IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.154187

AN OVERVIEW ON: CREDIT RISK ANALYSIS

Prof. Madhuri Parate[1], Meena Godghate[2], Hanisha Bulhe[3]

Professor, Department of Computer Science and Engineering,

Nagarjuna Institute of Engineering Technology and Management, Nagpur, Maharashtra, India[1]

UG Student, Department of Computer Science and Engineering (AIML),

Nagarjuna Institute of Engineering Technology & Management, Nagpur, Maharashtra, India[2,3]

Abstract: Credit risk analysis is essential for evaluating the likelihood of borrowers defaulting on loans. This study uses historical financial and customer data to develop models that assess creditworthiness. Various factors such as income, credit history, and repayment behavior are analyzed to identify risk patterns. Statistical and machine learning techniques are applied to improve prediction accuracy. The findings highlight the importance of data-driven approaches in minimizing financial risk and supporting effective lending decisions.

Keywords: Credit Risk, Creditworthiness, Default Prediction, Machine Learning, Financial Analysis, Risk Assessment, Logistic Regression, Data Analysis

I. INTRODUCTION

Credit risk analysis is a fundamental aspect of the financial industry, focusing on the evaluation of a borrower's ability to repay loans and meet financial obligations. With the rapid expansion of lending activities in banks and financial institutions, assessing credit risk accurately has become increasingly important to minimize losses and ensure financial stability. Ineffective credit risk management can lead to high default rates, negatively impacting both lenders and the broader economy.

Traditionally, credit risk assessment relied on manual evaluation methods and basic statistical techniques, considering factors such as income level, employment status, credit history, and collateral. However, with the growth of digital data and advancements in computational technologies, modern approaches now incorporate data-driven models and machine learning algorithms to enhance prediction accuracy and efficiency.

This project focuses on analyzing credit risk using historical financial data and borrower characteristics. By applying various analytical techniques, the study aims to identify patterns associated with high-risk borrowers and develop models that can predict the probability of default. The findings of this study can assist financial institutions in making informed lending decisions, reducing potential losses, and improving overall risk management strategies.

II.LITERATURE REVIEW

Credit risk analysis has been widely studied in the fields of finance and data analytics, with researchers focusing on improving the accuracy and efficiency of default prediction models. Early studies primarily relied on traditional statistical techniques such as logistic regression and discriminant analysis. These models were effective in identifying key financial indicators influencing creditworthiness, including income, debt ratio, and repayment history.

With advancements in technology, recent research has shifted toward the use of machine learning algorithms for credit risk assessment. Techniques such as decision trees, random forests, support vector machines, and neural networks have demonstrated higher predictive performance compared to traditional methods. These models are capable of capturing complex, non-linear relationships within large datasets, leading to more accurate risk classification.

Several studies have also emphasized the importance of feature selection and data preprocessing in improving model performance. Factors such as credit history, loan amount, employment status, and behavioral patterns have been identified as significant predictors of default risk. Additionally, the integration of alternative data sources, such as transaction records and digital footprints, has further enhanced credit scoring systems.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 4, April 2026

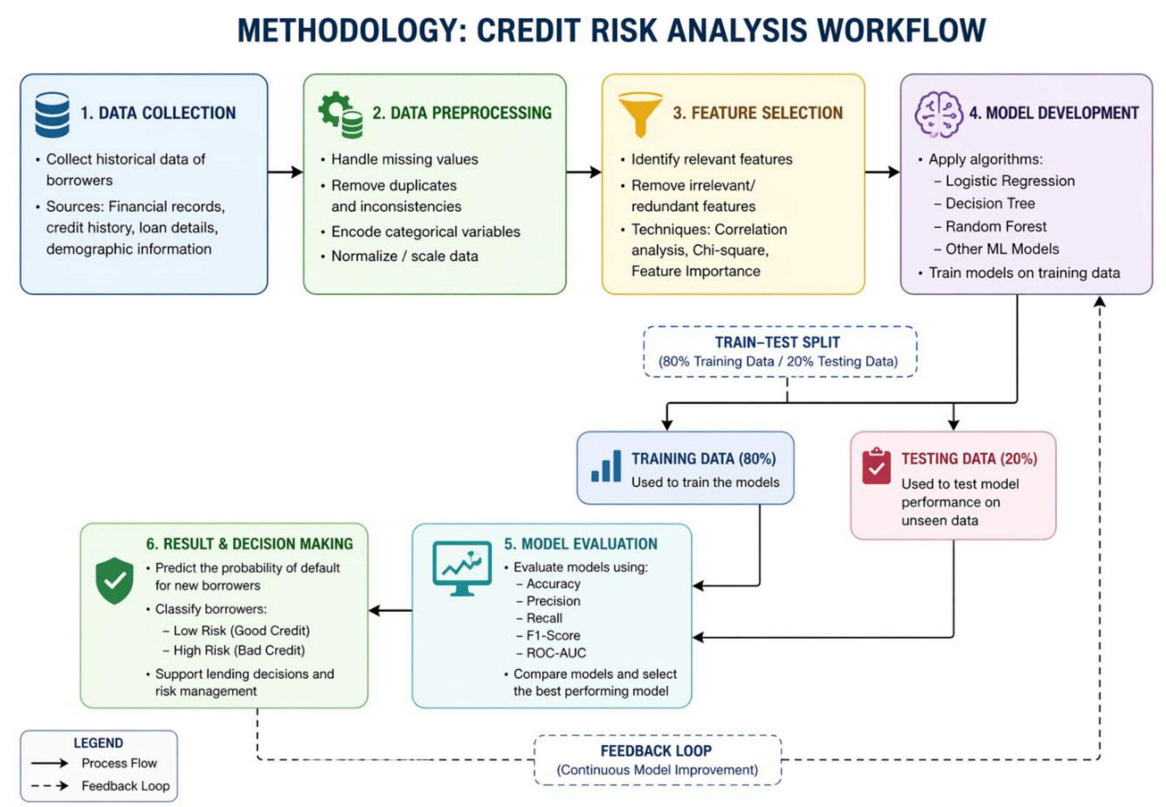
DOI: 10.17148/IJARCCE.2026.154187

Despite these advancements, challenges remain in terms of model interpretability, data privacy, and bias in decision-making. Recent research highlights the need for explainable artificial intelligence (XAI) to ensure transparency and fairness in credit risk models. Overall, the literature indicates a clear transition from traditional statistical approaches to more sophisticated, data-driven techniques for effective credit risk management.

III. METHODOLOGY

The methodology of this study involves a systematic approach to analyze and predict credit risk.The results are analyzed to identify the best-performing model and key factors affecting credit risk.

1. **Data Collection:** Gather borrower data such as income, credit history, loan details, and repayment behavior.
2. **Data Preprocessing:** Clean the data by handling missing values, removing duplicates, and converting categorical variables into numerical form.
3. **Exploratory Data Analysis:** Analyze data using statistical methods and visualizations to identify patterns and relationships.
4. **Feature Selection:** Select important variables that influence credit risk and remove irrelevant features.
5. **Data Splitting:** Divide the dataset into training and testing sets (typically 80:20).
6. **Model Development:** Apply models like logistic regression, decision trees, and random forests to predict credit risk.
7. **Model Evaluation:** Evaluate models using accuracy, precision, recall, F1-score, and ROC-AUC.
8. **Result Interpretation:** Identify the best model and key factors affecting default risk.
9. **Deployment:** Use the final model for real-world credit risk assessment (optional).



IV. MODELLING AND ANALYSIS

In this study, both statistical and machine learning models are used to analyze and predict credit risk. The dataset is first divided into training and testing sets to ensure unbiased evaluation of model performance. Logistic regression is applied as a baseline model due to its effectiveness in binary classification problems such as identifying defaulters and non-defaulters.

To improve prediction accuracy, advanced models such as decision trees and random forest classifiers are implemented. These models are capable of capturing complex and non-linear relationships between borrower attributes and credit risk.

DOI: 10.17148/IJARCCE.2026.154187

Feature selection techniques are used to identify the most relevant variables, including income, credit history, loan amount, and repayment behavior, which significantly influence the likelihood of default.

The performance of the models is evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The results indicate that machine learning models, particularly random forests, provide higher accuracy and better generalization compared to traditional methods. However, logistic regression remains useful for its interpretability and ability to explain the impact of individual variables.

Further analysis reveals that borrowers with low income, poor credit history, and higher loan amounts are more likely to default. Visualization tools such as confusion matrices and ROC curves are used to interpret model performance and validate the results.

Overall, the combined modeling and analysis demonstrate that data-driven approaches significantly enhance credit risk prediction and provide valuable insights for effective decision-making in financial institutions.

V. RESULTS AND DISCUSSION

The results of the study indicate that machine learning models provide effective predictions of credit risk based on borrower data. Among the applied models, the random forest classifier achieved the highest accuracy, followed by decision trees and logistic regression. Logistic regression, while slightly less accurate, offered better interpretability in understanding the influence of individual variables.

Evaluation metrics such as accuracy, precision, recall, and F1-score demonstrated that the models were capable of correctly classifying both defaulters and non-defaulters. The ROC-AUC score further confirmed the strong predictive performance of the models. The confusion matrix analysis showed a relatively low rate of misclassification, indicating the reliability of the developed models.

The analysis revealed that key factors such as low income, poor credit history, high loan amounts, and irregular repayment behavior significantly increase the probability of default. These findings are consistent with existing financial risk theories and highlight the importance of these variables in credit assessment.

Overall, the study demonstrates that data-driven and machine learning approaches enhance the accuracy and efficiency of credit risk analysis. The results support the adoption of advanced analytical techniques by financial institutions to improve decision-making, reduce default risk, and strengthen credit management systems.

VI. CONCLUSION

This study highlights the importance of credit risk analysis in evaluating the likelihood of borrower default and supporting effective lending decisions. By using historical financial data and applying both statistical and machine learning techniques, the study successfully identifies key factors influencing credit risk, such as income level, credit history, and loan characteristics.

The results demonstrate that machine learning models provide higher prediction accuracy compared to traditional methods, making them valuable tools for modern financial institutions. These models not only improve risk assessment but also help in reducing financial losses and enhancing decision-making efficiency.
Overall, the study emphasizes the significance of data-driven approaches in credit risk management. The findings suggest that the adoption of advanced analytical techniques can lead to more reliable and efficient credit evaluation systems, ultimately contributing to the stability and growth of the financial sector.

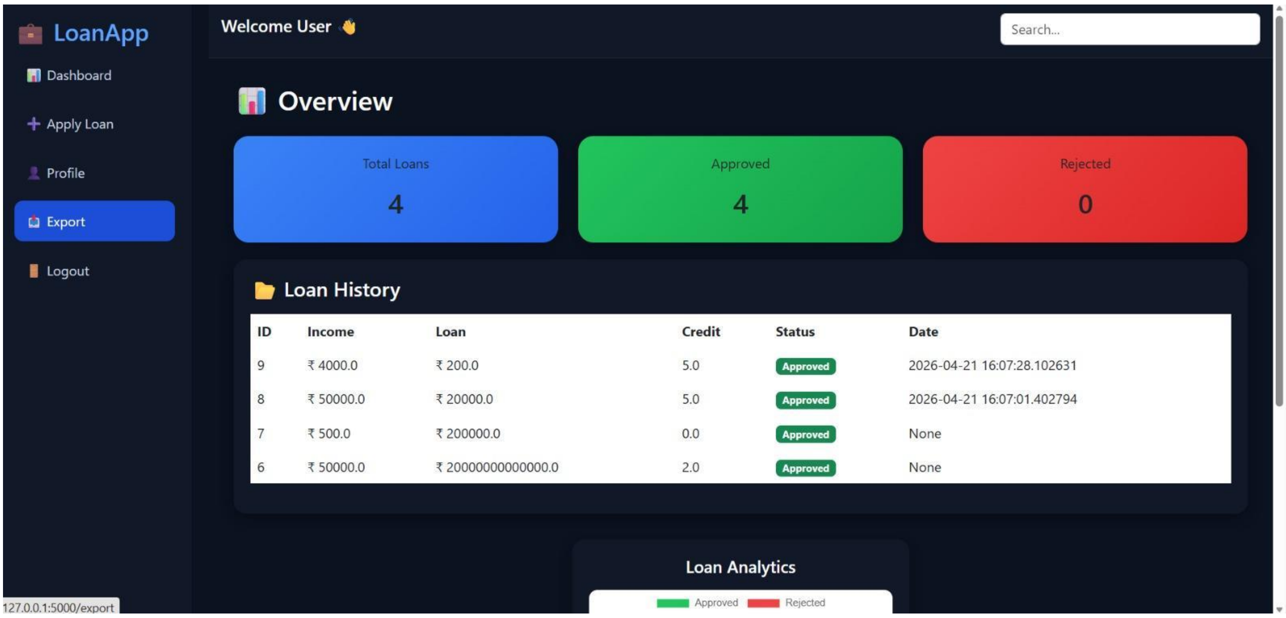
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.154187

VII. OUTPUT



REFERENCES

[1]. Altman, E. I., Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, 1968.
[2]. Hand, D. J., Henley, W. E., Statistical Classification Methods in Consumer Credit Scoring, 1997.
[3]. Basel Committee on Banking Supervision, Principles for the Management of Credit Risk, 2000.
[4]. Breiman, L., Random Forests, 2001.
[5]. Hastie, T., Tibshirani, R., Friedman, J., The Elements of Statistical Learning, 2009.