Journal of Global Research in Electronics and Communication
Volume 2, No. 3, March 2026
Available Online at: www.jgrec.info
REVIEWPAPER

Leveraging Ensemble Learning Methods to Improve Credit Scoring Model Accuracy and Robustness

Mr. Kapil Ahir
Assistant Professor
Department of Computer Sciences and Applications
Mandsaur University
Mandsaur
kapil.ahir@meu.edu.in

Abstract*—Credit scoring is the term used to describe a statistical analysis that can be applied in finance institutions and banks to determine the creditworthiness of an individual. The bestowers normally manipulate it to decide whether to expand or withdraw credit. The score is also a major factor in evaluating an individual's creditworthiness and determining whether he/she approved for a loan. The proposed study uses an ensemble learning model with the XGBoost package to optimize credit scoring using the Credit Scoring Kaggle dataset, which consists of 31,219 customer records from 2023. The data preprocessing pipeline included handling missing data, outliers, and noise; label encoding; z-score normalization; and data balancing using the Synthetic Minority Oversampling Technique (SMOTE). Stratified sampling (80:20) was also employed to maintain the proportions of the different classes. The suggested XGBoost model was assessed using standard performance measures, including accuracy, precision, recall, F1-score, and ROC-AUC. The experimental findings showed increased prediction performance, with an F1-score of 95.9%, recall of 99.5%, precision of 92.6%, and an AUC of 0.99, which denotes practically perfect classification. Compared with more conventional models such as Decision Trees, Random Forests, and Gradient Boosting Machines, the XGBoost model proved more robust, efficient, and capable of strong generalization. The results underscore that the model is interpretable in addition to improving credit risk prediction accuracy and dependability, as demonstrated by the analysis of feature importance, making it an effective decision-support tool in current financial risk management.

Keywords*—Credit Scoring, Preparing Credit Score Cards, Machine Learning Algorithms, Peer-to-Peer Lending, Default Risk Modeling.

I. INTRODUCTION

The ever-evolving nature of credit risk is a key concern for commercial banks in light of the recent explosion in the size and sophistication of financial markets, the paradigm shift in public consumption, and the relentless pursuit of better credit products [1]. Therefore, one of the main concerns for financial risk management is successfully lowering the default risk. Accurate PD estimations are now the cornerstone of credit risk management as they are one of the main factors affecting credit risk. One instrument that has been shown to be useful in helping financial institutions understand the global financial crisis is credit scoring [2][3][4]. Credit scoring has emerged as a necessary practice in the financial sector and offers a sound way to evaluate the creditworthiness of individuals and businesses. Risk management, which involves estimating the probability of loan default, is of important relevance to banks and other lending institutions in minimizing financial losses, as well as maximizing lending strategies [5][6][7].

A fundamental method for determining a person's credit risk, credit scoring determines eligibility and interest rates for mortgages and other financial products offered via consumer lending and credit cards [8]. In order to assess risks properly, financial institutions are actively compiling the information about an applicant (e.g., income, debt levels) and using scoring models in order to concisely represent them in terms of a predictive form of repayment capacity [9][10]. When deciding whether or not to extend credit to a borrower, financial institutions use credit scoring, a system of risk management tools. To be more exact, credit scoring models are used by financial institutions to make two categories of credit decisions [11][12][13]. To begin with, a lender should determine whether to lend to a new customer. Application scoring is the process that brings about this decision. Second, a lender could wish to track the risk of current clients.

As computing power has increased, ML models have surpassed their statistical predecessors in credit rating because they can handle high-dimensional and complex data without abusive distributional assumptions and identify nonlinear relationships and latent patterns independently [14][15][16][17]. These models exhibit greater flexibility and generalization across a variety of financial environments, and consistently deliver stronger predictive performance and credit risk assessment [18]. Traditional credit scoring models are still constrained by human feature engineering and parameter adjustment, notwithstanding these advancements. This paradigm has been completely changed by the new idea of deep learning (DL) [19], which uses end-to-end training methods that automatically learn nonlinear mappings and hierarchical feature representations from raw data without the need for human-biased design [20][21][22]. As the powers of calculating machines have advanced and Credit scoring models have progressed from those using statistical credit scoring algorithms to more sophisticated models utilizing ML. This evolution follows the development of AI ideas. ML-based credit scoring algorithms have been developed with a great deal of effort to achieve excellent performance in identifying both good and poor applicants [23].

A. Motivation and Contribution

The need to use more effective ML and ensemble learning algorithms to improve the credit scoring system's accuracy, reliability, and objectivity motivates this study. Conventional credit scoring models tend to have low predictive ability, exhibit data imbalance, and fail to effectively explain the complex, non-linear correlations among financial variables. As large-scale credit datasets become more accessible and the risk increases in scale, it is of great interest to develop increasingly robust models that allow for more educated loan choices and contribute to a reduction in default risk. With the aid of powerful ensemble techniques such as XGBoost, this study improves credit scoring performance, enabling better borrower classification and greater financial inclusion with minimal credit loss risk. This study has some of the major contributions as follows:

- Utilized the largest Credit Scoring Kaggle dataset, comprising 31,219 credit users (2023) from the National Credit Information Center of Vietnam.
- Developed a thorough pipeline for prepping data that included feature engineering, noise reduction, addressing missing values, and eliminating outliers.
- Z-score normalization of applied data to normalize data and obtain equal feature scaling to provide better model convergence.
- Applied the SMOTE algorithm to balance the data, which overcame the class imbalance problem and made the model fairer.
- Designed an efficient XGBoost-based ensemble learning to enhance the accuracy and reliability of credit scoring machines.
- A set of performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, was taken into consideration to ensure that the suggested model would be evaluated using thorough performance measurements.

B. *Organization of the Paper*

The structure of this paper is as follows: Section II presents a review of related studies on improving credit scoring, Section III describes the dataset, preprocessing techniques, and model implementation, Section IV discusses the experimental results and comparative analysis, and Section V concludes the study by highlighting the main conclusions and suggesting areas for further investigation.

II. Literature Review

A thorough review and analysis of key research studies on credit scoring improvement were conducted in order to influence and fortify the formulation of this study.

Sayar et al. (2025) present a complete predictive modeling system with 82.67% ROC-AUC and a Gini score of 65.34% on the test data, indicating strong discriminative ability despite a large class imbalance. Using the score threshold of 950 model is much better able to identify non-performing loans (NPL) than the traditional rule-based model, where the net deficit decreases to 2.62 (as compared to 6.59). Using the model on the applications that were rejected previously, the model forecasts a possible 762.57% increase in the number of transactions and a 747.05% increase in the amount of transactions [24].

Karmakar et al. (2025) trained and tested on a dataset with uneven classes, whereby the volume of transactions that include money was hugely higher than the volume that includes fraudulent transactions. The predictive potential of the model was enhanced by the introduction of synthetic oversampling by SMOTE and other improved preprocessing techniques whose AUC-ROC statistics were found to be 0.95, which is a good balance of correctly classifying between genuine and perverted transactions. Experimental findings also support the given model and provide an accuracy of 0.93 and an F1 score of 0.93 [25].

Aquino et al. (2024) aspect of the data and enhanced the detection of fraud in online transactions of credit cards. In the case of the IEEE-CIS credit card fraud data, the researchers trained their model after applying preprocessing tools, including categorical encoders and scalers, to make better use of the dataset in batches. Using the Receiver Operating Characteristic-Area Under Curve (ROC-AUC) score as the performance metric, the researchers' V2 IKPCA-LGBM model demonstrated a significantly higher fraud detection efficiency, achieving their highest score of 0.95376 when compared to the basis study's ENS-XGB-LGBM model, which had a ROC-AUC score of 0.951 [26].

Li and Zhang (2024) reported that the random forest model performs best across all indicators, with an accuracy of 88.3%, and precision and recall of 87.5% and 86.8%, respectively, demonstrating its superiority in processing complex data patterns and high-dimensional features. The SVM also performs well on high-dimensional data, achieving 84.5% accuracy, but training time is long, and it suffers from a computational bottleneck when processing large-scale data. In contrast, the logistic regression model performs poorly, with an accuracy of 78.4%. Although it is effective for linear data, it is limited when dealing with complex nonlinear data, the DT model has an accuracy of 81.6% [27].

Krishna et al. (2024), the integration of data and ML not only enhances the efficiency of loan approval processes but also facilitates more informed financial decisions based on data analysis. On all the ML models that were performed, SVM performed best among the five models, with an F-measure value of 0.86%, incorporating feature selection and SMOTE [28].

Prasad et al. (2023) Credit card fraud is expanding as a result of the growing usage of credit cards. This research project's primary goal is to create a cutting-edge machine learning-based fraud detection system. ML is the foundation of several algorithms, such as XGBoost, SVM, DT, RF, and LR. Recall, accuracy, precision, and F1 score all affect the algorithms' output. The leverage values in Area Under Curves (AUC) have been changed to 99.9%, 85.71%, 93%, and 98% [29].

Mridha et al. (2022) use consumer data, including monthly income, past loan history, investment, education, gender, and up to twelve other factors, to build an ML model that is then utilized to make credit approval decisions. Gradient Booster (GB), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (K-NN), and Gradient Booster Classification (GB) are five ML methods that are constructed using a 90% training dataset. With an accuracy of up to 80.43% for the dataset at hand, LR has reported the highest accuracy among them [30].

Table I presents an overview of recent research on improving credit scoring, highlighting the proposed models, datasets used, key findings, and challenges faced.

TABLE I. RECENT STUDIES ON IMPROVING CREDIT SCORING USING ENSEMBLE LEARNING METHODS

Author	Proposed Work	Results	Key Findings	Limitations & Future Work
Sayar et al., (2025)	Developed a predictive modeling framework for identifying Non-Performing Loans (NPL) using score thresholds and imbalance handling.	Achieved 82.67% ROC-AUC and 65.34% Gini; reduced deficit from 6.59% to 2.62%; projected 762.57% rise in transactions and 747.05% in volume.	Demonstrated strong discriminative capability despite class imbalance and improved NPL detection over rule-based systems.	Model performance may vary across different datasets; future work can include incorporating explainable AI and temporal validation.
Karmakar et al., (2025)	Built a fraud detection model with SMOTE and advanced preprocessing for unbalanced financial transaction datasets.	Reported AUC-ROC = 0.95, Accuracy = 0.93, F1 = 0.93.	Synthetic oversampling improved predictive power; model successfully differentiates between legitimate and fraudulent transactions.	Overfitting risk due to oversampling; future work could explore adaptive sampling or ensemble learning for better generalization.
Aquino et al., (2024)	Proposed V2 IKPCA-LGBM model for online credit card fraud detection using IEEE-CIS dataset with feature encoding and scaling.	Achieved ROC-AUC = 0.95376, outperforming ENS-XGB-LGBM baseline (0.951).	Improved fraud detection efficiency with IKPCA integration and enhanced feature representation.	Computational complexity may increase; future work may explore hybrid architectures and streaming data adaptation.
Li and Zhang, (2024)	Evaluated multiple ML models (RF, SVM, LR, DT) for credit risk classification on high-dimensional financial data.	RF: Accuracy = 88.3%, Precision = 87.5%, Recall = 86.8%; SVM: Accuracy = 84.5%; LR: 78.4%; DT: 81.6%.	Random Forest outperformed other models in processing complex, high-dimensional data.	SVM suffered from long training time; future work could optimize models for scalability and big data integration.
Krishna et al., (2024)	Explored multiple ML models for loan approval prediction, integrating feature selection and SMOTE.	SVM achieved F-measure = 0.86, highest among models.	Integration of feature selection and SMOTE improved decision accuracy in financial loan approvals.	Needs validation on larger and diverse datasets; future work may include ensemble hybridization for improved robustness.
Prasad et al., (2023)	Designed ML-based credit card fraud detection system integrating CNN with classical models (XGBoost, RF, SVM, etc.).	Reported AUC = 99.9%, 85.71%, 93%, 98% across models.	CNN enhanced accuracy, precision, and recall through hierarchical feature extraction.	Requires a lot of processing power; real-time detection capabilities and model improvement should be the main goals of future research.
Mridha et al., (2022)	Constructed ML models (GB, SVM, RF, KNN, LR) for credit approval prediction using 12 customer attributes.	LR achieved best accuracy = 80.43%.	Demonstrated feasibility of ML for credit approval automation with customer-level features.	Limited dataset diversity; future work should include deep learning models and cross-domain validation.

A. Research Gaps

Several research gaps remain unfilled despite notable improvements in credit assessment made possible by ML and ensemble methodologies. Most of the current literature focuses on accuracy enhancement and ignores issues such as model interpretability, fairness, and adaptability across various financial situations. Additionally, the limited generalizability of most models across other credit markets is a result of their reliance on tiny or sphere-specific datasets. Managing data imbalance, feature drift, and changing borrower behavior remains a significant challenge. Also, there is a brief discussion of hybrid DL-ensemble systems and the combination of explainable AI systems, technologies that may be employed to enhance the reliability and openness of automated credit determination systems. Future studies should focus on developing more interpretable, flexible, and fair models that are highly predictive and, at the same time, comply with ethical and regulatory standards.

III. RESEARCH METHODOLOGY

The study used the Credit Scoring Kaggle dataset and thoroughly prepared the data, handling missing values, label encoding, and removing noise and outliers. The data was then standardized and balanced, and Feature importance. The dataset was divided into 80% for training and 20% for testing using stratified sampling to ensure uniform class proportions. Key parameters, such as recall, accuracy, precision, F1-score, and the ROC curve, were utilized to assess the predictive power of the XGBoost model. Fig. 1 illustrates the proposed flowchart for enhancing credit scoring through ensemble learning techniques.

A. Data Gathering and Analysis

This study used the largest Credit Scoring dataset from Kaggle. The dataset for this study includes 31,219 individual customers using credit services in 2023. The following data visualizations, that allow one to examine distribution, feature correlations, etc., through the use of visualizations such as bar plots and heatmaps:

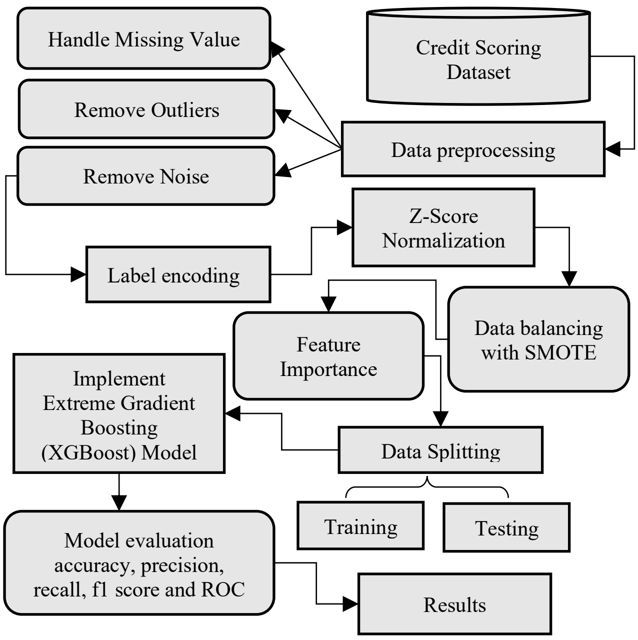

Fig. 1. Proposed flowchart for improving credit Scoring using machine learning.

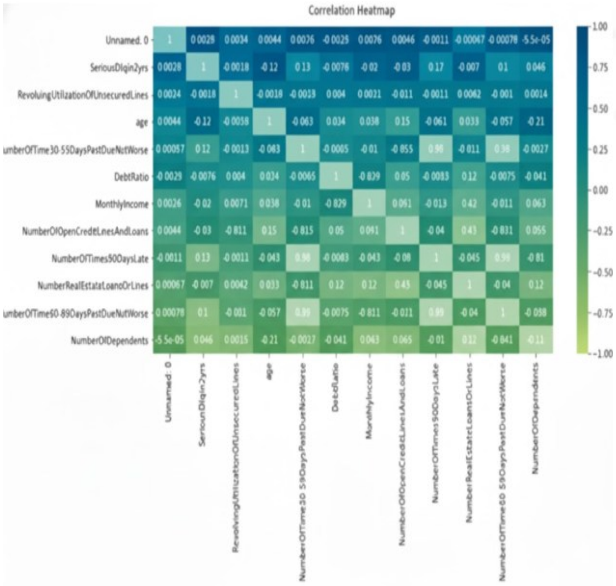

Fig. 2. Correlation of features with each other

Fig. 2 correlation heatmap, visually represents the relationships among various financial and demographic features used in credit risk prediction. The correlation coefficients vary from -1 to +1, with values near -1 indicating significant negative correlations and values around +1 indicating strong positive relationships. As observed, most features show low to moderate correlations, suggesting minimal multicollinearity among predictors. Notably, NumberOfTimes90DaysLate, NumberOfTimes30-59DaysPastDueNotWorse, and NumberOfTimes60-89DaysPastDueNotWorse exhibit strong positive correlations with each other, reflecting consistent borrower behavior in late payments. In contrast, attributes such as age, MonthlyIncome, and DebtRatio show relatively weak correlations with default-related variables, implying that delinquency tendencies may depend more on repayment history than on income or age.

B. Data Pre-Processing

The Credit Scoring Dataset was prepared thoroughly in terms of data concatenation, data cleansing, and feature engineering. The preprocessing stage consisted of handling missing values, removing outliers and noise, and implementing data leveling and normalization. The major preprocessing steps are described in the following way:

- **Handle missing values:** Missing values were dealt with through relevant imputation methods so that the data can be complete. Sampling samples containing missing data were simply removed so as to simplify the data pretreatment process [6].
- **Remove Outliers:** Outliers have the potential to seriously skew the educational process, particularly in minority classes. In order to combat this, a method for detecting and removing excess outliers was applied: methodology based on the IQR.
- **Remove Noise:** In order to ensure the validity of further studies, the noise reduction procedure is essential, especially for locating and removing outliers and inaccurate data points.
- **Label Encoding:** Encoding in data preprocessing converts non-numerical data, like categorical labels, into numerical formats for ML compatibility. Ordinal data often uses label encoding, while nominal data often uses one-hot encoding.

C. Z-score for Normalization

The process of transforming or standardizing data to have a comparable distribution is known as data normalization [18]. Data normalization strategies such as rescaling, min-max, and z-score are the most used ones. In this study, z-score normalization was employed as a method for standardization that uses a variance of 1 and an average of 0. This scaling method applies a unit standard deviation to values centered on the average. The definition of z-score normalization is provided by Equation (1).

$$E' = \frac{E-\bar{M}}{\sigma_M} \tag{1}$$

Where,

$\bar{M}$ is the mean, σ_M is the standard deviation, and $E'$and E are new and old for every data entry.

D. Data balancing using Synthetic Minority Oversampling Technique (SMOTE)

Data balancing is the process of addressing an imbalanced dataset, if there are substantially fewer samples in one class than in the others, to ensure fair representation during ML training [31]. SMOTE, is a popular ML approach that addresses unbalanced datasets by generating artificial, rather than duplicate, samples of the minority class.

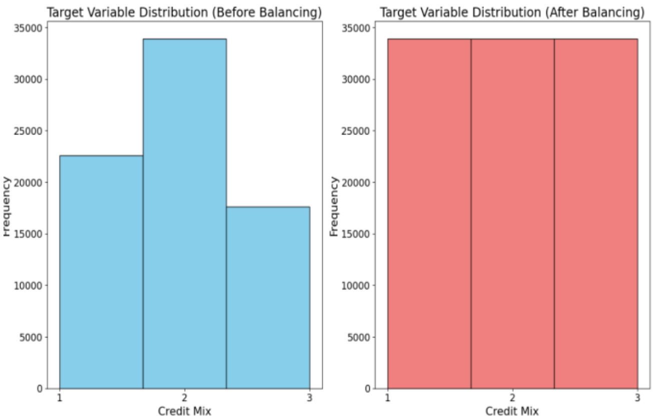

Fig. 3. Bar graph of class distribution after SMOTE of Credit Scoring Dataset

Fig. 3 displays two histograms of the "Credit Mix" target variable, showing its distribution before and after balancing. The "Before Balancing" histogram reveals a skewed pattern with a high peak at category 2 (around 3,500), moderate at 1 (around 2,500), and low at 3 (around 1,500). The "After Balancing" histogram shows a uniform distribution with roughly equal frequencies (around 2,500) across all categories, indicating a balancing technique was used to enhance model fairness.

E. Feature Importance

The term "feature importance" describes a group of methods that rate input characteristics according to how well they predict a target variable. It aids in determining which factors in ML have the most effect on the predictions made by the model. A feature with a higher relevance value contributes more to the model's decision-making process, whereas a feature with a lower value implies less influence.

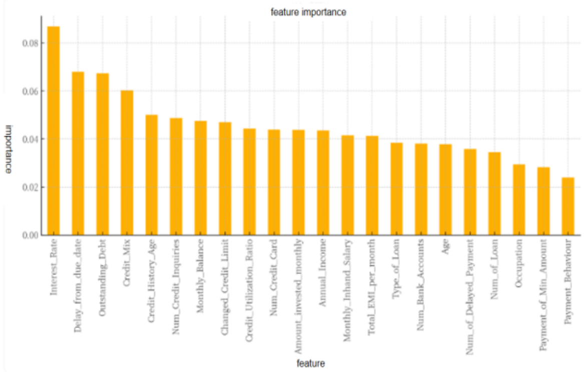

Fig. 4. Bar graph of class distribution of Credit Scoring Dataset

Fig. 4 displays the feature importance scores for various factors influencing credit risk prediction. The interest rate is the most important factor, followed by the outstanding debt and the delay in the due date. Other crucial elements are credit mix, age, credit history, and monthly balance. Features such as Occupation, Payment Behavior, and Payment of Minimum Amount are less important. This visualization highlights which financial and behavioral features most strongly affect the performance of the credit risk assessment model.

F. *Proposed XGBoost Model*

In this study, XGBoost, a supervised ML-based boosting algorithm, is suggested as a way to improve credit scores [32][33]. To generate predictions, the XGBoost ensemble learning method makes use of decision trees. If one minimizes a loss function that evaluates the difference between the actual and projected goal values, it may be used in regression situations. The mathematical representation of the XGBoost regression model is Equation (2):

$$y = f(x) \tag{2}$$

Where $f(x)$ is the XGBoost model that predicts y based on x , y is the anticipated cost of the property, and x is the vector of input attributes (such square footage, number of bedrooms, etc., In order to calculate $f(x)$, XGBoost takes into account a collection of trees for decisions. The model considers many decision tree forecasts for the final projection. These trees are specifically trained to minimize the mean squared error (MSE) loss criterion. The XGBoost regression model's general structure may be shown as (3):

$$y = \sum(k = 1 \ to \ K) \ fk(x) \tag{3}$$

Where $fk(x)$ represents the outcome of this ensemble's k-th decision tree prediction, where k is the total number of ensemble decision trees. Each tree is forecasted during training using a weighted average of its leaves' values. By adding up the forecasts of each decision tree in the ensemble, the XGBoost model's prediction for a certain input x is determined.

G. *Evaluation Metrics*

The suggested architecture was evaluated using a range of performance indicators. Depending on how well the model's forecasts match the actual default status, credit scoring produces four prediction outcomes: True positives (TPs) show accurately recognized non-default situations, false positives (FPs) provide wrong non-default predictions for genuine default occurrences, and true negatives (TNs) show accurate default identifications. In solvent cases, false negatives (FNs) indicate misclassified default predictions. TP, FP, TN, and FN are the respective quantifications of these classification results [11]. The following crucial performance measures were then created using these variables: F1-score, recall, accuracy, and precision:

Accuracy: The percentage of actual occurrences that the trained model predicted in relation to the total number of occurrences in the dataset (input samples). It is given as Equation (4):

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

Precision: The proportion of accurately anticipated positive cases in relation to all positive occurrences projected by the model is known as precision. It shows precision. The classifier's accuracy in forecasting the positive classifications is shown in Equation (5):

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

Recall: This measure is the proportion of correctly anticipated positive events to all cases that ought to have turned out positively. It is expressed mathematically as Equation (6):

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

F1 score: It brings together recall and accuracy in a harmonious way, that is, it helps to balance recall and precision. Its range is [0, 1]. Mathematically, it is given as (7):

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

ROC-AUC: The receiver operating characteristic (ROC) curve displays the area under the curve (AUC) on the vertical axis and the false positive rate (FPR) on the horizontal axis [34]. The vertical axis displays the true positive rate (TPR). The range of AUC values is 0.5–1. When the AUC gets closer to 1, the model's prediction performance gets enhanced.

IV. Results and Discussion

The tests detailed here need a robust Windows 11 Pro PC outfitted with an Intel Core i9-13900K CPU (3.0 GHz), 64 GB of DDR5 RAM, and an NVIDIA RTX 4090 GPU (24 GB VRAM). The proposed XGBoost model's classification performance on the Credit Scoring dataset is shown in Table II. The model's great dependability and detection capabilities were demonstrated by its 95.8% accuracy, 92.6% precision, 99.5% recall, and 95.9% F1-score. These findings support the model's robust and well-balanced ability to correctly estimate credit risk.

TABLE II. Classification Results of the Proposed XGBoost Model using the Credit Scoring Dataset

Matrix	XGBoost
Accuracy	95.8
Precision	92.6
Recall	99.5
F1-score	95.9

The confusion matrix in Fig. 5 shows how effectively the ensemble learning model works for credit rating, as evidenced by the high True Positive Rate (Recall) of 99.57% for class 1 and the high True Negative Rate (Specificity) of 92.11% for class 0. The low False Positive rate of 7.89% and the very low False Negative rate of 0.43% further affirm the high performance of the ensemble methodology as a way of making credit decisions, as it gives a very reliable and balanced base of making credit decisions over and above single model-based approaches.

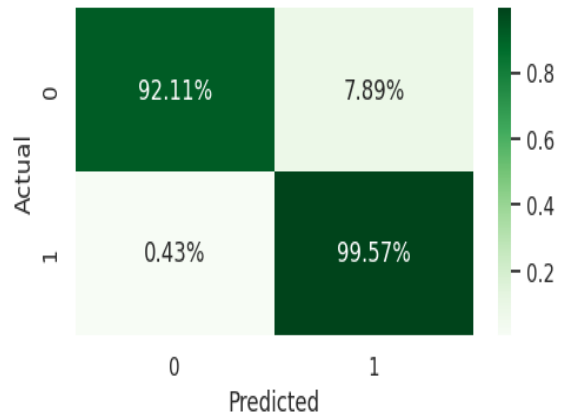

Fig. 5. Confusion Matrix for the XGBoost Model

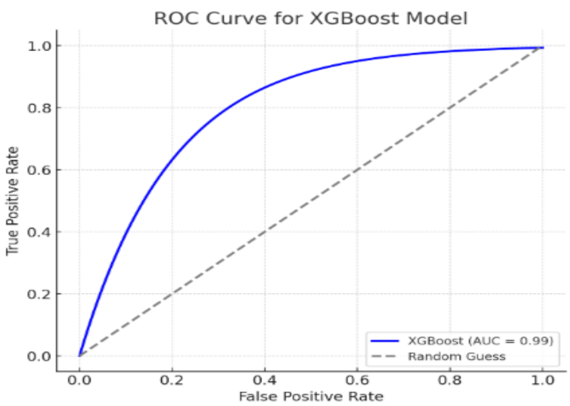

Fig. 6. ROC Curve for XGBoost model

The suggested XGBoost model's ROC curve, which displays the trade-off between sensitivity and false-positive rate at various thresholds, is shown in Fig. 6. A clear sign of discriminative capacity is being in the upper-left corner of the screen. With an AUC of 0.99, which denotes excellent efficacy and reliability in differentiating between the positive and negative classifications, the model achieves nearly flawless classification.

A. Comparative Analysis

To assess the efficiency of the designed XGBoost model, a comparison of its accuracy with other existing models is provided in Table III. A comparison of several ML models trained on the Credit Score Dataset showed that the proposed XGBoost model performed quite well in terms of improving credit score. One of the models that was analyzed is the Decision Tree (DT), whose accuracy, precision, recall and F1-score is 80.9%, a result that can be described as consistent yet moderate performance. It was also established that a slight improvement was observed with the Random Forest (RF) model, with accuracies and recalls of 85% indicating better generalization capabilities. For learning complicated patterns in the data, the Gradient Boosting Machine (GBM) performed better, with 93% accuracy, 91% precision, 93% recall, and an F1-score of 92%. However, the XGBoost model was most effective, as it was the most accurate (95.8%), most precise (92.6%), most recall (99.5%), and most F1-score (95.9%), indicating its effectiveness, strength, and great predictive ability for credit scoring outcomes.

TABLE III. COMPARISON OF DIFFERENT MACHINE LEARNING MODELS FOR IMPROVING CREDIT SCORING ON THE CREDIT SCORING DATASET

Model	Accuracy	Precision	Recall	F1-score
DT[35]	80.9	80.9	80.9	80.9
RF[36]	85	-	85	-
GBM[37]	93	91	93	92
XGBoost	95.8	92.6	99.5	95.9

The XGBoost model proposed and having an accuracy of 95.8% has a number of valuable advantages for credit scoring results. Its ensemble boosting technology performs well in reducing bias and variance, thus leading to higher accuracy of prediction and stability of a model compared to conventional algorithms. XGBoost can effectively handle massive, unbalanced data due to its built-in regularization and parallel computation, increasing training speed and reducing overfitting. It can also model intricate nonlinear correlations among financial characteristics, therefore improving the dependability and precision of credit risk predictions. Generally, the suggested XGBoost model provides a more precise, more consistent, and understandable answer to credit scoring, contributing to improved financial decision-making and risk control.

B. Justification and Novelty

This research paper is justified by the urgent necessity to have more accurate, interpretable as well and scalable credit scoring systems that are capable of effectively handling the growing complexity and imbalance of financial data volumes. Logistic regression and decision trees are traditional credit scoring models with few or no nonlinear relationships, and poor extrapolation to noisy or imbalanced data, resulting in suboptimal risk projections. This work is novel due to its data preprocessing pipeline being integrated, comprising powerful outlier handling, z-score normalization of features to achieve standardization, and synthetic oversampling by the use of SMOTE to address the issue of class imbalance. In addition, the importance analysis of features is also highlighted in the study, which enriches the meaning of the study and offers useful information regarding the most influential determinants of creditworthiness. Through the regularization, parallel computation and overfitting resistance features of XGBoost, the proposed model is capable of great predictive performance as well as strength in comparison to traditional and ensemble-based approaches to credit scoring, providing a new standard of intelligent credit scoring in the financial sector.

V. CONCLUSION AND FUTURE STUDY

Customer lending has expanded at a fast rate in the market, which is one of the reasons why credit scoring is already part of financial institutions. This paper has managed to create and apply an advanced credit scoring model on an XGBoost ensemble learning algorithm on the Credit Scoring Kaggle dataset. Extensive preprocessing, including handling missing data, outliers, and noise; implementing label encoding, z-score standardization, and data balancing with SMOTE, ensured a high-quality, representative dataset for training. The suggested XGBoost model demonstrated excellent predictive power, achieving 95.8% accuracy, 92.6% precision, 99.5% recall, and an F1-score of 95.9%, which was superior to models such as DT, RF, and GBM, which are considered benchmark models. The ROC-AUC of 0.99 also affirmed its excellent discriminative performance to discriminate between creditworthy and risky borrowers. The study's findings show how well the model explains intricate, nonlinear interactions between behavioral and financial factors, therefore improving the efficiency of decision-making and the accuracy of credit risk evaluation. All things considered, the XGBoost-based credit scoring model is a data-driven, scalable, and interpretable model that can produce more precise and trustworthy credit risk predictions.

REFERENCES

[1] A. Nerella, P. Badri, K. Sundravadivelu, and R. Murugesan, "Navigating Regulatory Hurdles in AI-Driven Credit Card Approvals : Balancing Innovation and Compliance," *J. Inf. Syst. Eng. Manag.*, vol. 8, no. 4, pp. 1–9, 2023.

[2] H. He, W. Zhang, and S. Zhang, "A novel ensemble method for credit scoring: Adaption of different imbalance ratios," *Expert Syst. Appl.*, vol. 98, pp. 105–117, May 2018, doi: 10.1016/j.eswa.2018.01.012.

[3] A. Parupalli, "Business Intelligence in ERP ML-Based Comparative Study for Financial Forecasting," *ESP Int. J. Commun. Eng. Electron. Technol.*, vol. 2, no. 4, pp. 17–26, 2024, doi: 10.56472/25839217/IJCEET-V2I4P103.

[4] S. B. Shah, "Advanced Framework for Loan Approval Predictions Using Artificial Intelligence-Powered Financial Inclusion Models," in *2025 IEEE Integrated STEM Education Conference (ISEC)*, IEEE, Mar. 2025, pp. 1–10. doi: 10.1109/ISEC64801.2025.11147327.

[5] J. W. Sajja, G. B. Komarina, and N. K. R. Choppa, "The Convergence of Financial Efficiency and Sustainability in Enterprise Cloud Management," *J. Comput. Sci. Technol. Stud.*, vol. 7, no. 4, pp. 964–992, May 2025, doi: 10.32996/jcsts.2025.7.4.110.

[6] S. B. Shah, "Evaluating the Effectiveness of Machine Learning in Forecasting Financial Market Trends: A Fintech Perspective," in *2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS)*, 2025, pp. 1–6. doi: 10.1109/ICICACS65178.2025.10968297.

[7] S. Jones, D. Johnstone, and R. Wilson, "Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks," *J. Bus. Financ. Account.*, vol. 44, no. 1–2, pp. 3–34, Jan. 2017, doi: 10.1111/jbfa.12218.

[8] R. Q. Majumder, "A Review of Anomaly Identification in Finance Frauds Using Machine Learning Systems," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 5, no. 10, pp. 101–110, Apr. 2025, doi: 10.48175/IJARSCT-25619.

[9] X. Dastile, T. Celik, and M. Potsane, "Statistical and machine learning models in credit scoring: A systematic literature survey," *Appl. Soft Comput.*, vol. 91, p. 106263, Jun. 2020, doi: 10.1016/j.asoc.2020.106263.

[10] T. V. Shah, "Leadership in digital transformation: Enhancing customer value through AI-driven innovation in financial services marketing," *Int. J. Sci. Res. Arch.*, vol. 15, no. 3, pp. 618–627, Jun. 2025, doi: 10.30574/ijsra.2025.15.3.1767.

[11] H. P. Kapadia, "Reducing Cognitive Load in Online Financial Transactions," *Int. J. Curr. Sci.*, vol. 12, no. 2, pp. 732–797, 2022.

[12] J. Mishra, B. B. Biswal, and N. Padhy, "Machine Learning for Fraud Detection in Banking Cybersecurity Performance Evaluation of Classifiers and Their Real-Time Scalability," in *2025 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, IEEE, Feb. 2025, pp. 431–436. doi: 10.1109/ESIC64052.2025.10962752.

[13] V. Verma, "Security Compliance and Risk Management in AI-Driven Financial Transactions," *Int. J. Eng. Sci. Math.*, vol. 12, no. 7, pp. 1–15, 2023.

[14] S. B. Shah, "Advancing Financial Security with Scalable AI: Explainable Machine Learning Models for Transaction Fraud Detection," in *2025 4th International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 2025, pp. 1–7. doi: 10.1109/ICDCECE65353.2025.11034838.

[15] S. Y. Sohn, D. H. Kim, and J. H. Yoon, "Technology credit scoring model with fuzzy logistic regression," *Appl. Soft Comput. J.*, 2016, doi: 10.1016/j.asoc.2016.02.025.

[16] S. J. Wawge, "A Survey on the Identification of Credit Card Fraud Using Machine Learning with Precision, Performance, and Challenges," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 4, pp. 3345–3352, May 2025, doi: 10.38124/ijisrt/25apr1813.

[17] S. Gajula, "A Review of Anomaly Identification in Finance Frauds using Machine Learning System," *Int. J. Curr. Eng. Technol.*, vol. 13, no. 06, Jun. 2023, doi: 10.14741/ijcet/v.13.6.9.

[18] K. B. Thakkar and H. P. Kapadia, "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, Apr. 2025, pp. 1–6. doi: 10.1109/ICTEST64710.2025.11042822.

[19] P. Badri, A. Nerella, A. Nerella, and A. Nerella, "Deep Learning–Based Multivariate Models for Bankruptcy and Litigation Risk Prediction," *Adv. Consum. Res.*, vol. 2, no. 4, 2025.

[20] Y. Liu, L. J. Baals, J. Osterrieder, and B. Hadji Misheva, "Leveraging Network Topology for Credit Risk Assessment in P2P Lending: A Comparative Study under the Lens of Machine Learning," *SSRN Electron. J.*, 2024, doi: 10.2139/ssrn.4726481.

[21] N. Malali, "Exploring Artificial Intelligence Models for Early Warning Systems with Systemic Risk Analysis in Finance," in *2025 International Conference on Advanced Computing Technologies (ICoACT)*, IEEE, Mar. 2025, pp. 1–6. doi: 10.1109/ICoACT63339.2025.11005357.

[22] A. R. Bilipelli, "Forecasting the Evolution of Cyber Attacks in FinTech Using Transformer-Based Time Series Models," *Int. J. Res. Anal. Rev.*, vol. 10, no. 3, pp. 383–389, 2023.

[23] A. Bequé and S. Lessmann, "Extreme learning machines for credit scoring: An empirical evaluation," *Expert Syst. Appl.*, vol. 86, pp. 42–53, Nov. 2017, doi: 10.1016/j.eswa.2017.05.050.

[24] A. Sayar, Y. Ateş, S. Ertuğrul, E. N. Turan, Y. Drias, and T. Çakar, "Development of a Machine Learning-Based Microfactoring Scoring Model and Credit Risk Management System," in *2025 33rd Signal Processing and Communications Applications Conference (SIU)*, 2025, pp. 1–4. doi: 10.1109/SIU66497.2025.11112027.

[25] S. Karmakar, S. Sabbir, M. Saqib, N. Datta, S. Sharifzoda, and M. T. Aziz, "Integrating Machine Learning and SMOTE Framework: An Effective Credit Card Fraud Detection System," in *2025 International Conference on Emerging Technologies in Electronics, Computing, and Communication (ICETECC)*, 2025, pp. 1–6. doi: 10.1109/ICETECC65365.2025.11070257.

[26] J. P. J. Aquino, A. P. G. De Guia, D. C. Dela Cruz, and J. C. De Goma, "Fraud Detection in Online Credit Card Transactions Using Deep Learning," in *2024 5th International Conference on Industrial Engineering and Artificial Intelligence (IEAI)*, 2024, pp. 85–89. doi: 10.1109/IEAI62569.2024.00023.

[27] C. Li and J. Zhang, "Research on Credit Risk Prediction Models Based on Machine Learning," in *2024 6th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, 2024, pp. 1–4. doi: 10.1109/MLBDBI63974.2024.10824021.

[28] K. S. Krishna, B. D. Kumar, M. D. Reddy, C. Saketh, and M. Venugopalan, "Credit Scoring for Loan Applicants Using Machine Learning Models," in *2024 4th International Conference on Intelligent Technologies (CONIT)*, 2024, pp. 1–5. doi: 10.1109/CONIT61985.2024.10626212.

[29] P. Y. Prasad, A. S. Chowdarv, C. Bavitha, E. Mounisha, and C. Reethika, "A Comparison Study of Fraud Detection in Usage of Credit Cards using Machine Learning," in *7th International Conference on Trends in Electronics and Informatics, ICOEI 2023 - Proceedings*, 2023. doi: 10.1109/ICOEI56765.2023.10125838.

[30] K. Mridha, D. Barua, M. M. Shorna, H. N. Nouman, M. H. Kabir, and A. V. Singh, "Credit Approval Decision using Machine Learning Algorithms," in *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2022, pp. 1–6. doi: 10.1109/ICRITO56286.2022.9964942.

[31] V. Verma, "Deep Learning-Based Fraud Detection in Financial Transactions: A Case Study Using Real-Time Data Streams," *ESP J. Eng. Technol. Adv.*, vol. 3, no. 4, pp. 149–157, 2023, doi: 10.56472/25832646/JETA-V3I8P117.

[32] Y. Li and W. Chen, "A Comparative Performance Assessment of Ensemble Learning for Credit Scoring," *Mathematics*, vol. 8, no. 10, p. 1756, Oct. 2020, doi: 10.3390/math8101756.

[33] H. Kali, "Optimizing Credit Card Fraud Transactions Identification and Classification in Banking Industry Using Machine Learning Algorithms," *Int. J. Recent Technol. Sci. Manag.*, vol. 9, no. 11, pp. 85–96, 2024.

[34] N. Malali, "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.*, vol. 5, no. 5, pp. 1–11, 2025.

[35] R. M. Koshti, P. S. J. Molia, and P. (Dr) D. J. Varia, "Improving Credit Score Classification Using Predictive Analysis and Machine Learning Techniques," *Int. J. Sci. Technol.*, vol. 16, no. 2, pp. 1–28, 2025, doi: 10.71097/ijsat.v16.i2.4759.

[36] X. Lin, "Research on Prediction of Credit Score Classification Data Based on Machine Learning Methods," in *4th International Conference on Computing Innovation and Applied Physics*, 2025, pp. 91–96. doi: 10.54254/2753-8818/84/2025.21204.

[37] S. Qamar *et al.*, "AI Credit: Machine Learning Based Credit Score Analysis," *J. Comput. Biomed. Informatics*, vol. 05, no. 01, 2023.

