2019

Credit rationing of SMEs

MINIMIZING INFORMATION ASYMMETRY IN A P2P CONTEXT WITH HELP OF DATA ANALYSIS

Author: SAMANTHA GRONICH JAKOBSEN (92246)

Supervisor: THOMAS RIISE JOHANSEN

Master's Thesis - MSC in Business Economics and Auditing (cand.merc.aud)

Copenhagen Business School

85 pages, 145.685 characters

May 15th, 2019

Table of contents

1 **CHAPTER 1 - INTRODUCTION** **1**

1.1 Introduction *1*
1.2 Problem Statement *2*
1.3 Motivation for this Thesis *2*
1.4 Delimitation *3*
1.5 Terminology *3*
1.6 Thesis Structure *4*

2 **CHAPTER 2 – METHODOLOGY** **6**

2.1 Introduction *6*
2.2 Methodology *6*
2.3 Theory and Data Sources *6*
2.4 Source criticism *8*

3 **CHAPTER 3 – THE RATIONALE FOR P2P LENDING** **10**

3.1 Introduction *10*
3.2 SMEs *10*
3.3 P2P Platforms: The Crowdlending Marketplace *14*
3.4 The Risks Lenders Face in P2P Crowdlending *18*
3.5 Summary *24*

4 **CHAPTER 4 – ASYMMETRIC INFORMATION** **25**

4.1 Introduction *25*
4.2 The origins of the study of asymmetric information *25*
4.3 Understanding important aspects of microeconomic theory *26*
4.4 Information asymmetry issues between lenders and borrowers *26*
4.5 Forms of asymmetric information *27*
4.6 The consequences of information asymmetry *28*
4.7 Mitigating information asymmetry *30*
4.8 Summary *33*

5 **CHAPTER 5 – USING DATA ANALYSIS FOR CREDIT SCORING** **35**

5.1 Introduction *35*
5.2 The lender's decision-making *35*
5.3 Credit assessment *37*
5.4 Credit scoring *44*
5.5 How is credit being scored in Danish P2P platforms *46*
5.6 Improving the current model and minimizing information asymmetry *47*
5.7 Summary *56*

6 **CHAPTER 6 – THE CHALLENGES FOR IMPLEMENTING DATA ANALYSIS TO DANISH P2P PLATFORMS?** **58**

6.1 Introduction *58*
6.2 Fundamentals for implementing data analysis *58*
6.3 The opportunities regarding implementation *62*
6.4 The challenges regarding implementation *62*
6.5 Search costs *66*

6.6 Summary ... *67*

7 CHAPTER 7 – CONCLUSION ... 69

8 CHAPTER 8 – PERSPECTIVATION ... 74

8.1 Introduction ... *74*
8.2 Providing other forms of crowdfunding ... *74*
8.3 Cross-border platforms ... *75*
8.4 Summary ... *76*

9 BIBLIOGRAPHY ... 77

10 APPENDIX 1: BIG DATA SCORING ... ERROR! BOOKMARK NOT DEFINED.

11 APPENDIX 2: NOITSO ... ERROR! BOOKMARK NOT DEFINED.

Abstract

Informationsasymmetri er et skadeligt element i markeder med ufuldstændige informationer. I disse markeder, er asymmetrien uundgåeligt for både långivere og låntagere og er dermed ødelæggende for samfundet.

SMV'er er rygraden i nationens økonomi, og de bidrager til jobskabelse, innovation og økonomisk vækst. Imidlertid, står SMV'er overfor en kredit rationering fra banker, og informationsasymmetrien er synderen bag dette finansieringsgab. Situationen forværres under finanskriser, fordi bankerne kommer i en kreditklemme og bliver endnu mere risikoavers.

Hvis SMV'er skal fortsætte deres enorme bidrag til samfundet, skal alternative finansieringsmuligheder være tilgængelige. Finansieringsgabet betyder ikke kun, at der er et behov for alternative finansieringsmuligherder, men også at der er en åbenbar mulighed for alternative långivere.

Crowdlending, er en lånebaseret form for crowdfunding, der er blevet en ganske populær alternativ finansieringsmulighed rundt om i verden, men her i Danmark er crowdlending stadig i et barnestadie. Der er et potentielt og uudnyttet marked for lån, hvor flere SMV'er er ivrige efter at få deres projekter finansieret og långivere, der er ivrige efter at investere deres kapital på rentabelvis. Crowdlending platforme kan fylde finansieringsgabet og hjælpe både SMV'er og långivere til at nå deres mål.

Crowdlending platforme skal hjælpe långivere, der er særligt sårbare overfor informationssymmetri, med at vælge, i hvilke rentable projekter de skal investere deres penge i. Dette betyder, at platformene skal bidrage til at minimere virkningerne af informationssymmetri mellem låntagere og långivere.

At vælge, hvilke projekter der skal investeres i, er en kompliceret sag, da långivere skal frasortere de "dårlige" låntagere. Denne sortering blev udelukkende udført på baggrund af låntagerens historisk økonomisk data. Nye teknologier og Big Data tillader nu udviklingen af forudsigende modeller, som med høj nøjagtighed kan forudsige sandsynligheden for misligholdelse.

Denne kandidatafhandling tager sigte på at undersøge, om disse nye teknologier inden for dataanalyse kan hjælpe crowdlending platforme med at minimere informationssymmetrien mellem långivere og låntagere og dermed bidrage til at udfylde finansieringsgabet, som SMV'erne står over for i dag.

Chapter 1 - Introduction

1.1 Introduction

According to the Danish Association for state-authorized auditors (FSR)[1], in 2018 around 30% of the members of the Danish Federation of Small and Medium-sized enterprises (SMVDanmark)[2] complained over a lack of financing opportunities[3].

The main issue regarding financing rationing is that it can stall the growth of companies, and in worse case-scenarios, result in the sudden death of an otherwise healthy company, that unable to grow due to lack financing, was forced to shut its activities down. Even healthy companies might disappear due to lack of financing, if the company is growing faster than its structure can deal with it, so the demand is higher than the production. In this case, companies either grow or they stagnate and eventually die.

Lack of financing can also result in the loss of new ideas and innovations, that could improve society as a whole, if those were allowed to move out of the ideas-realm. This shows that new forms of financing are really needed.

When banks, and other credit institutions refuse to offer credit to companies, an opportunity for alternative lenders is created. Crowdfunding or crowdlending is an alternative way of obtaining financing that has become quite popular nowadays around the world. The concept behind crowdlending is rather simple and works similarly to obtaining a loan in the bank. However, the middleman (the bank) is no longer a part of the equation, and lenders and borrowers are allowed to deal directly, through a peer-to-peer (P2P) platform. Crowdlending is though risky, not only due to the inherent risk of financial markets, but because of information asymmetry.

While banks and other lending institution have an army of analyst to help with the credit assessment of borrowers, lenders in P2P platforms, face an exacerbation of this asymmetry, because they cannot distinguish between high-quality and low-quality borrowers. This might make lenders shy away from crowdlending. Therefore, crowdlending platforms will only attract lenders if it finds ways to minimize this information asymmetry and improve the lender's decision-making.

[1] www.fsr.dk
[2] www.smvdanmark.dk
[3] https://www.fsr.dk/Nyheder%20og%20presse/Nyheder/2018-nyheder/Flere%20virksomheder%20faar%20finansiering%20 med%20revisors%20hjaelp

1.2 Problem Statement

In order to try and minimize, at least theoretically, the credit rationing that SMEs are facing in Denmark, this thesis will answer the following problem statement and sub-questions:

What are the opportunities and barriers for using data analysis to minimize information asymmetry in a crowdlending context?

1. What is the rationale for P2P lending?
    a. What are SME's and why they are relevant?
    b. What are the causes for SME's difficulty in obtaining traditional loans?
    c. Is crowdlending really a good financing alternative for SME's?
2. What are the risks, lenders are exposed to in crowdlending?
3. What are the causes and consequences of information asymmetry?
4. In which ways can data analysis mitigate information asymmetry?
5. What are the opportunities and challenges for the implementation of data analysis to Danish P2P platforms?

1.3 Motivation for this Thesis

The main aim of this thesis is to investigate, if data analysis can improve credit assessment, mitigating the effects of information asymmetry in a P2P context. The purpose for this investigation is to reduce the financing gap, that SMEs are facing in Denmark. But SMEs will only be positively affected by crowdlending, if P2P platforms manage to become more mainstream and attract more lenders. In other countries, like in the USA, P2P lending is already a great alternative source of financing, not only to SMEs but also to larger companies and individuals. American platforms like Prosper and Lending Club have already issued loans for over 6 and 20 billion dollars respectively. In Denmark, P2P platforms are at their baby steps, and only a tiny percentage (1%) of the members of SMEs, according the Danish Federation of Small and Medium-sized enterprises have used crowdlending, as an alternative form of financing[4].

[4] https ://www.fsr.dk/Nyheder%20og%20presse/Nyheder/2018-nyheder/Flere%20virksomheder%20faar%20finansierin g %20med%20revisors%20hjaelp

There is an untapped demand for loans, and most likely also a supply for capital, as interest rates in P2P platforms are significantly higher than the interest rate paid by many investing alternatives. P2P platforms need, therefore, to become more attractive to both lenders and borrowers, which only will be possible if lenders start perceiving crowdlending as a real and profitable possibility. If P2P platforms want to attract lenders, they will need to provide some degree of safety regarding credit assessment, so risks regarding information asymmetry are mitigated, and lender's decision-making is improved.
Newer technological advances within computing capacity and artificial intelligence have improved the process of data analysis, and with the help of Big data, they can develop highly-accurate models, that can predict future behaviour based on past behaviour. It is, therefore, the purpose of this thesis to investigate, whether data analytics can improve credit assessment and reduce asymmetry of information.

1.4 Delimitation

This thesis focus on C2B (client to business) crowdlending (loan-based crowdfunding), and therefore it will disregard the other forms of crowdfunding, that are presently available. Although there are some crowdfunding platforms in Denmark, this thesis will only discuss aspects of two of the biggest platforms: Lendino and Better rates. Additionally, the regulation of crowdfunding in the European market is still in its first steps, as a proposal has been presented and is currently under discussion. This probable new regulation will be mentioned but not discussed further.
Regarding the risks, lenders face in crowdlending, operational risks will be mentioned, but not further addressed, as it is also out of scope. data analysis, the focus is only in artificial intelligence and data mining. Implementation is only approached theoretically.
Moreover, all rules and regulations mentioned in this thesis have the sole purpose of exposing the complexity of the matter, and will not be further analysed, as this would be out of this thesis scope.
Furthermore, this thesis is presenting some resumed aspects of economic theory in a descriptive manner and will avoid using further mathematical or graphical explanations, nor will use any mathematical and graphical explanations of how statistical models are construed.

1.5 Terminology

This thesis is written in English but refers to Danish problems. Many of the words, names and terms used in this thesis are, therefore, translated from Danish. To guarantee the understanding of the text, and avoid that nuances are lost in translation, the first time a Danish word, term or name that has been translated to English appears in the text, it will be accompanied by the original Danish version within parentheses.

1.6 Thesis Structure

This thesis is grouped into 5 main areas as shown in figure 1.1 below.

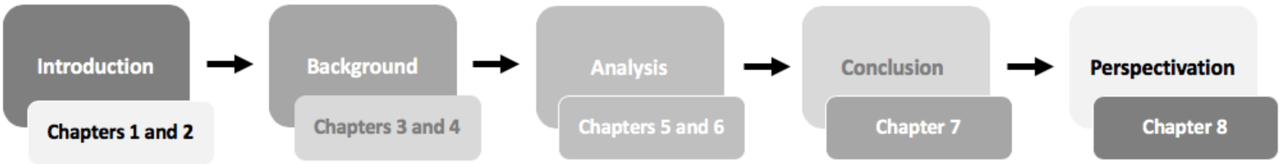

Figure 1.1 – Thesis structure distributed in areas and chapters

1.6.1 Introduction

This thesis is divided into 8 chapters. Chapter 1 and 2 are the introductory chapters. Chapter 1 presents this thesis motivation, problem statement, delimitation, terminology and structure. Chapter 2 deals with all methodological aspects, describing the chosen theories, data sources, and offering a criticism to the literature.

1.6.2 Background

Chapter 3 and chapter 4 introduce the background information for the thesis. Chapter 3 describes SMEs, P2P platforms and lenders, and exposes the relevance of SMEs, the need of alternative financing sources and how P2P platforms can fulfil this need and present the risks, lenders face when transacting via a P2P platform. Chapter 4 explore one of main sources for the risks, lenders face when transacting with P2Ps, namely information asymmetry. The chapter further explores how information asymmetry is the cause of the lack of financing, SMEs face, while discussing the different options to mitigate this asymmetry.

1.6.3 Analysis

Chapter 5 and 6 are the analytical chapters of this thesis. Chapter 5 introduces the concept of data analysis, describing how different methods of credit assessment work, and compare them to show how a combined model would be better in mitigating the effects of information asymmetry. Chapter 6 discusses the challenges regarding the implementation of data analysis in Danish P2P platforms and offer some suggestions to minimize those challenges.

1.6.4 Conclusion

Chapter 7 offers the conclusion of this thesis, where all the answers to the thesis' problem statement are added to answer the main question for the thesis.

1.6.5 Perspectivation

The final chapter 8 discusses other options, that P2P platforms have to increase supply and demand of lenders and borrowers, and therefore, increase the volume of available data.

Chapter 2 – Methodology

2.1 Introduction

The previous chapter introduced this thesis' problem statement. This chapter presents the methodology used to answer the problem statement, presenting the theories used by the author and the source of data. Furthermore, it presents a criticism to the chosen literature and describes, the analytical process used to develop each chapter.

2.2 Methodology

This thesis uses a qualitative method of research, based on literary theory and other available data to set the foundation and help in search of answering the problem statement. The work process follows the ground elements in knowledge production (figure 1.1). (Andersen, 2013)

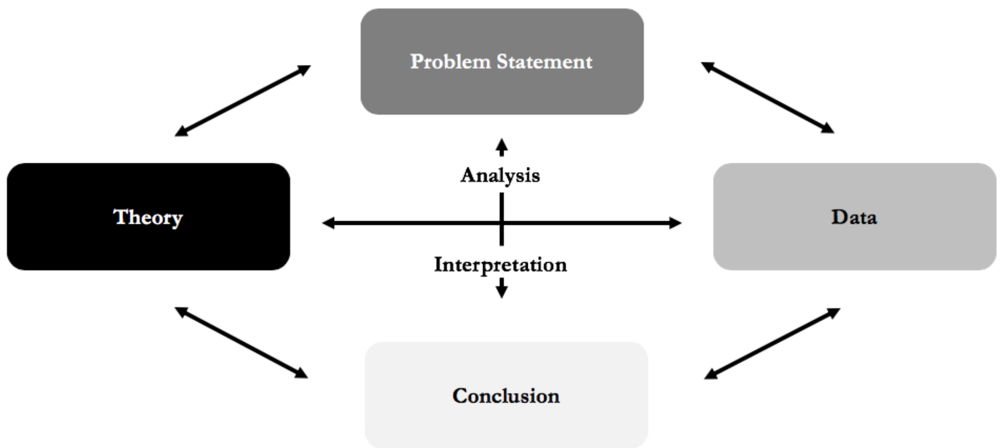

Figure 2.1 – Knowledge production ground elements and work process based on Ib Andersen‘s model

2.3 Theory and Data Sources

The theoretical background of this thesis is mostly comprised of academic articles, but academic books, practical guides, reports and other publications are also used.

The majority of the academic articles were found using sites such as Google Scholar, Research Gate and CBS's Libsearch. The search terms used were: SMEs, information asymmetry, signalling, screening, credit rationing, principal-agent problems, credit transactions, credit scoring, data analysis, data mining, artificial intelligence, credit assessment, risk assessment, risk identification, cost-monitoring, behaviour analysis, clickstream, social network data, mobile data, SMEs, and predictive patterns. The identification of relevance was done by evaluating content and only selecting those articles, that would fit within this thesis scope, which means that only articles fitting the parameters to answer the problem statement were chosen. Part of data regarding credit scoring was also acquired empirically through interviewing Erki Kert, CEO of Big Data Scoring, and Mohammed Azzouzi and Ronni Pedersen, CEO and Lead Senior Data Scientist respectively from Noitso. A resume of those interviews is added to the Appendix of this thesis.

Inspiration to write this thesis came from the site of FSR, in articles about SMEs and crowdlending.

Data sources and research for each chapter will be described below following this thesis structure (see figure 1.1)

2.3.1 Introduction

Research for chapters 1 and 2 was done based on Ib Andersen's book "Den skinbarlige virkelighed".

2.3.2 Background

Research for chapter 3 was done using articles published on both academic publications and regular homepages, having a significant part of those articles acquired on the pages of FSR, Confederation of Danish Industry (Dansk industri) and Danish Federation of Small and Medium-sized enterprises (SMVDanmark). Also, further information regarding SMEs was acquired through various OECD's publications such as Financing SMEs and Entrepreneurs 2017 and New approaches to SME and Entrepreneurship Financing: Broadening the Range of Instruments (2015), as well as on the site Statistics Denmark (Danmarks Statistik). All the different Danish legislations mentioned in this chapter were retrieved from the site Retsinfomation.dk.
Moreover, Information relating the production of financial statements in Denmark were retrieved from relevant laws, EY guide and books. Furthermore, information regarding rules and regulation for P2P platforms in Denmark was found on the Danish Financial Supervisory Authority (Finanstilsynet) homepage, while information regarding European suggestions for the regulation of those platforms were retrieved from the European Banking Authority (EBA).

Research for Chapter 4 was done using books on microeconomics and information asymmetry, as well as multiple academic articles from known experts in the topics.

2.3.3 Analysis

Research for chapter 5 was done using articles published on academic publications as well as information retrieved over the internet, by visiting the pages of different P2P platforms around the world and in Denmark. Furthermore, part of the information relating use of data analysis for credit scoring was acquired from webpages of companies that use develop credit scorings, like Big Data Scoring and Noitso, as well as from YouTube videos about those two companies. Additionally, the rest of information was acquired through interviews with Big Data Scoring's CEO Mr. Erki Kert and with the CEO and Lead Senior Data Scientist from Noitso, Mohammed Azzouzi and Ronni Pedersen, respectively. All the data regarding P2P platforms were acquired mostly on their homepages, although some information was also received through e-mail and phone contact with Lendino.

Chapter 6 uses most of the data presented in the previous chapters and just some extra information from sites and blogs discussing the implantation of data analysis are used.

2.3.4 Perspectivation

Research for chapter 8 uses data from blogs and articles online as well some academic articles.

2.4 Source criticism

The vast part of articles used in this thesis were, as already discussed, acquired from platforms such as Google Scholar, Research Gate and CBS's Libsearch. The articles come from sources such as Finance Journals, Accounting Journals, Economy Journals and Computer Science Journals, and some were also published by different universities. Part of the data was retrieved by books, the majority available at the CBS library. Data from OECD, governmental homepages, associations and sites containing legislation (retsinformation) were also used and are considered as data from reputable sources.

The majority of articles has a significant number of citations. Articles offering criticism of credit scoring using alternative data was also used. Many of the authors are known within their fields, some of them even having received prizes for their contribution.

Also, the data source for this thesis is very varied, so not only scientific articles were used, but articles from analysts, business people and specialists in data analysis were also used, giving a broader access of information, not only from the academic world, but also from the real world.

The source of all data is, therefore, considered reliable, as they come from respected writers, were published research from universities, and were provided by experts.

Chapter 3 – The Rationale for P2P Lending

3.1 Introduction

In this chapter, the rationale behind crowdlending is presented. As stated in chapter 1, this thesis is investigating, whether data analysis can help minimize information asymmetry in a P2P context, in order to attract lenders and improve the chances for SMEs in their search for financing. This chapter builds the background for the main discussion in this thesis, and presents SMEs, explain their relevance, and the challenges they face. Moreover, P2P platforms are presented and their function as an alternative financing source is discussed. Furthermore, the risks lenders face when transacting through P2P platforms are described, and The European Banking Authority's (EBA) suggestions to mitigate those risks are presented.

3.2 SMEs

Only companies that fit within the parameters established by the EU in their recommendation 2003/361 can register as an SME[5]. There are two main factors determining whether an enterprise is an SME or not: staff headcount and either turnover or balance sheet total[6].

Company class	Company type	Characteristics
D	Listed companies and State-owned limited liability companies	All despite their size
C	Larger companies	Balance sheet total >156 m. DKK Turnover >313 m. DKK Staff headcount > 250
	Middle-sized companies	Balance sheet total 44 - 156 m. DKK Turnover 89 - 313 m. DKK Staff headcount 50 - 250
B	Small companies	Balance sheet total 0 - 44 m. DKK Turnover 0 - 89 m. DKK Staff headcount 0 -50
	Micro-sized companies	Balance sheet total 0 – 2,7 m. DKK Turnover 0 – 5,4 m. DKK Staff headcount 0 - 10
A	Very small companies covered by the Law on certain commercial enterprises (LEV) and Personal companies (Personlige virksomheder)	All despite their size

Table 3.1 – The building block model (Byggeklodsmodelen) based on ÅRL § 7

5 https://ufm.dk/forskning-og-innovation/tilskud-til-forskning-og-innovation/typiske-sporgsmal/horizon-2020-typiske-sporgsmal/faq-smv-instrumentet/hvordan-finder-man-ud-af-om-ens-virksomhed-er-en-smv

6 Note that those ceilings can alone apply to the figures for individual firms

In Denmark SMEs are defined, in accordance with EU regulation, by the Financial Statement Law[7] (Årsregnskabslov[8]) §7, following a building block model (byggeklodsmodel) that is pictured in figure 3.1. A company should change class if 2 out of 3 criteria are overwritten for two consecutive years. Also, some middle-sized subsidiaries can choose to submit an annual report class B, if certain conditions are met.

The ÅRL divides commercial enterprises in two groups: those that have to submit a financial statement and those who are not compelled by law to doing that, but might choose nevertheless to submit a financial statement, if they choose to do so. This means that smaller companies are compelled to follow fewer general requirements, while bigger companies have to oblige to further and more detailed requirements. It is though always possible to choose following the requirements for a higher-class company.

3.2.1 SMEs Relevance

As already stated in the introduction to chapter 1, SMEs might be smaller in size, and therefore less visible, but they are the backbone of the economy. The vast majority of companies around the world is SMEs. In Europe, they represent 99% of all businesses[9]. The same is valid for Denmark, where 99,7% of all Danish business fit within the description for SMEs[10].

Firm size (employees)	Number	%	Jobs %	Value added %
SMEs	221.079	99,7	63,6	59
Micro	196.717	88,7	19,6	20.8
Small	20.510	9,2	22,8	19,5
Medium	3.852	1,7	21,2	18,7
Large	727	0,3	36,4	41
Total	221.806	100	100	100
Adaptation of table 1 produced by 2017 SBA Fact Sheet Denmark – European Commission. These are estimates for 2016 produced by DIW Econ, based on 2008-2014 figures from the Structural Business Statistics Database (Eurostat).				

Table 3.2 – Distribution of firms in Denmark by firm size

[7] The Financial Statement Law (Årsregnskabslov) applies to all Danish commercial enterprises, except for financial companies and certain public companies, cf. § 1.
[8] Årsregnskabslov (ÅRL) - https://www.retsinformation.dk/Forms/R0710.aspx?id=175792
[9] http://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition/
[10] https://read.oecd-ilibrary.org/industry-and-services/financing-smes-and-entrepreneurs-2017_fin_sme_ent-2017-en#page 118

The table above (table 3.2) shows the number of SMEs (including industry, construction, trade and services but excluding agriculture, forestry, fisheries and largely non-market service sectors such as education and health. SMEs are not only the vast majority of the companies in Denmark, but they also contribute altogether with 59% of the value added to the total Danish GDP.
SMEs not only comprise the majority of businesses but also represent 94% of high-growth companies, namely companies with an average annual increase in value added of at least 10% over the past three years[11]. High-growth companies contributed with more than 100 billion DKK, which corresponds almost entirely to all the increase in value growth created by all private companies in Denmark between 2013 and 2016. SMEs can be, therefore, considered the backbone of a country's economy, the engine behind a nation's economic growth.

SMEs have different reasons to search for financing, but mostly they require capital for four main reasons: operational financing, growth financing, emergency funding and conditional funding. They are predominantly financed by banks, mostly because SMEs lack access to public institutional debt and equity capital markets. (Mills & McCarthy, 2014; Mayer, 2016)

In Denmark SMEs can also acquire a loan from the Danish growth fund (Vækstfonden), a state financing fund, that offers Danish companies access to working capital. In 2017, the European Investing Fund (EIF) reached an agreement with Vækstfonden over a guarantee agreement of 1.6 billion DKK on loans to SMEs[12]. However, this is not enough to supply all the financing needs of SMEs[13].

3.2.2 The Economic Crisis of 2008

For the last decade, financing conditions for SMEs have worsened, due to the decline in bank financing to small business and entrepreneurs after the financial crisis of 2008. (Fenwick, et al., 2017).

The 2008 crisis affected the whole world and its effects were also severe in Denmark, that in 2009 experienced its most significant economic downturn in decades, resulting in a decline in the GDP by 5,1 per cent.[14] The slowed economic growth was also felt by financial institutions, in the form of tighter credit policy, due to higher risk aversion, affecting the Danish bank's availability to funding capital and herewith the availability for funding for SMEs. (Fenwick, et al., 2017)

---

11 https://www.danskindustri.dk/di-business/arkiv/nyheder/2019/2/smver-er-danmarks-vakstlokomotiver
12 https://www.danskindustri.dk/di-business/arkiv/nyheder/2017/12/rekordaftale-eu-stiller-16-mia.-kr.-garanti-pa-lan-til-danske-smver/
13 http://www.smvportalen.dk/Finansiering-tilskud-stoette/Finansiering-tilskud-stoette/Nyt-om-stoette-om-finansiering/2019-Nyt-om-stoette-og-finansiering/Kreditklemmen%20lever%20i%20bedste%20velgaaende
14 https://www.dst.dk/Site/Dst/Udgivelser/GetPubFile.aspx?id=24799&sid=smv2014

As a result of the international financial crisis, lending to SME's - for loans which amount to less than 1 million EUR - declined by around 30% between 2007 and 2009[15]. SMEs lending increased by 23% in 2010 and stagnated in 2011, dropping once more between 2012 and reaching its lowest levels in 2013. However, new lending to SMEs increased by 39,7% in 2015 on year by year basis, which points out that SME bank lending improved a bit, but in smaller share than among the larger enterprises[16], and even though it is an improvement, SME lending is still lower than before the financial crisis (Mills & McCarthy, 2014).

3.2.3 SME's Challenges

SMEs are smaller, less diversified and have a weaker financial structure than larger enterprises. (Fenwick, et al., 2017) It has been already established by academic literature, that SMEs are more affected by financial crisis, due to their dependency on bank investment. Therefore SMEs are generally more sensitive to changes in the economy, have higher failure rates and fewer assets, or assets of lower quality, that can be used as collateral for a loan (Mills & McCarthy, 2014). That results in an economic barrier, affecting SMEs creditworthiness and hence making SME lending riskier.
SME's creditworthiness is not improved by the fact that those companies face a greater risk in operation than larger companies. Estimates show that around 24% of SMEs disappear within two years of its creation and that nearly 53% of SMEs will leave the marked within four years of its inception, either due to failure or bankruptcy (Duan, et al., 2009).

3.2.3.1. Credit rationing

The main issue in hand is the asymmetric information between SMEs and banks, which makes assessing their creditworthiness much more complicated. The banking business model is characteristic for the financial institution assuming all the credit risk, and therefore banks have risk management departments. (Serrano-Cinca, et al., 2015) Also, there is not so much public information about the majority of SMEs, as it is rather uncommon that these enterprises will issue publicly trade equity or debt securities (Mills & McCarthy, 2014). Additionally, the ÅRL allows some enterprises to choose whether they want to submit

15 https://read.oecd-ilibrary.org/industry-and-services/financing-smes-and-entrepreneurs-2017_fin_sme_ent-2017-en#page118
16 https://www.danskindustri.dk/arkiv/analyser/2018/1/smverne-er-tilbage-pa-sporet-10-ar-efter-finanskrisen/

a financial statement (ÅRL §4-6), so, some Danish SMEs[17], may not have an income and balance statement available to offer. This results in a credit rationing to SMEs.
Also, SME's have a tendency of merging with the figure of the owner, acquiring his idiosyncrasies regarding his most likely informal relationship with eventual stakeholders. That exacerbates the information asymmetry between borrower and lender, and the heterogeneity in the SME sector does not help minimize this asymmetry, on the contrary[18]. (Caire & Kossmann, 2003)
In financially depressed times, such as during an economic crisis, banks become more risk-averse, due to the limited availability in funds they have. That leads banks to apply a strategy to contain adverse selection risks, in which they require more collateral to back up their investment. But as already exposed, SMEs do not have as strong assets as larger firms do, and therefore, the sparsely available funds will most likely be allocated to firms, that can offer the best collateral as a guarantee to the loans.
Moreover, SME lending also faces structural barriers in the form of high transaction costs. Although transaction costs of a big loan are comparable to transactions cost of small loans, the higher the value of the credit, the higher the profit margin for banks, and like in an economy of scale, higher loans will result in smaller unit transaction costs (Duan, et al., 2009).

In 2019 the FSR has done a survey relating to alternative financial options to SMEs. According to this survey, SMEs usually will finance their needs with the help of friends, family and the Vækstfond. As claimed by the survey, 26% of auditors, who have SMEs as clients, observe that SMEs are very much searching for alternative financing options to banks and other credit institutions, and the main reason for this search is due to the fact that those SMEs have had their loan requests rejected by banks. Another main reason leading SMEs to search for alternative financing possibilities is the difficulty they have in providing a security to back up the bank loan. The table below shows the main issues, auditors see as reasons driving SMEs to search for alternative financing.

3.3 P2P Platforms: The Crowdlending Marketplace

Although P2P lending was conceived to be free from mediation, crowdlending has become increasingly mediated by online intermediaries (P2P lending platforms). In those platforms, borrowers can place

[17] The majority of SMEs that do not have an obligation to submit a financial statement are companies with personal responsibility such as sole proprietorships, partnerships and limited partnerships, as long as not every participant is a company with limited liability (kapitalselskaber) -**Invalid source specified.**.
[18] http://www.oecd.org/cfe/smes/New-Approaches-SME-full-report.pdf

requests for loans and just as in an auction-like process, lenders can bid to fund those projects. (Fraunhofer, 2009).

Despite the fact that P2P platforms maintained a mediator, this marketplace offers many advantages. The expensive middleman (banks) is replaced by a more cost-effective platform, thus reducing transaction costs. That is the case because those platforms make their profits from commissions instead of the spread between deposit and loan (OECD, 2015).

P2P platform's activities consist in collect, score and distribute the credit assessment of prospective borrowers, report real-time bids on projects and supply online service and monitoring of loans (Kwok, et al., 2010). They collect loan pledges from the lenders for private projects and release them at the moment the target is reached. Platforms also collect repayment instalments from the borrower, forwarding them to the lenders (OECD, 2015). They usually develop a credit rating system for loan approvals and pricing and perform credit checks for borrowers.

3.3.1 Crowdlending as a Financing Alternative for SME's

The aftermath of the financial crisis exposed how vulnerable SMEs are to changing conditions in bank financing. The increased risk-aversion developed by banks, in association with a new regulatory environment made by governing demands, resulted in more rigorous procedures in lending capital, which might enhance the rationing of credit for SMEs.

The rationing of SME's lending has created a financing gap, that can have catastrophic consequences for the economic health of a nation. If SMEs cannot finance their operations and expansions, many will disappear, removing job opportunities from the market, as well as innovations and possibilities for improvement of life in general. SMEs are, as already mentioned, the major source of employment in the private sector, and this financing gap can, in the worst case, could lead to massive unemployment and henceforth a national depressive economic state.

Therefore, it is essential to broaden the access to different financing possibilities, so SMEs can continue to exert its critical role in competitiveness, growth, innovation and job creation. Banks will still be an important source of financing for SMEs, but by diversifying their financial options, SMEs will become less vulnerable to economic changes, and hence, the chances of long-term investments will be improved. Crowdlending or P2P lending is one of the possible options in diversifying financing possibilities for SMEs. P2P lending works just like bank financing, without the mediator, namely the bank. It is, therefore, a form of loan transaction between individuals, without any form for mediation, in which lenders and borrowers have direct contact and exchange information, using the internet as their platform. The availability of information and easy access to it makes the whole process highly transparent. Loans operate on the idea of "full financing", which means that a project will only be funded, if it receives enough bids for lenders to cover the entire requested amount, within the pledging time.

Lenders and borrowers establish a debt contract between themselves, in which the borrower promises to repay the principal as well as the interest rate accorded between the parts, depending on the risk level, within a certain period (Kwok, et al., 2010).

P2P loans are usually unsecured, which means that no collateral is required of borrowers. Additionally, lenders are most likely inexperient investors, without specific knowledge needed to analyse historical financial data. That makes P2P lending an attractive option for SMEs that lack both collateral and credit history. However, P2P is not only attractive to highly-risk borrowers that are refused by banks. Those platforms have been able to attract high-quality credit risk to companies and individuals, providing loans to refinance credit-card debt and other debts (OECD, 2015) .

3.3.1.1. The opportunities of crowdlending

P2P lending, if employed correctly and efficiently, offers the possibility of a win-win experience to both lenders and borrowers. It gives lenders the opportunity to increase their capital with a rate of return on investment higher than that provided by banks (Fraunhofer, 2009). It is important to emphasise here, that the opportunity of capital gains might attract both professional as well as inexperient lenders. While professional lenders are capable of evaluate their risks, inexperient lenders might not be aware of the risks involved in the transaction, which could turn this "win" into a financial loss. On the other hand, borrowers have the opportunity of receiving lower rates than those required by banks, because transaction costs regarding overhead and regulatory burdens are lower. Besides that, borrowers have the opportunity of presenting their projects in more detail, which is not the case in bank financing, that have standardized decision processes and ignore information that does not fit into their selected parameters (Fraunhofer, 2009). Additionally, there is a sense of fairness and transparency, as all bids are visible and traceable online. As the process popularizes, more and more SMEs will be able to benefit from this interaction while more and more lenders will acquire financial experience, that might be beneficial in future investments.

According to a survey made by FSR in 2019, SMEs choose crowdfunding as a financing option mostly due to not being able to acquire loans in the traditional financing sector. Although crowdlending is perceived as an alternative financing possibility, there is still a lack of regulation in most member states, which contrasts with the stringent rules applied to banks[19] (Ahern, 2018). The EU has, however, plans to regulate crowdlending, and a proposal for a regulation on crowdfunding has been published by the European Commission in March 2018. The proposal is currently being analysed in-depth[20].

3.3.2 Regulatory Framework in Denmark

In order to minimize risks, a company wishing to establish a P2P platform in Denmark is required by law to have an authorization from the Danish Financial Supervisory Authority (Finanstilsynet)[21] to operate either as a bank (pengeinstitut) or as a provider of payment services (betalingstjenesteudbyder).

3.3.2.1. License to Provide Payment Services

A platform will provide payment services if it handles and transfers the investment and the repayment, with interest, between lender and borrower, and the lender is free to choose, which projects he wishes to invest on. This type of license allows platforms to carry out all activities regarding payment transactions between two parties, but it is the concrete set-up of the payment system, that will be decisive in deciding which license is required by the Act on Payments[22]. This licensing type can be full or limited, depending on the platform's transaction volume and cross-border activities.

According to Act on Payments § 51, a P2P platform will only require a limited authorization as payment service provider if the total payment transactions (the value of the loans and interest and repayments that are transferred through the platform) does not exceed an amount corresponding to the value of € 3,000,000 per month (calculated for the previous 12 months). No further capital requirements are required, but the company will have to follow the requirements stipulated by § 52 of the same law. Those kinds of P2P platforms will only be allowed to provide their services inside of Denmark.

However, if the average total payment transactions exceed the limits, or in the case the platform wishes to expand their activities for areas outside of Denmark, the platform will require an authorization as a

[19] Ahern, D., 2018 - The EU's Opt-in Regulatory Framework for Crowdlending: Expediency at the Expense of FinTech Investor Protection? - https://www.law.ox.ac.uk/business-law-blog/blog/2018/09/eus-opt-regulatory-framework-crowdlending-expediency-expense-fintec
[20] https://eurocrowd.org/2018/03/13/ec-proposal-regulation-european-crowdfunding-services-providers/
[21] https://www.finanstilsynet.dk/Tilsyn/Information-om-udvalgte-tilsynsomraader/Fintech /Crowdfunding
[22] Lov om betalinger - https://www.retsinformation.dk/Forms/R0710.aspx?id=191823

payment institution, according to § 9. In this case the company be driven as a limited company, limited liability company or limited liability company, and must have sufficient starting capital, according to §10.

3.3.2.2. License to Operate as a Financial Institution

If the platform receives money from the investors and then provide loans for own account to projects selected by the platform, it will require an authorization as a financial institution (pengeinstitut). The conditions for obtaining this type of authorization are stated in §§ 7 and 14 of the Financial Business Act[23], i.e. the company must be driven as a public limited company and must have a share capital of at least €5 million.

There is other relevant legislation a P2P platform provider must take in consider, such as regulation regarding money laundering, taxation and marketing.

3.4 The Risks Lenders Face in P2P Crowdlending

The majority of risks, lenders are exposed to, derive from asymmetric information. Some risks are though caused by the uniqueness of P2P lending marketplace (Kwok, et al., 2010). The main issue is that lenders are typically inexperient individuals, that do not have enough financial knowledge to understand historical financial data. Furthermore, many SMEs do not necessarily have enough, if any, historical financial data to present to future lenders.

In the paragraphs above it was presented, that even banks, financial institutions that hire individuals with financial expertise to analyse historical financial data, suffer from asymmetric information, and therefore, choose to allocate scarce capital in more assured investments. That shows that asymmetric information is a core issue in financial markets and only seems to corroborate, that private lenders face a severe challenge when deciding to whom they will lend their money, as well as defining the risk involved with such transaction, especially when taking in consideration, that P2P lending usually is unsecured.

The European Banking Authority (EBA), an organ responsible, among other things, to monitor new and existing financial activities in Europe and adopt guidelines and recommendations to safeguard that those markets are sound and follow regulation, in its report on crowdlending "Report on lending-based

[23] Lov om finansiel virksomhed - https://www.retsinformation.dk/Forms/R0710.aspx?id=193767

crowdfunding: risks, drivers and potential regulatory approaches”, identifies six risks lenders face when loaning money through P2P (see figure 3.1): credit risk, the risk of fraud, lack of transparency or misleading information, legal risks, liquidity risks and operational risks.

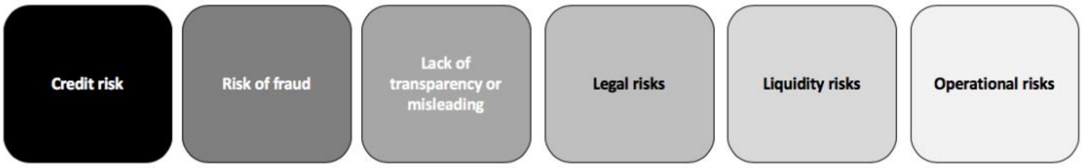

Figure 3.1 – Risks that lenders face in P2P transactions

3.4.1 Credit Risk

Credit risk can be divided into four categories, which will be described in the next sections.

3.4.1.1. Risk of Default

An inherent part of most investments is risk. It is the risk that creates the possibility for lenders to require a return on investment, and the higher the risk, the higher the rate on that return. The most obvious negative consequence derived from P2P lending is the risk, lenders face, of losing their investment. A default can occur due to both unexpected and expected causes. Unexpected causes are causes that cannot be predicted or prevented at the contractual moment between lenders and borrowers. Sudden economic changes, financial crisis and other unforeseen events may lead to unanticipated default. Expected causes are, however, possible to be foreseen by employing a comprehensive and high-quality credit risk evaluation.

3.4.1.2. Risks Regarding Credit Risk Evaluation

Evaluating credit risk is essential to making loan decisions, but a good credit risk evaluation requires enough credit information of good quality, and the ability to understand this information. Evaluating credit risk is, therefore, one of the main difficulties, lenders face in P2P lending, due to two issues: the inherent uncertainty regarding the borrower’s ability to repay and the even more complex issue of establishing the borrower’s willingness to repay the debt. It is thus a two-stage evaluation.

It is initially critical to establish whether the borrower can repay the debt, which means that the borrower has or is expected to have enough capital to repay the principal and the interest in a timely manner. Subsequently is necessary to determine, whether the borrower is willing to repay the debt, as not all defaults are caused by the inability to repay, but merely because borrowers choose to allocate their capital differently.

In theory, a lender could be more protected against this risk by lending through a P2P platform that employs credit rating and credit risk assessment for borrowers. However, even a P2P platform suffers from asymmetric information and may supply lenders with a faulty credit evaluation, that could be risky and misleading. P2P platforms usually rely on the information provided by the borrower, which creates an opportunity for fraud (Galloway, 2009/2010). Besides these platforms could also entice lenders to invest in certain unsafe investments with the promise of unrealistically high rate return on investment.

3.4.1.3. Risks Regarding P2P Platforms

Furthermore, lenders face risks due to mediation. A P2P platform might not repass the borrower's payment, and in worse circumstances default, due to either fraud or simple lack of adequate controls.

3.4.1.4. Behavioural Risks

Studies are pointing out that lenders exhibit herding behaviour in online business when facing the risk of uncertainty due to asymmetric information. This entails that lenders may have their decision-making influenced by the decision of others. There are two main explanations to this behaviour: excessive information available over the internet (information overload), and the fact that it is easy to observe, how others are choosing online (Pokorná & Sponer, 2016). The issue in hand is that mere fact that many people are bidding on a project does not qualifies the project as a safe option.

3.4.2 Fraud Risks

It is the lenders, and not the financial institutions (banks), who have the role of credit risk assessment in P2P lending. This creates an opportunity for misrepresentation by borrowers regarding their creditworthiness (Pokorná & Sponer, 2016). Such risk can be mitigated or increased by P2P platforms, depending on the quality of their credit risk evaluation of the borrower.

Fraud risks also encompass fraudulent P2P platforms, and therefore it is crucial that those platforms are under strict regulation. Lastly, lenders also face the risk of having their information stolen or misused, if P2P platforms to not implement controls regarding data security.

3.4.3 Risks Regarding Lack of Transparency or Misleading Information

Lenders might not be able to identify a conflict of interest between the P2P platform, or one of its employees, and a borrower. Moreover, other issues relating unclear or lacking terms and conditions as well as loan transaction contractual rules can cause uncertainty to lenders regarding their rights and obligations.

Another risk lenders face due to their inexperience is to trust that all loans and borrowers are good options, due to the assumption that the P2P has done an efficient and complete credit assessment of them and their projects, and only the best or safest options are available. Lenders are usually not capable of judging, whether the methods used by a P2P platform are valid or correctly implemented.

3.4.4 Legal Risks

Lenders might be uncertain of the contents of the debt contract between them and borrowers, if the P2P platform fails to disclose, in an understandable manner, the contractual rights and obligations applying to lenders and borrowers. Another issue that can lead to uncertainty regards the scope of the mediation service provided by the platform. All the information relating a service, or its terms and regulations, should, thus, be presented in an easy, clear and plain form, in order to guarantee full comprehension. Legal risks also cover the situation, in which the capital provided by the lender is not repassed in its entirely to the borrower, either due to fraud or error.

3.4.5 Liquidity Risks

Debt contracts usually have a contractual period of validity, which means that the lender can only expect the complete fulfilment of the borrower's obligation after this period expires. That per se can result in liquidity issues, as the lender cannot redeem his investment before it reaches its full term. Lenders can also face liquidity problems in case of delinquency (untimely payment).

3.4.6 Operational Risks

Technical issues with the P2P platform can lead lenders to experience economic loss. Those issues might be caused by unforeseeable and uncontrollable events, such as acts of nature, but might also be a result of lack of controls, such as a data backup procedure, disaster recovery plan and a business continuity plan, aiming to recover and resume normal operations.

3.4.7 EBA's Suggestion Regarding Regulatory Measures to Address the Risks and Risk Drivers

EBA has also listed some potential regulatory measures, that P2P platforms should implement to address the risks and their drivers[24]. P2P should investigate the risk factors – reasons behind risks – and establish procedures to minimize those risks and thus increase trust in loan transactions via P2P platform.

There are some measures P2P platforms can take regarding lenders asymmetric information and financial illiteracy, such as to conduct a risk assessment and analysis of crowdfunding initiatives and present the information in a clear, understandable and not misleading manner. Lenders should be informed on projects, borrowers, risks – including the risk of total or partial loss of invested capital, as well as risks regarding not obtaining the expected return or eventual liquidity issues – and financing mechanisms.

Platforms should also categorize lenders according to their expertise level, and only allow investments within the category, the lender is categorized in, or establish investment limits, so a lender would have a maximum amount per project, within a certain period, according to his or her income or wealth.

As already exposed, lenders may underestimate risk, assuming that all projects on a platform are safe. Therefore, P2P platforms should inform the lenders of all the different forms for assessment performed. All information regarding a project or borrower should be made available to potential lenders. Platforms should also be required by law to conduct an effective, proper and diligent procedure on any investment opportunity, as well as provide lenders will full transparency regarding their assessment process. Platforms should also reject projects from borrowers with insufficient creditworthiness. Some risks could be mitigated if P2P platforms cooperated with banks, using the bank's assessment processes as a basis for their credit assessment.

[24] https://eba.europa.eu/documents/10180/983359/EBA-Op-2015-03+(EBA+Opinion+on+lending+based+Crowdfunding).pdf

Lender protection services could be established, where the platform could retain a certain amount of the charges to ensure repayments in the event of default.

Other measures can be taken regarding issues that are unique to P2P platforms, such as preventing platform failure, where a P2P platform should be required to have arrangements in place to ensure continued service for current clients in case the platform bankrupts or somehow otherwise goes out of business. Those arrangements must also contain a compensation scheme, insurance coverage for default or other similar provision.

It should be required an authorization by a national financial supervisory authority or similar, that would ensure that the crowdfunding platform is managed following appropriate standards for competence, capability, integrity and financial soundness. That authorization should be expressed disclosed on the P2P's website.

Regarding data protection, P2P platforms should have a clear terms and conditions page stating their document-handling policies. Also, it is fundamental that platforms address issues regarding conflicts of interests, by prohibiting shareholders, managers and key employees from having or acquiring financial interests in a borrower's business. Measures should be implemented to identify and manage potential risks of interest.

Platform's terms and conditions should also clearly present the rights and obligations of the parts, informing them about the financing process, costs and other features applicable to contracting parties, as well as offering an appropriate complaints handling mechanism. Platforms should always ensure a money segregation between their money and the client's investments or repayments. Platforms can also help mitigate liquidity issues by ensuring that the transfer of agreed funds happens timely.

IT risks must be mitigated by stablishing effective IT controls, such as a data backup procedure, disaster recovery plan and a business continuity plan, aiming to recover and resume normal operations.

Lastly, it is important to address concerns regarding anti-money laundering, by establishing controls to address issues relating to borrower's anonymity.

3.5 Summary

In conclusion, SMEs are fundamental to the economic health of a nation, as they are the major source of jobs within the private sector. They are also responsible for a large part of the private sector's growth and value added, contributing immensely to the nations GDP. However, they face a rationing of credit, and this has only been exacerbated by the international financial crisis of 2008, which made if further difficulty for SMEs to obtain the essential loans to maintain their operations, expand and develop innovation. Even though the financing situation improved since the crisis started, it is still far from optimal. This financing gap shows how vulnerable SMEs are to economic changes, and therefore highlights the need to increase the range of financing options available to those companies.

Among alternative financing options is crowdlending. P2P platforms, that connect lenders and borrowers is a viable and interesting new financing possibility, that can offer a win-win approach to all parts involved in the transaction. Lenders will be able to make a higher rate of return on investments than the one offered by traditional bank savings, while borrowers will receive the financing they require, in order to maintain competitiveness, growth and innovation.

Crowdlending is however not without risks. While it can be a useful tool in counteracting credit rationing for SMEs, it creates its share of complications, and lenders face problems deriving from the asymmetric information that is inherent to credit markets as well as operational issues that are unique to P2P platforms.

Chapter 4 – Asymmetric information

4.1 Introduction

This chapter continues to build the background for the main discussion in this thesis, and presents information asymmetry, its causes and consequences. The previous chapter introduced, that SMEs suffer from credit rationing. This chapter explains, how credit rationing is a protective mechanism against information asymmetry. Additionally, this chapter presents some factors, that can mitigate this asymmetry, and that will be the grounding block for the next chapters analysis on how data analysis can be used to mitigate asymmetry of information between borrowers and lenders.

As already stated, information asymmetry is an inherent risk in financial markets, and the explanation is quite simple: in financial contracts, information asymmetry will occur, due to the fact that the lender does not have all the necessary information, nor control over, whether the borrower can and is willing to repay his debt. Also, the borrower, who is using the lender's capital, have an incentive to disguise the true nature of his project, or in some cases to use the invested capital in a different project, and not in the project he first proposed. The borrower has an incentive to announce lower-than-actual earnings as well, in order to reduce his financial obligations at the lender's expense (Bebczuk, 2003; Perloff, 2015).

4.2 The origins of the study of asymmetric information

The concept of asymmetric information is part of the study of microeconomics and its main proponents were awarded the Nobel Memorial Prize of Economics for their analyses of asymmetric information in markets[25].

The theory was first presented by George Akerlof in 1970, in an article called "Market for lemons", in which he analyses the effects of asymmetrical information on the automobile market, in which sellers have an incentive to sells goods of less than average market quality, also known as lemons, because they are aware of a disparity in the levels of information between themselves and buyers (Akerlof, 1970). Akerlof suggested that to reduce the asymmetry of information counteracting institutions (intermediary market institutions) could be used, to guarantee that the good is actually been priced accordingly to its qualities.

Akerlof's theory was complemented in 1973 by Michael Spence with his theory on signalling, in which

[25] https://www.nobelprize.org/prizes/economic-sciences/2001/press-release/

the author claimed that in imperfect information markets, individuals affected by adverse selection could use signalling in order to increase their trustworthiness. (Spence, 1973)
Finally, in 1975, Joseph Stiglitz contributed to the theory by explaining the rationale behind screening, describing the use of screening devices to identify the different qualities of goods, services and even borrowers. (Stiglitz, 1975)

4.3 Understanding important aspects of microeconomic theory

This subsection will shortly present some important aspects of microeconomics, in an attempt to further clarify the comprehension of the topic and pave a better background view of information asymmetry.
In a market in equilibrium, supply equals demand, which means that the production of a good or service will be equal to the demand that good or service has. The main component that hold this equilibrium is price, and if demand exceeds supply, that will result in an increase in price, which will push demand down by reducing it, so at this new price level, demand will equal supply. The same happens when supply exceeds demand, in which case the price will decrease, and henceforth demand will increase. (Perloff, 2015)
So, in theory, if price is fulfilling its function, there will be no rationing. However, rationing exists in reality, when markets are in disequilibrium, i.e. when the demand for loans is high. The disequilibrium might be a result of an exogenous shock (such as a financial crisis), in which case it will be a temporary disequilibrium (short-term). On the other hand, governmental constraints such as regulations could lead to a permanent disequilibrium (long-term), i.e. regulations trying to minimize the effects of crisis, by increasing regulation over bank loans. (Stiglitz & Weiss, 1981)

4.4 Information asymmetry issues between lenders and borrowers

The relationship between lenders and borrowers can be described as a principal-agent relationship, in which the lender is the principal and the borrower is the agent. Principal-agent relationships are considered problematic relationships because the two parties can have different interests (conflict of interest) and different levels of information. The agent has always more information than the principal and cannot be directly controlled by him. The principal, however, has the capital the agent needs. Therefore, there is a risk that the agent will not act in the principal's best interest, particularly if doing so is costly. (Akerlof, 1970. Janda, 2006)

In credit markets, lenders and borrowers contract with each other, through a debt contract[26] establishing the rules and obligations for both parties and defining whom will receive financing and for what, and whom will provide the capital. Although borrowers promise to repay the principal plus the required interest rate during a certain period of time, the debt contract is compromised by uncertainty.

Any investment project is surrounded by an intrinsic amount of uncertainty, as multiple variables are involved in a current economic decision based on events yet to come. In other words, investing in any project is risky, because it is impossible to predict with 100% accuracy, whether this project will succeed or not. Lenders are aware of this uncertainty, and being rational[27] and risk averse players, they match the risk to a required rate of return (RRR)[28]. The higher the risk of the project, the higher the interest rate. Lenders will also take in consideration opportunity costs, and between two loans with similar RRR, they will choose the safest one.

Borrowers, on the other hand, are risk neutral, especially when their liability is limited, because they only have invested in the transaction with a promise. (Ghatak & Guinnane, 1999) This means that if the project fails, they have little to lose, besides their time, work and of course, the opportunity cost connected to choosing one activity over another. However, borrowers are also rational players interested in financing terms that allow space for profit. This space for profit can be seriously affected by the lender's financing demands, as the lender's RRR should be lower than the project's RRR. The higher the lender's RRR is, the lower is the borrower's profit.

Lenders have, therefore, way more at risk, not being able to distinguish high-quality borrowers from low quality borrowers only enhances the inherent uncertainty of credit markets. One issue is not knowing if the project will be successful, another completely different is dealing with an opportunistic borrower, that either offers a lemon, knows that he does not have the economic means to repay a loan, or has absolutely no intention of repaying the loan, even if economic means to do so are present. (Bebczuk, 2003; Akerlof, 1973, Perloff, 2015)

4.5 Forms of asymmetric information

Information asymmetry in credit markets can be divided in three types: adverse selection, moral hazard and monitoring costs.

[26] https://bizfluent.com/facts-6800917-definition-debt-contract.html

[27] The term rational player is used to describe individuals that are driven by an individual rationality constraint. This means that no rational person will accept to be part of a transaction with either a negative expected return, or with a profit that does not even come close to the expected level of return.

[28] The required rate of return (RRR) is the minimum amount of profit that an investor will receive for his investment, it is in other words the minimum return for assuming the risks involved in an investment or project.

4.5.1 Adverse selection

This form of asymmetric information appears before the disbursement of the loan, because the lender is not capable of distinguishing between a high-quality borrower and a low-quality borrower when allocating his credit.

Since lenders would rather pick safer projects, low-quality borrowers have an incentive to camouflage themselves as high-quality borrowers, especially when taking in consideration that lenders cannot differentiate. So all borrowers appear as if they were high-quality borrowers, and lenders, that are aware of their inability to ascertain the quality of the borrower, choose to self-protect by establishing a single interest rate for all borrowers, treating them all as low-quality, in the expectation of securing the desired RRR.

4.5.2 Moral hazard

This form of asymmetric information appears after the loan disbursement has happened. It happens when the borrower chooses to apply the loaned investment differently that what was agreed upon with the lender, without the lender's consent or knowledge.

4.5.3 Monitoring costs

This form of asymmetric information also appears after the loan is disbursed. The borrower takes advantage of the asymmetric information to underrepresent his profits. The lender, who is hindered of any control over the borrower's behaviour, will be forced to monitor the borrower, whenever he declares himself unable to fulfil his obligation. Monitoring would require a special clause in the debt contract, in which the lender would have the right to audit the borrower and seize any verified cash flow, whenever the borrower announces default.

4.6 The consequences of information asymmetry

One of the characteristics of imperfect markets is the power struggle between principal and agent. No matter what advantage the borrower might have due to asymmetric information, the lender has capital power, that he uses in an attempt to deter opportunistic behaviour. Lenders can always select an alternative use for their money while borrowers, who might struggle to find alternative financing sources, are either forced to abandon their projects or to accept the contractual conditions imposed by lenders.

Both adverse selection and monitoring costs can have serious consequences for both lenders and borrowers.

4.6.1 The consequences of adverse selection

As exposed in subsection 6.5.1, lenders use a self-protective mechanism that is discriminatory against high-quality borrowers, because their lower-risk projects would require a lower interest rate. In some cases, the lender's RRR is higher than the project's RRR, which would eliminate the entrepreneur's profits. High-quality borrowers, that don't take advantage of their information power will desist of the project, leaving only low-quality borrowers on the market. (Stiglitz & Weiss, 1981, Bebczuk, 2003) This is another paradox, showing how defence mechanisms against information asymmetry increase risks for lenders. Low-quality borrowers, on the other hand, when affected by adverse selection, choose riskier projects with higher payoffs in case of success. (Stiglitz & Weiss, 1981) If the project is successful, the lender's expected return will be lower than the required return in the project, and therefore, the effects of adverse selection are worse to high-quality borrowers than to low-quality borrowers. (Bebczuk, 2003)
It is believed that opportunistic behaviour occurs as a result of the borrower's information power, so borrowers lie just because they can. That is however, just one of the aspects of opportunistic behaviour. Borrowers might choose to lie as a result of adverse selection, because they need the financing and the terms of the debt contract are too rigorous. If the lender's RRR is too high, borrowers might lie to acquire cheaper credit. So, another consequence of adverse selection is moral hazard.

4.6.1.1. Credit rationing

Credit rationing is a good example of how adverse selection can affect both lenders and borrowers. Credit rationing can be described as "circumstances in which either (a) among loan applicants who appear to be identical some receive a loan and others do not, and the rejected applicants would not receive a loan even if they offered to pay a higher interest rate[29]; or (b) there are identifiable groups of individuals in the population who, with a given supply of credit, are unable to obtain loans at any interest rate, even though with a larger supply of credit, they would". (Stiglitz & Weiss, 1981)
The reality is, that credit rationing might be a characteristic of the equilibrium in loan markets, due to the presence of asymmetric information, and the reason is quite simple: banks are constrained by rationality, and thus, interested in profits and concerned about credit risk and interest rates. This rationality

[29] Interest rate and RRR are used interchangeable.

constraint creates a paradox for banks, in which their self-protecting mechanism increase the probability of credit loss.

The paradox can be explained as follows: banks will require a higher RRR for higher risks. However, the higher the interest rate is, the riskier the project become, because if a project does not reach a positive profit threshold (the project's RRR), a higher interest rate will decrease the probability of repayment. (Bebczuk, 2003)

In subsection 6.3, it was presented that price is the mechanism used to maintain the equilibrium of the market's supply and demand. However, in this case, price, or more precisely RRR, might not be able to do its job in bringing the market to an equilibrium. (Stiglitz & Weiss, 1981). The market is not in disequilibrium due to endogenous or exogenous reasons. Asymmetric information changes the market equilibrium, because increasing price instead of minimizing risks, increase them. So, in credit markets, it is credit rationing, and not price, that brings the market to equilibrium.

This means that it may not be profitable for banks to raise interest rate or require more collateral when the demand for loans is high, and instead banks just deny loans, which explains why SMEs have been facing difficulties in acquiring the needed financing. As explained in 6.3, this is exacerbated by exogeneous shocks such as financial crisis, because banks have a reduced access to capital. New regulations imposed by governments, in an attempt to minimize the nefarious effects of the financial crisis will only make the matters worse for borrowers.

4.6.2 Consequences of monitoring costs

The need to constant monitor the borrower, to be certain that he is being honest regarding this economic situation is costly. Lenders require the help of accountants, or lawyers, or both, to do this job, so monitoring costs minimize the lenders profit.

4.7 Mitigating information asymmetry

One of the theories regarding how to minimize the effects of information asymmetry, is through signalling, where the borrower sends credible signals to the lender, as a way to guarantee the soundness of the project. (Spence, 1973)

The theory is, however, not an effective solution because low-quality borrowers can send the same signals as high-quality ones, so lenders will remain incapable of discerning high-quality borrowers from low-quality borrowers. (Bebczuk, 2003) Signalling through contractual clause is also ineffective, because the simple act of accepting the terms of contract does not imply that those terms will be followed or respected. In subsection 6.4 it was presented, that lenders have a high liability while borrowers have limited liability. Here is the heart of the problem: lenders have more to lose than borrowers. Trustworthiness will, thus, only be achieved if the borrower bears a cost as a payment for achieving lighter credit and accepts to give up his limited liability. In order for signalling to be effective, it should "be costly to all borrowers, but more importantly, it should be prohibitively costly to the riskier borrower". (Bebczuk, 2003)

4.7.1 Collateral

One way of increasing the borrower's liability is through collateral. Collateral is defined as an asset, that the borrower will transfer automatically to the lender, if the project revenues are not sufficient to result in the full reimbursement of the loan. By choosing to offer a collateral, the borrower gives up his limited responsibility in case of negative results. The benefit of collaterals is that they reduce lender's losses if the project fails. (Bebczuk, 2003) (Ghatak & Guinnane, 1999)

The borrower who chooses to use a collateral as a form of guarantee is implicitly signalling that he perceives the probability of success as being high. The cost is the potential loss of asset. Low-quality borrowers can also use collateral in order to pretend being high-quality borrowers, but they will only do so until the benefit disappears, because it diminishes their profit. Also, there is a higher probability of failure for the low-quality borrowers. When the incentive to undertake the project disappears, only high-quality borrowers are left in the market. Hence, collateral is an effective signal in eliminating information asymmetry, and its widespread use is proof of the practical significance of asymmetric information in financial markets. (Bebczuk, 2003)

4.7.2 Internal funds

A borrower can tie his personal fortune to a project, thus giving up his limited responsibility. This is called using internal funds, and it can be perceived by lenders as a signal of trustworthiness, helping borrowers to distinguish themselves. By taking such a big risk, the borrower signals that he perceives his own probability to repay as high. It is also an effective method in eliminating or at least minimizing information asymmetry.

This signalling method carries an opportunity cost to the borrower, as he loses the possibility of gaining a return on an alternative investment. (Bebczuk, 2003)

4.7.3 Reputational capital

Even though borrowers have an information superiority, that could give them an incentive to take advantage of the lender, there are many borrowers who choose not to make use of that possibility. That could be due to the borrower's natural moral behaviour or to real world dissuasive factors, that will act as a neutralizer of the asymmetric information.

If the borrower has a natural predisposition to behave ethically and abide by his word, he will just act in that manner, and the information asymmetry will be minimized, because the borrower will be open, and therefore, the transaction will work as if the market had perfect information. However, it is most likely that the main dissuasive factor enforcing a morally behaviour in borrowers is cause by real world incentives.

One such an incentive is reputation. Enterprises do not usually have a short-term planning horizon, and most likely there is a plan for running multiple projects during an enterprise's existence. Therefore, the enterprise will need to maintain its reputation, if it has any expectation of financing its future projects. Borrowers can therefore be kept honest by ethical control, because their reputability provides them a future earnings stream, a "price premium" for being honest. This will deter opportunistic behaviour, as long as the reputational capital is greater than the gains obtained through one-time cheating. (Parker, 2005)

Bankruptcy has a high cost, not only because of the direct costs, but also because it has high reputational consequences. Specially in smaller firms, in which the figure of the owner can somehow merge with the enterprise's image, a bankruptcy can affect the owner as well as the managers, and make it much more difficult for them to achieve any other financing in the future, as well as affecting their personal prestige. (Caire & Kossmann, 2003) On the contrary, in larger corporations, that handle with vast amounts of money, it is possible that the owner or manager will be tempted with incentives so expressive, that could result in them putting their good reputation aside, because the cost of their reputation would be inexpressive when compared to the potential economic gain.

Therefore, a one-time act of dishonesty will only pay off, if the benefits outweighs this reputational cost (Bebczuk, 2003; Parker, 2005)

4.7.4 Screening

To help with this identification, banks use different screening devices, such as for example interest rates. Willingness to pay a high interest rate can signify that a project is riskier, and that the borrower perceives his own probability to repay as low. (Stiglitz & Weiss, 1981)

Because information asymmetry is inherent to credit markets, screening is needed to help distinguish high-quality from low-quality borrowers. Credit assessment, or credit analysis, is the method used to screen the borrower and identify risks in a potential loan and calculate his creditworthiness. The main goal is determining, whether a borrower will repay his loans or not. (Stiglitz & Weiss, 1981) Also, the RRR can be a tool for screening: willingness to pay a high interest rate can signify that a project is riskier, and that the borrower perceives his own probability to repay as low.

4.8 Summary

Information asymmetry is an inherit part of credit markets. It is caused by the fact that the lender does not have information or control about the intentions of the borrower regarding his ability and willingness to repay the loan on time. This lack of information makes is very difficult for lenders to distinguish between a high-quality and a low-quality borrower, and as a protective measure, lenders will opt to treat all borrowers as being low-quality, which means that the project will be perceived as risky and a high RRR will be expected. This protective measure is called adverse selection.

Although adverse selection is aimed as a protective measure for lenders, it is more detrimental than beneficial to both lenders and borrowers. One of the detrimental effects is credit rationing. Banks would have the same behaviour as described above, requiring a higher RRR to compensate for their risk. However, the higher the RRR, the higher the risk, because the RRR will take a bigger chunk of the project's profit, and if the profit is below the expected, loans might not be paid. This proves that increasing interest is not an effective measure to reduce risk, so banks ration loans instead.

Another detrimental effect is that high-quality borrowers, whose projects would require a lower RRR, are discriminated due to adverse selection, and in some cases desist of the project altogether, leaving only low-quality borrowers on the market. Also, low-quality borrowers, when faced with high RRR, tend to choose riskier projects with high expected return, because if the project succeeds, their margin of profit will be higher. Furthermore, borrowers have an incentive to disguise the true nature of their projects in order to try and acquire cheaper credit (moral hazard).

Borrowers have also an incentive to announce lower-than-actual earnings as well, in order to reduce his financial obligations at the lender's expense, so lenders incur in high costs due to the need of monitoring the borrower.

Asymmetric information can, however, be minimized by using signalling in the form of collateral, internal funds or reputational capital, because those signals increase the liability of borrowers, and hence, make them more trustworthy to lenders.

Chapter 5 – Using data analysis for credit scoring

5.1 Introduction

The previous chapters have described the challenges resulting from information asymmetry in credit markets. It also discussed, that in order to handle information asymmetry, credit assessment is necessary to screen borrowers, and distinguish the high-quality from the low-quality ones. Also, collateral, internal funds (cash) and reputational capital were presented as ways to mitigate information asymmetry and increase the trustworthiness of borrowers. While collateral and cash increase the liability of the borrower, reputational capital reflects his moral character, and therefore, can be an excellent indicator to help only select high-quality borrowers.

This chapter presents a description of the lender's decision-making in credit markets and discuss how credit assessment can be achieved. A combined model using traditional historical financial data, alternative data and quantitative data is introduced, and the methods of credit assessment are individually compared to each other in other to show how a combination of them would render an enhancement of the credit assessment of borrowers, in order to find those whose reputational capital will deter opportunistic behaviour and therefore, mitigate the effects of information asymmetry.

5.2 The lender's decision-making

In market with imperfect information, the lenders decision-making is complicated by asymmetric information, so what could be a simple decision in a market with perfect information becomes a complex task. Lender's must decide, whether they will grant a loan or not, and to which RRR.

That decision is made based on the potential risk of credit loss, and both ability and willingness to repay a loan are to be considered. Figure 5.1 pictures the lender's decision-tree. The lender begins by evaluating, whether the borrower has the ability to repay the loan, considering that the borrower's economic situation remains unchanged. If the answer is yes, then the lender evaluates, whether the borrower is willing to repay the loan. If the evaluation reaches yes to both questions, the loan should be approved.

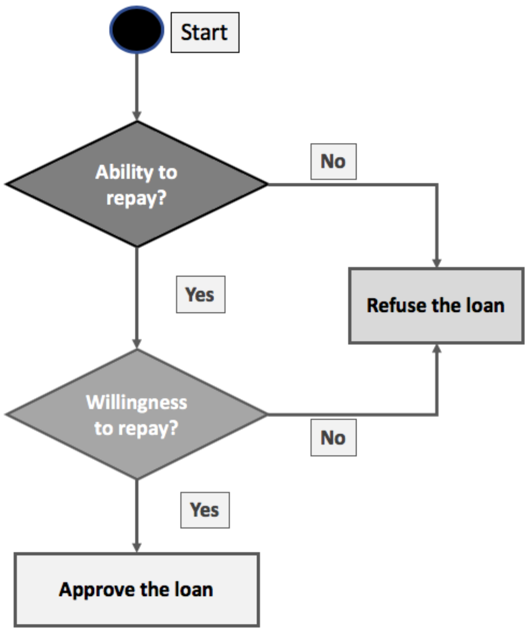

Figure 5.1 – Lender's decision tree

However, markets have imperfect information, forcing lenders to find ways to corroborate the statement of the borrower, that he can and is willing to repay the loan. It is here, that credit assessments are a vital tool for risk management.

5.2.1 Ability and willingness to repay a loan

The repayment of a loan is dependent on two important characteristics pertaining the borrower: his economic situation and his intentions.

5.2.1.1. The borrower's economic circumstances

This first characteristic defines the borrower's ability to repay a loan. It can be determined by analysing the borrower's financial information regarding his liquidity, profitability and leverage.

5.2.1.2. The borrower's intentions

This characteristic defines the borrower's willingness to repay a loan. It can be determined by analysing the borrower moral character, by observing his choices, as they reflect the borrower's intention, or lack thereof, of repaying a loan on time.

5.3 Credit assessment

Credit assessment[30] can be defined as the method used to determine the loan's safety, profitability, and the sustainability of its purpose (Sathye and Bartle, 2017), and hence determine the risk for default[31], distinguishing high-quality borrowers from low-quality ones.

Two kinds of data are involved in credit decisions: financial data and non-financial data. Financial data relates to information that can be expressed in number and pertains to a person or company, such as financial statements, auditor reports, bank statements, tax statements among other documents. Non-financial data cannot be expressed in numbers and have a more subjective nature, such as a person or clickstream, online behaviour, company's reputation, customer satisfaction, phone usage, social presence and so on. Credit assessment usually takes in consideration three main factors: external factors, internal factors, and borrower-specific factors.

5.3.1 Internal & External factors

External factors regard legislation, macro-economic factors and industry-specific factors, and internal factors refer to the lender's lending policy and loan budget. (Sathye and Bartle, 2017)

5.3.2 Borrower-specific factors

Borrower-specific factors should meet the "five Cs" criteria[32]: character, capacity, capital, collateral and conditions[33] (see figure 5.2).

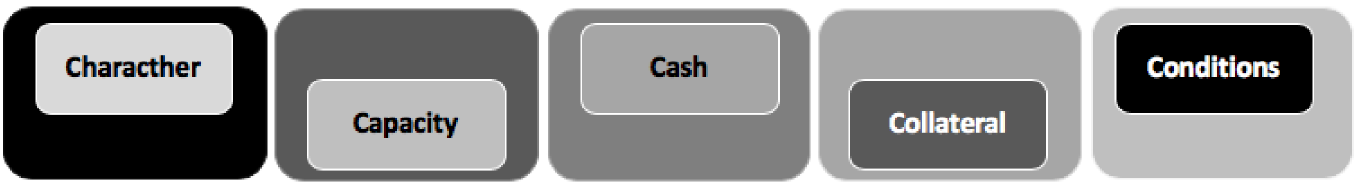

Figure 5.2: The five criteria of the borrower-specific factors

30 Credit assessment and risk assessment are used interchangeable
31 Risk of default and credit risk are used interchangeable.
32 https://www.wallstreetmojo.com/credit-analysis/
33 https://strategiccfo.com/5-cs-of-credit-5-cs-of-banking/

5.3.2.1. Character

This first criterium refers to the person of the borrower, in other words, who the borrower is, and depicts the borrower's willingness to repay or lack thereof. The goal of this criterium is to answer the following questions:

- Based on past transactions, the borrower is a good or bad payer? Delinquency[34] or default[35]?
- Is the borrower registered in a bad payer registry like RKI[36]?
- Has the borrower experienced bankrupting in the past, or is his company in the process of bankruptcy?
- Is the borrower forthcoming in regard to all the required information?
- Does the borrower look like an honest and truthful person?
- Was there a previous case of delinquency/default, and if so, how did the borrower dealt with this?
- First the hard data would be analysed, in order to understand the borrower's credit history, and check for eventual lawsuits due to default or registration on a bad payer registry. After the hard data is analysed, the lender will analyse the borrower's behaviour.

5.3.2.2. Capacity

This criterium refers to borrower's ability to repay and analyse different ratios, regarding profitability, leverage, liquidity and instant coverage, analysing, therefore, ratios such as debt-to-income ratio, ROA, net margin, EBIT, and debt-service coverage ratio. This analysis aims to ascertain, that the economic situation of a business to ascertain, whether the business can repay a loan on time.

5.3.2.3. Cash

This criterium refers the capital invested by the borrower on the business or the project. It affects both ability and willingness to repay. The fact that the borrower has invested his own capital on his business, in form of equity, or on the project, in form of part of the investment or a down payment for the loan,

---

34 Delinquency happens when the borrower is late with his payments for more than 90 days.
35 Default happens when the borrower does not fulfill his obligations at all. It is therefore, beyond being just late in payments.
36 RKI is a company owned by Experian that is the largest registry of bad payers in Denmark.

is a signalling of his willingness to repay. However, the presence of this capital is also increased indication of ability to repay. An analysis of the debt-to-equity ratio can also be done during this step.

5.3.2.4. Collateral

This criterium refers to whether the borrower has offered a collateral or guarantee, that can improve his chances of being granted a loan. Also, a collateral can result in lower interest rates and better terms for the borrower. As with cash, a collateral is a good indication of ability and willingness to repay.

5.3.2.5. Conditions

This criterium aims to answer questions regarding key risk areas in the industry and characteristics of the loan:

- What is the purpose of the loan?
- What are the conditions of the loan, i.e. interest rate, amount loaned, duration of loan?
- What is the state of economy?
- What are the industrial trends in the borrower's industry?
- Are we expecting any legal changes, that could affect this industry?

5.3.3 Traditional method of credit assessment

Traditional credit assessment can be defined as a method to effectively establish the economic conditions of the borrower. It is a subjective credit assessment method based on the judgemental appreciation of the borrower-specific data (historical financial data), made by an analyst with expertise and experience, and following the 5 "Cs". The extent of the credit analysis can be more or less comprehensive in proportion to the size and terms of the loan, and if the assessment is positive and the loan is considered a good investment, taking in consideration opportunity costs, the loan is granted. This method of credit assessment takes in consideration borrower-specific historical financial data[37], that contains mostly financial data but also some non-financial data in order to establish creditworthiness. (Volk, 2012)

The financial data is used to determine ability to repay and its indicators are calculated based on the borrower's income statement, balance sheet and tax statements.

[37] Historical financial data is all data pertaining the borrower's finances: Income, taxes, saving, past credit behaviour, registration in lists of bad-payers and so on.

Verification of income from both income statements and tax files (TastSelv)[38]

- Liquidity analysis – current, cash and operating cash-flow ratios
- Leverage analysis – debt ratio, debt service coverage ratio
- Profitability analysis – ROA, ROE, gross, operating and profit margin, net gearing, RAROC
- Efficiency ratio – Inventory conversion, DSO, asset turnover ratios

The non-financial data is used to both ability and willingness to repay and is based on analysis of the borrower's:

- Company age
- Industry
- Competition in the market
- Market size and share
- Growth potential
- CPR and CVR registration
- Check if the borrower is registered in bad-payers registries like RKI and Bisnode's debtor registry

It is important to note that micro, small and sometimes even medium-sized enterprises are managed by one "key" individual, so the credit quality of those kinds of business are usually mirroring the entrepreneur's financial behaviour. This means that "the likelihood of timely repayment is directly related to that entrepreneur's willingness to repay". (Caire & Kossmann, 2003) Therefore, default prediction can also be based on many of the key entrepreneur's personal characteristics.

5.3.4 Data analytical methods of credit assessment

Data analytical methods of credit are, as the name infer, the assessment of credit with the use data analysis. Data analysis is basically "all the ways you can break down the data, assess trends over time, and compare one sector or measurement to another. It can also include the various ways the data is visualized to make the trends and relationships intuitive at a glance"[39]. The purpose of the analysis is to transform good data into useful information to improve decision-making. Data analysis is possible due to data mining, text analytics, business intelligence, and data visualizations, among others.

[38] TastSelv allows a person to give another authorization to access or change tax information on SKAT. - https://skat.dk/skat.aspx?oid=1924748
[39] https://www.ngdata.com/what-is-data-analysis/

The technological advances in processing speed, cloud storage, and social networks, created the possibility of gathering and storing big data. (Earley, 2015; Yan, Yu & Zhao, 2010) Big data is characterized by the three "Vs"[40]: "high-volume, high-velocity and high-variety"[41].

Data analysis creates models of predictive nature based on patterns encountered through statistical analysis of large quantity of data. It predicts future behaviour based on past behaviour[42]. The process involved in the creation of a predictive model is pictured in figure 5.3. An objective for analysis is chosen and a goal is identified, so data is collected from the appropriate data sources, then the relevant variables are selected, and the dataset is cleaned for leakage. Using this new cleaned version of the dataset a predictive model is created, giving a degree of accuracy to the prediction, and the model is then evaluated.

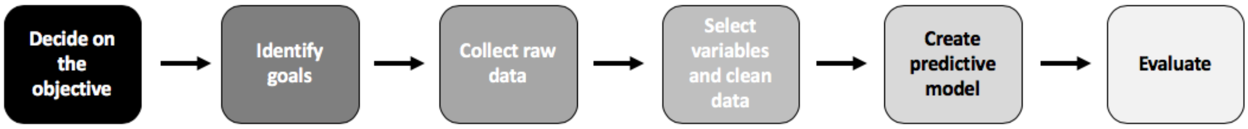

Figure 5.3 – Process for data analysis

One of the big complexities of data analysis is that the analyst must be able it is complex to engage in data analysis[43], requiring an analyst able to engage and understand the data (Earley, 2015). The main issue is selecting which data is appropriate to apply in the research and which data should be ignored. That selection of data is essential to avoid over-optimistic predictive models due to data leakage[44].

5.3.4.1. Behaviour analysis

Behaviour analysis can be defined as analyse of non-financial alternative data from a person or entity in order to identify a goal. In the case of credit assessment, the goal is the likelihood of default. Behaviour analysis can use a wide variety of data from different sources such as clickstream, online presence data and so on.

[40] Big data is characterized by the three "Vs": volume, velocity, and variety, but sometimes veracity is added to the description - https://www.dexlabanalytics.com/blog/the-opportunities-and-challenges-in-credit-scoring-with-big-data
[41] https://www.gartner.com/it-glossary/big-data
[42] https://halobi.com/blog/how-to-implement-predictive-analytics-into-your-company/
[43] Predictive models[43] are created by: deciding on the objectives, identifying goals, collecting data, cleaning for leakage, growing a data science team, optimizing and repeating to perfect the model. - http://www.oracle.com/us/corporate/profit/big-ideas/052313-gshapira-1951392.html
[44] Data leakage happens when the data you are using to train a machine learning algorithm happens to have the information you are trying to predict - https://insidebigdata.com/2014/11/26/ask-data-scientist-data-leakage/

5.3.4.2. Clickstream data

Clickstream data is data collected by the lender regarding the borrower's behaviour on the lender's website. A clickstream is a sequence of HTTP requests made by a borrower on the lender's website. Whenever the borrower clicks on something on the website, that click will receive a timestamp, an anonymised user ID, and activity and related variables, describing the behaviour of the borrower on the website. (Yang, Zhang, & Guo, 2018)

Clickstream can collect data regarding how the borrower got to the site, if he has already been on the site before[45], the duration of the visit, timestamp, IP address[46], session details such as the time he takes on each page, the sequence of pages he visited, visitor's demographic and so on.

The collected data can be used in two ways: the first is based on known indicators, so analysis is done automatically, based solely on the borrower-specific data, and the second is by finding statistical patterns on quantitative data to create predictive models for credit scoring.

There are many indicators regarding the borrower's behaviour during the navigation of a site, that can be used to establish likelihood for default, such as time in which the loan is requested, the choice of the borrower regarding reading terms and conditions of the loan, and the interest in knowing before requesting the loan, which is the RRR of the loan. So, if a borrower requests a loan on a Friday evening, after midnight, there is an indication that this is not a high-quality borrower. This points more towards the desperation act of a person so pressured economically, that he is having a sleepless night. The same is the case for a borrower that is not interested in knowing how much he will have to pay in interest rates for the requested loan – this is a serious indication of unwillingness to repay, as the borrower does not care if he has ability to repay the loan. One would assume that a borrower requesting a loan does not have a limitless amount of funds available, otherwise he would not need a loan. If his available funds are limited, he should be concerned of whether he can fit that loan on his budget or not.

Studies regarding the use of clickstream regarding credit scoring are rather promising. A Chinese study of a clickstream model called DeepCredit, based on data collected from a large Chinese P2P platform, has achieved AUC scores of 0.89 predicting delinquency and 0.90 in predicting default. The model is currently being implemented in China. (Yang, Zang & Guo, 2018)

---

45 Websites usually download a cookie to computers, so they can recognize visitants and remember what those visitants have done on the site on their last visit. Cookies can be removed, in which case the website will not be able to know, that the user has been on the website before.

46 Internet Protocol (IP) address is a numerical label given to every computer connected to the internet. It has two functions, identification and address location. This means, that is possible to find the location of the user based on this number. This information can, however, be manipulated by the use of a VPN (virtual private network) software.

Also, Big Data Scoring (see Appendix 01), a cloud-based service from UK, has been using clickstream for credit assessment for the past couple years, with a significant degree of success. In a case-study[47] from 2014, the company describe their results by adding a mixture of clickstream, web search, address investigation and other data points from various public data sources to a central European lender's in-house credit assessment, achieving a reduction of 34,7% of credit loss.

5.3.4.3. Social media data

Social media data is all the data regarding a person's or entity's social presence on the internet. Social networks like Facebook, Twitter, Instagram, Snapchat, YouTube, and so on, gathering huge amounts of data from more than two billion people worldwide[48]. Personal reviews or star-ratings on those sites, as well as sites such as Trustpilot, Yelp, and Amazon, among others, can also contribute to the volume of useful data. (Yan, Yu & Zhao) This means that each user has a large footprint, with his preferences, personal data, interests and list of friends available in the pool of big data.

Among the variables required are demographics, preferences and network (Tan & Phan, 2016), as well as number of posts (from the company and from visitors), comments, fan counts, credit ratings, overall star ratings, and likes and dislikes. Behaviour analysis has been proved in studies to being capable of predicting default with an accuracy of almost 85%, when used in association with historical financial data. (Zhang, et al., 2016)

Predictive models based solely on social media have already been implemented in the real world with great results. Big Data Scoring created in 2013 a credit scoring model using Facebook data alone, with a high degree of accuracy. In a case study[49] from July 2013, the company concluded that the model could also be used to enhance the existing in-house credit assessment of a bank or other lending company.

5.3.4.4. Other kinds of data

There are many other kinds of data, that can be used to analyse behaviour, such as utility data, mobile data and user preference.

---

[47] https://web.archive.org/web/20151022010402/http://www.bigdatascoring.com/2014/11/case-study-about-a-central-european-lender/index.html
[48] https://dazeinfo.com/2017/07/19/social-media-politics/
[49] https://www.bigdatascoring.com/study-of-credit-scorecard-using-only-facebook-data-3/

Utility data is basically credit history data, but very few credit bureaus use this data[50]. Mobile usage[51] is always stored in usage logs, and that information can be used as raw data to find behaviour patterns predicting the likelihood for default. (Pedro, et al., u.d.). Information regarding phone subscription and even the choice over mobile systems can be relevant in determining creditworthiness. However, that is only the beginning, data regarding the type of phone, whether iPhone or Android[52] is also relevant, as well as information regarding the kind of phone plan or lack thereof a borrower has.

Clickstream, mobile data and social data are mainly objective data, that mirrors the behaviour of the borrower online. It is not data provided by the borrowers, and therefore, data that could potentially be manipulated, but instead, data acquired proactively by the lender. Because the data is objective, it has a higher value and can contribute to the reduction of information asymmetry. (Yan, Yu & Zhao, 2015)

Behaviour analysis is therefore used to find the borrowers, whose reputational capital can deter opportunism, so those borrowers will act ethically, minimizing concerns regarding both adverse selection and moral hazard. If the borrower is not acting opportunistic, and the credit assessment is accurate, information asymmetry is minimized. Though, using this kind of data must be done with caution, as it can result in predictive patterns that are discriminating, due to the classification of individual or patterns in different groups. (Hurley & Adebayo, 2017)

5.4 Credit scoring

Credit scoring is "the process of assigning a single quantitative measure, or score, to a potential borrower representing an estimate of the borrower's future loan performance". (Feldman, 1997) This method creates statistical models based on the experience gained from performance and characteristics of previous observations, from both financial and non-financial data. (Ciampi & Gordini, 2013; Altman, Sabato & Wilson, 2012). In the credit market, a credit scoring can be made based on previous loans in order to predict the future performance of similar loans, based on relationships between application information and the likelihood for default. (Caire & Kossmann, 2003) Likelihood of default is usually presented as a binary option answering a question such as "Is this loan likely to default?", and the credit scoring will result in a yes or no, presented with a numerical weight, i.e. 1 for yes or 2 for no.

Credit scoring can enhance this assessment, by classifying borrowers into different groups. (Lancher, 1995; Hsieh, 2004) Those different categories of borrowers are analysed in search of similar behaviour

[50] https://www.fico.com/blogs/tag/utilities/
[51] http://archive.boston.com/business/technology/innoeco/2012/01/what_does_your_mobile_phone_us.html
[52] https://www.wired.com/story/your-smartphone-could-decide-whether-youll-get-a-loan/

patterns, categorizing loans as accepted or rejected based on borrower's features such as age, income, marital status, education, preferences, address etc. (Chen & Huang, 2003). The analysis of this classification and categorization will generate predictive models of shared behaviour patterns that will be able to assess, whether a borrower is likely to default.

Credit scoring is a two-phase process: the first phase comprises of quantitative analysis of a dataset, for example, a dataset about previous borrowers and their loan performance (historical financial data), or a dataset containing behaviour data, and the second phase is using the results to evaluate potential borrowers with similar information and predict his likelihood for default. The main advantage of data analysis is that the volume of data allows this predictive model to continuously improve itself.

Statistical models are extremely powerful, but they require a large "pool" of at least 1.000 negative outcomes (defaulted loans) in order to be able to create an effective predictive model. (Caire & Kossmann, 2003) Another important characteristic of credit scoring is that it is not capable of affirming that loan A or B is going to default or not. The model can only show that there are statistical patterns, predicting default, with a certain degree of accuracy. This degree of accuracy is given by the Area under the curve (AUC), or under the Receiver Operating characteristics curve (ROC), which is a graphical plot that shows the predictive ability by plotting true positives as a function of false positives[53]. An AUC of 1 gives 100% accuracy and is in theory perfect. An acceptable AUC is above 0.7 and a great AUC is higher than 0.85. Nearly perfect AUCs are most likely the result of data leakage – information added to the dataset, that were not available at the moment of the loan (see figure 5.4)

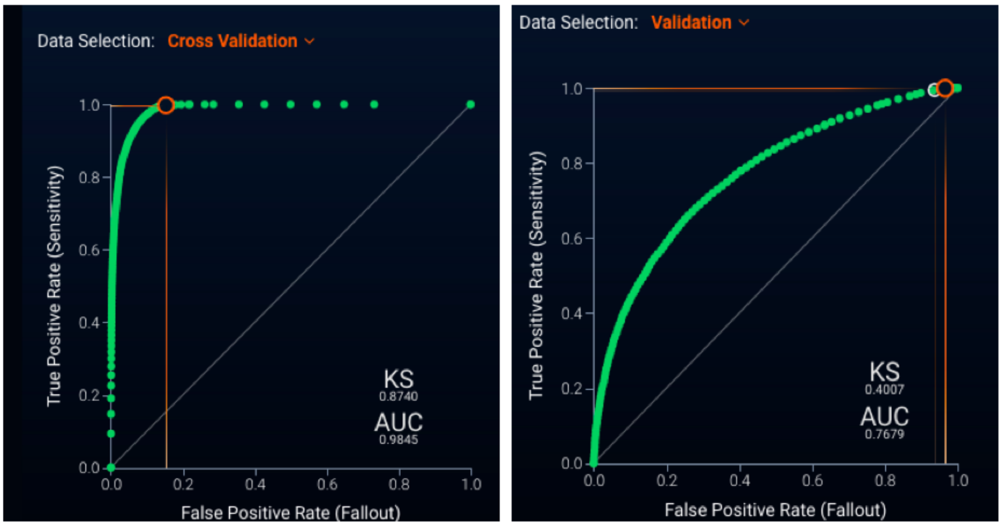

Figure 5.4 – The picture above shows two different ROC curve graphics. AUC is the area under the green curve. In the first graph, the AUC is 0,9845, so 98,45% accurate. This high degree of accuracy was cause by data leakage. The second graph has an AUC of 0,7679. Both AUC relates to the same dataset, however the dataset was cleaned for data leakage, resulting in a better and more reliable model.

[53] https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

An issue regarding credit scoring relates the data anonymization, that is required by some regulations such as the General data protection regulation (GDPR)[54]. If data is too anonymized the predictive model might lose the kind of valuable insights that come with more detailed information[55]. Also, data cannot be used for profiling, and if that happens, the borrower must be informed and have the option to opt out. Additionally, profiling cannot take in consideration any discriminatory factor such as race, politics or religious beliefs[56].

5.5 How is credit being scored in Danish P2P platforms

Lendino[57] was the first crowdlending platform in Denmark. Launched in 2013, the platform has loaned more than 105 million DKK in 358 loans at an average interest rate of 8.6%. Lendino's credit assessment is based solely on borrower-specific data (qualitative analysis) and the platform assigns different credit class and different credit risk to the borrowers, from A+ to C-. The qualitative data used is as follows:

- Company's credit history and research of publicly available information
- Is the company's audit report blank?
- What is the company's business model?
- How is the company design and who is the owner and top management?
- Was there any tax change?
- If the company has changed address frequently
- If the company has a separate finance function
- Interview or written correspondence with the borrower

Lendino takes in consideration for their credit assessment some behaviour indicators from borrowers, as for example, the time stamp of loans. However, the platform does not use clickstream for collecting data, nor other form of data analytics. According to Kristian M. Frederiksen, Lendino's credit officer, the platform does not outsource due to concerns regarding black-box predictions[58]. The platform also makes

[54] GDPR protects or regulates the processing of personal data, i.e.. any kind of information about an identified or identifiable natural person, which is the data subject
https://blog.instantor.com/ultimate-guide-gdpr-impacts-on-credit-ratings
[55] https://piwik.pro/blog/the-ultimate-guide-to-data-anonymization-in-analytics/
[56] https://dataconomy.com/2018/04/how-gdpr-will-affect-data-science/
[57] www.lendino.dk - the information regarding how Lendino credit assess borrowers was acquired in their homepage and confirmed during a phone call with Kristian M. Frederiksen, Lendino's credit officer.
[58] Supervised machine learning models using huge amounts of both numerical and categorical data are develop through a training process focused on non-linear relationships in the data, that can result in black-box predictive models, which are models that have a high degree of prediction accuracy, but are rather difficult to understand, and therefore, difficult to validate.

it clear on their website that their credit assessment is for guidance and that the lender is responsible to carry out his own credit assessment before investing.

Better rates[59] was founded in 2016 and has loaned more than 255 million DKK in 9092 loans. The company uses both borrower-specific data and quantitative data to create a credit scoring for each borrower based on the following information:

- The company's verified income through tax data source via TastSelv
- The company's verified debt ratios through the same tax data source
- The borrower's debt factor
- Data from Experian Segmentation Data. This data is prepared using both data from Experian and Statistics Denmark – the data divides the Danish population into different financial segments
- Interview with the borrower (by phone)
- Better rates credit algorithms
- Borrower information regarding jobs, housing, civil status etc.
- Check of borrower status on RKI and Debtor Register – they refuse all loans of borrowers registered in those bad-payer registers.

Additionally, Better rates only accepts borrowers, whose verified income is of at least of 240.000 DKK yearly.

5.6 Improving the current model and minimizing information asymmetry

This thesis hypothesis is, that a combined method will increase the accuracy of credit assessments. This means, that the elements of a traditional credit assessment, where mostly financial data, but also some non-financial data, is analysed will be done automatically and complemented by elements of behaviour analysis, in which only non-financial data is analysed.

[59] www.betterrates.dk – the information regarding how Better rates credit assess borrowers was acquired in their homepage

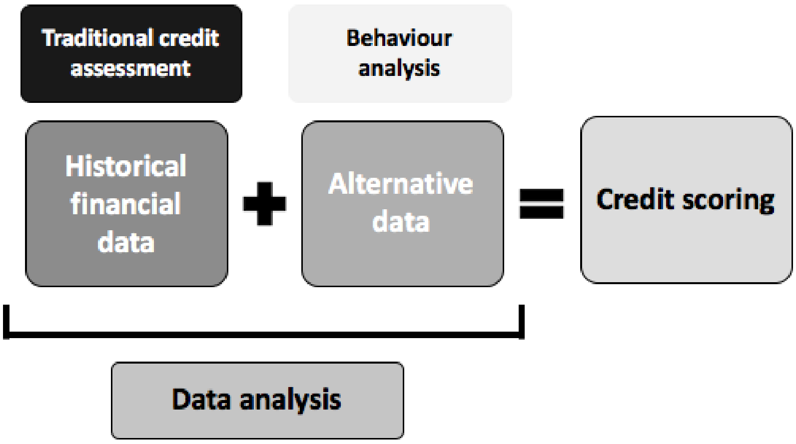

Figure 5.5 – Blueprint for minimizing information asymmetry

It has been already discussed, that to determine a borrower's creditworthiness it is necessary to assess both his ability and willingness to repay. In Chapter 4, some factors have been presented as ways to mitigate information asymmetry: collateral, cash and reputation.

Collateral and cash are ways of increasing the borrower's liability in an attempt to signal trustworthy, which means that those are actions a borrower take as signalling creditworthiness and therefore are rather straightforward. Either a borrower acts or not, and the only open discussion is if the size of the economic cost born by the borrower is enough to increase his creditworthiness to desired levels.

Reputation, on the other hand, is an excellent indicator of willingness to repay, but it is rather complicated to establish, because it does not derive from an action, and therefore, can only be mirrored by the borrower's behaviour. Herewith, the difficulty in establishing reputation; one would need to have a really good picture of the borrower's behaviour in different aspects of his life in order to be able to establish, whether the borrower has a good character or not.

This thesis aim is to find ways to better identify those borrowers, whose reputational capital can deter opportunism, because those borrowers will act ethically, and therefore, information asymmetry will be minimized. The following paragraphs will compare different forms of credit assessment introduced, to determine their strengths and weaknesses, in order to demonstrate, how a combined model using data analysis can be used to enhance the ability a lender has of identifying trustworthy borrowers.

5.6.1 Comparative analysis using SWOT

Figure 5.6 shows how the comparative analysis will be done. This subsection is, therefore, divided in four parts. In each part, an analysis of how this form of credit assessment affects collateral, internal funds and reputation is done, as those elements are important for the mitigation of information asymmetry.

Additionally, an analysis using SWOT (strengths, weakness, opportunities and threats) is done, to show how those different forms of credit assessment can complement each other, when combined together.

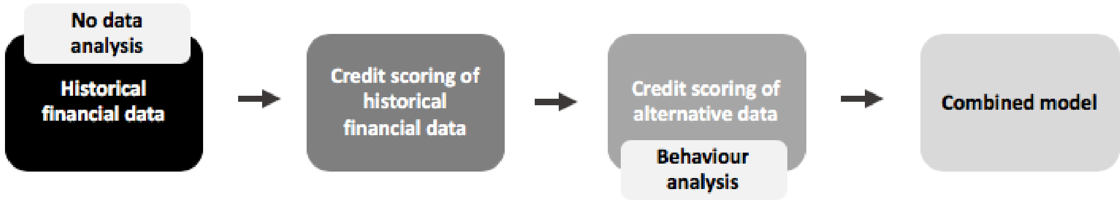

Figure 5.6 – Illustration of the model for comparative analysis

5.6.1.1. Traditional credit assessment with historical financial data

This method of credit assessment can determine both ability and willingness to repay. The analysis is subjective, made based on the conclusions and judgement of an experienced analyst. Although part of this analysis is done based on facts (hard data), some of it relies on soft data, that is difficult to analyse and interpret, and passive of mistakes based on the analyst's prejudices.

By analysing the financial data from the borrower, the lender can have a much better comprehension of

- management's performance (profitability ratios)
- the company's ability to pay its current obligations (liquidity ratios)
- how much the company is dependent of loans to finance its operations (leverage ratio)
- how effectively the company uses its credit, inventory and assets (efficiency ratios)

Also, adding non-financial data to the financial analysis can tell a lot about the risks and the future of the company. Indicators like market size, competition, age of the company can predict how the company is expected to develop in the future: new companies, for example, have a higher rate of bankruptcy ((Duan, et al., 2009), so the risk of default is bigger. Also, by analysing past events of default or delinquency[60], observed in finished or ongoing legal proceedings for non-payment, overdrafts, and registration in agencies for bad-payers, the analyst will be able to form a bigger picture of the moral standing of the borrower.

This method of analysis can also expose a little about the borrower's willingness to repay, by analysing the borrower's past behaviour, and use this as an indication of future behaviour.

[60] Default happens when the borrower fails to repay the loan while delinquency happens when the borrower misses the due date.

Collateral and internal funds

Traditional credit assessments are excellent at verifying this signal of trustworthiness. As said before, collateral and cash are signals that either are present or not. A loan backed up entirely on collateral, like for example a loan for a real estate backed up by the deed of the property, is a loan with very little risk. The lender will still bear the costs regarding the sale of the property in case of default, but most of the debt will be repay, as long as the property does not lose significantly value in the time between the contractual moment of the loan and the default. This proves that having a full coverage through collateral might be enough in determining, whether a loan should be issued[61]. The borrower's willingness to repay is already indicated by his choice in offering this guarantee and further determination of willingness to repay is not necessary.

However, when the collateral is partial, and therefore, does not cover the entire value of the loan, willingness to repay should still be determined.

Also, a loan with a value inferior to the value invested in the company in the form of equity has a bigger chance of being repaid, than loans who are not backed up by anything. Additionally, borrowers who invest in their projects have increased their own risk, and will be therefore, more cautious in their handlings. In all cases, the borrower has increased his liability, with signals trustworthiness and willingness to repay. However, further determination of willingness to repay is still needed, as capital invested in projects does not increase the chance of the lender being paid, just shows that the borrower is serious regarding his project.

Reputation

While a lender can check the past behaviour of a borrower and use this information as an indication of the borrower's character, this data can only show that the borrower has not acted opportunistically yet. This does guarantee that he will always behave ethically. It is possible, that the borrower did not have incentive to behave opportunistically until now. So, the prediction of willingness to repay is rather limited, when only this method is used, and establishing reputation with this model alone is not efficient.

[61] This is correct as long as the property does not lose value during the period between the loan and the default.

SWOT analysis of this model

In the following paragraphs, the strengths, weaknesses, opportunities and threats of using a traditional method of credit assessment based on borrower-specific historical financial data are discussed.

1. **Strengths**

This method is straightforward and has been used successfully for a long time. It is a good method in determining ability to repay and can give an input regarding willingness to repay.

2. **Weakness**

This method is based on the judgement and expertise of an analyst and therefore it is based on subjective analysis, which can be flawed. Additionally, the determination of willingness to repay is limited to the observation of past behaviour, which does not necessarily give the lender a good understanding of whom the borrower really is.

3. **Opportunities**

Many studies on the use of data analysis for credit scoring, especially regarding the use of alternative data such as behaviour analysis are made with the data from developing countries. The main reason for that is that those countries have less strict laws regarding the storing of data and privacy rights. In the western world, the individual privacy has received more legal protection, and this could make data analysis impractical, due to excessive anonymization of the stored information. So regulatory issues could, in worse case-scenarios, reduce the use of other methods, making this method a good alternative.

4. **Threats**

This is an efficient model, but it is outdated and can be improved. Newer and better models are most likely be developed, and unless, as already mentioned, regulation makes it impossible to store and share data, this model, when used alone, will render a weaker result in credit assessment in the future.

5.6.1.2. Credit scoring of financial historical data

This method is the association of the previous method and data analysis, and therefore, it will be based on both borrower-specific historical financial data and quantitative data based on past loans. Data analysis can significantly improve the credit assessment, because patterns, that are not easily noticed by an experienced analyst are easily caught by artificial intelligence. Another advantage of using this method is, that the analysis becomes less subjective and more objective, since the decision of whether to loan or not is no longer encumbered to a person.

Data analysis can find patterns describing better ratio ranges, better proportions between collateral/cash and loan value as well as which purposes or demographics regarding loans have a lower risk of default. This model can increase the number of loans because instead of using a one-value-fits-all model, it can adjust the different values and ratios to different industries and levels of market competition.

<u>Collateral and cash</u>

Credit scoring can discover the best proportions between collateral or cash, and herewith improve the determination of ability and willingness to repay.

<u>Reputation</u>

Find demographics and delimit the purposes that are less likely to default. However, this method only covers the data from past loans, in other words quantitative historical financial data, so unless the model is more complete, we cannot have a very good picture of the borrower.

<u>SWOT analysis of this model</u>

In the following paragraphs, the strengths, weaknesses, opportunities and threats of using data analytics to perform a credit scoring based on both quantitative and borrower-specific historical financial data are discussed.

1. **Strengths**

It is an objective approach to credit assessment, which increase the reliability of the results. Since the borrower is no longer dependent on the good impression he causes on the analyst, the risk of discrimination is decreased. The determination of both ability and willingness to repay is enhanced. Additionally, automated analysis can lower search costs.

2. **Weakness**

The main weakness it that data analysis requires a large volume of data to develop effective predictive models. Also, data must be cleaned and prepared with caution to avoid over optimistically results. Additionally, the kind of data available for this kind of credit scoring (historical financial data regarding previous loans) does not give a complete picture of the borrower, and therefore, may not contain sufficient information to better estimate willingness to repay. (Mester, 1997)

3. **Opportunities**

Both the volume of stored data and the capacity of computer systems has been increasing exponentially. New technological advances will most likely improve the quality of credit scoring while reducing costs.

4. **Threats**

Regulations such as the GDPR create an obligation to anonymize data, which can result in lower quality data, depending on how the anonymization is done, especially is many variables in a dataset are affected.

5.6.1.3. Credit scoring using alternative data (Behaviour analysis)

Because this method uses alternative data, it can be very effective at determining willingness to repay. A wide range of different data sources about the borrower's behaviour, can offer various different angles of the borrower's personality. The more data from different sources, the more angles you might be able to see. It helps creating a 360° profile of the borrower.

Collateral and cash

This model does not affect or improve the analysis of collateral and cash, nor does it help prove its presence or lack thereof.

Reputation

Due to the variety of data, behaviour analysis can help identifying those borrowers, whose reputational capital can deter opportunistic behaviour.

SWOT analysis of this model

In the following paragraphs, the strengths, weaknesses, opportunities and threats of using data analysis to perform a credit scoring based on both quantitative and borrower-specific alternative data (behaviour analysis) are discussed.

1. **Strengths**

Behaviour analysis is based on objective data, because this data mirrors the behaviour of the borrower online. It is data acquired, not given by the borrower, and therefore, the risk of manipulation is smaller. Data analysis with the use of this kind of data improves the determination of willingness to repay, due to the better view of the borrower's character.

2. **Weakness**

One of the main issues regarding data analysis using behaviour data regards ownership of data. Although data is continuously collected by different platforms, this data might not be available for credit assessment. This means that, though there is data enough to develop predictive models, it might not be available. Another core issue is that behaviour data can result in discriminatory models, due to the way borrowers are classified into different groups. An example to explain the eventual discriminatory nature of behaviour analysis is, that the model might identify a pattern regarding individuals who purchase strawberry yoghurt at Netto on Fridays as bad-payers. That does not mean that every individual buying strawberry yoghurt at Netto on Fridays are bad-payers, just that a significant number of bad payers have this desire to eat strawberry yoghurt on Fridays. But if this model is used on its own, and this particularly characteristic is considered of high-importance by the model, people who have urges for eating strawberry yoghurt on Fridays and have a Netto close-by will have their requests denied.

Another issue regards the privacy rights of the borrower. Social data, for example, is shared within a context[62], and are not meant to be used in another context, as for example for credit scoring.

3. Opportunities

Every single person in the western world use some kind of social media, and companies realize the value of having online social presence. This means that data is being continuously stored and is usually updated, so models will have the ability of constant improvement.

4. Threats

The main threat regards regulatory measures to protect individual privacy, that might restrict the storing of data, require its anonymization or totally ban access to social media data for credit scoring. Another threat is that, if people realize what variables are used in credit scoring, their social data will stop being objective, because borrowers will start manipulating their social media, by only connecting with the right kind of people or showing acceptable behaviour.
Furthermore, there is a risk that credit scoring companies might abuse their power by denying the right of loans to individuals or companies, that do not fit within certain parameters regarding acceptable opinion in order to silence dissent, as it is already occurring in the USA[63] and China[64].

5.6.1.4. The combined model

As shown above, by combining the strengths of traditional credit assessment, behaviour analysis and credit scoring, we can achieve a better determination of ability and willingness to repay. Each different model has its strengths, but they also have some weakness, that could compromise the results of credit scoring. By adding the three methods together, the strengths of one model will minimize the weakness of the other, creating a model that is less flawed.

62 https://eba.europa.eu/regulation-and-policy/consumer-protection-and-financial-innovation/discussion-paper-on-innovative-uses-of-consumer-data-by-financial-institutions;jsessionid=BDF5F2C46E63062802DDBA22DEDD1CC0?p_p_id=169&p_p_lifecycle=0&p_p_state=maximized&p_p_mode=view&_169_recordId=1542519&_169_struts_action=%2Fdynamic_data_list_display%2Fview_record
63 https://www.nationalreview.com/2019/04/chase-bank-conservative-customers/
64 https://www.businessinsider.com/china-social-credit-system-punishments-and-rewards-explained-2018-4?r=US&IR=T

Combined models have a superior capability in determining both ability and willingness to repay, but they also minimize the risk of discriminatory predictive models, as the number of variables taken into modelling is higher. This means, that in models using alternative data alone, those patterns that can result in discriminatory outcomes could have a higher importance in the modelling of predictions, due to the lower number of considered variables, than models combining a larger array of data sources, because the higher number of variables will prevent discriminatory patterns form having an excessive importance in developing a predictive model.

Also, this combined method is automated, so it can reach faster decisions, using alternative data, that is already being stored in large quantities and is continuously updated, improve underwriting outcomes, while potentially lowering costs. There is, therefore, a great opportunity for credit assessment improvement, especially when computing power is increasing and cloud computing becomes more and more cheap. Lastly, better and faster credit assessment can reduce the borrower's cost of the loan, as RRR will be lower.

5.7 Summary

The decision-making regarding the financing of a project should be simple: only finance projects from borrowers who are able and willing to repay the loan on time. However, information asymmetry makes this task rather complicated, as lenders can't distinguish between high-quality and the low-quality borrowers. Lenders need, therefore, to screen borrowers, and determine their ability and willingness to repay. This screening is done through a credit assessment. While determining ability to repay is straightforward, willingness to repay, that is reflected on the borrower's character, requires a better view and understanding, of whom the borrower really is.

Credit assessment has been done traditionally by analysing borrower-specific historical financial data. This method is efficient in determining ability to repay, but can only determine, whether the borrower has, until now, behaved morally. Therefore, its determination of willingness to repay could can be improved.

New technologies have allowed the use of machine power to find patterns or trends that can predict default (data analysis). The most common form of data analysis is credit scoring using quantitative historical financial data. This method is capable of predicting default based on statistical models capable of spotting patterns on quantitative data, creating predictive models that can, with high-accuracy, predict the likelihood for default. Those predictive models need to be evaluated against the borrower-specific historical financial data, resulting in a credit scoring based both on quantitative and qualitative data. A credit scoring can improve both the determination of ability and willingness to repay. However, even though the determination of willingness to repay is improved using this technique, it still has a very limited set variables to work with, as it is only based on quantitative historical financial data.

If understanding the borrower is fundamental to the determination of his character, alternative data should be taken in consideration while doing a credit scoring. This is also known as behaviour analysis. This data is already being continuously stored from many different sources and platforms over the net, such as clickstream data, social network data and even mobile data, so it can be used to develop credit scoring models to predict the likelihood of default based on its variables. Behaviour analysis is already been used and it is generating highly accurate predictive models around the world. However, this kind of data should never be used alone, as the only source for predicting likelihood for default. The main reason is that those models are better to describe willingness to repay, and as already discussed, a borrower might be willing to repay his debt timely, but still not be economically able to do so. Additionally, behaviour analysis can be discriminatory when used alone, due to the way data is trained and patterns are discovered.

By combining traditional historical financial data with alternative data regarding behaviour analysis in a credit scoring using both borrower-specific data (qualitative) and quantitative data, it will be possible to generate a predictive model that is both capable of determine ability and willingness to pay, giving the lender a 360° profile of the borrower. The most varied the data, the more angles of the borrower will be seen. This means that the lender will be able to distinguish among borrowers, those who are considered high-quality, namely, those whose reputational capital is capable of deterring opportunistic behaviour, minimizing therefore, information asymmetry.

Lastly, the combined model will be automated, resulting in faster decisions at a lower cost. As data keeps being stored, models can be continuously improved and newer technologies will help further improve the results. Lenders will be able to only choose high-quality borrowers, which in turn will result in cheaper loans to borrowers.

Chapter 6 – The challenges for implementing data analysis to Danish P2P platforms?

6.1 Introduction

In chapter 3, this thesis has presented the European Banking Authority's (EBA) suggestions for the regulation of P2P platforms, which included a topic regarding the investigation of risk factors and the establishing of procedures aiming to minimize those risks. The main goal of the EBA's suggestion is to guarantee that P2P platforms will strive to offer the best and most accurate credit assessment possible, so the information asymmetry lenders face would be minimized. Also, chapter 6 has showed, that data analytics can improve the accuracy of credit assessments, by helping the determination of both ability and willingness to repay a loan, which therefore, results in more accurate model predictors of credit risk. Additionally, chapter 6 presented how the two bigger crowdlending platforms in Denmark, Better Rates and Lendino do their credit assessment.

This chapter will discuss how data analysis can, at least viewed from a theoretical perspective, be implemented in Danish P2P platforms, while describing some suggestions for the data needed, and the data sources. This chapter will also discuss about the challenges P2P platforms might face when implementing data analysis and offer some suggestions to mitigate those challenges.

6.2 Fundamentals for implementing data analysis

The implementation of data analysis is done by following some steps as showed by figure 7.1.

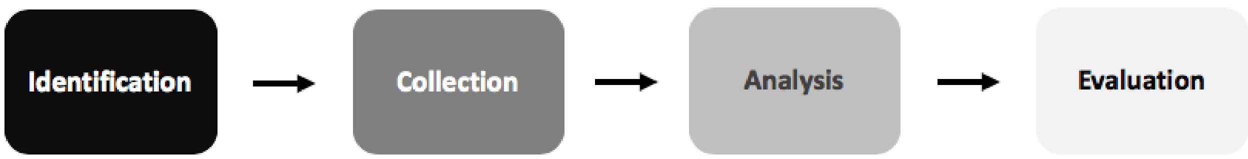

Figure 7.1 – Steps for implementing data analysis

6.2.1 Identification

This first step involves the identification of the objective, the goals, the required data and the data sources. The objective of a P2P platform is to perform the best credit assessment possible, predicting risk of default. In order to do that, it is important to identify the following steps:

1. Identify the goal
2. Identify the required data to achieve this goal
3. Identify what is the source of this data

The goal is determining what are the elements needed in assessing the credit of a borrower, which are determining ability and willingness to repay a loan's principal plus interest. Once the goal is identified, the platform needs to decide, which kind of data will be used in making their credit assessment: financial data and/or non-financial data. Note here that, as already mentioned in the previous chapter, credit assessments using both financial and non-financial data are better. If the platform chooses to use both financial and non-financial data, it would then need to define which non-financial data will be used: is it non-financial data regarding past loans or is it alternative data?

The next step is finding which are the data sources for this data.

- Historical financial data – borrower-specific data and quantitative data regarding previous loans
- Alternative data – online data collected by the platform (clickstream), data regarding the social presence and evaluation of the borrower (social media data), data regarding the borrower's use of phones/mobile phones (mobile data) and so on.

Determining the kind and source of data is only the beginning. The platform has to also assess the availability of this data, present expected outcomes, and delimit the inputs (variables) that are needed to achieve the goals. An example of useful variables to achieve the described goal above, using a combined method can be seen below:

6.2.1.1. Determining ability to repay

In order to determine ability to repay, the following information should be collected:

- Presence of collateral and proportion of collateral value in relation to loan value
- Presence of cash in form of equity or borrower own investment in the project, and its proportion in relation to loan value
- Evaluation of the borrower's company profitability, liquidity, leverage and efficiency

- Evaluation of the owner and top management's debt factor

6.2.1.2. Determining willingness to repay

- Past behaviour – determining whether the owner, the members of the top management and the company is registered as bad-payer in: RKI and Debtor register. Also, determining whether the company is in bankruptcy and it the owner and top management have had prior bankruptcies
- Time stamp of the loan request.
- The borrower's behaviour on the website, with emphasis on, whether the borrower has checked and read the terms and regulations of the loan contract as well as shown interest /curiosity in knowing before accepting the terms, which would be the RRR for the loan.
- Does the company has a fixed phone line, does the company's calls get answered and or returned?
- Determining the company's online social presence – is it registered on Facebook, Twitter, Linkedin, Instagram, YouTube, etc.?
    - Analyse the number of posts, comments, shares, likes and dislikes, as well as the time it takes for a company to answer comments/questions.
    - Analyse the star ratings of the company on those platforms as well as on other review platforms such as Trustpilot, Yelp, etc.

6.2.2 Collection

As already discussed, predictive analysis requires large volumes of data, with a certain number of defaults to show the trends and insights the data contain. Collection of different kinds of data coming from multiple sources should require a unitary approach to data. This is an issue because data islands are not all designed using only one specific key across systems. That means that issues regarding different format or categorization of data might arise. Additionally, it is not only the way the data is stored that should be taken in consideration, but also how the data is structured: some data will be presented in the form of tables (datasets from previous loans) while other data will be presented in a much less structured way (Facebook comments).

6.2.3 Analysis

Once all the data is acquired, the next step is to clean the data from all variables that should not be knows at the moment of the underwriting of a loan. After the data is cleaned from all leakage, the analytical process starts, trying to find patterns that indicate risk of default.

Based on quantitative data from past loans (historical financial data) and data regarding online behaviour as well as defaulted companies, patterns of the following data should be found:

- Find patterns from past loans that are indicative of default in all the dataset regarding variables such as age of company, age of borrower, gender of borrower, zip code of company, purpose of the loan, etc.
- Find patterns from past loans indicating the best range of financial ratios, collateral and cash invested in proportion to the loaned value
- Find patterns from defaulted companies and their social media data regarding posts, comments, shares, likes and dislikes, as well as the time it takes for a company to answer comments/questions, star ratings and reviews
- Find patterns regarding time of request for loan, and eventually also patterns referring to the company's choice of phone (fixed line or not), as well as borrower's choice of phone (IOS vs. Android), and subscription

6.2.4 Evaluation

When the patterns are found, predictive models can be created, and the borrower-specific data must be evaluated in comparison with those predictive models to result in a credit scoring of the borrower.

6.2.5 Implementation

After the previous steps are done and the best predictive model is chosen, it is permanently incorporated by the P2P platform. It would, however, be beneficial to keep searching for better models, as more data is added.

6.3 The opportunities regarding implementation

There are many opportunities for the implementation of data analysis in Danish P2P platforms. First, P2P are online platforms, making it easier for them to collect some types of alternative data such as clickstream data. Also, the amount of publicly available data on the web is in exponential growth. Additionally, most people on the western world has some kind of social presence, and there is a larger number of individuals, who share their data openly. Companies are also very aware of the need of keeping social presence. Furthermore, there are many different sites offering reviews and people are sharing valuable data regarding their experience with this or that company. The way companies respond to those reviews is also a great indicative of the company's customer-relationship management (CRM). Lastly, the implementation of credit scoring using both historical and alternative data would result in an automated process, reaching faster and more accurate decisions at a potentially lower cost.

6.4 The challenges regarding implementation

However, the implementation is not without challenges. There are three main challenges, P2P platforms face when introducing data analysis to their credit assessment verification. The first challenge regards creating enough supply and demand for crowdlending, the second regards the volume of data needed to implement quantitative analysis, and the third refers to regulatory issues. In a way, both the first and second issues are interconnected, as a P2P platform would need to have enough supply and demand for their services in order to acquire the needed volume of data.

6.4.1 Supply and demand

It has been presented in previous chapters, that there is a demand for alternative forms of financing, though not so many SMEs are making use of crowdlending. There can be many reasons to explain this, but one explanation could be that SMEs might be unaware of the real possibilities of crowdlending.
The curious paradox about crowdlending is that in order to attract borrowers, it need to have prospective lenders, and in order to attract lenders, it need to have interesting prospective borrowers, so there is a very important balance at hand. One of the hypotheses of this thesis is that, by minimizing information asymmetry between lenders and borrowers, P2P platforms would become a more interesting investing

venue for lenders, as they offer a significant higher rate of return[65] than other investing options, as savings accounts for example. But P2P platforms will only attract lenders, if there is someone interested in borrowing. In the beginning of May (2019), Lendino[66] had only one project available for lenders to invest in. Another interesting factor is that most of the already closed projects managed to acquire the desired funds in a question of minutes. An example is a SME requesting a loan for 150.000 DKK for the purpose of making some kind of extension construction to vacation apartments. This loan was financed by 74 lenders in only 7 minutes. Better rates, on the other hand, did not have any project available for loans in that same period.

This shows that Lendino has already enough lenders for the number of projects it offer. The platform can, however, improve their credit assessment, that currently is listed as simple guidance, and attract more lenders. Another approach can be marketing crowdlending, so more companies become aware of this financing possibility. Lastly, Lendino could start using clickstream analytics in order to increase conversion of visitors to lenders/borrowers.

A clear evidence that improved methods of credit assessment is efficient in attracting more lenders and subsequently borrowers to P2P platforms is that Better rates, a company that offers a credit assessment using credit scoring (and hence, data analytics), has in a period of only 3 years of existence financed 9122 projects, and acquired more than 9.000 members[67], while Lendino that has been in the market for 6 years only financed 358 projects, though 1.848 lenders. Also, the amount of capital (in DKK) financed for Better rates is 243% higher than the amount of capital financed by Lendino.

So, improving the quality of credit assessment can help attract more lenders to a platform. The question is why P2P platforms are not attracting more borrowers, and most importantly, if they are attracting only low-quality prospective borrowers, that do not pass the credit assessment, and therefore, never convert to real borrowers.

6.4.2 Volume

Another main issue faced by P2P platforms regards the presence or absence of volume of data. In previous chapters, it was discussed that the volume of stored data is immense. However, one thing is that there is data store, another is whether this data is available.

[65] Crowdlending platforms offer a rate of return between 6 and 15%

[66] www.lendino.dk/projects

[67] Better Rates have 9703 members. This does not mean that all those members are investing but considering the amount of projects and the volume of capital financed, it is expected that Better rates has a significant higher amount of lenders than Lendino. There is no information available at the site regarding to the conversion rate of lenders to borrowers. Also, there is not information on Lendino regarding the number of members, just the amount of lenders.

As already stated, creating quantitative models require a large volume of data, with at least 1.000 defaults. This shows, that volume can only be created with high supply and demand. A platform, like Lendino, would be unable to create a credit scoring using quantitative data, because there is not enough data to create a model. Lendino has 358 loans, out of which 3,8% suffered delinquency and 3,2% defaulted. The volume of defaults is irrelevant for model creation. Better rates have a higher volume of loans (over 9.000) but it is dubious that they will also have at least 1.000 defaults to create a model. Although Better rates use credit scoring, they use data from third-party companies to create their scoring (data from Experian Segmentation Data).

A P2P platform will only be able to create their own models, after collecting enough data to create them. They should, therefore, try and store as much data as possible from different angles of the borrower, so the models created in the future will also show different angles regarding to risk of default.

This leads to the question: should P2P platforms do their credit assessment in-house or outsource it? The answer seems to be simple, if the platform has enough data it is possible to start developing their own predictive models. However, in the absence of data, a platform has only two options: either credit assess using traditional method without the help of data analytics (which is the case of Lendino), or outsource the analytics process to a third-party company (which is the case of Better rates). This outsourcing can be full or partial: full outsourcing means that the platform will not have in-house credit assessment, and will only collect data, while partial outsourcing means that the platform will only outsource the data analysis to a third-party company.

This outsourcing can be done using renowned international credit bureaus as Experian or by using the services of companies such as Noitso (see Appendix 02) and Big Data Scoring (see Appendix 01).

If a P2P desires to implement credit scoring to their in-house assessment, but lacks volume of data for doing it, this implementation can be done following a two-phase process: the first phase is by enhancing the in-house traditional credit assessment with the services of a third-party company, while collecting as much data as possible, from as many sources as possible. The second phase can only start, when the data collected has the right volume to develop predictive models.

For credit scoring based on social media data, it is also possible to research defaulted companies and their online presence, rating and behaviour for the past 6 months before default, to try and find patterns between their social data and risk of default. This is conditional to the access of the required data. If the platform has no access to this data, it will need to build their own database, and only create predictive models when the volume is adequate to do so.

Clickstream data can be used immediately, at least regarding some known indicators, that give insight over credit risk. However, with volume, this data will be able to show many other unknown indicators and show different patterns, that might further enhance the credit risk prediction ability of this kind of data.

6.4.3 Regulatory issues

Another main issue regarding the implementation of data analysis refers to regulatory barriers, such as GDPR, because they limit the access and storage of data and could potentially difficult or could even prevent, the use of data analysis.

The main issue does not regard anonymization of data, although, as already stated in Chapter 6, if the data is too anonymized, the predictive model might lose valuable insights and become less efficient. It regards the strict need for consent[68].

GDPR requires that unless a legitimate interest[69] is present, strict consent from the borrower is required for profiling. Credit scoring is basically profiling[70], and the concept of "legitimate interest" is not so clear. Preamble 47 of the law states: "Such legitimate interest could exist for example where there is a relevant and appropriate relationship between the data subject and the controller in situations such as where the data subject is a client or in the service of the controller", so at first glimpse, it appears that P2P platforms would have a legitimate interest. However, this could be subjected to a different interpretation, as the same preamble continues: "At any rate the existence of a legitimate interest would need careful assessment including whether a data subject can reasonably expect at the time and in the context of the collection of the personal data that processing for that purpose may take place. The interests and fundamental rights of the data subject could in particular override the interest of the data controller where personal data are processed in circumstances where data subjects do not reasonably expect further processing."

Also, preamble 39 states that: "The principle of transparency requires that any information and communication relating to the processing of those personal data be easily accessible and easy to understand, and that clear and plain language be used." Once again, the text of law leaves room for interpretation regarding how much detail would be necessary to make the information easy to be understood. In Chapter 5, it was clearly demonstrated, that the value regarding behaviour analysis lie on the fact that the data used for it is objective by nature. It was also discussed, that if the borrower knew the variables regarding behaviour data used during a credit scoring, the data could lose its objectivity, because borrowers might change their online behaviour to fit within accepted parameters.

[68] GDPR article 4, 11
[69] GDPR article 6, 1f
[70] GDPR article 4, 4

Another issue regards the prohibition against discriminatory profiling. As already presented, behaviour analysis can be discriminatory, because of how the data is processed. Actually, it was presented in Chapter 5 that credit scoring should not be used alone, without an assessment on historical financial data first because it will lack important borrower-specific data, but also due to its possible discriminatory results. The question is that some of those discriminatory patterns might have valid insights. The credit scoring should not, however, be solely decided by them, so other parameters should be also taken in consideration to guarantee that no discrimination occur.

Additionally, GDPR is only one law protecting individual privacy. Other regulations might appear in the future, further affecting data analytics negatively.

So, the solution to this problem would have to be a change of regulation with less emphasis on individual rights protection, but the main question is, whether this is desirable. GDPR protects every individual against some of the abuses that are seen happening in other countries, regarding misuse of personal information to shut dissent. In this case, the only solution would be to improve the process, so all data is completely anonymized and still be relevant and reliable in creating predictive models. Whether this is possible, is yet to be seen.

6.5 Search costs

There are costs involved in this service, and it is most likely that the volume of credit scoring will influence the search cost. Both Big Data Scoring and Noitso claim to have a low search cost. In a case-study[71] from 2014, Big Data Scoring claimed that their service resulted in significantly gains to the client, because he could stop using the services offered by a renowned, but expensive, credit bureau. Noitso, on the other hand, claims that their credit assessment for a private lender is under 100 DKK, and this value would be presumably even lower for P2P platforms, requiring a higher volume of credit scoring. Also, Noitso provides the whole credit scoring, which would reduce further the costs of P2P platforms.

Also, credit scoring using a combined model will result in better predictive models, increasing the efficiency of the credit assessment and reducing costs involved in delinquency and default. Moreover, as the process is done electronically, cost is reduced. This requires, however, electronic access to data to make the process faster and cheaper. Some public data (i.e. Bilbogen, BBR) is electronic available, while some other have limited access, as for example eSkatData, that only grants access to banks and other

---

[71] https://web.archive.org/web/20151022010402/http:/www.bigdatascoring.com/2014/11/case-study-about-a-central-european-lender/index.html

credit providers. As already mentioned, credit analysts can access a borrower's tax data through TastSelv, but this method is manually done and cannot be automated by computers. So, having fully access to all important electronic databases would help drive search costs down.

6.6 Summary

In conclusion, P2P platforms can implement data analysis by identifying their goal, the required data, and the data source. After those first steps, data is collected, cleaned and analysed, in order to create predictive models, that will be used in evaluating borrower-specific data, and result in a credit scoring for the borrower. When this predictive model is chosen, it can be incorporated by the P2P platform.

P2P have many opportunities regarding the implementation of data analysis for credit scoring purposes. As those platforms are online, the task of collecting data is simplified. Also, there is a vast amount of publicly available data online, as most people and companies have some kind of social online presence. Online platforms offering star-ratings and reviews offer a great input about client's experience with the company, while also offering a better understanding of the company's CRM. By implementing a combined model, the automated process would result in more accurate predictive models, allowing a fast decision at a lower cost. However, the process can be made difficult by three challenging-factors: supply and demand creation (lenders and borrowers), volume acquisition and regulatory barriers.

P2P platforms require both lenders and borrowers, as lenders will attract borrowers, and borrowers will attract lenders. This thesis hypothesis is, that by minimizing information asymmetry between lenders and borrowers, more lenders will be attracted to crowdlending, as the rate of return offered by most platforms is attractive. This hypothesis is, at least apparently, confirmed in reality by Danish Platforms, as a newer platform with a stronger credit assessment method was capable of attracting a significant higher number of loans than a platform with twice the amount of years in the market, that only offers financial guidance to its lenders.

Moreover, projects seem to be financed without issue, once the borrower is accepted by the platform. The main issue the observed P2P platforms seem to have is that they are not attracting more borrowers. Whether this is caused because borrowers are unaware of this option or because only low-quality borrowers, that are not accepted by platforms, are attracted is yet to be evaluated. Anyhow, if platforms want to attract high-quality borrowers, it will beneficial to market their services as well as use clickstream data to enhance conversion of visitors to consumers.

The second challenging factor regards volume: even though volume might exist, it might not be available to the P2P platform. Creating volume requires first that supply and demand of loans is created. However, acquiring the correct volume of data can take time. In this case, a P2P platform has two options: wait until volume is sufficient, while storing as much data as possible, or outsource credit scoring to a third-party company. In any way, implementing data analysis, in the situation where volume is not present, would be done in two-stages: phase 1 where data is stored, and either the P2P does not make use of data analysis or outsource it to third-party companies and phase 2 in which the platform can develop its own models, based on their own data.

The third main challenge P2P's face when implementing data analysis refers to regulatory barriers. Currently the core concern is GDPR, however new regulation concerning privacy rights may be approved in the future, making it very difficult, if not even impossible, to use data analysis for credit scoring. Requirement of strict consent focused on transparency could render behaviour analysis useless, as borrowers would be informed of the important variables for credit scoring. This would render behaviour data useless, due to loss of objectivity.

Lastly, credit scoring should be cost-effective. The best way of achieving this is by having a fully automated system in place, which cannot be possible, if some data is only available electronic for some entities, which is the case of, for example eSkatData.

Solving the first two challenging-factors would require time and effort, but it is absolutely possible. Solving the third issue is more complicated. Although regulations like GDPR might seem to be overly restrictive, it might not be desirable to change them. The change should, in this case, be done in the way data is processed, so completely anonymized data would be still maintaining its reliability, relevance and insights, which might not be possible to achieve.

Chapter 7 – Conclusion

This thesis's aim is to investigate the opportunities and barriers for using data analysis to minimize information asymmetry in a crowdlending context. The reason behind this investigation is the fact that many Danish SMEs struggle to finance their needs, and crowdlending appears to be an alternative financing option, that could help those SMEs achieve their goals. However, investments are risky, and the information asymmetry between borrowers and lenders is significantly enhanced, when the borrower is an individual and not a bank or credit institution. So, if P2P platforms plan on fulfilling the need of SMEs, it has to be able to attract lenders, by offering them ways to minimize this information asymmetry.

Problem statement questions 1 (a, b and c) and 2

SMEs are the backbone of every country's economy, and that is also the case for Denmark. The Danish SMEs are the major source of jobs within the private sector and contribute significantly to the country's GDP. However, SMEs face a credit rationing, that has been further exacerbated by the economic crisis of 2008. From 2009 to now, the bank financing of SMEs has improved, but there are still many SMES struggling to acquiring the needed capital to finance their activities, and maintain competitiveness, growth and innovation. When banks refuse credit, this opens opportunities for alternative lenders, and crowdlending is a very used option all around the world. In Denmark, crowdlending is at its first steps, but it offers great opportunities for both lenders and borrowers. Crowdlending is however risky, not only due the inherent risk of credit markets, but because of P2P transactions are tainted by both information asymmetry and operational risks.

Problem statement question 3

Lenders do not have information nor control about, whether the borrower is able and willing to repay the loan on time. This makes is very difficult for lenders to distinguish between a high-quality and a low-quality borrower, and because lenders are rational player, they protect themselves by treating borrowers as low-quality borrowers, and all projects are perceived as having a high risk, which results in a high RRR. This is also called adverse selection.

Adverse selection has a rather detrimental effect to both lenders and borrowers. It can result, for example in credit rationing, if the lender is a bank or other kind of credit institution. This can be explained as follows: because banks are rational players, information asymmetry should result in higher RRR. However, banks are aware that the higher the interest, the higher the risk, as the interest be a significant part of the project's expected profit, and if the expected profit is not achieved, the borrower might default. If increasing the RRR does not reduce risk, credit is rationed and the capital, banks have at their availability, will be invested otherwise.

Also, treating all borrowers as low-quality and all projects as high-risk is discriminatory against high-quality borrowers, because their projects should require a lower RRR. In some cases, the high RRR will result in such low profit, that those borrowers will desist, leaving only low-quality borrowers on the market. Moreover, low-quality borrowers are also rational players, and when faced with high interest rates, they will choose riskier projects with high expected return, so potentially maximize their profits. And because they are aware, that higher risk results in higher interest rate, they might behave opportunistically and disguise the true nature of their projects in order to acquire cheaper credit, which is also known as moral hazard. Lastly, because low-quality borrowers want to maximize their profit, they have an incentive to announce lower-than-actual earnings, because, due to the information asymmetry, lenders would have to incur in monitoring costs to verify their earnings.

There are ways, however, to minimize asymmetric information. This can be done through signalling; in which case the borrower signals his creditworthiness by increasing their liability via an economic measure in the form of either collateral or internal funds (cash). Also, the borrower's reputational capital might deter opportunistic behaviour, because tainting his reputation would close future door for other business and investments.

Problem statement question 4

When deciding over an investment, lenders should only finance borrowers who both have ability and willingness to repay the loan on time. However, lenders are unable of distinguishing between high-quality and the low-quality borrowers, and need to screen them, through a credit assessment, in order to determine their ability and willingness to repay.
Traditionally, credit assessment has been done by analysts using borrower-specific historical financial data. Although this method is efficient in determining ability to repay, it could be improved regarding the determination of the borrower's willingness to repay.

Improving this traditional credit assessment is possible nowadays, thanks to new technologies, that use machine power to find patterns or trends that can predict default. Credit scoring using quantitative historical financial data is the most common used method of credit assessment using data analysis. Using statistical models based on quantitative historical financial data, credit scoring can spot patterns and trends from past behaviour and use those to predict future behaviour with high-accuracy. Although credit scoring cannot be used to determine, whether company A is going to default, it can predict the likelihood of company A defaulting. Hence, the determination of both ability and willingness to repay is improved. However, the determination of willingness to repay could be further improved, if the method also used alternative data, allowing for a behaviour analysis of the borrower.

Information asymmetry can be, as already said, mitigated by reputational capital, and therefore, understanding the borrower is fundamental to the determination of his character. By adding alternative data to the process of credit scoring (combined model), the predictive model would take in consideration many other angles from the borrowers, allowing for a better view of whom he is. This will significantly improve the determination of willingness to repay, and only those borrowers, whose reputational capital is present to deter opportunism will be accepted. Lenders will no longer be unable to distinguish between high-quality and low-quality borrowers, and therefore, information asymmetry is mitigated.
Also, combining historical financial data with alternative data for credit scoring helps minimize some side effects, that behaviour analysis can have, namely the creation of prediction models that can be discriminatory.

Finally, this combined model is fully automated, making the process more objective, while delivering faster decisions at a lower cost. As data keeps being stored, models can be continuously improved and newer technologies will help further improve the results. However, it is important that the process is fully automated, so access to different data sources should be granted to P2P platforms.

Problem statement question 5

This combined model can be implemented by P2P platforms following the following steps: identifying the goal of the model, the data and the data sources, collecting the data, cleaning and analysing the data, creating predictive models, evaluating and choosing the most accurate model. This best model is to be incorporated by the platform.

There are many opportunities for the implementation of data analysis in Danish P2P platforms. Those platforms operate online, simplifying the task of collecting data. Additionally, platforms has nowadays access to a vast amount of online publicly available data, because individuals and companies are aware of the necessity of having social online presence. Moreover, reviewing sites are an excellent source of information regarding the experience of the company's clients as well as the company's CRM. Also, due to automation, and the high-accuracy of predictive models, decisions are made faster and at a lower cost. Implementation is, however, not free from complications. P2P platforms face three main challenges when implementing data analysis to their credit assessment: difficulty attracting supply of capital and demand of loans, lack of relevant volume of data and regulatory barriers.

Supply and demand are essential for P2P platforms, not only because lenders are needed to attract borrowers and borrowers are needed to attract lenders, but also because it through a high supply and demand that volume of data is created. The hypothesis of this thesis hypothesis is, that to increase the supply of capital, P2P platforms must minimize the information asymmetry between lenders and borrowers. Observing two of the biggest Danish Crowdfunding platforms, this hypothesis seems to be confirmed. Best rates provides a credit scoring based on historical financial data (data analysis), which has attracted a significant higher number of both loans and lenders than what has been achieved by Lendino, a platform with twice the age on the market, but that does not use credit scoring and only offers a credit assessment as guidance to lenders.

Demand for loans, however, could be improved. In the beginning of May, there was only one project available at Lendino, and none available at Best rates. Loans seem to be financed rather fast, so lenders are eager to invest. P2P platforms need to attract more borrowers. It is difficult to assess, why so few borrowers are using crowdlending, when there is a clearly financing gap in Denmark. Borrowers could be unaware of this alternative form of financing, or perhaps, crowdlending is only attracting low-quality borrowers, that gets barred by the credit assessment. Whatever the reason is, P2P platforms should invest in market campaigns to inform prospective borrowers of this option, as well as use clickstream data to find out the reasons that explain why visitants do not apply for loans and improve this conversion rate from visitor to consumers.

Supply and demand are essential for the creation of volume of data. Modelling of data requires high amounts of data, and although this data might be stored, it might not be available to the P2P platform. Platforms need to create their own volume, which takes time. Meanwhile, P2P platforms can either use traditional methods of credit scoring or outsource data analysis to a third-party company. Data should, though, be continuously scored, so the P2P platform can use it in the future. Because of this volume issue, implementation of data analysis in the platform's credit assessment must occur in two phases: in the first phase data is stored for future use, while the platform either refrain from using data analytics or outsource the service, and in the second phase, when the volume of data is sufficient to use modelling of data, P2P platforms can develop in-house their predictive models.

Regulatory barriers also challenge the implementation of data analysis in Danish P2P. Regulations such as the GDPR, can make it difficult to store data, which would affect the quality of predictive models, and eventually totally prevent the use of data analysis. Also, the requirement in the GDPR regarding strict consent could render behaviour analysis useless, as most of the data could cease being objective, due to risk of manipulation from borrowers.

Finally, the main goal is to achieve a cost-effective credit scoring, and the best way to do this is by fully automating the process. This requires expanding the access of some data, that nowadays is only electronic available to some entities to P2P platforms.

While solving issues regarding supply and demand and volume of data requires only time and effort, solving regulatory issues demand a change in legislation, which is a much larger and complex task to embrace. Also, there is the question of whether changing regulations like GDPR is even desirable. While GDPR might appear restrictive, it protects the privacy rights of individuals. So perhaps, instead of removing individual protection, it would be better to change the way data analysis is done, so completely anonymized data can still be useful and reliable, resulting in predictive models that do no lose any important insight. Whether this is possible to achieve, is yet to be seen.

Chapter 8 – Perspectivation

8.1 Introduction

It has been presented in chapter 6, that P2P must attract a higher supply and demand in order to be able to not only minimize the financial gap for SMEs, but also increase the volume of data, that is required for modelling of data. This chapter will present some suggestions, that might help P2P platforms to attract more lenders and borrowers.

8.2 Providing other forms of crowdfunding

While this thesis was written based on crowdlending (loan-based crowdfunding) from consumer lenders to business (C2B), there are other forms of crowdlending that could be also implemented, increasing the number of potential borrowers in the platform. Platforms could start offering crowdfunding for start-ups, also known as equity-crowdfunding, and/or they allow individual to obtain private loans through their market place. Those two options are further discussed below.

8.2.1 Equity-crowdfunding

Equity-crowdfunding is characterized by the sale of securities (shares, convertible notes, debt, and revenue shares) belonging to a private company, that is not listed on stock exchanges, to raise capital[72]. Many start-ups need financing, and until some time ago, venture capital was the solution. However, equity-crowdfunding be rather beneficial for the company's shareholders.

When a start-up company enters the market place, it can test the attractability of their project beforehand. This means, that by trying to acquire financing, companies can find out, whether their product will become a success or not, based on how many lenders show interest in lending to the product. Since projects are only financed if the required value is achieved within a certain period, projects that do not attract lenders should be scrapped, as the project will most likely not become a success in the real world.

72 https://www.forbes.com/sites/howardmarks/2018/12/19/what-is-equity-crowdfunding/

Another benefit equity-crowdfunding offer is the easy access to capital. When companies depend on venture capitalists, the competition is enormous, and venture capitalists, that are investing high amounts of capital in a project are choosey and overdemanding. Lenders on a P2P marketplace are not necessarily as choosey, as they are investing a significantly smaller amount of capital individually in the project. Also, since lenders only purchase a small share or right, shareholders will remain having the majority of shares, and therefore, they can maintain their control over the company.
Those benefits make equity-crowdfunding particularly attractive to start-ups. Lenders will also be attracted by this kind of crowdfunding, as they will own small shares or rights from companies, in which projects they believe, and hope will become huge successes.

8.2.2 Personal loans

Many P2P platforms around the world are already tapping in this opportunity called personal loans, and there is a good reason for that. The market for personal loans is immense, and here in Denmark, quick-loans are marketed everywhere.
However, quick-loans are more harming that good, as the average investment cost (ÅOP) is extremely high, sometimes reaching values over 800% a year. Despite this high cost, borrowers still take quick-loans to finance some unexpected needs or help maintain a budget in order. Taking quick-loans can, though, backfire by causing havoc in a person's economy, so alternative sources of financing, that could offer significantly lower investment costs would be extraordinarily attractive to this group of borrowers. Lenders would also be attracted by this model, because as risk is high, the RRR lenders can expect is also high. This means that less risky-averse lenders would be interested in this form of investment. Also, the possibility of investing in pools of borrowers, with same credit risk, could improve the safeness of the investment, decreasing the risk of credit loss for borrowers.
Both lenders and borrowers, are therefore, benefited, with the adding effect that there are a significant higher number of individuals than companies in Denmark. Also, this data could be used to analyse the creditworthiness of SMEs, because as already presented in previous chapters, the person of the owner and or manager of a SME can merge with the company.

8.3 Cross-border platforms

Another way of increasing the supply and demand for loans, and therefore increase the volume of stored data is by creating cross-border platforms, which means P2P platforms with access to many different markets in different countries. The platforms would become available to a significant higher number of individuals, which most likely would result in a higher number of loans, lenders and borrowers. A European cross-border P2P[73] platform would also have the benefit of having a shared regulation cross-borders, when the regulation regarding crowdfunding in Europe is finally approved and implemented in member states.

8.4 Summary

The challenge regarding supply and demand for loans and access to volume of data can be minimized faster, if P2P platforms expand their offering of crowdfunding from crowdlending to equity crowdfunding and personal loans. By offering equity-crowdfunding, platforms can attract start-ups that will be fully benefited by this form of financing, as they can test the market interest of their project before any real investment is done. Also, because lenders invest individually a smaller amount, they purchase smaller shares and rights, and are less demanding than venture capitalists. This means, that the company's shareholders can hold the majority of the shares, and therefore, keep control over the company.

Additionally, by offering personal loans, platforms can tap into the market of quick loans, and attract borrowers by offering cheaper credit, while attracting lender with the promise of higher interest rates. Personal loans can give an excellent source of data for SME financing, because in SMEs, the person of the owner or key manager merges with the person of the company, so understanding the person of the owner/manager, can give excellent indications of whether a loan will default.

Finally, volume can be created through a cross-border European platform, that would increase the scope of the platform, while granting regulatory tranquility, because there is a regulation for crowdfunding on the making, which will render all legislation within the EU area compatible.

[73] https://ec.europa.eu/info/tender/identifying-market-and-regulatory-obstacles-cross-border-development-crowdfunding-eu_en

Bibliography

1. **Akerlof, G. A.**, 1970. *The market for "lemons": quality uncertainty and the market mechanism*. The quaterly journal of Economics, Volume. 84, n.° 4, August, pp. 488-500.
2. **Altman, E., Sabato, G. & Wilson, N**., 2010. *The value of non-financial information in small and medium-sized enterprise risk management.* The Journal of Credit Risk, Volume 6, n.° 2, pp. 1-33.
3. **Andersen, I**, 2013. *Den skinbarlige virkelighed - vidensproduktion i samfundsvidenskaberne.* 5. udgave: Samfundslitteratur.
4. **Bebczuk, R. N.**, 2003. *Asymmetric information in financial markets: introduction and applications.* Cambridge University Press, UK and US.
5. **Berger, A. N., Cowan, A. M. & Frame, S.**, 2011. *The surprising use of credit scoring in small business lending by community banks and the attendant effects of credit availability, risk and profitability*. Journal of Financial Services Research, Volume 39, issue 1-2, pp. 1-17.
6. **Bester, H.**, 1985. *Screening vs. Rationing in Credit Markets with Imperfect Information*. The American Economic Review, Volume 75, n.° 4, pp. 850-855.
7. **Caire, D. & Kossmann, R.**, 2003. *Credit Scoring: Is it right for your bank.* Material prepared for Bannock Consulting on a Technical Assistance engagement funded by the European Bank for Reconstruction and Development and the EU.
8. **Ciampi, F. & Gordini, N.**, 2013. *Small enterprise default prediction modelling through artificial neural networks: An empirical analysis of Italian small enterprises.* Journal of Small Business Management, Volume 51, issue 1, pp. 23-45.
9. **Duan, H., Han, X. & Yang, H.**, 2009. *An analysis of causes for SMEs Financing Difficulty.* Internation Journal of Business and Management, Volume 4, n.° 6, pp. 1-3.
10. **Earley, C. E.**, 2015. *Data analytics in auditing: Opportunities and challenges.* Business Horizons - Volume 58, issue 5, pp. 493-500.
11. **Fedders, J., Steffensen, H. & Lassen, K. T.**, 2017. *Årsrapport efter internationale regnskabsstandarder - fra dansk praksis til IFRS.* 5. udgave: Karnov Group.
12. **Feldman, R. J.**, 1997. *Small business loans, small banks and big change in technology called credit scoring.* The region, September, September, pp. 19-24.
13. **Fenwick, M., McCahery, J. A. & Vermeulen, E. P. M.**, 2017. *Fintech and the Financing of Entrepreneurs: From Crowdfunding to Marketplace Lending. ECGI Working Paper Series in Law*, September, pp. 2-53.

14. **Fraunhofer, M. K.**, 2009. *Online Peer-to-Peer lending: A Lender's perspective*. SSRN Electronich Journal, July, pp. 1-5.
15. **Galloway, I.**, 2009/2010. Peer-to-peer lending and Comunity Development Finance. *Community Investments,* Volume 21, Issue 3, Winter, Volume 21, issue 3, pp. 18-39.
16. **Ghatak, M. & Guinnane, T. W.**, 1999. *The economics of lending with joint liability: theory and practice.* s.l., Journal of Developement Economics, pp. 195-228.
17. **Hsieh, N. C.**, 2004. *An integrated data mining and behavioral scoring model for analyzing bank customer*. Expert Systems with Applications, Volume 27, issue 4, November, pp. 623-633.
18. **Hurley, M. & Adebayo, J.**, 2017. *Credit Scoring in the Era of Big Data*. Yale Journal of Law and Technology, Volume 18, issue 1, pp. 3 -70.
19. **Janda, K.**, 2006. *Lender and borrower as principal and agent.* IES Working paper 24 - *IES FSV*.
20. **Kwok, S. H., Lang, K. R. & Tam, K. Y.**, 2010. *Peer-to-peer Tecnology Business and Service models: Risk and Opportunities.* Electronic Markets Vol.12, No. 3, pp. 1-9.
21. **Maier, E.**, 2016. *Supply and demand on crowdlending platforms: connecting small and medium-sized enterprise borrowers and consumer investors.* Journal of Retailing and Consumer Service, Volume 33, pp. 141-153.
22. **Mester, L. J.**, 1997. *What is the point of credit scoring?* Business Review, 3 September/October , pp. 3-16.
23. **Mills, K. G. & McCarthy, B., 2014.** *The state of small business lending: credit access during the recovery and how technology may change the game*. Working paper 15-004.
24. **OECD,** 2015. *New Approaches to SME and Entrepreneurship Financing: Broadening the Range of Instruments,* s.l.: OECD.
25. **Parker, G. R.**, 2005. *Reputational capital, Opportunism, and self-policing in legislatures*. Public Choice, Volumw 122, n.° 3/4, March, pp. 333-354.
26. **Pedro, J. S., Proserpio, D. & Oliver, N.**, *MobiScore: Towards universal credit scoring from Mobile Phone Data.* Conference Paper, June pp.1-12.
27. **Pokorná, M. & Sponer, M.**, 2016. *Social lending and its risks*. Procedia - Social and Behavioural Science, Volume 220, pp. 330-337.
28. **Serrano-Cinca, C., Gutierrez-Nietto, B. & Lopes-Palacios, L.**, 2015. *Determinants of Default in P2P lending.* PLoS ONE*,* October, pp. 1-22.
29. **Spence, M.**, 1973. *Job market signaling.* The quarterly Journal of Economics*,* Volume 83, issue 3, August, pp. 355-374.

30. **Stiglitz, J. & Weiss, A.**, 1981. *Credit rationing in markedt with imperfect information*. The American economic review, Volume 71, nr° 3, pp. 393-410.
31. **Tan, T. & Phan, T.**, 2016. *Social media-driven credit scoring: the predictive value of social structures*, SSRN Electronic Journal, January, pp. 1-11.
32. **Volk, M.**, 2012. *Estimating probability of default and comparing it to credit rating classifications by banks*. Economic Business Review, Volume 14, nr.° 4, pp. 299 - 320.
33. **Yang, Z., Zhang, Y. & Guo, B. D. Y.**, 2018. *DeepCredit: Exploiting user clickstream for loan risk prediction in P2P Lending*. Association for the Advancement of Artificial Intelligence (AAAI) Publication.
34. **Zhang, Y., Jia H., Diao, Y., Hai, M. & Haifeng, L.**, 2016. *Research on credit scoring by fusing social media information in online peer-to-peer lending*. Procedia Computer Science, Volume 91, pp. 168-174.

Appendix 1: Big Data Scoring

Summary of interview with Mr. Erki Kert, CEO of Big Data Scoring made via Skype on the 28th of February 2019.

Big Data Scoring is a cloud-based credit scoring company located in UK with offices in the USA, Chile, Indonesia, Finland and Poland. They provide credit scoring for banks, telecoms and consumer lenders. Their purpose is not only to improve in-house credit scoring, but also help those individual without credit like immigrants and young people to acquire loans. Their solution is integrated via an API to the lenders platform, so the credit scoring is done in-house, combining alternative data to the lenders internal data. Their solution improves the in-house scoring between 5 to 15% as soon as installed and tested. After the solution is used for 3 to 6 months, the level of accuracy is improved to up to 30%.
Big Data Scoring collects vast amount data from multiple sources: web search results, behaviour tracking (clickstream), device technical details, mobile app data. Around 5.000 data points are collected from each borrower.

Some of the parameters they look for are:

- Neighbourhood of the applicant and geodata
    - Information about distance between parks, hospitals, number of unemployed in the area among many other variables are collected
- The device and/or system used by the applicant
    - How was the access done? Mobile or home computer, IOS or Android, etc.
- The applicant's IP address
    - Are you accessing from where you listed as your address, who is the internet provider.
- The applicant's E-mail address
    - E-mail provider, Has this e-mail been involved in fraud, etc.
- Information publicly available over the internet about the applicant
    - Google, Linkedin, etc

The credit scoring is all done automatically and it does not take more than a few seconds to be made.

Appendix 2: Noitso

Summary of interview with Mr. Mohammed Azzouzi and Mr. Ronni Pedersen, CEO and Lead Senior Data Scientist respectively at Noitso made on the 07th of March 2019.

Noitso is a Danish company located in Copenhagen, that produces scorecards based on statistical models to calculate likelihood of default. Because they bland historical financial data with alternate data sources, their statistical model is considered 100% objective and neutral.

The scorecards are made with a large number of data sources, both internal and external.

- Data entered by the applicant (internal source), metadata collected about the applicant, the applicant's behaviour and performance
- Data from public records and databases (external data) – CPR, CVR, RKI, Debtor Register, Statistics Denmark, BBR, Bisnode, Experian etc.

Figure A1 – Noitso's flow

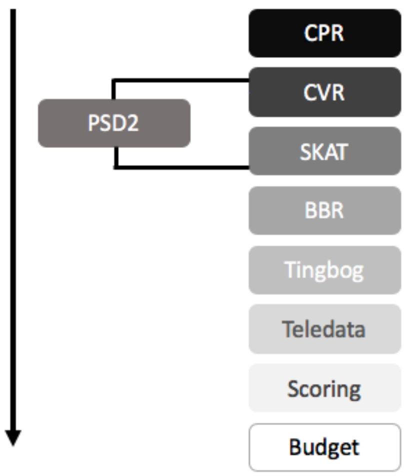

Figure A1 shows the flow Noitso follows when credit scoring SMEs. Whenever an applicant search for a loan, they request both the CPR of the applicant as well as the CVR of the company.

- From the CPR Noitso has access to personal information, such as name, address, age, etc.
- From the CVR they have access to the companies stamdata.
- They search for income statements on SKAT data, as well as information regarding the persons paycheck for the last 3 to 5 months
- BBR (Bygnings- and bolig register) – information regarding any property in Denmark
- Tingbogen – information about the registration of property ownership in Denmark
- Telephone data – information about type of phone, subscription plan, etc.

A scoring is done and Noitso also elaborates a budget, following the common expenses of the applicant. Currently this budget is based on a guess for values, but this is expected to change with PSD2. A scorecard for an individual buyer has a cost under 100 DKK.