ISSN: 2349-6002

Comparative Study of Traditional And ML-Based Credit Scoring Models

Aman Gautam[1], Saijalpreet Kaur[2], Ashutosh Rauniyar[3], Akhil Pandey[4]
[1,2,3,4]*Department of Computer Applications, Invertis University, Bareilly, India*
doi.org/10.64643/IJIRTV12I11-207109-459

***Abstract*—Credit scoring is a crucial component of lending decisions in financial institutions, as it helps banks, credit agencies, and other lenders evaluate the probability of borrower default before approving loans or credit facilities. An effective credit scoring system supports better risk assessment, reduces financial losses, and improves the overall quality of credit management. Traditional credit scoring models have been widely used for several decades because of their simplicity, transparency, interpretability, and acceptance within regulatory frameworks. These models, such as logistic regression, discriminant analysis, and scorecard-based approaches, allow decision-makers to understand how different borrower characteristics influence creditworthiness.**

***However, with the rapid growth of digital financial services, large-scale customer data, and complex borrowing patterns, traditional models often face limitations in capturing non-linear relationships and hidden patterns in credit data. In this context, Machine Learning (ML) techniques are increasingly being adopted for credit scoring because they can process high-dimensional datasets, handle complex variable interactions, and often provide better predictive accuracy than conventional statistical methods. ML models such as Random Forest, Support Vector Machine, XGBoost, and Artificial Neural Networks have shown strong potential in improving default prediction and credit risk classification.**

This paper presents a comprehensive review of traditional and ML-based credit scoring models. It examines their basic methodologies, advantages, limitations, interpretability, regulatory relevance, and comparative performance. The review also discusses recent developments, including hybrid credit scoring models and explainable artificial intelligence, which aim to balance prediction accuracy with transparency. Finally, the paper highlights major insights and future research directions for developing more reliable, fair, and explainable credit risk assessment practices in modern financial systems.

***Index Terms*—Credit scoring; Explainable artificial intelligence; Hybrid models; Logistic regression; Machine learning; Random Forest; XGBoost.**

I. INTRODUCTION

Credit scoring is one of the basic tools used by banks and financial institutions to estimate the possibility of borrower default. It supports loan approval, credit limit decisions, risk control, and compliance-related reporting. With the expansion of credit markets and the entry of different types of borrowers, the need for more reliable risk assessment has become stronger [9]. Conventional credit scoring models, particularly statistical and scorecard-based methods, have been used for a long time because they are simple to explain and are generally accepted by regulators [7]. However, these methods are often limited when the relationship between borrower characteristics and default behaviour is non-linear or when the available data are large, diverse, and partly unstructured.

In recent years, machine learning has created a new direction in credit scoring by allowing models to learn patterns from large datasets with less dependence on manually specified assumptions. These methods can handle complex variable interactions and, in many cases, produce better predictive accuracy than traditional models [8]. Tree-based ensembles, support vector machines, neural networks, and deep learning models have been tested in different credit risk studies, and several comparative works have reported their strong performance, especially where the dataset is large and imbalanced [2], [11]. Still, the use of machine learning in finance is not straightforward. Issues related to model explanation, fairness, bias, customer rights, data privacy, and regulatory auditability continue to restrict the direct adoption of black-box models in lending decisions [4].

This review discusses traditional and machine learning-based credit scoring approaches with respect to their methodology, prediction ability, interpretability, data requirements, and regulatory

suitability. It further considers hybrid modelling, explainable artificial intelligence, and the use of alternative data as emerging directions in credit risk assessment. The purpose is to present a balanced understanding of both approaches rather than treating machine learning as a complete replacement for established credit scoring practices.

II. TRADITIONAL CREDIT SCORING MODELS

Traditional credit scoring has largely been developed around statistical models and scorecard systems because these approaches are easy to implement, transparent, and suitable for routine lending practice. Linear Discriminant Analysis, used in early credit and bankruptcy prediction studies by Altman [1], and Logistic Regression, discussed widely in consumer credit scoring by Hand and Henley [7], are among the most common methods. Linear Discriminant Analysis separates good and bad borrowers by finding a linear combination of variables that gives better group separation. Logistic Regression estimates the probability of default or non-default through a logistic function, which makes it useful for binary credit decisions. Along with these models, scorecards are also common in banks and financial institutions. In scorecard-based systems, borrower-related variables such as income range, repayment history, age group, employment status, existing debt, and length of credit history are converted into score bands. The total score is then used to decide whether the applicant falls within an acceptable risk level. Their usefulness lies in the fact that credit officers, branch managers, customers, and regulators can easily understand how the final score has been produced [9].

The strongest advantage of traditional models is their interpretability. Since the effect of each variable can be examined through coefficients, weights, or score bands, these models are easier to justify during internal review and regulatory assessment. They also work well in institutions where documentation, audit trails, and customer-level explanation are necessary. In stable lending environments, where borrower profiles and economic conditions do not change sharply, these methods may remain reasonably reliable. However, their limitations are also clear. Most traditional models assume linear or near-linear relationships, and therefore they may fail to capture complex interactions among financial, behavioural, and demographic variables. Their performance depends heavily on manually prepared features, and this process may involve expert bias or loss of useful information. Another limitation is that these models are mainly designed for structured data and do not naturally handle text records, online behaviour, customer communication, image-based documents, or other alternative data sources.

III. MACHINE LEARNING-BASED CREDIT SCORING MODELS

Machine learning-based credit scoring models have gained Machine learning-based credit scoring models have received attention because they can detect complex and hidden patterns in large datasets. Random Forest and XGBoost are widely applied ensemble methods in this area. Random Forest generates many decision trees through bagging and combines their outputs to reduce variance and improve prediction stability. XGBoost, on the other hand, follows a boosting mechanism in which each new tree attempts to correct the errors of previous trees, making it effective for structured prediction tasks [3]. Support Vector Machines are useful for high-dimensional datasets because they try to identify an optimal separating boundary between default and non-default borrowers. However, their performance depends on suitable kernel selection and careful tuning of parameters. Artificial Neural Networks are also used in credit scoring because their layered structure can represent non-linear relationships among variables. At the same time, they need sufficient training data, proper regularization, and careful validation to reduce overfitting [6]. Advanced deep learning methods, including deep neural networks, convolutional neural networks, recurrent neural networks, and LSTM models, have also been explored where credit assessment involves sequential, image-based, or time-dependent financial information [11].

The main benefit of machine learning models is that they can capture relationships that are difficult to express through conventional statistical assumptions. Because of this, many empirical studies have found that machine learning models provide better prediction performance than traditional methods, particularly when the data contain non-linear patterns and a large number of variables [8]. These models are also flexible because they can include transaction histories, digital

footprints, customer behaviour, and text-based information. In some situations, they reduce the burden of manual feature construction by learning useful representations from raw or semi-processed data. However, these advantages also create practical difficulties. Many machine learning models, especially deep learning models, are difficult to interpret and may function as black-box systems. This creates concern for lenders, borrowers, auditors, and regulators because credit decisions should be explainable and contestable [4]. Bias is another important concern, as historical lending data may reflect social or institutional inequalities. If such data are used without proper checks, the model may produce unfair outcomes for sensitive groups. Privacy and data protection are also important because these models often require large amounts of personal and financial information, which must be handled according to legal and ethical requirements [5].

IV. COMPARATIVE ANALYSIS: TRADITIONAL VS. ML-BASED CREDIT SCORING

In terms of interpretability, traditional models such as scorecards, Logistic Regression, and Linear Discriminant Analysis are easier to explain because their variables, coefficients, and scoring rules are visible. This transparency helps in customer communication, internal audit, and regulatory compliance. Machine learning models, in contrast, often provide higher predictive strength but are more difficult to explain. To reduce this problem, explainable AI techniques such as SHAP and LIME have been introduced so that the contribution of individual features can be examined in a more understandable manner [4]. These tools are useful, but they do not fully remove the need for careful model governance in credit decisions.

The data requirements of both approaches are also different. Traditional models generally require structured numerical or categorical variables, and their performance depends strongly on expert-based feature engineering. Machine learning models can work with structured, semi-structured, and unstructured data, including transaction sequences, text logs, online behaviour, and other alternative information sources. Deep learning models are particularly useful when meaningful representations can be learned directly from large volumes of raw data [6]. However, the use of such data must be justified carefully because more data do not automatically mean better or fairer credit decisions.

Prediction performance is commonly evaluated through accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve [2]. In many benchmark studies, machine learning models, especially Random Forest and XGBoost, have shown better predictive performance than Logistic Regression when the dataset is large, imbalanced, and non-linear [8], [11]. However, traditional models should not be dismissed completely. When the dataset is small, the variables are well understood, and the relationship between predictors and default is mostly linear, Logistic Regression and scorecard models may perform competitively. Therefore, the choice of model should depend on the nature of the data, the purpose of lending, the required level of explanation, and the regulatory environment.

From the regulatory perspective, traditional credit scoring models are more aligned with existing compliance systems because their decision logic can be documented and reviewed. Machine learning models require stronger governance because lenders must demonstrate that the model is explainable, fair, stable, and free from unacceptable bias. Regulatory bodies and policy frameworks increasingly stress transparency, data protection, and accountability in automated decision-making [5]. For this reason, financial institutions using machine learning for credit scoring must maintain proper validation records, bias testing, model monitoring, and customer-level explanation mechanisms.

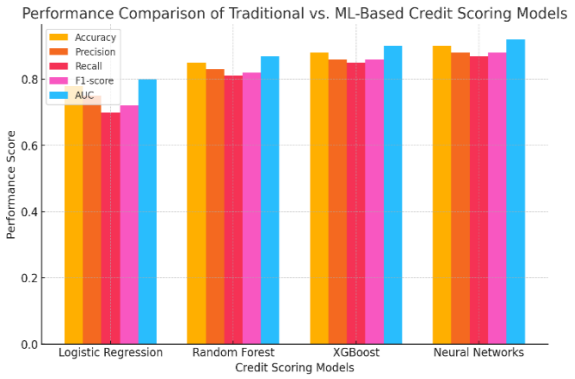

Fig 1: Performance Comparison of Traditional vs. ML-Based Credit Scoring Models

The performance comparison shown in Fig. 4.1 may be used to present the relative behaviour of Logistic

Regression, Random Forest, XGBoost, and Neural Networks across accuracy, precision, recall, F1-score, and AUC. Such comparison is useful for showing that machine learning models usually perform better on complex datasets, but the figure should be interpreted with caution because model performance depends on data quality, class imbalance, feature selection, and validation design.

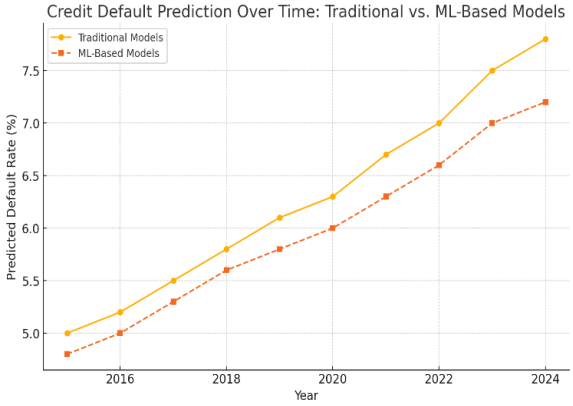

Fig2: Credit Default Prediction Over Time: Traditional vs. ML-Based Models (Adapted from: Brown & Mues, 2012; Zhang & Trubey, 2019).

Similarly, Fig. 4.2 may be used to explain the difference between traditional and machine learning-based default prediction trends over time. Machine learning models may adapt better to changing borrower behaviour and economic instability, while traditional models may show more rigid prediction patterns. However, any such graph should be clearly linked with the dataset, experimental design, and cited studies; otherwise, it should be presented only as a conceptual or adapted illustration [2], [11].

V. EMERGING TRENDS AND FUTURE DIRECTIONS

Recent credit scoring research is moving towards hybrid models that combine the clarity of traditional approaches with the predictive strength of machine learning. For example, Logistic Regression may be used with tree-based models or other machine learning methods to improve performance while retaining some level of explanation [10]. Explainable AI is another important direction, as methods such as SHAP, LIME, and counterfactual explanations can help lenders understand why a model has accepted or rejected a borrower [4]. These methods are particularly important in credit scoring because borrowers and regulators may require a clear reason for adverse decisions. Future work needs to develop explanation methods that are not only technically correct but also understandable for credit officers and customers.

The use of alternative data is also increasing, especially for borrowers with limited formal credit history. Transaction records, mobile behaviour, online activity, and other digital traces may help assess thin-file borrowers, but their use must be controlled carefully to avoid privacy violations and unfair profiling [11]. Deep learning and reinforcement learning are also being explored for dynamic credit risk assessment, especially where financial behaviour changes over time. Still, these advanced models should be adopted only when there is sufficient data quality, validation evidence, and regulatory clarity. In the coming years, credit scoring research should focus on fairness-aware modelling, privacy-preserving learning, model monitoring, and standard guidelines for the responsible use of machine learning in finance.

VI. CONCLUSION

Traditional credit scoring models remain important in financial decision-making because they are transparent, easy to document, and widely accepted within existing regulatory frameworks. Methods such as logistic regression, linear discriminant analysis, and scorecard models allow lenders and regulators to understand how borrower characteristics influence the final credit score. This makes them useful where accountability, compliance, and customer explanation are required. However, machine learning models, especially ensemble-based methods such as Random Forest and XGBoost, often provide better predictive performance when datasets are large, complex, and non-linear. These models can identify hidden patterns and improve default prediction more effectively than many traditional techniques.

Still, higher accuracy alone is not sufficient for credit decision-making. Since credit scoring directly affects borrowers, issues related to interpretability, fairness, bias, data privacy, and regulatory compliance must be carefully addressed. A practical way forward is to develop hybrid and explainable models that combine the predictive power of machine learning with the transparency of traditional methods. Future research should focus on responsible use of alternative data,

privacy-preserving techniques, bias control, and standardized regulatory procedures.

REFERENCES

[1] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, 1968.

[2] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," Expert Systems with Applications, vol. 39, no. 3, pp. 3446–3453, 2012.

[3] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 785–794.

[4] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017.

[5] European Parliament and Council of the European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)," Official Journal of the European Union, vol. L119, pp. 1–88, 2016.

[6] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[7] D. J. Hand and W. E. Henley, "Statistical classification methods in consumer credit scoring: A review," Journal of the Royal Statistical Society: Series A (Statistics in Society), vol. 160, no. 3, pp. 523–541, 1997.

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

[9] L. C. Thomas, J. N. Crook, and D. B. Edelman, Credit Scoring and Its Applications. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics (SIAM), 2002.

[10] C.-F. Tsai and M.-L. Chen, "Credit rating by hybrid machine learning techniques," Applied Soft Computing, vol. 10, no. 2, pp. 374–380, 2010.

[11] Y. Zhang and P. Trubey, "Machine learning and ensemble techniques for credit risk evaluation: A systematic review," *IEEE Access*, vol. 7, pp. 146182–146195, 2019.