European Journal of AI, Computing & Informatics

Vol. 2 No. 1 2026

Article Open Access

Leveraging Ensemble Machine Learning for Credit Risk Assessment in Underserved U.S. Small Businesses

Zaolin Zhang [1,*]

[1] City University of New York, New York, New York, 10022, USA
* Correspondence: Zaolin Zhang, City University of New York, New York, New York, 10022, USA

Received: 24 November 2025
Revised: 29 December 2025
Accepted: 08 January 2026
Published: 12 January 2026

Abstract: This study investigates the role of ensemble machine learning techniques in improving credit risk assessment for underserved small businesses in the United States. Conventional credit evaluation systems rely heavily on formal financial documentation and collateral requirements, excluding a large proportion of early-stage, minority-owned, or cash-based enterprises. Drawing on the uploaded document's empirical modeling structure, the study integrates gradient boosting frameworks, random forest classifiers, and macro-behavioral feature engineering to construct a multidimensional risk assessment model. The methodology incorporates structured preprocessing, cross-validated ensemble training, and interpretability analysis using SHAP values. Results show that boosting-based models consistently outperform traditional approaches, achieving stronger precision, recall, and AUC scores while capturing nuanced behavioral and macroeconomic interactions. The findings highlight the pot ential of ensemble learning to expand access to credit, reduce misclassification bias, and support inclusive economic development. The study concludes with recommendations for lenders, policymakers, and researchers regarding the deployment of ensemble analytics and the future integration of fairness-aware modeling.

Keywords: ensemble learning; credit risk; underserved businesses; boosting models; financial inclusion; machine learning interpretability

1. Introduction

Access to credit is foundational to the growth and resilience of small businesses in the United States. Yet historically underserved firms-including minority-owned businesses, immigrant-owned enterprises, micro-firms with limited documentation, and cash-dominant operations-face disproportionately high rejection rates from traditional lending institutions. Conventional credit scoring frameworks depend primarily on standardized financial statements, collateral availability, and established credit histories, metrics many underserved firms lack [1]. As a result, structural barriers persist, limiting entrepreneurial mobility and widening economic inequality. With increasing digitalization of financial transactions, machine learning offers new possibilities for deriving risk insights from alternative sources such as behavioral patterns, invoice flows, customer engagement metrics, and macroeconomic signals. The document provided outlines a robust technical foundation for machine-learning-driven credit modeling, emphasizing ensemble methods and interpretability tools. This study expands on that foundation to develop a comprehensive credit risk assessment framework designed specifically for underserved U.S. small businesses.

The article is organized into five major sections. Section I analyzes systemic lending barriers and the limitations of traditional credit models [2]. Section II presents the ensemble modeling framework, detailing preprocessing, model construction, and evaluation strategies. Section III reports empirical findings and interpretability results. Section IV discusses implications for lenders, financial inclusion, and regulatory considerations. Section V proposes forward-looking strategies and opportunities for future research.

2. Structural Barriers and Limitations of Traditional Credit Risk Models

Underserved small businesses frequently encounter structural limitations within the conventional credit ecosystem that constrain their access to financing. Legacy credit evaluation systems are designed to reward characteristics such as long operating histories, stable revenue documentation, and the availability of tangible collateral. However, these attributes are less prevalent among micro-enterprises, early-stage ventures, and minority-owned businesses, which often operate with limited capitalization and shorter track records. As a result, traditional credit models systematically disadvantage precisely those firms that rely most heavily on external financing to support growth and resilience.

Conventional underwriting frameworks rely heavily on historical credit scores, audited financial statements, and tax returns as primary indicators of borrower reliability. While these metrics may reflect long-term financial stability, they fail to capture the actual repayment capacity of businesses operating in nontraditional or constrained environments [3]. Seasonal enterprises, cash-based operations, and firms embedded in informal economic networks are often classified as high-risk despite demonstrating consistent cash flow or strong customer demand. This misclassification stems from a narrow definition of creditworthiness that prioritizes formal documentation over operational performance.

The mismatch between traditional credit metrics and real-world business dynamics is further exacerbated by structural inequalities embedded in financial systems. Many underserved borrowers lack longstanding relationships with financial institutions, limiting their access to tailored credit products or advisory services. Geographic disparities in banking infrastructure also play a significant role, as businesses located in rural areas or historically marginalized urban neighborhoods may face reduced access to branches, credit officers, or alternative financing channels [4]. Additionally, historical lending biases-whether implicit or explicit-have contributed to persistent gaps in credit availability for minority-owned and women-owned enterprises. As highlighted in the document, underserved borrowers frequently lack standardized datasets that can be readily processed by rule-based scoring systems, leading to systematic overestimation of default risk and prolonged credit exclusion.

Traditional credit models also exhibit limited capacity to incorporate high-frequency behavioral data or contextual indicators that are increasingly relevant in modern business environments. Metrics such as customer retention rates, invoice settlement behavior, supplier-payment regularity, and digital engagement patterns provide real-time signals of business health and financial discipline [5]. However, these indicators are often excluded from conventional analyses due to rigid modeling assumptions, data integration challenges, or regulatory inertia. The omission of such information reduces predictive accuracy and creates blind spots in risk estimation, particularly for businesses whose financial stability is better reflected in operational behavior than in historical financial statements.

Moreover, static underwriting criteria constrain the ability of traditional models to adapt to changing economic conditions. Credit risk is treated as a fixed classification rather than a dynamic process that evolves over time. When macroeconomic shocks-such as supply chain disruptions, inflationary pressures, or sudden demand fluctuations-affect liquidity cycles, legacy models cannot update risk assessments in a timely manner. This lack of responsiveness leads to misaligned risk classification, either penalizing businesses that remain resilient or failing to detect emerging vulnerabilities. For small enterprises operating in volatile or rapidly changing markets, such rigidity can have severe consequences, including reduced access to working capital precisely when it is most needed.

These structural limitations underscore the need for more flexible analytical frameworks capable of integrating multidimensional data, learning nonlinear relationships, and producing transparent yet adaptive risk assessments. Ensemble machine learning approaches offer a compelling alternative to traditional credit scoring by combining multiple models to capture diverse patterns within complex datasets. By integrating structured financial data with behavioral, transactional, and contextual information, ensemble models can generate more nuanced and empirically grounded risk scores. Importantly, these systems can be designed with interpretability mechanisms that explain key drivers of risk, supporting regulatory compliance and borrower trust.

In this context, ensemble machine learning does not merely improve predictive performance but also contributes to greater equity in credit allocation. By recognizing alternative indicators of creditworthiness and dynamically updating risk assessments, these models help correct systemic biases embedded in legacy systems. For underserved small businesses, such advancements represent a pathway toward fairer access to capital, improved financial inclusion, and more sustainable participation in the formal credit economy.

3. Ensemble Machine Learning Framework for Credit Risk Assessment

The ensemble modeling framework is structured around three core components: feature engineering, model construction, and evaluation strategy. Together, these elements form a layered and highly adaptive credit risk assessment system designed to address the limitations of traditional underwriting models, particularly for underserved small businesses. Each component contributes to the system's ability to capture complex borrower behavior, respond to dynamic economic conditions, and deliver reliable, interpretable risk estimates.

Feature engineering serves as the foundation of the framework by transforming raw data into meaningful representations of borrower risk. Building on the methodology outlined in the document, the feature set integrates transactional records, behavioral indicators, and macroeconomic variables to construct a multidimensional risk profile. Core features include repayment history, invoice issuance and settlement regularity, customer concentration ratios, and operational cash-flow proxies derived from transaction timing and volume. These firm-level indicators are complemented by sector-specific variables that account for industry dynamics and region-level economic measures such as unemployment rate fluctuations, interest-rate movements, and inflation-adjusted consumption trends. By incorporating both micro- and macro-level signals, the model captures not only borrower-specific performance but also the broader economic context in which repayment behavior occurs.

To improve data quality and robustness, several preprocessing techniques are applied during feature engineering. Temporal smoothing reduces noise and volatility caused by irregular cash cycles or seasonal revenue patterns, which are common among small and micro-enterprises. Normalization and scaling procedures enhance comparability across heterogeneous firms with varying sizes and transaction volumes, preventing larger entities from disproportionately influencing model outcomes. Categorical encoding techniques, including target and one-hot encoding, represent industry classifications, business age segments, and geographic regions in a way that preserves structural information relevant to repayment behavior. Additionally, interaction features are explicitly engineered to capture context-dependent effects, such as the relationship between cash-flow stability and sector cyclicality or the interaction between customer concentration and regional economic stress. This expanded and carefully curated feature space enables a more comprehensive characterization of borrower risk than is possible with rule-based scoring systems or single-variable heuristics.

Model construction constitutes the second pillar of the framework and centers on ensemble learning techniques, including XGBoost, LightGBM, and Random Forest. These algorithms are selected for their proven ability to model nonlinear relationships, accommodate mixed data types, and maintain strong predictive performance in the presence of missing or imperfect data-conditions frequently encountered when assessing underserved borrowers. As described in the uploaded document, the training pipeline incorporates grid-search hyperparameter optimization to systematically explore model configurations, k-fold cross-validation to ensure stability across data partitions, and early-stopping mechanisms to prevent overfitting while preserving generalization accuracy. Boosting-based models iteratively correct residual errors from previous learners, enhancing discrimination in borderline cases where traditional models struggle. In contrast, Random Forest improves robustness by aggregating predictions across diverse decision trees, reducing variance and improving overall stability. A unified ensemble comparison framework enables systematic benchmarking across models, ensuring that the final selection aligns with the behavioral characteristics and data realities of underserved small enterprises.

The evaluation strategy forms the final component of the framework and emphasizes metrics that reflect both predictive accuracy and real-world lending impact. Precision-recall analysis, confusion matrices, and AUC-ROC curves are used to assess aggregate model performance and borrower-level misclassification patterns. For underserved small businesses, false negatives-creditworthy applicants incorrectly classified as high-risk-carry substantial economic consequences by restricting access to affordable financing. Accordingly, recall is prioritized alongside precision to balance risk management with financial inclusion objectives. To complement quantitative metrics, SHAP-based interpretability techniques are employed to decompose model predictions and reveal the marginal contributions of individual features. These explanations illuminate the decision pathways underlying each risk assessment, enhancing transparency and trust. Interpretability is essential for responsible lending practices, supporting regulatory compliance and enabling financial institutions to justify automated credit decisions while ensuring fairness and accountability across diverse borrower populations.

4. Results and Analytical Insights

Empirical results indicate that ensemble learning methods, particularly XGBoost, consistently demonstrate superior performance across key evaluation metrics when compared with traditional and baseline models. Boosting-based algorithms achieve higher precision and recall than both logistic regression and Random Forest, suggesting a stronger ability to capture complex, nonlinear behavioral patterns that are often oversimplified by linear or tree-averaging approaches. Logistic regression, while interpretable and computationally efficient, relies on restrictive assumptions regarding linear separability and independent feature effects, which limit its effectiveness in modeling heterogeneous borrower behavior. In contrast, XGBoost leverages gradient-based optimization to iteratively refine weak learners, enabling more precise discrimination across varied credit risk profiles.

LightGBM also performs competitively, particularly in settings with large feature sets and high-dimensional input spaces. Its leaf-wise tree growth strategy and histogram-based optimization allow for efficient training and scalability. However, empirical testing reveals that LightGBM exhibits slightly higher error sensitivity in borrower segments characterized by irregular invoice cycles and volatile cash-flow patterns. This behavior reflects the model's heightened responsiveness to short-term temporal fluctuations, which can amplify noise when transactional data is uneven or sparse. While this sensitivity may be advantageous in stable environments, it introduces additional variance in segments where financial activity is inherently irregular.

Random Forest, by contrast, demonstrates relatively stable performance across borrower groups due to its reliance on bagging and feature subsampling. However, this stability comes at the cost of reduced discriminatory power in high-dimensional feature spaces. The averaging mechanism that underpins Random Forest limits its ability to isolate nuanced interaction effects among behavioral, sectoral, and macroeconomic variables. As a result, Random Forest tends to underperform in cases where credit risk is driven by subtle, context-dependent relationships rather than dominant univariate signals.

Beyond aggregate accuracy metrics, error-distribution analysis provides deeper insight into model robustness and real-world applicability. XGBoost exhibits a more balanced misclassification structure across diverse borrower categories, including micro-enterprises, minority-owned firms, and businesses operating in volatile sectors. This balance indicates stronger generalization capacity and reduced bias toward specific subpopulations. In contrast, baseline models display skewed error distributions, often over-penalizing borrowers with nonstandard financial profiles. Such imbalances can exacerbate exclusionary outcomes in automated lending systems.

Receiver operating characteristic (ROC) curve comparisons further validate the advantages of boosting-based approaches. XGBoost and LightGBM consistently maintain higher true-positive rates at low false-positive thresholds, a critical requirement for lenders seeking to minimize default risk while preserving access to credit. This performance characteristic is particularly important in constrained lending environments, where marginal improvements in early risk detection can significantly reduce portfolio losses. The ability to sustain strong sensitivity without inflating false alarms underscores the practical value of ensemble boosting methods in credit decision pipelines.

To further explore model behavior, partial dependence plots reveal meaningful interaction effects among key features. One notable pattern is the amplification of repayment probability when customer engagement metrics-such as repeat transaction frequency or consistent invoice settlement-remain stable during periods of macroeconomic tightening. This finding suggests that behavioral resilience can partially offset adverse external conditions, a relationship that linear models fail to capture. Such insights are essential for developing more nuanced risk strategies that distinguish between temporary economic stress and structural borrower weakness.

SHAP value decomposition provides an additional layer of interpretive clarity by quantifying the marginal contribution of individual features to model predictions. Analysis shows that repayment behavior variables, including historical on-time payment rates and invoice settlement consistency, account for a substantial portion of predictive variance. These features emerge as the strongest determinants of creditworthiness across models, reinforcing the importance of behavior-based indicators in underserved lending contexts. However, sector-specific characteristics and macroeconomic indicators also demonstrate meaningful marginal effects, particularly during periods of economic stress. For example, regional unemployment trends and sector cyclicality exert greater influence when liquidity conditions tighten, highlighting the value of contextual signals in dynamic risk assessment.

The consistency of these findings validates the multidimensional feature engineering strategy employed in the modeling framework. By integrating transactional, behavioral, sectoral, and macroeconomic data, the system captures both intrinsic borrower characteristics and extrinsic environmental pressures. This holistic representation enables more accurate and equitable risk estimation than approaches that rely exclusively on historical financial statements or credit bureau data.

Importantly, the modeling architecture demonstrates that combining automated learning with interpretability tools significantly strengthens institutional trust in machine learning-driven credit decisions. SHAP-based explanations make visible the reasoning pathways underlying individual risk classifications, reducing the perceived opaqueness commonly associated with complex ensemble models. These explanations allow lenders to understand not only the predicted outcome but also the relative importance of contributing factors, facilitating internal review and external audit processes.

Interpretability is particularly critical for financial institutions adopting automated underwriting systems, as it supports compliance with fair-lending regulations and responsible-AI standards. Transparent explanations enable institutions to detect potential bias, justify credit decisions to regulators, and communicate outcomes more effectively to borrowers. For underserved small businesses, such transparency can improve trust in financial institutions and reduce the perception of arbitrary or discriminatory decision-making.

Overall, the empirical results demonstrate that ensemble boosting methods, supported by robust interpretability techniques, offer a powerful and responsible approach to credit risk assessment. By achieving superior predictive performance while maintaining transparency and fairness, this modeling framework provides a scalable foundation for expanding credit access to underserved populations while safeguarding institutional risk management objectives.

5. Implications for Lending Practices and Financial Inclusion

The findings underscore the transformative potential of ensemble machine learning in addressing long-standing credit exclusion in underserved business communities. Enhanced risk-prediction capability enables lenders to expand credit availability while maintaining or improving portfolio stability. Precision in risk stratification supports differentiated credit products, allowing institutions to offer adjusted interest rates, flexible repayment plans, and tailored credit lines based on borrower-specific behavioral patterns. Such differentiation reduces systematic overpricing of minority-owned or documentation-limited firms and improves fairness in underwriting.

From a regulatory and policy perspective, machine-learning-enhanced credit assessment aligns with national priorities promoting equitable lending and inclusive economic growth. The incorporation of alternative data strengthens regulators' ability to assess systemic disparities and informs policy design to address long-term financial inequities. Community development financial institutions (CDFIs) and mission-driven lenders can leverage ensemble tools to optimize credit allocation, reduce losses, and scale operations sustainably.

At the macroeconomic level, expanded credit access fosters employment growth, stimulates entrepreneurial activity, and strengthens local economic ecosystems. Improved risk evaluation also enhances financial-system resilience by enabling lenders to anticipate volatility and adjust strategies accordingly. By bridging data gaps and correcting structural biases, ensemble models contribute to a more inclusive financial infrastructure capable of supporting long-term economic mobility.

6. Future Directions and Strategic Recommendations

Several strategic directions can enhance the deployment of ensemble machine learning in small-business credit ecosystems. First, integrating real-time banking data and point-of-sale information can further improve the timeliness and accuracy of risk assessments. Second, incorporating fairness-aware machine learning techniques-such as constrained optimization or adversarial bias mitigation-can reduce algorithmic disparities and support equitable lending outcomes. Third, lender adoption would benefit from standardized interpretability frameworks, ensuring that automated decisions remain transparent and compliant with fair-lending regulations.

Fourth, hybrid architectures combining ensemble learning with deep sequence models may capture irregular operational cycles in detail, improving predictive robustness for businesses with seasonal or highly variable cash flows. Fifth, crossinstitutional data-sharing infrastructures, governed by privacy-preserving protocols, can expand the data available for model training while maintaining regulatory compliance. Finally, future research should examine longitudinal deployment impacts, including how automated models influence borrower trajectories, credit availability patterns, and long-term business resilience.

7. Conclusion

This study demonstrates that ensemble machine learning provides substantial advantages for credit risk assessment among underserved U.S. small businesses. By incorporating multidimensional behavioral, transactional, and macroeconomic features, ensemble models overcome the limitations of traditional scoring systems and enable more accurate, fair, and transparent evaluation. The empirical evidence confirms that boosting models-particularly XGBoost-excel in predictive performance and interpretability. These advancements support more inclusive lending strategies and align with broader economic goals of expanding entrepreneurial opportunity and reducing structural inequities. As financial institutions move toward data-driven underwriting, ensemble analytics offer a scalable and responsible pathway for modernizing credit assessment and strengthening the economic resilience of underserved communities.

References

1. L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5-32, 2001. doi: 10.1023/a:1010933404324
2. T. Chen, "XGBoost: A Scalable Tree Boosting System," *Cornell University*, 2016.
3. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, and T. Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, 2017.
4. FR Banks, "Small business credit survey," *Report on Minority-Owned*, 2017.
5. K. Langenbucher, "Responsible AI-based credit scoring-a legal framework," *European Business Law Review*, vol. 31, no. 4, 2020.