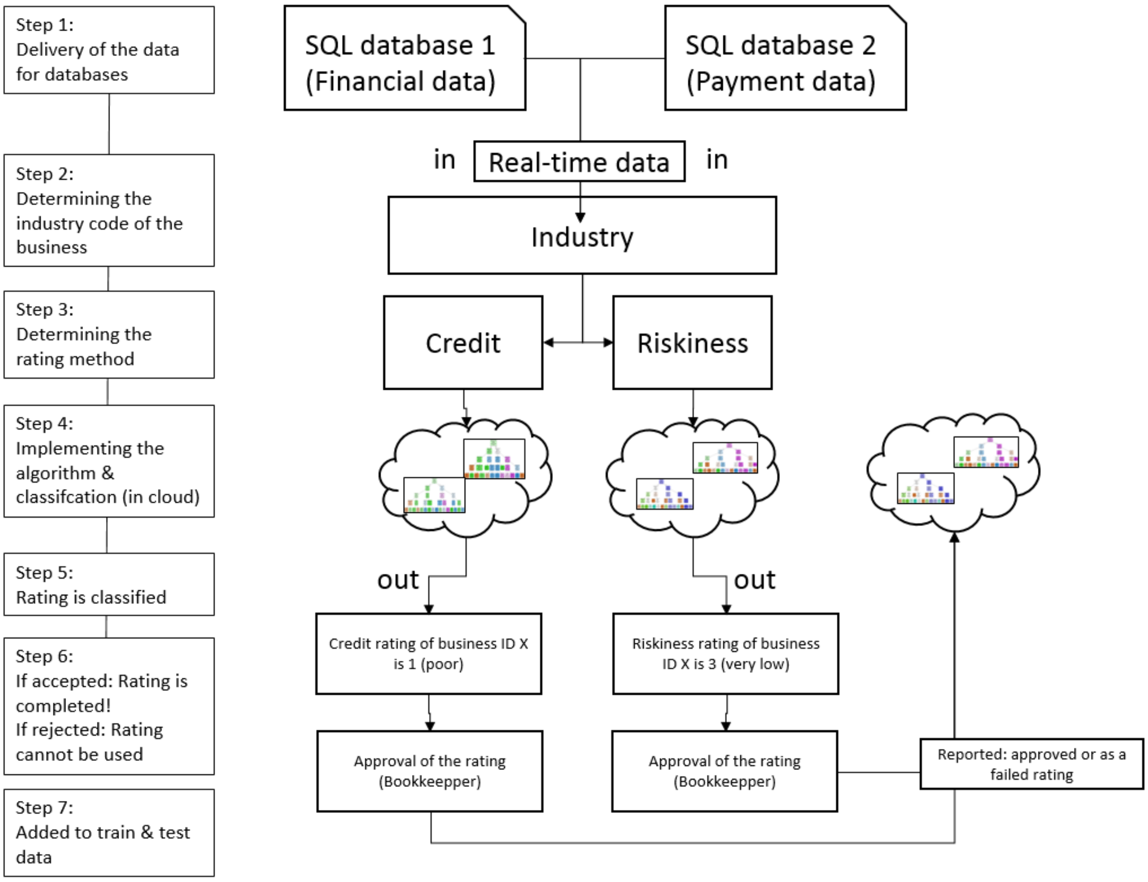


Figure 16. Process chart of automated credit rating and risk rating process with real-time data

5.2 Evaluation of the research and rating model

The aim theoretical evaluation was to provide understanding about the short-term ratings of SMEs that are commonly based on the financial statement analysis and examines the contents of the financial statement. The needs and inadequacies guided the research with the information of the demands for certain kind of rating model. The needs for ratings are seen as constantly changing and growing. The ratings that were examined are credit rating and risk rating. Credit rating is seen as an indicator for the capability to settle the payments, and risk rating is seen as an indicator of forecasting the financial difficulties and bankruptcy. Plenty of past research was studied, and the building of overall knowledge of rating processes and methods was built accordingly in order to understand the phenomenon related to the current state of rating the Finnish SMEs.

Publicly available data has been stated to cause issues when used in rating as the financial information of SMEs can be faulty or missing, it does not mean that real-time data is a complete and only solution for ratings. Both of the datasets do bear possible risks within. Although, real-time data benefits the rating process in terms of rating with financial information that provides an understanding of current financial status, instead of possible determining the rating with information that is seen providing incomplete information. Also, the manipulation of the financial ratios is prevented with current financial data. There is no room for errors in data, and it must be formed correctly according to the chart of accounts. As the real-time data is something that covers the financial information at any point of the fiscal year, from the last 12 months, there could be some problems related to the automated credit and risk rating processes.

It is essential to understand the rating methods of Finnish rating agencies, which might be specially built for analysing the past financial data. Model of agencies cannot be stated to be applicable in the classification of real-time data without comparing the ratings of different years as the data has significant differences when comparing. Also, the qualitative factors such as the interviews and weight of the person running the operations are a critical factor commonly when rating long-term ratings, which did not impact the ratings on this research.

Randomly chosen sampling can also cause an issue if the extreme impact values cause changes in deviation. In past researches, classification of models reached the accuracy of roughly 80%, which is seen to be improved with real-time data. However, it is to be considered if the original ratings are even correct ones to use in training and testing of the real-time classification model. As stated, the accuracies of the supervised logistic regression classification models are on a decent level, but there is room for improvement. Multicollinearity could have been caused a misleading coefficient of correlation if the variables are too similar, such as the current ratio and quick ratio. High impacts of net gearing and equity ratio in relation to ratings could have been causing faulty conclusions when variables are not analysed separately in multinomial logistic regression analysis.

It can be argued if the financial ratios and payment information is enough information to evaluate the creditworthiness and risk of the entity at all. Some of the falsities can also be found from the sources of the information. The material consists of the documents that are delivered for accounting. Therefore, it can be stated that the data is just as accurate as of the accuracy in the delivery of the material. The other factor that is seen problematic are the entries that bookkeeper has made. There could be issues related to the valuation of inventory if the list is not delivered or have not been deferred according to accrual accounting. There might be similar kinds of problems related to the reconciliation of revenue if the revenue consists of different projects and is paid in advance, noun, or does not match the expenses that are or are not running. All the other write-offs and depreciation are a problematic factor in not only when the ratings have consisted of real-time data, but also if the financial statement is correct when registered. The stated problem could have significant offsets in profitability, solvency, and liquidity, which could provide faulty ratings.

The research was executed with a massive amount of data. The mining of the data took a lot amount of time but was successful. It must be pointed out that mining requires resources and should not be considered if the dataset is not massive. Combining the data into financial ratios was done successfully, according to the account-specific formulas that had to be formed. Reserve-engineering of the rating models was done in order to evaluate the ratings and compare the two datasets, and high accuracies were accomplished. Differences in the comparison indicate that the accuracy of rating increases when customizing the rating method according to significant industry-specific variables. Late payments information provides benefits when evaluating the risk, but the evaluation needs to be improved if implemented into the rating models. The relation between clearing account to all transactions cannot be used as a measurement in rating models as it does not provide significant information in risk related to the payment behaviour. Some of the problems related to the current rating processes can be tackled with the suggested rating model, but the process must include the input of bookkeeper and cannot be completely automatic. It can be stated that real-time financial data benefits the automated rating processes when the built method is implemented in the process.

This thesis examined the possibility of determining the rating with financial indicators and payment behaviour. Factors are seen as a proper method for the rating according to past studies. Suggested model in this research is one of many ways to determine the ratings. It cannot be stated that any of the models used in ratings are correct as the studies and researches indicate a different kind of result on a different approach.

Researches done in continuous of this research could be to evaluate the financial variables on other industries and examine the variables with businesses that operate in other countries than Finland. The research can also be conducted on unsupervised Neural Network model, which would be an interesting subject for future research, especially when comparing the results.

This research confirmed that real-time data can be implemented to credit and risk rating process of the accounting firm, but also that there are benefits in real-time rating. Some of the issues related to the validation of data can be avoided when implementing real-time data in ratings. Payment behaviour in terms of late payments does partially have an impact on industry-specific ratings, although it could be misleading as an explanatory variable. Clearing account in relation to all transactions does not have a statistically significant influence on either credit rating or risk rating according to the result of this research.

References

Accounting Act 1336/1997.

Accounting Board 1653/2001.

Alessi, C. 2013. The Credit Rating Controversy. Council on Foreign Relations. URL: www.cfr.org/financialcrises/credit -rating-controversy. Accessed: 10 February 2020.

Alma Talent 2020. Menestyjäluokitus. URL: https://www.almatalent.fi/tietopalvelut/tun-nuslukuopas/kriisitunnusluvut/menestyjaluokitus. Accessed: 10 May 2020.

Altman, E. 1980. Commercial bank lending: Process, credit scoring, and costs of errors. Journal of Financial and Quantitative Analysis, 15, pp. 813-832.

Altman, E. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance, 23, 4, pp. 589-609.

Amato, J. & Furfine, C. 2004. Are credit ratings procyclical?. Journal of Banking and Fi-nance, 28, pp. 2641-2677

Asiakastieto 2020. Rating-Alfa. URL: https://www.asiakastieto.fi/yritykset/fi/suomen-asia-kastieto-oy/01110279/rating-alfa. Accessed: 10 May 2020.

Aziz, M. & Dar, A. 2006. Predicting Corporate Bankruptcy: Where we stand?. Corporate Governance International Journal of Business in Society, 16, 1, pp. 18-33.

Cifuentes, P. 2013. Get Technical to fix rating agencies. Financial Times. London.

Cox, M. & Ellsworth, D. 1997. Managing Big Data for Scientific Visualization. Communica-tions of the ACM 42, 8, pp. 82-90.

Creighton, E. 2014. Credit Rating Agencies: An Overview. Annual Review of Financial Economics, 5, 1, pp. 93-122.

Diebold, F. 2004. Modelling and Forecasting realized volatility. Econometrica, 71, pp. 579-625.

Duff, A. 2015. The Credit Industry: Stakeholders Perspectives of Competitive, Regulatory and Funding Issues. University of West Scotland.

Feldmann, H. & Read, P. 2013. Financial system sophistication and unemployment in industrial countries. International Journal of Finance and Economics, 18, 4, pp. 319-338.

Finnish patent and registration office 2019. Tilinpäätösilmoitus. URL: https://www.prh.fi/fi/kaupparekisteri/tilinpaatos.html. Accessed: 10 May 2020.

Foss, M. 2010. Enron and the Energy Market Revolution. University of Houston. Law Center.

Frost, C. 2007. Credit rating agencies in capital markets: A review of research evidence on selected criticism of the agencies. Journal of Accounting, 22, pp. 469-492.

Gonzalez, F., Haas, F., Johannes, R., Persson, M., Toledo, L., Violi, R., Wie land, M. & zins, C. 2004. Market dynamics associated with credit ratings - a literature review. ECB Occasional Paper. No 16. European Central Bank. Frankfurt.

Gropelli, A. & Nikbakht, E. 1990. The Credit Rating Agencies. Journal of Finance, pp. 431-451.

Holopainen, M. & Pulkkinen, P. 1999. Tilastolliset menetelmät. Helsinki: WSOY.

Horrigan, J., Peck, L. & West, R. 1966. The Determination of Long-Term Credit Standing with Financial Ratios. Journal of Accounting Research, pp. 44-62.

Hyvärinen, M. 1995. Pienten ja keskisuurten yritysten reittaus. Kauppa- ja teollisuusministeriö, Tutkimuksia ja raportteja.

Ikäheimo, S., Laitinen, E. & Laitinen, T. 2014. Yrityksen taloushallinto tänään. Vaasan Yritysinformaatio.

Jollineua, S., Tanlu, L. & Winn, A. 2014. Evaluation Proposed Remedies for Credit Rating Agency Failures. The Accounting Review. 89, 4, pp. 1399-1420.

Junkkari, M. 2008. Markkinoiden diktaattori toimii viimein. Helsingin Sanomat. URL: https://www.hs.fi/paakirjoitukset/art-2000002517256.html. Accessed: 3 April 2020.

Kagan, J. 2020. Credit Rating. URL: https://www.investopedia.com/terms/c/creditrating.asp. Accessed: 9.5.2020.

Kenton, W. 2020. Pearson Coefficient. Investopedia. URL: https://www.investopedia.com/terms/p/pearsoncoefficient.asp. Accessed: 9 May 2020.

Koulu, R. 2009. Konkurssioikeus. Helsinki. Talentum.

Krahnen, J. & Weber, M. 2001. Generally accepted rating principles: A primer. Journal of Banking and Finance, 25, 1, pp. 3-23.

Kronwald, C. 2009. Credit Rating and Impact on Capital Structure. GRIN. Verlag.

Laitinen, E. & Laitinen, T. 2004. Yrityksen rahoituskriisin ennustaminen. Jyväskylän Kirjapaino Oy. Jyväskylä.

Laitinen, E. 2015. A probability tree model of audit quality. European Journal of Operational Research, 243, pp. 665-677.

Laitinen, E. 1992. Konkurssin ennustaminen. Vaasan Yritysinformaatio. Vaasa.

Laitinen, E. 2001. Maksuhäiriöyritykselle löytyi viisi varmaa tunnusmerkkiä. Luottolista. 20. pp. 12-14.

Layney, D. 2001. Application Delivery Strategies. Meta Group Research. Note 6.

Leppiniemi, J., Leppiniemi K. & Kaisanlahti, T. 2017. Tilinpäätöksen tulkinta. ALMA. Finland.

Mashey, J. 1998. Big Data and the Next Wave of Infrastress. Usenix. Monterey. USA.

Metsämuuronen, J. 2009. Tutkimuksen tekemisen perusteet ihmistieteissä 4th edition. Gummerus Kirjapaino Oy. Jyväskylä.

Moody's 2020. Issuer/Investor Requested Ratings Process Detail. URL: https://www.moodys.com/sites/products/ProductAttachments/SP32056_MIS_Ratings%20Process%20Diagram_v13.pdf. Accessed: 20 March 2020.

NRSRO USSEC 2019. Annual Report on Nationally Recognized. URL: https://www.sec.gov/files/2018-annual-report-on-nrsros.pdf . Accessed. 5 April 2020.

O'Brien, R. 2007. A Caution Regarding Rules of Thumb for Variance Inflation. Quality & Quantity. 41, 673–690

Ross, S., Westerfield, R. & Jordan, R. 2020. Essentials of Corporate Finance 10th edition. McCraw Hill. London.

Servigny, D. & Renault, O, 2004. Measuring and Managing Credit Risk. McGraw Hill Professional. London.

Sohn, S. & Kim, Y. 2013. Behavioral credit scoring model for technology-based firms that considers uncertain financial ratios obtained from relationship banking. Small Business Economics, 41,4, pp. 931-943.

Standard & Poor's rating services. 2019. Guide to Credit Rating Essentials. McGraw Hill Financial. London.

Talenom 2019, Talenom julkaisee tilitoimistoasiakkailleen uuden laskurahoituspalvelun. URL: https://www.talenom.fi/talenom-julkaisee-tilitoimistoasiakkailleen-uuden-laskurahoituspalvelun/. Accessed: 23 February 2020.

Statistics Finland 2003. Käsitteet: PK-Yritys. URL: https://www.stat.fi/meta/kas/pk_yritys.html. Accessed: 14 May 2020.

Treacy, W.F. & CAREY, M. 2000. Credit risk rating systems at large US banks. Journal of Banking and Finance, 24, 1/2, pp. 167-201.

Troy, S. 2019. Investopedia Stock Analysis: How Big Data Has Changed Finance. Investopedia. URL: https://www.investopedia.com/articles/active-trading/040915/how-big-data-has-changed-finance.asp. Accessed: 14 April 2020.

Uusitalo, H. 1991. Tiede, tutkimus ja tutkielma: johdatus tutkielman maailmaan. WSOY. Helsinki.

Weiss, S. & Indurkhya, N. 1998. Predictive Data Mining. Morgan Kaufmann.

Ylijoki, O. and Porras, J. 2016. Conceptualizing Big Data: Analysis of Case Studies. Intelligent Systems in Accounting, Finance and Management, 23, 4, pp. 70-79.

Yritystutkimusneuvottelukunta 2011. Yritystutkimuksen tilinpäätösanalyysi 9th edition. Gaudeamus. Helsinki.

Appendices

Appendix 1. Ranges of account of financial indicators

Indicator:	Accounts:
Turnover	3000 – 3599
Operating earnings	3000 – 8999
Net profit (loss)	3000 – 9997
Depreciation, amortization, and reduction in value	6800 – 6999
Other operating income	3650 – 3999
Expenses Operating Expenses	4000 – 9694 4000 – 6690
Financial expenses	9300 – 9699
Taxes	9900 – 9989
Balance sheet total	1000 -1999 or 2000 -2999
Stocks	1500 – 1599
Short-term receivables	1700 – 1849
Cash at bank and in hand	1900 – 1999
Investments	1860 – 1899
Short-term liabilities	2800 – 2979
Advances received	2860 – 2869
Receivables	1600-1632 + 1641 + 1642 + 1741 + 1742 + 1700-1732
Equity	2000 - 2389 + 3000 – 9997
Provisions	2500 - 2599
Cumulative accelerated depreciation	2400 – 2499
Taxation-based reserves	2770 - 2799 (short-term) 2980 - 2999 (long-term)
Subordinated debt	2380 - 2389 + 2600 - 2609 + 2800 - 2809
Capital	2000 - 2389 + 3000 – 9997 + 2400 – 2499 + 2981 + 2380 – 2389 + 2600 – 2609 + 2800 – 2809
Interest-bearing liabilities	2611 – 2651 + 2671-2691 + 2811 – 2851 + 2891

Non interest-bearing liabilities	2660-2668 + 2701 – 2751 + 2860 – 2889 + 2901 – 2979
Liquid assets	1700 – 1849
Debt on the balance sheet	2601 – 2979
Clearing account (other receivables)	1777

Ratio:	**Formula (ranges from chart of accounts):**
Revenue Growth %	(3000-3599) / (3000-3599)
Operating profit %	100% * [(3000-8999) + (6800-6999)] / (3000-3599)
Profit margin %	[(3000-3599) + (3650-3999) – (4000 -6690) – (6800 – 6999)] / (3000-3599)
ROCE (Return on capital employed)	100% * [(3000 – 6999) + (9300 – 9699) + (9900 – 9989)] / [(2000 – 2389) + (3000 – 9997)] + [(2611 -2651) + (2671-2691) + (2811 – 2851) + (2891)] (avg. of beginning balances and ending balances)]
ROA (Return on assets)	100% * [(9998 – 9999) + (9300 – 9699) + (9900 – 9989)] / (1000 - 1999)
Current ratio	[(1500 – 1599) + (1900 – 1999) + (1860 + 1899)] / (2800 – 2979)
Quick ratio	[(1700 – 1849) + (1900 + 1999) + (1860 – 1899)] / [(2800 – 2979) – (2860 – 2869)]
Turnover of receivables	365 * [(1600 – 1632) + (1641 – 1642) + (1741 – 1742) + (1700 – 1732)] / (3000 – 3599)
Equity ratio	(2000 - 2389) + (3000 - 9997) - (2400 - 2499) - (2770 - 2799) - (2980 - 2999) + (2380 - 2389) + (2600 - 2609) + (2800 - 2809)] / [(1000-1999) - (2860-2869)]
Net gearing	100% * [(2611 - 2651) + (2671 - 2691) + (2811 - 2851) + 2891 - (1700 - 1849)] / [(2000 - 2389) + (3000 - 9997) - (2400 - 2499) - (2770 - 2799) - (2980 - 2999) + (2380 - 2389) + (2600 - 2609) + (2800 - 2809)]
Relative indebtedness	(1700 - 1849) / (3000 - 3599)

Appendix 2. Results of logistic regression analysis and algorithms

RELATION OF FINANCIAL RATIOS TO RATINGS:

Relation of variables to credit rating	**(Public data)**	
Variable	**Correlation coefficience**	**p-value**
Revenue growth %	-0,0102753806897221	0,7777492776494460
Operating margin %	0,2797547985978470	0,0000000000000045
ROCE	0,2016124724927320	0,0000000220010700
ROA	0,1625006402696990	0,0000070063995866
Current ratio	-0,0480212944802195	0,1868956586789980
Quick ratio	0,0131790309197665	0,7173374312977990
Turnover of re-ceivables	0,0617602114172876	0,0894962518156533
Equity ratio	0,4421272185400550	0,0000000000000000
Net Gearing	-0,3301770633351620	0,0000000000000000
Relative Indebtedness	-0,1149778007530050	0,0015309928123781
Riskiness	0,4474206869995950	0,0000000000000000
Rating class	1,0000000000000000	0,0000000000000000

Relation of variables to riskiness rating	**(Public data)**	
Variable	**Correlation coefficience**	**p-value**
Revenue growth %	-0,0457560583127497	0,2085749274309820
Operating margin %	0,1365494429111770	0,0001642791457217
ROCE	0,0729373852805509	0,0448425115824569
ROA	0,1282856012602980	0,0004025591254314
Current ratio	-0,0618623282696297	0,0889666467980829
Quick ratio	-0,0023128033790964	0,9493456231186510
Turnover of re-ceivables	0,0394056371929073	0,2788877985702080
Equity ratio	0,2544918434550490	0,0000000000011793
Net Gearing	-0,1845655334861110	0,0000003159575176
Relative indebtedness	-0,0404312239083048	0,2665572023443960
Riskiness	1,0000000000000000	0,0000000000000000
Rating class	0,4474206869995950	0,0000000000000000

Relation of variables to credit rating	**(Public data: industry co-des 41-43)**	
Variable	**Correlation coefficience**	**p-value**
Revenue growth %	0,0134872670311853	0,7578468104922300
Operating margin %	0,3936092648379950	0,0000000000000000
ROCE	0,1780309723216690	0,0000409071341425
ROA	0,1670264987698570	0,0001205630984629
Current ratio	0,0686603954724691	0,1161101503089300
Quick ratio	0,0127138424027802	0,7713357553895530

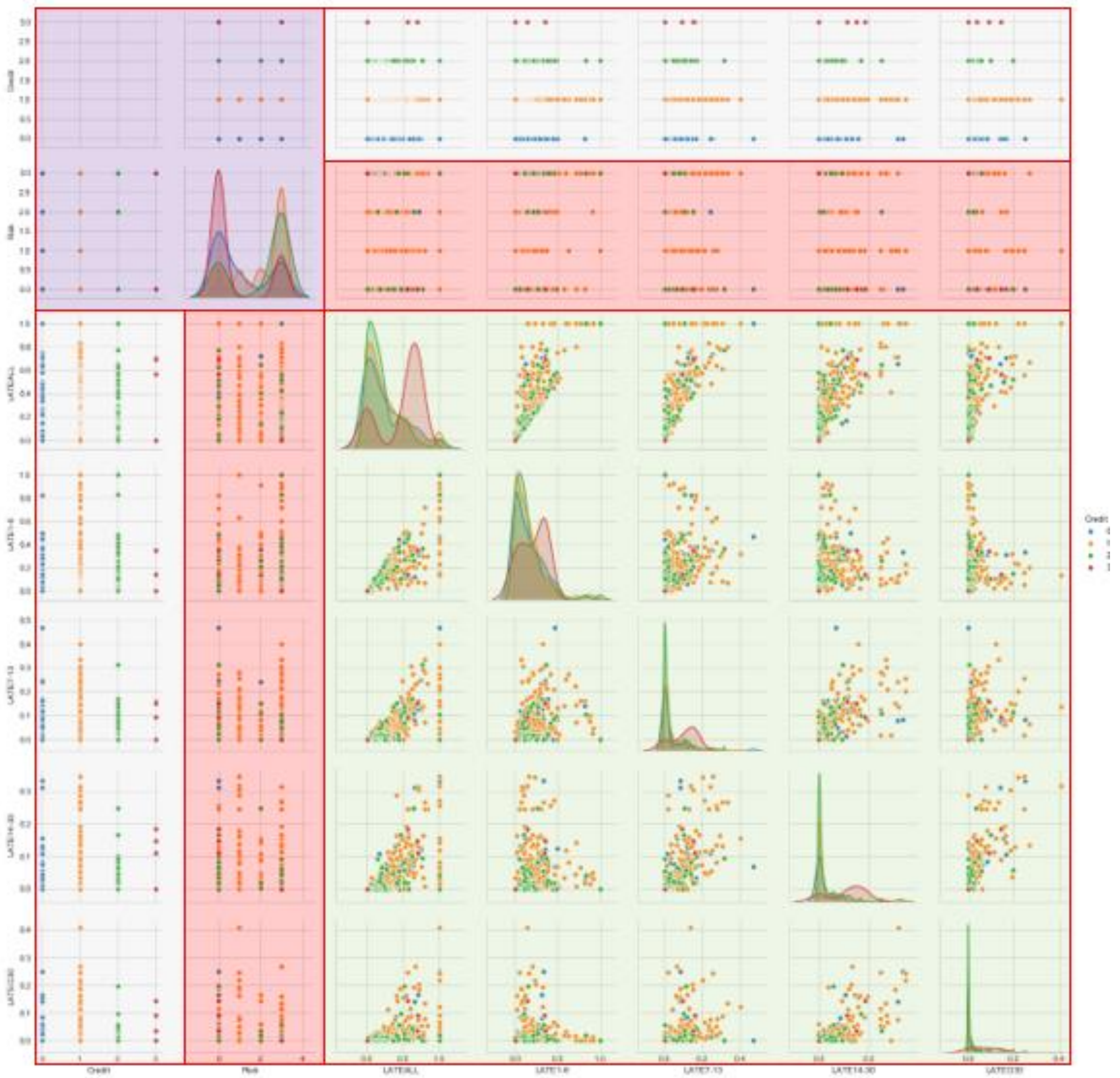
Turnover of receivables	0,0608958548027409	0,1635391249254580
Equity ratio	0,4579038907927070	0,0000000000000000
Net Gearing	0,3720906850252400	0,0000000000000000
Relative indebtness	0,1459657114279570	0,0007950690449228
Relation of variables to riskiness rating	**(Public data: industry codes 41-43)**	
Variable	**Correlation coefficience**	**p-value**
Revenue growth %	0,0557972097218427	0,2018064186435010
Operating margin %	0,2110712482178520	0,0000010619087041
ROCE	0,0448173219290011	0,3053814726298020
ROA	0,1256157199962270	0,0039417763328703
Current ratio	0,0838925048585958	0,0547286113415298
Quick ratio	0,0273102106045962	0,5323777448847550
Turnover of receivables	0,0325209178173194	0,4571376127495400
Equity ratio	0,2484840161700160	0,0000000078959974
Net Gearing	0,1821912707552910	0,0000267178186145
Relative indebtness	0,0652404271419261	0,1354702557712510
Relation of variables to credit rating	**(Public data: industry codes 55-56)**	
Variable	**Correlation coefficience**	**p-value**
Revenue growth %	0,0423391659679390	0,5210709009754080
Operating margin %	0,2406996615843850	0,0002147636227534
ROCE	0,2579148634889740	0,0000704748681620
ROA	0,1621483773722020	0,0134052888336351
Current ratio	0,1352877777286520	0,0467106785680385
Quick ratio	0,1286082401527470	0,0494117021711800
Turnover of receivables	0,0819965057426728	0,2133950627232750
Equity ratio	0,4539738468445920	0,0000000000003390
Net Gearing	0,2578090363714330	0,0000709764537181
Relative indebtness	0,0589210354986349	0,3716497201340160
Relation of variables to riskiness rating	**(Public data: industry codes 55-56)**	
Variable	**Correlation coefficience**	**p-value**
Revenue growth %	0,0210939038736196	0,7492769367039100
Operating margin %	0,1020311908055390	0,1212025537178750
ROCE	0,1384110815754560	0,0351193629476458
ROA	0,1378611980904980	0,0358585523079465
Current ratio	0,1048933261555990	0,1110577803581110
Quick ratio	0,1059393460071910	0,1075209126170710
Turnover of receivables	0,0696590729485374	0,2907085705703100
Equity ratio	0,2810905491499930	0,0000138473054236

Net Gearing	0,1913944441857060	0,0034269959144059
Relative indebtness	0,0056025159009891	0,9323609877882680

RELATION OF LATE PAYMENTS TO REAL-TIME RATINGS:

Relation of late payments to credit rating (Real-time: industry codes 41-43)
(Correlation coefficient, p-value)
Credit (1.0, 0.0)
Risk (0.16526326997922236, 1.7348225871777886e-15)
LATEALL (-0.01569070535816934, 0.4529546001207578)
LATE1-6 (0.005446890477035332, 0.7944666913885748)
LATE7-13 (-0.027251433152381425, 0.1923636412715462)
LATE14-30 (-0.03348607475052033, 0.10915227003508976)
LATEO30 (-0.02657436741807774, 0.20365094034794112)

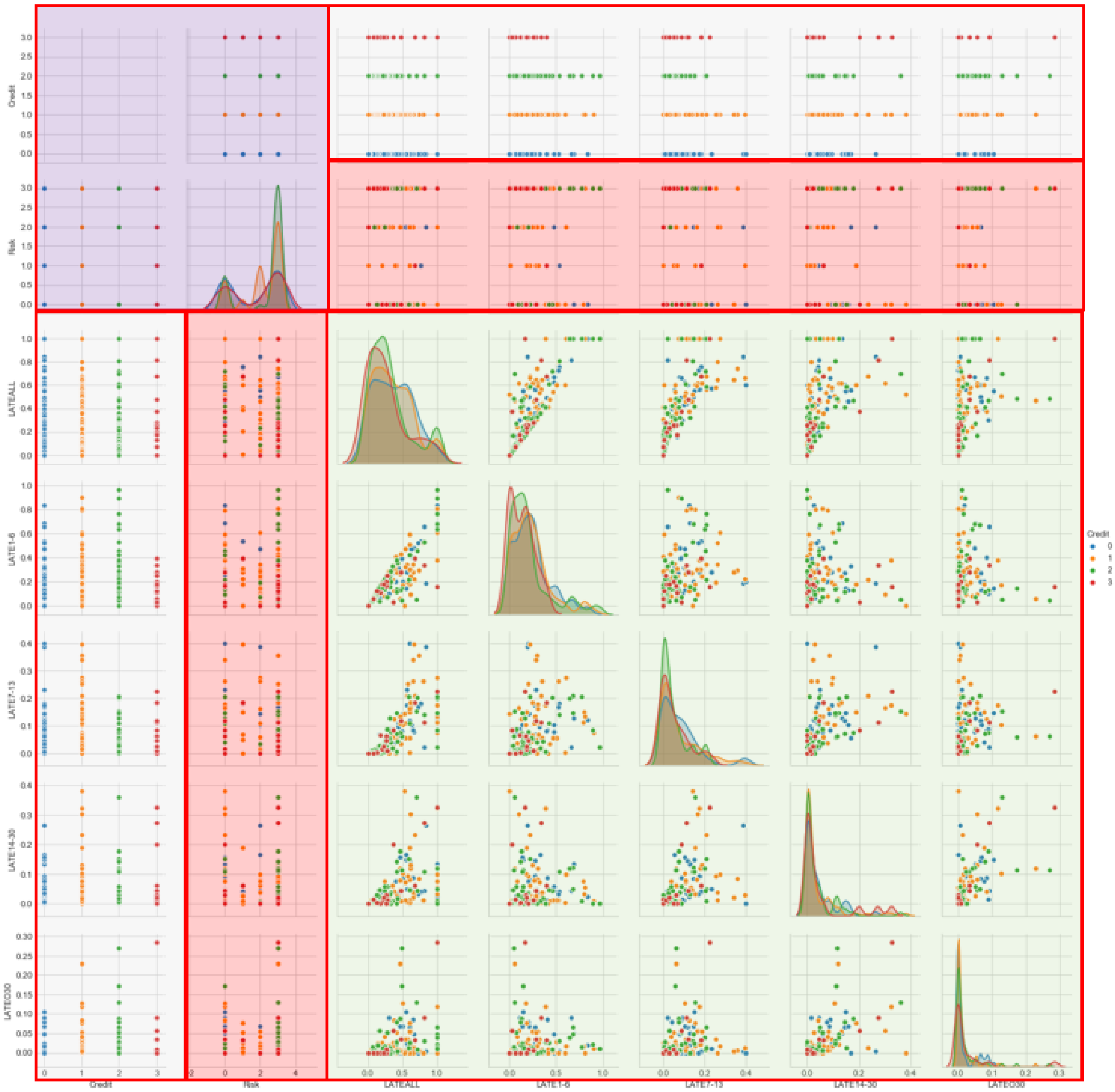
Relation of late payments to risk rating (Real-time: industry codes 41-43)
(Correlation coefficient, p-value)
Credit (0.16526326997922236, 1.7348225871777886e-15)
Risk (1.0, 0.0)
LATEALL (-0.11474973245483747, 3.66190918822138e-08)
LATE1-6 (-0.024971618710060414, 0.2322729547220757)
LATE7-13 (-0.14510004057065362, 3.020439111706248e-12)
LATE14-30 (-0.14824172829538337, 1.0073235309212601e-12)
LATEO30 (-0.15710002528655811, 4.0077083398247024e-14)



Relation of late payments to credit rating (Real-time: industry codes 55-56)
(Correlation coefficient, p-value)
Credit (1.0, 0.0)
Risk (0.10705778320360931, 0.0006964320622221689)
LATEALL (-0.10302096148313397, 0.0011046605197664491)
LATE1-6 (-0.11295194515600496, 0.00034511564274019397)
LATE7-13 (-0.12632125869582408, 6.183699418472358e-05)
LATE14-30 (0.01739015261915351, 0.5828142405485378)
LATEO30 (0.061791152118289984, 0.05076857650801949)

Relation of late payments to risk rating (Real-time: industry codes 55-56)
(Correlation coefficient, p-value)
Credit (0.10705778320360931, 0.0006964320622221689)
Risk (1.0, 0.0)
LATEALL (-0.12599603095866957, 6.461436303931268e-05)

LATE1-6 (-0.00236194894100706, 0.9405345099324868)
LATE7-13 (-0.23095293633825503, 1.4201638946437103e-13)
LATE14-30 (-0.19350553057657166, 6.831816960421743e-10)
LATEO30 (-0.07209577953928606, 0.022608593968243065)



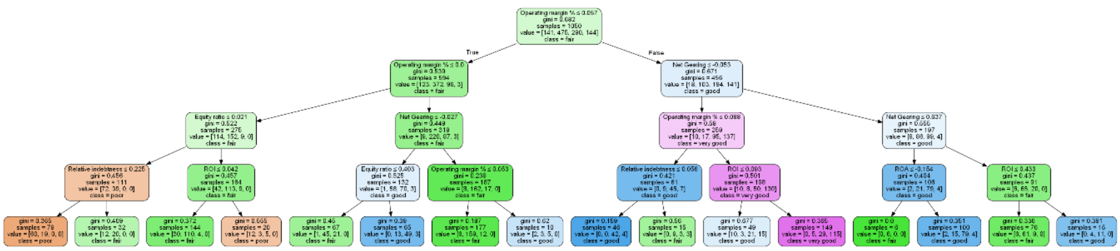
RELATION OF CLEARING ACCOUNT TO REAL-TIME RATINGS:

Relation of late payments to ratings (Real-time: industry codes 41-43)
(Correlation coefficient, p-value)
Credit (0.005777652961771537, 0.7822909039947935)
Risk (-0.002236039172894084, 0.9148326086950281)
Clearing (1.0, 0.0)

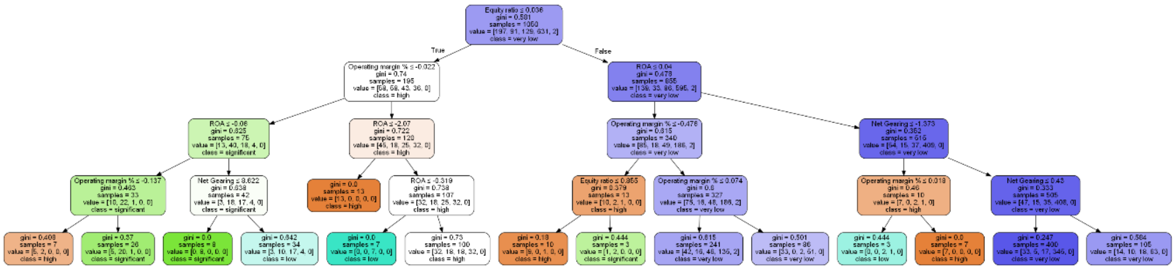
Relation of late payments to ratings (Real-time: industry codes 55-56)
(Correlation coefficient, p-value)
Credit (0.027246833605875702, 0.3894018173329197)
Risk (0.0205668052412201, 0.5159281762177698)
Clearing (1.0, 0.0)

SUPERVISED LOGISTIC REGRESSION CLASSIFICATION ALGORITHMS USED IN REAL-TIME RATING:

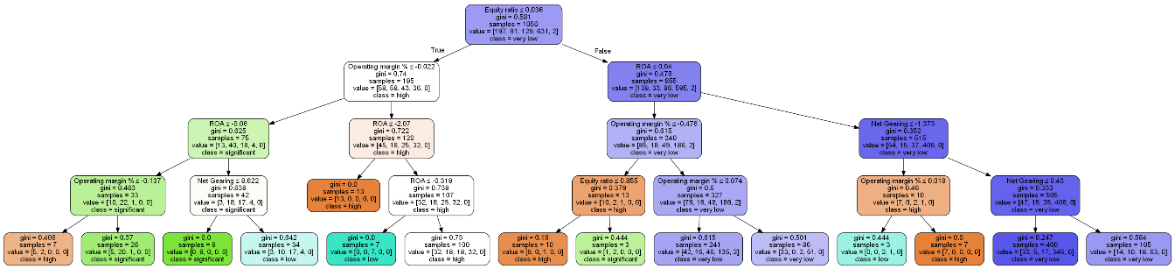
Supervised logistic regression classification algorithms of credit rating (industry codes 41-43)



Supervised logistic regression classification algorithms of risk rating (industry codes 41-43)



Supervised logistic regression classification algorithms of credit rating (industry codes 55-56)



Supervised logistic regression classification algorithms of risk rating (industry codes 55-56)