Exploring the Evolution and Impact of Credit Scoring in the European Union: An Analysis of Alternative Data, Individual Consumer Data, and Regulatory Frameworks

Author: Marcus Müller
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT

This study aims to explore the potential of using alternative data, such as social media profiles and web browsing history, to assess the credit risk of individuals with minimal or no credit history. The study aims to address the existing knowledge gap and provide insights for policymakers, financial institutions, and consumers in the context of the usage of alternative data sources in the credit scoring landscape.
The paper will also discuss the role of Credit Rating Agencies (CRAs) in the European Union (EU), with a focus on the structure of credit reporting agencies. The analysis will highlight how we can divide the types of Credit Rating Agencies depending on the information they collect and process.
Furthermore, the paper will examine the current laws related to credit risk scoring in the EU, including the General Data Protection Regulation (GDPR), the Consumer Credit Directive, and the Proposal for a Regulation on Artificial Intelligence (AI). The study will analyze how credit risk scoring, particularly when utilizing alternative data sources, fits within in the current legislative context, and identify potential shortcomings in EU legislation in protecting consumers against using alternative data in credit scoring.
By investigating the potential findings and comparisons of alternative data sources in credit risk assessment, this paper aims to contribute to the ongoing discussions on credit scoring transparency, consumer protection, and the promotion of a stable and equitable financial system in the European Union.

Graduation Committee members:

Marcos Machado
Jörg Osterrieder

Keywords
Alternative Data, Credit Risk, Credit Rating Agencies, EU Legislation, GDPR, Transparency, Consumer Protection, Financial System

1. INTRODUCTION

Assessing credit risk is a critical component of lending decisions, as it determines the creditworthiness of individual consumers. In the European Union (EU), the credit rating landscape involves a combination of internal bank customer credit systems and credit reporting agencies (CRAs) (Stellinga, 2019). These systems are instrumental in determining individual customers creditworthiness, with many lenders relying on third-party CRAs for comprehensive credit risk evaluations of individual customers.

This paper aims to explore the impact of alternative data on evaluating individual credit risk in the EU. Specifically, it will address the following research questions:

1. How do CRAs approach credit risk assessment?
2. What is the effect of data sources on credit risk predictions?
3. What are the practical implications of using alternative data in credit risk evaluation?
4. What rights do EU individuals have regarding credit risk assessment with the use of alternative data?

To address these questions, the paper will first examine the existing landscape of credit rating systems in the EU, highlighting the differences in credit reporting methodologies across member states. Building on this, it will delve into the distinction between big data and alternative data, discussing how each type of data can impact credit risk predictions. The paper will also explore the practical implications of using alternative data in credit risk assessments, focusing on the potential benefits and challenges associated with the usage of alternative data in credit risk assessments of individuals.

Furthermore, the paper will examine the regulatory landscape surrounding the usage of alternative data in credit risk assessments, specifically analyzing the General Data Protection Regulation (GDPR), the Consumer Credit Directive, and the proposed regulation on harmonized rules for artificial intelligence. The analysis will aim to uncover potential shortcomings in the current EU legislation and identify areas where improvements could be made to protect consumers and ensure transparency and fairness in credit risk assessments.

Overall, this paper seeks to provide a comprehensive understanding of the consequences of using alternative data in assessing credit risk for individual consumers in the EU. The findings will contribute to the ongoing debates surrounding credit risk assessment methodologies and inform future policy decisions to enhance transparency, fairness, and consumer protection in the EU's credit landscape.

The rest of the paper is organized as follows. Section 2 contains a literature review that explores the existing research on credit risk assessment methodologies, the use of alternative data in credit scoring, and the regulatory landscape in the EU. Section 3 presents the findings of the study, including an analysis of the different types of credit reporting agencies in the EU, the potential benefits and challenges of using alternative data in credit risk assessment, and the implications of EU legislation on the use of alternative data in credit scoring. Section 4 discusses the implications of the findings and provides recommendations for policymakers, financial institutions, and consumers. Finally, Section 5 concludes the paper by summarizing the main findings and highlighting the importance of transparency, fairness, and consumer protection in the EU's credit landscape.

2. METHODOLOGY

The study adopted a predominantly qualitative research approach to enhance understanding of the credit rating landscape in the European Union. This approach was used to understand the functionality of internal bank customer credit systems, Credit Rating Agencies (CRAs), and the potential usage of alternative data sources. A part of this study delved into examining how alternative data, such as information sourced from social media profiles and web browsing history, was utilized in evaluating the credit risk of individuals, specifically those with little or no credit history.

A narrative literature review was conducted as part of the study, serving as a vital tool to ensure the breadth and depth of the research.

Scopus, Web of Science, and Google Scholar were the databases used to source articles related to credit scoring, alternative data, and its associated challenges.

Search queries included in the literature review process were:

1. ("credit scoring" OR "credit rating") AND ("European Union" OR "EU")
2. ("credit scoring" OR "credit rating") AND ("methodology" OR "models" OR "algorithms")
3. ("credit scoring" OR "credit rating") AND ("fairness" OR "bias" OR "discrimination")
4. ("credit scoring" OR "credit rating") AND ("alternative data" OR "non-traditional data")
5. ("credit scoring" OR "credit rating") AND ("GDPR" OR "data protection" OR "privacy")
6. ("credit scoring" OR "credit rating") AND ("financial inclusion" OR "access to credit")

During the literature review, these search queries were regularly tweaked and revised to make sure all relevant articles could be found. Special attention was paid to literature on the structure and role of CRAs in the EU, the utilization of alternative data sources in credit risk scoring, and current laws related to credit risk scoring in the EU including the General Data Protection Regulation (GDPR), the Consumer Credit Directive, and the Proposal for a Regulation on Artificial Intelligence (AI). This was done to provide a comprehensive view of the legislative context surrounding credit scoring in the EU and to pinpoint potential shortcomings in EU legislation in terms of protecting consumers when alternative data is used in credit scoring.

3. FINDINGS

3.1 Credit risk

Credit risk assessment is a key determinant of individual borrowers creditworthiness, traditionally involving internal bank systems and Credit Reporting Agencies (CRAs) (White, 2013). However, the advent of Big Data and Alternative Data provides new avenues for credit risk assessment. This new data ranges from online behavior to social media interactions (Kulkarni & Dhage, 2019), using digital footprints (Berg et al., 2020), and web browsing patterns (Rozo, Crook, & Andreeva, 2023).

While these new data sources enhance credit risk predictions, they also pose challenges, such as issues surrounding data privacy, security, and fairness (Langenbucher & Corcoran, 2021). The use of such data could potentially violate the EU's General Data Protection Regulation (GDPR) (Goodman & Flaxman, 2017).

Existing legislation like the Consumer Credit Directive focuses on consumer protection and financial literacy in the era of traditional credit scoring methods, but there are gaps in the current legislation concerning modern credit risk assessment methodologies (Langenbucher & Corcoran, Patrick, 2021). The rise of alternative data thus calls for continuous review and adaptation of EU legislation to protect consumers and ensure transparency and fairness in credit risk assessments.

In summary, the incorporation of alternative data in credit risk assessments presents both opportunities and challenges. As such, EU legislation should evolve to meet these emerging trends, ensuring consumer rights and protection are upheld in the evolving credit risk assessment landscape.

3.1.1 Credit Ratings and the Role of Credit Rating Agencies (CRAs) in Europe

To better understand how alternative data may affect credit risk and scoring in the EU, we have to understand how the credit rating agencies and systems work in the context of credit scoring and classify them by how they work. We can first classify the credit risk analysis from where the financial institutions get their information from, which is from internal and external sources (Grunert et al., 2005).

Internal Credit Rating Systems are one way in how banking and financial institutions can understand and analyze a customer's credit risk. These systems utilize the data gathered from customers during their interactions with the institution, including account balances, transaction histories, repayment behaviors, and other financial activities. This data, often obtained through the customer's existing relationship with the bank, forms the basis of an internal credit rating that reflects the customer's credit risk in relation to their direct interactions with the institution.

This approach has its benefits. By leveraging their internal credit scoring systems, financial institutions can tailor their credit scoring model to their specific customer base and risk tolerance. The use of internal data can also expedite loan application processing as there's no need to wait for data from an external credit rating agency (Mileris & Boguslauskas, 2011). The development of large digital databases offering individuals financial histories has further enhanced these internal systems, allowing banks to have an in-depth understanding of a customer's financial behavior. However, these systems can also present a narrower view since they are based on interactions with one institution and do not account for the customer's dealings with other financial institutions.

In particular, internal credit scoring systems may disadvantage new customers who lack history with financial institutions. Due to these limitations, many financial institutions combine their internal data with information from third-party vendors of standardized financial information or external credit rating agencies. These third-party resources can provide additional insights about certain types of loans, especially those to individuals, new customers, or larger loan amounts. Therefore, by using a mix of internal and external data, financial insitutions such as banks can get a more comprehensive assessment of a customer's credit risk.

External Credit Rating Systems: These external credit agencies are used by financial institutions such as banks to complement their internal credit system data to get a better picture of a customer and make a better risk assessment. Some of these agencies in the EU are CIRBE, Schufa, ASNEF, FICP, Equifax, and TransUnion. These agencies are not standardized, vary across countries and by the information that they hold, and how they process their information (De Haan & Amtenbrink, 2011).

In the current banking landscape, a combination of both systems is widely used. The reason for this hybrid approach is that it allows banks to gain a comprehensive understanding of a borrower's credit risk. By combining internal data with external credit scores, banks can form a better perspective of a borrower's creditworthiness. This integrated approach is instrumental in making informed lending decisions.
Nevertheless, within the context of the Credit Reporting Agencies in the EU we can differentiate three types of agencies/bureaus based on the type of information they deal with. We can sort them by if they provide scoring or a risk profile, if they collect only negative information of customers, and if they collect all types of credit movements, positive and negative

1. Negative-Information Bureaus: This category of agencies mainly accumulates and reports negative credit events, such as missed payments, defaults, or bankruptcies. These databases serve as a warning to potential lenders about individuals who have previously demonstrated high credit risk. The Asociación Nacional de Establecimientos Financieros de Crédito (ASNEF) in Spain is a good example of a Negative-Information agency (Carrasco Conesa, 2017). Another instance is the "Fichier des Incidents de Remboursement des Crédits aux Particuliers" (FICP, 2013) in France, maintained by the Banque de France, which records all incidents of non-payment of loans by individuals.
2. Comprehensive Credit Bureaus: These agencies gather a wide range of both positive and negative financial data, such as the regularity of payments on loans and credit cards, as well as instances of late payments or defaults (Equifax, n.d.). They generate credit scores based on this data, helping lenders to assess an individual's creditworthiness. The Schutzgemeinschaft für allgemeine Kreditsicherung (SCHUFA) in Germany is a good example of a Comprehensive Credit Bureau (Horen, n.d.). Similarly, in the Netherlands, the Bureau Krediet Registratie (BKR) records all loans and credits and provides a risk profile, which serves a similar function to a credit score (Ortlepp, 2019).
3. Credit Registries: Typically, public or government-managed entities, these agencies maintain records of individuals' credit obligations and provide this information to lenders. They generally do not produce a credit score but offer comprehensive data on an individual's credit history for lenders to assess risk. The Central de Información de Riesgos del Banco de España (CIRBE) is an example of this type of agency. CIRBE is regulated by the Banco de España (Spain's national central bank) and its main objectives are to give Banco de España data and to provide banks and other entities that give out loans with the necessary information to develop their credit risk assessment. Contrary to the ASNEF, CIRBE does not only collect negative credit events such as defaults, but includes debts, credits, loans, and guarantees of the clients of financial institutions, whether they are up-to-date with payments or not. Banks and other financial institutions must declare almost all their credit risks and the holders to whom they correspond to CIRBE monthly. Any individual or legal entity can access all the information in their name in the CIRBE for free (as per GDPR guidelines) (Carrasco Conesa, 2017).

In this context, we might observe how different countries in the EU implement and utilize these three types of agencies or bureaus, based on the analysis of some, but not all, countries.

For example, Spain has both a Negative-Information Agencies, ASNEF, and a Credit Registry, CIRBE. ASNEF, which is privately owned, mainly records negative credit information such as defaults and missed payments. Conversely, the publicly managed CIRBE collects a broader range of credit data, offering a fuller understanding of a customer's credit history (Carrasco Conesa, 2017).

The primary credit rating system in France is a Negative-Information Bureau known as FICP, or "Fichier National des Incidents de Remboursement des Crédits aux Particuliers". Managed by the country's central bank, the Banque de France, the FICP exclusively logs negative credit events (FICP, 2023). Payments that are late by 90 days or more, along with any credit-related legal or administrative procedures, are stored. The access to this database is limited to credit companies and individual debtors who want to review their own records (Trumbull, 2008).

On the other hand, some countries primarily rely on Comprehensive Credit Bureaus, like Germany's SCHUFA and the Netherlands Bureau Krediet Registratie (BKR). These privately owned entities compile both positive and negative credit data and generate credit scores or risk profiles (Horen, n.d.). In countries where these Comprehensive Credit Bureaus exist, there is often less reliance on publicly owned agencies, as the former already provides a detailed overview of credit behavior.

This diversity in credit reporting across the EU using Negative-Information Bureaus, Comprehensive Credit Bureaus, and Credit Registries signifies the complexity of credit risk assessment and management in the European Union. It further highlights the unique ways each country gathers and analyzes data to generate credit scores and assess creditworthiness. These traditional methods of credit reporting have their strengths. However, the increasing development of technology and the vast amount of data available suggest new possibilities for credit risk analysis. In particular, we're seeing a growing interest in leveraging big data and alternative data for credit risk assessment (LexisNexis, 2013).

The diversity of traditional credit reporting agencies in the EU, including Negative-Information Bureaus, Comprehensive Credit Bureaus, and Credit Registries, paints a comprehensive picture of how credit risk is assessed and managed. With this understanding in mind, we can fully appreciate the potential of big data and alternative data. These emerging data resources can enhance, or even revolutionize, the existing methods of credit reporting and risk assessment, providing new avenues and opportunities for financial institutions.

3.2 Defining Big Data and Alternative Data in Credit Risk Context

Before we delve into how alternative data might help customers without credit history get access to credit, we have to differentiate Big Data from what we define in this paper as Alternative Data. While both are similar in the sense that they both depend on Artificial Intelligence and Machine Learning to assess and determine credit risk, the information that is processed comes from different sources.

According to Hurley et al. (2016), Big Data in credit scoring encompasses a vast amount of information about consumers' daily lives, including but not limited to their browsing habits, social media activities, and geolocation data, all of which contribute to determining credit risk. The application of Big Data extends to multiple sources of information, including structured data such as transaction records and customer service interactions, as well as unstructured data such as social media posts and geolocation data.

By processing and analyzing this data, lenders can obtain a richer understanding of a consumer's creditworthiness, thereby enabling more informed credit risk assessments.

There are significant advantages to using big data to assess credit risk, especially in its capacity to deliver diverse insights and predictive power. The broad range of information that big data offers can provide a comprehensive picture of a consumer's financial behavior, while machine learning and AI tools can analyze past patterns to make predictions about future behaviors (Hurley, Mikella & Adebayo, Julius, 2016.).

However, the use of big data is not without its challenges. Concerns among them are the significant privacy and security concerns it raises, given the personal and sensitive nature of the data it often involves. Additionally, algorithms that rely on big data for credit risk scoring can unintentionally discriminate based on factors that are associated with protected characteristics such as race, gender, or age. Furthermore, Hurley, Mikella & Adebayo, Julius, (2016) warn of the potential for inaccuracies in Big Data, as well as the risk of spurious correlations due to the overabundance of data points.

In contrast, alternative data that comes directly from consumers, such as social media profiles and web browsing history, which are used to evaluate creditworthiness. This information is obtained directly from consumers, and importantly, with their explicit consent. The use of alternative data presents a different perspective and opportunity for creditors to gain unique insights into consumer behavior that may not be evident from traditional credit data or even Big Data.

The potential of alternative data for increasing financial inclusion is significant. Lenders can extend credit to individuals with thin or non-existent credit files by leveraging this data, providing an opportunity for those typically excluded by traditional credit scoring methods. This will be further delved into the following section.

The use of alternative data can also address some privacy concerns, as it is obtained with explicit consumer consent. However, this practice is not without its own challenges. The quality and relevance of alternative data can vary greatly, requiring careful interpretation and analysis. Furthermore, even with explicit consent, there are still potential privacy concerns related to the use of personal data from sources such as social media.

Nevertheless, both Big Data and Alternative Data involve handling large and complex data sets and information, requiring machine learning and IA for their processing. Additionally, they both present the risk of introducing unintentional bias into credit decisions if not used carefully. However, as stated before, there are key differences between the two. In particular, while alternative data is obtained with explicit consumer consent, the use of big data may not always be transparent or involve explicit consent. Furthermore, big data involves much larger and more diverse datasets, while alternative data typically has a narrower focus on a specific set of data types and sources.

In the following sections, the focus will be on the potential of alternative data in credit scoring. This includes its advantages, challenges, and ethical considerations, particularly in the context of evolving data protection laws such as those in the EU and the context of how traditional credit scoring agencies from section 3.3 fit in with this development.

3.3 Practical Applications of Alternative Data in Credit Risk Assessment

Before continuing, it's important to highlight how the use of alternative data in credit risk evaluation represents potential benefits for all involved. For lenders, this novel methodology unlocks a whole new world of potential customers - individuals who, although creditworthy, have been largely invisible to traditional credit systems due to their limited or absent credit history (LexisNexis, 2013; Rozo et al., 2023).

For the posible borrowers, especially young individuals or those with lacking or no credit history, this alternative data credit risk scoring can help lead the way to financial inclusion in an otherwise impossible way. The probability of a young individual possessing any substantial record in an external credit rating system tends to be quite low. This lack of credit history can be attributed to their limited financial interactions and commitments. However, certain countries, like Germany and the Netherlands, are somewhat better off with their comprehensive credit agencies, which could deduce some semblance of credit risk even in the absence of alternative data sources.

On the other hand, other nations relying heavily on Negative-Information agencies such as ASNEF (Spain), FICP (France), or Credit Registries like CIRBE (Spain) typically lack this advantage and rely extensively on established lending history. This is particularly relevant in countries with a high labor force participation rate among young individuals and where the trend of youngsters leaving their parental households is notably high. The dependency of these individuals on their family for essentials like housing and sustenance further compounds the difficulty in establishing an independent credit profile.

There has been some research conducted already into the usage of Alternative Data to estimate credit risk. Research by Rozo et al. (2016) has shown the potential of an unconventional data source - web browsing history, in credit risk assessment. It has shown that the use of web browsing history could empower lenders to extend credit offerings to potential borrowers who might be lacking a conventional credit history, thereby expanding the reach of financial services to previously untapped demographics.
Their detailed analysis has further uncovered some interesting trends: high-intensity website users tend to exhibit a higher probability of default compared to their less frequent counterparts, irrespective of other factors.

Research done by Berg et al (2016) has analyzed the potential of using digital footprints to predict the likelihood of consumer default. Observing that by just using digital footprints the information gathered can match the information gathered by credit bureaus.

In the context of using Alternative Data to assess credit risk, information asymmetry also has to be taken into account. Information asymmetry occurs when one party, typically the lender, does not have the same level of information about the borrower as the borrower does about themselves. This can lead to adverse selection, where high-risk borrowers are more likely to apply for loans, and moral hazard, where borrowers may take on more risk after the loan is granted. Traditional credit risk agencies often rely on a limited set of data, such as credit history, income, and employment status, which may not fully capture a borrower's creditworthiness, especially for those with no or limited credit history.

In this context, Choudhary (2020) discusses the impact of adverse selection in credit markets, particularly in the context of a Pakistani banking reform that reduced public information. The Akerlof unraveling effect is mentioned in the context of the misallocation of capital and the complete unraveling of the credit market due to asymmetric information in credit markets. This is based on the theoretical literature that describes how asymmetric information in credit markets causes lenders to alter what contracts they offer. Compared to an environment with full information, this can lead to the misallocation of capital and even to the complete unraveling of the credit market, as proposed by Akerlof (1970).

The use of alternative data and big data in credit scoring can provide a more comprehensive and accurate assessment of a borrower's credit risk. For example, it can capture information about a borrower's online behavior, social media activity, and other non-traditional data points that can provide insights into their creditworthiness. This can help reduce information asymmetry and improve the accuracy of credit risk assessment.

Given these developments, countries like Spain have implemented strategic measures to increase the financial independence of their young population. One such initiative involves offering guarantees to banks to finance up to 95% of a mortgage for buying a house, with the intent to ease housing accessibility (Rocha, 2023). Therefore, it becomes increasingly significant for customers, banks, and EU governments to accurately gauge individual credit risk using alternative data, while ensuring explicit consent for the usage of such data. This strategy helps optimize credit issuance decisions and the determination of interest rates.

This approach of using alternative data is therefore beneficial: it enhances access to credit for individuals with minimal or no credit history and fosters a competitive edge for lending institutions by targeting a new customer base (Berg et al., 2020; Rozo et al., 2023).

Coupling customer data with Artificial Intelligence (AI) to assess credit risk shares similarities with utilizing Big Data for the same purpose. The usage of both big data and alternative data present both advantages and potential drawbacks. On the one hand, it allows for improved precision in risk assessment, while on the other, it may inadvertently introduce biases.

3.4 The Application of Laws on Alternative Data in Credit Risk Assessment

In the context of credit risk assessment of individuals, the existence of Artificial Intelligence (AI) systems has brought about significant changes, enabling automated decision-making processes. However, this also prompts concerns regarding potential discrimination and fairness issues. To navigate this evolving landscape, understanding relevant legal frameworks is critical. This section will be focusing on three major legal instruments within the European Union (EU) that deal with this issue: the General Data Protection Regulation (GDPR), DIRECTIVE 2008/48/EC also known as the Consumer Credit Directive, and the recent updated proposal released on May 9th that deals Proposal for a Regulation on harmonized rules on artificial intelligence (AI). After that a comparison is made between the European legal framework and the American ECOA and potential legal shortcomings are discussed.

3.4.1 General Data Protection Regulation

Any organization that deals with personal data within the European Union deals with the GDPR. The GDPR, or General Data Protection Regulation, is a comprehensive data privacy law that came into effect on May 25, 2018. It was designed to harmonize data protection laws across EU member states, giving individuals greater control over their personal data while simplifying the regulatory environment for businesses.

The GDPR applies to any organization, regardless of its location, that processes the personal data of EU residents. This includes businesses, non-profits, and public sector organizations. The regulation outlines several key principles, such as lawfulness, fairness, transparency, data minimization, and accuracy. Organizations must also adhere to the principles of purpose limitation, storage limitation, and accountability.

Under the GDPR, individuals have several rights, including the right to access, rectify, and erase their personal data, as well as the right to data portability and the right to object to the processing of their data. Due to the nature of processing personal data, the relevance to automated decision-making, specifically in the context of credit and loans, is high. Kaminski's 2018 paper, "The Right to Explanation Explained," highlights a hypothetical scenario where an algorithm predicts the probability of loan repayment. Here, Kaminski explores the risk of "uncertainty bias," where a risk-averse algorithm might deny credit to certain groups based on the inherent uncertainty of small sample sizes. This scenario presents a potential issue regarding discriminatory practices under GDPR (Kaminski, 2018). GDPR's Article 22 specifies that individuals should not be subjected to decisions solely based on automated processing, including profiling, which could significantly impact them. Hence, GDPR provides a framework for addressing potential discrimination in automated credit assessment (Kaminski, 2018).

There is where the distinction of big data and alternative data comes into play, while big data does indeed fall in what can be considered automated, as what Comprehensive Credit Bureaus do when assessing credit and giving a score. Alternative data gathered given by consumers to lenders with the explicit consent to use it to assess credit does not fall into this case.

Under GDPR's Article 22 exception 1(c), automated decision-making is allowed if the person whose data is being used gives clear consent. In the context of alternative data, this means that individuals willingly provide detailed information with the clear understanding that it will be used for assessing credit risk. This explicit consent allows the use of Alternative Data within the bounds of the GDPR.
On the other hand, big data, which comes from a wide variety of sources and includes large amounts of information, often lacks this clear consent. It's difficult to ensure that every person included in such large datasets has given explicit consent for their data to be used, particularly for specific purposes like credit scoring. Thus, big data usually falls outside the exception provided by Article 22's exception 1(c).

3.4.2 DIRECTIVE 2008/48/EC (Consumer Credit Directive)

DIRECTIVE 2008/48/EC, also known as the Consumer Credit Directive, is another key piece of legislation that comes into play when dealing with credit risk assessment. Unlike GDPR, this directive does not explicitly mandate creditors to provide detailed reasons for credit rejection upon a consumer's request. However, it requires that if a credit application is denied based on a database consultation, the creditor must inform the consumer of the database consultation's outcome and the particulars of the database consulted.

The directive further requires creditors to provide comprehensive explanations to consumers to assess if the proposed credit agreement suits their needs and financial situation. Although it does not demand detailed explanations for credit rejections, it provides a legislative basis for transparency in credit decision-making processes.

3.4.3 The Proposal for a Regulation on Harmonized Rules on Artificial Intelligence

Lastly, the Proposal for a Regulation on harmonized rules on artificial intelligence (AI) (European Commission, 2021) and the DRAFT Compromise Amendments on the Draft Report (European Parliament, 2023) offer a more recent perspective on how the EU aims to govern the use of AI in various applications, including credit scoring.

The 2021 proposal classifies AI systems used for credit scoring as high-risk due to their potential for discrimination and the significant impact they can have on individuals' access to essential services. It emphasizes the need for these systems to be developed based on high-quality training, validation, and testing data sets that meet certain criteria, including relevance, representativeness, and freedom from errors. However, it does not explicitly state that institutions using AI for credit scoring are required to inform consumers about specific reasons for credit rejection (European Commission, 2021).

The DRAFT Compromise Amendments on the Draft Report maintains the classification of AI systems used for credit scoring as high-risk and further emphasizes the importance of data governance and the mitigation of possible biases in the data sets used for training these systems. It also highlights the need for transparency in the use of AI systems, including those used for credit scoring (European Parliament, 2023).

Langenbucher & Corcoran (2021) discuss the EU's 2021 proposal in the context of AI credit scoring. They note that the proposal does not explicitly address issues of algorithmic fairness, historic bias, or discrimination as such. Instead, its approach is more product-oriented, focusing on certification procedures, data and model quality checks, technical documentation, and ex post monitoring duties. They also highlight a fundamental tension between the proposal's anti-discriminatory policy goal and its risk-based, formalistic regulatory design (Langenbucher & Corcoran, 2021).

Both documents necessitate a summary of the findings from the fundamental rights impact assessment and the data protection impact assessment, which should include analyses of potential impacts on individual rights, freedoms, and personal data protection. They also stress the importance of ensuring that AI systems do not perpetuate or amplify existing discrimination or create new forms of discriminatory impacts (European Commission, 2021; European Parliament, 2023; Langenbucher & Corcoran, 2021).

While neither document explicitly requires institutions to inform consumers about specific reasons for credit rejection, the emphasis on transparency, data governance, and the mitigation of biases in these proposals suggests a broader push towards greater accountability and fairness in the use of AI for credit scoring.

3.4.4 Comparative Analysis – The Equal Credit Opportunity Act (ECOA) and the European Union Legislation

A noteworthy contrast to the EU legislation is observed in the United States with the Equal Credit Opportunity Act (ECOA), which explicitly obliges creditors to provide specific reasons for credit rejection (Langenbucher & Corcoran, Patrick, 2021). This provision ensures a degree of transparency and fairness in the credit approval process, allowing applicants to understand and potentially challenge the decision. In comparison, the EU legislation, as examined in the preceding sections, does not impose a similar requirement on creditors.

DIRECTIVE 2008/48/EC (Consumer Credit Directive) only requires that creditors inform consumers if their credit application has been rejected due to a database consultation. However, it stops short of obliging creditors to provide a specific explanation for the rejection. Similarly, the proposed EU regulation on AI emphasizes transparency but does not demand institutions provide detailed reasons for credit refusal. This contrast indicates a difference in legislative approaches towards transparency and individual rights in automated credit decision-making between the EU and the US.

This comparative analysis takes on greater significance considering the increasing reliance on alternative data and big data in credit risk assessment. These data types are increasingly being integrated into AI systems for credit scoring, bringing about an era of 'Big Data' in credit risk assessment.

In this new context, the specificity of reason for credit rejection (as required by ECOA) becomes even more relevant. The use of Alternative Data might lead to rejections based on non-traditional factors, making the reason for rejection less obvious to applicants. As such, the absence of a similar provision in EU legislation could potentially create a transparency gap, as consumers may be left in the dark about why they have been denied credit based on complex AI-driven risk assessments. This underscores the importance of revisiting and possibly revising legislative measures in the EU to ensure transparency, fairness, and the protection of individual rights in the era of big data and AI-driven credit risk assessment.

3.4.5 Shortcomings of EU Legislation in Protection Consumers Against Alternative Data Use in Credit Scoring

Despite the rigorous data protection regime in the EU, it seems that none of the existing regulations fully protect consumers with respect to the use of Alternative Data in credit scoring. There may be several reasons for this.

1. Limited Scope: The GDPR, the Consumer Credit Directive, and the Proposal for a Regulation on AI provide broad protections against data misuse. However, their scope is more general and does not specifically address the use of Alternative Data in credit scoring.
2. Protection of Comprehensive Credit Bureaus: By not necessitating detailed reasons for credit denial, EU laws may also be aimed at protecting the functioning of Comprehensive Credit Bureaus. These agencies gather a wide range of financial data, and the specific methodologies they use to calculate credit scores might be proprietary. Providing detailed reasons for credit denial could reveal these methodologies, making the system susceptible to manipulation or 'gaming'.
3. No Explicit Requirement to Provide Reasons for Credit Denial: None of the EU laws demand that credit providers explain the reasons for credit denial in detail. This aspect, covered in the ECOA in the US, helps promote transparency in the credit decision-making process, allowing consumers to challenge unfair credit denials.
4. Insufficient Focus on Algorithmic Fairness: While EU regulations, especially the Proposal for a Regulation on AI, emphasize the importance of fairness, accountability, and transparency, they lack explicit guidance on addressing the issues of algorithmic fairness, historic bias, or discrimination in AI credit scoring. These gaps could potentially allow unfair or biased practices in credit scoring, using Alternative Data.
5. Consent-Based Approach: The GDPR's consent-based approach allows consumers to willingly provide detailed information, with the understanding it will be used for assessing credit risk. However, this can also create situations where consumers might feel compelled to provide data, not realizing the full implications of their consent or its impact on credit decisions.
6. Rapid Technological Advancements: Finally, the fast-paced development in Big Data and AI technologies have created new challenges that current legislation may not have fully anticipated. The usage of Alternative Data is a recent trend that is not explicitly covered in any existing EU laws, which were drafted before these technologies and data sources became so widespread.

As such, while these EU regulations have undoubtedly increased data protection, they may not be adequately designed to address the specific issues related to the use of Alternative Data in credit scoring. Hence, there may be a need for revisions or new legislative measures that specifically target these issues, ensuring transparency, fairness, and protection of individual rights in this new era of credit risk assessment.

4. DISCUSSION

The first research question focused on how credit rating agencies (CRAs) approach credit risk assessment, necessitating a review of the distinct types of credit reporting agencies across the EU. These encompass Negative-Information Bureaus, Comprehensive Credit Bureaus, and Credit Registries. Each country in the EU has its unique credit reporting agency market, indicating a highly diverse credit reporting landscape.

While specific criteria or statistical models employed by CRAs to assess credit risks are generally not made public, primarily due to the risk of gaming the system, distinct methodologies and types of data are utilized. Each type of CRA processes different amounts of data. For instance, some may only use negative credit events, while others might consider a wider range of factors, including positive and negative financial events. Although the Consumer Credit Directive establishes certain guidelines for creditors, it does not set out a universal standard for credit risk assessment across the EU, contributing to these differences. Each country has a distinct process influenced by various factors such as data privacy laws, economic status, technological progress, among others. Consequently, identifying similarities or pinpointing common practices across this diverse landscape proves to be a considerable challenge.

The second research question, concerning the effect of data sources on credit risk predictions, prompted a discussion on the distinction between big data and alternative data. While both types of data can enhance credit risk predictions, differences exist in their sources and usage. Notably, empirical studies highlight the efficacy of alternative data in forecasting credit risk. Berg et al. (2020) found that digital footprints can be effective in predicting consumer payment behavior and improving credit risk predictions, particularly for unbanked individuals. Meanwhile, Rozo et al. (2023) revealed that certain web browsing variables significantly enhance the predictive accuracy of a probability of default (PD) model, further validating the potential of alternative data in credit risk assessment.
This exploration, however, lacks a direct comparison of the effectiveness of big data usage versus alternative data usage in credit risk assessments, suggesting an intriguing area for future research. Nonetheless, this discussion acknowledges the potential benefits and challenges of using alternative data in credit risk assessments, including the potential for increased financial inclusion and the need to address privacy concerns and potential biases.

The third research question on the practical implications of using alternative data in credit risk evaluation underscored the potential benefits of alternative data in expanding access to credit for individuals with limited or no credit history. It was noted how alternative data can provide insights into consumer behavior that may not be evident from traditional credit data, which in turn allows lenders to make more informed credit decisions. Nonetheless, the challenges associated with alternative data, such as ensuring data quality and relevance and addressing privacy concerns, were acknowledged.

In response to the fourth research question regarding the rights of EU individuals in credit risk assessment with alternative data, it becomes clear that the existing regulations - General Data Protection Regulation (GDPR), Consumer Credit Directive, and the proposed regulation on harmonized rules for artificial intelligence - provide broad protections. However, they may not comprehensively address the specific challenges associated with alternative data in credit scoring.
The GDPR guards against potential discriminatory practices in automated credit assessments by granting individuals rights to control their data and restrict automated decisions that significantly impact them. For instance, alternative data used for credit risk assessment, given explicit consent by the individuals, falls within the bounds of the GDPR. However, large datasets categorized as big data often lack this explicit consent, presenting potential GDPR compliance issues.
The Consumer Credit Directive provides transparency in credit decision-making processes but doesn't mandate detailed explanations for credit rejections. This brings into focus the potential for opacity when credit decisions are based on non-traditional data sources.
The proposed AI regulation classifies AI systems used for credit scoring as high-risk and emphasizes data governance, bias mitigation, and transparency. However, it doesn't necessitate institutions to inform consumers about specific reasons for credit rejection, possibly leading to a transparency gap.
A comparative analysis with the U.S. Equal Credit Opportunity Act, which obliges creditors to provide specific reasons for credit rejection, demonstrates the potential shortcomings of EU legislation. As the era of AI-driven credit risk assessment unfolds, legislative measures in the EU may require revisiting to ensure transparency, fairness, and the protection of individual rights.

To conclude, this discussion has demonstrated the relationship between the topics covered in the prior subsections and provided answers to the research questions stated in the introduction. It has highlighted the diversity of credit reporting agencies in the EU, the potential benefits and challenges of using alternative data in credit risk assessments, and the need for revisions or new legislative measures to ensure transparency, fairness, and the protection of individual rights in the era of alternative data and AI-driven credit risk assessment. These findings contribute to the ongoing debates surrounding credit risk assessment methodologies and inform future policy decisions to enhance transparency, fairness, and consumer protection in the EU's credit landscape.

5. CONCLUSION

The usage of alternative data presents a promising opportunity to expand financial inclusion and extend credit chances to those with little or no credit history. By using non-traditional data like social media profiles and web browsing history, lenders can gain a new understanding of consumer behaviors, leading to more informed credit decisions. However, this fresh approach prompts questions about data privacy, security, and potential biases in evaluating credit risk.

In the European Union, existing laws like the General Data Protection Regulation (GDPR), the Consumer Credit Directive, and the Proposal for a Regulation on Artificial Intelligence (AI) set a foundation for addressing data protection and transparency in credit risk evaluation. Despite these, the regulations may not thoroughly handle the unique challenges that come with using alternative data in credit scoring. Compared to the United States Equal Credit Opportunity Act (ECOA), which mandates lenders to explain credit rejection reasons, EU laws don't impose similar rules, potentially leaving a gap in consumer transparency.

For a fair and open credit environment in this era of abundant data and AI-powered credit risk evaluation, the EU needs to rethink and revise its stance on credit scoring transparency. Policymakers, financial institutions, and consumers need to cooperate to find the right balance between safeguarding consumers and maintaining market access, supporting a stable and fair financial system. By improving transparency and consumer protection in credit scoring, the EU can assist consumers to make better financial decisions and promote wider access to credit through innovative use of alternative data.

5.1 Future research

Building upon the topics that were set out in this paper, there are several aspects that could be further researched, on how alternative data may affect credit risk assessment in the European Union:

1. Addressing Bias and Ensuring Fairness in Algorithms: With AI and machine learning increasingly important in credit risk assessments, future investigations should include the potential biases and discriminatory tendencies in algorithm-driven decisions in the context of using Alternative Data as a base source. This research may help shed light on how alternative data can affect fairness in AI-powered credit risk evaluations.
2. Consumer Consent and Awareness: The usage of alternative data creates the concern about consumers awareness and their consent granting in credit risk evaluation scenarios. Therefore, an area needing further scrutiny is the consumer's perception of using alternative data in credit scores, their grasp of the implications of giving consent, and the potential risks of being coerced or manipulated into consenting to credit assessments.
3. Exploring Direct Access to Consumer Social Media Profiles: Utilizing Application Programming Interfaces (APIs) to directly access consumers social media profiles for alternative data opens a new set of possibilities in credit risk assessment. But it also brings forth concerns around data privacy and security. Future research should delve into the practicality, benefits, and hurdles of using APIs to access consumer social media data for credit scoring.

To sum up, the application of alternative data in credit risk assessment brings to the table a combination of promising opportunities and significant challenges. Focused research into these areas will contribute to a more profound understanding of the advantages and limitations of using alternative data in credit scoring. This, in turn, will be crucial in shaping policy decisions and best practices, with the goal of fostering transparency, fairness, and consumer protection in the evolving credit landscape.

6. ACKNOWLEDGMENTS

I would like to express my gratitude to my parents for their support, to my supervisor Marcos, and to the weather for staying cool.

7. REFERENCES

Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the Rise of FinTechs: Credit Scoring Using Digital Footprints. The Review of Financial Studies, 33(7), 2845–2897. https://doi.org/10.1093/rfs/hhz099

Carrasco Conesa, M. del P. (2017). La morosidad en España: El sistema FACe como herramienta de gestión empresarial. https://idus.us.es/bitstream/handle/11441/64751/La%20morosidad%20en%20Espa%C3%B1a_%20el%20sistema%20FACe_Pilar%20Carrasco%20Conesa.pdf

Choudhary, M. A., & Jain, A. K. (2020). How public information affects asymmetrically informed lenders: Evidence from a credit registry reform. Journal of Development Economics, 143, 102407. https://doi.org/10.1016/j.jdeveco.2019.102407

Crone, S. F., & Finlay, S. (2012). Instance sampling in credit scoring: An empirical study of sample size and balancing. International Journal of Forecasting, 28(1), 224–238. https://doi.org/10.1016/j.ijforecast.2011.07.006

De Haan, J., & Amtenbrink, F. (2011). Credit Rating Agencies. In S. Eijffinger & D. Masciandaro (Eds.), Handbook of Central Banking, Financial Regulation and Supervision. Edward Elgar Publishing. https://doi.org/10.4337/9781849805766.00026

Directive 2008/48/EC of the European Parliament and of the Council of 23 April 2008 on credit agreements for consumers and repealing Council Directive 87/102/EE, Pub. L. No. Directive 2008/48/EC, OJ L 133, 22.5.2008, p. 66–92 (2008). https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:133:0066:0092:EN:PDF

Draft Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9 0146/2021 – 2021/0106(COD)). (n.d.). European Parliament. https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf

Equifax. (n.d.). Decision360 Brochure. Equifax. https://assets.equifax.com/legacy/pdfs/corp/Decision-360-Brochure_051010.pdf

FICP. (2023). Fichier des incidents de remboursement des crédits aux particuliers (FICP) [Data set]. https://www.service-public.fr/particuliers/vosdroits/F17608

Florez-Lopez, R. (2010). Effects of missing data in credit risk scoring. A comparative analysis of methods to achieve robustness in the absence of sufficient data. Journal of the Operational Research Society, 61(3), 486–501. https://doi.org/10.1057/jors.2009.66

Goodman, B., & Flaxman, S. (2017). European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation." AI Magazine, 38(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Grunert, J., Norden, L., & Weber, M. (2005). The role of non-financial factors in internal credit ratings. Journal of Banking & Finance, 29(2), 509–531. https://doi.org/10.1016/j.jbankfin.2004.05.017

Horen, T. (n.d.). Rechtliche Grundlagen des SCHUFA-Scoring-Verfahrens. https://www.itm.nrw/wp-content/uploads/schufa_scoring_verfahren.pdf

Hurley, Mikella, & Adebayo, Julius. (2016). CREDIT SCORING IN THE ERA OF BIG DATA. 2016. https://openyls.law.yale.edu/bitstream/handle/20.500.13051/7808/Hurley_Mikella.pdf

Kaminski, M. E. (2018). The Right to Explanation, Explained. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3196985

Kulkarni, S. V., & Dhage, S. N. (2019). Advanced credit score calculation using social media and machine learning. Journal of Intelligent & Fuzzy Systems, 36(3), 2373–2380. https://doi.org/10.3233/JIFS-169948

Langenbucher, K., & Corcoran, Patrick. (2021). Responsible AI Credit Scoring – A Lesson from Upstart.com. https://pdfs.semanticscholar.org/ed77/860177ab254b7e03e2c0cf0a8b243c36bb5c.pdf

LexisNexis. (2013). Evaluating The Viability Of Alternative Credit Decisioning Tools—A $3.6 billion opportunity for the auto- and credit card-lending markets. https://risk.lexisnexis.com/-/media/files/financial%20services/white-paper/alternative-credit-decisioning%20pdf.pdf

Mak, V., & Braspenning, J. (2012). Errare humanum est: Financial Literacy in European Consumer Credit Law. Journal of Consumer Policy, 35(3), 307–332. https://doi.org/10.1007/s10603-012-9198-5

Mileris, R., & Boguslauskas, V. (2011). Credit Risk Estimation Model Development Process: Main Steps and Model Improvement. Engineering Economics, 22(2), 126–133. https://doi.org/10.5755/j01.ee.22.2.309

Ortlepp, B. (2019). A Feasibility Study on Using the Blockchain to Build a Credit Register for Individuals Who Do Not Have Access to Traditional Credit Scores [University of Cape Town]. https://open.uct.ac.za/bitstream/handle/11427/30884/thesis_com_2019_ortlepp_bryony.pdf?isAllowed=y&sequence=1

Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM/2021/206 final). (2021). European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Pub. L. No. Regulation (EU) 2016/679, OJ L 119, 4.5.2016, p. 1–88 (2016). https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04

Rocha, C. (2023, May 16). Moreno entra en la pugna por la vivienda: Aval del 15% a 1.000 jóvenes sin límite de renta. El Confidencial. https://www.elconfidencial.com/espana/andalucia/2023-05-16/moreno-ayudas-vivienda-jovenes-avales_3630451/

Rozo, B. J. G., Crook, J., & Andreeva, G. (2023). The role of web browsing in credit risk prediction. Decision Support Systems, 164, 113879. https://doi.org/10.1016/j.dss.2022.113879

Trumbull, G. (2008). Consumer Protection in French and British Credit Markets. https://www.jchs.harvard.edu/sites/default/files/media/imp/ucc08-17_trumbull.pdf

White, L. J. (2013). Credit Rating Agencies: An Overview. Annual Review of Financial Economics, 5(1), 93–122. https://doi.org/10.1146/annurev-financial-110112-120942

Wiedemann, K. (2018). Automated Processing of Personal Data for the Evaluation of Personality Traits: Legal and Ethical Issues. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3102933

Xiao, J. J., Chatterjee, S., & Kim, J. (2014). Factors associated with financial independence of young adults: Financial independence of young adults. International Journal of Consumer Studies, 38(4), 394–403. https://doi.org/10.1111/ijcs.12106

Zhang, S., Xiong, W., Ni, W., & Li, X. (2015). Value of big data to finance: Observations on an internet credit Service Company in China. Financial Innovation, 1(1), 17. https://doi.org/10.1186/s40854-015-0017-2

(N.d.).
Consumer Studies, *38*(4), 394–403.
https://doi.org/10.1111/ijcs.12106

Zhang, S., Xiong, W., Ni, W., & Li, X. (2015). Value of big data to finance: Observations on an internet credit Service Company in China. *Financial Innovation*, *1*(1), 17. https://doi.org/10.1186/s40854-015-0017-2