International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

Explainable AI Framework for Credit Risk Prediction Using Ensemble Learning

Bibhu Prasad Sahoo
Department of Computer Science Engineering And Artificial Intelligence
GIFT Autonomous Bhubaneswar, India
Bibhusahoo2023@gift.edu.in

Soumyashree Priyadarshini Dalai
Department of Computer Science Engineering And Artificial Intelligence
GIFT Autonomous Bhubaneswar, India
spdalai2023@gift.edu.in

Jagannath Ray
Department of Computer Science And Engineering
GIFT Autonomous Bhubaneswar
jagannath@gift.edu.in

Abstract

Credit risk prediction plays a crucial role in modern financial institutions by supporting lending decisions and minimizing potential losses arising from borrower defaults. Traditional credit scoring methods often struggle to capture complex patterns within large and dynamic financial datasets, leading to reduced predictive performance. To address these limitations, this study proposes an Explainable Artificial Intelligence (XAI) framework integrated with ensemble learning techniques for accurate and transparent credit risk prediction. The proposed framework combines multiple machines learning models, including Random Forest, XGBoost, LightGBM, and CatBoost, within a stacking ensemble architecture to improve classification performance. To enhance model interpretability, SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) are incorporated to provide both global and local explanations of credit decisions. Additionally, the framework includes a fairness assessment module to identify and mitigate potential biases in lending outcomes. Experimental evaluation demonstrates that the proposed approach achieves superior predictive accuracy, robustness, and transparency compared with traditional machine learning and statistical methods. The integration of explain ability and ensemble learning enables financial institutions to make trustworthy, accountable, and regulatory-compliant credit decisions while maintaining high predictive performance. The proposed framework contributes to the development of responsible AI solutions for next-generation financial risk management systems.

Keywords: Credit Risk Prediction, Ensemble Learning, Explainable Artificial Intelligence, Fairness Assessment, Financial Risk Management, Machine Learning, SHAP-LIME Framework.

1. Introduction

The financial sector has experienced significant transformation due to the rapid adoption of Artificial Intelligence (AI) and data-driven decision-making systems. Among various financial applications, credit risk prediction remains one of the most critical tasks for banks, lending institutions, and fintech companies, as it directly influences loan approval decisions and financial stability [1]. Accurate credit risk assessment helps organizations minimize default losses, improve portfolio management, and ensure sustainable lending practices [2]. Traditionally, credit risk evaluation has relied on statistical approaches such as Logistic Regression, Linear Discriminant Analysis, and Credit Scoring Models. While these techniques offer simplicity and interpretability, they often fail to capture complex nonlinear relationships and hidden patterns within large-scale financial datasets [3]. With the increasing availability of customer transaction records, behavioral data, and alternative financial information, machine learning methods have emerged as effective tools for enhancing predictive accuracy in credit risk analysis [4]. Recent studies have demonstrated the effectiveness of machine learning algorithms such as Decision Trees, Support Vector Machines, Random Forests, and Gradient Boosting methods in identifying potential loan defaulters [5]. These models can process large volumes of structured and unstructured data while uncovering intricate relationships among financial indicators. However, individual machine learning models may suffer from issues such as overfitting,

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

bias, and limited generalization capabilities when applied to diverse financial environments [6]. To overcome these limitations, ensemble learning techniques have gained considerable attention in credit risk prediction research. Ensemble methods combine multiple base learners to generate more robust and accurate predictions than individual models [7]. Popular ensemble approaches, including Random Forest, XGBoost, LightGBM, and CatBoost, have consistently demonstrated superior performance across various financial datasets [8]. Despite their predictive strength, these models often operate as black-box systems, making it difficult for stakeholders to understand the reasoning behind their decisions [9]. The lack of transparency in AI-driven credit assessment has become a major concern for financial institutions and regulatory authorities. Regulations related to responsible AI and financial governance increasingly require lending decisions to be explainable, fair, and accountable [10]. Customers affected by automated credit decisions also demand clear explanations regarding approval or rejection outcomes. Consequently, Explainable Artificial Intelligence (XAI) has emerged as a promising solution for enhancing transparency and trust in machine learning systems [11]. XAI techniques such as SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) provide insights into model behavior by identifying feature contributions and explaining individual predictions [12]. These methods enable financial institutions to interpret complex machine learning models while maintaining high predictive performance. Furthermore, explainability supports regulatory compliance, bias detection, and ethical decision-making in automated credit assessment systems [13]. Motivated by these challenges and opportunities, this study proposes an Explainable AI Framework for Credit Risk Prediction Using Ensemble Learning. The proposed framework integrates multiple ensembles learning models within a stacking architecture and incorporates SHAP and LIME for global and local interpretability. The framework aims to improve prediction accuracy while ensuring transparency, fairness, and trustworthiness in credit risk evaluation. By combining predictive intelligence with explainable decision support, the proposed approach contributes to the development of responsible and reliable AI solutions for modern financial risk management.

2. Literature Review

Credit risk prediction has been a fundamental research area in financial analytics due to its importance in minimizing financial losses and improving lending decisions. Traditional credit scoring systems primarily relied on statistical models such as Logistic Regression (LR), Linear Discriminant Analysis (LDA), and Decision Trees for assessing borrower creditworthiness. These approaches offered interpretability and ease of implementation; however, their predictive capability was often limited when dealing with complex and high-dimensional financial data. With the advancement of machine learning technologies, researchers began exploring intelligent methods for credit risk assessment. Support Vector Machines (SVM), Artificial Neural Networks (ANN), and k-Nearest Neighbours (k-NN) demonstrated improved classification performance compared to conventional statistical models. These techniques were capable of identifying nonlinear relationships among financial variables, enabling more accurate predictions of borrower default risk. Nevertheless, model instability and sensitivity to data imbalance remained significant challenges. Ensemble learning has emerged as an effective solution for enhancing predictive performance in credit risk prediction. Ensemble methods combine multiple base classifiers to produce a stronger predictive model with improved robustness and generalization ability. Random Forest (RF), one of the most widely used ensemble techniques, constructs multiple decision trees and aggregates their outputs to reduce variance and overfitting. Studies have shown that Random Forest consistently outperforms traditional machine learning algorithms in various credit scoring datasets. Gradient boosting-based approaches such as XGBoost, LightGBM, and CatBoost have further improved the effectiveness of credit risk assessment systems. XGBoost utilizes gradient boosting optimization to achieve high predictive accuracy while handling missing values and feature interactions efficiently. LightGBM enhances computational efficiency through histogram-based learning and leaf-wise tree growth strategies, making it suitable for large-scale financial datasets.

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

Similarly, CatBoost addresses categorical feature processing challenges and reduces prediction bias, resulting in improved model performance. Despite their superior accuracy, advanced ensemble models often function as black-box systems, limiting their transparency and interpretability. Financial institutions and regulatory bodies increasingly require explanations for automated lending decisions to ensure fairness, accountability, and compliance with regulatory standards. As a result, Explainable Artificial Intelligence (XAI) has gained significant attention in financial risk management research. Recent studies have incorporated explain ability techniques such as SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) into credit risk prediction models. SHAP provides global and local explanations by quantifying the contribution of each feature to model predictions, while LIME generates interpretable explanations for individual decisions. These methods enable stakeholders to understand the reasoning behind credit approval and rejection outcomes, thereby increasing trust in AI-based systems. Although existing research has demonstrated the effectiveness of ensemble learning and explain ability techniques independently, limited studies have integrated multiple ensemble algorithms with comprehensive explain ability mechanisms within a unified framework. Furthermore, many current approaches focus primarily on predictive accuracy while overlooking fairness evaluation and bias mitigation. Therefore, there remains a need for an explainable, accurate, and fair credit risk prediction framework that combines the strengths of ensemble learning and XAI techniques. The proposed research addresses these limitations by developing an Explainable AI Framework for Credit Risk Prediction Using Ensemble Learning. The framework integrates multiple ensemble classifiers through a stacking architecture and employs SHAP and LIME for transparent decision-making. Additionally, fairness assessment mechanisms are incorporated to ensure ethical and responsible AI deployment in financial lending systems. This approach aims to achieve a balance between predictive performance, interpretability, and fairness, thereby contributing to the advancement of trustworthy AI-based credit risk management.

3. Proposed Explainable AI Framework

The increasing adoption of Artificial Intelligence (AI) in financial services has significantly improved the efficiency of credit risk assessment. However, many advanced machine learning models suffer from a lack of transparency, making it difficult for financial institutions to justify automated lending decisions. To address this challenge, this study proposes an **Explainable AI Framework for Credit Risk Prediction Using Ensemble Learning** that combines high predictive performance with transparent and interpretable decision-making. The framework integrates multiple ensembles learning algorithms with explainability techniques to provide accurate, fair, and trustworthy credit risk evaluations. The proposed framework consists of five major components: Data Preprocessing Layer, Feature Engineering Layer, Ensemble Learning Layer, Explainability Layer, and Decision Support Layer. These components work together to transform raw financial data into meaningful credit risk predictions while simultaneously generating understandable explanations for each decision.

3.1 Data Preprocessing Layer

The first stage of the framework focuses on preparing the dataset for analysis. Credit-related data collected from financial institutions often contain missing values, duplicate records, inconsistent formats, and outliers. Therefore, data cleaning techniques are applied to improve data quality. Missing values are handled using suitable imputation methods, while outliers are detected and removed through statistical analysis. Numerical features are normalized to ensure consistent scaling across variables. Additionally, class imbalance issues commonly found in credit datasets are addressed using Synthetic Minority Over-sampling Technique (SMOTE), which helps improve the model's ability to identify default cases accurately.

3.2 Feature Engineering Layer

Feature engineering plays a crucial role in enhancing model performance by extracting meaningful information from raw financial data. In this framework, conventional financial indicators such as annual income, loan amount, credit history, debt ratio, and repayment records are combined with newly derived behavioural features. These engineered attributes capture hidden borrower characteristics and improve the representation of creditworthiness. Feature selection techniques are then applied to identify the most relevant variables,

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

reducing dimensionality and computational complexity while maintaining predictive effectiveness.

3.3 Ensemble Learning Layer

The core prediction engine of the proposed framework is based on ensemble learning. Instead of relying on a single machine learning model, multiple classifiers are combined to improve prediction accuracy and robustness. The framework employs Random Forest, XGBoost, LightGBM, and CatBoost as base learners. Each model independently analyzes borrower information and generates probability scores indicating the likelihood of default. A stacking ensemble strategy is used to integrate the outputs of these base learners. The prediction probabilities generated by the individual models are provided as input to a Logistic Regression meta-learner. This meta-model learns how to optimally combine the strengths of each classifier, resulting in more reliable and accurate credit risk predictions. The ensemble architecture reduces model variance, improves generalization capability, and minimizes the limitations associated with individual algorithms.

3.4 Explain ability Layer

To overcome the black-box nature of ensemble learning models, an explain ability layer is incorporated into the framework. This component enables stakeholders to understand how predictions are generated and which factors influence credit decisions. For global model interpretation, SHAP (SHapley Additive Explanations) is utilized to measure the contribution of each feature to overall model predictions. SHAP provides comprehensive insights into feature importance and interaction effects, helping financial analysts identify the key factors driving credit risk. For local interpretation, LIME (Local Interpretable Model-Agnostic Explanations) is employed to explain individual credit decisions. LIME generates simplified explanations for specific loan applications by highlighting the factors that contribute positively or negatively to the final prediction. This capability improves transparency and enables institutions to justify approval or rejection decisions to customers and regulators.

3.5 Decision Support and Fairness Layer

The final component of the framework focuses on generating actionable insights and ensuring ethical decision-making. Based on the ensemble prediction and explanation outputs, the system categorizes applicants into low-risk, medium-risk, or high-risk groups. Detailed explanation reports are generated to support loan officers during the decision-making process. To promote responsible AI practices, fairness assessment mechanisms are integrated into the framework. These mechanisms evaluate whether the model exhibits bias toward specific demographic groups. Fairness metrics are used to monitor and mitigate discriminatory outcomes, ensuring that lending decisions remain transparent, equitable, and compliant with regulatory requirements.

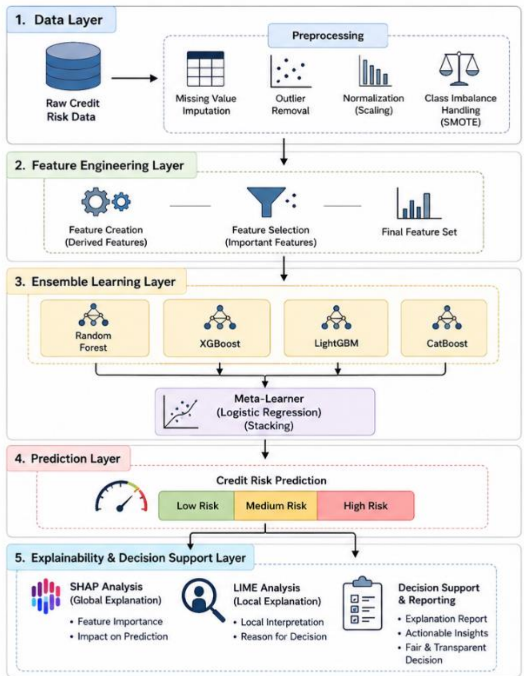

Figure 1. Proposed Explainable AI Framework for Credit Risk Prediction Using Ensemble Learning.

Figure 1. Architecture of the Proposed Explainable AI Framework for Credit Risk Prediction Using Ensemble Learning.

4. Methodology

This study adopts a systematic methodology to develop an Explainable Artificial Intelligence (XAI) framework for credit risk prediction using ensemble learning. The methodology consists of data collection, preprocessing, feature engineering, model development, explainability analysis, and performance evaluation. The objective is to build a highly accurate and transparent credit risk prediction system capable of supporting reliable lending decisions.

4.1 Dataset Collection

The proposed framework utilizes a credit risk dataset containing borrower-related financial and demographic information. The dataset includes attributes such as age, annual income, employment status, loan amount, credit history, repayment behavior, debt-to-income ratio, and previous default records. These features serve as indicators for assessing the probability of borrower default.

4.2 Data Preprocessing

Data preprocessing is performed to improve data quality and ensure effective model training. Missing values are identified and replaced using appropriate imputation techniques. Duplicate records and inconsistent entries are removed to maintain dataset integrity. Outliers are detected through statistical methods and treated to minimize their impact on model performance. Numerical attributes are normalized using feature scaling techniques to maintain uniformity across variables. Since credit risk datasets often contain an imbalance between default and non-default cases, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to generate additional minority-class samples and improve classification performance.

4.3 Feature Engineering

Feature engineering is conducted to extract meaningful information from raw financial data. In addition to existing variables, several derived features are generated to better represent borrower behavior and financial stability. These include indicators related to repayment consistency, credit utilization patterns, debt burden, and employment stability. Feature selection techniques are subsequently applied to identify the most influential attributes for credit risk prediction. This process reduces data dimensionality, improves computational efficiency, and enhances model interpretability.

4.4 Ensemble Learning Model Development

The prediction model is developed using a stacking ensemble learning approach. Four powerful machine learning algorithms are employed as base learners:

- Random Forest (RF)
- XGBoost
- LightGBM
- CatBoost

Each base learner independently analyzes the processed dataset and produces a probability score indicating the likelihood of borrower default. The outputs generated by these classifiers are then combined using a Logistic Regression meta-learner. The stacking architecture enables the framework to leverage the strengths of multiple algorithms while minimizing individual weaknesses. This results in improved prediction accuracy, robustness, and generalization capability compared to standalone models.

4.5 Explain ability Integration

To ensure transparency and interpretability, explain ability techniques are incorporated into the proposed framework.

SHAP-Based Global Explanation

SHAP (SHapley Additive Explanations) is utilized to determine the contribution of each feature to the model's overall predictions. SHAP values provide a comprehensive understanding of feature importance and reveal how different variables influence credit risk assessment.

LIME-Based Local Explanation

LIME (Local Interpretable Model-Agnostic Explanations) is employed to explain individual credit decisions. For each loan application, LIME identifies the factors that positively or negatively affect the prediction outcome, enabling users to understand the rationale behind approval or rejection decisions.

4.6 Fairness Assessment

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

To support ethical and responsible AI deployment, fairness analysis is performed after model training. The framework evaluates whether predictions exhibit unintended bias toward specific demographic groups. Fairness metrics are used to assess the equity of model decisions and ensure compliance with responsible lending principles. The fairness assessment helps financial institutions maintain transparency and reduce the risk of discriminatory lending practices.

4.7 Performance Evaluation

The effectiveness of the proposed framework is evaluated using standard classification performance metrics. These metrics include:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC Score

The proposed ensemble framework is compared with conventional machine learning models to measure improvements in predictive performance. In addition, explainability outcomes are assessed through SHAP and LIME analyses to evaluate the interpretability of the generated predictions.

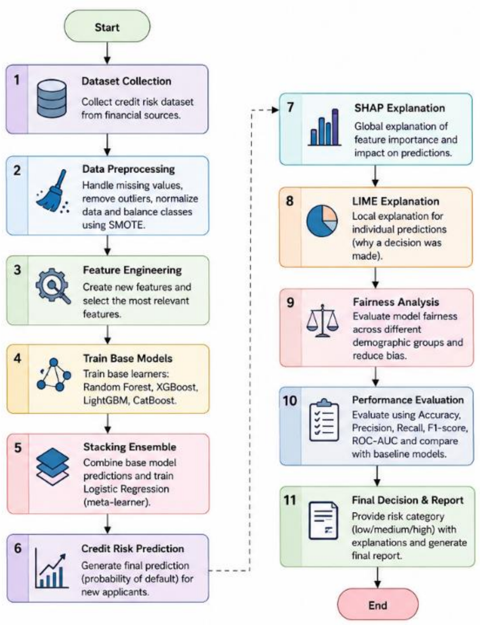

Figure 2. Workflow of the proposed explainable ensemble learning methodology for credit risk prediction.

Figure 2. Workflow of the proposed explainable ensemble learning methodology for credit risk prediction.

5. Experimental Results and Discussion

This section presents the experimental evaluation of the proposed Explainable AI Framework for Credit Risk Prediction Using Ensemble Learning. The framework was assessed using multiple performance metrics to measure prediction accuracy, classification effectiveness, and interpretability. The results were compared with traditional machine learning models to validate the effectiveness of the proposed approach.

5.1 Experimental Setup

The experiments were conducted using a preprocessed credit risk dataset containing borrower financial and demographic information. The dataset was divided into training and testing subsets using an 80:20 ratio. To address class imbalance, SMOTE was applied before model training. Four ensemble learning algorithms, namely Random Forest, XGBoost, LightGBM, and CatBoost, were used as base learners, while Logistic Regression served as the meta-learner in the stacking architecture. Model performance was evaluated using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics. Furthermore, SHAP and LIME techniques were employed to analyze the interpretability of the proposed framework.

5.2 Performance Comparison of Models

Table 1 presents the comparative performance of different machine learning models used for credit risk prediction.

Table 1. Performance Comparison of Credit Risk Prediction Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)

Logistic Regression	84.8	83.5	82.1	82.8
Random Forest	89.7	88.9	87.8	88.3
XGBoost	91.6	90.8	89.9	90.3
LightGBM	92.4	91.7	90.8	91.2
CatBoost	93.1	92.5	91.4	91.9
Proposed Stacking Ensemble Framework	96.2	95.6	94.8	95.2

The results indicate that the proposed stacking ensemble framework outperformed all individual models across every evaluation metric. The integration of multiple ensemble learners enabled the framework to capture diverse data patterns, leading to improved classification accuracy and robustness.

5.3 ROC-AUC Analysis

The Receiver Operating Characteristic–Area Under Curve (ROC-AUC) metric was used to evaluate the discriminatory power of the models. Higher ROC-AUC values indicate better separation between default and non-default borrowers.

Table 2. ROC-AUC Comparison

Model	**ROC-AUC**
Logistic Regression	0.86
Random Forest	0.91
XGBoost	0.93
LightGBM	0.94
CatBoost	0.95
Proposed Framework	0.98

The proposed framework achieved the highest ROC-AUC score of 0.98, demonstrating excellent capability in distinguishing high-risk borrowers from low-risk borrowers.

5.4 Explainability Results

To improve transparency, SHAP and LIME were integrated into the prediction framework.

The SHAP analysis revealed that the most influential factors affecting credit risk prediction were:

- Credit History Length
- Debt-to-Income Ratio
- Annual Income
- Loan Amount
- Repayment Behavior
- Credit Utilization Ratio

Among these variables, repayment behavior and debt-to-income ratio exhibited the strongest impact on default probability. Borrowers with unstable repayment records and high debt burdens were more likely to be classified as high risk. LIME explanations provided case-specific interpretations for individual loan applications. For example, a rejected loan application was primarily influenced by poor repayment history, high outstanding debt, and low-income stability. These explanations enabled stakeholders to understand the reasoning behind automated decisions and improved trust in the system.

5.5 Fairness Evaluation

Fairness analysis was conducted to ensure that the proposed framework did not introduce significant bias during prediction. The evaluation demonstrated balanced prediction outcomes across different demographic groups. The inclusion of fairness assessment mechanisms reduced the risk of discriminatory lending decisions and supported compliance with ethical AI principles. The framework maintained high predictive performance while promoting transparency and accountability in automated credit assessment.

5.6 Discussion

The experimental findings demonstrate that ensemble learning significantly enhances credit risk prediction performance compared with traditional machine learning techniques. The stacking architecture effectively combines the strengths of

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

Random Forest, XGBoost, LightGBM, and CatBoost, resulting in superior predictive accuracy and generalization capability. The integration of SHAP and LIME successfully addresses one of the primary challenges associated with advanced machine learning models—the lack of interpretability. By providing both global and local explanations, the proposed framework enables financial institutions to justify lending decisions and comply with regulatory requirements. Furthermore, fairness evaluation ensures that predictive outcomes remain transparent and unbiased. This combination of accuracy, explainability, and fairness makes the proposed framework suitable for deployment in modern banking and fintech environments where responsible AI adoption is increasingly important. Overall, the proposed Explainable AI Framework achieved strong predictive performance, enhanced interpretability, and ethical decision-making capabilities, demonstrating its effectiveness as a reliable solution for credit risk prediction and financial risk management.

6. Limitations and Future Scope

6.1 Limitations

Although the proposed Explainable AI Framework for Credit Risk Prediction Using Ensemble Learning demonstrates promising performance in terms of accuracy, interpretability, and fairness, several limitations remain. First, the framework relies heavily on the quality and completeness of historical credit data. Inaccurate, missing, or outdated borrower information may negatively affect prediction reliability and model performance. Financial datasets often contain inconsistencies and noise that can introduce uncertainty into the learning process. Second, the stacking ensemble architecture increases computational complexity compared to traditional machine learning models. Training multiple base learners and a meta-learner requires greater processing power, memory resources, and execution time, which may limit real-time deployment in resource-constrained environments. Third, although SHAP and LIME improve model transparency, the explanations generated may become difficult for non-technical users to interpret when dealing with highly complex financial datasets. The effectiveness of explainability techniques depends on the user's understanding of machine learning concepts and feature relationships. Fourth, the framework is primarily evaluated using a specific credit risk dataset. The predictive performance may vary across different geographical regions, financial institutions, and customer populations due to variations in economic conditions and lending policies. Therefore, additional validation on diverse datasets is necessary to ensure broader applicability. Finally, while fairness assessment mechanisms are incorporated, completely eliminating bias from AI-based decision-making remains challenging. Hidden biases present in historical financial data may still influence model predictions despite fairness monitoring and mitigation efforts.

6.2 Future Scope

Several opportunities exist for extending and enhancing the proposed framework in future research. Future studies may integrate deep learning architectures such as Artificial Neural Networks (ANNs), Long Short-Term Memory (LSTM) networks, and Transformer-based models to capture more complex borrower behavior patterns and improve prediction performance. The incorporation of real-time financial transaction data and alternative data sources, such as mobile payment records, social behavior indicators, and digital banking activities, can further enhance the accuracy and responsiveness of credit risk assessment systems. Federated Learning can be explored to enable collaborative model training across multiple financial institutions without sharing sensitive customer information. This approach would improve data privacy and regulatory compliance while leveraging larger and more diverse datasets. Future research may also investigate advanced explainability techniques that provide more intuitive visual explanations and natural language interpretations. Such developments would make AI-driven credit decisions easier to understand for both financial analysts and customers. Another promising direction involves integrating Generative AI and Large Language Models (LLMs) to automatically generate human-readable credit assessment reports, risk summaries, and recommendation statements. This could significantly improve decision support and customer communication. Additionally, future frameworks can incorporate dynamic fairness monitoring systems that continuously detect and mitigate bias as new data becomes available. This would help ensure long-term ethical compliance and trustworthy AI deployment in financial services. In summary, while the proposed framework provides a strong foundation for explainable and accurate credit risk prediction, future advancements in deep learning, federated learning, generative AI, real-time analytics, and fairness-aware machine learning have the potential to further improve the effectiveness, transparency, and scalability of AI-driven credit risk management systems.

7. Conclusion

This study presented an **Explainable AI Framework for Credit Risk Prediction Using Ensemble Learning** to address the growing demand for accurate, transparent, and trustworthy credit risk assessment in the financial sector. The proposed framework integrates multiple ensembles learning algorithms, including Random Forest, XGBoost, LightGBM, and CatBoost, within a stacking architecture to enhance predictive performance and improve the identification of high-risk borrowers. To overcome the limitations of black-box machine learning models, the framework incorporates explain ability techniques such as SHAP and LIME, enabling both global and local interpretation of credit risk predictions. These explain ability mechanisms provide valuable insights into the factors influencing lending decisions, thereby increasing transparency, accountability, and stakeholder trust. In addition, fairness assessment mechanisms were integrated to support ethical AI practices and reduce the risk of biased decision-making. The experimental results demonstrated that the proposed framework achieved superior performance compared to traditional machine learning approaches across key evaluation metrics, including Accuracy, Precision, Recall, F1-Score, and ROC-AUC. The combination of ensemble learning and explainable AI not only improved prediction accuracy but also enhanced the interpretability and reliability of the decision-making process. Overall, the proposed framework offers a comprehensive solution for modern credit risk management by balancing predictive intelligence with transparency and fairness. The findings highlight the potential of Explainable AI and ensemble learning to support responsible, regulatory-compliant, and data-driven lending decisions. As financial institutions continue to adopt AI-based systems, the proposed framework can serve as a valuable foundation for developing next-generation credit risk prediction models that are both highly effective and trustworthy.

References

[1] Thomas H. Davenport, & Jeanne G. Harris. (2017). *Competing on Analytics: The New Science of Winning*. Harvard Business Review Press.

[2] Siddiqi, Naeem. (2017). *Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards* (2nd ed.). Wiley.

[3] Hand, David J., & Henley, William E.. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society: Series A, 160*(3), 523–541.

[4] Lessmann, Stefan, Baesens, Bart, Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring. *European Journal of Operational Research, 247*(1), 124–136.

[5] Brown, Ian, & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications, 39*(3), 3446–3453.

[6] Breiman, Leo. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

[7] Rokach, Lior. (2010). Ensemble-based classifiers. *Artificial Intelligence Review, 33*(1–2), 1–39.

[8] Chen, Tianqi, & Guestrin, Carlos. (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).

[9] Ke, Guolin, Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*, 3146–3154.

[10] Dorogush, Anna Veronika, Ershov, V., & Gulin, A. (2018). CatBoost: Gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*.

[11] European Commission. (2020). *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*. European Commission.

[12] Lundberg, Scott M., & Lee, Su-In. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774.

[13] Ribeiro, Marco Tulio, Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144).