International Journal For Business, Management And Accounting

GREEN PUBLICATION

ISSN(Online): 2455-6114
Volume 10 Issue 02 June 2024
Available on-line at: https://ijrbma.com/index.php/bma/issue/archive

DOI:*https://doi.org/10.53555/bma.v10i2.2486* *Received: 12-Mar-2024 Accepted:18-May-2024*

MACHINE LEARNING-BASED CREDIT RISK MODELING IN RETAIL BANKING

Rohan Malhotra[1]*, Kavya Menon[2], Sandeep Patil[3]

[1] Department of Banking and Financial Services School of Management NMIMS University Mumbai, Maharashtra, India
[2] Department of Business Analytics and Data Science SRM Institute of Science and Technology Chennai, Tamil Nadu, India
[3] Department of Finance and Quantitative Methods Institute of Management Studies Devi Ahilya University Indore, Madhya Pradesh, India

***Corresponding Author:** Rohan Malhotra*
**Department of Banking and Financial ServicesSchool of Management, NMIMS University Mumbai, India*
Email: rohan.malhotra@nmims.edu

Abstract
Machine learning applications are becoming more and more important in the improvement of credit risk assessment in the field of retail banking. In detecting the default of the borrowers, the paper will compare the predictive power of different classification models using structured borrower level data, which will be made up of borrower characteristics, variables of credit history and facts of the loan. The applied models were Decision Tree, Random Forest and Gradient Boosting models that were implemented and evaluated in a single validation framework. The following process made up the data preprocessing tasks: the process of missing values, replacement of categorical variables, elimination of post-loan performance variables to prevent the leakage of data and the problem of the use of class imbalance to ensure that sound estimation was obtained. Primary performance measure that was adopted to test the performance of the model was the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) that was supported by accuracy, precision, recall, and F1-score. These results show that the ensemble-based methods are more effective than the single-tree methods with Gradient Boosting that has the greatest ability to discriminate followed closely by the Random Forest. The importance of features analysis has indicated that the most important predictors of the risk associated with the default of the borrower are the interest rate, credit utilization, and debt-to-income ratio. The findings confirm the argument that group learning approaches can be very useful in improving classification and predictive stability of retail credit risk modeling.

Keywords: *Credit Risk Modeling, Machine Learning, Retail Banking, Gradient Boosting, AUC-ROC*

1. Introduction

Credit risk assessment is viewed as one of the most significant tasks of the retail banking as it has a direct impact on the lending decisions, profitability, and solvency. Identification of the possible default risk would allow the banking institutions to raise capital in the most efficient way with the least exposure to non-performing loans. Traditionally, the use of statistical and logistic regression scoring was used in credit scoring due to their simplicity and interpretation. However, the explosive growth of the electronic financial information and the creation of the computing methods have affected the practice of credit risk modeling to a significant level. Machine learning algorithms now have the capability of having better predictive results in learning and capturing non-linear relationships and complex interactions of borrower characteristics.

Empirical literature on this suggests a great deal of evidence that shows that the new classification algorithms are more effective than the old statistical models in credit scoring. A long benchmarking experiment of various state-of-the-art algorithms also demonstrated that in the predictive accuracy, the ensemble methods are by all means more effective as compared to the traditional models (Lessmann et al., 2015). This kind of a revelation has prompted researchers and practitioners to adopt better machine learning models in the process of evaluating credit risk more accurately.

Among them has been boosting and the technique has been very helpful in credit scoring. A larger impact on the classification performance was found to be caused by an improved decision tree model that was optimized using the Bayesian hyper-parameter optimization resulting in the effective detection of nonlinear trends among the data of the borrowers (Xia et al., 2017). Similarly, the comparative analysis of the base classifiers in the ensemble configurations indicate that several learning algorithms are resistant and less likely to be affected by the prediction variance (Abellan and Castellano, 2017). These findings identify the importance of model optimization and selection of algorithm in the retail credit modeling.

Among the ensemble-based approaches, gradient Boosting and extreme gradient boosting (XGBoost) were regarded as rather popular. Empirical evidence that supports the claim includes XGBoost having a superior discriminatory power and stability in credit risk evaluation of financial institutions (Chang et al., 2018). Later research that integrated deep learning with gradient boosting models demonstrated that combination models improve the predictive accuracy, particularly when dealing with large-scale loan level data (Petropoulos et al., 2019). These developments amuse the increased significance of advanced group practices in the modern retail banking environments.

There are other machine learning paradigms that have also been explored to assist credit scoring systems along with ensemble learning. It has proposed that self-organizing maps can serve as a transfer learning model to not only enhance the generalization of the models, but also the precision of classification in the credit risk scenarios (AghaeiRad et al., 2017). In addition, it has also been shown that unsupervised clustering is enhanced through supervised classification to be used in promoting prediction and a better borrower segmentation (Bao et al., 2019). These approaches are concerned with the applicability of feature engineering and data representation to credit risk modeling.

Even though predictive accuracy is crucial, the remaining factors that are important in the finance decision making process are transparency and interpretability. The practices of regulation are making automated decisions of credit to be justifiable and explainable. The explainable AI research on credit risk management reveals that the interpretability can enhance trust, regulatory compliance, and acceptance by the managers without causing a significant decrease in the performance of the models (Ariza-Garzon et al., 2020; Bussmann et al., 2021). Such findings suggest that there is a trade off between performance improvement and model transparency in retail banking applications.

The other important factor in credit scoring research work is associated with data pre-processing and research assessment procedure. The impact of sample selection and class imbalance on the performance of an ensemble is thoroughly investigated, and the outcomes of well-balanced datasets help to realize that the proper management of the imbalanced datasets can impact the model performance and discrimination ability to a considerable extent in a positive manner (Garcia et al., 2019). Therefore, it must welcome effective validation processes and performance measures that are relevant to make meaningful model comparison.

The overall literature indicates that machine learning-based algorithms, in particular, the ensemble models, such as random forest and gradient boosting, can greatly surpass the traditional statistical models in terms of credit risk prediction. However, the empirical evaluation will be required in the future so as to measure the model performance under different conditions and formats of retail lending. Also, predictive accuracy and interpretability are a pair whose combination are still two of the primary concerns in the credit risk management.

Relying on such developments, the contemporary study will evaluate and compare the different machine learning solutions to retail credit risk modeling using structured data at the level of loan. This paper, which uses a logistic regression, decision tree, random forest and gradient boosting technique in the context of a standard validation framework, confirms the growing body of evidence of the application of data-driven credit risk assessment. It is also a study on the patterns of importance of features to exude viable knowledge on the determinants of borrower default. This general evaluation assists the research to add to the current information on machine learning in retail banking, and to create more accurate and reliable credit rating frameworks. The study objectives are:

1. To assess and evaluate the predictive performance of different machine learning models in retail credit risk assessment.
2. To assess the key financial and credit-related determinants influencing borrower default risk.
3. To evaluate the effectiveness of ensemble learning techniques in improving credit risk prediction accuracy.

2. Methodology

2.1 Research Design

This research paper took a quantitative empirical design of research to design and test machine learning models to predict credit risk in retail banking. In the study, the authors adopt a predictive modelling design based on systematized loan level data to analyse the characteristics of borrowers that are linked to default behaviour. It is a design that is observational and analytical, attempting to find statistically significant relationships between financial data as opposed to comparing two or more variables through the establishment of experimental causality.

Controlled machine learning methods were used to categorize borrowers as either default or non-defaulters. This strategy was also in line with the current trends in risk management practices in banking, where predictive analytics are used to assist in credit judgement and risk management of portfolios. The quantitative design was a guarantee of objective evaluation of the model performance and enhances the validity of the empirical results.

2.2 Data Source and Sample

The empirical study used a secondary data consisting of about 10,000 retail consumer loan data acquired through the Lending Club Loan Dataset (Singh, 2023). The data set comprised of the demographic details of the borrower, credit history variables and loan specific variables, and the variables are in the real world retail lending environment.

Loans whose final repayment status was not in place were eliminated to bring about clarity in the outcome measurement. Loans that were coded as non- default (0) were those that were fully repaid and loans that were coded as default (1) were charged-off or severely delinquent. In the case of loans in the status of continuous repayment, exclusion was made to eliminate classification ambiguity. This dichotomous operationalization is in accordance with the prevailing credit risk modeling tradition in retail banking.

2.3 Variables and Measures

The dependent variable was the credit default status which is a binary variable indicating default or non-default borrowers. There were three categories of independent variables. Borrower attributes are annual earnings, years of employment, ownership of a home and ratio of debt to income which reflect financial ability and repayment predictability. The variables used in credit history are past delinquency, public record bankruptcies, credit utilization ratio, total and open credit lines, and recent credit inquiries that are indications of past credit behavior and exposure to risks. The characteristics of loans were loan amount, interest rate, loan term, loan purpose, and credit grade assigned. The variables indicating the post-loan repayment activity were not included so that the information may not leak and predictions can be based on factors that are available at the time of loan inception.

2.4 Data Processing and Preparation

The data was preprocessed systematically before analysis in order to maintain accuracy and internal consistency. Median imputation was used in cases where the numerical values were missing whereas appropriate methods of substitution were used in case of the categorical variables. Numerical values were used to represent categorical characteristics so that it can be used with machine learning algorithms.

The continuous variables were standardized where needed to improve the computation speed. Since the percentage of default cases was less than no-default cases, the imbalance in classes was evaluated and adjusted with the help of resampling and class-weight-adjustments.

The dataset was further separated into the training (80) and testing (20) samples to test out-of-sample predictive accuracy, and reduce overfitting.

2.5 Data Analysis Techniques

The initial statistical analysis was descriptive and was performed to summarize borrower characteristics, credit indicators and loan attributes. Correlation was then conducted to analyse relations between variables and identify multicollinearity.

Several of the supervised machine learning models were carried out, among them being logistic regression, decision tree, random forest, and gradient boosting algorithms. Accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) were used to evaluate the model performance.

It is a rigorous empirical approach to evaluating the performance of machine learning methods in retail credit risk modelling.

3. Results

3.1 Descriptive Statistics

The descriptive statistics is calculated to summarize the borrower characteristics, credit history indicators and loan characteristics. Table 1 indicates that new borrowers earned an average of 75,420 annually, although the variance was high between the observations. The debt-to-income ratio is 17.84 on average and the average credit utilization rate was 53.21 which shows that borrowers have moderate exposure to credit. The mean loan balance is 14,850.00 and the average interest rate was 12.48 per cent.

The complete sample displayed a complete default rate of about 18.6, indicating the moderate imbalance in classes that would be appropriate in the prediction modelling.

Table 1: Descriptive Statistics of Key Variables (N = 10,000)

Variable	Mean	Std. Dev.	Min	Max
Annual Income ($)	75,420	48,315	10,000	850,000
Debt-to-Income Ratio	17.84	8.92	0.0	45.6
Credit Utilization (%)	53.21	24.17	0.0	150.3
Total Credit Lines	22.64	11.58	2	95
Delinquencies (2 yrs)	0.32	0.84	0	13
Loan Amount ($)	14,850	8,950	1,000	40,000
Interest Rate (%)	12.48	4.35	5.32	30.99

3.2 Correlation Analysis

In order to test any relationship among the predictor variables and to determine possible multicollinearity, correlation analysis was carried out. Based on Table 2, it is found that the largest correlation was between debt to income ratio and credit utilization (r = 0.42). All of the correlation coefficients is less than 0.70, which is an indication that the issues of a strong multicollinearity are not severe.

Table 2: Correlation Matrix

Variable	DTI	Credit Util.	Delinq.	Loan Amt	Int. Rate
Debt-to-Income	1.00	0.42	0.18	0.09	0.21
Credit Utilization	0.42	1.00	0.27	0.05	0.24
Delinquencies	0.18	0.27	1.00	0.02	0.19
Loan Amount	0.09	0.05	0.02	1.00	0.14
Interest Rate	0.21	0.24	0.19	0.14	1.00

Figure 1. Correlation Heatmap of Key Credit Risk Variables

The correlation heatmap presented in Figure 1 shows the strength and direction of the relationship between important variables of credit risk. The most significant correlation is seen between debt to income ratio to credit utilization (r = 0.42) and the rest of the correlations are moderate, which means that there is no severe problem of multicollinearity.

3.3 Model Performance Comparison

Different predictive performance is assessed through the implementation of four supervised machine learning models. Table 3 shows that ensemble based models were better than traditional logistic regression and decision tree models. Gradient Boosting had the greatest accuracy (0.88) and AUC-ROC (0.92) closely followed by the Random Forest model whose AUC-ROC was 0.90. There is moderate predication performance of logistic regression (AUC-ROC = 0.84) with the decision tree having a relatively lower overall performance.

Table 3: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC

Logistic Regression	0.81	0.69	0.61	0.65	0.84
Decision Tree	0.78	0.63	0.66	0.64	0.76
Random Forest	0.86	0.74	0.71	0.72	0.90
Gradient Boosting	0.88	0.77	0.74	0.75	0.92

3.4 Feature Importance Analysis

The Gradient Boosting model is used to analyze the importance of features. Table 4 indicates that interest rate, credit utilization, and debt-to-income ratio were the most predictive indicators of default of borrowers.

Table 4: Top Predictors of Default (Gradient Boosting Model)

Rank	Variable	Importance Score
1	Interest Rate	0.19
2	Credit Utilization	0.16
3	Debt-to-Income Ratio	0.14
4	Annual Income	0.11
5	Delinquencies (2 yrs)	0.10
6	Total Credit Lines	0.09
7	Loan Amount	0.07
8	Credit Inquiries	0.06

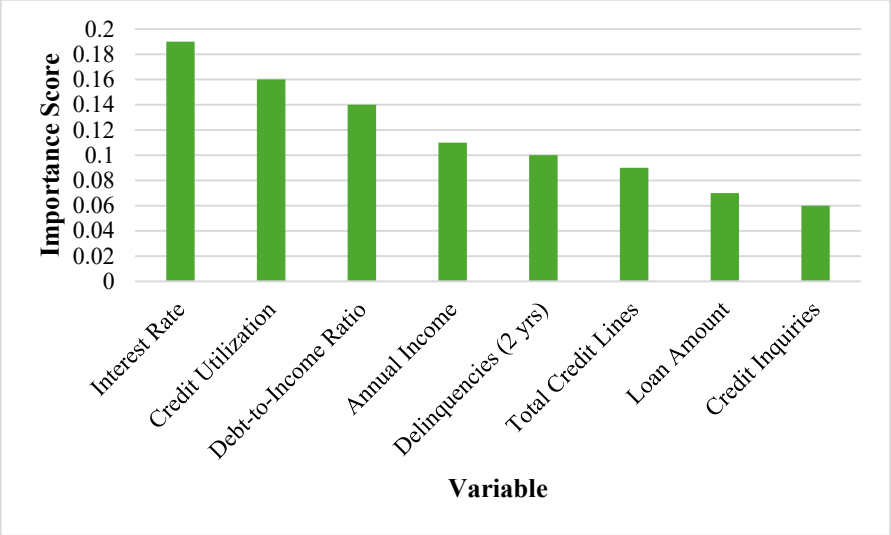

Figure 2. Key Predictors of Borrower Default Risk

Figure 2 identifies the importance of each variable to the prediction model graphically, through the feature importance plot. The most influential variables on the default risk are financial stress indicators and credit behavior variables.

4. Discussion

The increasing complexity of the retail lending environment has contributed to the necessity to adopt new tools of analysis that can be used to evaluate non-linear borrower behaviors and complex financial risk profiles. The machine learning algorithms have therefore emerged to be a favorite choice in the credit risk modeling since the algorithm is more practical in addressing high-dimensional data sets and non-linear relationships among variables. The ensemble-based research confirms that ensemble-based techniques are universal relative to the traditional classification strategies in the context of the credit score prediction problems (Moscato et al., 2021). The attained results of comparative performance also substantiate this point of view as Gradient Boosting and Random Forest models reach a higher discriminatory force than logistic regression and single decision tree models do.

The enhancement in the capabilities of the ensemble algorithms can be explained by the fact that it is an iterative nature of learning and it can minimize the errors in its prediction by aggregation. The methods applied in boosting, particularly, improve weak learners in a gradual manner and the model can identify small trends in borrower credit profiles. Generalized research of machine learning-based models of credit risk is concerned usually with the fact that ensemble and hybrid approaches are stronger and more likely to work more effectively in matters of generalization in the financial datasets (Shi et al., 2022). The given conclusion is corroborated by the existing empirical similarity, according to which ensemble frameworks provide greater forecasting permanence in retail banking implementations.

The analysis may give a very important revelation on the relative contributions made by the financial stress indicators and the credit behavior variables. Interest rate, credit utilization and debt to income ratio were determined to be the strongest predictors of the borrower default. All these variables embody the ability to repay, leverage and credit exposure- basic

spheres of the risk assessment models embraced. These predictors being constant regardless of the machine learning model means that the algorithms evolve, however, the determinants of credit risks depend to a very large degree on the financial behavior of the person taking a loan. The importance of the financial leverage and repayment measures in credit scoring models is also indicated in other previous benchmarking research works.

Even though sophisticated algorithms are useful in terms of performance, transparency and interpretability must be taken into consideration by the application of machine learning in banking in practice. The standards of regulation are pushing towards explainability and audibility of credit decisions. Explainable artificial intelligence tools were found to be widely usable in order to make complex models more interpretable by measuring feature contributions, and improve traceability in automated decision systems (Gramegna and Giudici, 2021). More so, systematic ways of characterizing AI explainability in credit risk management demonstrate that explainability may be compatible with a big predictive accuracy that augments the institutional trust and compliance (Misheva et al., 2021). Although the given study is primarily focused on predictive comparison, the importance of features analysis can contribute to the interpretability aspect of the research because it demonstrates that there are significant risk drivers in the model.

The other methodological factor is in the aspect of the class imbalance, which is characteristic of the retail credit data, whose default events are very minimal. Without appropriate alterations, the skewed data will result in skewed models providing excess precision but limited default-detecting capability. According to recent work, there are new extensions of boosting specific to class-imbalance credit scoring, where there are modifications to the loss terms to better predict majority classes (Mushava and Murray, 2022). Performance of the model in this paper was ensured by the use of the evaluation plan and the AUC-ROC, recall, and F1-score metrics because they ensure that the performance of the model is based on the discrimination capacity rather than the overall accuracy. The recall rates of the ensemble methods are high which indicates that the methods are practical in the prediction of the high-risk borrowers in an unequal environment.

Even in the case of credit risk model based on machine learning, there is still a need of methodological rigor. Systematic reviews emphasise the importance of ensuring that there are standardised validation processes, performance and close preprocessing measures are the same to avoid false comparisons. This scientific study gives to the clarity of the procedures and the believable evaluation of the models with the help of various algorithms which are incorporated in one system and maintenance of a consistent data preparation procedure.

Overall, the comparative analysis shows that ensemble machine learning is becoming more applicable when it comes to retail credit risk management. Financial stress indicators and credit exposure remain to be the primary predictive variables of default, and boosting-based algorithms enhance the further sophistication of the classification and stability. In the meantime, the interpretability and class imbalance can also be relevant to the sustainable implementation of machine learning systems into the controlled banking environments. Integrating both good performing predictive frameworks and explainable frameworks is, therefore, a just and progressive solution to the modern day retail credit risk assessment.

5. Conclusion

The increasing application of machine learning techniques in retail banking has made a significant transformation in the credit risk assessment practices. The comparative analysis of classification models indicates that approach of ensemble-based algorithms such as Gradient Boosting and random forest are more predictive than the traditional approach of logistic regression and single decision trees. Enhanced discrimination power and improved stability of the classifications evidences the success of the boosting and bagging techniques to process complex data structure of borrowers. The analysis also indicates the relevance of the variables of financial stress and credit behavior such as the interest rate, credit utilization and credit debt to income ratio as the determining variables that indicate the default of a borrower. In spite of the complexity of algorithms developed, the core characteristics of finances remain at the heart of appropriate risk forecasting. It means that effective credit risk modelling not only depends on advanced computational instruments but it also depends on any representation of financial characteristics that is not trivial. Such aspects of methodological rigor as imbalance treatment in classes, high measures of evaluation, including AUC-ROC and F1-score, are significant in facilitating trustworthy model comparisons. The ensemble techniques in conjunction with organized validation process enhance accuracy of predictions and strength of the model. At the same time, it is critical to ensure that it can be interpreted and made compliant with the regulations to be introduced in a sustainable way into the existing banking system.

References

1. Abellán, J., & Castellano, J. G. (2017). A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Systems with Applications, 73*, 1–10. https://doi.org/10.1016/j.eswa.2016.12.020
2. AghaeiRad, A., Chen, N., & Ribeiro, B. (2017). Improve credit scoring using transfer of learned knowledge from self-organizing map. *Neural Computing and Applications, 28*(6), 1329–1342. https://doi.org/10.1007/s00521-016-2567-2
3. Ariza-Garzón, M. J., Arroyo, J., Caparrini, A., & Segovia-Vargas, M. J. (2020). Explainability of a machine learning granting scoring model in peer-to-peer lending. *IEEE Access, 8*, 64873–64890. https://doi.org/10.1109/ACCESS.2020.2984412
4. Bao, W., Lianju, N., & Yue, K. (2019). Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Systems with Applications, 128*, 301–315. https://doi.org/10.1016/j.eswa.2019.03.033
5. Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics, 57*(1), 203–216. https://doi.org/10.1007/s10614-020-10042-0

6. 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. https://doi.org/10.1016/j.asoc.2018.09.029
7. García, V., Marques, A. I., & Sánchez, J. S. (2019). Exploring the synergetic effects of sample types on the performance of ensembles for credit risk and corporate bankruptcy prediction. *Information Fusion, 47*, 88–101. https://doi.org/10.1016/j.inffus.2018.06.004
8. Gramegna, A., & Giudici, P. (2021). SHAP and LIME: An evaluation of discriminative power in credit risk. *Frontiers in Artificial Intelligence, 4*, 752558. https://doi.org/10.3389/frai.2021.752558
9. Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research, 247*(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030
10. Misheva, B. H., Osterrieder, J., Hirsa, A., Kulkarni, O., & Lin, S. F. (2021). Explainable AI in credit risk management. *arXiv Preprint arXiv:2103.00949*. https://doi.org/10.48550/arXiv.2103.00949
11. Moscato, V., Picariello, A., & Sperlí, G. (2021). A benchmark of machine learning approaches for credit score prediction. *Expert Systems with Applications, 165*, 113986. https://doi.org/10.1016/j.eswa.2020.113986
12. Mushava, J., & Murray, M. (2022). A novel XGBoost extension for credit scoring class-imbalanced data combining a generalized extreme value link and a modified focal loss function. *Expert Systems with Applications, 202*, 117233. https://doi.org/10.1016/j.eswa.2022.117233
13. Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Klamargias, A. (2019). A robust machine learning approach for credit risk analysis of large loan level datasets using deep learning and extreme gradient boosting. *IFC Bulletins Chapters, 49*.
14. Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. (2022). Machine learning-driven credit risk: A systemic review. *Neural Computing and Applications, 34*(17), 14327–14339. https://doi.org/10.1007/s00521-022-07472-2
15. Singh, U. (2023). *Lending Club Loan Dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/utkarshx27/lending-club-loan-dataset
16. Xia, Y., Liu, C., Li, Y., & Liu, N. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications, 78*, 225–241. https://doi.org/10.1016/j.eswa.2017.02.017