



Original Article

Comparative Analysis of Machine Learning Models for Loan Approval: Accuracy, Bias, and Regulatory Implications

Manjiri Abhay Bilurkar

ATSS CBSCA College, Chinchwad, Pune, Computer Science Department

Manuscript ID:
IBMIRJ -2026-030104

Submitted: 06 Dec. 2025

Revised: 10 Dec. 2025

Accepted: 05 Jan. 2026

Published: 31 Jan. 2026

ISSN: 3065-7857

Volume-3

Issue-1

Pp. 15-18

January 2026

Correspondence Address:
Manjiri Abhay Bilurkar
ATSS CBSCA College, Chinchwad,
Pune, Computer Science Department
Email:manjiri.bilurkar@atsscollege.edu.in



Quick Response Code:



Web: <https://ibrj.us>



DOI: 10.5281/zenodo.18918126

DOI Link:

<https://doi.org/10.5281/zenodo.18918126>



Creative Commons

Abstract

Machine learning (ML) models have become a central component of modern loan approval systems as financial institutions seek to improve prediction accuracy, automate decision-making, and reduce operational costs compared to traditional statistical approaches. By leveraging large and complex datasets, ML techniques can capture nonlinear relationships among borrower attributes that conventional models often fail to identify. However, the increased adoption of these models has raised significant ethical and regulatory concerns, particularly regarding algorithmic bias, fairness, transparency, and accountability in credit decisions. This paper provides a comparative analysis of three widely used ML models—logistic regression, random forests, and gradient boosting—with a focus on their predictive performance, susceptibility to bias, and implications for regulatory compliance. Using evidence from publicly available loan datasets and existing empirical studies, the analysis demonstrates that more complex ensemble-based models generally outperform simpler models in terms of accuracy and risk prediction. Nevertheless, these performance gains are frequently accompanied by reduced interpretability and a higher likelihood of disparate outcomes across protected demographic groups. The findings emphasize that accuracy alone is insufficient for responsible credit decision-making. To ensure ethical integrity and compliance with financial regulations, the integration of fairness-aware learning methods and explainable artificial intelligence (XAI) techniques is essential when deploying ML-based loan approval systems.

Keywords: machine learning, loan approval, credit scoring, algorithmic bias, fairness, explainable AI, financial regulation

Introduction

Loan approval represents one of the most consequential decision-making processes in the financial sector, shaping access to credit for individuals and businesses and influencing broader patterns of economic inclusion and stability. Historically, credit institutions have relied on manual underwriting procedures and relatively simple statistical models, most notably logistic regression, to assess borrower creditworthiness (Hand & Henley, 1997). These traditional approaches are valued for their transparency and ease of interpretation; however, they are often limited in their ability to model complex, nonlinear relationships and interactions within increasingly large and heterogeneous financial datasets. Recent advances in machine learning (ML) have transformed credit risk assessment by enabling data-driven models that can process high-dimensional data and uncover subtle patterns associated with default risk and repayment behavior (Zhou, 2012). As a result, ML-based loan approval systems have demonstrated improved predictive accuracy and operational efficiency compared to conventional methods. Despite these advantages, the growing reliance on complex ML models has introduced new challenges related to algorithmic fairness, transparency, and accountability. Because these models learn from historical data, they may inadvertently replicate or intensify existing social and economic biases, potentially leading to discriminatory lending outcomes for protected groups (Barocas & Selbst, 2016). In parallel, regulatory frameworks such as the U.S. Equal Credit Opportunity Act (ECOA) require that credit decisions be nondiscriminatory and justifiable, placing strict constraints on the use of opaque or poorly interpretable models (Lambrecht & Tucker, 2020). Against this background, this paper synthesizes existing empirical evidence to compare the predictive performance of widely used ML models in loan approval, assess their fairness implications, and examine the regulatory considerations associated with their deployment in highly regulated financial environments.

Creative Commons (CC BY-NC-SA 4.0)

This is an open access journal, and articles are distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License](https://creativecommons.org/licenses/by-nc-sa/4.0/), which allows others to remix, tweak, and build upon the work noncommercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

How to cite this article:

Bilurkar, M. A. (2026). Comparative Analysis of Machine Learning Models for Loan Approval: Accuracy, Bias, and Regulatory Implications. *InSight Bulletin: A Multidisciplinary Interlink International Research Journal*, 3(1), 15–18. <https://doi.org/10.5281/zenodo.18918126>

Literature Review

Machine Learning Models in Credit Decisioning

Machine learning techniques have been extensively explored in the domain of credit decisioning due to their capacity to process large-scale financial data and identify complex, nonlinear relationships among borrower characteristics (Lessmann et al., 2015). Early credit scoring systems primarily relied on statistical models, with logistic regression emerging as the dominant approach because of its mathematical simplicity, stability, and clear interpretability. The transparency of logistic regression allows lenders and regulators to directly assess the influence of individual variables on credit outcomes, making it particularly suitable for regulated financial environments (Hosmer et al., 2013). However, its linear structure limits its effectiveness when borrower behavior is influenced by intricate interactions and nonlinear patterns. To address these limitations, ensemble-based machine learning models have gained increasing attention in credit risk assessment. Random forest algorithms combine the outputs of multiple decision trees generated from bootstrapped samples of the data, thereby reducing overfitting and enhancing predictive robustness (Breiman, 2001). This ensemble strategy enables random forests to capture complex feature interactions while maintaining relatively stable performance across diverse datasets. Similarly, gradient boosting machines (GBMs) employ an iterative learning process in which successive models focus on correcting the errors of previous ones, leading to highly accurate predictions, particularly in high-dimensional settings (Friedman, 2001). Despite their demonstrated performance advantages, both random forests and GBMs present challenges related to model transparency and interpretability. Their decision-making processes are not easily traceable, which complicates explanation and validation in credit approval contexts. Consequently, while these models offer superior predictive capability, their adoption in financial institutions necessitates additional governance mechanisms, including explainability tools and rigorous validation frameworks, to ensure compliance with regulatory and ethical standards.

Algorithmic Bias and Fairness Metrics

A growing body of research highlights that ML models may reproduce or amplify societal biases present in historical data, leading to disparate outcomes across demographic groups (Barocas & Selbst, 2016). Fairness in machine learning encompasses various metrics, such as demographic parity, equal opportunity, and disparate impact, each reflecting different notions of fairness (Hard et al., 2016; Feldman et al., 2015). Demographic parity requires equal acceptance rates across groups regardless of actual default risk, a stringent requirement that may not always align with business goals (Hardt et al., 2016). Equal opportunity focuses on equal true positive rates across groups, balancing fairness with predictive relevance (Hardt et al., 2016). Disparate impact measures the ratio of favorable outcomes between groups and is often applied in legal contexts to assess discrimination (Feldman et al., 2015).

Explainability and Regulatory Compliance

As models grow in complexity, interpretability becomes harder to achieve without specialized techniques. Explainable AI (XAI) methods such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) provide tools to interpret complex models by quantifying feature contributions (Ribeiro et al., 2016; Lundberg & Lee, 2017). These techniques are especially valuable in regulated sectors where lenders must justify adverse credit decisions to consumers and regulators (Doshi-Velez & Kim, 2017). Regulatory frameworks like ECOA in the United States and similar anti-discrimination laws internationally require that credit decisions be free from bias and that explanations be provided to affected individuals (Lambrecht & Tucker, 2020). Consequently, lenders must balance predictive gains from complex models with requirements for fairness and explainability.

Methodology

This study employs a secondary-data research approach, synthesizing results from publicly available loan datasets and peer-reviewed comparisons of ML models in loan approval contexts. Key public datasets include Lending Club and German Credit, which have been widely used in empirical studies to benchmark ML performance in credit risk (Kaggle, n.d.; Dua & Graff, 2019). Performance metrics examined include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC). Fairness evaluations rely on reported demographic parity differences, equal opportunity gaps, and disparate impact ratios where available. No primary model training was conducted; instead, the focus is on evidence reported in the literature.

Comparative Analysis of Modeling Approaches in Credit Decisioning

Predictive Accuracy

A substantial body of empirical research demonstrates that ensemble learning methods, particularly random forests and gradient boosting machines (GBMs), outperform traditional statistical models such as logistic regression in predicting binary credit outcomes. These gains stem from the ability of ensemble models to capture complex, non-linear relationships and high-order feature interactions that are difficult to model explicitly using linear techniques (Lessmann et al., 2015). Gradient boosting methods, in particular, iteratively refine predictive performance by focusing on misclassified observations, allowing them to achieve superior discrimination power. Empirical comparisons frequently report higher Area Under the Receiver Operating Characteristic Curve (AUC-ROC) scores for GBMs relative to both logistic regression and random forests, especially in large, feature-rich credit datasets (Brown & Mues, 2012). Higher predictive accuracy can translate into tangible business benefits, such as improved risk differentiation, reduced default rates, and more efficient capital allocation. However, these accuracy improvements should be interpreted with caution. Credit datasets are often noisy, imbalanced, and subject to temporal drift. As a result, gains in out-of-sample performance may not always persist in operational settings without robust validation and monitoring. Moreover, marginal improvements in predictive metrics may come at the cost of increased model complexity and governance burdens, raising questions about their overall value in regulated financial environments.

Fairness and Bias Risks

Despite their predictive advantages, complex machine learning models present heightened risks related to fairness and discrimination. In their unconstrained form, ensemble methods can exacerbate disparate impact across protected groups, even when sensitive attributes such as race or gender are excluded from the training data. This occurs because non-linear models can exploit proxy variables—such as ZIP code, employment history, or educational background—that are correlated with protected characteristics (Zliobaite, 2017). Logistic regression models, by contrast, are often considered more amenable to fairness interventions due to their transparent structure and limited functional form. Regularization techniques, constrained optimization, and explicit exclusion or transformation of sensitive features can be more easily applied and interpreted in linear models. Furthermore, the marginal effects of individual variables in logistic regression can be directly inspected for potential bias, facilitating compliance reviews and internal audits. Random forests and GBMs, due to their depth, feature interactions, and ensemble aggregation, make it more difficult to identify and control discriminatory mechanisms without explicit fairness-aware techniques. As a result, their deployment typically necessitates additional bias mitigation strategies, such as pre-processing (e.g., reweighting or resampling), in-processing (e.g., fairness-constrained learning), or post-processing adjustments to decision thresholds. Without such safeguards, the use of these models may increase legal and ethical risks for financial institutions.

Explain ability Challenges

Explain ability is a central requirement in credit decisioning, where adverse actions must be justified to consumers and regulators. Logistic regression has long been favored in this domain because its coefficients offer a direct and intuitive interpretation of feature influence on predicted outcomes. This transparency enables institutions to generate clear, human-readable explanations for loan denials and to document decision logic for regulatory examinations. In contrast, ensemble methods are often described as “black-box” models, as their internal decision processes are not readily interpretable. To address this limitation, post-hoc explain ability techniques such as Shapley Additive explanations (SHAP) have been developed. SHAP values provide both global explanations—highlighting the overall importance of features across the dataset—and local explanations that clarify why a specific individual received a particular prediction (Lundberg & Lee, 2017). While these tools significantly enhance interpretability, they also introduce additional complexity into model development and governance. Explain ability frameworks must be validated, maintained, and communicated effectively to non-technical stakeholders. Moreover, post-hoc explanations do not alter the underlying model behavior and may not fully capture causal relationships, which can limit their usefulness in fairness assessments and regulatory disputes.

Ethical and Regulatory Implications

The use of machine learning in credit decisioning operates at the intersection of technological innovation, ethical responsibility, and legal compliance. Algorithmic systems that systematically disadvantage certain demographic groups may violate anti-discrimination laws, even if they optimize predictive accuracy or profitability (Barocas & Selbst, 2016). This creates a fundamental tension between efficiency and equity that must be carefully managed. Regulatory bodies increasingly expect financial institutions to adopt proactive approaches to bias detection and mitigation throughout the model lifecycle. These expectations include routine fairness testing, documentation of model design choices, and the implementation of fairness-aware learning algorithms or post-processing corrections where disparities are identified (Corbett-Davies & Goel, 2018). Institutions are also expected to demonstrate ongoing monitoring to ensure that deployed models remain compliant as data distributions and economic conditions evolve. Explainability requirements further complicate the adoption of opaque models. Consumer protection regulations mandate that borrowers receive understandable and actionable reasons for adverse credit decisions. Without robust explainable AI (XAI) mechanisms, complex machine learning models may struggle to meet these standards, exposing institutions to reputational damage and regulatory sanctions.

Discussion

The findings of this review highlight the trade-offs inherent in applying advanced machine learning techniques to loan approval systems. While ensemble models offer measurable improvements in predictive performance, they introduce challenges related to fairness, transparency, and regulatory compliance. Logistic regression remains a viable and often preferred option in many credit contexts due to its interpretability, stability, and alignment with long-standing regulatory practices. Nevertheless, dismissing advanced models entirely may result in missed opportunities for improved risk management and financial inclusion. When deployed responsibly, ensemble methods can enhance decision quality, provided they are accompanied by rigorous fairness constraints, explain ability tools, and governance frameworks. A hybrid or tiered modeling strategy—where complex models inform decisions while simpler models support explanation and oversight—may offer a pragmatic compromise.

Conclusion

This paper examined the trade-offs associated with applying machine learning models to loan approval decisions, focusing on predictive accuracy, fairness, and interpretability. Although complex ensemble models such as random forests and GBMs demonstrate superior predictive performance, they pose significant ethical and regulatory challenges related to bias and explainability. Logistic regression continues to play a critical role due to its transparency and ease of alignment with fairness requirements. To ensure responsible deployment, financial institutions should integrate fairness-enhanced machine learning practices and robust explainability techniques into the entire model lifecycle. Future research should empirically evaluate the effectiveness of fairness interventions in real-world credit systems and explore automated mechanisms for continuous auditing and monitoring of machine learning models in production environments.

Acknowledgment

I would like to express my sincere gratitude to the faculty members of the Computer Science Department, ATSS CBSCA College, Chinchwad, Pune, for their valuable guidance, encouragement, and academic support throughout the completion of this research work. Their insights and constructive feedback greatly contributed to shaping the direction and quality of this study.

Financial support and sponsorship

Nil.

Conflicts of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732.
2. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
3. Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39(3), 3446–3453.
4. Corbett-Davies, S., & Goel, S. (2018). The measure and mis measure of fairness. *arXiv*. <https://arxiv.org/abs/1808.00023>
5. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *ArXiv*. <https://arxiv.org/abs/1702.08608>
6. Dua, D., & Graff, C. (2019). UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml>
7. Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkata subramanian, S. (2015). Certifying and removing disparate impact. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
8. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
9. Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*, 29, 3325–3333.
10. Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society: Series A*, 160(3), 523–541.
11. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. Wiley.
12. Kaggle. (n.d.). Lending Club Loan Data. <https://www.kaggle.com>
13. Lambrecht, A., & Tucker, C. (2020). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management Science*, 66(7), 2946–2969.
14. Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124–136.
15. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
16. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.
17. Zliobaite, I. (2017). Measuring discrimination in algorithmic decision making. *Data Mining and Knowledge Discovery*, 31(4), 1060–1089.
18. Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*. Chapman & Hall.