https://doi.org/10.59865/abacj.2025.27

IMPROVING CREDIT DECISIONS THROUGH MACHINE LEARNING AND ALTERNATIVE DATA: EVIDENCE FROM NBFIS

Jinnajate Achalapong[1] and Manit Satitsamitpong[2,*]

Abstract

Non-banking financial institutions (NBFIs) often struggle to make accurate credit decisions, especially for customers with insufficient traditional credit histories. Conventional models, such as logistic regression, primarily depend on credit bureau data and fail to capture the full credit potential of underserved populations—thereby hindering business expansion and financial inclusion. This study investigates how NBFIs can enhance credit decision-making by applying advanced machine learning techniques—namely XGBoost and neural networks—alongside alternative data sources, including mobile phone usage patterns, utility bill payments, and social media activity. Utilizing a real-world dataset of over 300,000 individuals, the findings demonstrate that machine learning models significantly outperform traditional approaches, particularly when alternative data is incorporated. These improvements lead to more precise risk classification, enabling institutions to reduce default rates, expand lending to previously overlooked borrowers, and improve portfolio profitability. In addition, the study addresses critical ethical and privacy considerations surrounding alternative data use. The results provide actionable insights for NBFIs aiming to adopt data-driven credit strategies that balance predictive power with responsible data governance—ultimately enhancing credit operations and promoting inclusive growth.

Keywords: Credit Decision-Making, Machine Learning, Alternative Data, NBFIs, Financial Inclusion

1. INTRODUCTION

The rapid evolution of digital economies has transformed the financial services industry, expanding access to credit and accelerating the adoption of data-driven decision-making (World Bank, 2022). Within this shifting landscape, non-banking financial institutions (NBFIs) have emerged as crucial players in extending financial services to underserved populations—particularly those lacking formal credit histories (Garg & Agarwal, 2014). Despite their growing importance, NBFIs continue to face significant challenges in making

[1]Jinnajate Achalapong (First Author) is a credit risk modeler with a specialization in the retail banking industry. He holds a bachelor's degree in Computer Engineering from Kasetsart University, Thailand. He is pursuing a master's degree in Digital Business Transformation (Data Science) at Thammasat University, Thailand. His professional interests include credit risk modeling, data analytics, and the application of machine learning in financial services. Email: jinnajate.a@gmail.com

[2,*] Asst. Prof. Dr. Manit Satitsamitpong (Corresponding Author) is currently working as a lecturer in the Master of Digital Strategy Program at Thammasat University, Thailand. He holds a bachelor's degree in Computer Science and Engineering from the University of Pennsylvania, an MBA from the University of Illinois, and an MS in Information Technology Management from Georgia Institute of Technology. He received his doctoral degree in Global Information and Telecommunication Studies at Waseda University, Japan. His research interests are in economics and policy issues on digital technology utilization in developing nations as well as in business and technical issues on the impact of digital technology within an organizational context. Email: manit@citu.tu.ac.th

accurate credit decisions due to their limited access to comprehensive financial data and the inherent limitations of traditional credit scoring methods (Óskarsdóttir, Bravo, Sarraute, Vanthienen, & Baesens, 2019). Conventional credit assessment techniques, particularly logistic regression, remain widely used due to their interpretability and simplicity. These models typically rely on historical financial indicators such as repayment behavior, debt-to-income ratios, and credit bureau scores (Thomas, Crook, & Edelman, 2017). However, their linear structure and dependence on traditional financial records hinder their effectiveness in evaluating thin-file or first-time borrowers. Logistic regression assumes linear relationships between input variables and default probabilities, which restricts the ability of this method to capture the complex and often non-linear patterns prevalent in real-world credit data (Baesens, Roesch, & Scheule, 2016). Moreover, when used to assess applicants without established credit records, these models often lead to inaccurate predictions and unnecessary credit exclusions (Blanco, Pino-Mejías, Lara, & Rayo, 2020).

To address these limitations, financial institutions are increasingly exploring the use of alternative data and machine learning (ML) techniques to enhance credit decision-making. Alternative data—such as mobile phone usage patterns, utility bill payment histories, social media activity, and other non-traditional behavioral indicators—offer new ways to assess a borrower's creditworthiness (Berg, Burg, Gombović, & Puri, 2020). These data sources can supplement or, in some cases, replace conventional credit bureau information, particularly for individuals who are financially active but remain unbanked or underbanked. Simultaneously, the development of advanced ML algorithms has enabled the construction of more flexible, robust, and accurate predictive models. ML models such as Extreme Gradient Boosting (XGBoost) and neural networks are well-suited for high-dimensional, imbalanced, and non-linear datasets commonly found in credit applications (Chen & Guestrin, 2016; LeCun, Bengio, & Hinton, 2015). These techniques can uncover complex interactions among features and detect subtle patterns that traditional models overlook. Empirical studies have shown that ML significantly improves prediction accuracy when combined with alternative data, especially for customers with limited or no credit history (Bazarbash, 2019; Lessmann, Baesens, Seow, & Thomas, 2015; Óskarsdóttir et al., 2019).

Despite global advancements in data science and artificial intelligence, the adoption of machine learning (ML) methods and alternative data by non-bank financial institutions (NBFIs) remains limited, particularly in emerging markets. Recent studies identify regulatory uncertainties, resource constraints, data privacy, and model transparency concerns as primary barriers (Alliance for Financial Inclusion [AFI], 2025; International Finance Corporation [IFC], 2020). Even in advanced economies, adoption is modest due to uncertain benefit-cost ratios and privacy concerns, with only 21% of non-bank lenders prioritizing innovation and alternative data (Bradford, 2023; HFS Research & Cognizant, 2025). In Thailand, the Bank of Thailand's Credit Risk Database (CRD) primarily supplements traditional credit scoring methods, indicating continued reliance on legacy practices (Tangsawasdirat, Tanpoonkiat, & Tangsatchanan, 2021). Addressing this gap, this study evaluates the predictive performance of traditional logistic regression versus advanced ML models (XGBoost, neural networks), using data from over 300,000 customers of an NBFI, guided by these research questions:

RQ1. Does the application of machine-learning models (XGBoost and neural networks) yield a statistically significant improvement in discriminatory power over a baseline logistic-regression scorecard for NBFI credit scoring?

RQ2. Does the integration of alternative data (mobile-usage, utility-payment, and social-behavioral features) enhance model discrimination compared to models relying solely on traditional data?

The motivation for this study is both theoretical and practical. Theoretically, it contributes to the growing literature on the integration of non-traditional data in predictive analytics, particularly in the context of the underbanked market segment and constrained institutional environments. Practically, it offers insights for NBFIs seeking to strengthen credit decisions, reduce default risks, and expand access to finance through intelligent, data-driven approaches. The study also highlights best practices for the responsible use of alternative data in compliance with ethical and regulatory standards.

2. LITERATURE REVIEW

Effective credit risk assessment is essential for non-banking financial institutions (NBFIs), particularly given their focus on market segments often underserved by traditional banks. These customers frequently lack sufficient credit histories, making it difficult for institutions to accurately assess repayment potential using conventional methods. This literature review explores the conceptual foundations and strategic importance of credit risk evaluation, the strengths and weaknesses of traditional scoring models, and the transformative role of machine learning and alternative data. It also reviews the ethical and operational challenges involved in modernizing credit decision-making systems and highlights key gaps in the literature which motivate the present study.

2.1 Concept and Importance of Credit Risk Assessment

Credit risk refers to the potential for financial loss when borrowers fail to meet loan repayment obligations (Thomas, Crook, & Edelman, 2017). Sound credit risk assessment is central to financial performance and institutional resilience, typically focusing on three dimensions: probability of default (PD), exposure at default (EAD), and loss given default (LGD) (Baesens, Roesch, & Scheule, 2016). Credit scoring models support risk-informed lending decisions by quantifying borrower creditworthiness using variables such as repayment history, debt-to-income ratios, and demographic indicators (Anderson, 2007). The logistic regression model, a standard in financial services, remains popular due to its simplicity, interpretability, and ease of implementation when sufficient historical data are available (Siddiqi, 2017).

The broader significance of credit risk assessment lies in its implications for economic inclusion and development. By enabling lenders to extend credit to reliable borrowers, it fosters consumption, entrepreneurship, and growth (Blanco, Pino-Mejías, Lara, & Rayo, 2020). In contrast, overly rigid or data-dependent models risk excluding applicants who lack conventional credit records but may be creditworthy, thereby reinforcing financial exclusion and inequality (Garg & Agarwal, 2014). NBFIs are particularly vulnerable to this issue, given their outreach to clients such as the self-employed, informal sector workers, and first-time borrowers—groups underrepresented in mainstream credit bureau datasets (Óskarsdóttir, Bravo, Sarraute, Vanthienen & Baesens, 2019). Consequently, traditional assessment methods often prove inadequate in these contexts, prompting the need for innovative techniques.

2.2 Traditional Credit Scoring Models and Their Limitations

Traditional credit scoring models—particularly logistic regression—serve as the backbone of many financial institutions' risk assessment frameworks. These models estimate the likelihood of loan default based on historical repayment behaviors, income stability, and other credit bureau data (Anderson, 2007). Their advantages include clarity, consistency, and regulatory familiarity, making them suitable for institutional settings that prioritize explainability (Siddiqi, 2017). Yet, despite their prevalence, logistic regression models carry several limitations that weaken their suitability for modern and inclusive credit assessment.

First, they rely on assumptions of linearity and variable independence, which oversimplify the complexity of borrower behavior (Baesens et al., 2016). Second, they depend heavily on structured historical financial records—data often unavailable for underbanked populations (Blanco et al., 2020). This limits model performance when evaluating applicants without credit bureau histories, resulting in high rejection rates for otherwise creditworthy individuals. Third, logistic regression struggles with class imbalance, as default cases typically constitute a minority in credit datasets. This can lead to biased predictions favoring the non-default class, thereby undermining the model's risk classification accuracy (Lessmann, Baesens, Seow & Thomas, 2015). Finally, traditional models are not adaptive to fast-changing market conditions, digital financial behaviors, or shifts in consumer patterns (Bazarbash, 2019). Their rigidity makes them insufficient for responding to emerging credit risk signals in dynamic environments, especially for NBFIs that operate in fluid economic settings.

2.3 Machine Learning in Credit Risk Assessment

In contrast, machine learning (ML) offers advanced tools that can address the above limitations. Unlike logistic regression, ML models can capture non-linear interactions and dependencies among a wide range of variables, enabling more nuanced risk classification. Techniques such as decision trees, random forests, XGBoost, and neural networks are well suited to the diverse, high-dimensional, and often imbalanced data encountered in credit modeling (Chen & Guestrin, 2016; LeCun, Bengio, & Hinton, 2015). XGBoost, for instance, applies gradient boosting to combine multiple weak learners into a robust predictive model, while neural networks use layered processing nodes to identify patterns that are not evident through traditional statistical analysis.

Empirical research strongly supports ML's performance benefits. Lessmann et al. (2015) found that ensemble and deep learning models outperform logistic regression in both predictive accuracy and sensitivity to rare default events. These models can also incorporate a broader array of variables—including behavioral and transactional indicators—enabling more inclusive and accurate assessments. Moreover, ML models dynamically improve over time as more data becomes available, making them adaptable to evolving borrower behaviors. For NBFIs seeking to extend credit access while controlling default risk, these capabilities are highly valuable.

2.4 The Use of Alternative Data in Credit Risk Assessment

Alternative data has gained prominence as a complementary or even standalone source for credit risk evaluation, particularly in markets with limited traditional credit coverage. This includes mobile phone usage patterns, utility payment records, social media interactions, online transactions, and digital wallet activity (Berg, Burg, Gombović & Puri, 2020). These datasets provide behavioral insights that can proxy for financial reliability. For example, mobile data such as call consistency, top-up behavior, and bill payment frequency, has demonstrated strong predictive value for repayment outcomes, particularly in the absence of formal credit records (Óskarsdóttir et al., 2019). Utility bill histories similarly reveal patterns of responsibility and cash flow regularity (Aitken, 2017), while e-commerce transactions and digital payments capture spending habits and liquidity (Jagtiani & Lemieux, 2019).

Social media data, although controversial, offers contextual signals about lifestyle and risk tolerance that can complement conventional metrics (Wei, Yildirim, Van den Bulte & Dellarocas, 2016). When integrated with ML algorithms, these alternative data sources significantly boost predictive power and allow institutions to score previously unscorable individuals. Studies confirm that such integration not only improves classification accuracy but also broadens credit access, thereby advancing financial inclusion (Berg et al., 2020).

Nevertheless, the use of alternative data raises concerns regarding data ownership, consent, and regulatory compliance. Users may not be aware that their non-financial behaviors are being evaluated in credit decisions. Without appropriate safeguards, the use of such data can violate privacy norms and lead to reputational risk or legal penalties (Wei et al., 2016).

2.5 Challenges and Ethical Considerations in Using ML and Alternative Data

The integration of ML and alternative data into credit scoring frameworks introduces both operational and ethical challenges. From an operational perspective, alternative data sources often lack standardized formats, requiring significant cleaning and preprocessing. Inaccurate or irrelevant variables may introduce noise and compromise model performance (Berg et al., 2020; Jagtiani & Lemieux, 2019). Moreover, advanced ML models such as neural networks or ensemble trees are often difficult to interpret, earning the label of “black-box” models. This lack of transparency complicates regulatory approval, customer communication, and internal audit procedures (Lessmann et al., 2015).

Ethically, the use of ML and behavioral data presents risks of privacy infringement and algorithmic discrimination. Alternative data often contains sensitive personal information, raising concerns under data protection laws such as GDPR. Additionally, historical biases embedded in training data can lead to discriminatory outcomes, particularly against marginalized groups (Óskarsdóttir et al., 2019). As models increasingly influence credit access, ensuring fairness, transparency, and accountability becomes essential.

Addressing these concerns requires a comprehensive governance approach. Institutions must implement explainable AI (XAI) methods, establish data quality and bias monitoring protocols, and adopt robust data privacy policies. Regular model audits and impact assessments should become standard practices, particularly for institutions like NBFIs that serve vulnerable populations.

2.6 Research Gaps and Justification for the Current Study

Although a growing body of research supports ML and alternative data in credit assessment, significant gaps remain. Most studies emphasize model performance without investigating the combined operational and ethical implications of deploying such systems in real-world NBFI environments. Furthermore, research often overlooks the unique challenges that NBFIs face—such as limited digital infrastructure, regulatory constraints, and the high proportion of unscorable clients.

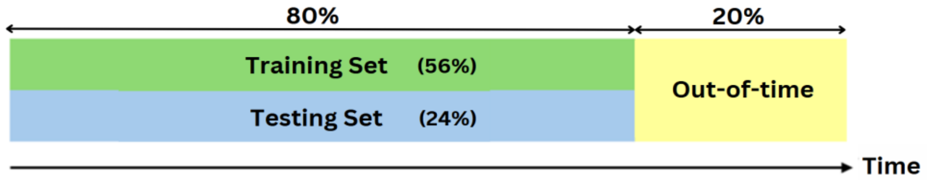
There is also a shortage of empirical work directly comparing traditional and ML-based models using real NBFI datasets with both conventional and alternative data. Ethical guidance remains mostly theoretical, with limited actionable recommendations for mitigating risks related to bias, transparency, and data privacy. This study seeks to address these gaps by conducting an empirical analysis of three specific credit models—logistic regression, XGBoost, and a neural network—on an NBFI dataset of over 300,000 borrowers. It also explores implementation strategies and ethical considerations, offering a comprehensive perspective for institutions aiming to modernize their credit risk management systems responsibly.

3. RESEARCH METHODOLOGY

This study employs data from a non-bank financial institution (NBFI), comprising over 300,000 customer records. The dataset includes traditional credit information, such as repayment history and credit utilization, alongside alternative behavioral data, including mobile usage, utility payments, and social media activity. Data exploration assessed dataset dimensions, data types, and missing values, revealing a default rate of approximately 8%. Missing values were addressed using mean imputation for traditional variables and K-Nearest Neighbors (KNN) for alternative data to preserve data integrity.

Feature engineering, informed by domain knowledge, generated relevant financial ratios and behavioral indicators. Recursive Feature Elimination (RFE) was employed for feature selection, retaining the most predictive variables. The final dataset was divided into training, testing, and out-of-time (OOT) validation sets to evaluate model performance across different temporal segments and ensure generalizability.

Figure 1 Data Preparation Step



Four predictive models were developed. A Random Model served as a baseline, assigning random probabilities to establish a minimum benchmark. Logistic Regression, trained using Stochastic Gradient Descent (SGDClassifier) with L2 regularization, was optimized through Randomized Search Cross-Validation (15 iterations) using Stratified K-Fold sampling. The best parameters were selected based on ROC-AUC scores, and the J-statistic was used to fine-tune the classification threshold, improving the true positive rate (TPR) while minimizing false positives. XGBoost, a tree-based ensemble method, was implemented for its robustness in handling large, imbalanced datasets. Its hyperparameters were tuned using Bayesian Optimization (10 iterations), again prioritizing ROC-AUC.

Artificial neural networks (ANN) were also used to model the complex, non-linear relationships in the data sets. Dropout regularization was applied to mitigate overfitting, and hyperparameters—including node count, dropout rate, and learning rate—were optimized via a random search. The best-performing configuration for traditional data featured a three-layer structure (64-32-32 units) with a dropout rate of 0.2 and learning rate of 0.00162. For models incorporating both traditional and alternative data, the optimal architecture expanded to 128-96-48 units with a learning rate of 0.00038. Model performance was assessed using confusion matrices, precision, recall, F1-score, and ROC-AUC to ensure both statistical robustness and practical relevance for credit risk assessment in NBFIs.

4. RESEARCH FINDINGS

4.1 Evaluation Framework and Metrics

This section presents the empirical findings from evaluating the four credit scoring models—Random Model, Logistic Regression, XGBoost, and Neural Network—across two data environments: traditional financial data and alternative data sources. Each model was tested on training, testing, and out-of-time (OOT) validation datasets to assess both their learning capacity and robustness. Model performance was evaluated using standard classification metrics, including Accuracy, Precision, Recall, F1-score, and ROC-AUC. The Gini coefficient, derived from ROC-AUC, was also used as a key measure of discriminatory power in financial contexts. The Random Model was introduced as a non-learning benchmark to validate that all supervised models provided substantial improvements beyond chance-level prediction.

4.2 Results Using Traditional Financial Data

As shown in Table 1, the Random Model performed at chance level, producing ROC-AUC values around 50 percent and F1-scores below 14 percent across all datasets. These results confirm the model's lack of learning capability and its sole utility as a performance baseline. Logistic Regression, implemented with L2 regularization, demonstrated a clear improvement over the random benchmark. The model yielded ROC-AUC values between 72.85 and 73.45 percent and accuracy scores around 68 percent. However, its precision and recall values remained modest, reflecting its limitations in capturing complex borrower behavior. Notably, the model identified features such as interest burden, credit history duration, and delinquency records as significant predictors.

Table 1 Results for Traditional Data

Model	Dataset	ROC-AUC	Accuracy	Precision	Recall	F1-Score
Random Model	Training	49.87%	49.98%	8.13%	49.81%	13.98%
	Testing	50.03%	50.00%	7.99%	50.11%	13.79%
	Out-of-time	49.69%	49.66%	7.85%	49.69%	13.55%
Logistic Regression	Training	73.45%	67.63%	15.50%	66.61%	25.14%
	Testing	73.23%	67.86%	15.25%	66.46%	24.81%
	Out-of-time	72.85%	68.92%	15.37%	64.68%	24.84%
XGBoost	Training	77.19%	72.16%	18.03%	68.04%	28.51%
	Testing	77.17%	67.48%	16.26%	74.12%	26.67%
	Out-of-time	76.80%	69.19%	16.66%	71.95%	27.05%
Neural Network	Training	78.63%	86.42%	27.79%	41.59%	33.32%
	Testing	74.49%	82.86%	21.73%	44.12%	29.11%
	Out-of-time	73.65%	83.54%	21.67%	41.01%	28.36%

XGBoost demonstrated stronger performance across all metrics, particularly in terms of recall and F1-score. Its ROC-AUC values peaked at 77.17 percent, and recall exceeded 74 percent in the test set. The model's ability to handle nonlinear interactions and high-dimensional data enabled it to better leverage features such as debt-to-credit ratios and regional credit risk indicators. XGBoost thus proved to be a more behaviorally sensitive model compared to logistic regression.

The neural network delivered the highest training accuracy, reaching 86.42 percent, and attaining the top F1-score at 33.32 percent on the training dataset. While it maintained solid performance on the testing and OOT datasets, a decline in recall and slight drop in precision suggested some degree of overfitting. Nevertheless, it consistently outperformed the other models across several metrics, although its lack of interpretability poses a concern for real-world credit decision environments.

4.3 Results Using Traditional Financial Data and Alternative Data

The inclusion of alternative data—such as mobile phone usage, utility payments, and social media behavior—resulted in marked performance improvements across all predictive models. These datasets provide behavioral insights that can proxy for financial reliability. For example, mobile data such as usage consistency, top-up behavior, and utility bill payment behavior demonstrated strong predictive value for repayment outcomes.

Table 2 Results for Traditional Data combined with Alternative Data

Model	Dataset	ROC-AUC	Accuracy	Precision	Recall	F1-Score
Random Model	Training	50.41%	49.98%	8.23%	50.52%	14.15%
	Testing	49.61%	49.59%	7.79%	49.06%	13.44%
	Out-of-time	49.63%	50.21%	7.97%	49.96%	13.75%
Logistic Regression	Training	77.86%	70.66%	17.68%	70.99%	28.31%
	Testing	77.81%	68.84%	16.79%	73.43%	27.33%
	Out-of-time	78.19%	65.70%	15.96%	77.83%	26.49%
XGBoost	Training	79.37%	71.54%	18.39%	72.39%	29.33%
	Testing	79.58%	70.22%	17.76%	75.28%	28.74%
	Out-of-time	79.26%	69.57%	17.33%	75.10%	28.16%
Neural Network	Training	84.97%	90.46%	42.42%	47.30%	44.72%
	Testing	77.01%	85.78%	25.26%	39.92%	30.94%
	Out-of-time	77.08%	83.95%	23.37%	44.78%	30.71%

As illustrated in Table 2, the Random Model continued to yield ROC-AUC values close to 50 percent, underscoring its utility as a baseline rather than as an effective predictive tool.

Logistic Regression exhibited substantial gains with the enriched dataset, achieving an ROC-AUC of 78.19 percent and recall of 77.83 percent on the Out-of-Time (OOT) set. These metrics underscore the value of behavioral data in accurately identifying creditworthy borrowers. However, the moderate precision and F1-score suggest that, although alternative data enhanced discriminatory power, the linear structure of Logistic Regression constrained its capability to capture intricate, nonlinear interactions.

XGBoost demonstrated further improved performance when augmented by alternative data. ROC-AUC increased notably to 79.6 percent, with recall consistently exceeding 75 percent across all evaluation datasets. The model's strengths lie in effectively synthesizing diverse data types and uncovering subtle behavioral patterns, such as community repayment habits and peer influences. These insights contributed to more balanced and dependable credit risk predictions.

The neural network reached its highest accuracy of 90.46 percent and F1-score of 44.72 percent in the training dataset, underscoring its potential when leveraging extensive and nuanced alternative data inputs. The results from the testing and OOT datasets confirmed enhanced generalization compared to traditional data approaches, as evidenced by increased precision and recall. Nonetheless, the inherent complexity and limited interpretability of neural networks poses challenges for practical deployment, particularly in regulated financial environments where model transparency is paramount.

4.4 Summary of Empirical Results

In summary, all supervised learning models significantly outperformed the random baseline across both data environments, validating their capacity to learn meaningful credit risk patterns. The transition from traditional to alternative data consistently enhanced model performance, particularly in terms of recall and F1-score, which are critical for minimizing credit losses and opportunity costs. While logistic regression benefited from the additional data, its linear structure constrained its performance ceiling. XGBoost consistently offered a compelling balance between predictive power and interpretability, making it a practical choice for deployment in non-banking financial institutions. The neural network delivered the strongest predictive results but remained less suitable for high-stakes decision-making contexts due to interpretability limitations. These results lay the foundation for deeper strategic and managerial discussions, which are explored in the following section.

5. DISCUSSION AND CONCLUSION

5.1 Interpretation of Empirical Results

The findings from the previous section confirm that machine learning models, particularly XGBoost and neural networks, substantially outperform logistic regression in predicting credit risk. These performance differences are most evident in recall and F1-score metrics, which are directly related to the cost of misclassification. Notably, the addition of alternative data yielded considerable gains across all models, reinforcing the notion that behavioral signals—such as mobile usage or peer repayment patterns—capture borrower dynamics not visible through traditional financial records alone.

Among the models tested, XGBoost consistently delivered the best balance between accuracy, robustness, and interpretability. The neural network achieved the highest predictive performance but exhibited signs of overfitting and presented practical challenges due to its black-box nature. Logistic regression, while easy to interpret, demonstrated limited adaptability to complex borrower behavior. Taken together, these results highlight the transformative potential of both machine learning and alternative data in advancing credit decision-making for NBFIs.

5.2 Responses to the Research Questions

For RQ1, the results indicate that both XGBoost and neural networks demonstrate statistically significant improvements in discriminatory power over logistic regression in the context of credit scoring for NBFIs. Using the Hanley & McNeil test for comparing correlated ROC-AUC values at a significance level of $\alpha = 0.05$, both machine learning models yielded p-values below 0.05, confirming that the observed performance differences are unlikely to have occurred by chance. XGBoost consistently exhibited superior ROC-AUC, F1-scores, and recall across the testing and out-of-time (OOT) datasets, while also providing greater model stability and business interpretability. Although neural networks were shown to achieve comparable or even higher predictive accuracy in some settings, they pose challenges in terms of transparency and explainability—factors critical to risk governance and regulatory compliance. These findings suggest that machine learning models, particularly XGBoost, can serve as statistically and operationally superior alternatives to traditional credit scoring approaches.

Regarding RQ2, the inclusion of alternative data—such as mobile usage, utility payments, and social behavioral features—led to statistically significant enhancements in model discrimination across all modeling approaches. This was evidenced by improved ROC-AUC scores, which were again tested using the Hanley & McNeil method and confirmed at the 5% significance level. The enriched input space provided by alternative data allowed models to better capture behavioral patterns, particularly for thin-file borrowers who lack extensive credit histories. Even logistic regression benefited from the additional data, though the gains were most pronounced in the XGBoost and neural network models. These results underscore the complementary value of alternative data in improving both the accuracy and inclusivity of credit scoring systems for NBFIs.

5.3 Data Privacy Governance for NBFIs

To implement responsible data practices, NBFIs should integrate privacy governance directly into their modeling processes. Clear, user-friendly consent mechanisms at onboarding and when introducing new data sources can inform customers about data collection purposes. By collecting only essential attributes—such as abstracted device-usage metrics rather than full logs—and deleting raw data promptly after deriving necessary features, institutions can uphold data minimization principles. Sensitive identifiers should be protected via pseudonymization or hashing to prevent re-identification. Internally, strict access controls and immutable audit logs ensure only authorized staff are able to access personal data, with regular reviews conducted. Adopting an automated "privacy-by-design" approach, validating datasets against approved schemas before modeling, helps NBFIs maintain compliance efficiently, balancing innovation and customer trust.

5.4 Theoretical Implications

This study contributes to the growing body of research at the intersection of machine learning, behavioral data, and credit risk. Empirically validating the role of alternative data in improving model performance, helps bridge theoretical gaps between traditional econometric modeling and modern data science. The results support a broader reconceptualization of creditworthiness that includes both financial and behavioral dimensions, particularly relevant in emerging markets and thin-file populations.

5.5 Managerial and Strategic Implications

This research provides profound managerial and strategic insights for leaders at non-banking financial institutions (NBFIs), underscoring the transformative potential of integrating alternative data and advanced analytical methods such as XGBoost. The Swap-set Analysis (Table 3) illustrates a striking strategic advantage: adopting alternative data significantly reduces default risk by more than 50% in band 5 (50% of applicants), lowering it from 3.31% using traditional methods to 1.68% when leveraging XGBoost analytics. Alternatively, maintaining an established risk appetite at 2.8% enables institutions to expand customer acceptance dramatically by 75%, moving from band 4 (40% of applicants) up to band 7 (70% of applicants). This expanded customer reach represents a major growth opportunity without elevating risk levels.

Table 3 Swap-set Analysis

Credit Score Band	Logistic Regression (Traditional Data only)				XGBoost (Traditional Data with Alternative Data)			
	Non-Default	Default	Cum. Account	Cum. Default	Non-Default	Default	Cum. Account	Cum. Default
Band 1	30,283	468	10.00%	1.52%	30,612	139	10.00%	0.45%
Band 2	29,976	775	20.00%	2.02%	30,449	302	20.00%	0.72%
Band 3	29,740	1,010	30.00%	2.44%	30,309	441	30.00%	0.96%
Band 4	29,488	1,263	40.00%	2.86%	30,069	682	40.00%	1.27%
Band 5	29,174	1,577	50.00%	3.31%	29,732	1,019	50.00%	1.68%
Band 6	28,703	2,047	60.00%	3.87%	29,301	1,449	60.00%	2.19%
Band 7	28,224	2,527	70.00%	4.49%	28,725	2,026	70.00%	2.81%
Band 8	27,585	3,165	80.00%	5.22%	27,652	3,098	80.00%	3.72%
Band 9	26,288	4,463	90.00%	6.25%	25,923	4,828	90.00%	5.05%
Band 10	23,221	7,530	100.00%	8.07%	19,910	10,841	100.00%	8.07%

The practical and strategic implications for marketing and customer acquisition are particularly compelling. Utilizing detailed behavioral insights—such as mobile phone usage, utility payment consistency, and social media interactions—enables institutions to craft precisely targeted marketing strategies. Such tailored approaches facilitate entry into market segments previously overlooked by traditional credit evaluation processes, substantially increasing both market share and the breadth of the customer base. This nuanced understanding allows for personalized financial products that resonate deeply with customer needs and lifestyles, transforming traditional customer engagement into a proactive, customer-centric model.

Moreover, this approach positions financial technology not merely as a competitive tool, but as a powerful lever for societal improvement. By providing underserved populations access to fair and affordable credit, NBFIs enable individuals and households to make critical investments in education, healthcare, housing, and small business ventures—investments that fundamentally enhance their quality of life and economic mobility. Consequently, this approach not only fosters significant social impact but also enhances corporate image, allowing NBFIs to build a reputation as responsible and socially committed.

The integration of innovative credit methodologies aligned with societal benefits creates a sustainable competitive advantage. Institutions adopting this strategic direction not only achieve tangible financial growth but also contribute directly to community stability and socioeconomic development. Thus, the findings of this study offer a blueprint for institutions aiming to merge commercial success with meaningful social responsibility. Ultimately, embracing smart, innovative credit solutions promotes sustainable business growth, reinforces organizational resilience, and significantly elevates institutional brand equity through sustained, impactful community engagement.

5.6 Limitations and Future Research

This study's reliance on data from a single NBFI may limit the generalizability of the findings. As the findings are based on a specific customer demographic and operational context, results such as predictive accuracy and feature importance might differ in other institutions or markets. Future research should involve cross-institutional validation by applying similar machine learning methods to datasets from different NBFIs within comparable emerging markets. Such studies could refine guidelines for feature selection, privacy adherence, and model calibration. Additionally, exploring more interpretable neural network designs, such as attention-based models, and assessing the long-term stability of machine learning credit scores are essential areas for further investigation. Ethical considerations surrounding fairness and privacy, particularly concerning alternative data use, also warrant deeper empirical exploration.

5.7 Conclusion

This research demonstrates that machine learning—especially XGBoost—and alternative data sources can meaningfully enhance credit risk assessment for NBFIs. These innovations improve risk prediction, reduce misclassification, and support the provision of broader access to credit, particularly for underserved populations. For institutions aiming to modernize their lending strategies, the strategic integration of machine learning with expanded data signals represents a compelling path toward both business growth and financial inclusion. Embedding such models within a responsible governance framework ensures that innovation remains aligned with transparency, fairness, and long-term sustainability.

REFERENCES

Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management,* 18(2-3), 59-88. https://doi.org/10.1002/isaf.325

Aithal, V., & Jathanna, R. D. (2019). Credit risk assessment using machine learning techniques. *International Journal of Innovative Technology and Exploring Engineering,* 9(1), 3570-3575.

Albanesi, S., & Vamossy, D. F. (2019). Predicting consumer default: A deep learning approach (NBER Working Paper No. 26165). *National Bureau of Economic Research*. https://doi.org/10.3386/w26165

Alliance for Financial Inclusion. (2025). Alternative data for credit scoring. *AFI*. https://www.afi-global.org/publication/2025-alternative-data-for-credit-scoring/

Avery, R. B., Calem, P. S., & Canner, G. B. (2004). Consumer credit scoring: Do situational circumstances matter? *Journal of Banking & Finance*, 28(4), 835-856. https://doi.org/10.1016/j.jbankfin.2003.10.009

Anderson, R. (2007). *The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation*. Oxford University Press.

Baesens, B., Roesch, D., & Scheule, H. (2016). *Credit risk analytics: Measurement techniques, applications, and examples in SAS*. John Wiley & Sons.

Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2019). Consumer-lending discrimination in the FinTech era (NBER Working Paper No. 25943). *National Bureau of Economic Research*. https://doi.org/10.3386/w25943

Bazarbash, M. (2019). FinTech in financial inclusion: Machine learning applications in assessing credit risk (IMF Working Paper No. 19/109). *International Monetary Fund*. https://doi.org/10.5089/9781498314428.001

Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of intechs: Credit scoring using digital footprints. *The Review of Financial Studies*, 33(7), 2845-2897. https://doi.org/10.1093/rfs/hhz099

Blanco, A., Pino-Mejías, R., Lara, J., & Rayo, S. (2020). Assessing the fairness of credit scoring models. *Journal of Financial Services Research*, 57(2), 153-177. https://doi.org/10.1007/s10693-019-00322-8

Bradford, T. (2023). Give me some credit!: Using alternative data to expand credit access (Payments System Research Briefing). Federal Reserve Bank of Kansas City. Retrieved June 27, 2025, from https://www.kansascityfed.org/research/payments-system-research-briefings/give-me-some-credit-using-alternative-data-to-expand-credit-access/

Bravo, C., Maldonado, S., & Weber, R. (2015). Methodologies for the construction of models for credit granting based on utility functions. *Annals of Operations Research*, 233(1), 237-264. https://doi.org/10.1007/s10479-013-1492-3

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). *ACM*. https://doi.org/10.1145/2939672.2939785

De Cnudde, S., Moeyersoms, J., Stankova, M., Tobback, E., Javaly, V., & Martens, D. (2019). What does your Facebook profile reveal about your creditworthiness? Using alternative data for microfinance. *Journal of the Operational Research Society*, 70(3), 353-363. https://doi.org/10.1080/01605682.2018.1434402

Fourcade, M., & Healy, K. (2013). Classification situations: Life-chances in the neoliberal era. *Accounting, Organizations and Society*, 38(8), 559-572. https://doi.org/10.1016/j.aos.2013.11.002

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139. https://doi.org/10.1006/jcss.1997.1504

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance*, 77(1), 5-47. https://doi.org/10.1111/jofi.13090

Gambacorta, L., Huang, Y., Qiu, H., & Wang, J. (2019). How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm (BIS Working Paper No. 834). *Bank for International Settlements*. https://www.bis.org/publ/work834.pdf

Garg, N., & Agarwal, P. (2014). Financial inclusion in India–a review of initiatives and achievements. *IOSR Journal of Business and Management*, 16(6), 52-61.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Haykin, S. S. (2009). *Neural networks and learning machines* (3rd ed.). Pearson.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering,* 21(9), 1263-1284. https://doi.org/10.1109/TKDE.2008.239

HFS Research & Cognizant. (2025). Reinventing non-bank mortgage lending journey in the age of AI. *HFS Research*. https://www.hfsresearch.com/research/cognizant-non-bank-mortgage/

International Finance Corporation. (2020). Artificial intelligence innovation in financial services (EM Compass Note 85). *World Bank Group*. https://www.ifc.org/content/dam/ifc/doc/mgrt/emcompass-note-85-ai-innovation-in-financial-services.pdf

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. https://doi.org/10.1038/nature14539

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. https://doi.org/10.1016/j.ejor.2015.05.030

Leyshon, A., & Thrift, N. (1995). Geographies of financial exclusion: Financial abandonment in Britain and the United States. *Transactions of the Institute of British Geographers*, 20(3), 312-341. https://doi.org/10.2307/622654

Losing, V., Hammer, B., & Wersing, H. (2018). Incremental on-line learning: A review and comparison of state of the art algorithms. *Neurocomputing*, 275, 1261-1274. https://doi.org/10.1016/j.neucom.2017.06.084

Omarini, A. E. (2018). Banks and fintechs: How to develop a digital open banking approach for the bank's future. *International Business Research*, 11(9), 23-36. https://doi.org/10.5539/ibr.v11n9p23

Óskarsdóttir, M., Bravo, C., Sarraute, C., Vanthienen, J., & Baesens, B. (2019). The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. *Applied Soft Computing*, 74, 26-39. https://doi.org/10.1016/j.asoc.2018.10.004

Panjwani, M., Joshi, S., & Sinha, S. (2019). Application of machine learning techniques for stress testing. *Journal of Risk Management in Financial Institutions*, 12(4), 365-380.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215.

https://doi.org/10.1038/s42256-019-0048-x
Samek, W., Wiegand, T., & Müller, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv*. https://arxiv.org/abs/1708.08296
Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), 210-229. https://doi.org/10.1147/rd.33.0210
San Pedro, J., Proserpio, D., & Oliver, N. (2015). MobiScore: Towards universal credit scoring from mobile phone data. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 195-207). Springer. https://doi.org/10.1007/978-3-319-20267-9_16
Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing*, 5(4), 13-22.
Shrestha, Y. R., Krishna, V., & von Krogh, G. (2021). Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. *Journal of Business Research*, 123, 588-603. https://doi.org/10.1016/j.jbusres.2020.09.068
Siddiqi, N. (2017). *Intelligent credit scoring: Building and implementing better credit risk scorecards* (2nd ed.). John Wiley & Sons.
Tangsawasdirat, B., Tanpoonkiat, S., & Tangsatchanan, B. (2021). Credit risk database: Credit-scoring models for Thai SMEs (PIER Discussion Paper No. 168). Puey Ungphakorn Institute for Economic Research. https://www.pier.or.th/en/dp/168/
Thomas, L. C., Crook, J., & Edelman, D. (2017). Credit scoring and its applications (2nd ed.). *Society for Industrial and Applied Mathematics*.
Verma, S., & Rubin, J. (2018). Fairness definitions explained. In Proceedings of the International Workshop on Software Fairness (FairWare ,18) (pp. 1-7). *ACM*. https://doi.org/10.1145/3194770.3194776
Wei, Y., Yildirim, P., Van den Bulte, C., & Dellarocas, C. (2016). Credit scoring with social network data. *Marketing Science*, 35(2), 234-258. https://doi.org/10.1287/mksc.2015.0949
World Bank. (2022). The Global Findex Database 2021: Financial inclusion, digital payments, and resilience in the age of COVID-19. *World Bank*. https://doi.org/10.1596/978-1-4648-1897-4
Zhang, W., He, H., & Zhang, S. (2019). A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: An application in credit scoring. *Expert Systems with Applications*, 121, 221-23
Zhou, Z. H. (2012). *Ensemble methods: Foundations and algorithms*. CRC Press.