

# **The opportunities and challenges of utilizing alternative data in the assessment of creditworthiness in the Finnish consumer finance**

Johannes Liuska

Master's Thesis  
International Business Management  
2021

Johannes Liuska



MASTER'S THESIS	
Arcada	
Degree Programme:	International Business Management
Identification number:	23690
Author:	Johannes Liuska
Title:	The opportunities and challenges of utilizing alternative data in the assessment of creditworthiness in the Finnish consumer finance
Supervisor (Arcada):	Andreas Stenius
Commissioned by:	
<p>Abstract:</p> <p>The modern financial network facilitates profound credit risk assessment. However, currently the Finnish finance industry uses conventional methods to examine the creditworthiness of a customer. Luckily, this will change since the full-file credit information register is in progress and expected to be launched in 2024. Nonetheless, the current trends such as transformation of working life, digitalization, increasing urbanization and consumer over-indebtedness might have major impact in the Finnish society and ultimately on creditworthiness. Moreover, new alternative data sources, increased capacity of big data analytics and open banking concepts could provide even better ways to incorporate smarter methods to assess the consumer creditworthiness. Thus, the financial service providers might need to develop their credit risk management to align with the variables of the modern society and economics. Therefore, this qualitative research investigates the opportunities and challenges of utilizing alternative data in the assessment of creditworthiness in the Finnish finance industry. The thesis reviews the underlying theories of information asymmetry (Stiglitz and Weiss 1981) and data-driven decision making (Provost and Fawcett 2013), and definitions of creditworthiness (Caplinska 2020). It also explores the current credit reporting environment in Finland and European Union to create a clearer picture of the current state of creditworthiness assessment.</p> <p>As the research topic is rather a novelty concept and industry-specific, it was meaningful to seek the information empirically by interviewing experts in the field. Altogether, four interviews were conducted and analyzed by thematic analysis. The findings indicate that the finance industry yearns for stable, unified, and verified data that the full-file credit register might provide. In addition, certain alternative data sets appeared potentially useful such as automated account transaction data. Yet, it seems that our current society is too far from open data economy and there are many constraints to overcome.</p>	
Keywords:	credit risk assessment, alternative credit data, creditworthiness, reduction of asymmetry, open banking, big data.
Number of pages:	52
Language:	English
Date of acceptance:	28.05.2021

# CONTENTS

<b>1</b>	<b>Introduction.....</b>	<b>7</b>
1.1	The research question and purpose of the thesis .....	7
1.2	The scale and delimitations .....	9
1.3	Terminology.....	11
<b>2</b>	<b>Assesment of Creditworthiness .....</b>	<b>11</b>
2.1	Information asymmetry .....	11
2.2	Data-driven decision making with imperfect information .....	12
2.3	Creditworthiness.....	13
2.3.1	<i>Credit scoring</i> .....	14
2.3.2	<i>Data for the assessment of creditworthiness</i> .....	16
2.3.3	<i>The conventional data</i> .....	16
2.3.4	<i>The alternative data</i> .....	17
2.3.5	<i>Review of example companies in alternative data</i> .....	20
2.4	Information Exchange in EU Retail Financial Markets .....	21
2.4.1	<i>The mess of the European credit data sharing</i> .....	22
2.4.2	<i>Credit refence reporting in Finland</i> .....	23
2.4.3	<i>Full-file credit information register in Finland in 2023</i> .....	24
2.5	Regulation .....	26
2.5.1	<i>Consumer Credit Directive</i> .....	26
2.5.2	<i>The General Data Protection Regulation</i> .....	27
2.5.3	<i>Credit Information Act 2007/527</i> .....	28
<b>3</b>	<b>Methodology .....</b>	<b>29</b>
3.1	Sampling.....	30
3.2	Interview .....	30
3.3	Data analysis.....	32
<b>4</b>	<b>Findings.....</b>	<b>33</b>
4.1	Data quality and methods.....	35
4.2	Improvements in the data .....	36
4.2.1	<i>Full-file credit register</i> .....	36
4.2.2	<i>Transactional Data</i> .....	38
4.2.3	<i>Big Data</i> .....	39
<b>5</b>	<b>Discussion .....</b>	<b>40</b>
<b>6</b>	<b>Conclusions .....</b>	<b>44</b>

<b>7 Recommendations for future research .....</b>	<b>45</b>
<b>References .....</b>	<b>46</b>
<b>Appendices .....</b>	<b>53</b>

## **Tables**

Table 1: Impacts of the full-file credit information

Table 2: Description of the interviewees

Table 3: Results of the interviews

Table 4: Summary of the opportunities and challenges of alternative data in the assessment of creditworthiness

Table 5: Synthesis of comprehensive credit information

# 1 INTRODUCTION

In the era of digital finance, all data is credit data (Ramsay & Williams 2019, p. 222). Financial industry is in a transformation and challenged in the cross-pressure due to technology, legislation, and customer expectations. The unknown variables of the recent economic turmoil might have caused even more new challenges for financial service providers to predict the future. Yet, financial institutions have an underutilized potential at their disposal - a substantial catalogue of data.

However, currently in Finland the method of collecting and utilizing the data for the consumer credit risk assessment is concise compared to many EU countries and lags the international practices (Saarenpää 2013, ACCIS 2017 and Kontkanen 2018). Typically, a consumer's creditworthiness is assessed by utilizing conventional information which is obtained from the credit application and the bank's internal customer information database. In addition, an external information is obtained from the credit information register in case of any possible payment defaults that could be called negative credit information. This method stems by the EU directives and Finnish national regulation (European Banking Authority 2015 and Finnish Financial Supervisory Authority 2017).

However, in many other developed countries, the highly advanced financial infrastructure and network facilitate a more profound full-file credit reporting system that, in addition to the negative data, holds also positive credit information such as existing debts and credit information (The World Bank Group 2019). Moreover, new alternative data sources increased capacity of big data analytics and open banking concepts could provide even better ways to incorporate smarter methods to assess the consumer creditworthiness (Siddiqi 2017).

## 1.1 The research question and purpose of the thesis

The thesis studies the opportunities and challenges of utilizing alternative data in credit risk assessment in the Finnish finance industry. The aim is to get familiarized with the current credit reporting landscape on international level, mostly focusing on the EU level

and in the national framework: what kind of data is used and could be used in assessment of creditworthiness?

On national and international level, there has been discussion already for the past 30 years how the credit reporting should be developed (Saarenpää 2013, Kontkanen 2018, World Bank Group 2019). Yet, the potential of full-file credit reporting and alternative credit scoring data in Finland remains unclear. Although certain initial political and law preparation measures are taken concerning the development of a Finnish credit information register that could be launched in 2024, however, the operational model is still under investigation (The Ministry of Justice 2020). Considering the ongoing law preparation project and the international practices, the thesis reviews very current and relevant topic.

The profound credit information has also a larger economic and political aspect. Major financial authorities such as European Central Bank and International Monetary Fund are stressing the need for development of macroprudential instruments to correspond the changes in the market environment. As an example, new credit service providers and their operational models might enable overly high-risk activities; therefore more profound information concerning the consumer indebtedness and factors affecting the creditworthiness are needed (Israël et al., 2017; Kontkanen 2018, p.14). Since consumer credits, together with other household loans, can endanger the financial stability, ultimately, smarter credit risk management practices are one part of the macro-financial stability in relation to credit reporting and overall risk management.

In addition, in 2017, the European Commission has initiated an action plan for a strategy to reinforce the EU single market in the retail financial services that seeks to harness the digitalization and fintech development to improve consumer access to financial services across the EU (EU Commission 2017). In the business perspective, more profound and comprehensive credit risk assessment may create an essential business potential for financial institutions, as they could reduce the information asymmetry (Salamina et al. 2019) and use the comprehensive data for competitive purposes (Minelli 2012, p.39).

The Finnish financial service providers are not immune to the megatrends either. The Finnish credit market has changed, and digitalization is expected to increase this phenomenon (Kontkanen 2018, p. 53). According to Aaltonen and Koskinen (2019), the market share of foreign banks that provide credits to Finnish households has increased rapidly



and amounts already 27% of the total outstanding credit capital. In addition, it is unclear how the conventional credit assessment methods comply with the future requirements as the socioeconomic changes and urbanization continues. For example, the increasing level of urbanization and residential transition from home ownership to rental apartments could lead to a shortage and impairment of conventional collaterals (e.g., house or apartment) which can negatively affect the credit rating of Finnish finance system as a whole (Koste et al. 2020, citing: Brotherus 2017). Other megatrends and societal changes, such as climate change (Association for Financial Markets in Europe 2020), transformation in the forms of working and employment relationships (Finnish Government 2018), and the over-indebtedness of Finnish households (Nykänen 2018) have impact on the financial environment. Especially the over-indebtedness and increased amount of registered payment defaults are considered as major issues in Finnish consumer finance (Kontkanen 2018) which, in the authority perspective, appears to be the priority concern that has initiated the development of a profound credit reporting system.

On these grounds, the financial service providers may need to develop their credit risk management concepts to align with the variables of the modern society and economics. Consequently, if the Finnish finance industry holds on the conventional information and methods, the industry might be in sub-optimal competitive position compared to the international counterparts that have stronger competence of utilizing the cutting-edge analytics and comprehensive data.

## **1.2 The scale and delimitations**

Credit risk assessment has a vast scale of viewpoints. Although regulation is a major part of financial operations, the idea of the thesis is to focus on the data itself and thus postpone the profound debate of legislation aspects for the later studies. Among other perspectives, the thesis also excludes technological and customer viewpoints. In addition, the thesis does not consider the internal statistical credit risk management practices of the financial service providers. Therefore, it does not give any specific statistical and mathematical recommendation how the credit risk management should be developed. The sole aim is to study the possibilities and challenges of utilizing alternative data in the credit risk

assessment and investigate the subject empirically by interviewing experts in the Finnish finance industry.

In order to create a solid theoretical framework, several open databases were explored to gain insight of the prior research. Based on the prior studies, existing companies and model examples, the thesis also aims to create synthesis of the existing data and data sources in order to build a clearer picture of the novelty concept. The empirical part focuses on expert interviews to gain deeper understanding on the matter and industry-specific knowledge to comprehend the opportunities and challenges of alternative data in the Finnish finance industry.

In the thesis, consumer loan is considered as a non-secured consumer credit and loan. Secured loans, home loans, mortgages and corporate loans are intentionally excluded, since they include a greater number of regulatory and macroprudential aspects such as EU Mortgage Directive and capital requirements (International Monetary Fund, 2017, p. 23-29). In addition, in Finland, mortgage finance is currently centralized to two major banks, OP Financial Group and Nordea, that possess nearly 70% market share in the consumer lending sector (Bank of Finland 2021). Therefore, there might not be any major applicable fintech companies operating on this field. However, the non-secured credit sector might have more competition since so called digital banks have already almost 20% market share in the non-collateral credit business (Wirman 2020), which could provide more opportunities for a research.

The limited amount of Finnish industry-related research steers to seek prior research on international context, thus, the literature review may exclude relevant matters related to the Finnish national legislation and industry qualities. However, the empirical part aims to fill in these gaps. Recently, in Finland, a few bachelor's degree theses have been conducted that are related to the topic of this thesis such as "Social Media Data in Credit Risk Management (Sinelampi 2020) and "Utilizing Artificial Intelligence in Credit Processes (Maatraiva 2019).

## **1.3 Terminology**

During the thesis research, it occurred that terminology used in credit risk assessment and credit reporting is rather incoherent. In order to emphasize the objectivity of the research and to avoid misleading terminology, the thesis uses an internationally established term “full-file credit reporting” (Saarenpää 2013, p. 24) when referring to a so called “positive credit register” which includes both positive and negative credit information. In addition, to clarify, when referring to credit reporting organization that obtain a credit reporting register, the thesis refers to a credit reference agency (CRA). There are certain differences between CRAs and credit information bureaus based on the public and private ownership, although that is not critical difference in the scope of this thesis.

## **2 ASSESMENT OF CREDITWORTHINESS**

### **2.1 Information asymmetry**

Based on the classical theory of economy, the information is the key element in the economy. Also, the concept of information asymmetry is the underlying problem in the credit industry. The information asymmetry creates a problem where one party of the contract does not have the same information as the other party (Ferretti and Vandone 2019, p. 107-108, citing: Admati & Pfleiderer 2000; Akerlof 1970; Berger and Udell 1995; Diamond 1991; Stiglitz and Weiss 1981). In the perspective of credit service providers, this could be considered as the different knowledge or level of data in the credit relationships: for example the payment behavior of the customer or ability or willingness to repay a debt (Ferretti et al. 2019, p.108). The information asymmetry leads to credit risk that means the possibility of default, the uncertainty related to the repayments, and the possible amount of loss which a bank may suffer in case of default (Ghosh 2017, p.6). One of the key elements in the credit risk management is to minimize the possible losses. For the banks, the non-secured loans such as credit cards and payday loans are usually higher risk loans, since there is no collateral posted against the loan to cover the possible losses in the case of default (Ghosh 2017).

The reduction of asymmetric information is a part of managing the lending business and, for instance, has multiple effects on risk management, credit pricing, customer's credit-worthiness, credit application processing, customers' segmentation, and product development (Ferretti and Vandone 2019, p.108, citing: Admati & Pfleiderer 2000; Akerlof 1970; Berger and Udell 1995; Diamond 1991; Stiglitz and Weiss 1981). In order to balance the asymmetry, personal data and credit history data are collected from various sources to create credit profiles and make optimal lending decisions (World Bank Group 2019, p.6).

## **2.2 Data-driven decision making with imperfect information**

According to Verbeke et al. (2017, p. 3), the character of data as a valuable raw material is pinpointed by a popular quote "data is the new oil". As stated, the modern risk management would not exist without advanced data analytics (Minelli et al. 2012, p. 37).

Based on Provost and Fawcett (2013, p. 53), data-driven decision making (DDDM) refers to the practice of establishing decisions on the analysis of data instead of intuition. According to Provost et al. (2013, p. 58), data science contributes to data-driven decision making and allows automated decision making at a vast scale, although it depends on the technologies for big data storage and engineering. Provost et al (2013, p. 58) states that the principles of data science should be considered explicitly to employ the full potential of data science. Correspondingly, according to Abdou et al. (2011, p. 8), obtaining data is a critical issue in building a robust credit score model that can provide competitive advantage for the financial institutions. By gathering and isolating all relevant data, the credit analysts should be able to decide which variables should be included and excluded in the final model that fits the real field requirements (Abdou 2011, p.8). The selection of data is important, since it will have a deterministic impact on the utilized analytical models (Verbeke et al. 2017, p. 14-15).

According to Bertsimas and Thiele (2006, p. 95), traditionally the models of decision making under uncertainty assume perfect information, for example, accurate values and specific probability distributions for the random variables. Whereas such specific knowledge is rarely available, and in many real-life applications, the decision maker must

work with imperfect information. Today, there are robust and data-driven decision-making techniques proposed by for example Doumplos et. al. (2016 and 2019) and Gaganis et al., (2020).

## 2.3 Creditworthiness

Theoretical definitions of creditworthiness are not unified, as the concept derives and develops from credit theory creating different interpretation of credit and its economic environment such as the culture of credit and development level of the market. Currently, the creditworthiness is perceived emphasizing the regulating role of credit in the economy. Based on these prior interpretation and theories, borrower's credit worthiness could be approached as: *borrower's judicial and financial capacity to attract credit assets as well as willingness and ability under the conditions of uncertainty and specific branch, regional, and individual peculiarities to pay back the received credit (with interest) in the time specified in the agreement making profit also for one's development from the credit resources* (Caplinska et al. 2020, p.2551-2552).

Models of prognoses based on statistical methods are utilized for the assessment of potential credit customer's creditworthiness. Nonetheless, the use of mathematical models prevents the financial institution to take the impact of qualitative data into consideration, since the classification models emphasize the quantitative factors which are generated mainly from the internal statistical data. However, in the complex models both quantitative and qualitative data could be utilized (Caplinska et al. 2020, p. 2549-2550). According to Rosenblatt (2020, p. 1), historically a credit decision was based on the three Cs: Capacity, Collateral and Character. Contemporarily, according to Caplinska et al. (2020, p. 2550, Altman et al. 1998), the banking practices in the USA use the six C-rule: Character, Capacity, Cash, Collateral, Conditions and Control. In addition, according to Caplinska et al. (2020, p. 2550, Hollander 1979), in the UK, the rule for giving credit is formed as PARTS: Purpose, Amount, Repayment, Terms and Security, or PARSER: Person, Amount, Repayment, Security, Expediency and Remuneration.

According to Caplinska et al. (2020), there are multiple methods to assess the creditworthiness, although the underlying principles are to multiply the value of specific financial

indicators by the weight of their indicator significance in the resulting indicator. The resulting indicator depends on the nature of the method which could be the credit rating method, prognosis method (probability of default) and creditworthiness assessment in complex methods (Caplinski et al. 2020, p. 2552). Pickert (2017, p 50, citing Sinclair 1994) also reminds that creditworthiness is a dynamic condition where the quality of the rating can alter as new events occur that impact on the borrower's creditworthiness.

### **2.3.1 Credit scoring**

It is vital to point out that the term “credit scoring” has different connotation across the world. In the US, the credit scoring typically represents a well-established numerical score generated by privately owned standards such as FICO® Score (FICO 2020 website, Rosenblatt 2020, p. 7). However, as learnt later, the credit reference system across the EU is incoherent and the credit score does not imply the same as in the US. Therefore, in the context of EU and Finland, it could be clearer to use a general term “assessment of creditworthiness” instead of “credit scoring”. Since the present thesis aims to focus on the data itself, it would be meaningful to use the same terminology as used in the prior research. Therefore, the term, credit scoring, is used despite the different implications.

As the lending process of non-secured loans is highly automated, the importance of profound customer information in the risk management is underlined. To lower the risks, the credit providers have implemented credit scoring systems using statistical methods assessing the creditworthiness (Kontkanen 2018, p.11). According to Ferretti (2014, p.7, citing: Thomas 2000, p. 149-172), the credit scoring was originally used to minimize the number of customers who default, but nowadays the scoring is used more for the risk-based pricing, and according to Minelli et al. (2012, p.39), for the marketing purposes targeting to the most profitable customers.

The information that is acquired by the credit provider of a loan applicant is used to create a numerical score for each applicant (Abdou 2011; Thomas, 2002; Hand & Jacka, 1998; Lewis, 1992). The evaluation of the creditworthiness is a crucial element in credit management decisions (Siddiqi, 2017; Abdou and Pointon, 2011). The evaluation process includes obtaining, analyzing, and classifying several credit data and variables (Abdou et al. 2011). The credit score could be presented as numbers to represent a single quality or

grades as letters or labels to represent one or more qualities (Anderson 2007, p.6). Thus, the credit scoring is a statistical probability whether the loan applicant will be good or bad (Siddiqi 2017) or more specifically, as usage of statistical model converts relevant data into a measure that guide credit decisions (Abdou et al. 2011). The credit scoring involves data mining techniques that include, for instance, statistics, artificial intelligence, and machine learning aiming at obtaining knowledge from large databases (Ferretti 2017, p. 16, citing: Bigus (1996); Desai, Conway, Crook, and Overstree (1997), pp. 323-346; Handzic, Tjandrawibawa, and Jeo (2003), pp. 97-109; Jensen (1992), pp. 15-26; Yobas and Crook (2000), pp. 111-125.)

Credit scoring can be only as good as the original design, and the scoring model is also limited by the quality of the data. Inaccuracies of the models can be reduced if the data is frequently updated (Abdou et al, 2011; Heffernan, 2005). According to Abdou et al. (2011), there is no optimal number of variables that must be used in the scoring models. The choice of data varies based on the data qualities and cultural and economic aspects, since different models are appropriate to a particular market and country (Abdou et al., 2011).

In addition, there is no generalized best statistical method or technique for creating a scoring model. However, in terms of higher predictive ability, there is evidence that more advanced techniques such as genetic programming and neural networks perform better than traditional techniques such as logistic regression and discriminant analysis. In addition, the hybrid methods, such as the hybrid neural discriminant techniques, appears to be promising opportunity for better predictive capabilities and classification (Abdou et al., 2011). Naturally, the credit score is not only the figure measured, but, along with other business, consideration such as expected loan approval rates, profit, customer exits, and losses are also measured (Siddiqi 2017, p. 9). Nonetheless, despite of the more advanced techniques, the issues related to so called black box-approach steers to utilize the original approaches in the credit rating sector, since in the advanced techniques, the link from input to output could be untraceable (Qiu et al. 2020, p.2, citing Altman et al., 2017).

### **2.3.2 Data for the assessment of creditworthiness**

Before explaining each separate dataset more closely, it is important to clarify the sources of information. Conventionally, the main sources are the loan application filled by the loan applicant and banks own internal customer information databases as well as the external CRA information concerning the possible payment defaults. In Finland, there are two credit reference agencies: Suomen Asiakastieto Oy and Bisnode Oy, that hold the credit information register concerning unfilled financial obligations and defaults. This information is called as negative credit information. Later as the Ministry of Justice proceeds with the legislative project, the Finnish Taxation authority could hold positive credit register that keeps reports on outstanding credit information, such as income, assets, and open credit accounts. Since the credit information forms an entity and, using the terms “negative and positive data”, could lead to misleading connotation, the credit information holding both positive and negative information could be referred as full-file credit information (Saarenpää 2013, p.24). The any other internal or external information which can be considered useful for credit risk assessment, this information is called as alternative data.

The data could be classified and grouped in various approaches, depending on the perspective. In general, credit data can be classified as structured and unstructured information (World Bank Group 2019, p.27) and even more specifically as semi-structured information (Siddiqi 2017, p. 150). However, this type of classification could be more beneficial in the perspective of data analytics. In the era of digital finance, all data is credit data (Ramsay & Williams 2019, p. 222). Also, in the perspective of the research, it is vital to clarify what data is considered as conventional or alternative data.

### **2.3.3 The conventional data**

The data contain customer information, demographics, data from authorities, and credit information (Salamina et al. 2019, p.21 Kontkanen 2018, p.24, Vojtek et al. 2006, p.164). More precisely, data that is widely used in the scoring are: age, marital status, dependents, educational level, occupation, time at present address, and having a credit card (Abdoy et al, 2011, citing: Sustersic et al, 2009; Hand et al. 2005; Lee and Chen 2005; Sarlija et al., 2004; Banasik et al. 2003; Chen & Huang, 2003; Lee et al., 2002; Orgler 1971;



Steenackers and Goovarts 1989). In addition, the information related to: time at present job, loan amount, loan duration, house ownership, monthly income, bank accounts, having a car, mortgage, purpose of loan, and guarantees have been utilized in the scoring models (Abdoy et al, 2011, citing: Ong et al. 2005; Lee and Chen, 2005; Greene 1998; Sarlija et al., 2004; Orgler 1971; Steenackers and Goovarts 1989). These figures are rather easy to collect and process in the loan application or revise in the bank accounts or obtain from the customer information system, and in some extend, from the CRAs (Siddiqi 2017).

#### **2.3.4 The alternative data**

Usually, the easily readable and usable structured data represent 20% of all data available (World Bank Group 2019, p.27). Due to the progress of credit risk assessment and technologies as well as the changing economic cycles and demographics, new and alternative data are emerging in the credit risk analysis (Ferretti 2017, p.37).

The new alternative sources could be classified as nontraditional data; non-credit financial data and non-financial information (Salamina et al., 2019, p.21). The nontraditional sources of data could be, for instance, retailers, utility providers, public and private entities that collect information on consumers (World Bank Group 2019, p. 27). In addition, nontraditional banking data in the unstructured format can include information from sources such as mobile banking, applications, social media data such as connections on LinkedIn, internet browsing, application programming interfaces (APIs), and voice response logs (Siddiqi 2017, p. 153). The data could be also from mobile payment companies, transactional data, and even psychometric data (World Bank Group 2019, p. 28).

The non-traditional data could also be referred as big data (Salamina et al, 2019, p.9). According to Siddiqi (2017), the big data phenomenon has made larger data sources available for credit analysts to obtain information regarding a credit applicant. This increase in the variety of data sources provides banks opportunities to build more detailed view of the person, compared to the more traditional data sources, such as credit bureaus and credit reference agencies.

In addition, the transactional data can be mined for behavioral analysis. A great amount of data could be collected from mobile phones: location, browsing patterns, social media profiles. These data can produce information about owner's identity, habits, financial wellbeing, relationships and even personality. These data sets could potentially add value to assess the creditworthiness, which is helpful to both incumbent and nontraditional creditors.

However, not all alternative data mean sensitive and personal social media information. Hibbeln et al. (2019) points out that there is a tendency to think over the top regarding the alternative data that is not only social media network data but may also mean existing underutilized data such as bank account data. Alternative data can also be something easily collectable such as payment data from products that are not traditionally reported to credit bureaus such as utility and cell phone bills. These kinds of continuous payments can also be considered as a sign of trustworthiness (Siddiqi 2017) and provide objective information on cash flows (Salamina et al., 2019). Based on Salamina et al. (2019, p. 9), the structured transactional information from e-commerce, electronic payments and utility services are relatively novel and fast-growing data sets.

According to Siddiqi (2017, p. 159), the ability of many financial institutions to access unstructured big data sources, may significantly change the traditional credit scoring. The big data provides a substantial digital catalogue of data that can be analyzed and categorized using machine-learning algorithms (Ferretti 2019, p. 37, citing: De Mooy 2017). By embracing digitalization and using modern IT and data analysis, financial service companies can provide customized products and dynamic pricing (Ferretti 2017, p.37). The analysis of the big data points can detect hidden behavior patterns that could be utilized to predict the future performance (Salamina et al. 2019) and comprehend the trends in consumer behavior and to develop products and services for better customer experience (Kaufman et al., 2013).

The data that is available and relevant to credit decisions can supplement or even replace the traditional data (Salamina et al, 2019). For example, according to Berg, Burg, Gombović & Puri (2018), digital footprints contain useful information in predicting the default

rate. Based on their study, even the simple and accessible variables generated from digital footprint are as relevant or even better than the credit score data from credit bureaus. The study analyzed more than 250 000 observation from a European E-commerce company that contain rather simple digital footprints as people access and register on the website. The data set contained the variables, such as the device type (e.g. tablet or mobile), the operating system (iOS or Android), the channel where the customer enters the website (price comparison site or search engine), the time of purchase, email service provider and the form of the email address (Berg et al p. 3). Based on the findings, Berg et al (2018) states that a credit score utilizing the digital footprint serves as a benchmark for other models which use more complicated data sources. However, as Berg et al (2018) states, the discriminatory power of a dual model using both credit bureau score and the digital footprint data, exceeds the models as their own. Based on this conclusion, Berg et al (2018) suggest, that the digital footprint complements the credit bureau data that could help lenders to make improved lending decision.

Another study related to alternative credit scoring data was conducted by using information generated on a Chinese e-commerce platform and users' phone data. Based on the study, Gambacorta, Huang, Qiu & Wang (2019) concluded that credit scoring model based on machine learning and big data was better predicting losses and defaults than the traditional models. According to Gambacorta et al. (2019, p.20), the non-traditional information from alternative sources such as phone application and e-commerce platforms has high predictive value. However, the advantage of alternative model tends to decline for those customers with a longer credit history.

According to a survey by the Institute of International Finance and McKinsey & Company (2017, p.43-44), advanced analysis and usage of the richer and nontraditional data are high priority capability in credit risk management. The survey also states that all relevant data should be identified and harnessed in order to make timely and precise decision. In the digital era, these data points for profiling attributes could be such as click-streams, chat transcripts, payment patterns, purchasing history, location patterns, behavioral irregularities, and even data from Internet of Things (IoT) sources. However, in order to learn which attributes have the predictive power, considerable volume and variety of data should be processed and analyzed. In addition, the survey also states even though

the incumbent banks are encouraged to seek alternative data sources, the banks' IT legacy systems restrain the development because even internal data can not be mined properly. These circumstances accelerate partnering and ecosystems withing the incumbent and fintech companies (Institute of International Finance and McKinsey & Company 2017).

### **2.3.5 Review of example companies in alternative data**

During the last decade, financial landscape has evolved rapidly as new technologies enable fintech companies to enter the market. Internationally, there are numerous benchmarking companies how the big data is utilized. According to Cornelli et al. (2020 p. 11), there are empirical evidence that big tech and fintech lenders have achieved lower default rates by using alternative data and machine learning than by conventional data models would have provided.

For example, the companies called Tala and Kabbage obtain insight from consumer behavior on social media and payment platforms to make credit decisions (World Bank Group 2019, p.5). Tala collects credit scoring data from applicants' phone, such as location, money transfers, network diversity and communication, along with thousands of other data points (Shivani Siroya, TED Talk 2016). In addition, in credit scoring various methodologies are used to assess credit risk. For instance, Lenddo and Friendlyscore use social media data for credit scoring (Salamina 2019), however the use of such data is controversial. Thus, its use will be most likely restricted due to increasing regulation stemming from social media platform's privacy concerns (Packin, 2019). In addition, companies such as Trusting Social, Tiixa and First Access use telecommunication data: top-up patterns, mobile money use and contacts. CreditVidya utilize analytics and machine learning to examine the phone data: messaging content, telephone usage, browsing data and GPS location. Verde International obtain data also from conventional credit bureau in addition to alternative data (Salamina 2019).

A company called ZestFinance provides analyzing tool for other banks to assess their customers by artificial intelligence-based credit scoring solution called ZAML® (Zest Automated Machine Learning) that relies solely on unstructured data sources (Zest-Finance website, referred 14.5.2021). A European example would be a company called

Monedo, which is a German based fintech company that operates in Spain and Poland and in the emerging markets such Russia, India, and Thailand. Monedo provides micro-loans and Points of Sale service utilizing advanced data analysis creating numerous data points to assess a loan applicant without the existing credit score. The company uses the data analysis together with machine learning to decide whether a loan applicant is credit-worthy (Monedo website 14.4.2020).

Recently, in Finland, there have been innovative development project concerning automated credit decision-making and models. For example, OP Financial Group has created an advanced automated credit decision-making system (OP Credit Engine) that is the first system of a European bank that can provide a dynamic credit pricing of home loans since 2018. However, the legislation sets strict boundaries in which companies can operate. The information and the variables fed to the models are basically the same data that are used in manual credit decisioning since the companies need to carefully consider very what data can be modelled, for example age, gender nor mother tongue can not categorically determine the credit decision. Thus, so called “black box” modeling is prohibited by regulation. This is due to the fact that one should be able to trace back the variables in the model that has a certain effect on the automated decision and the service provider representatives should be able to explain how the model actually works (Kaisa Parkkila, OP Tech Podcast #23, 22.4.2020). The black box related issues are also noted by Koskinen (2018).

## **2.4 Information Exchange in EU Retail Financial Markets**

The information exchange concerning the customers is common in the retail finance and it has become extensively used practice of finance industry to underwrite decisions on borrowings and other services. Practically the exchange of information means the access to credit information register and related information that are managed by third party providers (such as CRAs) in order to evaluate the risk of financing and borrower’s credit-worthiness (Ferretti 2014, p.7). Since the credit providers utilize the information supplied by the credit reporting systems, the CRAs are a primary factor when evaluating the creditworthiness of a credit applicant (World Bank 2019, p. 4).

In general, CRAs obtain information on consumer's current payment behavior, past credits and also collect information from public sources and distribute the information to credit service providers prior their credit decision-making. Some CRAs add value as they merge and analyze the collected data to provide analytics and scoring tools to support the credit process. Despite the important role as the information suppliers, CRAs are not involved in decision-making in the credit process and the information is used only by the lenders (ACCIS 2018).

#### **2.4.1 The mess of the European credit data sharing**

Despite the coherent privacy rules and consumer protection, the European credit reporting space is very diverse. The EU member countries and other European countries do not have any centralized agencies or authorities which would provide any specific credit reference data nor score (ACCIS survey 2017). Naturally, within EU, the credit reporting is conducted according to the EU regulation and directives, but national legislation, principles and traditions set the framework in each country (Kontkanen 2018, p. 29; European Commission 2018). In addition, the definitions and assessment of which data should be in the databases differ from a country to country as the terminologies and significance of the data varies greatly (Ferretti 2014, p. 65). Some of the EU countries and CRAs utilize information from wide range and depth of records as several agencies obtain both positive and negative data from various sources and organizations (ACCIS 2017 survey). The full-file credit register is applied in various of forms in all EU countries except in France. Rather commonly, also information concerning commodity related credit accounts such as part payments and commodities such as tele, water and electric services are collected in the register. In Sweden, the income and wealth information are collected, as commodity related credit data in the United Kingdom. In addition, in Ireland, the register contains information concerning consumers' rental contracts (Kontkanen 2018, p. 24, 27, 30).

As the EU level reporting system is extremely incoherent, in 2017, the EU Commission has initiated an action plan for common creditworthiness assessment standards and principles under the directive 2008/48/EC. The aims of the action plan are to reinforce the development of the EU single market strategy in financial services, and practically, to facilitate the sharing of credit information that financial services could be provided across

the EU (EU Commission). The Commission is interested in profiting from various mechanisms and categories of data, such as those which could be provided through FinTech and big data, credit reporting systems, and loan data standardization to enable the creditors to assess the borrowers' creditworthiness on the basis of the same level and quality of information. (Livada, 2019, p. 228)

In addition, the EU has initiated a project called *AnaCredit that is a shared multipurpose database containing loan-by-loan information on credit extended by credit institutions to companies and other legal entities* (Israël et al., 2017, p.3). Currently, the analytical credit dataset is used for financial policy and central banking purposes (European Central Bank website, 7.4.2020), but in the future it could be enlarged to household exposures but most likely to be limited to mortgages (Israël et al., 2017, p. 38).

Majority of the EU countries' CRAs collect the full-file credit data (positive and negative credit data) from variety of sources but the policies differentiate across the EU. The development of credit reporting systems is important in the countries that do not yet support the comprehensive credit reporting system. Inadequate depth and scale of data decrease the quality of the service, increases cost of credit, and contributes overly restrictive and unaffordable lending. With optimal credit data credit providers can improve responsible lending principles such as compliance, favorable terms, and inclusive lending for the underserved customers (ACCIS 2018).

#### **2.4.2 Credit refence reporting in Finland**

According to ACCIS, the Finnish CRA, Suomen Asiakastieto Oy, holds only negative data acquired from public sources such as court or a similar public authority (ACCIS 2017 survey). Although, since 2013 there have been preliminary initiatives to enrich the credit reference reporting (Kontkanen 2018, p.29). According to Suomen Asiakastieto Oy, the CRA obtains and shares a full-file credit data in a credit reference system on a customer's consent ("Positiiviset luottotiedot kuluttajaluottojen kyselyjärjestelmässä - Suomen Asiakastieto Oy," 7.4.2020).

The full-file credit reporting is expected to provide more comprehensive overview of a customer which enables the lender e.g. to set optimal credit pricing which is also an advantage for the lower risk customers as they might benefit from lower credit expenses (Kontkanen 2018, p. 10). However, the current system that Suomen Asiakastieto Oy provides is not complete and does not include all financial service providers since there are only 39 member organizations in the system. In addition, 8 of the member companies provide only concise information. As a reference, two major banks OP Financial Group and Nordea are not members of the system. Although it is noteworthy that foreign banks and service providers are gaining an increasing market share in the Finnish non-secured credit sector as mentioned earlier in this thesis.

### **2.4.3 Full-file credit information register in Finland in 2023**

According to Kontkanen (2018, p. 64) the need for full-file credit information is obvious considering the increased supply of credit services, digitalization, household indebtedness and international practices. A centralized credit information register that collects and shares, in addition to negative data of payment defaults, also other types of positive data, would demonstrate the creditworthiness more profoundly and thus significantly complement the overall view of customer's creditworthiness (Kontkanen 2018, p.10&61).

Development of legislation and information technology contributes the validity to establish the register. Especially, the national income register (*in Finnish: Tulorekisteri*) enables the possibility to include income information into the credit register, although the current legislation does not support the development (Kontkanen 2018, p. 61).



Impact on credit providers	Impact on credit applicants	Impact on regulators and supervisors
<ul style="list-style-type: none"> <li>• Reduces information asymmetry</li> <li>• Reduces lending risk</li> <li>• Contributes responsible lending</li> <li>• Reduces need for collateral</li> <li>• Helps for business development</li> <li>• Simplifies credit decision-making processes</li> </ul>	<ul style="list-style-type: none"> <li>• Enables more dynamic pricing</li> <li>• Reduces need for guarantees</li> <li>• Prevents over-indebtedness</li> <li>• Could ease the access to credit</li> <li>• Could simplify taking credits</li> </ul>	<ul style="list-style-type: none"> <li>• Increases information about the financial market</li> <li>• Increases financial stability</li> <li>• Contributes risk management</li> <li>• Contributes competition</li> <li>• Prevents over-indebtedness</li> </ul>

*Table 1. Impacts of the full-file credit information by Kontkanen (2018, p10).*

Accordingly, the Ministry of Justice has initiated a preparational project concerning the centralized credit register. The aim is to plan the operational model of the register and administrative arrangements within the Finnish public authorities. The positive credit information register could be launched in 2023 and the authority responsible for the register is expected to be Finnish Tax Office. Currently the formulation of the positive credit register contains only information of consumers' income and existing credits, as the general objective of the register is to prevent the over-indebtedness (Ministry of Justice 2020).

The sensitive data requires special attention to data protection concerns. However, the earlier initial research cleared that the EU directive, General Data Protection Regulation (GDPR) is not an obstacle to process positive credit information and establish the register (Lång 2018, p. 124).

According to ACCIS (2018) the GDPR directive articulates certain views on creditworthiness assessment that ACCIS has formed into five principles. Two of the principles consider the data qualities as follows (ACCIS 2018):

*Principle 2. Creditworthiness assessments should be based on all data that is relevant, accurate and necessary for each specific credit application and deriving from all relevant data sources.*

*Principle 3. To effectively support creditworthiness assessments, credit reporting systems should contain optimal depth and breadth of data, including positive data.*

Therefore, in creditworthiness assessment lenders can collect data from multiple existing sources with a certain flexibility. Although that flexibility does not mean that credit providers can ignore data sources that fundamentally contribute the obligation of any creditworthiness assessment. However, in some EU countries, it is required to consider the traditional data such as payment defaults and credit history as they are considered to have predictive power of the possible default. Credit data from other non-traditional sources should be however differentiated from the traditional sources. In addition, the relevance and necessity of the non-traditional data should be carefully evaluated (ACCIS 2018).

## **2.5 Regulation**

Regulation concerning the credit assessment can be examined from various perspective, such as consumer rights and protection (Zuboff 2019 and Quirk et al. 2018), technology and automated decision making (Kroll et al. 2017), and competition law (Ferretti 2014). In this thesis the regulation is covered in the perspective of the EU regulation such as, Consumer Credit Directive (CCD), General Data Protection Directive (GDPR) and Finnish national law since they provide relatively firm legal framework regarding the assessment of creditworthiness.

### **2.5.1 Consumer Credit Directive**

In the EU, credit assessment has a strong regulatory and responsible lending perspective as credit service providers are compelled to assess loan applicant's creditworthiness. These obligations are included in the European directives: Mortgage Credit Directive (MCC) and Consumer Credit Directive (Ferretti 2014, p.63-64 and ACCIS 2018). In

contrary, according to Aliyev (2019, p.14), the CCD is less strict concerning responsible lending obligations than the MCC. The CCD for instance, does not make any clear proposal how the creditworthiness should be tested and therefore the EU member countries have different perceptions and designs related to creditworthiness assessment.

According to Ferretti (2017, p. 25), only recently the European Banking Authority (EBA) has started to investigate the possible advantages and disadvantages of the use of consumer data by financial service providers.

### **2.5.2 The General Data Protection Regulation**

The European Union's General Data Protection Regulation (GDPR) impose detailed requirements for organizations on collecting, storing and managing personal data. Any information about the data subject (an identified or identifiable person) is personal data. That information is: name, address, identification card, passport number, income, cultural profile, Internet Protocol address and data held for medical purposes that identify a person. In addition, there are classified special personal data categories that are prohibited to be processed: racial or ethnic origin, sexual orientation, political opinions, religious or philosophical beliefs, trade-union membership as well as genetic, biometric or health data, except in specific cases such as consumer's explicit consent or when processing is needed for reasons of substantial public interest, on the basis of EU or national law, as well as personal data related to criminal convictions and offences, unless this is authorized by EU or national law (Your Europe website, Data protection, 29.4.2020)

Data processing is allowed in a fair and legitimate manner for a specific lawful purpose. For instance: data can be processed if given a consent from the individual concerned, personal data is needed to fulfil a contractual or legal obligation or crucial interest of the individual. The general rule is that the individual's fundamental rights and freedom override the organization's interest. The consent to process personal data only permits to process the data for the reason for which consent was given. The individual also must have an opportunity to withdraw the consent. (Your Europe website, Data protection, 29.4.2020).

Nonetheless, GDPR does not contain any specific regulation referring to the assessment of creditworthiness and the regulatory framework is not harmonized withing EU legislation (Kontkanen 2018, p.19). Although, certain specifications related to the automated decision-making and profiling can be discovered. Based on the GDPR, individuals have the right not to be subject to solely automated decision processing which also applies in the financial services. That rule can be overruled by a customer consent. Nevertheless, the individual has a right to contest the automated decision and ask the automated decision to be examined by a person (Your Europe website, Data protection, 29.4.2020). However, the legal existence and feasibility of the claimed “right to explanation” of all decisions made by automated or AI algorithmic systems, is questioned. For instance, Wachter et al. (2017) claim that the GDPR lacks explicit and definite rights and safeguards against automated decision-making.

However, according to Koskinen (2018), applying GDPR into modeling of personal information for credit service purposes is difficult because for example machine learning techniques are not compliant with the law regarding the transparency. The big data fed to algorithms can embed correlations that can be discriminative since the algorithms are fundamentally design by humans and the data is collected from the society. Therefore, the data can include biases and are not free from the impact of the society where the prejudice occurs which can lead to systematic and broad discrimination (Koskinen 2018, p. 241-251).

### **2.5.3 Credit Information Act 2007/527**

The regulation applies in collecting, producing, storing, sharing, using, and processing credit data. The aim of the regulation is to ensure access to reliable credit data, protect privacy and protect the right to be assessed by correct and proper information (Credit Information Act 2007/527 1§, 2§)

The regulation restricts what data can be used for assessment of creditworthiness. The law states that information that is obtained from reliable sources and is necessary and appropriate to define solvency (ability to pay), willingness to pay or ability to meet the commitments, can be used as credit data. The source of information can be collected only

from the credit applicant, credit reference agencies or from authoritative registers (Credit Information Act 2007/527 5§, 6§). Consumer protection related to credit data is also referred in the regulations regarding the personal data protection and consumer protection (Consumer Protection Act 1978/38).

### **3 METHODOLOGY**

This thesis utilizes qualitative research method. According to Mligo (2016, p.7) the underlying question is: what is the needed form of data for the research? Since the core of the thesis is to gain industry specific insight regarding a relatively new concept, the suitable method would be a qualitative research. More specifically, the thesis tries to provide interpretation of a phenomenon (assessment of creditworthiness and the role of new data sources) in order to acquire new knowledge (possibilities and challenges in the context of Finnish finance industry) that could facilitate evaluation of existing credit assessment practices and even advancements in the industry and legislation.

Choosing the research approach can be debatable. Saunders (2019) suggests that if the research topic is new and generates debate but there is little related literature, the inductive approach could be suitable to generate data and reflect upon what theoretical themes the data proposes. On the other hand, if there is a lot of information about the research topic but far less in the context in which the research is conducted, the appropriate approach could be abductive. The aim of the thesis is not to deduce any specific hypothesis nor to make merely generalizations from the researched topic but to study exploratorily the possible opportunities and challenges of alternative data in the assessment of creditworthiness in the Finnish finance context. Saunders (2019) explains that applying the abductive approach to the research needs obtaining data that are sufficiently detailed and rich to allow to explore the phenomenon and identify and explain themes and patterns regarding the research questions. These themes and patterns that occur from the empirical research are intended to be integrated in an overall conceptual framework of assessment of creditworthiness.

On these following grounds, it is stated that the chosen research methodology and approach enable to reach the goal of the research: 1) the absent of relevant research in the desired context of Finnish finance industry, 2) the ongoing process of investigation by the Ministry of Justice to reform the credit information register and 3) access to well-established financial theories and 4) a vast amount of relevant quantitative research regarding the alternative data in credit risk management and 5) a vast amount of secondary data related to credit assessment and practices.

### **3.1 Sampling**

The sample were selected by the judgement of the researcher, thus following the principles of purposive sampling. The criterion was that the respondents are experienced in the field and are expected to have adequate information of the research topic. It was aimed to conduct at least five interviews that was assumed to provide the adequate amount of relevant insight. The researcher contacted by email over 10 companies that eventually resulted five responses and four conducted interviews. One of the representatives declined the interview invitation referring to the current state of legislation concerning consumer credits.

In this research, any generalization based on the discoveries can be debatable since the sample size was small and respondents were selected by the researcher. Although, according to Dudovskiy (2018), when taking into consideration the nature of research design and aims of the research, it could be argued that the purposive sampling may prove effective by in-dept interviews. This is also the case of the thesis since the respondents are expected to provide industry specific primary information. In addition, it is worth mentioning that the sample included representatives from both incumbent and emerging finance companies as well as an industry representative and scholar.

### **3.2 Interview**

The data was collected through interviews as it was assumed to be the most suitable data collection technique to access the industry related deep and rich knowledge concerning a rather novelty concept. The aim was to gain insight from both incumbent and smaller

emerging financial service providers as well as regulatory perspective from related authorities. The interviewees are described below in the table 2. In order validate the relevancy of the chosen interviewees, their background, position, and organization are described as well. The names of the organizations nor interviewees are not presented for the sake of promised anonymity. The interviews lasted from 40 to 60 minutes, and were recorded and transliterated.

	<b>Title/Position</b>	<b>Experience</b>	<b>Organization</b>
Interviewee 1	Head of Consumer Credits	16 years in financial sector	A Nordic financial group
Interviewee 2	Leading Legal Counsel	5 years in financial sector	A representative organization in the Finnish financial sector
Interviewee 3	Head of Credit Automation	15 years in financial sector	A major financial group in Finland
Interviewee 4	Adjunct Professor, Researcher & Board Member	Over 30 years in financial sector	An academic in the Finnish universities. A board member in a Finnish financial group and in an organization representing Finnish Fintech sector

*Table 2. Description of the interviewees.*

The interviews were conducted as semi-structured interview to gain wide range of insight for exploratory research. The interview was planned to be flexible in both questioning and responding, although it was guided by the interviewer through predetermined set of questions. The list of questions can be reviewed as appendix in the thesis. According to Mligo (2016, 87). the open-ended questions also reduce the bias of the interviewer and facilitate adequate exploration of viewpoints of the interviewees. Since the aim was to

have in-depth exploration of the informants' experience and view on the researched matter, the unstructured interview could be applied. However, presumably, it could be more convenient to apply a semi-structured thematic interview technique because it also facilitates a certain freedom to the respondents. The semi-structured technique helps the researcher to explore further by probing follow-up questions and provides a certain framework guided by the interview questions. Correspondingly, in the perspective of the abductive research approach, it also worth mentioning that the theories of credit risk assessment and the information asymmetry are used to design broadly the interview questions. As said, the view of the respondents is desired, therefore, the predetermined questions guide the interview but do not provide inflexible structure. Respectively, the inductively occurring themes from the interviews can be mapped against the theoretical framework, as the framework can be modified in the process. This way the research can serve both theoretical and practical contributions.

### **3.3 Data analysis**

The collected data is organized, analyzed, and interpreted accordingly. The obtained data is formed as scripts and recordings from phone interviews and one-on-one video conferences. The recorded interviews are systematically transliterated to the form suitable for analysis. In the research report the data is coded and arranged into subthemes for easier interpretation keeping in mind the research problem. Ultimately, the codes that appear the most relevant are presented in the thesis. The analysis follows the principles of thematic analysis in order to identify the patterns in the data that are important and address the research. The most important categories are reviewed, and subsequently, the most important categories are labelled, and the interconnection of the categories are described as they base the core of the discoveries.

Some data from the interviews are be shown as extracts to comprehend and interpret the view of the respondents, considering the meaning and significance as well as understanding what the interviewees have expressed. The further discussion of data involves arguing, comparing, and evaluating data as they appear.



## **4 FINDINGS**

The findings are divided into different sub-sections originating from the interview questions in order to find the themes that appeared relevant for the research question. These themes are 1. Data quality and methods, 2. Improvements in the data quality, which is discussed separately as 2.1. Positive Credit Register, 2.2 Transactional data and 2.3 Big data. Each sub-section is summarized as well. These themes are formed as a synthesis in the Discussion chapter as Opportunities and Challenges. The themes are presented below in the table 3.

	<b>Interview1</b>	<b>Interview2</b>	<b>Interview3</b>	<b>Interview4</b>	<b>Summary</b>
<b>1. Data quality and methods</b>	Regulative requirements  Imperfect data  Manual data collection	Slight differences across the sector  Stems from legislation, competition, and risk profiles	Process is not ideal  Current methods give satisfactory results  Data is imperfect	Old-fashioned  New data is needed to enrich the current models	Regulative requirements  Provides adequate results  Not ideal for good customer experience
<b>2. Improvements on data quality</b>	Credit register  Wealth- & property information  Account transaction data  Data quality is crucial	Credit Register  Housing company loans (in the credit register)  Utility related debts and part payments (in the credit register)	Credit Register  Income Register  Account transaction data	Open Banking concepts  Full understanding of digitalization	Credit register  Income and wealth information  Account transaction data  Reliable and efficient data processing
<b>2.1. Full-file credit register</b>	Essential  Slow progress  Decreases credit prices  Better customer experience and credit processes	Essential  Coverage and scope  Legitimacy  Useful tool for the risk assessment  Slow progress	Essential  Equality among all credit providers  Scope  Slow progress  Better customer experience	N/A	Essential  Enhance better credit decision process  Better customer experience  Legitimacy  Slow progress
<b>2.2. Account transactional data</b>	PSD2  Requires change in customer behavior  Open Banking development is slow	Data Protection  Consumer protection	PSD2  Change in consumer behavior  Uncertainty in the practical implementation	PSD2  Awareness of data sharing  Open banking enables  Transactional data is crucial	Potential Data  Better customer experience and credit processes  Uncertain practical implementation  Requires change in customer behavior.
<b>2.3. Big data</b>	Requires massive resources  GDPR  Focus on data quality, relevancy & database maintenance	Interesting data set  Customer behavior and data sharing  Regulative boundaries	The impact value might be low  Expensive, limited resources  Open data economy  Could be used in AML and Fraud operations	Data analytical tools are highly developed  Understanding the full capacity of digitalization  Fair Data Economy	Unrealistic  Regulatory and technological constraints  Lack of innovative mind set  Fair data economy

*Table 3: Results of the interviews*

## 4.1 Data quality and methods

The interviews quite clearly indicated that the current methods for the creditworthiness assessment are old fashioned and relies on the conventional methods of collecting data. The main sources of data are typically the internal customer data and the credit application filled by the customer. The financial institutions also collect information from the external sources such as Population Information System of Finland (VRK) and private sources such as CRAs.

According to the interviews, the manual data collection leads to inefficient credit process, low customer experience and inefficient credit decision-making. Interviewee 1. commented that their credit process includes manual work for both the bank and customer when managing additional information related to, for example, taxation or property documents, which the respondents felt unfortunate. The interviewee 3. had similar opinion and stated that:

"In my opinion, the credit decision process is ready when nothing else is necessary to ask except the customer's need for the loan... .."I would like to obtain all the information from a reliable source, statistically given in the same format because if a person him/herself fills in the information, the customer experience is usually poor."

The imperfect data can also lead to higher credit prices. The interviewee 1. commented that:

"There is a lot of competition and competition situation has itself lowered the credit prices especially in the past two three years, but if the impact of the positive credit register is considered, it will lower the prices at some level, at least in non -collateral credits. I am very certain that it will lower the prices."

Although the current method can be considered as old-fashioned, the interviewee 1 and 3 considered that current methods are still adequate and provide satisfactory results. According to the interviewee 1, the collected information is sufficient up to a certain point if a bank is willing to take the risk. The interviewee 1 also commented that the evaluation of the risk of customer's insolvency is based on the historical data and certain stability factors, such as life situation, which the statistical models are based on, have not changed fundamentally.

The interviewee 3 commented that the sources of information are adequate as the bank is making relatively good profit and the amount of bad credits is very moderate compared to the industry average. However, the respondent emphasized the impact of imperfect credit process from the customer's point of view and argued that the bank can not make as good customer selection without lowering the customer experience by asking detailed information on the credit application.

The interviewee 4 commented more broadly that it is surprising that the assessment of creditworthiness still relies on ancient methods. The respondent also contemplated the way of thinking in the banking industry by stating that:

” We know banking is a bit out-of-date at the moment and has old-fashioned approach when thinking that since we have managed in the past, we will manage also now”.

## **4.2 Improvements in the data**

The interviewees clearly pointed out certain data sets that are essential for the financial industry in the future. These data sets are the full-file credit register and account transactional data. Also big data could have a role in the future as well. Nonetheless, based on in the interviews, the question of the essential data sets emphasizes the efficiency of processing the data and reliability of the data. In the following chapters these three main findings are explained further.

### **4.2.1 Full-file credit register**

The ongoing legislative project regarding the full-file credit register is anticipated within the financial sectors and is expected to provide data to ease some of the current constraints in credit decision-making. All respondents agree that the register is highly needed since it would provide structured, verified, and reliable information to all credit providers.

The interviewee 3 argued that the credit register is not only essential for the quality of the credit decision-making but also for the better customer experience. The interviewee estimated that, as the regulations tightens and more detailed information about customer's

depts needs to be included in the credit application, it would be burdensome for the customer to fill in all these detailed data about housing loans, credit card depts and so on. As an example, the interviewee 3 commented:

If I needed to dug [details concerning] the mortgage payments, credit card interest percentages and all, in order to get a small credit, I would lose my temper”

However, since the project is still ongoing there are many open questions related to the actual data sets that are included in the final register. Respondents hoped that certain types of creditworthiness indicators such as Income register, Wealth information and related depts should be included in the register because it would provide relevant and verified information in standard form to the credit providers. It was also noted that credit can be considered in various forms and it is yet unclear which kind of loans are considered as registered loans. For example if utility related loans and part payments are under the radar, the register might not provide the best possible coverage.

However, the respondents (2,3) were quite clear that the register should only serve the purpose as a credit register, that it would not include other types of information such as rental contracts, utility payments nor should be it used for the marketing purposes. The respondents (2,3) thought that register is well scoped, and the register should be legit. Respondents were rather hesitant to broaden the scope of the register to include other types of information as in many other European registers. Also, they saw important that the register should be used only for the credit related purposes. The other kind of usage would perhaps endanger the legitimacy and efficiency of the register.

In addition, the interviewee 3 also estimated that even though the full file credit register will be launched and put banks in equal position in terms of data, the register itself will not trigger any competition by foreign banks or so-called big tech companies since the Finnish financial market is very challenging considering the market size and tight regulation.

### 4.2.2 Transactional Data

One data set clearly emerged from the interviews which is the account transaction data. The respondents saw potential in automated transactional data that could give clearer insight on the customer's mindset because banks could indicate signs of creditworthiness based on the account transactions. The interviewee 1 commented that:

“Account transaction data is interesting... ..it would be good to be able to build categories and verify income and debts and for example risk factors such as gambling...this could be significant thing in the credit assessment”.

According to the interviews, the open banking concepts and especially the PSD2 regulation was expected to increase the possibilities for banks to utilize the account data. However, the respondents felt that the progress has been rather slow. The interviewee 3 explained that the PSD2 includes some constraints, for instance, the data collection is not ideal since the implementation is expensive and imperfect. The interviewee 3 clarified that the requirements of Strong Customer Authentication (SCA) forces consumers to verify his or her identification in many parts along the different processes. As an example the interviewee 3 emphasized the credit application process which can be complicated if a customer needs to identify him/herself multiple times that can lead to a poor customer experience. The interviewee also pointed out that the PSD2 allows consumers to choose from which accounts the information is provided, and in some cases, relevant information might be hidden in purpose. In addition, the bank account might have multiple owners such as spouses, which disrupts the data quality and usability of the data, since the data can not be specified to a certain person. According to the interviewee 3, these constraints could lead to sub optimal analysis, and therefore, the practical implementation to obtain account transaction data is almost impossible.

In the contrary, the fourth respondent pointed out that, in Finland, there is a lack of innovative mind set as the mental models restrict the development and new kind of thinking. The respondent felt that although banks have been putting a lot of effort in the development for the PSD2 enabling services, the results have been poor because the investments might have been targeted in wrong places. The respondent 4 argued that there is a

common misunderstanding how the banking should work as the customer is expected to join the bank's process, whereas, the direction should be the opposite, so that the APIs are applied in the customer's process. The respondent 4 continued pondering:

These [developments] are often considered as automizing an old process, but it is not really thought could this process be renewed totally. I don't have any facts, but it could be that many people think, also the newcomers, that this is the process, this is automatized and digitalized, but it is not thought that could this process be something else or is the whole process useless today.

The interviewee 1 also pointed out that even though the PSD2 regulation forces banks to open their data, and it will change the business in the future, the progress requires a change in the customer behavior so that the customers allow their account data to be reviewed, and even if the process is automatized, it still requires customer's intervention in the data collection.

#### **4.2.3 Big Data**

Respondents considered so called big data or digital footprint data as a buzz word in the perspective of credit process. Based on the interviews, utilizing big data seems unrealistic in the short term because the data is considered inaccurate, unstable that might not provide satisfactory results. Also the investment cost into analytics and IT would be costly, therefore the business potential is unclear.

The respondent 1 commented that with limited resources it is hard to figure out any additional relatively stable and coherent data that is processable. The respondent also commented that data from social media and internet browsing requires substantial amount of analytics and computing power and making conclusion based on that information goes to very personal level. The respondent 2 commented that alternative data could be interesting pool of data if it can be used without limitations, and big data might be useful especially when acquiring new customers whose credit behavior is yet unknown. However the respondent 2 also believed that, in lending business, it is understood that not all data is worth using, not only because of privacy rules but also because of the fundamental idea that personal data ought to be personal.

The respondent 3 pointed out that after the implementation of the full-file credit register and income register, there might not even be need for any other alternative data in credit business. However, the respondent 3 concluded that big data could be more useful in financial crime preventions, such as anti-money laundering and fraud detections.

The respondent 4 explained that nowadays the data analytics is highly developed which can provide accurate conclusions, but it requires right data. The respondent also stated that if using some data is not against any regulation nor ethically questionable, the data should be used if it is beneficial for the customer and the service provider. The respondent continued that the current constraints in the development are both technical and mental, that companies do not dare to make bold initiatives. The interviewee 4 pointed out the mental models that can restrict understanding the full potential of technology and the innovative way of thinking in a managerial level. The interviewee explained that if the incumbent banks do not emphasize the importance of the payment services, they might lose the relevant data from their customers, for example the account transaction data.

## **5 DISCUSSION**

Based on the prior studies (Saarenpää 2013 and Kontkanen 2018), there is a proven need for developing the assessment of creditworthiness which also occurs in the thesis interviews. Based on the interviews, there is a lot of expectation for the full-file credit register which could deliver relevant, accurate and unified information for all banks and loan providers, as well as, setting the banks in equal position in terms of data. The full-file register is a step towards the ideal method that all the relevant information should be accessible from the relevant source automatically, not manually from the customer, thus more applicable for data-driven decision-making.

As mentioned, the discussion related the full-file credit register has been around already for the past 30 years, and finally, the project has proceeded to law drafting. There are still many open questions and unsolved practical problems related to the actual implementation. However, after the register is launched, it will most probably provide rich data for the banks to reduce the underlying problem of information of asymmetry in the consumer credit sector.



Despite the imperfect information and credit granting process involving manual data sharing, banks can select the customers adequately and generate sufficient results. This naturally might cause low customer experience as mentioned in the interviews. Developing the credit assessment methods could improve the customer experience and credit processes. Also the interviewees indicated that the more profound information about the consumer's creditworthiness could also result decreasing credit prices. By utilizing timely and accurate credit information, the companies could also use dynamic credit models that are adaptable to changing nominators of creditworthiness. Although some of the credit registers in EU countries have a broad scope and coverage containing various information sources such as, rental contracts and utility related payments, the register in Finland is expected to serve only as a credit related register so that the legitimacy is not compromised.

Consequently, after the full file credit register is utilized providing enough information, there is doubt if any other new data set would even deliver more relevant and processable information that could be used in the risk assessment. It could be that the full-file credit register dominates the discussion in the credit sector and real-life alternative data solutions are neglected. However, it is understandable that the major challenges of developing the assessment of creditworthiness relate to the lack of resources and the data quality, since the development can be costly and requires investment in the analytics and IT systems.

The interviews indicate that the transactional data can provide valid information for the assessment of creditworthiness. Although the automated account transaction data could be valuable information, one crucial point was raised in all interviewees which was the customer point of view and how is the customer experience understood in the credit process. Regulation might set boundaries on the development of new services and innovation, but ultimately, it depends on the customers perception if the service is considered valuable. For example, the PSD2 regulation and the open banking concepts enable to share information across the banking network, however, it occurred that regulation implements poorly in the customer process, since the customers need to give a permission

and authentication along the process. Therefore, the compatibility in the credit process and customer interface is considered as a challenge.

So called big data or digital footprints as a sensitive social media and internet behavior data, is most likely limited since the privacy regulation set strict rules. The interviewees indicated that the big data and the digital footprints are not expected to cause any major competition nor possibilities in the credit risk assessment. However, certain payment behavior and transaction related big data could be utilized in crime prevention such as anti-money laundering and fraud detection. Even though big data might not be valued as much as the full-file credit register and the account transaction data in the credit risk assessment, it is still worth acknowledging the potential role of big data. If all the data is considered as a credit data, the banks might need to ensure that they do not lose such valuable information. Transactional data is formed by purchases and other account transaction. If big data companies start to operate more on payment solutions, even if they are not directly operating on finance and credits, they obtain the transactional data, therefore, the incumbent banks may lose the transactional data.

However, the future scenarios ought to bear in mind. The EU aims to reinforce the EU single market also in the retail finance, fintech companies harness the digitalization, and the foreign banks are increasingly winning a bigger share in the credit market in Finland. Therefore, hopefully the incumbent domestic financial companies are aware of the future scenarios. One of the interviewees pointed out that the mental models can restrict understanding the full potential of technology and the innovative way of thinking in the managerial level.

The respondents saw that the society might change over time towards open data economy and people might be more willing to share their personal data. This could facilitate using data from various of sources. However, this kind of data economy might take generations to be fully in use.

On these grounds, the financial service providers may need to develop their credit risk management concepts to align with the variables of the modern society and economics. If the Finnish finance industry holds on the conventional information and methods, the

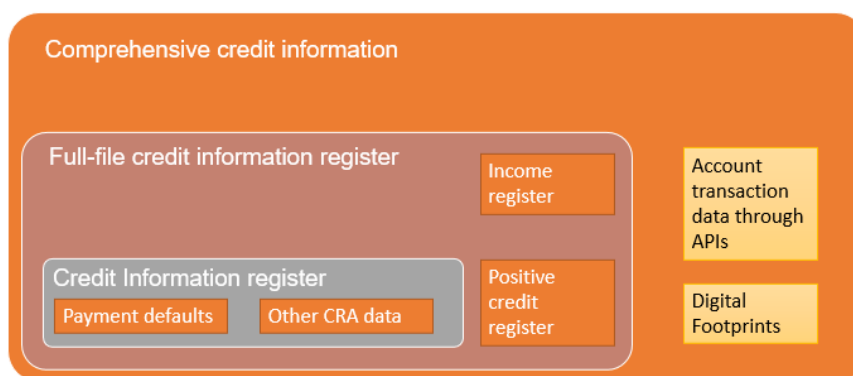
industry might be in sub-optimal competitive position compared to the international counterparts that have stronger competence of utilizing the cutting-edge analytics and comprehensive data.

The opportunities and challenges of utilizing the alternative data in the assessment of creditworthiness are presented below in the table 4.

<b>Opportunities</b>	<b>Challenges</b>
Automated account transactional data	Impractical implementation of PSD2
Open data economy	Legal and technical constraints
Anti-financial crime prevention	Innovative way of thinking

*Table 4. Summary of the opportunities and challenges of alternative data in the assessment of creditworthiness.*

As a synthesis, combining the researched data sets: the current credit information register, the full-file credit register and the possible alternative data sets all together could be formed as comprehensive credit information described below in the table 5.



*Table 5: Comprehensive credit information. Writer's own elaboration based on World Bank Group (2019 p.6), Kontkanen (2018, p.24), Siddiqi (2017), Saarenpää (2013) and the interview results.*

## 6 CONCLUSIONS

As learnt, the methods to assess the creditworthiness might be adequate today but not perhaps optimal in the future. The development of the full-file credit reporting is under investigation and the operational model will be formed in a few years. This change is highly anticipated in the financial sector. After the implementation of the full-file credit register, financial service providers could utilize a substantial digital database that will help the companies in their credit decision-making, reduce information asymmetry and improve the customer experience.

The development is also undergoing process in the EU level as the credit reporting could be more harmonized and leading to expansion of available data. The EU level legislation can enhance the development of the creditworthiness assessment. For example, the PSD2 regulation can facilitate the data sharing within the financial network enabling customers to open their bank account data for the service providers through the APIs. According to the thesis findings, account transaction data is interesting and possible alternative new data set which could be used in the credit assessment. However, there are many practical matters that slow down the development since the PSD2 might implement poorly in the efficient and customer friendly credit process. Also the account data might be difficult to be specified to a certain person if the account has multiple owners which can lead to inaccurate decision-making.

Utilizing other types of alternative data in the credit granting such as, the digital footprints, might be distant from our current society. Naturally, using sensitive personal data that is not directly linked to financial information, raises concerns related to the privacy rules and personal data security. Also retrieving, processing, and analyzing such a vast data set, requires computing power that might not be realistic with limited resources. However, it is not only the legislation nor companies' resources that set the boundaries. People as consumers are in a key role in this development. The increasing awareness of data sharing over the generations can enhance the open data economy.

All in all, based on the conducted research, the current methods that assess the creditworthiness are far from the highly advanced world described by many data scientist. It seems

that the struggles are still around the same underlying problems of the information asymmetry and the data quality. The full-file credit register is a major step towards the more efficient and customer friendly credit decision-making. Nevertheless, the financial companies might need to seek beyond conventional data and be aware of the changing operating environment and the future megatrends.

## **7 RECOMMENDATIONS FOR FUTURE RESEARCH**

The topic of the thesis stems from the financial concepts that are often-times statistical and quantitative in nature. The relevant credit assessment related research seem to lean solely on the quantitative methods. However, understanding the varieties and techniques of the modern social science research and the ability to analyze both quantitative and qualitative data are crucial in the global financial industry (Paterson et al. 2016). Correspondingly, when searching the possible recent and related studies, it was apparent that qualitative finance research surpasses many social science boundaries such customer behavior. On these grounds it could be beneficial to study more deeply the consumer perception on the open banking and open data economy. Furthermore it could be beneficial to research the practical implementation of the PSD2 related requirements in the automated credit decision-making in order to harness the substantial data that the bank account contains. In addition, the impact of the full-file credit can be studied after the implementation. That will naturally provide an interesting research area.

## REFERENCES

- Aaltonen, M. Koskinen, K. 9.5.2019. Uudet keinot tarpeenkulutusluottojen hillitsemiseksi Eurojatalous.fi — Suomen Pankin ajankohtaisia artikkeleita taloudesta. [https://www.eurojatalous.fi/fi/2019/2/uudet-keinot-tarpeen-kulutuluottojen-hillitsemiseksi/?print\\_page=1](https://www.eurojatalous.fi/fi/2019/2/uudet-keinot-tarpeen-kulutuluottojen-hillitsemiseksi/?print_page=1) (accessed 7.4.2020)
- Abdou, H.A., Pointon, J., 2011. CREDIT SCORING, STATISTICAL TECHNIQUES AND EVALUATION CRITERIA: A REVIEW OF THE LITERATURE: CREDIT SCORING, TECHNIQUES & EVALUATION CRITERIA: A LITERATURE REVIEW. *Intell. Sys. Acc. Fin. Mgmt.* 18, 59–88. <https://doi.org/10.1002/isaf.325>
- Aliyev, F. 2019. REVIEW OF THE CONSUMER CREDIT DIRECTIVE. The European Consumer Organisation. Also available: [https://www.beuc.eu/publications/beuc-x-2019-019\\_review\\_of\\_the\\_consumer\\_credit\\_directive.pdf](https://www.beuc.eu/publications/beuc-x-2019-019_review_of_the_consumer_credit_directive.pdf)
- Anderson, R., 2007. *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation* (OUP Catalogue). Oxford University Press. Also available at: <https://archive.org/details/TheCreditScoringToolkitTheoryAndPracticeForRetailCreditRiskMgmt.DecisionAutomati/page/n63/mode/2up?q=label>
- Association for Financial Markets in Europe. 2020. State of Play: Status of European Regulatory developments on Sustainable Finance. <https://www.afme.eu/Reports/Publications/Details/State-of-Play-Status-of-European-Regulatory-developments-on-Sustainable-Finance> (accessed 11.20.20).
- Bank of Finland. 2021. Suomessa toimivien luottolaitosten markkinaosuudet [https://www.suomenpankki.fi/fi/Tilastot/rahalaitosten-tase-lainat-ja-talletukset-ja-korot/taulukot/rati-taulukot-fi/markkinaosuudet\\_luottolaitokset\\_fi/](https://www.suomenpankki.fi/fi/Tilastot/rahalaitosten-tase-lainat-ja-talletukset-ja-korot/taulukot/rati-taulukot-fi/markkinaosuudet_luottolaitokset_fi/) (accessed 7.5.2021)
- Bertsimas, D., Thiele, A., 2006. Robust and Data-Driven Optimization: Modern Decision Making Under Uncertainty, in: Johnson, M.P., Norman, B., Secomandi, N., Gray, P., Greenberg, H.J. (Eds.), *Models, Methods, and Applications for Innovative Decision Making*. INFORMS, pp. 95–122. <https://doi.org/10.1287/educ.1063.0022>
- Caplinska, A., Tvaronavičienė, M., 2020. Creditworthiness place in Credit Theory and methods of its evaluation. *JESI* 7, 2542–2555. [https://doi.org/10.9770/jesi.2020.7.3\(72\)](https://doi.org/10.9770/jesi.2020.7.3(72))
- Christopher Curley. Many countries don't use credit scores like the US — here's how they determine your worth. *Business Insider*. 20.8.2018. Also available: <https://www.businessinsider.com.au/credit-score-around-the-world-2018-8> (referred 1.4.2020)

- Cornelli, G., Frost, J., Gambacorta, L., Rau, R., Wardrop, R., Ziegler, T. 2020. Fintech and big tech credit: a new database. Also available at: <https://www.bis.org/publ/work887.pdf>
- Doumpos, M., Lemonakis, C., Niklis, D., Zopounidis, C., 2018. Analytical Techniques in the Assessment of Credit Risk: An Overview of Methodologies and Applications. Springer. Also available at: <https://books.google.fi/books?id=rOBwDwAAQBAJ>
- Doumpos, M., Zopounidis, C., Grigoroudis, E., 2016. Robustness Analysis in Decision Aiding, Optimization, and Analytics. Springer. Also available at: <https://books.google.fi/books?id=I7OqDAAAQBAJ>
- European Banking Authority. EPV:N OHJEET LUOTTOKELPOISUUDEN ARVIOINNISTA. EBA/GL/2015/11 19.08.2015. [https://www.finanssivalvonta.fi/globalassets/fi/saantely/maarayskoko-elma/2017/03\\_2017/2017\\_03\\_liite1.pdf](https://www.finanssivalvonta.fi/globalassets/fi/saantely/maarayskoko-elma/2017/03_2017/2017_03_liite1.pdf) (accessed 7.5.2021)
- European Central Bank. Mikä on AnaCredit? URL <https://www.ecb.europa.eu/explainers/tell-me-more/html/anacredit.fi.html> (accessed 7.4.2020).
- European Commission. COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN CENTRAL BANK, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Consumer Financial Services Action Plan: Better Products, More Choice. COM/2017/0139 final.Brussels, 23.3.2017. [https://eur-lex.europa.eu/resource.html?uri=cellar:055353bd-0fba-11e7-8a35-01aa75ed71a1.0003.02/DOC\\_1&format=PDF](https://eur-lex.europa.eu/resource.html?uri=cellar:055353bd-0fba-11e7-8a35-01aa75ed71a1.0003.02/DOC_1&format=PDF) (accessed 7.4.2020)
- European Commission. Consumer protection in financial services. URL [https://ec.europa.eu/info/live-work-travel-eu/consumers/financial-products-and-services/consumer-protection-financial-services\\_en](https://ec.europa.eu/info/live-work-travel-eu/consumers/financial-products-and-services/consumer-protection-financial-services_en) (accessed 4.22.20).
- Ferretti, F., 2014. EU Competition Law, the Consumer Interest and Data Protection: The Exchange of Consumer Information in the Retail Financial Sector. Springer.
- Ferretti, F., 2017. *The Never-Ending European Credit Data Mess*. The European Consumer Organisation. URL: [http://www.beuc.eu/publications/beuc-x-2017-111\\_the-never-ending-european-credit-data-mess.pdf](http://www.beuc.eu/publications/beuc-x-2017-111_the-never-ending-european-credit-data-mess.pdf). (Accessed 20.4.2020).
- Ferretti, F., Vandone, D., 2019. Personal Debt in Europe: The EU Financial Market and Consumer Insolvency. Cambridge University Press.
- FICO website [https://www.fico.com/en/about-us#our\\_history](https://www.fico.com/en/about-us#our_history). (accesses 20.4.2020)
- Financial Supervisory Authority. Määräykset ja ohjeet 3/2017. Luottokelpoisuuden arviointi. [https://www.finanssivalvonta.fi/globalassets/fi/saantely/maarayskoko-elma/2017/03\\_2017/2017\\_03.m1.pdf](https://www.finanssivalvonta.fi/globalassets/fi/saantely/maarayskoko-elma/2017/03_2017/2017_03.m1.pdf). (accessed 7.5.2021)

- FINLEX® - Ajantasainen lainsäädäntö: Luottotietolaki 527/2007 [WWW Document]. URL <https://finlex.fi/fi/laki/ajantasa/2007/20070527> (accessed 5.14.20).
- Finnish Government. Government Report on the Future, Part 2 Solutions to the Transformation of Work. Finnish government publication series 30/2018. <https://julkaisut.valtioneuvosto.fi/handle/10024/161091> (accessed 20.11.2020)
- Gaganis, C., Papadimitri, P., Tasiou, M., 2020. A multicriteria decision support tool for modelling bank credit ratings. Ann Oper Res. <https://doi.org/10.1007/s10479-020-03516-9>
- Gambacorta, L., Huang, Y., Qiu, H., Wang, J., 2019. How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm. Also available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3506945](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3506945)
- Ghosh, A., 2012. Managing Risks in Commercial and Retail Banking. John Wiley & Sons, Incorporated, New York, SINGAPORE. Also available at: <https://ebookcentral-proquest-com.ezproxy.arcada.fi:2443/lib/arcada-ebooks/detail.action?docID=818434&query=amalendu#>
- Hibbeln, M., Norden, L., Usselmann, P., Gürtler, M., 2019. Informational synergies in consumer credit. Journal of Financial Intermediation 100831. <https://doi.org/10.1016/j.jfi.2019.100831>
- International Monetary Fund. Monetary and Capital Markets Department, 2017. Finland: Financial sector Assessment Program. INTERNATIONAL MONETARY FUND, Washington, D. C., UNITED STATES. Available from: ProQuest Ebook Central. [4.5.2020].
- Israël, J.-M., Damia, V., Bonci, R., Watfe, G., 2017. A magnifying glass for analysing credit in the euro area 44. Also available at: <https://www.ecb.europa.eu/pub/pdf/scpops/ecb.op187.en.pdf>
- Kaufman, M., Halper, F., Nugent, A., Hurwitz, J.S., Nugent, A., Hurwitz, J.S., 2013. Big Data for Dummies. John Wiley & Sons, Incorporated, Somerset, UNITED STATES.
- Kontkanen, Erkki; Lång, Jukka. Selvitys positiivisia luottotietoja koskevan järjestelmän edellytyksistä. Oikeusministeriö. 4.9.2018. Also available at: <http://urn.fi/URN:ISBN:978-952-259-701-4>
- Koskinen, I. 2018. Koneoppiminen ja EU:n yleisen tietosuojasetuksen vaatimus lainmukaisesta, kohtuullisesta ja läpinäkyvästä käsittelystä. Asianajotoimisto Mercurius. URL <https://merkuriuslaw.fi/koneoppiminen-ja-eun-yleisen-tietosuojasetuksen-vaatimus-lainmukaisesta-kohtuullisesta-ja-lapinakyvasta-kasittelysta/> (accessed 5.14.20).



- Koste, Otto-Wille. Lehtovuori, Panu. Neuvonen, Aleks. Schmidt-Thomé, Kaisa. MIKSI SUOMEN KAUPUNGISTUMINEN JATKUU?. URMI. Poliitikkipaperi 2020:1. Helsinki 6.2.2020. [https://www.aka.fi/globalassets/33stn/materiaaleja/politiikkasuositukset/2020-02\\_miksi-suomen-kaupungistuminen-jatkuu-.pdf](https://www.aka.fi/globalassets/33stn/materiaaleja/politiikkasuositukset/2020-02_miksi-suomen-kaupungistuminen-jatkuu-.pdf) (Accessed 24.4.2020).
- Livada, C.K., 2019. Assessment of consumers' creditworthiness. ERA Forum 20, 225–236. <https://doi.org/10.1007/s12027-019-00574-w>
- Minelli, M., Chambers, M., Dhiraj, A., Rajaram, D., Chambers, M., 2012. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses. Wiley, Somerset, UNITED STATES. Available from: ProQuest Ebook Central. [4.5.2020].
- Mligo, E.S., 2016. Introduction to Research Methods and Report Writing: A Practical Guide for Students and Researchers in Social Sciences and the Humanities. Wipf and Stock Publishers, Eugene, UNITED STATES.
- Monedo. URL <https://www.monedo.com/> (accessed 14.4.2020).
- Nykänen. M. 16.5.2019. Kotitalouksien velkaantuminen on riski koko kansantaloudelle. Eurojalous.fi — Suomen Pankin ajankohtaisia artikkeleita taloudesta. [https://www.eurojalous.fi/fi/2018/2/kotitalouksien-velkaantuminen-on-riski-koko-kansantaloudelle/?print\\_page=1](https://www.eurojalous.fi/fi/2018/2/kotitalouksien-velkaantuminen-on-riski-koko-kansantaloudelle/?print_page=1) (accessed 20.11.2020)
- Packin, N.G. 2019. Social Credit: Much More Than Your Traditional Financial Credit Score Data. Forbes. Available at: <https://www.forbes.com/sites/nizangpackin/2019/12/13/social-credit-much-more-than-your-traditional-financial-credit-score-data/> (accessed 14.4.2020).
- Parkkila, Kaisu. (2020). OP Tech Podcast #23 – OP:n Credit Engine ja automaattinen päätöksenteko. Available at: <https://optech.transistor.fm/episodes/23-op-n-credit-engine-ja-automattinen-paatoksenteko>. (accessed 23.4.2020).
- Paterson, A., Leung, D., 2016. Research Methods for Accounting and Finance. Goodfellow Publishers Ltd.
- Pickert, J., 2017. Risk Assessment of unsecured loans – example of entering a new market. The Central European Review of Economics and Management 1. <https://doi.org/10.29015/cerem.449>
- Portilla, Andrés; Vazquez, Jaime; Harreis, Holger; Pancaldi, Luca; Rowshankish, Kayvaun; Samandari, Hamid; Dash Rajdeep. October 2017. *The future of risk management in the digital era*. The Institute of International Finance and McKinsey & Company, Inc. URL <https://www.mckinsey.com/~media/McKinsey/Business%20Functions/Risk/Our%20Insights/The%20future%20of%20risk%20management%20in%20the%20digital%20era/Future-of-risk-management-in-the-digital-era-IIF-and-McKinsey.ashx> (accessed 20.4.2020)

- Provost, F., Fawcett, T., 2013. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data* 1, 51–59. <https://doi.org/10.1089/big.2013.1508>
- Puri, M., Berg, T., Burg, V., Gombović, A., 2018. On the Rise of the FinTechs - Credit Scoring using Digital Footprints. *SSRN Journal*. <https://doi.org/10.2139/ssrn.3259901>
- Qiu, W., Rudkin, S., Dłotko, P., 2020. Refining Understanding of Corporate Failure through a Topological Data Analysis Mapping of Altman's Z-Score Model. *Expert Systems with Applications*. Available from: <https://doi.org/10.1016/j.eswa.2020.113475>
- Quirk, P., Rothchild, J.A., 2018. Consumer protection and the Internet. *Handbook of Research on International Consumer Law, Second Edition*. Also available: <https://www.elgaronline.com/view/ed-coll/9781785368202/9781785368202.00017.xml>
- Ramsay, I., Williams, T., 2020. Peering Forward, 10 Years After: International Policy and Consumer Credit Regulation. *J Consum Policy* 43, 209–226. <https://doi.org/10.1007/s10603-019-09436-x>
- Rosenblatt, E., Wu, W., 2020. *Credit Data and Scoring: The First Triumph of Big Data and Big Algorithms*. Elsevier Science & Technology, San Diego, UNITED STATES.
- Saarenpää, Ahti. Näkökohtia luottotietojen kokoamisen ja käytön sääntelystä. 18.4.2013. Also available at: <http://urn.fi/URN:ISBN:978-952-259-290-3>
- Salamina, Luz Maria; Chhabra, Pratibha; Sankaranarayan, Shalini; Masunda, Collen. 2019. *Disruptive Technologies in the Credit Information Sharing Industry : Developments and Implications (English)*. Fintech Note; No. 3. Washington, D.C. : World Bank Group. <http://documents.worldbank.org/curated/en/587611557814694439/Disruptive-Technologies-in-the-Credit-Information-Sharing-Industry-Developments-and-Implications> (accessed 11.4.2020)
- Saunders, M., Lewis, P., Thornhill, A., Bristow, A., 2019. “Research Methods for Business Students” Chapter 4: Understanding research philosophy and approaches to theory development. pp. 128–171. Also available: [https://www.researchgate.net/publication/330760964\\_Research\\_Methods\\_for\\_Business\\_Students\\_Chapter\\_4\\_Understanding\\_research\\_philosophy\\_and\\_approaches\\_to\\_theory\\_development](https://www.researchgate.net/publication/330760964_Research_Methods_for_Business_Students_Chapter_4_Understanding_research_philosophy_and_approaches_to_theory_development)
- Siddiqi, N., 2017. *Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards*. John Wiley & Sons, Incorporated, Somerset, UNITED STATES.
- Siroya, S. 2016. A smart loan for people with no credit history (yet). TED Talk. Also available: [https://www.ted.com/talks/shivani\\_siroya\\_a\\_smart\\_loan\\_for\\_people\\_with\\_no\\_credit\\_history\\_yet](https://www.ted.com/talks/shivani_siroya_a_smart_loan_for_people_with_no_credit_history_yet)

- Stiglitz, Joseph E., and Andrew Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *The American Economic Review* 71(3):393–410.
- Suomen Asiakastieto Oy. Positiiviset luottotiedot kuluttajaluottojen kyselyjärjestelmässä - Suomen Asiakastieto Oy. <https://www.asiakastieto.fi/web/fi/positiivinen-tieto/positiiviset-luottotiedot-kuluttajaluottojen-kyselyjarjestelmassa.html> (accessed 7.4.2020).
- Suomen Pankki. Vakuudettomien kulutusluottojen koroissa on suuria eroja luottolaitosten välillä. Suomen Pankki. URL <https://www.suomenpankki.fi/fi/Tilastot/rahalaitosten-tase-lainat-ja-talletukset-ja-korot/tiedotehistoria/2019/vakuudettomien-kulutusluottojen-koroissa-on-suuria-eroja-luottolaitosten-valilla/> (accessed 4.7.2020).
- The Association of Consumer Credit Information Suppliers. ACCIS 2017 Survey of Members. Analysis of Credit Reporting in Europe. March 2018. Also available at: [https://accis.eu/wp-content/uploads/2018/11/AUG18\\_ACCIS-Survey-of-Members-2017\\_FINAL.pdf](https://accis.eu/wp-content/uploads/2018/11/AUG18_ACCIS-Survey-of-Members-2017_FINAL.pdf) (accessed 1.4.2020)
- The Association of Consumer Credit Information Suppliers. PRINCIPLES ON CREDIT-WORTHINESS ASSESSMENTS. 6 November 2018. <https://accis.eu/wp-content/uploads/2018/11/Final-Proofread-Principles-on-Creditworthiness-Assessments-and-the-Role-of-CRAs.pdf> (accessed 16.4.2021)
- The Ministry of Justice. Positiivisen luottotietorekisterin valmisteleminen ja maksuhäiriömerkintöjen säilytysaikojen lyhentäminen. Oikeusministeriö. URL <https://oikeusministerio.fi/hanke?tunnus=OM022:00/2020> (accessed 6.4.2020).
- Uudet keinot tarpeen kulutusluottojen hillitsemiseksi [WWW Document], n.d. . Euro ja talous. URL <https://www.eurojatalous.fi/fi/2019/2/uudet-keinot-tarpeen-kulutusluottojen-hillitsemiseksi/> (accessed 7.4.2020).
- Verbeke, W., Baesens, B., Bravo, C., 2017. Profit Driven Business Analytics: A Practitioner's Guide to Transforming Big Data into Added Value. John Wiley & Sons. Available from: Google Books. URL: [https://books.google.fi/books?id=ahk0DwAAQBAJ&dq=profit+driven+business+analytics&hl=fi&source=gbs\\_navlinks\\_s](https://books.google.fi/books?id=ahk0DwAAQBAJ&dq=profit+driven+business+analytics&hl=fi&source=gbs_navlinks_s)
- Wachter, S., Mittelstadt, B., Floridi, L., 2016. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. <https://doi.org/10.1093/idpl/ix005>
- Wirman, J. 21.12.2020. Digipankit haastavat konttoripankit kulutusluottojen tarjonnassa. Eurojatalous.fi — Suomen Pankin ajankohtaisia artikkeleita taloudesta. [https://www.eurojatalous.fi/fi/blogit/2020/digipankit-haastavat-konttoripankit-kulutusluottojen-tarjonnassa/?print\\_page=1](https://www.eurojatalous.fi/fi/blogit/2020/digipankit-haastavat-konttoripankit-kulutusluottojen-tarjonnassa/?print_page=1) (accessed 7.5.2021)
- World Bank. 2019. *Credit Reporting Knowledge Guide 2019 (English)*. Washington, D.C. : World Bank Group.

<http://documents.worldbank.org/curated/en/262691559115855583/Credit-Reporting-Knowledge-Guide-2019> (accessed 10.4.2020)

Your Europe. Data protection. Your Europe - Business. URL [https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index\\_en.htm](https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_en.htm) (accessed 29.4.2020).

Zest AI. URL <https://www.zest.ai/> (accessed 14.4.2021).

Zuboff, S., 2019. Surveillance Capitalism and the Challenge of Collective Action. *New Labor Forum* 28, 10–29. <https://doi.org/10.1177/1095796018819461>

## APPENDICES

### Interview questions

ENG

	Questions	Respondents	
	<b>Background information</b>		
1.	Name, profession, position		
2.	Name and position of the company/organization in the finance sector		
	<b>Business perspective</b>		
3.	What are the current methods of assessing the creditworthiness of a credit applicant? <ul style="list-style-type: none"><li>• What do you want to assess?</li><li>• Why do you assess?</li></ul>		
4.	What are the sources the data are collected from?		
5.	Is the collected data sufficient and why? If not, why?		

6.	What kind of data would you like to utilize?		
7.	What are the barriers of utilizing the wanted data?		
8.	What is your view on the “positive credit register” that is under investigation by the Ministry of Justice?		
9.	<p>What are the possible impacts of new alternative data* sources on the;</p> <ul style="list-style-type: none"> <li>• credit risk assessment</li> <li>• lending business</li> <li>• competition?</li> </ul>		
10.	What are the future possibilities of new data assessment sources/ methods?		
	<b>Regulatory perspective</b>		
11.	In the regulatory context, what are the possibilities and challenges of developing the current consumer credit risk assessment methods?		
12.	What is the current perception of alternative data sources as an		

	appropriate data in creditworthiness assessment?		
13.	Any other additional comments, follow-up?		

## Interview questions FIN

	<b>Kysymys</b>		
	<b>Taustatieto</b>		
1.	Nimi, ammatti, asema		
2.	Yrityksen nimi ja asema finanssisektorilla		
	<b>Liiketoimintanäkökulma</b>		
3.	Mitkä ovat nykyisin henkilöasiakkaan luottokelpoisuuden arviointimenetelmät? <ul style="list-style-type: none"><li>• Mitä asioita haluatte arvioida?</li><li>• Miksi haluatte arvioida?</li></ul>		
4.	Mistä lähteistä data kerätään?		
5.	Onko kerättävä data riittävää? Perustelut		
6.	Millaista dataa haluaisitte hyödyntää?		



7.	Mitkä ovat esteet halutun datan hyödyntämiseen?		
8.	Oikeusministeriö aloitti hiljattain ”positiivisen luottorekisterin” säädösvalmistelun, Mikä on näkemyksesi ko. rekisteristä?		
9.	Millaisia vaikutuksia vaihtoehtoisella datalla* olisi; <ul style="list-style-type: none"> <li>• luottokelpoisuuden määrittämiseen</li> <li>• liiketoimintaan</li> <li>• kilpailuun?</li> </ul>		
10.	Mitkä ovat tulevaisuuden mahdollisuudet ja haasteet vaihtoehtoisen datan hyödyntämiseen?		
	<b>Regulaatio näkökulma</b>		
11.	Lainsäädännön näkökulmasta mitkä ovat mahdollisuudet ja haasteet kehittää luottokelpoisuuden määrittäystä?		
12.	Millaisia ehtoja Suomen omat kansalliset lait asettavat		

	luottokelpoisuuden määrittämiseen?		
13.	Mikä on yleinen suhtautuminen vaihtoehtoisen datan hyödyntämiseen luottokelpoisuuden määrittämisessä?		