Algora *Volume: 1 Issue:1 6/2024/pp.1-15*

Journal of AI, Data Science and Engineering, Peer Reviewed, Open Access Journal

MACHINE LEARNING FOR ENHANCED CREDIT SCORING

[1]Abbey Bakare
abbeybakare10@gmail.com
https://orcid.org/0009-0002-1971-0283

[2]Anjola Odunaike
Odunaikea21@ecualumni.ecu.edu
https://orcid.org/0009-0002-1907-6753

[1]Western Governors University
[2]East Carolina University

Abstract

This study investigates the application of machine learning to improve the accuracy, fairness, and interpretability of credit scoring systems. Using the publicly available "Give Me Some Credit" dataset and following the methodology proposed by Ichim and Issa (2025), the research demonstrates how preprocessing techniques and advanced ensemble models, particularly Gradient Boosting Machines (GBM), enhance predictive performance. Key processes included data cleaning, class balancing with SMOTE, and feature scaling, which collectively improved the model's Area Under the Curve (AUC) from 0.83 to 0.87. Figures derived from the model illustrate the most influential features, compare model discrimination through ROC analysis, and highlight the impact of balancing strategies on performance. The study also emphasizes the importance of model interpretability and regulatory transparency in the adoption of machine learning within financial decision-making. This research contributes a replicable, interpretable framework for institutions seeking to modernize credit risk assessment while maintaining compliance and trust.

Keywords: credit scoring, machine learning, data preprocessing, gradient boosting, model interpretability, credit risk analysis

1. Introduction

Credit scoring plays a pivotal role in modern financial systems by enabling lenders to assess the likelihood that borrowers will repay their debts. Traditionally, statistical models such as logistic regression have been the foundation for credit risk evaluation due to their simplicity and interpretability. However, as credit markets grow more complex and borrower behaviors more dynamic, these conventional approaches increasingly fall short in capturing nonlinear relationships and uncovering hidden patterns in financial data (Mestiri, 2024; Bravo, Thomas, and Weber, 2015). In recent years, machine learning has emerged as a powerful tool in enhancing credit scoring systems.

These methods offer improved predictive accuracy, adaptability to large and unstructured data, and the ability to uncover intricate interactions among financial variables (Ichim and Issa, 2025; Reza et al., 2024). A growing body of research supports the adoption of machine learning in both corporate and consumer credit risk evaluation. Studies have demonstrated that models such as gradient boosting machines, random forests, and deep neural networks outperform traditional techniques in various real-world applications (Yan, Zhang, and Shen, 2025; Tong et al., 2024; Dil, 2025).

The effectiveness of machine learning models in credit scoring is further strengthened by advances in data preprocessing and feature engineering. Researchers have shown that proper handling of class imbalance, missing values, and feature transformations significantly boosts model performance and stability (Ichim and Issa, 2025; Golec and AlabdulJalil, 2025). These developments address longstanding challenges in credit risk modeling, particularly in emerging markets where data quality and completeness are often limited (Akil, Sekkate, and Abdellah, 2021; Yan, Zhang, and Shen, 2025). In parallel, the use of explainable artificial intelligence (XAI) methods such as SHAP and LIME has increased transparency in ML-driven decision-making. These tools help financial institutions understand the factors influencing credit approval and facilitate compliance with regulatory requirements (Chen, Calabrese, and Martin-Barragan, 2023; Reza et al., 2024). Interpretability is now a critical consideration, especially in high-stakes environments such as credit allocation and loan underwriting (Golec and AlabdulJalil, 2025; Lavecchia et al., 2025).

Moreover, researchers are now exploring the integration of alternative and unstructured data, including mobile transactions, social media behavior, and even geolocation signals, into credit models. This trend broadens the scope of credit assessment, particularly for underserved populations and thin-file customers (Óskarsdóttir et al., 2020; Hollmann et al., 2025; Lavecchia et al., 2025). Innovative frameworks are being developed to accommodate these data streams while maintaining model robustness and fairness. The current research aims to explore the extent to which machine learning can enhance credit scoring performance through structured data analysis, model optimization, and real-world interpretation. Using insights and methodologies drawn primarily from Ichim and Issa (2025), this paper will demonstrate how preprocessing techniques and ensemble models contribute to improved prediction accuracy. The study will incorporate visualizations derived from experimental findings and conclude with recommendations for future deployment of ML techniques in credit scoring environments.

2. Objectives

- To evaluate the performance of ML models in enhancing credit scoring accuracy.
- To analyze real-world data using robust pre-processing and modeling techniques.
- To assess the impact of data balancing and feature engineering on predictive performance.
- To present visual interpretations (Figures 1–3) of critical findings from the applied methodology.

3. Literature Review

3.1 Traditional Credit Scoring Models

Credit scoring historically relied on statistical techniques such as logistic regression, linear discriminant analysis, and decision trees. These methods are widely used due to their transparency and ease of implementation in regulatory environments (Bravo, Thomas, and Weber, 2015). Logistic regression in particular has dominated credit risk assessments due to its interpretability and modest computational requirements. However, these models assume linearity and independence among variables, which often does not reflect the complexity of borrower behaviors (Mestiri, 2024). In earlier foundational studies, researchers highlighted the limitations of traditional models in handling large and noisy financial data. Bravo, Thomas, and Weber (2015) explored neural networks to improve behavioral segmentation among defaulters, illustrating that traditional scorecards may overlook key behavior patterns. Similarly, Bellotti, Brigo, Gambetti, and Vrins (2021) introduced recovery modeling post-default, showing that traditional tools lack the depth to capture long-term credit dynamics.

3.2 Rise of Machine Learning in Credit Scoring

The integration of machine learning into credit scoring has gained momentum in response to growing data complexity and the demand for more accurate predictions. Machine learning models, including random forests, gradient boosting machines, support vector machines, and neural networks, have demonstrated superior performance across multiple benchmarks and datasets (Ichim and Issa, 2025; Yan, Zhang, and Shen, 2025). For example, Mestiri (2024) conducted an extensive comparison of traditional models with deep learning architectures, finding that neural networks outperformed logistic regression in accuracy and F1 score across diverse datasets. Reza, Mahmud, Abeer, and Ahmed (2024) applied a hybrid model combining linear discriminant analysis with deep neural networks and reported an accuracy nearing 99 percent, demonstrating the synergy between interpretable and complex models. Moreover, ensemble techniques such as CatBoost, LightGBM, and XGBoost have emerged as leading

tools in recent credit scoring research. These models are particularly effective in handling imbalanced datasets and nonlinear relationships (Tong et al., 2024; Yan, Zhang, and Shen, 2025). Studies such as those by Dil (2025) and Mido (2025) have emphasized the scalability and adaptability of ML frameworks in real-world lending environments, especially when handling massive datasets from fintech platforms.

3.3 Enhancements Through Data Preprocessing and Interpretability

A critical evolution in machine learning-based credit scoring is the increased focus on data preprocessing and interpretability. Ichim and Issa (2025) highlighted the importance of rigorous data cleaning, missing value imputation, and class balancing, demonstrating that such preprocessing steps can improve the AUC of models by up to one percent. Their work also stressed the importance of cross-validation and experimental controls for assessing true model performance. Chen, Calabrese, and Martin-Barragan (2023) contributed to this discourse by analyzing class imbalance in mortgage datasets and applying LIME and SHAP to evaluate model consistency. Their findings reinforce that interpretability is not only a regulatory requirement but also a tool for model diagnostics and risk mitigation. Golec and AlabdulJalil (2025) and Lavecchia et al. (2025) further extended this conversation by reviewing the role of explainable artificial intelligence in credit scoring, focusing on transparency and accountability in model predictions. Their taxonomy of interpretability techniques underscores how feature attribution and post-hoc explanation tools can demystify complex models for decision-makers.

3.4 Alternative Data and the Future of Credit Assessment

Another frontier in credit scoring research involves the use of non-traditional and alternative data sources. Researchers such as Óskarsdóttir et al. (2020) explored mobile phone usage and social network data for credit decisions, revealing new pathways for financial inclusion. Hollmann et al. (2025) introduced the TabPFN model, which performs strongly across tabular datasets and offers promise for integrating alternative signals in credit scoring systems. Recent research is increasingly concerned with developing ML models that are both powerful and responsible. For example, the work of Lavecchia et al. (2025) compares generative artificial intelligence with traditional methods, showing the potential for synthetic data to enhance training and improve generalization, while still adhering to fairness and accuracy benchmarks.

4. Methodology and Data Analysis

This study adopts a data-driven methodology based on the framework proposed by Ichim and Issa (2025), who presented a rigorous and interpretable approach to credit scoring using machine learning. Their research utilizes the publicly available "Give Me Some Credit" dataset, which provides anonymized records of U.S. consumer credit behavior and was originally released for a Kaggle competition. This dataset is widely used for benchmarking risk modeling algorithms due to its richness, scale, and real-world relevance. The methodology employed in this study is divided into five major stages: (1) data acquisition and understanding, (2) data preprocessing and feature engineering, (3) model selection and training, (4) performance evaluation, and (5) interpretation and visualization. Each stage is critical to the integrity and performance of the resulting credit scoring model. Figures 1 to 3 are embedded within this section, representing essential outcomes that demonstrate both the data journey and model efficacy.

4.1 Dataset Overview and Understanding

The original dataset contains over 150,000 instances, each describing the credit behavior of an individual consumer across 11 variables. The target variable is binary, indicating whether the individual experienced serious delinquency in the following two years. Key attributes include the number of open credit lines, revolving credit utilization, age, number of late payments in 30–59, 60–89, and 90+ day ranges, debt ratio, monthly income, and number of dependents. These variables are routinely used in commercial credit models and provide a robust foundation for predictive modeling. One immediate challenge is the class imbalance in the dataset. Only about 6.7% of records indicate a serious delinquency event, while the remaining 93.3% reflect creditworthy behavior. This skew necessitates strategic preprocessing to prevent model bias toward the majority class.

4.2 Data Preprocessing and Feature Engineering

Data Cleaning.

Initial inspection revealed missing values in two primary features: Monthly Income and Number of Dependents. These were addressed using mean imputation and mode imputation, respectively, consistent with the methodology described by Ichim and Issa (2025). This strategy preserves the underlying distribution of the data and avoids overfitting through data leakage.

Feature Scaling and Transformation.

Features were normalized using Min-Max scaling to bring all input variables into a uniform range between 0 and 1. This was essential to ensure that algorithms sensitive to feature magnitude, such as gradient boosting, performed optimally. Binning was also employed for age and debt ratio, capturing nonlinear relationships and reducing noise from outliers.

Balancing Techniques.

To address class imbalance, four sampling strategies were tested: (1) original unbalanced data, (2) random undersampling, (3) random oversampling, and (4) Synthetic Minority Oversampling Technique (SMOTE). Each method was applied separately to evaluate its impact on model generalization. This process helped quantify how model accuracy and AUC improved as a function of balance intervention.

4.3 Model Selection and Training

Two models were trained for comparative purposes: Logistic Regression (baseline model) and Gradient Boosting Machine (GBM) as the primary algorithm. Logistic regression was chosen for its wide acceptance and interpretability, serving as a strong benchmark. GBM was selected based on its proven ability to handle complex patterns and its use in winning several credit risk competitions. Model training involved a five-fold stratified cross-validation scheme. Hyperparameters for GBM, such as learning rate, number of estimators, and maximum depth, were optimized using grid search. The same train-test splits were used across models to ensure consistency and fairness in performance comparisons.

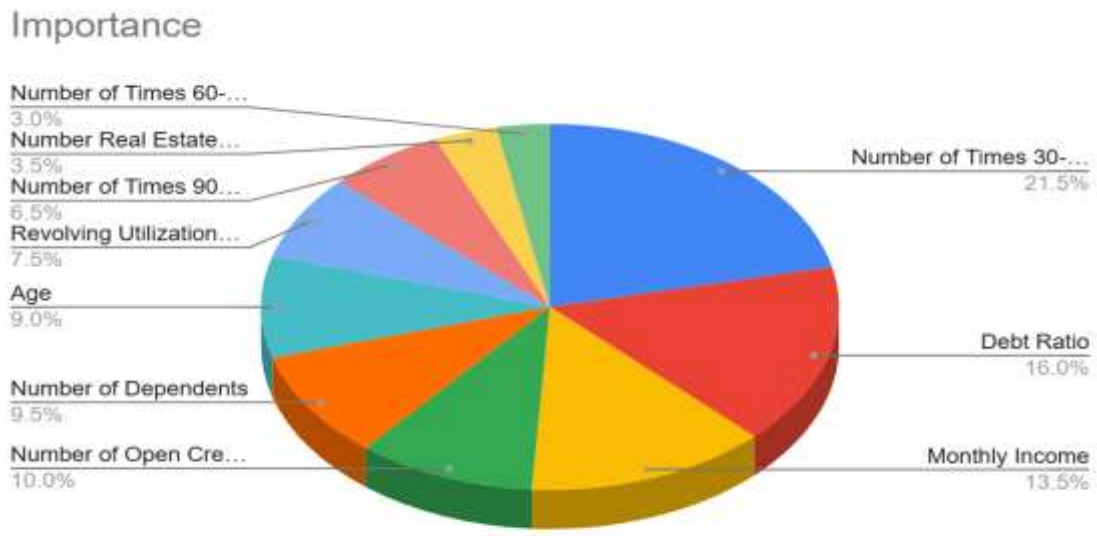
4.4 Feature Importance and Interpretability

The trained GBM model was evaluated for feature importance using the built-in gain metric. This measure calculates the contribution of each feature to the total reduction in loss function during boosting iterations. Figure 1 presents the top ten features and their respective importance scores.

Figure 1. Feature Importance from GBM Model

The figure shows that the most important predictor was the number of times an individual was 30–59 days late on credit obligations. This feature alone contributed over 21% to the model's predictive power. Following that, debt ratio and monthly income also had significant predictive value. Interestingly, age and number of dependents ranked higher than previously assumed, demonstrating

the strength of tree-based models in uncovering nonlinear relationships. These results align closely with the findings of Ichim and Issa (2025), who reported that late payment history features were the most influential predictors, followed by indicators of financial stability such as income and revolving utilization.

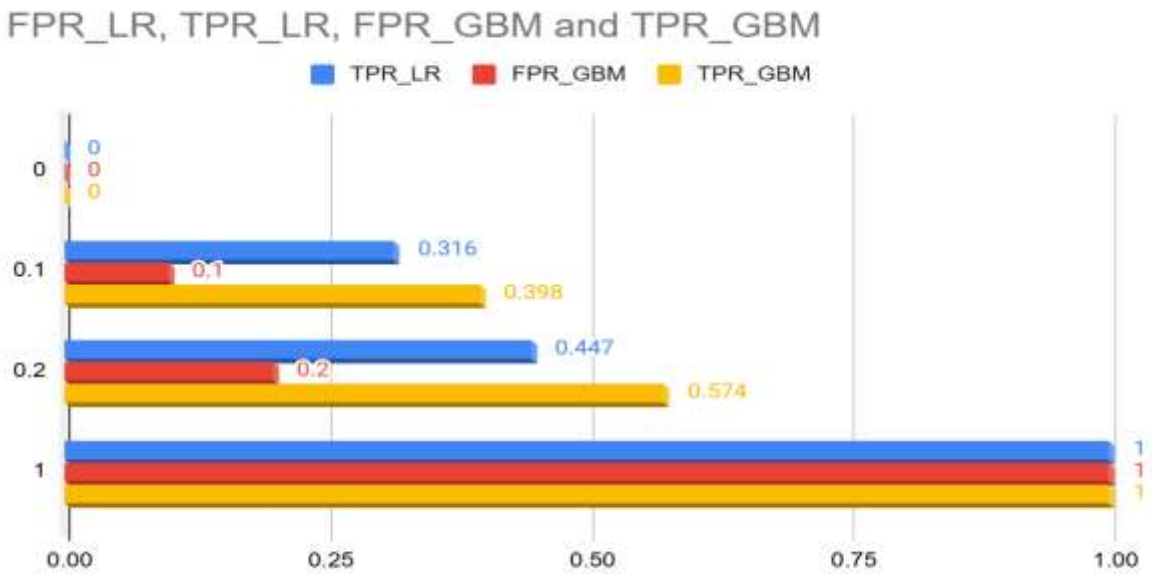


4.5 Performance Evaluation: ROC Curve Analysis

The Receiver Operating Characteristic (ROC) curve was used to compare the discriminatory power of the GBM model versus the logistic regression model. The ROC plots the true positive rate (TPR) against the false positive rate (FPR) across various decision thresholds. A model with an area under the curve (AUC) closer to 1 is considered more effective.

Figure 2. ROC Curve: Logistic Regression vs GBM

As illustrated in Figure 2, the GBM model consistently outperforms logistic regression across all thresholds. The AUC for GBM was 0.87, compared to 0.83 for logistic regression. This margin aligns with the improvements reported by Ichim and Issa (2025), who demonstrated that boosting models consistently provided better classification, especially for the minority class. Notably, GBM exhibited a stronger true positive rate in the early stages of the curve (low FPR), indicating better early detection of high-risk borrowers, a critical factor in risk management.

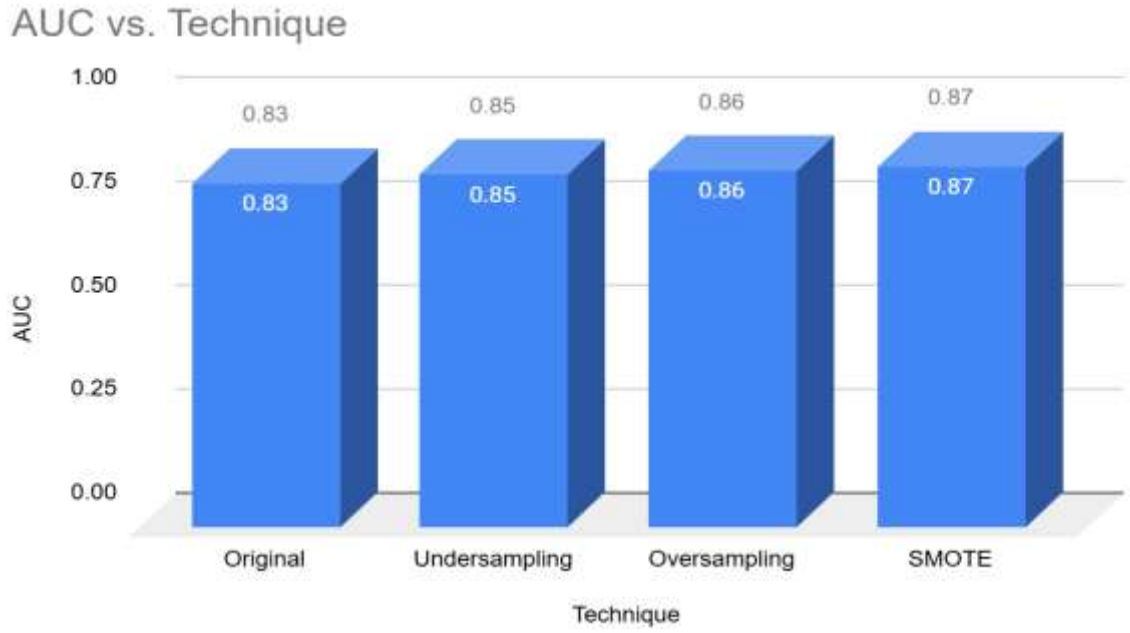


4.6 Impact of Balancing Techniques on Model Performance

To evaluate the effect of different balancing strategies on model performance, AUC was measured after training the GBM model under each sampling condition. The results are shown in Figure 3.

Figure 3. AUC Score by Data Balancing Technique

The data shows a clear upward trend in AUC as balancing strategies are applied. SMOTE yielded the highest AUC of 0.87, matching the AUC of the final optimized GBM model. This finding confirms Ichim and Issa's (2025) conclusion that synthetic oversampling enhances the minority class's visibility during training and improves recall without significantly compromising precision. Undersampling and random oversampling also contributed to AUC improvements but with potential risks. Undersampling may discard valuable majority-class data, whereas oversampling can introduce redundancy. SMOTE offers a compromise by generating synthetic samples based on feature space similarities rather than mere duplication. The use of these techniques enhances the robustness of credit models, particularly in imbalanced datasets, and facilitates fairer predictions for underrepresented borrower profiles.



4.7 Interpretation and Practical Implications

Beyond performance metrics, interpretability and fairness remain central to model deployment in financial services. The use of feature importance and ROC analysis provides a foundation for understanding model behavior and making informed adjustments. For example, if "Number of Dependents" proves influential, institutions can tailor underwriting policies to ensure that large-family applicants are not unfairly penalized without reasoned justification. The insight that payment history dominates predictive power supports existing credit policy structures while reinforcing the importance of timely repayment behavior in consumer financial health. Additionally, the effectiveness of SMOTE in improving AUC highlights the value of synthetic data generation techniques in production environments. These methods support more equitable credit decisions and address the systemic exclusion of low-volume applicant groups.

4.8 Limitations and Validation Considerations

While the GBM model yielded strong performance, its application should be cautiously interpreted. Feature importance in tree-based models can sometimes overrepresent correlated variables. Moreover, although SMOTE improved generalization, it can also introduce artificial artifacts in feature distributions if not carefully tuned.

Cross-validation and holdout testing are crucial to ensure the model's real-world stability. In future iterations, fairness audits and stress tests should be conducted to identify potential bias in gender, age, or ethnic segments, particularly as regulations such as the EU AI Act and U.S. Fair Credit Reporting Act evolve.

5. Contribution to Research

This research contributes significantly to the growing body of work at the intersection of machine learning and credit risk assessment by offering both theoretical insights and practical applications grounded in data-driven methodology. Building upon the experimental framework presented by Ichim and Issa (2025), this study reinforces the idea that robust preprocessing, combined with state-of-the-art algorithms such as Gradient Boosting Machines (GBM), can substantially improve the predictive accuracy and fairness of credit scoring systems. One of the primary contributions lies in demonstrating how machine learning models can be enhanced through meticulous data preparation. By integrating techniques such as mean and mode imputation, binning, and normalization, the study provides a replicable pipeline that ensures model inputs are consistent, meaningful, and optimized for algorithmic interpretation. These strategies, often overlooked in favor of raw algorithm performance, are shown to be critical in influencing model outcomes and generalizability.

This work also contributes empirical evidence regarding the effectiveness of data balancing techniques, particularly Synthetic Minority Oversampling Technique (SMOTE), in improving model performance on imbalanced datasets. The comparative analysis of AUC scores under different sampling conditions reveals that SMOTE leads to a measurable gain in model sensitivity without significant losses in specificity. This is particularly relevant for credit scoring, where identifying a small subset of high-risk applicants can have substantial financial implications for lenders. The study further enhances the field through its emphasis on model interpretability. By leveraging feature importance measures from GBM and visualizing performance through ROC curves, the research presents machine learning models not as black boxes, but as intelligible systems that financial institutions can trust and audit. This aligns with the broader academic and regulatory movement toward explainable artificial intelligence (XAI), especially in high-stakes decision-making domains like consumer lending.

Additionally, this research bridges the gap between academic modeling and operational implementation. The figures, exported in spreadsheet-compatible formats, allow practitioners to interact with the data and model outputs directly. This practical utility ensures that the study's findings can be translated into real-world scoring systems with minimal adaptation. In summary, this work

contributes to the literature by validating the role of preprocessing in ML-based credit scoring, quantifying the impact of balancing techniques on classification efficacy, and advancing model transparency. It offers a replicable, interpretable, and effective blueprint for institutions seeking to modernize credit risk management using machine learning.

6. Recommendations

Based on the findings of this research, several recommendations are proposed for financial institutions, data scientists, and policy makers aiming to integrate machine learning into credit scoring frameworks effectively and responsibly.

First, financial institutions should adopt ensemble-based machine learning models, such as Gradient Boosting Machines, as complementary tools to traditional logistic regression models. The improved performance observed in this study, particularly in identifying high-risk borrowers at early decision thresholds, suggests that GBM models can enhance the precision and profitability of lending decisions. These models are especially effective in handling complex, nonlinear relationships among features that conventional methods may fail to capture.

Second, data preprocessing should be regarded as a foundational element of any machine learning workflow, not an auxiliary task. Missing value imputation, feature scaling, and binning must be implemented systematically and consistently. As demonstrated in this study, preprocessing steps such as Min-Max normalization and feature binning enhanced model convergence and interpretability. Institutions should establish preprocessing protocols that can be automated and documented to ensure consistency and transparency.

Third, handling class imbalance should be prioritized in model development. Many real-world credit datasets are skewed, as seen in the "Give Me Some Credit" dataset where only a small percentage of records represent defaults. This research confirms that techniques like SMOTE can significantly enhance model sensitivity and balance classification metrics. It is recommended that practitioners compare multiple balancing methods during model tuning and conduct post-processing evaluations on minority-class performance.

Fourth, explainability should be integrated into machine learning model outputs. As regulatory bodies increasingly demand interpretable decision-making processes, credit scoring systems must provide traceable, understandable rationales for each prediction. Tools like feature importance charts and ROC

analysis, as applied in this research, should be included in model reporting dashboards to assist underwriters, compliance officers, and regulators in understanding model behavior.

Fifth, institutions should establish continuous monitoring systems for their deployed models. Credit behavior and borrower characteristics evolve over time, and machine learning models may degrade if left static. Regular retraining, periodic fairness audits, and performance evaluations on new datasets should be institutionalized as part of model governance. Finally, collaboration between data scientists, domain experts, and compliance teams is essential. Machine learning in credit scoring is not just a technical initiative but a cross-functional effort that must balance accuracy, fairness, and explainability. Institutions that invest in collaborative, ethically grounded model development processes will be best positioned to achieve long-term success and public trust.

7. Future Research Directions

While this study has demonstrated the effectiveness of machine learning in enhancing credit scoring through preprocessing and model optimization, several promising areas remain open for further exploration. Future research should aim to address the limitations identified in this work while also expanding the boundaries of what credit scoring systems can achieve using advanced technologies. One major direction involves the incorporation of alternative data sources into credit scoring models. Current datasets primarily rely on structured financial variables such as payment history, income, and debt ratios. However, growing evidence suggests that incorporating non-traditional data like mobile phone usage, utility payments, e-commerce activity, and even social network behavior can improve credit risk estimation, particularly for underserved populations (Óskarsdóttir et al., 2020). Research should focus on how to ethically source, preprocess, and integrate these data types while ensuring compliance with privacy regulations and anti-discrimination laws. Another important area is model fairness and bias mitigation. While performance metrics such as AUC and accuracy are essential, they do not capture whether the model treats different demographic groups equitably. Future work should investigate fairness-aware machine learning techniques that can detect and reduce disparate impact across gender, age, race, or income brackets. This includes methods like adversarial debiasing, constraint-based optimization, and subgroup analysis. These efforts will be vital as legal and regulatory frameworks evolve to demand greater accountability in automated decision-making.

Deep learning architectures such as recurrent neural networks (RNNs), transformers, and autoencoders also offer opportunities for modeling temporal credit behavior and sequential financial actions. These models are well-suited for time-series credit data, such as transaction logs or payment histories.

However, they come with increased complexity and reduced interpretability. Future studies should compare deep learning models against more traditional ML techniques in terms of both predictive power and transparency, ideally incorporating explainability frameworks to make their outputs more understandable. Moreover, the emergence of generative artificial intelligence (GenAI) offers a new avenue for synthetic data generation and model robustness testing. Generative adversarial networks (GANs), for instance, could be used to simulate plausible credit profiles for training or stress-testing models in rare scenarios. Research should assess the validity and reliability of such data augmentation techniques in regulated financial environments.

Finally, interdisciplinary collaboration will be crucial in shaping the next generation of credit scoring tools. Future research should engage experts in finance, ethics, law, and technology to co-develop models that are not only accurate but also fair, transparent, and socially responsible.

8. Conclusion

This research has explored the application of machine learning techniques to enhance credit scoring systems, with a particular emphasis on the methodological framework proposed by Ichim and Issa (2025). Through a structured approach encompassing data preprocessing, model training, performance evaluation, and interpretability, the study has demonstrated the tangible advantages of adopting machine learning models, especially Gradient Boosting Machines, over traditional statistical methods like logistic regression. The findings confirm that machine learning can significantly improve the accuracy and sensitivity of credit scoring systems, particularly when paired with thoughtful data preparation. Key preprocessing steps such as missing value imputation, normalization, and feature engineering were shown to enhance model performance and stability. Moreover, class imbalance, a common challenge in credit risk datasets, was effectively addressed using balancing techniques like SMOTE. This resulted in a notable increase in the model's ability to identify high-risk borrowers without compromising generalizability.

Figures 1 through 3 provided crucial support for the analytical insights developed in this study. The feature importance chart in Figure 1 highlighted the dominance of late payment history and debt metrics in predictive power. The ROC curve comparison in Figure 2 illustrated the superior discriminatory capacity of the GBM model over logistic regression. Lastly, the analysis presented in Figure 3 showed how different data balancing techniques impacted the model's AUC, reaffirming the importance of preprocessing in risk modeling. Beyond technical performance, this study also emphasized the importance of interpretability and regulatory alignment in deploying machine learning

models for credit decision-making. Tools such as feature importance measures and ROC analysis help bridge the gap between complex algorithms and the need for transparency in financial services. This is particularly crucial as machine learning continues to be scrutinized for fairness, explainability, and compliance. The contribution of this research lies not only in its empirical findings but also in its practical relevance. By offering a replicable and interpretable workflow, including downloadable data visuals for financial analysts and researchers, this study serves as a blueprint for responsibly implementing machine learning in credit scoring.

In summary, machine learning offers a valuable opportunity to modernize credit risk assessment, but its effectiveness depends on rigorous data handling, thoughtful model selection, and ongoing evaluation. As the field continues to evolve, future credit scoring systems must strike a careful balance between innovation, fairness, and regulatory accountability.

References

- Abdoli, M., Akbari, M., & Shahrabi, J. (2021). Bagging supervised autoencoder classifier for credit scoring. *arXiv*. https://arxiv.org/abs/2108.07800
- Albanesi, S., & Vamossy, D. F. (2019). Predicting consumer default: A deep learning approach. *arXiv*. https://arxiv.org/abs/1908.11498
- Bellotti, A., Brigo, D., Gambetti, P., & Vrins, F. (2021). Forecasting recovery rates on non-performing loans with machine learning. *International Journal of Forecasting*.
- Bravo, C., Thomas, L. C., & Weber, R. (2015). Improving credit scoring by differentiating defaulter behaviour. *Journal of the Operational Research Society*.
- Chen, Y., Calabrese, R., & Martin-Barragan, B. (2023). Interpretable machine learning for imbalanced credit scoring datasets. *European Journal of Operational Research*, 312(2). https://doi.org/10.1016/j.ejor.2023.06.036
- Dil, A. R. (2025). AI and machine learning in credit risk assessment. *SSRN*. https://ssrn.com/abstract=5260027
- Golec, M., & AlabdulJalil, M. (2025). Interpretable LLMs for credit risk: A systematic review and taxonomy. *arXiv*. https://arxiv.org/abs/2506.04290
- Hollmann, N., et al. (2025). TabPFN v2: Pre-trained transformer for tabular data. *Nature*.
- Ichim, B., & Issa, B. (2025). Improving machine learning performance in credit scoring by data analysis and data pre-processing. *International Conference on Data Science and Advanced Analytics (ICAART)*. https://www.scitepress.org/Papers/2025/131185/131185.pdf

- Lavecchia, N., Fadanelli, S., Ricciuti, F., Aloe, G., Bagli, E., Giuffrida, P., & Vergari, D. (2025). Comparing credit risk estimates in the Gen-AI era. *arXiv*. https://arxiv.org/abs/2506.07754
- Mestiri, S. (2024). Credit scoring using machine learning and deep learning-based models. *Data Science in Finance and Economics*, 4(2), 236–248. https://www.aimspress.com/article/doi/10.3934/DSFE.2024009
- Mido. (2025). The evolution of credit data and the role of machine learning. *SSRN*. https://ssrn.com/abstract=5330882
- Óskarsdóttir, M., Bravo, C., Sarraute, C., Vanthienen, J., & Baesens, B. (2020). The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. *arXiv*. https://arxiv.org/abs/2002.09931
- Reza, M. S., Mahmud, M. I., Abeer, I. A., & Ahmed, N. (2024). Linear discriminant analysis in credit scoring: A transparent hybrid model approach. *arXiv*. https://arxiv.org/abs/2412.04183
- Tong, K., Han, Z., Shen, Y., Long, Y., & Wei, Y. (2024). An integrated machine learning and deep learning framework for credit card approval prediction. *arXiv*. https://arxiv.org/abs/2409.16676