Credit Risk Prediction with Explainable AI

Pallavi Prakash Madarakhandi[1], Dr. Sakshi Kathuria[2], Dr. Ekta Soni[3]
[1]*M .Tech (DS), Amity University Haryana, Gurugram, Haryana*
[2,3]*Assistant Professor, ASET, Amity University Haryana Gurugram, Haryana*

Abstract: **The credit risk prediction is a significant aspect of financial risk management since it assists banks and other financial institutions in assessing the chance of a borrower falling to fulfil loan repayment obligations. The right prediction models are required to minimise the financial losses as well as to facilitate good lending decisions. Statistical methods that have been used traditionally in credit risk assessment include logistic regression, discriminant analysis, and classical methods of credit scoring. These methods utilise historical data of both financial and demographic data of a borrower to determine a probable default. As artificial intelligence and machine learning have developed, increasingly sophisticated predictive models have been presented to predict credit risk. Decision trees, random forests, support vector machines, gradient-boosting and neural networks are algorithms that have demonstrated better predictive accuracy than the classical algorithms. These models can extract complicated patterns in high-volume financial data, and this enhances the precision of risk analysis. Nevertheless, a lot of machine learning models are not transparent, and they are black boxes. To solve this problem, Explainable Artificial Intelligence methods like SHAP and LIME can be used to identify the influence of various features on model predictions and, therefore, achieve enhanced interpretability and enable more appropriate financial decisions.**

Keywords: Credit Risk Prediction, Machine Learning, Explainable AI, SHAP, LIME, Financial Risk Management.

I. INTRODUCTION

A credit risk prediction is a component of financial risk management since financial institutions must engage in assessing whether borrowers can repay their loans. A proper credit risk evaluation system can assist financial institutions in preventing financial loss and remaining stable in financial operations. Traditionally, financial institutions used such statistical techniques such as logistic regression and discriminant analysis to estimate the likelihood of default risk of borrowers. The elements of financial dimension, such as income level, credit history and employment, are taken into consideration in estimating the credit risk of borrowers through statistical procedures [1], [2].

As much as statistical methods are popular because they are easy to understand and interpret, they are not commonly in a position to analyse complex financial data. The emergence of digital financial services has necessitated the complexity of financial data, and it is imperative to employ more potent financial data analysis techniques.

Machine learning techniques have gained much interest in credit risk prediction. Large financial data can be analysed using machine learning techniques, and patterns associated with the actions of the borrowers can be detected, which is not possible using statistical techniques. It has been reported that machine learning techniques such as random forests, support vector machines, and gradient boosting can be of great help to enhance prediction accuracy in credit risk assessment [3], [4].

Most of such models are, however, black boxes, i.e. their decision-making process cannot easily be understood. This has been an issue in the financial sector because the necessity of regulation compliance and accountability can not be ignored. Explainable Artificial Intelligence (XAI) has been identified as a new branch of study, and it tries to enhance the interpretability of machine learning models. This assists the financial institutions in getting to know and have confidence in the automated credit scoring models [5], [6].

Statistical models have traditionally been used to carry out research on credit risk prediction. As an example, logistic regression has been extensively applied in credit scoring models as the obtained results can be readily interpreted, and the model is used to estimate the likelihood of defaulters based on the financial

background of the borrowers. The other model that is applied in credit scoring is the discriminant analysis, which assists in classifying borrowers based on their risks.

The other area where recent research has been directed is on enhancing the transparency and understanding of credit risk prediction systems. Demajo et al. [7] studied the usage of explainable artificial intelligence in credit scoring and highlighted the necessity of the ability to present interpretable predictions when making financial decisions. In their research, they noted that explainable machine learning models can enhance the level of trust and accountability towards automated credit evaluation systems. The financial institutions will be able to get a better understanding of the factors that affect the risk of the borrowers and make more credible lending decisions by employing the techniques of explainability.

On the same note, Abedin et al. [8] have investigated how machine learning algorithms can be applied to forecast credit risk and showed that the accuracy of predictive models can be substantially improved by using modern models in assessing the risk of borrowers. Their study revealed that the high-tech machine learning models can detect the latent trends in financial information that can hardly be reflected by traditional statistical models. These results indicate that incorporating machine learning and explainable AI technology may be an important step in creating more reliable and transparent credit risk forecasting systems.

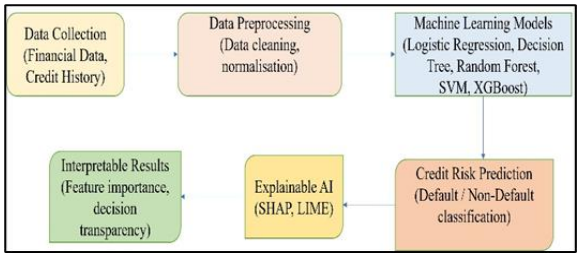

Fig. 1. Framework of credit risk prediction using machine learning and explainable AI.

II. LITERATURE REVIEW

A. Traditional Credit Risk Models

The machine learning methods have been studied by numerous researchers to enhance the accuracy of predicting credit risk. Saeed et al. [9] examined machine learning models, including decision trees and random forests, to score credit and discovered that the ensemble learning models are far more effective at classifying data when compared to the more traditional statistical models.

Collins and Emmanuel [10] used the gradient boosting algorithms to forecast loan defaults and achieved superior results over the logistic regression models. On the same note, Eshan et al. [11] put forward a self-supervised machine learning framework to predict credit risk that can increase the prediction accuracy by learning unobservable patterns in financial data.

Goel et al. [12] studied the importance of artificial intelligence in predictive analytics in financial risk management and revealed that the model based on AI is much more effective at enhancing the reliability of the predictions. The article by Al Shiam et al. [13] suggested an interpretable prediction framework of credit risk using XGBoost with SHAP explanations to learn about borrower financial characteristics.

de Lange et al. [14] have created explainable credit scoring with LightGBM and SHAP values to enhance transparency in the banking decision-making processes. In their research, Brown [15] examined the role of artificial intelligence in banking risk estimation and discovered that explainable models enhance trust in automated credit scoring models.

Edunjobi and Odejide [16] introduced theoretical ways on how to use artificial intelligence in managing credit risk and the relevance of explainable models to enhance transparency and regulate compliance.

Table I Summary of Literature Review

Author (s) [Ref]	Year	Algorithm Used	Dataset Used (Name + Type)	Description
Bussman n et al. [1]	2021	Explainable Machine Learning	Banking Credit Dataset (Financial data)	Discussed the role of explainable ML models in Credit risk
Misheva et al. [2]	2021	Explainable AI Techniques	Financial Lending Dataset	Highlighted interpretability in credit risk prediction.
Lessman n et al. [3]	2015	Classification Algorithms	German Credit Dataset (UCI)	Compared ML algorithms for credit scoring.

Brown & Mues [4]	2012	Machine Learning Models	Credit Card Default Dataset	Showed ML models outperform traditional models.
Ribeiro et al. [5]	2016	LIME	General ML datasets	Proposed LIME for explaining model predictions.
Lundberg & Lee [6]	2017	SHAP	Various ML datasets	Introduced SHAP for model interpretability.
Demajo et al. [7]	2022	Explainable ML	Credit scoring dataset	Applied XAI for credit scoring.
Abedin et al. [8]	2023	Machine Learning	Financial dataset	Used ML for credit risk prediction.
Saeed et al. [9]	2024	Random Forest	Banking Credit Dataset	Improved credit scoring using ensemble learning.
Collins & Emmanuel [10]	2023	Gradient Boosting	Loan Default Dataset	Applied boosting models for loan prediction.
Eshan et al. [11]	2025	Self-Supervised Learning	Financial Behaviour Dataset	Proposed advanced ML for credit risk prediction.
Goel et al. [12]	2024	AI Predictive Analytics	Financial Risk Dataset	Applied AI techniques for financial risk analysis
Al Shiam et al. [13]	2024	XGBoost + SHAP	Credit Risk Dataset	Developed explainable credit risk model.
de Lange et al. [14]	2022	LightGBM + SHAP	Banking Credit Dataset	Proposed transparent credit scoring system.
Brown [15]	2024	AI Risk Models	Banking Risk Dataset	Studied AI applications in banking risk analysis.
Edunjobi & Odejide [16]	2024	AI Framework	Financial Risk Dataset	Proposed AI frameworks for credit risk systems.
Breiman [17]	2001	Random Forest	ML Classification Dataset	Introduced the Random Forest algorithm.
He & Garcia [18]	2009	Imbalanced Data Learning	Classification Dataset	Addressed class imbalance problems.
Barocas et al. [19]	2019	Fairness in AI	Social & Financial Data	Discussed fairness in AI decision systems.
Friedman [20]	2001	Gradient Boosting	ML Dataset	Proposed gradient boosting algorithm.

B. Machine Learning Models for Credit Risk Prediction

Credit risk prediction systems have greatly enhanced performance due to the use of machine learning algorithms. The popularity of decision tree models lies in the fact that such models produce understandable decision rules, which cluster risky borrowers based on their riskiness. The Random Forest algorithm, originally proposed by Breiman [17],

Random forest models are an extension of decision trees, which combine several trees to enhance a higher prediction accuracy and reduce overfitting. XGBoost and LightGBM gradient boosting algorithms also enhance the performance of prediction by adding decision trees sequentially and correcting the errors of previous models.

Credit risk prediction also makes extensive use of Support Vector Machines since this type of machine can identify nonlinear relationships between financial data. Fig. 2 illustrates the machine learning models that are popular in credit risk prediction.

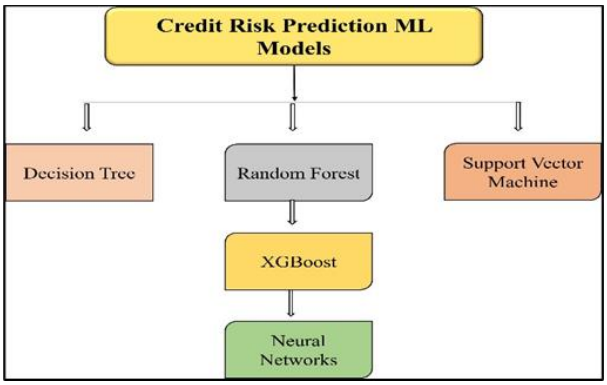

Fig. 2. Machine Learning Algorithms for Credit Risk Prediction

C. Explainable Artificial Intelligence in Credit Risk Prediction

Elucidable Artificial Intelligence methods have been proposed to enhance eminence in machine learning models. The SHAP values are used to assess the impact of every feature on the model predictions and give both global and local interpretations.

LIME describes individual predictions by estimating complicated models with simpler interpretable models. These methods are used to get a feel of how credit scoring models can make a prediction, as well as to determine significant features of borrowers that influence credit risk. Fig. 3 demonstrates explainable AI methods that are used to interpret machine learning predictions.

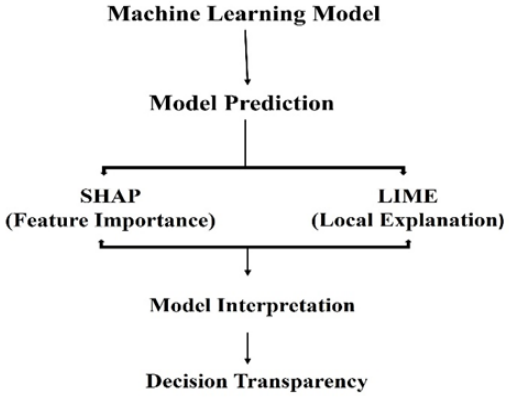

Fig. 3. Explainable AI Techniques

III. COMPARATIVE ANALYSIS

Credit risk prediction has been extensively applied with different statistical and machine learning models in financial organisations. All models are associated with various advantages and disadvantages based on the characteristics of the financial data and the purpose of the prediction task. Conventional statistical techniques like logistic regression are popular since they are easy and their output can be interpreted. However, they rarely reflect the intricate relationships existing in massive financial data. On the one hand, the use of different machine learning algorithms, including random forests, support vector machines, decision trees, gradient boosting algorithms, and neural networks. has been known to provide better predictive. credit risk prediction activities' accuracy. These financial data can be analysed using models to process large volumes of data. data and identify hidden patterns concerning the behaviour of borrowers. Table II provides a comparison of popular credit risk prediction models with their benefits and flaws.

Table II: Comparison of Credit Risk Prediction Models

Model	Advantages	Limitations
Logistic Regression	Simple and interpretable	Limited predictive power
Decision Tree	Easy to understand	Overfitting
Random Forest	High accuracy	Less interpretable
Support Vector Machine	Handles nonlinear data	Parameter tuning required
Neural Networks	Captures Complex Patterns	Difficult to interpret

According to Table II, the traditional statistical models like the logistic regression are more interpretable and simpler to implement in financial institutions. Nevertheless, these models can be ineffective in the case of large and complicated financial data. Random forests and gradient boosting algorithms are machine learning models that are typically higher in predictive accuracy than other models by being able to model nonlinear relationships in financial data. Although these models perform better, they are usually afflicted by a low level of interpretability, and explainable artificial intelligence techniques have been developed to enhance transparency in credit risk prediction systems.

IV. CHALLENGES IN CREDIT RISK PREDICTION

Even with the benefits of machine learning and explainable AI methods, there are still a number of challenges with credit risk prediction systems. A significant issue is to balance prediction accuracy and interpretability. Complex machine learning models can perform better in predictive performance, but cannot be interpreted.

The data imbalance is another problem in which default cases form a small part of the financial data sets relative to non-default cases.

Discrimination and equity are other significant issues with AI-based credit scoring models since discriminatory predictions can be made using biased data.

V. RESEARCH GAPS AND FUTURE WORK

A. Research Gaps

In spite of major advances made in this field of credit risk prediction, there is still a research gap associated with this research area. The other conventional credit score methods, such as logistic regression and discriminant analysis, provide results that are easily

explainable. It is not possible to easily explain complex nonlinear relationships that may be present in financial data. Although there has been an increase in the accuracy of prediction using various machine learning models, it is not easy to comprehend how the conclusion has been arrived at by the model [6], [13].

The second research gap is associated with the lack of transparency and interpretability of the system of machine learning-based credit risk predictions. Financial institutions' policies are heavily regulated and require transparent results. Although explainable AI is now being proposed to improve the interpretability of various models such as credit risk prediction models using SHAP and LIME methods, this has not been applied to any of the financial systems [13].

In addition, credit risk data is also faced with data imbalance problems, whereby the number of defaults is significantly less than the number of non-defaults. This can potentially impact the performance of prediction models and may lead to biased prediction results [18]. In addition, the issue of fairness and ethics in AI-powered credit rating has not been adequately addressed. This, therefore, creates the need for fairness-conscious machine learning models [19] since biased data can potentially lead to discriminatory credit behaviour.

B. Future Work

Future studies on credit risk forecasting must be concerned with the construction of models that are highly predictive and interpretable. The middle ground between traditional methods of statistic and emerging machine learning algorithms can be offered by hybrid methods that ensure the enhancement of the model performance and transparency.

The other avenue of research that is likely to yield positive returns is the combination of deep learning and explainable AI methods. Deep learning models have shown powerful applications in the analysis of complex data; but their inability to explain their results is a significant weakness. The explainable deep learning models should be developed, which might enhance its use in the financial risk management systems [20].

Moreover, it is recommended that future research should aim at addressing the problem of data imbalance in credit risk data using the complex sampling techniques and cost-sensitive learning models. They should also do research on fairness-conscious machine learning-based algorithms that diminish bias and discrimination in automated credit score systems [19]. Big data analytics and real-time financial data could also increase the accuracy and reliability of credit risk forecasting models.

VI. CONCLUSION

The prediction of credit risk plays a vital role in financial risk management since it allows financial institutions to assess the chances of the default of a borrower and to mitigate possible monetary losses. Credit scoring has been extensively performed using traditional statistical models such as logistic regression and discriminant analysis, but these models tend not to perform well on complex financial datasets.

The invention of machine learning has enhanced the credit risk prediction system performance tremendously. Algorithms including decision trees, random forest models, support vector machine techniques, gradient boosting approaches have proven to be more predictive than the traditional methods in statistics [3], [4].

Regardless of these strengths, most machine learning models are black-box and therefore untransparent. Explainable Artificial Intelligence models in the form of SHAP and LIME have thus been available to enhance the interpretability of models and enable the stakeholders of the system to comprehend how predictions are made [5], [6].

Even though a significant breakthrough has occurred in this area, there are still a number of issues such as the interpretability of the models and the imbalance of data, along with the fairness of AI-based credit scoring models. The further studies ought to be aimed at ensuring transparent, valid, and equitable credit risk prediction models, which would help make reliable decisions in financial institutions.

REFERENCES

[1] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable machine learning in credit risk management," *Computational Economics*, vol. 57, no. 1, pp. 203–216, 2021.

[2] B. Hadji Misheva, H. Giesecke, A. Amini, and S. Chawla, "Explainable AI in credit risk management," *Journal of Financial Data Science*, vol. 3, no. 4, pp. 1–15, 2021.

[3] S. Lessmann, B. Baesens, H. V. Seow, and L. C. Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring," *European Journal of Operational Research*, vol. 247, no. 1, pp. 124–136, 2015.

[4] T. Brown and J. C. Mues, "An experimental comparison of classification algorithms for credit scoring," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3446–3453, 2012.

[5] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. ACM SIGKDD*, 2016, pp. 1135–1144.

[6] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, 2017.

[7] A. N. Demajo, D. Briffa, and J. Dingli, "Explainable artificial intelligence for credit scoring," *IEEE Access*, vol. 10, pp. 109040–109052, 2022.

[8] A. T. Abedin, M. S. Rahman, and M. H. Kabir, "Credit risk prediction using machine learning algorithms," *IEEE Access*, vol. 11, pp. 23145–23156, 2023.

[9] S. Saeed, A. Mustafa, and M. Ali, "Machine learning techniques for credit risk assessment," *International Journal of Finance and Economics*, 2024.

[10] N. Collins and I. Emmanuel, "Machine learning algorithms for credit risk prediction," *International Journal of Financial Studies*, 2023.

[11] A. N. Eshan, M. Rahman, and S. Islam, "Credit risk prediction using self-supervised learning techniques," *Journal of Financial Analytics*, 2025.

[12] M. Goel, R. Sharma, and S. Gupta, "Artificial intelligence for predictive analytics in financial management," *International Journal of Data Science and Analytics*, 2024.

[13] S. A. Al Shiam, M. Hossain, and T. Rahman, "Explainable credit risk prediction using XGBoost and SHAP," *IEEE Access*, 2024.

[14] P. E. de Lange, J. Meijer, and S. Bhulai, "Explainable AI for credit scoring using LightGBM," *Expert Systems with Applications*, 2022.

[15] M. Brown, "Artificial intelligence applications in banking risk assessment," *Journal of Banking and Financial Technology*, 2024.

[16] T. Edunjobi and O. Odejide, "Artificial intelligence frameworks for credit risk management," *International Journal of Financial Technology*, 2024.

[17] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[18] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[19] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning*. Cambridge, MA, USA: MIT Press, 2019.

[20] J. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.