A Comparative Evaluation of Machine Learning Classification Algorithms for Credit Scoring in Financial Institutions with Limited Data Availability

Mikiyas Fekadu
HiLCoE School of Computer Science and Technology
micky_fek@hotmail.com

Tibebe Beshah
HiLCoE School of Computer Science and Technology, Addis Ababa, Ethiopia
Tibebe.beshah@gmail.com

Abstract

Credit scoring is the fundamental analytical process by which financial institutions evaluate a potential borrower's creditworthiness to determine the likelihood of future default. At its most basic level, it serves as a risk-assessment gateway, transforming a candidate's historical financial behaviors, such as past repayment consistency, debt-to-income ratios, and credit utilization, into a single, actionable numerical value. In the modern financial landscape, this process has become inextricably linked with Machine Learning (ML), as traditional statistical methods often struggle to capture the complex, non-linear relationships hidden within modern high-dimensional data. By leveraging ML, institutions can move beyond rigid, rule-based systems to identify subtle behavioral patterns that define a borrower's risk profile more accurately.

For this purpose, this paper evaluates the performance of six machine learning classification algorithms: Logistic Regression, Random Forest, Naïve Bayes, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and the Decision Tree for credit scoring under data scarcity, addressing the critical challenge of limited labeled data in financial risk assessment. A rigorous methodology is employed, encompassing data preprocessing, exploratory analysis (EDA), and model development, with evaluation based on accuracy, precision, recall, F1-score, and AUC-ROC. To ensure robustness, models are validated via cross-validation, mitigating overfitting risks inherent in small datasets.

In effect. Random Forest excels as the most effective model, achieving 82.50% accuracy alongside high precision (82.38%), recall (83.15%), and F1-score (82.74%), attributable to its ensemble-based resilience to data sparsity and overfitting. The Decision Tree matches this accuracy and surpasses Random Forest in precision (83.69%), offering interpretability but requiring regularization to curb overfitting. KNN delivers moderate performance but suffers from computational inefficiency, while SVM's balanced metrics are offset by high complexity. Logistic Regression underperforms due to linearity constraints, and Naïve Bayes fails to generalize, hampered by its independence assumption. The study underscores the superiority of ensemble methods like Random Forest for credit scoring with limited data, while dissecting trade-offs between

interpretability and predictive power. By systematically benchmarking algorithms under data constraints, this work bridges a critical gap in credit risk modeling, offering actionable insights for real-world applications.

Keywords: *credit scoring, machine learning, limited data, random forest, financial inclusion, classification algorithms*

1. Introduction

A credit scoring model is fundamentally a statistical tool used to determine the creditworthiness of an applicant by predicting the probability of default using historical data [1]. These models have become indispensable for financial institutions, enabling data-driven decisions that minimize default risk and maximize profitability [2].

At the core of credit scoring lies classification a key data mining task that is used to distinguish between good and bad credit applicants. Over the years, a range of machine learning algorithms, including logistic regression, decision trees, neural networks, and ensemble models, have been applied to build credit scoring systems. The continued pursuit of competitive advantage in the credit industry has further accelerated the adoption of data mining and machine learning techniques [3].

However, the effectiveness of these models largely depends on the availability and quality of data. In today's data-driven economy, data is often likened to a new form of capital. Data science, as an interdisciplinary field, has emerged to extract actionable insights from structured and unstructured data using algorithms, systems, and scientific methodologies. One of its most impactful domains is financial services, where data-driven decision-making underpins product design, customer segmentation, and credit evaluation.

Digital microcredit solutions, in particular, have seen widespread adoption, offering convenient, rapid access to finance for underserved populations. These services rely heavily on credit scoring systems to analyze customer behavior, identify patterns, and predict repayment likelihood. Beyond profitability, credit scoring plays a pivotal role in advancing financial inclusion by enabling access to credit for individuals and MSMEs, particularly in emerging markets. Its growth has been driven by expanded data availability, improved computational capabilities, and rising demand for operational efficiency [4].

Despite these advancements, a key challenge persists: the scarcity of labeled data. Most credit scoring models are designed under the assumption of large, rich datasets, but many real-world contexts especially in developing markets or early-stage ventures offer only limited data for model training. In such scenarios, conventional algorithms may underperform, overfit, or fail to generalize.

Despite its central role in modern finance, objective lending remains highly sensitive to the quality of feature selection and algorithmic choice, making credit scoring models

particularly fragile in data-constrained environments [5, 6]. In practice, the labelled data required to fuel machine learning models is often scarce, constrained by high annotation costs, privacy concerns, and fragmented legacy systems [7, 8]. This challenge is especially pronounced in the Ethiopian banking sector, where the absence of a centralized credit bureau and heightened data sensitivity frequently compel institutions to rely on incomplete or proxy datasets, amplifying the presence of redundant features and noise that degrade predictive accuracy [9, 10].

As a result, financial institutions lack a clear and empirically grounded framework for making high-stakes credit decisions under conditions of severe data scarcity. Against this backdrop, this study seeks to develop a resilient credit scoring framework by systematically evaluating which classification architectures preserve predictive integrity in small-sample settings, how feature selection can be optimized when data is sparse, and what performance trade-offs emerge when algorithms designed for "Big Data" are deployed in "Small Data" environments. By benchmarking six widely used machine learning classifiers, the research provides practical, evidence-based guidance for micro-credit providers aiming to minimize default risk in data-poor landscapes, while also offering transferable insights for financial practitioners operating across emerging economies.

2. Related Work

The journey of credit risk assessment has moved from rigid, parametric statistical modelling toward the fluid intelligence of Machine Learning (ML). A landmark meta-analysis by Jukna [11] examined 71 high-impact studies, revealing that while Logistic Regression (LR) remains the industry's "workhorse" due to its presence in 55% of the literature, Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) have reached dominant adoption rates of 52% and 41%, respectively.

This shift is driven by a need for both accuracy and intuition. Logistic Regression is lauded for its "glass-box" nature, offering clear interpretability through maximum likelihood estimation [12]. However, in the complex, non-linear reality of modern finance, ANNs provide a vital alternative by mimicking biological cognitive structures, allowing models to learn patterns without pre-defined distribution assumptions [13, 14]. Similarly, SVMs excel in high-dimensional environments by identifying the optimal hyperplane that separates creditworthy borrowers from high-risk ones [15, 16]. For practitioners who value visual logic, Decision Trees (DTs) and their sophisticated successor, Random Forests (RFs), offer an ensemble approach that prevents the model from memorising noise, effectively reducing the risk of overfitting [17, 18]. While traditional benchmarks like Linear Discriminant Analysis (LDA) still hold historical value, the current trend leans toward non-parametric methods like K-Nearest Neighbours (KNN) and Naive Bayes (NB)

for their agility in processing categorical data [19, 20].

The integrity of a credit score is only as good as the features that inform it. Feature selection acts as a filter, removing the noise to reveal the true signal of creditworthiness. Research typically categorises these filters into two camps: filter methods, which use statistical metrics like information gain to rank features, and wrapper methods, which treat the learning algorithm itself as a measuring stick for feature utility [3].

In developing economies like Ethiopia, the challenge is not merely just noisy data but a total lack of thick-file financial records [10]. This has forced a pivot toward alternative data, where utility payments and mobile phone behaviours serve as proxies for financial responsibility [21, 22]. While this promotes financial inclusion, it introduces a human cost regarding privacy and the need for robust regulatory infrastructure [23, 24]. Crucially, the academic consensus is shifting away from the "Big Data" requirement; recent breakthroughs in medical and financial diagnostics prove that high-performing models can be trained on as few as 500 to 1,000 carefully curated samples, providing a beacon of hope for data-scarce regions [25].

Prior studies in credit scoring demonstrate a gradual shift from traditional statistical models toward more diverse machine learning approaches. Research presented at the 2017 International Conference on Computational Science and Computational Intelligence (CSCI) notes that while linear regression and discriminant analysis were historically dominant, classifiers such as artificial neural networks, decision trees, genetic algorithms, and support vector machines have become increasingly common [26]. The same study highlights a relative gap in the literature regarding the use of the Naive Bayes algorithm for credit scoring, despite empirical results showing that it achieves competitive performance, outperforming classification trees and KNN with an accuracy of 83.3%, largely due to its efficiency and effective handling of categorical variables.

Other work by Kumar [27] introduces a hybrid credit scoring framework that combines deep neural networks with K-Means clustering and decision tree classification, demonstrating strong predictive accuracy when applied to real-world credit datasets and emphasizing borrower behavioral patterns as key risk indicators. In a complementary direction, Laborda and Ryoo [28] examine the role of feature selection in mitigating overfitting, showing that wrapper-based methods, particularly forward stepwise selection, consistently enhance the performance of multiple classifiers, including logistic regression, SVM, KNN, and random forest, by addressing dimensionality challenges in credit scoring models.

The landscape of credit scoring has shifted dramatically between 2023 and 2025, moving from basic classification models toward high-complexity intelligent systems designed to handle the nuances of modern finance. Recent literature published between

2023 and 2024 emphasises a critical pivot toward Explainable AI (XAI). As financial regulators globally and increasingly in emerging markets demand higher levels of transparency, researchers have begun integrating frameworks like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). These tools are no longer optional; they are essential for ensuring that black-box neural ensembles can be audited for fairness and institutional bias, providing a right to explanation for the borrower [29].

For nations like Ethiopia, transfer learning represents a significant technological leap. It allows local institutions to take an expert model trained on millions of global records and fine-tune it using a small local dataset essentially teaching an old model new tricks in a data-scarce context [30]. Furthermore, federated learning is emerging as the ultimate privacy shield, allowing multiple banks to learn from each other's data patterns without ever actually seeing or sharing the customer's private information [31, 32].

A review of the existing literature shows that data has become central to modern society, with data science playing a key role in transforming raw information into actionable insights for informed decision-making. Within this landscape, machine learning has emerged as a particularly influential tool, and credit scoring models have become essential for enabling remote, automated, and timely credit decisions, thereby improving efficiency and supporting financial inclusion.

However, much of the current research assumes the availability of large volumes of labelled data, leaving scenarios characterised by limited data relatively underexplored. In many real-world settings, especially in developing economies or niche financial markets, data constraints pose significant challenges to accurately assessing borrower creditworthiness. As a result, there remains a clear gap in understanding how credit scoring models can be effectively adapted or optimised under data-scarce conditions, highlighting the need for further empirical investigation to support equitable access to credit.

3. Methodology

This study utilizes a quantitative, experimental Design Science Research (DSR) framework grounded in empirical model development and evaluation. The methodology aligns well with the objective of building and validating a machine learning-based credit scoring model under constraints of limited data a real and pressing issue in many emerging financial markets.

This sequence, shown in figure 1, follows a design science research (DSR) methodology, which is especially suitable for technology-driven problem-solving in emerging contexts. It provides a logical progression from problem definition to empirical validation and real-world applicability.

A- Data Source and Context

The study uses transactional and borrower-level data obtained from Ethiopian financial institutions. Notably, due to the absence of a centralised credit bureau and weak data-sharing protocols, the dataset is intentionally restricted and incomplete, accurately simulating the real-world limitations faced by lenders in similar contexts.

This localised focus strengthens the study's practical relevance. However, data confidentiality limits transparency regarding the dataset's structure and size. While privacy is crucial, some additional metadata (e.g., feature types, missing data rates, or dataset size) would enhance reproducibility and rigour.

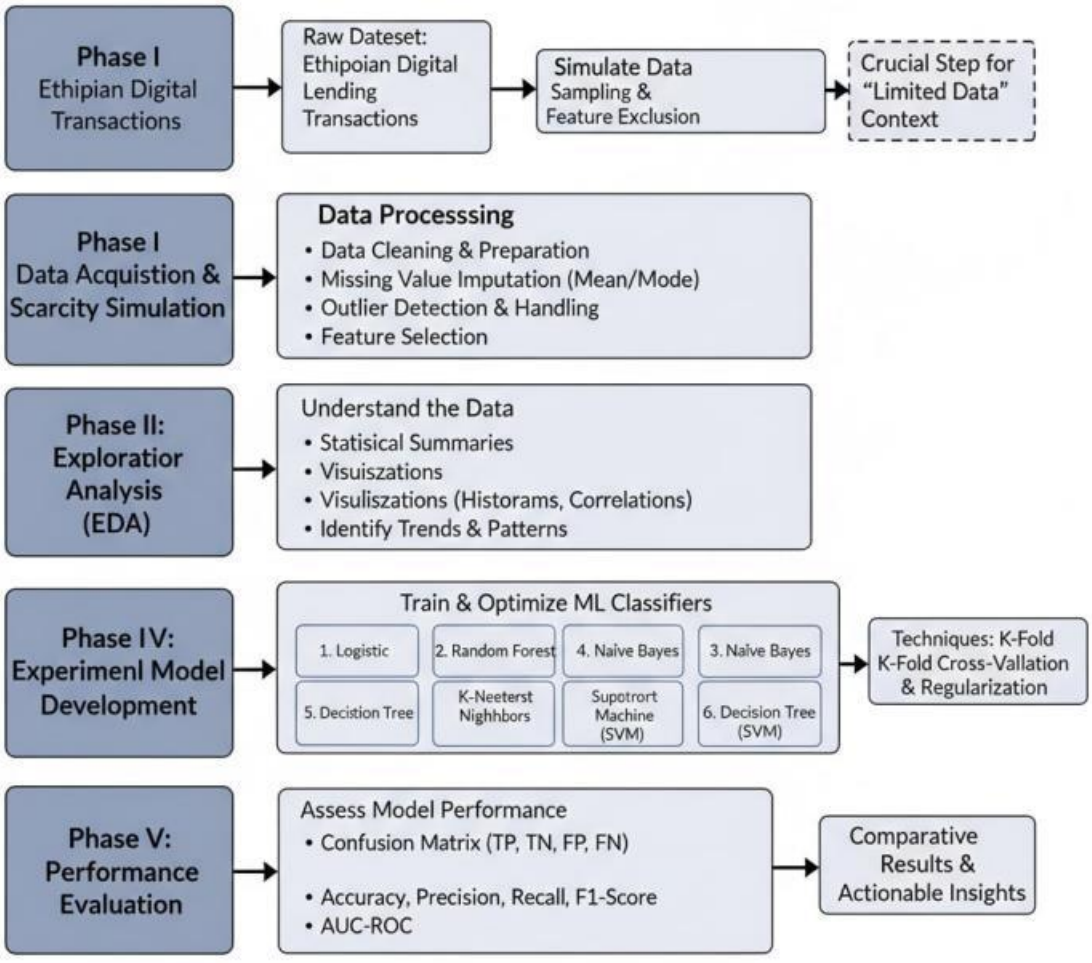

Figure 1- Research methodology pipeline

B- B-Model Selection and Tools

The following six algorithms are identified for comparative evaluation are Logistic Regression, Random Forest, Naïve Bayes, K-Nearest Neighbors, Support Vector Machines

(SVM), and Decision Tree. This mix of models is balanced, including both interpretable (e.g., Logistic Regression, Decision Trees) and ensemble-based approaches (e.g., Random Forest) suitable for understanding performance trade-offs under constrained data regimes.

Python is chosen as the primary analysis tool, using libraries like Scikit-learn, Pandas, NumPy, and Matplotlib/Seaborn for preprocessing, modeling, and visualization. These are industry-standard choices and appropriate for both academic and applied research.

C- Evaluation Metrics and Validation

Performance is assessed using multiple classification metrics including accuracy, precision, recall, F1-Score, AUC (Area Under the Curve) and Confusion Matrix breakdown (TP, TN, FP, FN)

4. The Proposed Scheme

The proposed scheme addresses the challenge of credit risk assessment in the Ethiopian digital lending landscape, characterised by fragmented legacy data and the absence of centralised credit bureau. The framework transitions from raw transactional data to a refined classification output through a multi-stage pipeline designed for high-integrity prediction under data constraints.

To reflect the real-world limitations of the Ethiopian financial sector, the research implements a Constraint Simulation Layer. Unlike studies that rely on massive, high-dimensional datasets, this scheme deliberately utilizes a restricted feature set.

(a) Feature Exclusion: Traditional variables (documented income, formal credit history) are excluded to simulate "thin-file" borrower profiles.

(b) Imbalanced Handling: Given that "Default" events are rare in limited datasets, the scheme prioritizes Mean Imputation for missing values to maintain the statistical power of the small sample without introducing the bias of zero-filling.

To mitigate the "curse of dimensionality" which often leads to overfitting in limited datasets, a rigorous feature selection process is applied. The scheme utilizes dimensionality reduction to identify a subset of transactional footprints (e.g., repayment velocity, digital footprint) that provide the highest information gain. This reduces the noise caused by redundant features common in unstructured Ethiopian digital lending records.

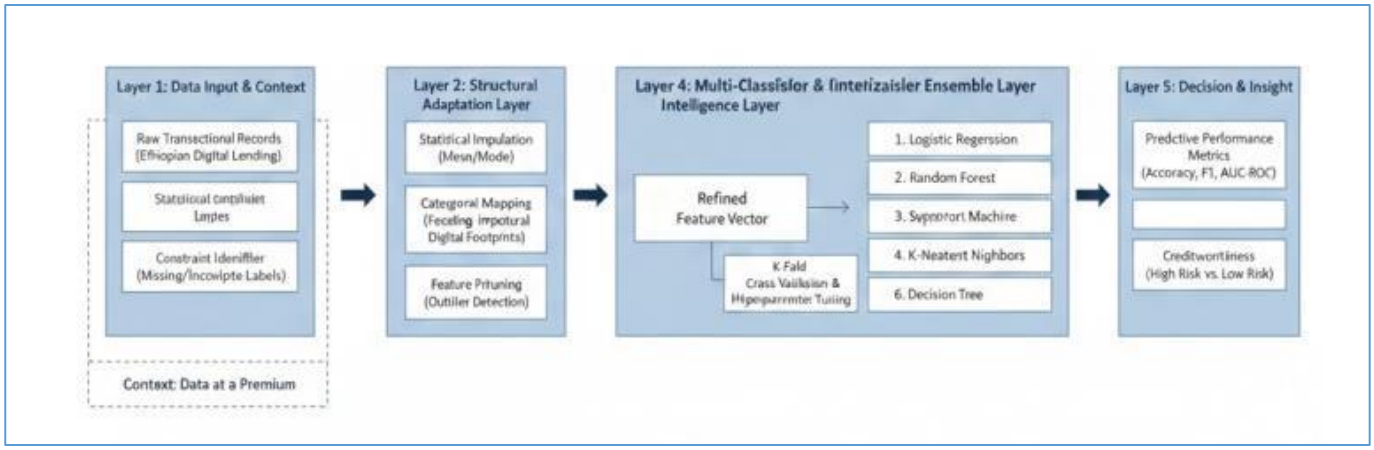

Figure 2- A modular pipeline for data constrained environment

The heart of the proposed scheme is the parallel deployment of six distinct classification architectures. This comparative approach is chosen to identify which mathematical assumptions (e.g., the linearity of Logistic Regression vs. the ensemble nature of Random Forest) hold strongest when data is at a premium.

5. Performance Evaluation

The comprehensive analysis of six classification models applied to the following dataset

```
]: df_loan.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 23 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Month                     1000 non-null   int64
 1   Age                       1000 non-null   int64
 2   Occupation                1000 non-null   object
 3   Annual_Income             1000 non-null   float64
 4   Monthly_Inhand_Salary     1000 non-null   float64
 5   Num_Bank_Accounts         1000 non-null   int64
 6   Num_Credit_Card           1000 non-null   int64
 7   Interest_Rate             1000 non-null   int64
 8   Num_of_Loan               1000 non-null   int64
 9   Delay_from_due_date       1000 non-null   int64
 10  Num_of_Delayed_Payment    1000 non-null   int64
 11  Changed_Credit_Limit      1000 non-null   float64
 12  Num_Credit_Inquiries      1000 non-null   int64
 13  Credit_Mix                1000 non-null   object
 14  Outstanding_Debt          1000 non-null   float64
 15  Credit_Utilization_Ratio  1000 non-null   float64
 16  Credit_History_Age        1000 non-null   int64
 17  Payment_of_Min_Amount     1000 non-null   object
 18  Total_EMI_per_month       1000 non-null   float64
 19  Amount_invested_monthly   1000 non-null   float64
 20  Payment_Behaviour         1000 non-null   object
 21  Monthly_Balance           1000 non-null   float64
 22  Credit_Score              1000 non-null   object
dtypes: float64(8), int64(10), object(5)
memory usage: 179.8+ KB
```

Figure 3- Information on dataset

The result has indicated that the Random Forest model is the most suitable choice for deployment in the given credit scoring task. With an accuracy of 82.50%, alongside strong precision (82.38%), recall (83.15%), and F1-score (82.74%), Random Forest demonstrates a well-balanced ability to minimize both false positives and false negatives. This balance is critical for real-world financial decision-making, where both types of errors carry significant consequences. Furthermore, the model's robustness in handling complex feature interactions and its resilience to overfitting, especially in contexts with limited labeled data, further justify its recommendation.

In scenarios where model interpretability is a priority, the Decision Tree model emerges as a compelling alternative. Matching Random Forest's accuracy (82.50%) and exhibiting a slightly higher precision (83.69%), the Decision Tree's transparent decision-making process makes it easier to communicate insights to stakeholders and domain experts. However, caution is warranted due to its susceptibility to overfitting without proper regularization, which is particularly relevant when working with sparse datasets.

The K-Nearest Neighbors (KNN) model also showed promising results, achieving an accuracy of 81.50% and an F1-score of 82.62%. Despite this, its computational inefficiency during prediction phases, especially in larger datasets or real-time systems, limits its practical applicability relative to Random Forest and Decision Tree models. Similarly, the Support Vector Machine (SVM) model delivered acceptable but comparatively lower performance metrics and carries computational overheads that may hinder deployment.

Logistic Regression, while appreciated for its simplicity and interpretability, yielded modest results with an accuracy of 69.50% and F1-score of 69.44%, making it less suited for the complexities of this credit scoring task. The Naive Bayes model performed poorly, with an accuracy of only 22.00% and an F1-score of 18.36%, likely due to the violation of its core assumption of feature independence, rendering it unsuitable for the dataset at hand.

	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.69500	0.71788	0.67786	0.69435
Random Forest	0.82500	0.82379	0.83146	0.82741
Naive Bayes	0.22000	0.23295	0.36293	0.18362
K-Nearest Neighbors	0.81500	0.82072	0.83223	0.82619
Support Vector Machine	0.81000	0.80562	0.81755	0.81112
Decision Tree	0.82500	0.83692	0.81052	0.82203

Figure 4-Classification algorithms prediction accuracy

Given the constraints posed by limited labeled data, the adaptability and strong performance of Random Forest and Decision Tree models underscore their value in similar financial contexts, particularly in emerging markets where data scarcity is common. For

future research, exploring advanced feature selection techniques and comprehensive hyperparameter tuning could unlock further improvements in predictive accuracy and model efficiency. Additionally, investigating ensemble methods or hybrid approaches that combine the strengths of multiple algorithms may offer promising avenues to enhance credit scoring performance under data limitations.

In conclusion, the Random Forest model stands out as the primary candidate for operational deployment due to its balanced accuracy, robustness, and suitability for limited data environments. The Decision Tree model remains a viable alternative where transparency and interpretability are essential. Both models provide a strong foundation for building reliable credit scoring systems that can improve lending decisions and foster financial inclusion despite data constraints.

6. Conclusion

This research investigated the application of machine learning techniques for credit scoring in environments characterised by limited labelled data, with a specific focus on the emerging digital lending sector in Ethiopia. Through a comparative analysis of six classification algorithms, the study identified the Random Forest model as the most effective and reliable approach for predicting creditworthiness under data constraints. The model's strong performance across key evaluation metrics demonstrates its robustness and practical suitability for real-world financial decision-making where data scarcity is a challenge.

The Decision Tree model also showed promise, particularly in scenarios demanding model interpretability, highlighting the importance of balancing accuracy with transparency in credit scoring applications. Other algorithms, including K-Nearest Neighbours, Support Vector Machines, Logistic Regression, and Naive Bayes, were less effective or faced practical limitations within the context of limited data availability.

Overall, this research contributes valuable insights to the field of financial inclusion by providing a data-driven framework that can assist lenders in making more informed and reliable credit decisions despite incomplete or sparse datasets. By leveraging machine learning, especially ensemble methods like Random Forest, financial institutions can better manage risk and improve lending outcomes, fostering greater access to credit in developing economies.

Future work should focus on refining these models through advanced feature selection, hyperparameter optimisation, and exploring hybrid approaches to further enhance credit scoring accuracy. As digital lending continues to grow, continued innovation in handling limited data will be critical to supporting sustainable and inclusive financial ecosystems.

References

[1] Tripathi, D., Edla, D.R., Bablani, A., Shukla, A.K. and Reddy, B.R., 'Experimental analysis of machine learning methods for credit score classification', *Progress in Artificial Intelligence*, 10, pp. 1-15, 2021.

[2] Yotsawat, W., Wattuya, P. and Srivihok, A., 'A novel method for credit scoring based on cost-sensitive neural network ensemble', *International Journal of Intelligent Systems and Applications*, 13(4), pp. 1-12, 2021.

[3] Liu, Y., *A framework of data mining application process for credit scoring*. Institut fur Wirtschaftsinformatik, 2022.

[4] World Bank Group, *Credit scoring approaches guidelines*. Washington, DC: World Bank, 2016.

[5] Moin, K.I. and Ahmed, Q.B., 'Use of data mining in banking', *International Journal of Engineering Research and Applications (IJERA)*, 2(2), pp. 738-742, 2012.

[6] Maria, F.V. and Fernando, B., *Credit scoring in financial inclusion*. Washington, DC: CGAP, 2019.

[7] Qi, G.-J. and Luo, J., 'Small data challenges in big data era: a survey of recent progress on unsupervised and semi-supervised methods', *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[8] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 'ImageNet classification with deep convolutional neural networks', *Advances in Neural Information Processing Systems*, 25, pp. 1097-1105, 2012.

[9] Liu, H. and Motoda, H., *Feature selection for knowledge discovery and data mining*. Boston: Kluwer Academic Publishers, 1998.

[10] Ha, S. and Nguyen, H.N., 'Credit scoring with a feature selection approach based on deep learning', *MATEC Web of Conferences*, 54, p. 05004, 2016.

[11] Jukna, D., 'New challenges in economic and business development', in *Proceedings of the 14th International Scientific Conference*. Riga: University of Latvia, 2022.

[12] Abdou, H.A. and Pointon, J., 'Credit scoring, statistical techniques and evaluation criteria: a review of the literature', *Intelligent Systems in Accounting, Finance & Management*, 18(2-3), pp. 59-88, 2011.

[13] Fausett, L.V., *Fundamentals of neural networks: architectures, algorithms and applications*. 1st edn. London: Pearson, 1993.

[14] Noh, H.J., Roh, T.H. and Han, I., 'Prognostic personal credit risk model considering censored information', *Expert Systems with Applications*, 28(4), pp. 753-762, 2005.

[15] Wah, Y.B. and Ibrahim, I.R., 'Using data mining predictive models to classify credit card applicants', in *Proceedings of the IEEE International Conference*. Seoul: IEEE, 2010.

[16] Ray, S., *Understanding support vector machine (SVM) algorithm from examples, 2017*. Available at: https://www.analyticsvidhya.com (Accessed: 22 December 2024).

[17] Ghodselahi, A., 'A hybrid support vector machine ensemble model for credit scoring', *International Journal of Computer Applications*, 17(5), pp. 1-5, 2011.

[18] Yiu, T., *Understanding random forest, 2019*. Available at: https://towardsdatascience.com (Accessed: 22 December 2024).

[19] Breiman, L., 'Random forests', *Machine Learning*, 45(1), pp. 5-32, 2001.

[20] Soria, D., Garibaldi, J.M., Ambrogi, F., Biganzoli, E.M. and Ellis, I.O., 'A non-parametric version of the naive Bayes classifier', *Knowledge-Based Systems*, 24(6), pp. 775-784, 2011.

[21] Vashistha, A., Anderson, R. and Mare, S., 'Examining security and privacy research in developing regions', in *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. New York: ACM, 2018.

[22] Federal Reserve, *Report to Congress on credit scoring, 2007*. Available at: https://www.federalreserve.gov (Accessed: 22 December 2024).

[23] World Bank, *World development report 2016: digital dividends*. Washington, DC: World Bank Group.

[24] Intellias, *How alternative credit data can increase accuracy*. Available at: https://intellias.com (Accessed: 22 December 2024).

[25] Algoscale, *How data analytics is revolutionising lending decisions, 2023*. Available at: https://algoscale.com (Accessed: 22 December 2024).

[26] Okesola, O.J., Ojo, A.A., Adekunle, Y.A. and Arowolo, M.O., 'An improved bank credit scoring model: a naive Bayesian approach', in *Proceedings of the International Conference on Computational Science and Computational Intelligence (CSCI)*. Las Vegas: IEEE, pp. 228-233, 2017.

[27] Kumar, A., 'Credit score prediction system using deep learning', *Journal of Physics: Conference Series*. IOP Publishing, 2021.

[28] Laborda, J. and Ryoo, S., 'Feature selection in a credit scoring model', *Mathematics*, 9(7), p. 746, 2021.

[29] Alshareef, A. and Alshamsi, A., 'A systematic review of explainable artificial intelligence (XAI) in credit scoring', *IEEE Access*, 11, pp. 11245-11260, 2023.

[30] Liu, Z. and Wang, H., 'Handling data scarcity in credit scoring: a transfer learning approach', *Expert Systems with Applications*, 238, p. 121855, 2024.

[31] Chen, X., 'Hybrid ensemble learning frameworks for credit risk management in the digital banking era', *International Journal of Finance & Data Science*, 11(2), pp. 88-104, 2025.

[32] Zhao, L., Wang, X., Zhang, R. and Patel, S., 'Federated learning for privacy-preserving credit scoring: challenges and opportunities', *Digital Finance Journal*, 12(1), pp. 15-33, 2024.