INNOVARE JOURNAL OF ENGINEERING & TECHNOLOGY

Vol 13, 2025

ISSN - 2347-1573

Research Article

COMPREHENSIVE PERFORMANCE ANALYSIS OF MACHINE LEARNING BASED ALGORITHMS FOR CREDIT DEFAULT PREDICTION USING LENDING CLUB DATASET

ANOOP SUNARTY*

Department of Computer Sciences and Applications, Mandsaur University, Mandsaur, Madhya Pradesh, India.
Email: anoop.sunarty@meu.edu.in

Received: 19 September 2025, Revised and Accepted: 12 November 2025

ABSTRACT

Objective: To develop and evaluate a machine learning–based framework for accurate prediction of credit default using the Lending Club dataset.

Methods: The study applies extensive data preprocessing including missing value treatment, outlier removal using Z-score, one-hot encoding of categorical variables, Min–Max normalization, and class balancing using SMOTE. Several models were evaluated, with Extreme Gradient Boosting (XGBoost) used as the primary classifier. Data were split into 80% training and 20% testing. Performance was measured using accuracy, precision, recall, and AUC.

Results: XGBoost achieved the best performance with accuracy of 91.56%, precision of 94.78%, recall of 93.30%, and AUC of 97.07, outperforming decision tree, logistic regression, artificial neural networks, gradient boosting, and convolutional neural networks.

Conclusion: The proposed XGBoost-based framework demonstrates strong potential for real-world credit risk assessment by effectively handling large and imbalanced datasets and improving early identification of default risk.

Keywords: Credit default prediction, Machine learning, XGBoost, Lending club dataset, Credit risk assessment, Predictive analytics.

© 2026 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/) DOI: http://dx.doi.org/10.22159/ijet.2025v13.57906. Journal homepage: https://innovareacademics.in/journals/index.php/ijet

INTRODUCTION

The rapid digital transformation in the past several years has been quite positive to the financial services sector and compelled commercial banks and lending institutions to embrace new financial technologies (FinTech) that are largely effective, larger, and more satisfactory to consumers. The application of digital tools caused the number of clients to grow tremendously, the models of the services became more diverse, and the operating costs of the financial sector reduced [1,2]. Conversely, escalating reliance on technology has brought in new vulnerabilities, such as the security of the data, the information gap, and the accelerated flow of financial risks through the digital medium. Since Fintech is altering the credit landscapes, it is crucial to study how digital technology affects the decision-making process and the spread of risk.

Financial institutions, credit card firms, and peer-to-peer lending platforms all face the same major risk: The possibility that a borrower may not be able to repay the loan [3,4]. Every type of loan, from mortgages to personal loans to credit cards, carries the possibility of credit risk, which threatens the viability, profitability, and very existence of financial organizations [5]. Banks grant millions of loans each year, but not every applicant has the ability or the desire to pay the debt back, which eventually becomes a source of great financial loss if the borrower defaults on the loan [6]. The reliance on credit cards and digital lending platforms continues to rise, driven by the factor of ease of use and availability, which consequently places more emphasis on the necessity for solid credit risk assessment systems.

The classical risk evaluation methods are based on rating systems that are meant to limit credit exposure to a specific level considered acceptable and at the same time to obtain the highest risk-adjusted returns [7,8]. These techniques often consider a number of risk categories, including institutional risk, default risk, concentration risk, nation risk, and downgrade risk.

In the financial sector, machine learning (ML) has become a strong and promising tool for credit risk modeling in present lending systems [9,10]. ML methods are very good at discovering intricate, non-linear correlations, working with multi-dimensional data, and being flexible with different kinds of borrowers [11,12]. Traditional statistical models usually do not possess these capabilities. The correct prediction of credit default is vital for different areas, such as risk-based pricing, loan approval, early-warning systems, and overall portfolio management [13]. At the same time, the prediction process is still quite complicated due to the problems of class imbalance, differences in customer behavior, and the occurrence of noisy or incomplete data.

Significance and contribution

This study is motivated by the harsh realities of credit risk assessment that have become very complex in the modern-day lending process. The existence of very large datasets, the constant change in the behavior of borrowers, and the extremely high-class imbalance make it impossible for the traditional methods to be effective. As a result, it is not be long before the financial institutions give totally their trust to data-driven decision-making and find it more urgent to have advanced ML methods that can distinguish the risky borrowers precisely and at the same time, keep the costs of misclassifications at the least level possible. Thus, the prediction framework, which is robust, scalable, and interpretable, has to be developed to facilitate lending, prevent losing money through bad debts, and allow credit systems that are both secure and inclusive of the needy. The key contributions are:

- A detailed pre-processing pipeline consisting of missing value treatment, outlier elimination, feature encoding, normalization, and data balancing through SMOTE was built.
- An optimized Extreme Gradient Boosting (XGBoost) classifier that catered to large and imbalanced credit datasets was designed and developed.
- Accuracy (ACC), precision (PRE), recall (REC), and area under the curve (AUC) were utilized as measures in the comprehensive

evaluation, which demonstrated that the new approach outperformed conventional ML and deep learning (DL) models.
- Presented a comparative analysis demonstrating the benefits of ensemble learning techniques for real-world credit risk prediction.

Justification and novelty

The increasing need for credit default prediction models with ACC, scalability, and reliability that can manage the financial institutions' enormous and unbalanced datasets, such as the Lending Club, justifies this study. The traditional statistical models are not able to depict the borrower behavior's complexity and non-linearity; therefore, the risk assessment is not optimal. The originality of this study lies in the association of a complete data pre-processing pipeline removing outliers, balancing with SMOTE, and normalizing features with a tuned XGBoost classifier, which leads to better predictive performance than the present ML and DL methods.

Structure of the paper

The structure of the study is as follows. A review of the literature on ML methods for credit risk assessment is provided in Section II. The dataset, feature selection, pre-processing techniques, and modeling methodology are covered in Section III. Section IV discusses experimental data and comparative examination of models. Section V ends the study and proposes the next research directions.

LITERATURE REVIEW

This section discusses some review articles on credit default prediction using ML approaches.

Yu (2025) proposes a two-stream network-based intelligent credit assessment technique called XGBoost. After pre-processing both traditional and non-traditional data, the AutoEncoder algorithm is used to minimize the dimension of non-traditional data. For effective credit evaluation and risk prediction, the fused feature data are input into the XGBoost model. This method can achieve an average AUC value of 0.820 and an ACC of over 89%. In addition, the ACC of risk prediction exceeds 90% [14].

Chauhan (2024) aims to explore the application of sophisticated ML classification methods to forecast non-performance of bank loans. They produced ACCs of 87, 89, and 88, respectively, with the different classifiers, such as LR, RF, and DT. RF was the best performing classifier with the very high scores of ACC and strength in its prediction of loan defaults [15].

T and Ahadit (2024) strategies involve an emphasis on minimizing false negatives, false predictions of low-risk loans with high monetary implications. They fine-tune the decision threshold with the help of the AUC analysis and the ROC curve. Empirical analysis shows the effectiveness of the model selected, which is supported by ROC and AUC, in lowering false negatives and securing the loan approval process against possible defaults [16].

Smadi *et al.* (2024) a dataset of actual credit card transactions and identifying various customer behaviors. To enhance the PRE of identifying scams, synthetic datasets are created based on consumer behavior. The LR model outperformed the other in both evaluations. In the first experiment, it achieved an impressive 96.4% ACC with remarkable time efficiency; in the second experiment, it maintained excellent time efficiency while achieving an impressive 94.5% ACC [17].

Nancy Deborah *et al.* (2023) loan requests and control the default risk. In order to forecast a customer's loan approval, this study uses algorithms, such as KNN and DT. Moreover, provide the SVC, a cutting-edge approach that is used to deploy the model and exhibits exceptional ACC. The use of ML techniques to forecast loan status. Among the algorithms examined was SVM, which achieved an ACC rate of 83% [18].

Chen (2022) examines the efficacy of different ML approaches in credit risk prediction and thoroughly assesses the performance of models, such as logistic regression, DT, RF, and SVM. The results show that the random forest model is the best at handling complex data patterns and high-dimensional features, with an ACC of 88.3%, PRE of 87.5%, and REC of 86.8%, respectively. The support vector machine has an ACC of 84.5; hence, this is also very good in processing information with high dimensions, but processing a large volume of data is very slow due to a long period of training and a bottleneck in computing [19].

Karim *et al.* (2022) attempt to use the SMOTE, which is an oversampling method to address these issues. One real-world imbalanced dataset that has been investigated for model training is Lending Club. Out of the 151 features in the dataset, 113 are numerical, and the remaining features are categorical. There were 2260701 instances in all. With PRE, REC, and F1 of 0.81, 0.97, and 0.89, respectively, RF outperforms the others and reaches 87% ACC. Additionally, the RF's ROC-AUC score was 0.93 [20].

METHODOLOGY

The Lending Club dataset in Fig. 1 is the basis for the methodology employed in this study, which aims to create a solid prediction framework for loan defaults. The first stage is to perform a thorough pre-processing of the data, which involves handling missing values, eliminating superfluous features, and identifying outliers. Transforming categorical variables into numerical values is done via one-hot encoding, and numerical features are normalized by keeping all values within the same range using Min-Max scaling. Since the dataset is highly imbalanced, SMOTE is employed to create fictional cases for the minority class in an effort to enhance the model's equity. The data are divided into two parts: 20% for testing and 80% for training. As a result, XGBoost is a ML model that is trained and evaluated using metrics, including ACC, PRE, REC, and AUC.

The following steps of the proposed methodology are briefly discussing in below:

Lending club dataset

The Lending Club dataset is a gold standard for credit risk research. It has 2,925,297 records and 141 attributes collected from April 2007 to April 2020. Lending Club is one of the largest P2P lending platforms in the US. Excluded loans made before 2014 from the dataset to guarantee completeness of outcome data and loan maturity. Among the many financial and borrower-related details covered are their demographics, credit histories, interest rates, and loan amounts. Essential components include seven loan statuses, such as "Fully Paid," "Charged Off," "Default," and overdue phases.

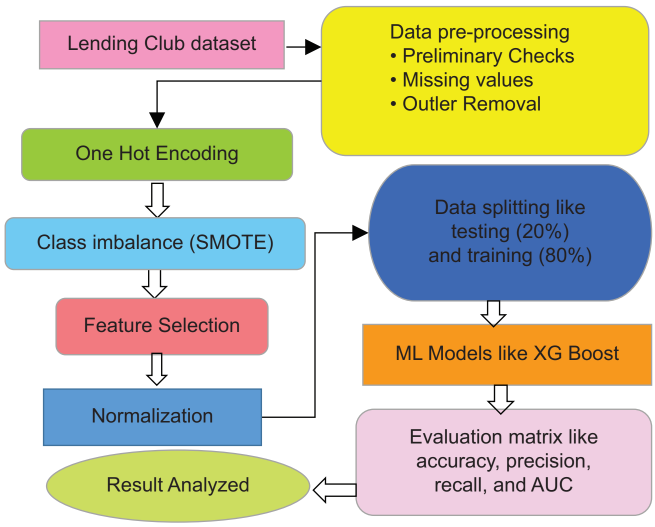

Fig. 1: Flowchart for credit default prediction using machine learning models on the lending club dataset

The distribution of the dataset's goal variable, loan default status, is shown in Fig. 2. The non-defaults and the default frequencies are presented in the histogram. The x-axis (0: Non-defaults, 1: Defaults) indicates the target classes, and the y-axis shows the number of instances in each of the classes. Having a 80/20 split to non-defaults and 20/80 split to defaults, the distribution is very uneven, which illustrates the heavily class -imbalanced dataset.

The heatmap of correlation, as illustrated in Fig. 3, shows the linear relationship of the main financial variables in the Lending Club database. The various colors reflect the power and type of correlation, where the lightest colors depict the very strong positive correlations and the darkest ones depict weak correlations or negative correlations. It is curious to mention that the loan amount received the best correlation with the installment (large loans mean large monthly payment), with total account count having a moderate correlation with the revolving balance. Most of the other variables are weakly correlated, indicating that the multicollinearity is not high and the non-linear trends of credit default prediction can be identified by ML models with ease.

Data pre-processing

Data pre-treatment solved fundamental difficulties in lending club datasets, including missing values, duplication, and class imbalance [21]. These steps follow standard practices from existing literature to enhance model performance and generalizability:

- Preliminary checks: To get a feel of the data, the first couple of checks were performed using the. head ().,info (), and. describe () commands to understand the structure of the data, the type of data, and its summary.
- Missing values: As a first step in cleaning the data, filling in any blanks is essential. The isnull (). The sum() technique was employed to identify any missing data [22]. The dropna () function was used to eliminate records that had a null value to ensure that data integrity was preserved.
- Outlier removal: Real-world datasets generally have a limited number of outliers due to the presence of artificial or natural causes of outliers, such as low probability occurrences or data input errors in the data collection system, which impair the ML model's performance (Equation (1)). To solve this issue, researchers have created a variety of outlier identification methods to locate the outliers in the datasets [23].

$$Z = \frac{(V - \mu)}{\sigma} \tag{1}$$

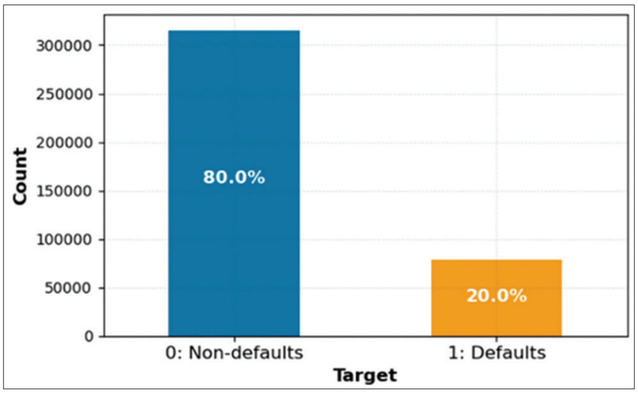

Fig. 2: Distribution of targets

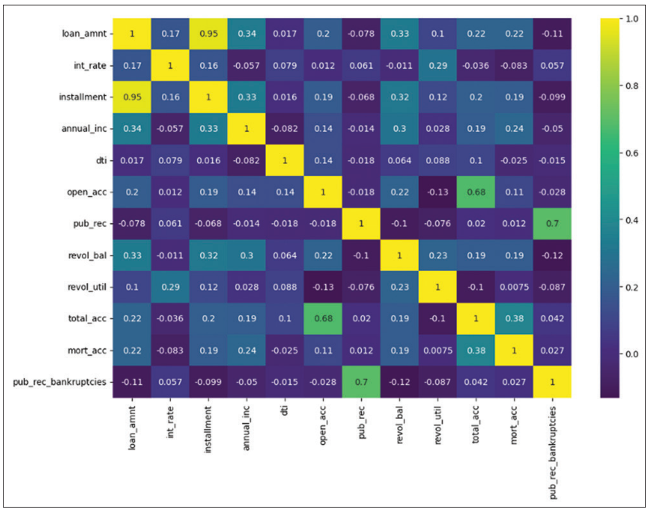

Fig. 3: Correlation matrix of lending club variables

One-hot encoding

One-hot encoding generates a brand-new binary attribute, set to zero, for each distinct value represented by the categorical feature. After that, the categorical feature is omitted and the binary feature linked to the value of the categorical feature is turned to one [24]. Concerning one-hot encoding, the categorical variables homeownership, verification status, and purpose are changed into various binary (0/1) columns that signify every category.

Data balancing with SMOTE

SMOTE tackles the issue of unbalanced categories in the original data set by examining and simulating microscales of category samples, followed by manually adding the simulated new samples to the dataset [7]. However, one of the SMOTE method's drawbacks is the generalization of the minority class space. This suggests that if the selected minority samples are surrounded by majority samples, there may be "noise"; that is, it may be challenging to classify the freshly synthesized samples because they overlap with the nearby majority samples.

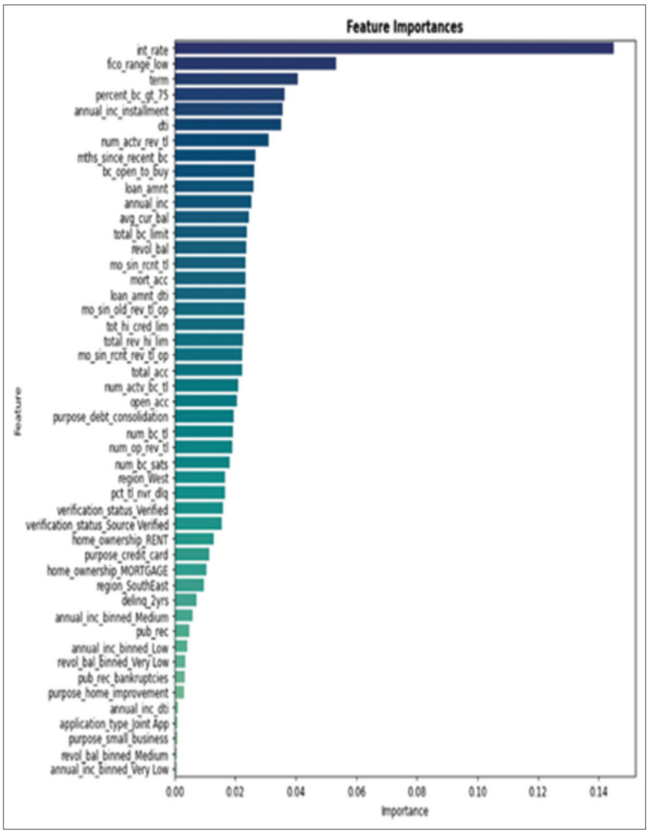

Fig. 4: Feature importance

Feature selection

Feature selection is crucial for improving model ACC, decreasing training time, and increasing interpretability. Given the abundance of features in the Lending Club dataset [25]. Fig. 4 displays the feature relevance ranking for a credit default prediction model. It emphasizes the influence of every characteristic on the predictive ability of the model. The most significant ones are the loan period (term), interest rate (int_rate), and the range of FICO score (fico_range_low), which implies that they have a significant influence on default risk. There are other significant features, such as credit utilization ratios, annual installments, and percent bc gigt 75. The lower-ranked variables that have the least impact are region, home ownership, and purpose categories. It is all said and done that the risks of default of a loan are highly pre-determined by the credit quality, the structure of the loan, and the ability to pay.

Data normalization

Features of a dataset that have a different range can affect some models. To address this issue, the features were scaled using the following normalizing methods:

- Min-max scaler: Data are scaled to the interval [0, 1]. It can be affected by outliers, even though it is not as sensitive to them as the regular scaler. The calculation is as follows (Equation (2))

$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{2}$$

Proposed XG boost classifier

XGBoost, is a scalable and effective boosting system. On multiple ML tasks, it has demonstrated state-of-the-art performance [26,27]. In contrast to the conventional gradient boosting algorithm, the XGBoost algorithm approaches the addition of weak learners in parallel using a multithreaded pattern, which results in proper hardware resource utilization and increased speed and efficiency.

The module is the regularization term, which improves generalization to fresh data and lowers model complexity to prevent overfitting [28]. Equation (3), which integrates regularization and the training error term, shows the two elements of the XGBoost objective function.

$$Obj(\phi) = Min\sum_{i=1}^{n} l(y_i, \widehat{y_i}) + \sum_{k=1}^{K}\Omega(f_k) \tag{3}$$

Where, $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$

Where T is the number of leaves in the tree, l is a differentiable loss function that measures the difference between the observation (y_i) and prediction ($\widehat{y_i}$), $\Omega(f_k)$ is a regularization term that controls the model's complexity and overfitting, γ is a regularization parameter, and the second part of the regularization term is $\lambda\sum_{j=1}^{T} w_j^2$ the regularization coefficient, where j is the weight of its corresponding leaf.

Evaluation metrics

A confusion matrix that captures the results of expected versus actual classifications is used to assess how well the classification models applied to the lending club dataset performed [29]. There are four main parts to the matrix. The confusion matrix is listed as follows: The numbers of samples for the positively and negatively classed classes are TP and TN, respectively, compared to the negatively and positively classified classes that were wrongly counted [30]. The following performance metrics are:

ACC

The ratio of "positive" defaults and "negative" non-defaults to all forecasts is known as ACC. The formula in Equation (4) defines it:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

PRE

PRE, which calculates the percentage of real defaults out of all default forecasts [31] (Equation (5)).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

REC

REC is the percentage of actual defaults that are accurately identified; it is also known as sensitivity or the TPR. This is how it is calculated in Equation (6):

$$\text{Recall} = \frac{\text{TP}}{TP + FN} \tag{6}$$

ROC-AUC

Particularly important in an unbalanced dataset is the AUC, which gauges the model's capacity to separate between defaulters and non-defaulters at all classification thresholds. The TPR is plotted against the false positive rate (FRP) on the ROC curve.

RESULT ANALYSIS AND DISCUSSION

This section shows the results of an experiment using the Lending Club dataset to test ML models for credit default prediction. It used an Intel dual-core i6 processor, operating Windows 10 with 3.3 GHz and 1 TB of RAM. The ACC, PRE, REC, and AUC measurements are used to access the performance. The data presented in Table 1 show the credit default prediction performance of the XG Boost classifier model on the Lending Club data.

The XGBoost model's performance evaluation for Credit Default prediction using Lending Club data is presented in Table 1 above. To find out how well the model worked, the main evaluation metrics were ACC, PRE, REC, and AUC. The results of the XG boost model have an ACC of 91.56, PRE of 94.78, REC of 93.30, and AUC of 97.07, showing that not only does the XG boost model have the ability to predict accurately, but it also correctly classifies default and non-default borrowers.

Fig. 5 shows the ROC curve of the XGBoost model that demonstrates the classification performance of the model. TPR against the FPR of the different threshold values is plotted against the curve. In addition, the model has an AUC of 0.97, which is excellent since it indicates that it can distinguish between the positive and negative classes with ease.

The confusion matrix for the XGBoost model is depicted in Fig. 6, which reveals its efficiency as a classification model for predicting loan defaults. The model was right in classifying 6,676 non-defaults and 419 defaults. Conversely, there were 1,541 FN (defaults mislabeled as non-defaults) and 364 FP (non-defaults mislabeled as defaults).

Table 1: XGboost on the lending club dataset for credit default prediction

Performance measures	XGBoost
Accuracy	91.56
Precision	94.78
Recall	93.30
AUC	97.07

AUC: Area under the curve, XGBoost: Extreme Gradient Boosting

Table 2: ML and DL models comparison on the lending club dataset for credit default prediction

Performance measures	XGBoost	DT [32]	ANN [33]	LR [34]	Gradient boosting [35]	CNN [36]
Accuracy	91.56	86.10	72.97	84.3	65.31	67.27
Precision	94.78	63.78	59.73	84.6	64.76	68.24
Recall	93.30	86.91	60.55	84.3	67.17	65.66
AUC	97.07	92.50	5031	84.1	65.94	67.62

ML: Machine learning, DL: Deep learning, CNN: Convolutional neural networks, XGBoost: Extreme Gradient Boosting, ANN: Artificial neural networks

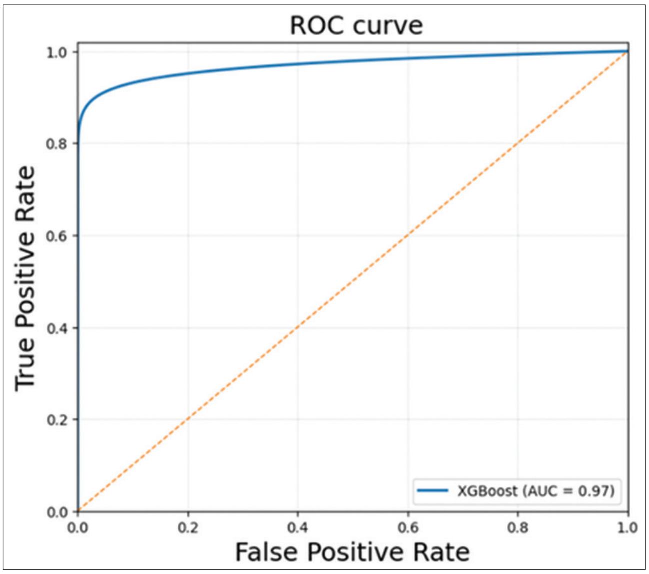

Fig. 5: ROC graph of XGBoost model

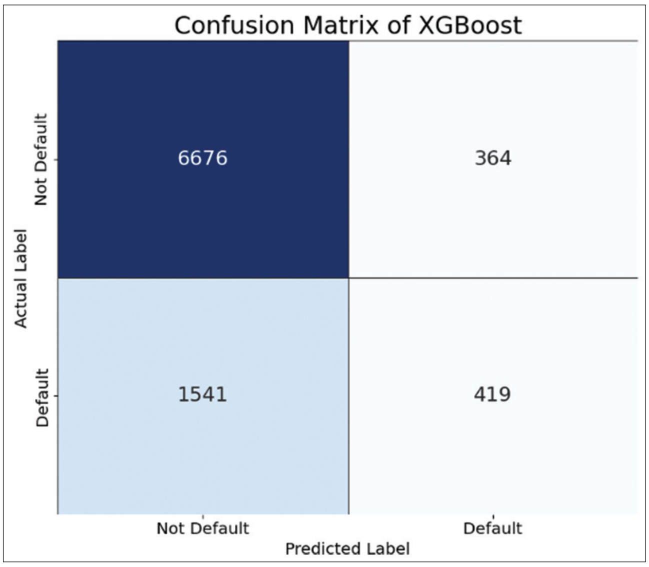

Fig. 6: Confusion matrix of XGBoost model

Comparative analysis and discussion

Using the Lending Club dataset, this section gives the results of the comparative study for Credit Default Prediction. Based on performance matrices, including ACC, PRE, REC, and AUC, Table 2 compares the performance of ML models including XGBoost.

A comparison of several ML and DL models performed on the Lending Club data on credit default prediction is provided in Table 2. The above results are clear in the sense that all other models perform poorly compared to the XGBoost in terms of all measures of ACC, PRE, REC, and AUC, that is, XGBoost is highly efficient in capturing the complex patterns in borrower behavior. DT is also relatively good, especially in REC, but it compares poorly with XGBoost in general reliability. Conventional models, such as the LR demonstrate a stable but mediocre performance, whereas artificial neural networks (ANN), Gradient Boosting, and convolutional neural networks (CNN) exhibit lower predictive power, with Gradient Boosting predicting the weakest results of all. On the whole, this comparison indicates that ensemble-based learning algorithms, in particular, XGBoost, are much more efficient to predict credit defaults on large and unbalanced data, such as Lending Club.

Limitations and future work

Even though the proposed XGBoost-based framework demonstrated excellent ACC, there are nevertheless some limitations of this research that should be pointed out. In order to guarantee the highest level of maturity, the dataset was limited to loans; thus, the model might not reflect the most recent socio-economic developments, altering consumer habits or new lending regulations. On top of that, although SMOTE is beneficial in ruling out class imbalance, it could also produce artificial data that hinders the model's generalization to real-life situations. In addition, traditional feature engineering was the method used in the study instead of DL which is capable of unearthing even more complex and deeper patterns from financial data. In the future, researchers might experiment with hybrid DL models, cost-sensitive learning that treats the unbalanced data more appropriately without synthetic sampling, and explainable AI methods with SHAP that help make the decision-making more transparent. Moreover, the dataset can be augmented by adding recent years, macroeconomic indicators can be incorporated, and the model can be utilized in a real-time loan approval setting, thereby increasing its reliability and practical use even more.

CONCLUSION

The forecast of credit defaults is an important aspect of financial risk management, particularly for large-scale online lending companies, such as Lending Club. The rise in loan amounts and the variety in borrower characteristics make it necessary for banks and financial institutions to have trustworthy, data-driven systems that can precisely classify high-risk borrowers and those with good credit. A consistent machine-learning-based framework was created in this research through the use of an intensive data-preprocessing phase, a feature engineering phase, a data balancing phase that made use of SMOTE, and a normalization phase, all coated with the application of the XGBoost classifier. With 91.56% ACC, 94.78% PRE, 93.30% REC, and 97.07% AUC, the results show that XGBoost has performed exceptionally well, providing the classifier with a grip over complex non-linear connections in the given data with high dimensions and imbalances. Subsequent analyses revealed that XGBoost was the best model for credit default prediction in this field, outperforming both tree-based and neural-based algorithms, such as DT, LR, ANN, Gradient Boosting, and CNN. The study as a whole demonstrates how ML techniques can be used to risk assessment, resulting in better lending decisions and financial stability.

REFERENCES

1. Xu J. Research on the impact of credit risk in commercial banks under the background of fintech. Highlights Business Econ Manag 2025;55:198-214.
2. Chandrashekar P. Data-driven loan default prediction: enhancing business process workflows with machine learning. Int J Emerg Res

Eng Technol 2025;6:198-214.
3. Lin J. Research on loan default prediction based on logistic regression, randomforest, xgboost and adaboost. SHS Web Conf 2024;181:02008.
4. Majumder RQ. A review of anomaly identification in finance frauds using machine learning systems. Int J Adv Res Sci Commun Technol 2025;5:101-10.
5. Wawge SJ. A survey on the identification of credit card fraud using machine learning with precision, performance, and challenges. Int J Innov Sci Res Technol 2025;10:3345-52.
6. Verma V. Deep learning-based fraud detection in financial transactions: A case study using real-time data streams. ESP J Eng Technol Adv 2023;3:149-57.
7. Cui X, Huang Y, Lai S, Keong CI, Wu D. Machine learning on credit risk for lending club business. BCP Bus Manag 2023;23:894-902.
8. Thakkar KB, Kapadia HP. The Roadmap to Digital Transformation in Banking: Advancing Credit Card Fraud Detection with Hybrid Deep Learning Model. In: 2025 2nd International Conference on Trends in Engineering Systems and Technologies (ICTEST), IEEE; 2025. p. 1-6.
9. Ande BR. Federated learning and explainable AI for decentralized fraud detection in financial systems. J Inf Syst Eng Manag 2025;10:48-56.
10. Bilipelli AR. Forecasting the evolution of cyber attacks in fintech using transformer-based time series models. Int J Res Anal Rev 2023;10:383-9.
11. Zhang X, Zhang T, Hou L, Liu X, Guo Z, Tian Y, *et al*. Data-driven loan default prediction: A machine learning approach for enhancing business process management. Systems 2025;13:581.
12. Mantha G. Transforming the insurance industry with salesforce: Enhancing customer engagement and operational efficiency. North Am J Eng Res 2024;5:???.
13. Macha Y, Pulichikkunnu SK. An explainable AI system for fraud identification in insurance claims via machine-learning methods. Int J Adv Res Sci Commun Technol 2023;3:1391-400.
14. Yu G. Intelligent Credit Assessment and Risk Prediction Model based on Machine Learning. In: 2025 3rd International Conference on Data Science and Information System (ICDSIS), IEEE; 2025. p. 1-7.
15. Chauhan S. Machine Learning Models for Loan Default Forecasting: Accuracy Comparison. In: 2024 Second International Conference Computational and Characterization Techniques in Engineering and Sciences (IC3TES), 2024. p. 1-5.
16. Saravana Kumar T, Ahadit AB. Machine Learning Applications in Banking for Credit Rating and Risk Evaluation. In: 2024 Third International Conference on Artificial Intelligence, Computational Electronics and Communication System (AICECS), 2024. p. 1-6.
17. Al Smadi B, Glisson WB, Tahat M, Alamleh H, AlQahtani AA. Credit Card Transactions Fraud Detection for Multiple Consumer Behaviors. In: 2024 International Conference on Computing, Networking and Communications (ICNC); 2024. p. 26-32.
18. Deborah RN, Rajiv SA, Vinora A, Devi CM, Arif SM. An Efficient Loan Approval Status Prediction Using Machine Learning. In: Proceedings of 3rd International Conference on Advanced Computing Technologies and Applications, ICACTA; 2023.
19. Chen H. Prediction and analysis of financial default loan behavior based on machine learning model. Comput Intell Neurosci 2022;2022:7907210.
20. Karim M, Samad MF, Muntasir F. Improving Performance Factors of an Imbalanced Credit Risk Dataset Using SMOTE. In: 2022 4th International Conference on Electrical, Computer and Telecommunication Engineering (ICECTE), IEEE; 2022. p. 1-4.
21. Gupta A, Gulati P, Chakrabarty SP. Classification-based credit risk analysis: The case of lending club. Lect Notes Netw Syst 2024;964:77-86.
22. Tiwari SK. Quality assurance strategies in developing high-performance financial technology solutions. Int J Data Sci Mach Learn 2025;5:323-35.
23. Qian H, Zhang S, Wang B, Peng L, Gao S, Song Y. A comparative study on machine learning models combining with outlier detection and balanced sampling methods for credit scoring; arXiv, 2021.
24. Malali N. Predictive AI for identifying lapse risk in life insurance policies: Using machine learning to foresee and mitigate policyholder attrition. Int J Adv Res Sci Commun Technol 2025;5:1-11.
25. Shah TV. Leadership in digital transformation: Enhancing customer value through AI-driven innovation in financial services marketing. Int J Sci Res Arch 2025;15:618-27.
26. Chen Y, Dong Y, Liu W. Prediction of credit default based on the XGBoost model. Appl Comput Eng 2024;96:85-92.
27. Kurakula SR. The role of AI in transforming enterprise systems architecture for financial services modernization. J Comput Sci Technol Stud 2025;7:181-6.
28. Deva MR, Jain N. Utilizing Azure Automated Machine Learning and XGBoost for Predicting Cloud Resource Utilization in Enterprise Environments. In: 2025 International Conference on Networks and Cryptology (NETCRYPT), IEEE; 2025. p. 535-40.
29. Kabade S. Smart pension payroll management enhances accuracy and efficiency through AI and cloud integration. TIJER Int Res J 2024;11:915-21.
30. Kali H. Optimizing credit card fraud transactions identification and classification in banking industry using machine learning algorithms. Int J Recent Technol Sci Manag 2024;9:85-96.
31. Omogbhemhe MI, Momodu IB. Model for predicting bank loan default using XGBoost. Int J Comput Appl 2021;183:1-4.
32. Yang R. Machine learning-based loan default prediction in peer-to-peer lending. Highlights Sci Eng Technol 2024;94:310-8.
33. Robinson N, Sindhwani N. Loan Default Prediction Using Machine Learning. 2024 11th ICRITO 2025. 12th International Conference on Reliability, Infocom Technologies and Optimization. (Future Trends and their Directions). Vol. 25. ICRITO; 2024. p. 15-8.
34. Chang V, Sivakulasingam S, Wang H, Wong ST, Ganatra MA, Luo J. Credit risk prediction using machine learning and deep learning: A study on credit card customers. Risks 2024;12:174.
35. Suhadolnik N, Ueyama J, Da Silva S. Machine learning for enhanced credit risk assessment: An empirical approach. J Risk Financ Manag 2023;16:496.
36. Ko PC, Lin PC, Do HT, Huang YF. P2P lending default prediction based on AI and statistical models. Entropy 2022;24:1-23.