The Application of Alternative Data in Credit Scoring

Xuanrui Zhang

College of Liberal Arts, University of Minnesota, MN 55455, Minneapolis, USA
zhan9719@umn.edu

Abstract. This paper mainly combs the ideas and methods of various researches on the application of alternative data in credit scoring and classifies, summarizes and summarizes the research results, so as to provide more comprehensive information for the current researches on this subject and clearer research direction in the future. Using the literature review method, this paper mainly summarizes the relevant researches on the application of alternative data in credit scoring into three aspects: Application of digital footprint, Application of telecommunications data source, and Assessing credit through social networks data, and the important role and feasibility of alternative data in credit scoring. In addition, by sorting out relevant literature, the advantages and disadvantages of each study and the model itself are found, hoping to provide scholars with a more comprehensive understanding of this topic, and provide ideas and research directions for subsequent research by pointing out the shortcomings of current research and application.

Keywords: Alternative data, Credit score, Digital footprint.

1 Introduction

With the development of financial technology, the concept of inclusive finance is increasingly popular, and the popularization and efficiency improvement of financial services have become the focus of social attention. For ordinary consumers, credit is the most important financial behavior, and credit score is an important basis for evaluating whether consumers can obtain credit. Traditional credit scoring systems rely heavily on structured data such as historical credit history and repayment performance. Many groups that lack credit data, such as start-ups, small and micro business owners, and young consumers, often have difficulty accessing financial services due to a lack of adequate traditional credit history [1]. Therefore, the application of fintech and alternative data to the calculation of personal credit scores to make up for the shortcomings of the traditional credit scoring system, improve the popularity and efficiency of financial services, and pay attention to possible problems and drawbacks is one of the important forms of integrating fintech into practical work and promoting the realization of inclusive finance.

Alternative data refers to other information, such as news public opinion, social platform data and e-commerce data, that is collected and processed by credit investigation agencies and data service agencies and used for credit granting decisions

J. Lu (ed.), *Proceedings of the International Workshop on Navigating the Digital Business Frontier for Sustainable Financial Innovation (ICDEBA 2024)*, Advances in Economics, Business and Management Research 315,
https://doi.org/10.2991/978-94-6463-652-9_84

of lending institutions outside the traditional collection scope of lending information [2]. In recent years, more and more scholars have begun to pay attention to the application potential of alternative data in credit scoring, believing that it can make up for the shortcomings of traditional data and improve the accuracy and efficiency of credit evaluation. At the same time, with the major changes brought about by the maturity of artificial intelligence technologies such as fintech and machine learning, alternative data and machine learning models have proven to be functions and tools that allow people to make financial assessments more accurately [3].

The evaluation of credit scores by machine learning enables automated decision-making process [4]. The main research methods include logistic regression, decision tree algorithm, neural network algorithm and so on. Logistic regression, as a classification algorithm, is often used to deal with binary outcomes (for example, whether a borrower will default). The model provides a transparent and interpretable way to understand the impact of different factors on reputation [5]. Decision trees and random forests are effective in capturing complex relationships in datasets, providing flexibility and accuracy for credit risk assessment [6]. Deep learning performs well in the face of large, unstructured data, providing a powerful framework for credit risk assessment.

In the context of individual credit risk assessment, a growing number of approaches are proving useful for some groundbreaking research, with the potential value of alternative data such as social media data, Internet browsing behavior, consumption history, etc. to provide more comprehensive and accurate information for credit assessment. It provides a new method for the optimization and innovation of credit scoring model [7].

This paper will extract methods from existing research literature, summarize the current mainstream research on how to use alternative data for credit scoring, and comprehensively compare its advantages and disadvantages. In addition, attention will be paid to the ethical issues raised by the birth of various new methods and models, privacy concerns, model accuracy and innovation.

2 Research on the Specific Ways of Applying Alternative Data to Credit Scoring

2.1 Application of Digital Footprint

An important source of alternative data is a customer's digital footprint, including mobile payment bills, online payment history, and more. These data can reflect customers' consumption habits, behavior patterns and economic conditions, providing a new dimension for credit assessment. By adding multiple types of bill payments (including mobile, cable, utility, and streaming services) to a person's debt service credit, the study found that its correlation increased. It can be concluded that a positive bill payment history could streamline the underwriting process, help more consumers get loans, and potentially lower mortgage rates for borrowers who might otherwise be overcharged [8].

2.2 Application of Telecommunication Data Resources

More detailed phone data such as call details (metadata for incoming and outgoing calls and text messages, airtime balances and top-ups), as well as data collected from customers' phones such as installed Apps, device models and manufacturers. Add data used in traditional credit scoring systems, such as a consumer's age, gender, marital status, and loan characteristics; Include the amount of the loan, as well as the length and reason of the loan. The authors' most reliable model, based on the support vector mechanism, almost halved the number of delinquent loans while more than doubling the approval rate [9]. Researchers from Edinburgh use psychometric data and data related to email usage characteristics to improve the predictive accuracy of credit scoring systems. Using the area under the curve (AUC) to measure the enhancement of the model's performance reveals the nonlinear nature of the dataset, which challenges traditional methods such as logistic regression. In contrast, neural networks have demonstrated superior accuracy in capturing complex nonlinear relationships inherent in the data [10].

2.3 Assessing Credit Through Social Network Data

Use social data to assess credit by assessing individual social networks with similar credit scores. When predicting loan default, logistic regression analysis is used to find that social media attributes, especially those on Weibo, have a significant impact [11].

Academics from Russia developed two credit scorecards specifically based on social data. The first scorecard uses the classic definition of default: 90 days of default within 12 months from the date the loan was originated. The second scorecard allows the use of scripture for false definition, i.e., breach of the covenant within the first 3 months. Both scorecards change the input data into woe, and then the logic returns. It is found that social data can predict fraud cases better than Putong breach of contract, and social data can be used in the score card of Fengfu Scripture. Even if the used input data is completely sourced from the social network, the performance of the score card is also located in an accessible water level [12].

3 Existing Advantages and Shortcomings of Credit Scoring by Alternative Data

3.1 Advantages

Alternative data covers a wider range of information, such as an individual's online shopping history, social media behavior, mobile payment bills, phone records, and travel information. For example, if a person often buys high-priced goods online and pays on time, his consumption pattern can reflect certain economic strength and credit habits; Frequent travel records may indicate a stable job or life status, which can provide an additional reference for a credit assessment.

Traditional credit scores may rely on long periods of historical data that are updated relatively infrequently. Alternative data, which can be accessed in real or near real time, can more quickly reflect the dynamics of an individual's credit profile.

For example, by monitoring the payment of mobile payment bills in real time, recent consumption and payment anomalies can be detected in time, providing financial institutions with the latest credit risk signals. Also, in today's fast-moving economic and social environment, an individual's financial situation and behavior patterns can change rapidly due to a variety of factors. The real-time nature of alternative data enables it to better adapt to this change when evaluating credit risk, adjust credit score in time, and improve the accuracy and timeliness of credit evaluation.

Because of the variety of alternative data sources, it is possible to dig deep into an individual's unique behavior patterns and preferences. Different people have different consumption habits, social circles, lifestyles and other aspects, and these personalized characteristics are reflected by alternative data, making credit scores more accurate to reflect the real credit level of individuals, rather than just based on some common standards and indicators to evaluate. At the same time, financial institutions can tailor more appropriate credit products and services for different customers according to their individual credit characteristics. For example, for customers with specific consumption preferences or behavior patterns, provide loan amounts, interest rates and repayment methods that match their needs to improve customer satisfaction and the business competitiveness of financial institutions.

3.2 Disadvantages

The enormous complexity of credit scoring systems based on alternative data raises significant accuracy and transparency issues - much of which stems from their secretive, legally protected status - as well as heightened concerns about discriminatory and biased scoring practices using non-traditional behavioral data. Because the data itself may contain biased or discriminatory information (such as gender, race, etc.), this information may be amplified or passed on during the credit scoring process. In addition, AI algorithms operate in a "black box" that cannot provide detailed decision-making basis and explanation, which may also exacerbate the impact of potential discrimination [13]. Therefore, strengthening mechanisms for individual consent prior to the collection and processing of personal data, as well as the establishment of specialized agencies to manage data, are particularly important to address discriminatory issues.

The application of data mining algorithms in credit scoring often has problems of opacity and lack of explanation. In particular, complex algorithms such as deep learning have internal mechanisms and decision-making processes that are difficult for humans to understand. This may cause financial institutions to lack adequate explanations and basis when making credit decisions, increasing potential risks and uncertainties. Therefore, when applying data mining algorithms for credit scoring, it is necessary to balance the accuracy and interpretability of the model. Simultaneously tapping the potential of algorithmic lending and containing its possible risks requires a reassessment of the existing regulatory framework [14].

Before the large-scale application of alternative data, banks do not have mature technology and professional talents to implement it, and the development of technology, training of talents and the development of standardized systems require high costs. In addition, small and medium-sized banks are less attractive to high-end technical talents, which also restricts the development of alternative data to a certain extent [15].

4 Conclusion

This paper summarizes various current studies on the application of alternative data in credit scoring, summarizes the research methods and directions of current studies, drawing a conclusion that the important role and feasibility of alternative data in credit scoring. Also, summarizes the problems still existing in the current model algorithm and scoring system, as well as possible effective solutions. With the increasing number of studies on alternative data, this paper focuses on the study of alternative data on credit scores, which is more conducive to the scholars who want to enter this research field or are in the research stage to fully understand the research process under this background, clarify the research ideas, and find alternative research directions and topics. This paper does not make statistics and analysis on the number of classified papers, which cannot guarantee the accuracy of classification and cannot compare the accuracy and practicability of the research among different categories. It is hoped that subsequent scholars can make further improvements on this basis.

References

1. Crook, Jonathan, et al.: Credit scoring with alternative data. (2019), https://cer.business-school.ed.ac.uk/wp-content/uploads/sites/55/2019/08/A1-Use-of-Alternative-Data-for-Credit-Scoring-Crook.pdf
2. Fei X. J., Wang Y., Zhu Q. Q., You B. F., Bu Z. X.: Research on the application of alternative data from the perspective of credit investigation. Chinese Journal of Credit Investigation 38 (5): 8-12+51 (2020).
3. Roa Ballén, Luisa Fernanda.: Machine Learning Models and Alternative Data in Credit Scoring: Statistical and Financial impact. Universidad de los Andes, (2021).
4. Müller, M. M.: Exploring the Evolution and Impact of Credit Scoring in the European Union: An Analysis of Alternative Data, Individual Consumer Data, and Regulatory Frameworks. BS thesis. University of Twente, (2023).
5. Addy, Wilhelmina Afua, et al.: AI in credit scoring: A comprehensive review of models and predictive analytics. Global Journal of Engineering and Technology Advances **18** (2), 118-129 (2024).
6. Golbayani, Parisa, Ionuţ Florescu, and Rupak Chatterjee.: A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees. The North American Journal of Economics and Finance **54,** 101251 (2020).
7. Shema, Alain.: Effective credit scoring using limited mobile phone data. Proceedings of the tenth international conference on information and communication technologies and development, (2019).
8. Mertasari, N. M. S., I. M. Candiasa.: Application of partial credit models in testing performance assessments for programming course. Journal of Physics Conference Series **1810** (1), 012055 (2021).

9. Ruiz, Saulo, et al.: Credit scoring in microfinance using non-traditional data. Progress in Artificial Intelligence: 18th EPIA Conference on Artificial Intelligence, EPIA 2017, Porto, Portugal, September 5-8, 2017, Proceedings 18. Springer International Publishing, (2017).
10. Djeundje, Viani B., et al.: Enhancing credit scoring with alternative data." Expert Systems with Applications **163**, 113766 (2021).
11. Ge, Ruyi, et al.: Predicting and deterring default with social media information in peer-to-peer lending. Journal of Management Information Systems **34** (2), 401-424 (2017).
12. Masyutin A. A.: Credit scoring based on social network data, Бизнес-информатика **3** (33), 15-23 (2015).
13. Chopra, Sahiba.: Current regulatory challenges in consumer credit scoring using alternative data-driven methodologies. Vand. J. Ent. & Tech. L. **23**, 625 (2020).
14. Ahmed, Shehnaz.: Alternative Credit Scoring-A Double Edged Sword. J. Indian L. & Soc'y **11**, 1 (2020).
15. Fan L., Wang J. H., Shu M. Y., Gao R., Li S. Y.: Application of alternative data in bank loan risk management under the background of "Internet Plus". Advances in Social Sciences **10**, 1048 (2021).