The Use of Alternative Data in Consumer Credit Scoring

Ali Enver Arslan, Sabanci University, MSc in Data Analytics.

Contact: aliarslan@sabanciuniv.edu

Index

Abstract

The paper discusses the recent trend in credit scoring systems that uses alternative data like social media and data reveals the consumer habits, lifestyle, and social status. A brief history of the credit scoring systems is explained with the theory lying behind the credit scoring algorithms. Afterward, the shift in the credit scoring methods towards using big data opportunities and some of the examples are discussed. Lastly, the potential risks that come with the use of alternative data and, the measures and precautions that regulatory and legislative authorities are trying to take are tried to be explained.

Keywords: Credit Scoring, creditworthiness, alternative data, big data, fintech

Introduction

Financial institutions have been in our lives for a long time. The need for unearned money for our needs has always been a problem for people all over the world. Banks are the main actors centered in the middle of all the financial activities in the system. To get credit from a bank, one must be assessed by a credit scoring mechanism that is run within the bank. Banks assess their customers and assign a credit score for each of them. The credit score is used to decide the creditworthiness of an applicant.

The most famous credit scoring is based on the one that FICO developed in the 1950s. FICO (Fair, Isaac, and Company) is a data analytics company and they are focused on credit scoring services. The FICO's ranking formula has been in use for consumer credit of all kinds almost everywhere since then. Due to this established assessment method of credit applications, creditworthiness is associated with our financial behaviors. Some of these financial behaviors that this classical approach takes into consideration are the applicant's income, existing loans, payment history, and assets s/he possessed.

However, this procedure has not yet incorporated all of the society in itself. There is a large number of people who are invisible to the financial system. These people are out of the system due to never using credit/credit cards before or anything valuable in the assessment process of credit scoring. Without a financial record history, one cannot take a loan from the financial institutions easily since it makes it impossible to assess the default risk of that loan. In addition to the people without a financial history, there are also people with very poor financial records that are also not eligible for getting credit in the system. According to the report published by Oliver Wyman in 2017, around 60 million adult people were credit-invisible in the United States alone (Carrol & Rehmani, 2017). These people are not considered eligible to get credit since they are unable to prove their creditworthiness to the lenders.

On the other hand, advances in technology created new tools to assess creditworthiness and increase the penetration of the financial system. Increasing digital footprints of individuals left on the web and massive enhancements made in machine learning and artificial intelligence have created a new actor in the financial system named FinTech companies. The emergence of the

FinTech companies accelerated the pace of the innovations made in the industry. In the competition between the banks and FinTech companies, FinTechs focused on the services that are not provided by the banks sufficient enough and customers that struggle with the existing financial system. However, during this competition, as innovations and new methods are brought on the stage, controversies also started to occur. The use of alternative data sources to assess the creditworthiness of customers has become one of the controversial topics in the industry.

The Shift in Consumer Credit Scoring

In the assessment of applicants' financial reliability, credit scores are mostly calculated by using data generated by the applicant throughout his/her financial history. This data includes income statement, bank account, and payment history, borrowings and repayment activity, approval, and terms of the previous loans, including the loan amount and interest rate. All this financial information can be found during the assessment process from the credit bureaus like Equifax, Experian, and TransUnion that collect information from the creditors. However, this traditional assessment method has its consequences. First, banks tend to follow tight credit policies during a downturn. However, it is the time when consumers need credit most. Second, it can be difficult for companies and individuals without credit histories to be assessed as reliable debtors. The situation is called one of the catch-22* of today's financial system. To borrow, you need a score; but to generate a score, you need to have borrowed before (Carrol & Rehmani, 2017).

With the help of advances in technology, increasing use of big data, machine learning, and artificial intelligence in almost every industry, FinTech companies started to challenge the financial services that have been dominated by the banks for a long time. In almost every area, Fintechs have been continuing to show up with an innovative product or service. credit assessment techniques are also one of the popular areas and there have been developments in recent years regarding the better assessment of the credit applications and reaching out to the consumers that fail to engage with the traditional financial system.

**A dilemma or difficult circumstance from which there is no escape because of mutually conflicting or dependent conditions. (Oxford Dictionary)*

According to research conducted by IMF, the rise of the internet permits the use of new types of nonfinancial customer data, such as browsing histories and online shopping behavior of individuals, or customer ratings for online vendors.

The literature suggests that such non-financial data are valuable for financial decision-making. Berg et al. (2019) show that easy-to-collect information such as the so-called “digital footprint” (email provider, mobile carrier, operating system, etc.) performs as well as traditional credit scores in assessing borrower risk. Moreover, there are complementarities between financial and non-financial data: combining credit scores and digital footprint further improves loan default predictions. Accordingly, the incorporation of non-financial data can lead to significant efficiency gains in financial intermediation (Mr.Lev Ratnovski et al., 2020).

In the last decade, credit scoring methodologies have been changing dramatically in terms of data sources (Chui, 2013; Jenkins, 2014; Lohr, 2015). An increasing number of firms rely on network-based data to assess consumer creditworthiness. One such company, Lenddo, is reported to assign credit scores based on information in users’ social networking profiles, such as education and employment history, how many followers they have, who they are friends with, and information about those friends (Rusli, 2013). Similar to Lenddo, a growing number of start-ups specialize in using data from social networks. Such firms claim that their social network-based credit scoring and financing practices broaden opportunities for a larger portion of the population and may benefit low-income individuals who would otherwise find it hard to obtain credit (Wei et al., 2015).

Other startups were founded focusing on various customer segments that are ignored by the traditional system. Nova Credit, founded in 2016, focuses on credit scoring for immigrants in the United States. Credolab calculates real-time credit scores based on evidence; shown in behavioral data from mobiles and web use. There are other startups founded all around the world working on innovative alternatives to the traditional credit scoring methodology and their numbers are growing rapidly. The data sources that can be used in the alternative credit scoring systems have been increasing too. Facebook, which is the owner of the two largest social media platforms, obtained a patent for a credit scoring system based on social networks. According to this patent, “when an individual applies for a loan, the lender examines the credit ratings of members of the individual's social network who are connected to the individual through authorized nodes. If the average credit rating of these members is at least a minimum credit score, the lender continues to process the loan application. Otherwise, the loan application is rejected." (Lunt, 2006)

Concerns About the Alternative Credit Scoring

In the traditional credit scoring systems, consumers already do not have much control over their creditworthiness. They are not even aware of most of the factors that affect their credit scores. These automated processes make it even more difficult to object to the results for the consumers. In this environment that creates unfair and inaccurate assessment, innovative solutions that use alternative data sources for credit scoring might seem very attractive for the consumers who struggle with the traditional scoring systems. However, alternative credit scoring systems, using the benefits of the technology, are not innocent as its' proponents claim. Alternative data sources may bring harm while creating opportunities for the consumers.

The "creditworthiness by association" system which means considering the consumers' social environment, and people with similar stereotypes together to predict the consumer is not a very new concept. The famous story of Kevin Johnson and American Express happened in 2009. According to the story, Kevin was an Afro-American successful businessman from Atlanta, United States who had deep knowledge about credit systems and was careful about his spending since his father was in the credit industry (Cuomo et al., 2009). However, American Express, one day, unexpectedly, lowers his credit card limit by more than 65%. The reason for the company for their action is the poor performance of other consumers who shop from the same stores as Kevin (Lieber, 2009). This story solely can explain the possible threats the alternative data scoring might cause. Kevin Johnson was a victim of "the creditworthiness by association" concept and with the advanced use of social media, now, people are much more exposed to the companies that collects their data and use for countless purposes.

ZestFinance is one of the most well-known credit scoring companies. They offer big-data credit scoring tools for financial institutions. "all data is credit data" approach they take combines the traditional credit scoring with alternative data they collected from various data sources (Hurley & Adebayo, 2017). ZestCash, an affiliate of ZestFinance as a lender, uses the same data sources for the assessment of the applicants. ZestCash uses uncommon data in its scoring algorithms. The situation is clarified in Quentin Hardy's article in New York Times as below.

> "For example, paying half of one's income in an expensive city like San Francisco might be a sign of conventional spending, while paying the same amount in cheaper Fresno could indicate profligacy. Giving up a prepaid cellphone, often the phone of necessity for the poor or credit-risky, meant losing a phone number, which indicates a willingness to lose social connections. That was a warning sign. A careful reading of ZestCash's terms and conditions, which the company could watch by tracking cookies, meant that someone was taking a loan seriously, not just rushing to get the money. That was a positive sign." (Hardy, 2012)

ZestFinance is only one of the fintech companies that are producing products and services using big data, and the methods they use are even more poorly understood than the traditional credit scoring systems. These companies first emerged with the claim of enabling the credit-invisible consumers to engage with the financial instruments and having more accurate creditworthiness assessments. However, since technology advances much faster than the regulations and there is a new company appearing with another product every day, dealing with the problems that might occur with the alternative credit scoring systems becomes even harder.

In the data collection and processing parts of the alternative credit scoring systems, various data sources are combined to build a model which assesses the consumers' creditworthiness. In the algorithms using that much data, there is a risk of false relationships between the target and the explanatory variables. Data scientist James Kobielus explains this situation as "One of the bedrock truths of statistics is that, given enough trials, almost any possible occurrence can happen... The more possible events that might take place, the more potential, albeit unlikely, 'fluke' events there are." (Hurley & Adebayo, 2017).

Another potential risk of using big data in credit scoring algorithms is that data collected might include biases in itself. One should be able to find out those biases carefully such as underrepresentation, social status, geography, race and gender discrimination, etc. However, it is a highly difficult task to solve. If the credit scoring algorithm uses data from social media activities, older people who are less engaged with social media platforms would be misrepresented. In addition to this, if the algorithm uses historical data with the location information of the applicants, there might be biases towards some social and/or ethnic groups.

The accuracy of the alternative credit scoring systems is also a matter of debate. Although more accurate assessments are one of their main claims, a study by the National Consumer Law Center in the United States, examined the consumer information held by several major data brokers likewise concluded that the data sources used by alternative credit scorers were "riddled with inaccuracies," ranging from "the mundane" to the "seriously flawed." (Hurley & Adebayo, 2017). While the accuracy is, on one hand, another part is the data quality concerns. Even banks who use officially gathered and approved financial data deal with incorrect or problematic data and considering the wide range of data sources that FinTech companies use, it is a much harder task to gather clean data that would not mislead the algorithms.

Finally, as mentioned before, as the data gets larger the transparency will be much more unclear and the consumers will have even less control over their creditworthiness compared to traditional credit scoring systems.

Regulations For the Alternative Data

The Fair Credit Reporting Act (FCRA) was enacted in the United States in 1970. Its main purpose is to promote the accuracy, fairness, and privacy of consumer information collected and kept by the consumer reporting agencies like TransUnion, Experian, and Equifax. The act gives consumers to get access to how their data is processed and used by the 3rd parties. On the other hand, companies that use consumer data they own, are out of the scope of FCRA due to the limitation in the statutory definition of Consumer Reporting Agencies (CRA)(Chopra, 2021).

The Equal Credit Opportunity Act (ECOA), which was enacted in the United States in 1974, prohibits discrimination based on race, color, religion, national origin, sex, marital status, age, receipt of public assistance, or good faith exercise of any rights under the Consumer Credit Protection Act. The Act also requires creditors to provide applicants, upon request, with the reasons underlying decisions to deny credit (Equal Credit Opportunity Act, 1974).

FCRA and ECOA are the two main regulatory frameworks in the United States. However, they are not updated accordingly with the current technological developments in the credit scoring industry and alternative credit scoring algorithms that are fed by big data. In the FCRA, there is no limitation regarding the data that can be used for creditworthiness assessment of the consumers which allows for companies to feed their algorithm with as many data types as they can (Hurley & Adebayo, 2017).

In 2019, "The Accurate Access to Credit Information Act of 2019" was enacted by the House of Representatives in the United States to amend the FCRA to improve the accuracy and fairness of the information that is provided by the CRAs (Accurate Access to Credit Information Act of 2019, 2019). The legislation emphasizes finding better ways of assessing the consumers' reliability and combining the use of big data to understand changing consumer habits (Guler-Fuechec, 2020).

Apart from the United States, there have been ongoing debates and actions in other countries, as well. The World Bank and Consultative Group to Assist the Poor (CGAP) have published a discussion paper in 2018. The paper discusses the current inadequacies of the traditional credit scoring systems and point outs the need for the use of alternative data. Moreover, there is guidance for the issues that are needed to be considered while using the alternative data. Major risks perceived by Mexican authorities about the use of alternative data in credit scorings are inconsistent, incomplete or inaccurate data, reliability of the data sources, violation of consumer privacy, the potential of discrimination, lack of transparency, and lack of the possibility to correct the data ("Data Protection and Privacy for Alternative Data," 2018).

On the European side of the discussions, The European Data Protection Supervisor (EDPS) declared its opinion for the directive on consumer credits proposed by the European Commission (EC). The EC's proposal aimed to ensure the appropriate use of consumer data and fair access to credit around the European Region (Pollet, 2021). The EDPS emphasizes that the Proposal protects strongly individual rights and freedom in terms of personal data processing for the creditworthiness assessments. On the other hand, EPDS suggests that the requirements, roles, and responsibilities of credit databases or third parties providing 'credit scores' should also be addressed in the proposal. If the assessment processes involve customer profiling or another automated processing of consumer data, consumers should be able to receive meaningful information and request a human assessment. The EC's proposal mentions that the data from social media platforms or data related to consumers' health conditions are not supposed to be processed. The EDPS also recommends extending such prohibition to the use of any special categories of personal data under "Article 9 of the GDPR" which is related to the processing of special categories of personal data, as well as information concerning individuals' online browsing behavior (*Opinion 11/2021 on the Proposal for a Directive on Consumer Credits*, 2021).

Conclusion

Considering the developments in the concept of credit, businesses will further develop new tools using different approaches that both people and regulatory authorities were unaware of. Companies have already discovered the potential of big data and the opportunities it provided so, the creditworthiness assessments have been shifting towards another field that uses more and more alternative data such as revealing race, gender, social habits, and social media usage. On the other hand, regulatory authorities are following this shift from behind and trying to keep the credit industry under control. Legislative and regulatory authorities must act fast to find a balance between the promise of better financial inclusion, more accurate assessment, and significant risks they might cause like racial or gender discrimination, unjust treatment, and violation of privacy. However, keeping the balance of the regulations is a very hard task. Regulations should not be made in a way that blocks innovation and financial inclusion. Innovation and new methods that promise to provide better services should be promoted while protecting the consumers from the potential risks they might be exposed to.

Bibliography

Carrol, P., & Rehmani, S. (2017). *Alternative Data and The Unbanked*. Oliver Wyman.

Chopra, S. (2021). Current Regulatory Challenges in Consumer Credit Scoring Using Alternative Data-Driven Methodologies. *Vanderbilt Journal of Entertainment and Technology Law*, *23*(3), 625.

Cuomo, C., Shaylor, J., McGuirt, M., & Francescani, C. (2009, January 28). "GMA" Gets Answers: Some Credit Card Companies Financially Profiling Customers. *ABC News*.

Data Protection and Privacy for Alternative Data. (2018). *The World Bank and CGAP*, 36.

Equal Credit Opportunity Act, Pub. L. No. 15 U.S.C. §§ 1691-1691f (1974).

Accurate Access to Credit Information Act of 2019, no. 739145, 6, 20 (2019).

Guler-Fuechec, S. (2020, January 28). Alternative Data in America: Policy and Concerns. *Center for Financial Inclusion*. https://www.centerforfinancialinclusion.org/alternative-data-in-america-policy-and-concerns

Hardy, Q. (2012, July 5). Big Data for the Poor. *The New York Times*. https://bits.blogs.nytimes.com/2012/07/05/big-data-for-the-poor

Hurley, M., & Adebayo, J. (2017). CREDIT SCORING IN THE ERA OF BIG DATA. *Big Data*, *18*, 69.

Lieber, R. (2009, January 30). American Express Kept a (Very) Watchful Eye on Charges. *The New York Times*.

Lunt, C. (2006). *Authorization and authentication based on an individual's social network*. https://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&d=PALL&s1=8302164.PN.

Mr.Lev Ratnovski, Peter Hoffmann, Mr.Arnoud W.A. Boot, & Mr.Luc Laeven. (2020). *Financial Intermediation and Technology: What's Old, What's New?*

Opinion 11/2021 on the Proposal for a Directive on consumer credits. (2021). European Data Protection Supervisor.

Pollet, M. (2021, August 27). EU privacy watchdog calls for stricter data use to guarantee fair access to credit. *Euractiv.Com*.

Wei, Y., Yildirim, P., Van den Bulte, C., & Dellarocas, C. (2015). Credit Scoring with Social Network Data. *Marketing Science*, *35*(2), 234–258.