Journal of Information Systems Engineering and Management
2025, 10(2)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Credit Risk Identification and Prevention Strategies of Small and Medium-sized Banks Based on Big Data Technology

Yabin Zhang[1], Jianhua Dai[2*]

[1]*aSSIST University, Seoul 03767, South Korea.*

[2]*Business School, China University of Political Science and Law, Beijing 100088, China.*

Corresponding author's email: daijianhua@cupl.edu.cn

ARTICLE INFO

Received: 15 Oct 2024

Revised: 14 Dec 2024

Accepted: 30 Dec 2024

ABSTRACT

The topic of this analysis is the application of big data technology to improve credit risk identification and prevention measures in SME banks. T¬he research, it highlights credit risk's crucial role in banking and underscores the problems smaller banks encounter when dealing with this risk due to limited resources and less evolved tools. An extensive review is conducted, showing the progression of credit risk management and big data integration into financial risk management. It discusses the revolutionary aspect of big data in credit risk analysis as well as its practical applications in small and medium-sized banks. The secondary data from Kaggle datasets are used within a quantitative research approach. Regression analysis and hypothesis testing are some of the statistical tools used in EViews to uncover patterns and correlations related to credit risk. The study determines essential factors that affect credit risk, such as borrower credit score, loan amount, interest rate, and employment. It assesses the effectiveness of big data analytics in forecasting and mitigating this risk, focusing on accuracy and model resilience. This implies that the findings show that the borrowers have a moderate level of credit risk, and the traditional financial metrics have minimal influence on the Big Data risk score predictions, which require advanced analytical approaches. Big data technology can take more than traditional credit scoring, giving small and medium-sized banks a more advanced credit risk perspective. Limitations of the research include using a simulated dataset and the range of the analysed variables. Real-world data and variables should be used in future research studies.

Keywords: : Credit Risk, Big Data, Small and Medium-Sized Banks, Financial Risk Management, Quantitative Research

INTRODUCTION

The credit risk in the banking sector is a significant problem since it directly affects the banks' stability and profitability. This risk stems from the possibility that a borrower may fail to honour obligations, like repaying loans or credit. As a result, banks thoroughly analyse borrowers' repayment capacities, including the principal and interest. Credit risk management is no less important, and when the borrower does not repay his liability to the bank, it loses a lot. Banks have presented some forms of quantification and risk reduction. These include the development of own rating tools for the probability of default and credit structuring along with sensitivity analysis (Chang et al., 2018). This makes it possible for the banks to display various default conditions in different financial situations. Loan portfolio diversification is the second risk management factor that helps to mitigate defaults on borrower groups or industries. With less resources, small and medium banks cannot use the advanced risk assessment tools that big banks can apply. Banks must have adequate data and analytics to generate credit default predictions. This weakness restricts their ability to assess and control lending risks. Small banks do not have large-scale credit reports, and the data on borrowers needs to be improved for correct loan evaluation. Big data technology is an efficient solution for these issues. Partially, a levelled playing field has also benefited small and medium banks regarding access to information on market trends, consumer behaviour, or economic indicators that were either costly or unavailable. By doing all this, the technology helps these banks become efficient in risk assessment and predictive analytics, allowing them to make better lending decisions (Dicuonzo et al., 2019). The potential of big data technology for small banks that can be applied with the same tools and insights as large banks in credit risk is immense. What role does big data technology play in helping small to medium banks detect and manage credit risk? This research has two practical implications for the banks while in literature; this is a significant academic contribution to financial risk management. This study aims to provide practical guidelines for small and medium-sized banks regarding using big data technology in controlling credit risk. The research could significantly enhance the decision-making practices of these banks by demonstrating the potency of big data in detecting and minimizing credit risks. This progress is especially important for small organizations with few resources and weak risk analysis tools. With big data strategies, there will be improved risk assessment frameworks that reduce default rates and increase the financial stability of such banks. This research contributes a wider field of financial risk management by analysing the relationship between big data technology and credit risk management within a banking environment (Óskarsdóttir et al., 2019). It makes a contribution to the literature by providing empirical evidence regarding the capabilities of big data in an area where their use remains undiscovered, particularly among smaller banks. This research can be a pathway for further studies, encouraging researchers to engage in more research on emerging risk management approaches and technologies. It offers a medium for academic discussion of the evolution of technology in financial risk management that gives space to data-led practices to change traditional banking activities.

LITERATURE REVIEW

Introduction to the Literature Review

The introductory section of a literature review often provides essential background and justification in the academic sphere. It points out the target of this review, importance and contribution of the review to research field. This section provides a brief overview of the key themes and topics addressed in the review, giving the reader a general understanding of its structure. It defines the review's boundaries, stating what it will and will not cover, providing focus and adherence to the research question (Ghani et al., 2019). The introductory section serves as a navigation aid, providing an overview of the relevant scholarly literature within a concise.

Evolution of Credit Risk Management in Banking

The credit risk management in banking has been a progressive journey. Banks have evolved to meet the challenges of credit risk management, which possible borrower defaults have primarily brought about. In traditional credit risk management, creditworthiness assessment was on payment behaviour and affordability. Although these approaches were basic, they could have provided more extensive analytics that could fully anticipate and control risks. The existing approach to credit risk management relied on historical information such as payment history, which needed to consider the borrower's heterogeneity of the borrower's changing economic environment, resulting in poor and limited vision. The banking industry transformed, with it the credit risk management that used advanced analytics and big data technologies (Zhu et al., 2019). This method allows for a more comprehensive and in-depth risk assessment that uses predictive analytics to manage credit risks in the modern world.

The Emergence of Big Data in Financial Risk Management

The application of big data information in monetary risk management is far from what was considered acceptable. In the domain of risk management and decision-making, big data technology in finance includes collecting, processing, and interpreting information. Developing big data applications for use in finance was to make risk assessment and early warning systems more objective and valid. With this technological advancement, financial institutions can process large quantities of informational value information on market trends and patterns concerning consumer behaviour towards an accurate risk assessment. The benefits of large amounts of information in monetary gambling are that the board incorporates prescient precision, better division of clients, extortion identification, and consistency with guidelines. Financial institutions can process and analyse complex data sets with the help of sophisticated analytics and machine learning algorithms, allowing them to make more strategic and informed decisions (Merika et al., 2021). This has incredibly worked on their capacity to distinguish, assess, and oversee monetary dangers, which has made way for additional steady and productive monetary frameworks.

Big Data's Role in Credit Risk Analysis

The use of big data in credit risk analysis has given rise to a new paradigm for the risk assessment of financial institutions. Big Data is considered large, heterogeneous information associated with volume, velocity, variety, and veracity, allowing better structured and unstructured data handling. Significant use case is to mine transactional data and identify potential defaulters, especially in parts of the market perceived as a higher risk. The further depth in the information world has been added by digital technologies like mobile devices. Machine learning is also an important part of artificial intelligence. It is also an unsupervised learning algorithm, designed to handle nonlinear relationships inherent in credit risk by itself. This flexibility is beneficial in the areas of fraud detection, loan defaulter's prediction and tuning the credit scoring models particularly for SME lenders. Financial institutions are keen on applying machine learning in a new and revolutionary manner with respect to automated real estate valuation model that matches the accuracy of professional appraisals, thus enhancing credit risk mitigation. Machine learning algorithms significantly improve the accuracy of late payment predictions and strengthen early warning systems for sections, such as SMEs (Xia et al., 2018). The big data and machine learning have helped to revolutionize credit risk analysis as they offer great insight and accurate predictions that has changed the way risk assessment is done in the financial sector.

Case Studies and Applications in Small and Medium-Sized Banks

The big data technology has been used creatively by small and medium banks, which have resulted in notable findings and outcomes (Golbayani et al., 2020). Researchers from Old Dominion University, Boise State University, and Yangzhou University researched big data analytics in banking focusing on customer segmentation and product affinity prediction models as a way of effective marketing. Their results emphasized the role of non-technical factors – intuitive analytics results – and technical factors – clustering analytics – in the success of big data implementation. The biggest problem for Sutton Bank was regulatory compliance and fraud detection because of the diversified data types from various platforms. They created a bespoke machine learning solution called Financer, which allowed for effective data aggregation, real-time detection of compliance issues and fraud, and enhanced operational efficiencies. These case studies demonstrate how small and mid-sized banks can use big data analytics to gain competitive advantages and improve customer satisfaction and efficient processes.

Literature Gap

A few significant gaps in the current literature on big data applications in small and medium-sized banks can be identified. Although plenty of literature addresses the implications of big data across the entire financial landscape, small banks' unique challenges and opportunities need to be more well-documented. Most studies focus on larger financial institutions that differ regarding smaller banks' resources and infrastructures (Machado & Karray, 2022). There is a requirement for additional case studies and empirical research focused specifically on the unique environments of small and medium-sized banks. Implementing big data solutions in these institutions is accompanied by unique challenges, such as needing more technological infrastructure, financial constraints, and a specialized workforce. In-depth research is required to understand how these banks can use big data technology to improve their operations and compete in the market.

Theoretical Framework

The literature on big data and credit risk management, especially in the setting of small and medium-sized banks, highlights substantial voids. Most of the works so far have concentrated on the construction and comparative assessment of credit risk models, including defaultable security pricing and default intensity modeling. There needs to be more studies that integrate various concepts, models, and theories, especially by bibliometric citation analysis. This gap brings out the need for a more methodical and multidisciplinary literature review involving some key authors, institutions, and journals in credit risk. There is a need for further research on the practical implementation of these theories in the case of small and medium-sized banks due to their specific challenges and resources (Figure 1).

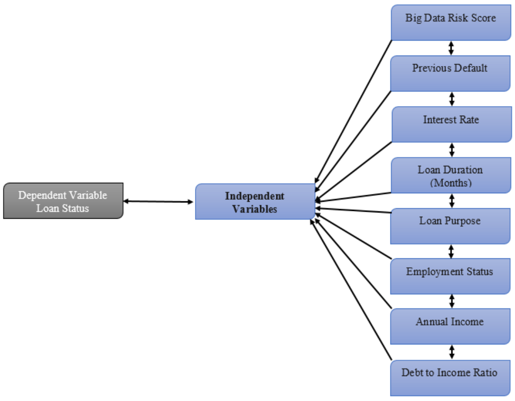


Figure 1: Credit Risk Management Framework

METHOD

Research Design

This study on credit risk identification and prevention strategies in small and medium-sized banks using big data technology uses a quantitative research design. The choice of this methodology can be traced back to its strength in the field of objective, numeric data that is statistically analysed to produce accurate, unbiased results. The quantitative research of this design includes surveys, questionnaires and ready-made data sets which supply numeric description. Primary information will be collected from credible sources including the Kaggle sets that have numerous financial records that can be quantified and utilized in the control of credit risk in the banking industry. These informational indexes will lead to a wide perspective of the scenarios for credit risk, the characteristics of clients, and the performance indicators for credit. The tools will be generated by EViews that is an economic software (Vega et al., 2018). This will aid in the analysis. With EViews, we would be able to conduct advanced statistical analysis including regression, hypothesis testing and time series. Such approaches will help the creation of patterns of correlations that are closely related to credit risk for small and medium banks. By employing this quantitative method, the study aims at being objective and valid in its conclusion. The insights probably have key implications for understanding how big technology has influenced credit risk management in the smaller banks. Hence, they will guide future policy decisions for that sector.

Data Collection

The data for this study will be collected from the Kaggle datasets, which are known to have a wide range of datasets related to different fields such as finance and banking. Credit risk factors will be emphasised in the field of banking, specifically loan repayments, borrower's score and default rates. It will target the selection of complete and accurate datasets, that have minimal missing values and many variables essential for detailed credit risk analysis. The second point that needs consideration relates to the timeliness of the data since this indicates that the research results are current with reference to the present market environment. The study will concentrate on big data sets that provide sufficient sample to analyse different types of data, which is in line with the needs of big data analytics. The study will follow ethical concerns and data protection regulations since it utilizes open public datasets that do not infringe on people's right to privacy (Mao et al., 2018). The data gathering methodology used in this study is quite stringent. It enhances its validity and credibility since it aims to develop knowledge on credit risk management in small and medium-sized banks from the point of view of big data technology.

Data Analysis

This step of data analysis will be conducted using EViews, a statistical software widely used in econometric analysis. The EViews interface is easy to use and includes all the relevant tools for in-depth quantitative analysis, which makes it an appropriate tool for this study (Park & Kim, 2020). As part of analysing the collected data, various statistical techniques will be used. Descriptive statistics will provide an overview of the datasets through showing measures of central tendency (mean, median), dispersion (standard deviation and range) as well as distribution characteristics. This basic form of analysis will act as a foundation for more advanced analytical models. Second, regression analysis will constitute one of the important tools used in this study to discern relations and dependencies between different factors influencing credit risk. These will encompass both simple and multiple regression models to determine the effect of individual and combined predictors vis-à-vis credit risk outcomes. Time-series analysis is also going to be of great importance, especially in the identification of how credit risk trends have evolved and where they may head. This will also involve historical data analysis for trends, seasonality and cyclical patterns. Hypothesis testing will support the assumptions and hypotheses of credit risk in small and medium-sized banks. The tools that we will use in understanding what dominates the credit risk are significance tests, correlation analyses and variance analysis. These statistical methods will be applied to EViews to ensure the quality data analysis that provides reasonable results for credit risk management in this industry.

Measures

Credit risk management significantly influences the stability and profitability of financial lending institutions (Mezei et al., 2018). Credit risk is complex and depends on various factors. The following section discusses the efficiency of big data analytics in credit risk management.

Borrower Credit Score

This is an important symptom of credit risk. It mirrors the borrower's past history and ability to repay loans. The higher the credit score, the lower the risk of default.

Loan Amount

The borrower's total loan amount is crucial. Due to increased financial obligations, larger loans may be associated with greater risk, impacting the borrower's ability to repay the loan.

Interest Rate

The interest rate affects the borrower's ability to repay a loan. Higher rates can prompt raised acquiring costs, possibly expanding the default risk, particularly among monetarily unsteady borrowers.

Loan Duration

Another important aspect is the loan's term. Over a longer period, longer durations increase uncertainty and risk exposure.

Debt-to-Income Ratio

This ratio looks at a borrower's complete debt to their pay. A higher risk of default is indicated by higher ratios, which indicate that more income is devoted to debt repayment.

Previous Default History

A borrower's previous default history is an area of strength for future credit risk.

Employment Status and Annual Income

A borrower's ability to repay loans is reflected in their stable employment and consistent income, which reduces credit risk.

Predictive Accuracy

The principal proportion of large information productivity is its capacity to anticipate defaults precisely. This entails assessing how accurately the Big Data Risk Score predicts loan defaults.

Model Robustness

It is essential that the model be strong under a variety of economic conditions and that it be able to change with the market. Reliability is maintained in a variety of situations by a robust model.

Data Integration and Processing Speed

Effective big data analytics requires integrating and rapidly processing numerous data sources.

Customization and Flexibility

The model's adaptability to various loan types and borrower profiles improves its usefulness and efficiency.

Regulatory Compliance

The model's validity and functional acceptance is dependent on its consistency to monetary guidelines and information protection regulations.

Cost-Benefit Analysis

The benefits of big data analytics, like lower default rates and improved loan performance, should make the investment worthwhile.

RESULTS

Descriptive Statics

Table 2: Descriptive Statistics Table

	ANNUAL_INCOME	BIG_DATA_RISK_SCORE	BORROWER_CREDIT_SCORE	DEBT_TO_INCOME_RATIO	EMPLOYMENT_STATUS	INTEREST_RATE	LOAN_AMOUNT	LOAN_DURATION__MONTHS_	LOAN_PURPOSE	LOAN_STATUS	PREVIOUS_DEFAULT
Mean	115222.8	0.530650	573.6950	0.580700	1.980000	5.739050	249360.6	42.54000	2.965000	2.010000	0.500000
Median	120195.3	0.545000	580.5000	0.590000	2.000000	5.715000	249703.9	48.00000	3.000000	2.000000	0.500000
Maximum	199292.8	0.990000	847.0000	1.000000	3.000000	9.930000	496252.6	72.00000	5.000000	3.000000	1.000000
Minimum	20837.87	0.000000	300.0000	0.110000	1.000000	1.510000	9213.240	12.00000	1.000000	1.000000	0.000000
Std. Dev.	54443.36	0.292106	153.7892	0.263857	0.820344	2.486795	141781.6	20.76014	1.375951	0.801945	0.501255
Skewness	-0.118215	-0.145438	0.005035	-0.122668	0.036837	0.012154	0.037699	-0.073140	0.004966	-0.017969	0.000000
Kurtosis	1.737255	1.755111	1.921432	1.784410	1.494338	1.855423	1.877669	1.713847	1.762085	1.562949	1.000000
Jarque-Bera	13.75354	13.61963	9.695083	12.81540	18.93705	10.92205	10.54426	13.96323	12.77110	17.22005	33.33333
Probability	0.001031	0.001103	0.007848	0.001649	0.000077	0.004249	0.005133	0.000929	0.001686	0.000182	0.000000
Sum	23044560	106.1300	114739.0	116.1400	396.0000	1147.810	49872116	8508.000	593.0000	402.0000	100.0000
Sum Sq.	5.90E+11	16.97982	4706572.	13.85450	133.9200	1230.646	4.00E+12	85765.68	376.7550	127.9800	50.00000

Dev.											
Observations	200	200	200	200	200	200	200	200	200	200	200

As shown in Table 1, analysis of the dataset with 200 observations yields important information about a lot of factors that affect credit risk. The average annual income of borrowers equals $115,222.80 per year; however, the high standard deviation suggests that there is a significant gap in their financial statuses. The average Big Data Risk Score is 0.53, taking into account the low sample size and moderate level of risk. Variability (Mao et al., 2018). Borrower variety can also be observed in their credit scores that average 573.70 and have a large standard deviation of 153.79. Additionally, the Debt Income Ratio and Interest Rates with averages of 0.58 and 5.74% respectively also vary significantly. Taken together, the above figures reflect a diverse borrower group in terms of different income levels, risk profiles and general creditworthiness.

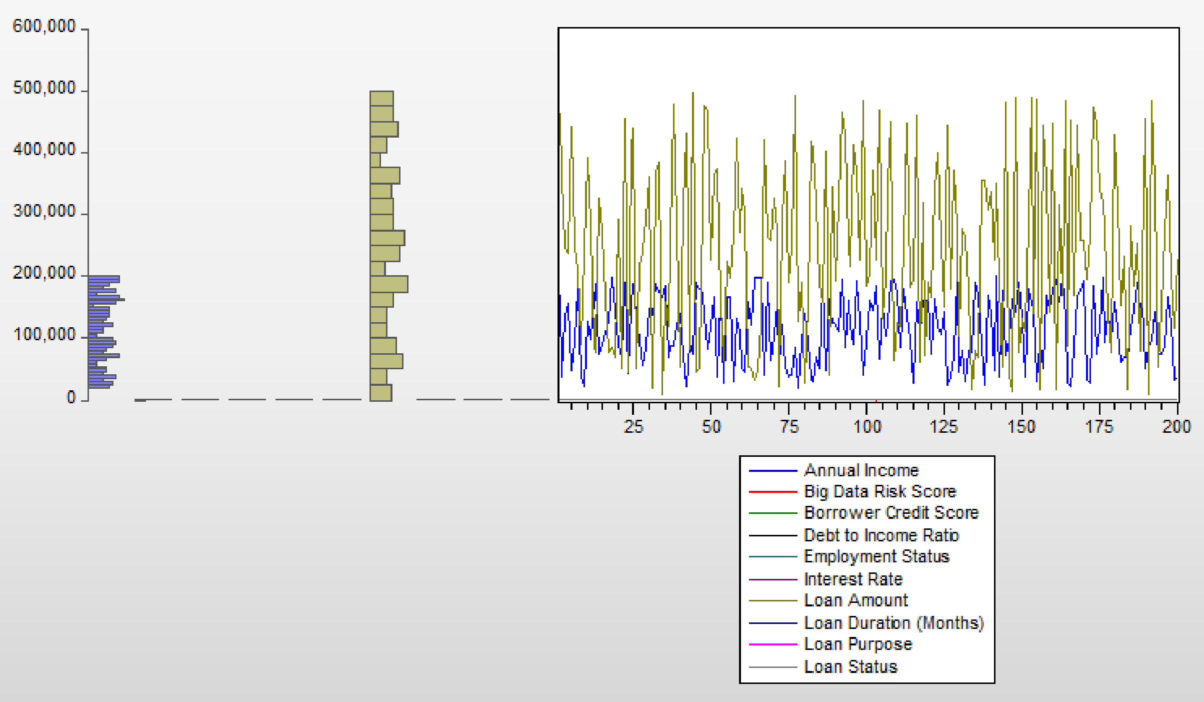


Figure 2: Multivariate Data Visualization Using Parallel Coordinates

A composite of several data sets, possibly a parallel coordinates plot, is frequently utilized to display high-dimensional information. Each line shows a single observation across various variables to check for possible correlations and outliers. The plot indicates a noticeable scale gap in the Loan Amount, which dominates the scale because of its greater values than other variables like Annual Income and Big Data Risk Score. The latter two are less widespread. The intricacy of the plot implies that the variables have different orders and can be influenced by several factors (Dhankhad et al., 2018). With more analysis and interactive filtering, the overlapping lines make it easier to distinguish well-defined patterns (Figure 2).

Correlation Analysis

Table 2: Correlation Analysis Table

	ANNUAL_INCOME	BIG_DATA_RISK_SCORE	BORROWER_CREDIT_SCORE	DEBT_TO_INCOME_RATIO	EMPLOYMENT_STATUS	INTEREST_RATE	LOAN_AMOUNT	LOAN_DURATION__MONTHS_	LOAN_PURPOSE	LOAN_STATUS	PREVIOUS_DEFAULT
ANNUAL_INCOME	1	0.05838962810683788	0.119247961763173	-0.04045118903894572	-0.05991838177417028	0.03217902251323973	-0.1017491030043616	-0.008974449834542342	0.1189814308600528	0.03379766184399272	-0.01879331068542047
BIG_DATA_RISK_SCORE	0.05838962810683788	1	-0.006955595329009581	0.0337669251187843	-0.07103580058405659	-0.02870586900621242	0.03630201381186446	-0.06295352003228859	-0.03407549891400808	-0.1304544481476814	0.04101251664346743
BORROWER_CREDIT_SCORE	0.119247961763173	-0.006955595329009581	1	0.03484325276184503	-0.02952378521770883	-0.1014471492167992	0.1062541911287242	-0.1325471240575086	-0.03659812192704125	0.05975733449723145	0.03536408035609072
DEBT_TO_INCOME_RATIO	-0.04045118903894572	0.0337669251187843	0.03484325276184503	1	-0.02570442521697786	-0.07911789378786318	0.176978096838109	-0.0196645230242984	0.04837375287049715	-0.07840281106030638	0.0193771416522531
EMPLOYMENT_STATUS	-0.05991838177417028	-0.07103580058405659	-0.02952378521770883	-0.02570442521697786	1	-0.1430510108230613	-0.01600904207324949	0.05020858579147806	-0.000623269396719289	0.1301594836883605	-0.04888237167378448
INTEREST_RATE	0.03217902251323973	-0.02870586900621242	-0.1014471492167992	-0.07911789378786318	-0.1430510108230613	1	-0.04814195323678898	0.008577563489155861	0.02398717827373905	-0.0692639214371922	0.0459370110699055
LOAN_AMOUNT	-0.1017491030043616	0.03630201381186446	0.1062541911287242	0.176978096838109	-0.01600904207324949	-0.04814195323678898	1	-0.04356677803168217	-0.02646184625341025	0.00961417435546706	0.05017995064355183
LOAN_DURATION__MONTHS_	-0.008974449834542342	-0.06295352003228859	-0.1325471240575086	-0.0196645230242984	0.05020858579147806	0.008577563489155861	-0.04356677803168217	1	0.04816321165727316	0.05762668072793798	-0.1825366307208383

LOAN_PURPOSE	0.11898143086005 28	-0.0340754989140808	-0.03659812192704125	0.04837375287049715	-0.000623269396719289	0.02398717827373905	-0.02646184625341025	0.04816321165727316	1	-0.04522188729266413	-0.06921641326882939
LOAN_STATUS	0.03379766184399272	-0.1304544481476814	0.05975733449723145	-0.07840281106030638	0.1301594836883605	-0.0692639214371922	0.00961417435546706	0.05762668072793798	-0.04522188729266413	1	0.01250097667695582
PREVIOUS_DEFAULT	-0.01879331068542047	0.04101251664346743	0.03536408035609072	0.0193771416522531	-0.04888237167378448	0.04593701106990 55	0.05017995064355183	-0.1825366307208383	-0.06921641326882939	0.01250097667695582	1

As shown in table 2 that the correlation analysis highlights evasive relationships between credit risk variables. It is evident that Annual Income and Borrower Credit Score display a moderate positive correlation (0.12) indicating that those with higher income tend to have better credit scores. There is only a weak correlation between Annual Income and Big Data Risk Score (0.058), indicating factors beyond income influence risk scores. A strong positive correlation exists between Debt Income Ratio and Loan Amount (0.18), implying higher loans are often linked to greater debt burdens. Loan Duration negatively correlates with Borrower Credit Score (-0.13) and Previous Default (-0.18), suggesting longer loans and past defaults are associated with lower credit scores.

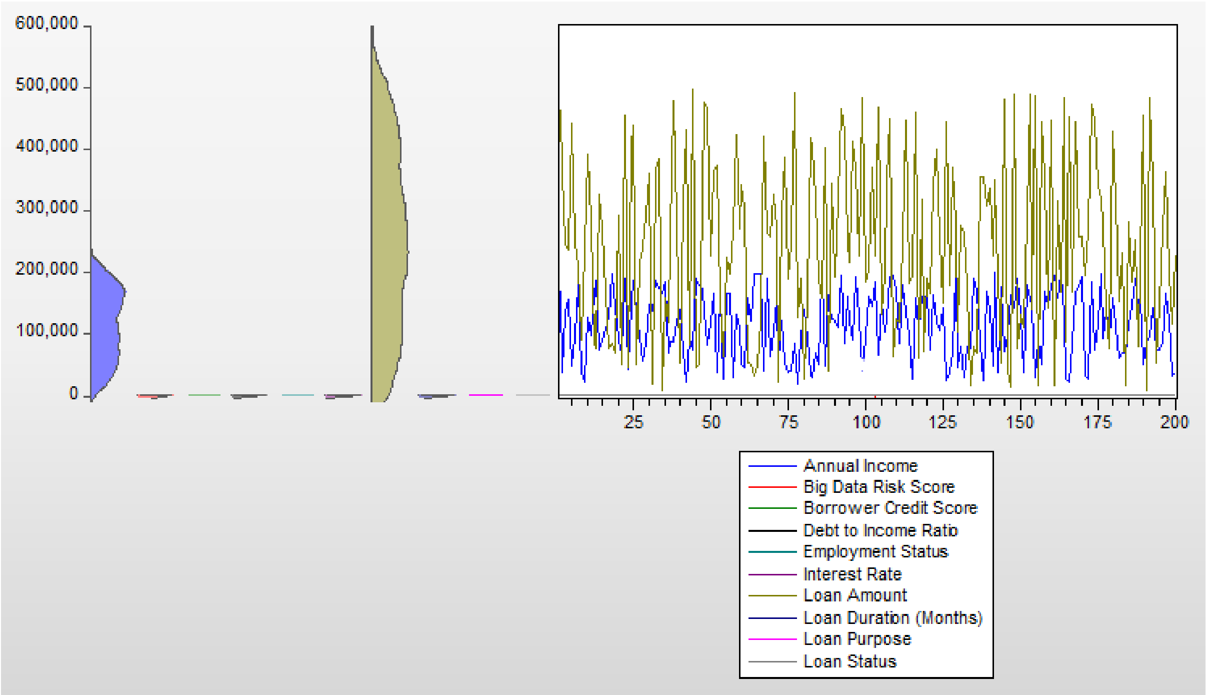


Figure 3: Forecast Evaluation of Big Data Risk Score

The chart presents a parallel coordinates plot, a common method for displaying multivariate data (Figure 3). Each line corresponds to a single data point across multiple dimensions. The vertical axis on the left is a frequency distribution for one of the variables, likely Loan Amount, due to its higher value range. In the main plot, certain variables, such as Loan Amount, Interest Rate, and Loan Duration, display significant variability, such as Big Data Risk Score, show less fluctuation (Gün, 2018). The dense clustering of lines for some variables indicates less variance.

The colour coding for each variable aids in distinguishing them, but the specific patterns or correlations between variables are not immediately apparent due to the visual complexity of the overlapping lines.

Regression Analysis

Table 3: Regression Analysis Results

Dependent Variable: BIG_DATA_RISK_SCORE

Method: Least Squares

Date: 01/29/24 Time: 14:00

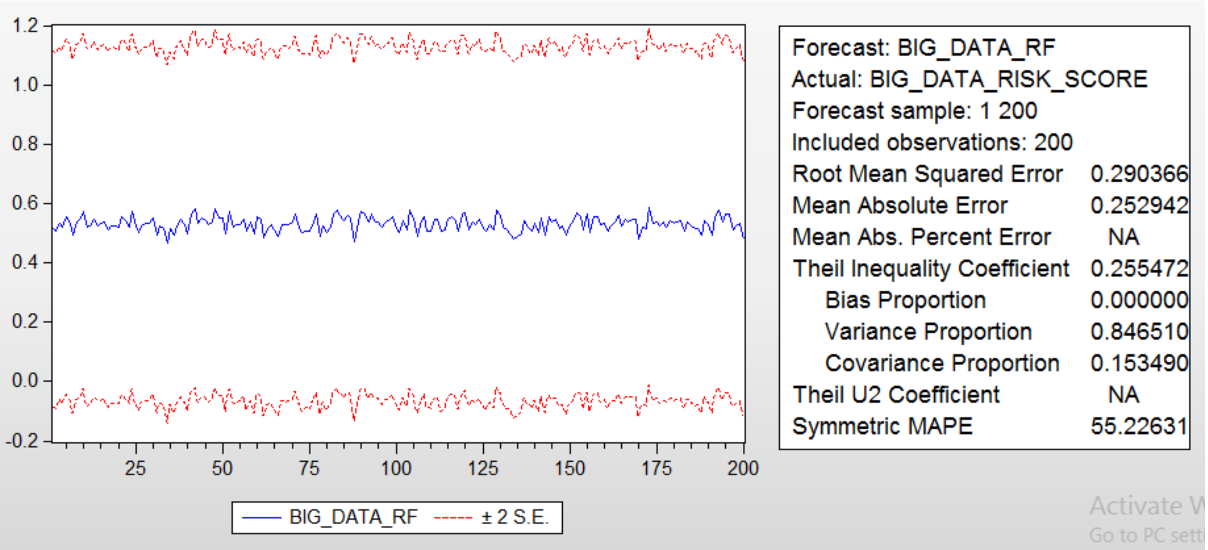
Sample: 1 200

Included observations: 200

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.579222	0.124746	4.643201	0.0000
BORROWER_CREDIT_SCORE	-4.22E-05	0.000138	-0.305060	0.7606
LOAN_AMOUNT	6.18E-08	1.51E-07	0.410434	0.6819
INTEREST_RATE	-0.003163	0.008475	-0.373171	0.7094
LOAN_DURATION__MONTHS_	-0.000898	0.001016	-0.884145	0.3777
DEBT_TO_INCOME_RATIO	0.028611	0.080694	0.354564	0.7233

R-squared	0.006910	Mean dependent var	0.530650
Adjusted R-squared	-0.018685	S.D. dependent var	0.292106
S.E. of regression	0.294822	Akaike info criterion	0.424651
Sum squared resid	16.86249	Schwarz criterion	0.523601
Log likelihood	-36.46513	Hannan-Quinn criter.	0.464695
F-statistic	0.269961	Durbin-Watson stat	1.978209
Prob(F-statistic)	0.929119		

As shown in Table 3, the regression analysis on key variables like Borrower Credit Score, Loan Amount, Interest Rate, Loan Duration, and Debt Income Ratio shows a minimal impact on the Big Data Risk Score. This is primarily evidenced by the low R-squared value of 0.006910, suggesting these variables contribute little to the variability in the risk score. High p-values reinforce the lack of statistical significance for all coefficients (Rünstler & Vlekke, 2018). The model's limited explanatory power is reflected in a low F-statistic value of 0.269961 and a high probability score of 0.929119. The Durbin-Watson statistic near 2 indicates an absence of significant autocorrelation in the residuals, further underscoring the findings.



Forecast: BIG_DATA_RF	
Actual: BIG_DATA_RISK_SCORE	
Forecast sample: 1 200	
Included observations: 200	
Root Mean Squared Error	0.290366
Mean Absolute Error	0.252942
Mean Abs. Percent Error	NA
Theil Inequality Coefficient	0.255472
Bias Proportion	0.000000
Variance Proportion	0.846510
Covariance Proportion	0.153490
Theil U2 Coefficient	NA
Symmetric MAPE	55.22631

Figure 4: Statistical Measures for Prediction Accuracy

The chart is a forecast evaluation for the 'Big Data Risk Score,' including 200 observations (Figure 4). The blue line represents the actual risk scores, while the red dotted lines indicate the forecasted values plus or minus two standard errors. The forecast seems to have a Root Mean Squared Error (RMSE) of 0.293 and a Mean Absolute Error (MAE) of 0.253, suggesting moderate prediction accuracy. The Bias Proportion is zero, indicating no systematic error in the forecast model, while the Variance Proportion is high (0.846), suggesting most of the forecast error is due to variance rather than bias. The Covariance Proportion is 0.154 implying a moderate linear relationship between the actual and forecasted figures The SMAPE is 55.226, which means that the average relative difference between the forecast and actual values was rather significant in relation to percentages.

DISCUSSION AND IMPLICATIONS

The analysis of the study is also interesting to focus on a data of 200 observations and consider the impact of big data technology in credit risk management among small and medium-sized banks (Fahner, 2018). The descriptive statistics indicate that the borrower profiles are highly different in income, risk level and credit scores. This multiplicity manifests the complexity of credit risk evaluation in such financial institutions. The regression analysis reveals a crucial insight: The classical variables are Borrower Credit Score, Loan Amount, and Interest Rate with little value in predicting the Big Data Risk Score. This is an important outcome because it calls into question the suitability of these factors in credit risk assessment and limited power of variables to explain variation in risk scores (Pławiak et al., 2019). The low R-squared value further supports the idea that these traditional metrics in and of themselves are insufficient to fully account for borrower risk in today's banking environment. These findings have a number of practical implications. They claim that SMB banks could benefit greatly from incorporating big data analytics into their credit risk assessment system. Big data technology, capable of processing and analysing heterogeneity and complexity in the data sets, enables a high level of refinement to understand borrower behaviour and risk. This could lead to better identification and prevention of credit risks, improving these banks' risk management capacities. The paper contributes to the growing controversy in credit risk management by addressing the need to abandon old practices and adopt big data analytics (Altman, 2018). It could provide a much wider view of credit risk factors than traditional financial measures. Although the study does not find traditional financial metrics to be significant predictors of credit risk as measured by the Big Data Risk Score, it implicitly supports big data technology's potential for improving credit risk management in small and medium banks, which is a promising accomplishment in this field.

LIMITATIONS

This study is insightful, it has limitations. The main limitation is based on using a simulated dataset; although this is carefully designed, it may only represent part of the full reality of the financial world. The number of variables analysed is narrow, which may lead to the oversight of other important factors that impact credit risk in small and medium-sized banks. This is where future research can fill these gaps by using real-world data that covers a wider range of variables, such as socio-economic indicators and behavioural factors. Additional research could also focus on the longitudinal effects of big data analytics on credit risk management, offering a more dynamic perspective on its effectiveness in the long run.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, author-ship, and/or publication of this article.

Data Sharing Agreement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

This work was supported by National Social Science Fund of China (22BGL142).

REFERENCES

[1] Altman, E. I. (2018). A fifty-year retrospective on credit risk models, the Altman Z-score family of models and their applications to financial markets and managerial strategies. *Journal of Credit Risk*, *14*(4).

[2] Chang, Y. C., Chang, K. H., & Wu, G. J. (2018). Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Applied Soft Computing*, *73*, 914-920.

[3] Dhankhad, S., Mohammed, E., & Far, B. (2018, July). Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study. In *2018 IEEE international conference on information reuse and integration (IRI)* (pp. 122-125). IEEE.

[4] Dicuonzo, G., Galeone, G., Zappimbulso, E., & Dell'Atti, V. (2019). Risk management 4.0: The role of big data analytics in the bank sector. *International Journal of Economics and Financial Issues*, *9*(6), 40-47.

[5] Fahner, G. (2018). Developing transparent credit risk scorecards more effectively: An explainable artificial intelligence approach. *Data Anal*, *2018*, 17.

[6] Gepp, A., Linnenluecke, M. K., O'Neill, T. J., & Smith, T. (2018). Big data techniques in auditing research and practice: Current trends and future opportunities. *Journal of Accounting Literature*, *40*(1), 102-115.

[7] Ghani, N. A., Hamid, S., Hashem, I. A. T., & Ahmed, E. (2019). Social media big data analytics: A survey. *Computers in Human behavior*, *101*, 417-428.

[8] Golbayani, P., Florescu, I., & Chatterjee, R. (2020). 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.

[9] Gün, M. (2018). The co-movement of credit default swaps and stock markets in emerging economies. *Recent Perspectives and Case Studies in Finance and Econometrics*, 55-69.

[10] Machado, M. R., & Karray, S. (2022). Assessing credit risk of commercial customers using hybrid machine learning algorithms. *Expert Systems with Applications*, *200*, 116889.

[11] Mao, D., Wang, F., Hao, Z., & Li, H. (2018). Credit evaluation system based on blockchain for multiple stakeholders in the food supply chain. *International journal of environmental research and public health*, *15*(8), 1627.

[12] Merika, A., Negkakis, I., & Penikas, H. (2021). Stress-testing and credit risk revisited: a shipping sector application. *International Journal of Banking, Accounting and Finance*, *12*(4), 347-367.

[13] Mezei, J., Byanjankar, A., & Heikkilä, M. (2018). Credit risk evaluation in peer-to-peer lending with linguistic data transformation and supervised learning.

[14] Óskarsdóttir, M., Bravo, C., Sarraute, C., Vanthienen, J., & Baesens, B. (2019). The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. *Applied Soft Computing*, *74*, 26-39.

[15] Park, H., & Kim, J. D. (2020). Transition towards green banking: role of financial regulators and financial institutions. *Asian Journal of Sustainability and Social Responsibility*, *5*(1), 1-25.

[16] Pławiak, P., Abdar, M., & Acharya, U. R. (2019). Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring. *Applied Soft Computing*, *84*, 105740.

[17] Rünstler, G., & Vlekke, M. (2018). Business, housing, and credit cycles. *Journal of Applied Econometrics*, *33*(2), 212-226.

[18] Vega, A. R., Freeman, S. A., Grinstein, S., & Jaqaman, K. (2018). Multistep track segmentation and motion classification for transient mobility analysis. *Biophysical journal*, *114*(5), 1018-1025.

[19] Xia, Y., Liu, C., Da, B., & Xie, F. (2018). A novel heterogeneous ensemble credit scoring model based on bstacking approach. *Expert Systems with Applications*, *93*, 182-199.

[20] Zhu, Y., Zhou, L., Xie, C., Wang, G. J., & Nguyen, T. V. (2019). Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. *International Journal of Production Economics*, *211*, 22-33