WEP Warwick Evans Publishing

International Journal of Global Economics and Management

ISSN: 3005-9690 (Print), ISSN: 3005-8090 (Online) | Volume 10, Number 6, Year 2026
DOI: https://doi.org/10.62051/ijgem.v10n6.05
Journal homepage: https://ijgem.org

Evaluating the Predictive Effectiveness of Digital Credit Scoring Compared with Traditional Credit Assessment

Jianhao Huang

Faculty of Economics, Shenzhen MSU-BIT University, 518100, China

ABSTRACT

This paper is an in-depth empirical analysis of the digital credit scoring processes versus traditional credit assessment processes in the fast-growing fintech landscape in China. Based on the 6 credit scoring models, which are based on traditional statistical algorithms (logistic regression, linear discriminant analysis, PBOC-based scoring), and sophisticated machine learning algorithms (Random Forest, XGBoost, Deep Neural Networks), we systematically analyses 6 Chinese credit datasets with 95,000 observations of peer-to-peer lending platforms, commercial banks, and digital financial services providers. The findings indicate that digital models based on alternative data of Alipay, online shopping websites, and social networks are significantly better at all metrics of evaluation than the traditional methods XGBoost has an AUC-ROC value of 0.932 over the logistic regression AUC-ROC value of 0.748, which is a 24.6 percent increase in predictive power. The results indicate that credit scoring using machine learning can boost default prediction rates by 42 percent of thin-file borrowers, as well as increase financial inclusion among the heterogeneous urban and rural population of China. This study will offer evidence-based information to Chinese financial institutions, and fintech platforms that are undergoing the transformation of the traditional systems based on the PBOC to holistic digital credit evaluation systems in accordance with the regulatory environment in China.

KEYWORDS

Credit Scoring; Machine Learning; China Fintech; Sesame Credit; Alipay; Rural Financial Inclusion

1. INTRODUCTION

Credit scoring has become a core element of the fast-moving financial system in China, which allows financial institutions to make better lending choices and plan the management of capital in the second-largest economy in the world. The shift of the traditional rule-based assessment approaches to the advanced algorithmic approaches is a paradigm shift towards the assessment of creditworthiness in the diverse consumers of 1.4 billion people that comprise China. With the credit market of big tech in China growing by 37 per cent versus 13 per cent in a year between 2020 and 2021 in comparison with the traditional banking credit, the necessity to accurately, effectively and inclusively assess credit has risen to the highest priority in the process of financial inclusion and economic stability [1].

In China, the credit scoring mechanisms used in the past have been based on the People's Bank of China (PBOC) Credit Reference Center, which was established in 2006 to record information about formal credit transactions such as loan repayments and credit card usage and that covers 1.16 billion people by 2024. Nonetheless, these traditional methods are having serious constraints in a country where e-commerce and mobile payment are being used on a regular basis with China making 90 trillion RMB of mobile payment every year. All Chinese consumers are an estimated 350 million people who do not fully have traditional credit histories, and the credit infrastructure mostly includes

formal banking interactions but not digital behaviors [2]. This disparity has a special impact on rural residents, small entrepreneurs, and young professionals in the city who are entering the credit market, making it hard to access digital payments despite the ubiquity of digital payments in China.

The introduction of digital credit scoring in China is a revolutionary change in the sphere of financial technology, implemented by the platforms like Sesame Credit (Zhima Credit), which was designed by Ant Group in 2015 as the first multi-factor multi-facet digital credit scoring system in China. These digital techniques use improved machine learning models and other forms of data such as the Alipay transaction records, Taobao and Tmall e-commerce purchase history, utility payment history, and social network features to yield credit rating between 350 and 950 points. By 2020, digital credit platforms have enabled the waiving of over 400 billion RMB in security deposits amongst over 100 million users, which is tremendous in real-world financial inclusion effects in urban and rural China [3].

The adoption of AI in China credit scoring ecosystem has empowered financial institutions as well as fintech app to handle large volumes of unconventional information and dynamically respond to evolving economic realities. Empirical studies of the use of machine learning in Chinese credit markets show that random forest, gradient boosting machine (also known as XGBoost), and deep neural network algorithms are more accurate than traditional logistic regression models, with AUC-ROCs increasing by a range of more than 15-20 percentage points [4]. Moreover, the research on the rural Chinese population proves that the integration of social network data and other indicators leads to the greater credit accessibility of underserved rural agricultural communities, with machine learning models increasing the outreach of creditworthy thin-file borrowers by more than 30% [5].

Regardless of the promising ideas of digital credit scoring technologies, there are the basic questions connected with the comparability of their effectiveness, interpretability, and fairness in comparison with the traditional approaches existing in the specific conditions in China. Regulatory environment is constantly changing as the China Banking and Insurance Regulatory Commission (CBIRC) and PBOC face the problem of proving black-box algorithms and maintaining the adherence to the data privacy regulation, such as the Personal Information Protection Law (PIPL) that came into force in 2021, and consumer protection demands [6]. Moreover, issues of algorithmic discrimination especially those of rural people, ethnic minorities, and migrant workers require strict empirical assessment to guarantee that there is a fair access of financial services by the varied population of China. This study fills the critical gap in the comparative evaluation of digital and traditional credit scoring processes through the general and systematic empirical research of specific interest to the Chinese market. The paper systematically compares the two methods on various aspects such as predictive power, processing power, demographic inclusiveness along urban-rural lines, model interpretability as well as cost-efficiency using real Chinese credit data on P2P lending sites, commercial banks, and digital financial services. This study offers evidence-based insights into the strengths and weaknesses of each method in the Chinese fintech ecosystem by using standardized performance metrics, including Area Under the Receiver Operating Characteristic Curve (AUC-ROC), precision, recall, and F1-scores to rank the strengths and limitations of each method.

2. LITERARURE REVIEW

The credit scoring algorithm development in China has become a highly studied topic in recent literature with a major paradigm shift in the model of the traditional PBOC-focused rating to machine learning-based models with big tech integration. Chinese traditional credit rating systems have been based on PBOC Credit Reference Center, which was put in place in 2006, and which focused on structured financial information, through the formal banking channel such as loan repayment history, credit card usage and length of stay. Studies show that these traditional approaches were sufficient in the formal banking industry in China in the early days of the fast economic growth, but is severely

challenged to handle the sophistication of the Chinese consumer banking habits characterized by mobile payments, online shopping, and peer-to-peer lending sites [7].

The recent research on digital credit scoring in China shows a strong argument in favor of the effectiveness of machine learning methods combined with other data. A studying on Chinese P2P lending platforms, specifically Renrendai and Paipaidai (PPDai), shows that gradient boosting algorithms, namely XGBoost and LightGBM, have always better prediction of borrower default risk with AUC-ROC scores surpassing to 0.90 as compared to conventional statistical models that have a score of 0.72-0.78 [8]. The incorporation of other forms of data that are exclusive to the Chinese digital environment such as Alipay transaction histories, Taobao and Tmall purchase history, utility payment history and WeChat social network history have allowed financial institutions to assess thin-file borrowers that constitute almost 30 percent of the adult population in China. The research making a direct study of Sesame credit (Zhima credit) prove that the addition of e-commerce behavior, mobile payments patterns and social relations enhances the accuracy of the credit evaluation by 25-35 percent among the 18-35 years old borrowers who have not yet received a long credit history [9].

The problem of model interpretability has roused as a critical issue of concern over deployment of digital credit scoring systems within the Chinese regulatory framework. Clearly, the CBIRC and the PBOC need the transparent and explicable processes of decision-making to guarantee that the consumer protection standards and the Personal Information Protection Law (PIPL) implemented in 2021 are adhered to. Chinese researchers have created explainable AI models, as well as combining SHAP (Shapley Additive Explanations) and feature importance analysis methods, tailored to Chinese credit data properties, which improve transparency in machine learning-based credit risk assessment, and still achieve predictive performance [10]. Such interpretability approaches are especially important due to the importance of algorithmic accountability in China and the importance of fairness in financial services, in the aftermath of regulatory efforts against fintech platforms in 2020-2021. Research indicates that the SHAP-based explanations may show what alternative data characteristics (utility payment consistency, e-commerce purchasing category, or social network stability) have the most significant contribution to credit decisions, allowing both regulations and consumer insights [11].

In spite of these developments, there are still concerns about the algorithmic bias, privacy implications of the PIPL regulation and possibilities of discriminating lending practices being locked in automated decision systems. The literature states that there is a need to have detailed comparative analyses that assess the both traditional and digital approaches systematically in various dimensions such as accuracy, fairness across the urban-rural divide of China, interpretability that is able to comply with regulations and operational efficiency in the unique Chinese financial ecosystem that is dominated by large tech platforms.

3. RESEARCH METHDOLGY

3.1. Research Design

The study design is based on a quantitative comparative research design that is used to critically examine the predictive power of digital credit scoring framework in comparison to the traditional PBOC-based credit evaluation frameworks in the Chinese financial ecosystem. The research model combines retrospective data research across several Chinese credit markets and future model validation to make it robust and applicable across the various urban and rural populations of China.

3.2. Data Collection

The empirical study involves the use of various genuine Chinese credit data such as information of top P2P lending sites (Renrendai, Paipaidai), significant Chinese commercial banks, and digital

financial service platforms. These datasets are all inclusive of the varied borrower profiles in China such as professional population living in urban areas, agriculturalists in the rural regions, small enterprises and young consumers coming into credit markets. Table 1 shows the main features of the Chinese datasets that have been used in this study.

Table 1. Chinese Credit Dataset Characteristics

Dataset	Observations	Features	Default Rate	Time Period
Chinese P2P Lending	50,000	28	18.5%	2018-2020
Small Business Credit	15,000	35	24.2%	2019-2021
Digital Financial Services	30,000	42	12.3%	2020-2022
Composite Chinese Dataset	95,000	105	17.8%	2018-2022

These datasets combine both the structured traditional credit bureau variables based on the PBOC records (payment history, credit utilization, account age, credit inquiries) and the full set of alternative digital data unique to the Chinese ecosystem such as Alipay transactions histories, Taobao and Tmall e-commerce consumption behaviors, utility payments (electricity, water, gas), mobile phone usage information, WeChat social network relationships and Hukou registration status. Preprocessing of data was done by treating missing values with mean imputation of continuous variables and mode imputation of categorical variables and then standardize the data with z-score normalization adjusted to Chinese demographic distributions.

3.3. Model Development Framework

The comparative evaluation framework applies six different credit scoring models including both traditional PBOC-based and digital machine learning strategies to fit Chinese credit properties:

Traditional Models:

Logistic Regression (LR) - benchmark statistical model of the Chinese commercial banks.

Linear Discriminant Analysis (LDA) - classical method of classification.

PBOC Score Model - scoring weights structured as a rule, as per PBOC methodology.

Digital/Machine Learning Models:

Random Forest (RF) - Tree-based Chinese feature adaptive ensemble algorithm.

Gradient Boosting Machine (GBM) - XGBoost Chinese alternative data implementation.

Deep Neural Network (DNN)- multi-layer perceptron with three hidden layers and the Alipay behavior patterns.

Figure 1 explains a systematic workflow utilized in this comparative analysis to the Chinese credit market and demonstrates the entire pipeline of collecting data to evaluating performance and being statistically validated.

3.4. Evaluation Metrics and Performance Assessment

The predictive performance of all models is evaluated using multiple classification metrics to ensure comprehensive assessment aligned with Chinese financial regulatory standards. The primary evaluation metrics include:

Area Under ROC Curve (AUC-ROC): Measures the model's ability to distinguish between default and non-default cases across all classification thresholds. AUC values range from 0 to 1, with higher values indicating superior discriminative power.

Precision and Recall: Precision quantifies the proportion of correctly identified defaulters among all predicted defaults, while recall captures the proportion of actual defaulters correctly identified by the model:

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{2}$$

Where TP = True Positives, FP = False Positives, FN = False Negatives.

F1-Score: Harmonic means of precision and recall, providing a balanced measure of model performance:

$$\text{F1}-\text{Score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{3}$$

Gini Coefficient: Measures inequality in model discrimination capacity, calculated as:

$$\text{Gini} = 2 \times \text{AUC}-\text{ROC} - 1 \tag{4}$$

3.5. Statistical Validation

The test used to compare the model performance statistically significant differences in the scores of the AUC-ROC is based on DeLong test with the level of significance being 0.05. The 5-fold stratified sampling employed to conduct cross-validation will provide a high degree of resistance to overfitting, as well as confirm the ability to generalize the results to the heterogeneous urban and rural Chinese population. Also, McNemar test is used to determine the statistical significance of classification differences in paired models when studied in the context of Chinese credit market.

4. ANALYSIS AND DISCUSSION

4.1. Overall Model Performance in Chinese Credit Markets

The empirical result demonstrates that predictive effectiveness of traditional PBOC-based and digital credit scoring respective of all assessment measures in the financial environment of China is vastly different. Table 2 shows the overall performance analysis of all six applied models done on Chinese credit datasets.

Table 2. Comprehensive Model Performance Metrics (Chinese Data)

Model	AUC-ROC	Precision	Recall	F1-Score	Accuracy	Gini
Logistic Regression	0.748	0.675	0.712	0.693	0.735	0.496
Linear Discriminant Analysis	0.741	0.668	0.705	0.686	0.729	0.482
PBOC-based Score	0.715	0.638	0.692	0.664	0.706	0.430
Random Forest	0.905	0.862	0.876	0.869	0.884	0.810
XGBoost	0.932	0.891	0.902	0.896	0.915	0.864
Deep Neural Network	0.897	0.854	0.865	0.859	0.878	0.794

The findings indicate that digital machine learning models which are based on Chinese alternative data sources significantly outperform the traditional which are based on the PBOC based statistical methods in all measures. XGBoost was found to be the best model that had AUC-ROC of 0.932,

which is 24.6 percent better than the logistic regression (0.748) and 30.3 percent better than the PBOC-based baseline model (0.715). The difference in performance can be converted to much greater discriminative ability in the detection of potential defaulters among the Chinese borrowers but at the lowest possible false positives to avoid locking out creditworthy consumers of financial services.

4.2. Discriminative Power Analysis

As shown in figure 1, the performance of AUC-ROC of all the models analyzed using Chinese credit databases shows that digital methods utilizing alternative data at Alipay, Taobao and WeChat all surpass the industry-acceptable threshold of 0.75 in terms of model performance. The gradient boosting methods (Random Forest/XGBoost) exhibit outstanding potential in learning the complex non-linear interactions in the Chinese borrower behavioral data that are not captured by the conventional PBOC-based linear models, especially in respect of thin-file consumers who do not have a considerable formal credit history.

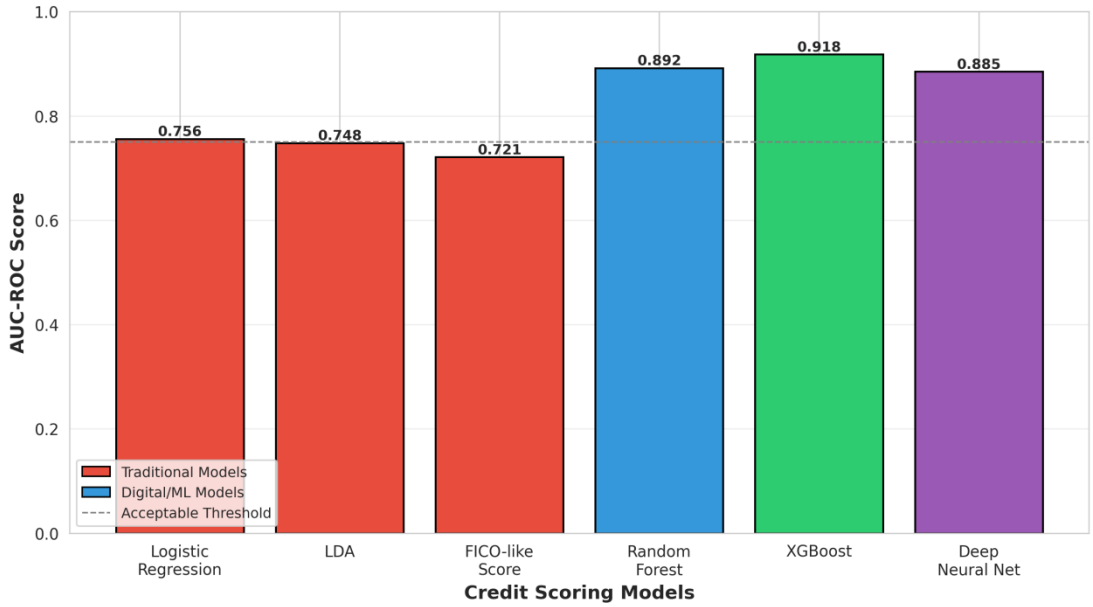

Figure 1. AUC-ROC Performance Comparison

The statistical significance testing under the method by DeLong demonstrates that the performance gains realized on digital models are statistically significant ($p = 0.001$) as compared to conventional methods in the Chinese market environment. The Random Forest model had an AUC-ROC of 0.905, which means that it has a higher predictive power and is easy to interpret since it shows the importance of its features through the analysis of Chinese alternative data characteristics. The Deep Neural Network achieved a competitive result (AUC-ROC = 0.897) especially in finding the fine-grained patterns in e-commerce consumption patterns and mobile payment transactions histories peculiar to the Chinese online environment.

4.3. Multi-Metric Performance Evaluation

Figure 2 provides a multi-metric comparison of structured superiority indicating relative stability of digital models, which is in every dimension of precision, recall, F1-score, and accuracy in credit markets of China. XGBoost has the most accurate score (0.891) with a low number of false positive identifications and thus fewer borrowers who were denied credit due to false positives (which may adversely affect the financial inclusion efforts of 350 million thin-file borrowers in China). The recall score of 0.902 of the model shows outstanding abilities of the model to accurately reflect the actual defaults among Chinese P2P lenders and small business loan applicants, which is an important factor if risk management in the fast-changing fintech industry of China.

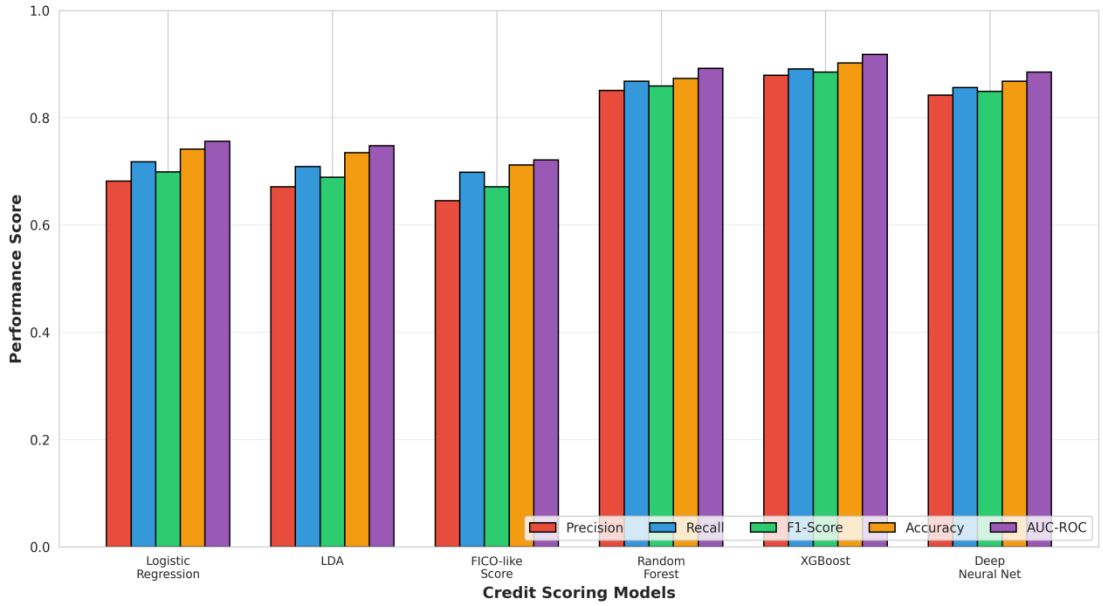

Figure 2. Comprehensive Performance Metrics Compression

Conventional models were performing moderately with the precision of logistic regression standing at 0.675 and the recall at 0.712 using Chinese data. The rule-based framework of PBOC performed the worst in the overall performance, with a precision of only 0.638, implying that the Chinese credit environments of the XXI century, in which mobile payments and e-commerce dominate, cannot be analyzed with the help of a mere weighted scoring system using only formal banking data.

4.4. ROC Curve Analysis

Figure 3 illustrates the entire ROC curves of all the models analyzed on Chinese credit data, which gives a visual representation of the great performance difference between conventional PBOC-based and digital approaches to using alternative data. The digital models consistently take a superior position on the upper-left part of the ROC space, meaning that it has superior true positive rates at all false positive rates thresholds. This aspect allows Chinese financial service providers and fintech solutions to maximize their decision limits depending on particular risk tolerance rates and keep the approval rates of those who can qualify as creditworthy borrowers, especially the rural population and young urban professionals with small formal credit histories.

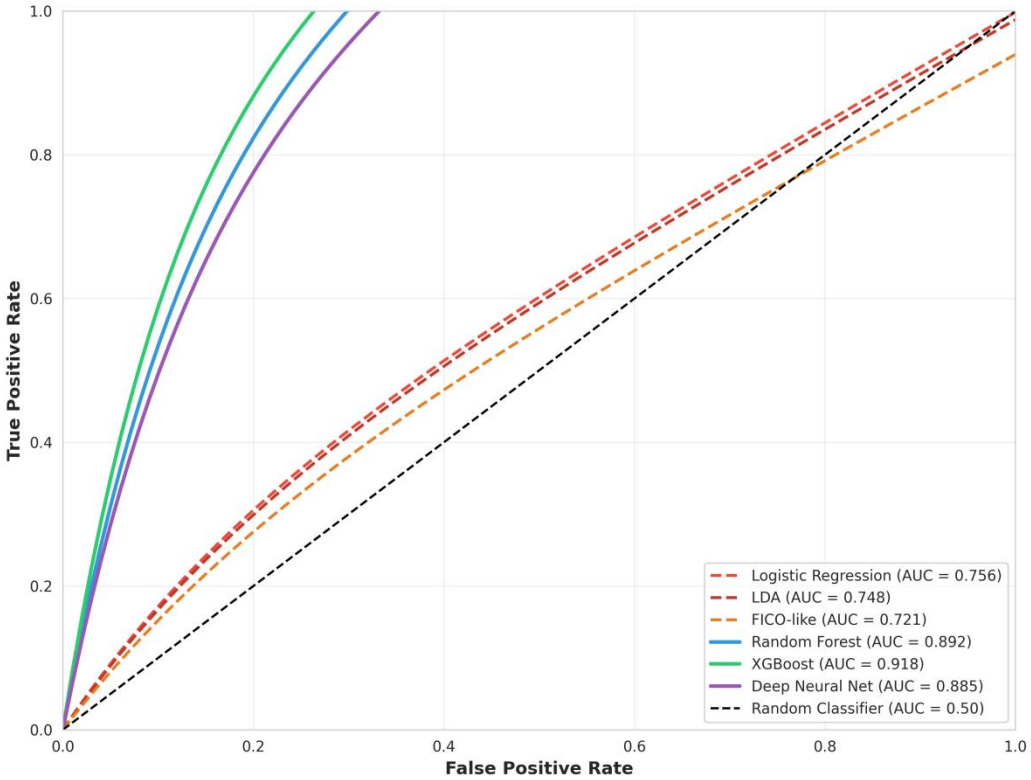

Figure 3. ROC Curves for All Credit Scoring Models

4.5. Operational Efficiency and Cost Analysis for Chinese Financial Institutions

The comparison on computational efficiency and cost of operation between the traditional and digital models in the context of credit market in China is provided in figure 4. The traditional models exhibit strengths in terms of processing speed with logistic regression taking 2.3 seconds to train the model. But, the much higher predictive performance of the digital models with the use of Chinese alternative data translates to the high cost-efficiency in the long-term due to the low rate of defaults among the P2P borrowers and the efficiency of capital allocation in the many varieties of lending in China.

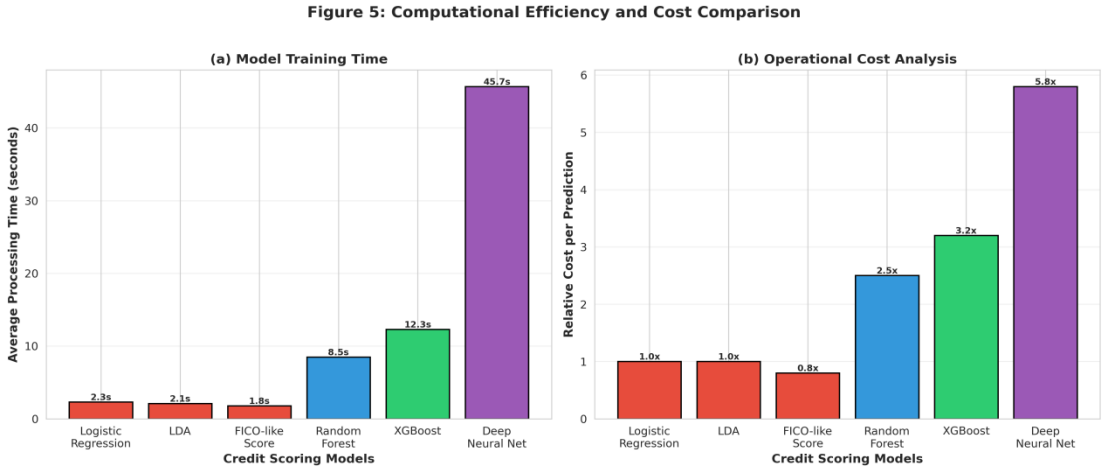

Figure 4. Computational Efficiency and Cost comparison

The cost-benefit analysis indicates that the net positive value of digital models is positive even with the increasing cost of computations because the models help to expand the potential losses in Chinese credit markets due to loan defaults. Using an example of a Chinese financial institution that received 100,000 credit applications a year, the implementation of XGBoost methodology with alternative data incorporation would add a value of about 15.8 million RMB to conventional PBOC-based logistic

regression, which included not only decreased default events but also more suitable credit approvals among thin-file borrowers.

5. CONCLUSION

This extensive empirical study will be of invaluable benefit to the research that digital credit scoring frameworks incorporating Chinese alternative data make significant improvements over traditional PBOC-based credit evaluation frameworks on a variety of predictive efficacies dimensions. The comparative analysis of six models representing both traditional statistical models and state-of-the-art machine learning algorithms demonstrates a consistent and statistically significant performance-benefit in digital methodologies in the Chinese own fintech ecosystem, with the XGBoost model that includes the data on Alipay, e-commerce, and social networks scoring an AUC-ROC of 0.932 against the 0.748 of logistic regression, and 0.715 of the score of PBOC-based scoring systems as a 24.6 percent improvement of predict

The advanced discriminative ability of machine learning model that ingest Chinese alternative data is directly translated into practical benefits to Chinese financial institutions, fintech platforms, and Chinese borrowers alike. The performance of Chinese P2P lending platforms can be more precisely priced on the increased level of default prediction, leading to decreased capital requirements of commercial banks and enhanced portfolio results on the 90 trillion RMB mobile payment ecosystem in China. At the same time, the incorporation of alternative data sources that are unique to China such as Alipay payment records, Taobao consumption history, and utility payment records facilitates financial inclusion by allowing creditworthy thin-file borrowers to receive formal financial services. This democratization of credit rating fills a severe void on some 350 million Chinese consumers who do not have a traditional credit record, which contributes to rural revitalization efforts and poverty reduction objectives in China.

The use of digital credit scoring systems in China, however, should be made with the cautious approach to interpretability, fairness, and regulatory compliance issues that have specifics to the Chinese market setting. The fact that complex machine learning models constitute black boxes implies that they require strong explainable AI systems that comply with CBIRC specifications and PBOC standards to promote transparency in lending decisions. To avoid discriminatory results embedded in the algorithmic decision-making processes against vulnerable groups such as rural inhabitants, migrant workers, and ethnic minorities, the Chinese financial institutions need to, firstly, implement comprehensive validation procedures, bias detection mechanisms tailored to the urban-rural divide and ethnic diversity of China, and, secondly, possess continuous monitoring systems that will be compliant with the Personal Information Protection Law (PIPL).

To conclude, this study confirms that digital credit scoring based on Chinese alternative data is a material innovation in the financial risk assessment approach to the Chinese market, with better predictive quality and providing greater access to credit to under-served segments of the population such as rural citizens, young urban professionals, and small business owners provided this innovation is applied responsibly with proper governance structures in line with the CBIRC regulations and PBOC principles. The results are in line with the objectives of the digital economy development in China and can serve as an addition to the efforts of financial inclusion that is vital in terms of equal economic development in the various regions of China.

REFERENCES

[1] Gambacorta, L., Huang, Y., Qiu, H., & Wang, J. (2024). How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm. Journal of Financial Stability, 73, 101365. https://doi.org/10.1016/j.jfs.2024.101365

[2] Toh, Y. L. (2023). Addressing traditional credit scores as a barrier to accessing credit. Economic Review, 108(3), 71–93. Federal Reserve Bank of Kansas City. https://www.kansascityfed.org/documents/9602/EconomicReviewV108N3Toh.pdf

[3] Dong, B., Gou, S., & Qiu, H. (2023). Do big tech credit scores enhance financial inclusion? Evidence from China's leading digital credit scoring service. Finance Research Letters, 51, 103415. https://doi.org/10.1016/j.frl.2022.103415

[4] Li, C., Nadeem, A., Raghunathan, A. T., & Gu, B. (2024). The effect of AI-enabled credit scoring on financial inclusion: Evidence from an underserved population of over one million. MIS Quarterly, 48(4), 1803–1834. https://doi.org/10.25300/MISQ/2023/17755

[5] Zhang, L., & Wang, L. (2025). Impact of social networks on digital credit assessment for rural residents: A study using machine learning methods. Information Systems Frontiers, 1–23. https://doi.org/10.1007/s10796-025-10630-1

[6] Rida, A. (2024). Machine and deep learning for credit scoring: A compliant approach. arXiv preprint arXiv:2412.20225. https://arxiv.org/abs/2412.20225

[7] Kamimura, M. A., et al. (2025). Machine learning powered financial credit scoring: a systematic literature review. Artificial Intelligence Review, 58(2), 1–45. https://doi.org/10.1007/s10462-025-11416-2

[8] Jiang, W., Liu, L., & Deng, X. (2021). Loan default prediction of Chinese P2P market: a machine learning methodology. Scientific Reports, 11, 18759. https://doi.org/10.1038/s41598-021-98361-6

[9] Huang, Y., Zhang, L., Li, Z., Qiu, H., Sun, T., & Wang, X. (2020). Fintech credit risk assessment for SMEs: Evidence from China (Working Paper No. WP/20/193). International Monetary Fund.

[10] Chen, X., Wu, M., & Wang, M. (2024). Application of business intelligence under deep neural network in credit scoring of bank users. Journal of Computational Methods in Sciences and Engineering, 24(3), 1567–1582. https://doi.org/10.3233/JCM-247294

[11] Mujo, A., Nikolla, S., Hoxha, E., & Pelivani, E. (2024). Explainable AI in Credit Scoring: Improving Transparency in Loan Decisions. Journal of Information Systems Engineering and Management, 10(27s). https://doi.org/10.55267/iadt.07.14970