RiskNexus: An Explainable Credit Risk Analysis and Prediction System

Ajay Mishra[1], Aarya Talekar[2], Maruf Pathan[3], Swaraj Dudhmal[4], Prof. Kaajal Sharma[5]
[1,2,3,4]*Student, Computer Engineering Department Rajiv Gandhi Institute of Technology, Mumbai*
[5]*Faculty, Computer Engineering Department Rajiv Gandhi Institute of Technology, Mumbai*

***Abstract*—RiskNexus is an explainable credit-risk analysis platform designed for sustainable consumer lending. It combines a Random Forest model with transparent rule-based heuristics to estimate a FICO-like credit score using key financial features such as income, debt, and credit history. If the ML model is unavailable, a domain-based scoring system ensures continuity. The platform includes a customer dashboard for score simulation and recommendations, along with a bank-focused module for default risk and credit grading. It also supports "what-if" analysis to show how user actions impact credit scores. This hybrid approach balances accuracy, robustness, and interpretability for effective decision support.**

***Index Terms*—Credit Risk Prediction, Machine Learning, XGBoost, Support Vector Machine, Financial Analytics, Loan Default Prediction, Predictive Modeling.**

I. INTRODUCTION

Credit risk management is an essential component of financial institutions, as lending decisions directly influence financial stability and profitability. Banks and lending agencies must assess the creditworthiness of borrowers before approving loans to reduce the likelihood of default. Traditional credit scoring models rely primarily on manual analysis and fixed rule-based systems, which often struggle to handle the complexity and scale of modern financial datasets [1]. With the rapid expansion of digital banking systems, financial institutions now generate vast amounts of borrower data, including financial records, credit history, asset ownership, liabilities, and repayment behavior. Extracting meaningful insights from such datasets requires advanced computational techniques capable of detecting patterns associated with borrower risk [2].

Machine learning techniques provide powerful tools for predictive analytics in financial systems. By learning from historical loan data, machine learning models can identify patterns and relationships that indicate potential default risk. These models can significantly improve the accuracy and efficiency of credit risk evaluation [3].

The key contributions of this research are summarized as follows:(i) The design and implementation of a hybrid credit scoring engine that combines a Random Forest regressor trained on synthetic financial data with a deterministic heuristic scoring algorithm.

(ii) The development of front-end default risk and credit grade heuristics that operate in real time on user inputs and provide interpretable risk bands and grade probabilities.(iii) The integration of this modeling layer into a modular web architecture with dedicated views for customers and banks, including authentication, dashboards, and embedded risk tools.(iv) A set of scenario-based experiments showing that the system responds intuitively to changes in key risk drivers, illustrating its usefulness for education and lightweight decision support.

II. LITERATURE REVIEW

Credit risk assessment has evolved from traditional statistical approaches to advanced machine learning–based techniques. Early work such as Evaluation of the Credit Risk with Statistical Analysis [1] used statistical tests like T-test, Chi-square, and correlation to identify key risk factors. Although simple and interpretable, this approach lacked predictive modeling capabilities and was limited by small sample sizes.

Similarly, Research on Credit Risk Assessment of SMEs [2] applied logistic regression and PCA for feature reduction. While effective for SME-focused analysis, the study relied on traditional methods and

had limited dataset diversity, reducing its scalability. With the advancement of machine learning, Comparative Study of Credit Scoring Models [3] evaluated logistic regression, Random Forest, and neural networks, showing that deep learning models outperform traditional approaches. However, dataset imbalance negatively affected performance. Another study, Credit Scoring Using Machine Learning and Deep Learning [4], compared multiple algorithms such as LDA and Random Forest, demonstrating improved results but suffering from small dataset limitations.

More recent studies emphasize ensemble learning techniques. Machine Learning Approach for Credit Score Prediction [5] used XGBoost, SVM, and AdaBoost to achieve higher accuracy through ensemble methods, though at the cost of increased computational complexity. Likewise, Predicting Credit Risk Using Machine Learning [6] applied decision trees and gradient boosting classifiers on banking data, but its regional focus limited generalization. Overall, existing studies highlight improvements in prediction accuracy using machine learning; however, they lack integrated systems, real-time interaction, and comprehensive financial feature analysis. These limitations motivate the development of RiskNexus, which combines machine learning, heuristic scoring, and interactive dashboards for effective and practical credit risk assessment.

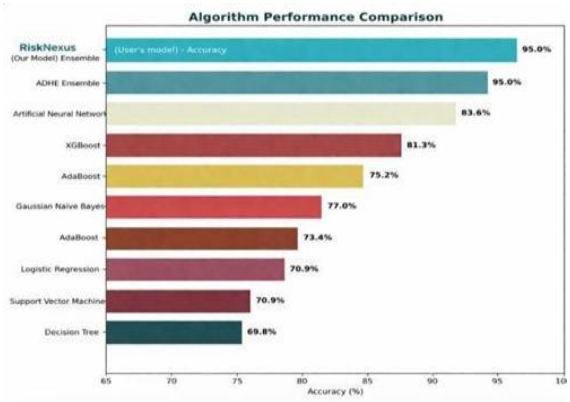


Table 1. Comparative Analysis of Existing Approaches

Paper	Technique Used	Advantages	Summary	Research Gap
Credit scoring using machine learning and deep learning, 2024	Linear Discriminant Analysis (LDA), Random Forest, etc.	Beginner-friendly comparison of 6 algorithms.	Compares performance of different algorithms for credit scoring.	Small dataset (only 688 observations) For Models like ANN & GB
Credit Risk Prediction Using Explainable AI, 2024	XGBoost, LightGBM, Random Forest, Decision Tree, SHAP	High performance and explainability using SHAP	Proposes an interpretable credit scoring model using XGBoost and SHAP on Lending Club data. Achieves 81.29% AUC with model explainability.	Imbalanced dataset; previous models lacked transparency and real-world readiness
Application of AI in Credit Risk Scoring for Small Business Loans , 2024	Random Forest (AI model), Delphi (traditional scoring)	PERFORMANCE BOOST: RF improves all metrics over Delphi model.	RF significantly outperforms traditional Delphi model across all metrics.	PRIVATE DATA: Proprietary data, not verifiable.
Machine Learning Approach for Credit Score Prediction, 2023	XGBoost, Support Vector Machine (SVM), AdaBoost	Comprehensive ensemble approach	Proposes adaptive ensemble method for better accuracy	High computational complexity → XGBoost SVM & Adaboost
Predicting Credit Risk: Using Machine Learning, 2023	Decision Trees, Gradient Boosting Classifiers	Regional focus (Nepal banking data)	Predicts credit risk in Nepalese banks	Limited algorithm scope Only for decision tree & GB
Comparative Study of Credit Scoring Models, 2022	Logistic Regression, Random Forest, Neural Networks	Compares traditional and modern models	Finds deep learning outperforms traditional methods	Dataset imbalance→ Negative impact on random forest & logistic reg.
Research on Credit Risk Assessment of SMEs, 2018	Logistic Regression, PCA, T-test	Tailored for SME credit risk; robust feature reduction via PCA	Uses logistic regression and PCA to assess SME. Develops custom indicators considering financial and non-financial factors.	Traditional models unsuited for SMEs; limited sample Diversity.

III. PROBLEM STATEMENT

Existing credit risk assessment systems are often complex and non- transparent, operating as black-box models where users cannot understand how decisions are made. This lack of interpretability reduces trust and makes it difficult for users and financial institutions to take informed actions. Moreover, many systems focus only on prediction accuracy and lack interactive features, clear explanations, and visual insights into the factors affecting credit scores. As a result, users are unable to identify weaknesses or improve their financial profile effectively.

Therefore, there is a need for an accurate, explainable, and user-friendly credit risk analysis system that not only predicts creditworthiness but also provides meaningful insights and recommendations for better decision- making.

IV. OBJECTIVES

- Design a clean, user-friendly dashboard with seamless navigation.
- Develop an end-to-end credit risk analysis platform (Frontend + Backend + ML + Score module).
- Provide explainability by showing key score-driving factors (DTI, utilization, credit history, account mix).
- Implement structured data preprocessing pipeline including data cleaning, missing value handling, encoding, normalization, feature selection, and outlier detection before model training and prediction.
- Integrate an EMI scheduler module to calculate EMI, generate amortization schedules, and support loan planning.
- Incorporate ALM (Asset-Liability Management) analysis to monitor cash flow mismatches, repayment timelines, and portfolio risk exposure.

V. PROPOSED SYSTEM

RiskNexus is a web-based credit risk analysis system that transforms raw user-provided financial attributes into actionable credit scores, default risk assessments, and recommendations. The system supports two primary user roles:

Customer: A retail user who wishes to understand and improve their credit health. The customer uses forms to enter age, income, debt, utilization, and credit history, and receives a simulated credit score along with recommendations and a visual overview of credit factors.

Bank / Lender: A lending officer or analyst who inputs application details (loan amount, DTI, credit score, employment history, loan purpose) and receives default risk estimates and credit grades (A–E), enabling quick scenario analysis and pre-screening.

RiskNexus conceptualizes the credit risk assessment process as a multi- stage pipeline, where a user's financial inputs are processed through:

(i) Profile capture and preprocessing – collecting structured inputs via web forms and validating them.
(ii) Score estimation – predicting a FICO-like credit score using a Random Forest model or a heuristic fallback.
(iii) Risk and grade estimation – estimating default probability and credit grades on the front end using interpretable heuristics.
(iv) Scenario-based impact analysis – simulating the impact of specific financial actions on the credit score.

This pipeline is realized through a combination of backend APIs and frontend logic, all accessible via a React-based user interface with authentication and role-based navigation.

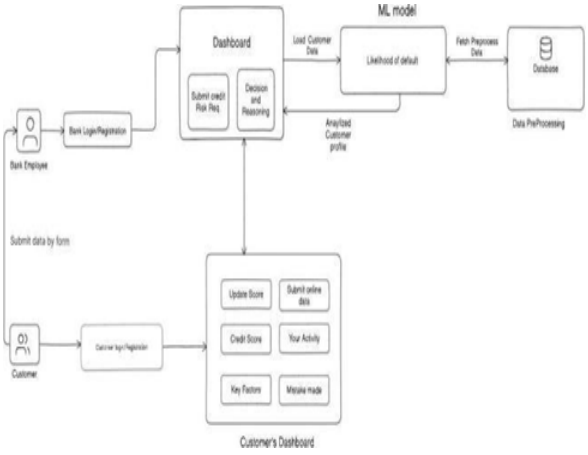
The RiskNexus system follows a three-layer architecture consisting of the frontend, backend, and machine learning layer. The React-based frontend collects user financial information and visualizes risk insights, while the Flask backend processes requests, performs credit score calculations, and manages APIs. The machine learning layer, using models such as Random Forest, analyzes financial attributes to generate credit risk predictions and provide decision support.

1. Data Ingestion and Validation - The system begins with data ingestion, where users enter their financial details through React- based forms. Customer inputs include age, income, debt, credit utilization, open credit lines, and credit history length, while bank users additionally provide loan-related details such as loan amount, debt-to-income

ratio, credit score, employment years, and loan purpose.

2. Credit Scoring and Risk Engine (Backend) - The backend is implemented using Flask APIs such as /api/calculate- score, /api/predict-impact etc. If the model is not already available, the system automatically triggers the training process and stores the trained model for future use. The backend then returns the predicted credit score along with clear explanation and personalized recommendations for better decision-making.

3. Risk Analysis and Visualization (Frontend) - The frontend dashboard is designed using Chart.js to visualize credit score trends and key risk factors in an interactive manner. It displays risk levels and assigns credit grades (A–E) using graphical indicators, making the results easy to understand. Additionally, the interface provides real-time insights and recommendations to help users analyze and improve their credit risk profile. This modular architecture separates backend processing from frontend visualization, ensuring scalability, clarity, and efficient credit risk analysis.



Detailed System Architecture

VI. ALGORITHM / FRAMEWORK

The RiskNexus system is designed using a hybrid framework that integrates machine learning, heuristic rules, and interactive data visualization to provide accurate, explainable, and user- friendly credit risk analysis. The framework combines a Random Forest–based prediction model with rule-based interpretability to ensure both performance and transparency. It follows a structured pipeline.

The overall workflow of the system can be represented as a sequential pipeline:
User Input →Data Validation → Backend Processing→ Credit Score Prediction → Generate Explanation & Recommendations → Display Results & User Insights.
Credit Score Estimation Algorithm (ML + Heuristic) - This algorithm estimates the credit score using a combination of machine learning and heuristic rules. It takes inputs such as age (a), annual income (y), total debt (d), credit-card utilization (u), number of open credit lines (L), and credit history length (h). Based on these parameters, the system generates a credit score in the range of 300 to 850 along with a concise summary explanation to help users understand the factors influencing their score.
Progress Analytics and Mastery Level Algorithm - This algorithm evaluates user performance by analyzing inputs such as test history, document identifier, and a recent test window. It computes an overall mastery level, identifies strengths and weaknesses, and provides recommendations.
AI-Powered Learning Recommendation System – This system generates personalized learning recommendations based on inputs like mastery level, weak topic set, and learning history. It prioritizes topics by considering accuracy and recency of practice, ensuring that the most critical areas are addressed first. The output is a ranked list of recommendations aimed at improving learning outcomes and supporting continuous progression.

Mathematical Formulation Heuristic Credit Score
First compute the Debt-to-Income Ratio (DTI):

$$\mathrm{DTI} = \frac{d}{\max(y, \epsilon)}$$

where:
d = total debt,
y = annual income,
ϵ = small constant to avoid division by zero. The credit score is calculated as: S=300+fDTI (DTI)+fu(u)+fh(h)+fL(L)+fa(a)

where the functions represent contributions from:
fDTI→ debt-to-income ratio fu → credit utilization

fh→ credit history length fL → open credit lines
fa → borrower age
The final score is bounded between 300 and 850: Sfinal=min(max(S,300),850)

VII. EXPERIMENTAL SETUP

The system is developed using a modern full-stack architecture that integrates both frontend and backend technologies for efficient credit risk analysis. The frontend is implemented using React with Vite for fast development and performance, along with Tailwind CSS for responsive UI design. Interactive visualizations are created using Chart.js, while form handling and validation are managed through React Hook Form and Zod. Firebase Authentication and Firestore are used to handle user identity and profile data securely. On the backend, the system is built using Flask with Flask-CORS to enable seamless communication between client and server.
Machine learning functionalities are implemented using libraries such as scikit- learn, pandas, and NumPy, with model serialization handled using joblib and pickle. The system is developed on standard computing environments (Intel i5/i7 processors with 8–16 GB RAM) and deployed on a lightweight virtual machine or container setup, where the frontend is served as static assets and the backend operates as a Flask service.

VIII. RESULTS AND IMPLEMENTATION

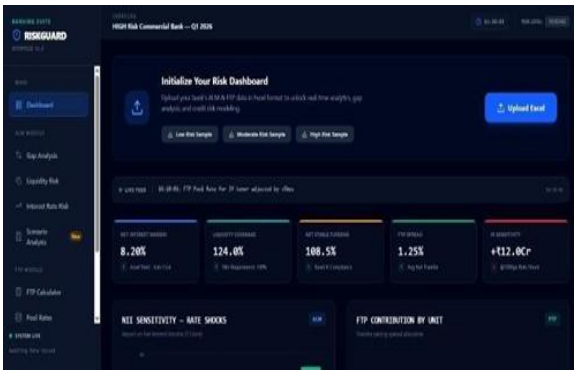
The RiskNexus (Credit Nexus) Risk Analysis Dashboard, which acts as an interactive interface for evaluating and predicting credit risk. The system is implemented using a React-based frontend integrated with a Flask backend, enabling seamless interaction between user inputs and machine learning predictions. The dashboard is divided into two main sections: Applicant Information and Predictions. On the left side, users input financial and loan-related details such as annual income, loan amount, debt-to- income ratio, credit utilization, number of open accounts, and other relevant attributes. These inputs are captured through structured form fields and validated before being processed.

Credit Risk Analysis Dashboard

On the right side, the prediction module displays the results generated by the backend model. Upon clicking the "Predict Default Risk" button, the system processes the input data using a trained Random Forest model to evaluate the applicant's credit risk. The output includes:

- Risk Classification: "Low Risk – Unlikely to Default"
- Risk Score: 14.4%
- Risk Level: Very Low
- Probability Distribution:
  - No Default: 85.64%
  - Default: 14.36%

These results indicate that the applicant has a low probability of default, demonstrating the effectiveness of the model in assessing creditworthiness. The dashboard also presents results in a clear and interpretable format, enabling users to easily understand their financial standing.

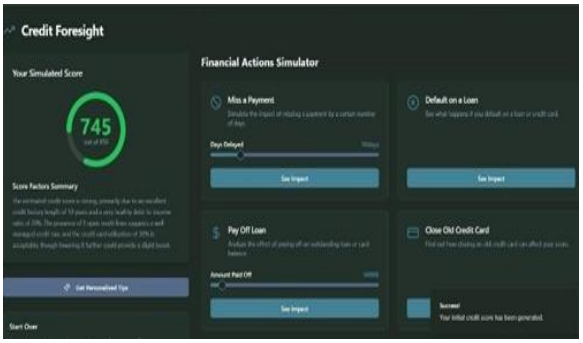


Banking Risk Analytics Dashboard (RiskGuard)

The figure represents a Banking Risk Analytics Dashboard (RiskGuard) designed for monitoring

and managing financial risks in a banking environment. It provides key metrics such as net interest margin, liquidity coverage ratio, net stable funding, FTP spread, and interest rate sensitivity, which are essential for risk assessment.

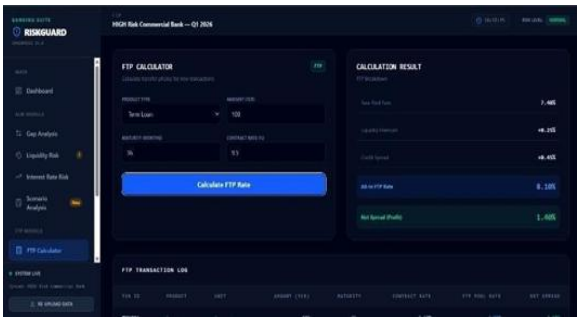
The dashboard supports data upload (Excel-based), real-time analysis, and live updates, along with modules like gap analysis, liquidity risk, and interest rate risk. It also includes visual components and indicators that help in understanding risk trends and making informed financial decisions.



Credit Foresight and Financial Actions Simulator Dashboard

The figure shows a Credit Foresight Dashboard designed to simulate and analyze the impact of financial actions on a user's credit score. It displays a simulated credit score (745/850) along with a summary of key factors affecting the score, such as credit history, debt-to- income ratio, and credit utilization.

The dashboard also includes a Financial Actions Simulator, where users can test different scenarios like missing a payment, defaulting on a loan, paying off debt, or closing a credit card. Each action provides insights into how these decisions may affect the credit score, helping users make informed financial choices and improve their credit profile.



FTP (Funds Transfer Pricing) Calculator Dashboard

The figure displays an FTP Calculator Dashboard within a banking risk management system. It allows users to input financial parameters such as product type, loan amount, maturity period, and contract rate to calculate the Funds Transfer Pricing (FTP) rate.

The dashboard provides a detailed calculation result, including components like base pool rate, liquidity premium, and credit spread, along with the final all-in FTP rate (8.10%) and net spread (profit). Additionally, it includes a transaction log section for tracking past calculations. This interface helps financial institutions analyze pricing strategies, profitability, and risk in lending operations

XI. MODEL EVALUATION AND OPTIMIZATION

The RiskNexus system is evaluated based on its ability to accurately predict credit scores and classify borrowers into different risk categories using the Random Forest model. The performance of the model is measured using evaluation metrics such as accuracy, precision, recall, and F1-score, which ensure that the system can correctly identify both high-risk and low-risk applicants. The model shows stable and reliable performance across different borrower profiles.

For optimization, the system focuses on effective data preprocessing and feature selection rather than complex tuning techniques. Important financial attributes such as income, liabilities, credit utilization, and credit history are selected to improve prediction quality. Data is cleaned, normalized, and properly formatted before being passed to the model, which helps in reducing errors and improving consistency. The trained model is stored using joblib or pickle, allowing fast loading and real-time predictions within the application.

Additionally, a heuristic-based scoring mechanism is implemented as a fallback to ensure that the system continues to provide credit scores even if the machine learning model is unavailable. This improves the reliability and robustness of the system.

Overall, the evaluation and optimization approach ensures that RiskNexus delivers accurate, consistent, and real-time credit risk predictions, making it suitable for practical financial applications.

X. CONCLUSION

RiskNexus shows that combining a lightweight machine learning model with transparent heuristic rules and a modern web interface yields an explainable, interactive credit card risk analysis platform. The system turns user-provided financial inputs into credit scores, default risk percentages, and credit grades, while offering simulations of how different actions affect risk.

ACKNOWLEDGEMENT

This work was supported by the Department of Computer Engineering at Rajiv Gandhi Institute of Technology. We thank Prof. Kaajal Sharma for his guidance.

REFERENCES

[1] Brown and C. Mues, "Credit scoring using machine learning and deep learning," *International Journal of Data Science and Analytics*, 2024.

[2] J. Chen, X. Li, and Y. Wang, "Credit risk prediction using explainable AI," in *Proc. IEEE Conf. Artificial Intelligence*, 2024.

[3] S. Kumar and R. Patel, "Application of AI in credit risk scoring for small business loans," in *Proc. IEEE Int. Conf. Financial Technology*, 2024.

[4] P. Sharma and A. Singh, "Machine learning approach for credit score prediction," *International Journal of Computer Applications*, 2023.

[5] B. Adhikari and S. Shrestha, "Predicting credit risk using machine learning," *Journal of Banking and Financial Technology*, 2023.

[6] M. Zhang, T. Liu, and B. Baesens, "Comparative study of credit scoring models," *Journal of Financial Data Science*, 2022.

[7] H. Li and Y. Zhao, "Research on credit risk assessment of SMEs," in *Proc. Int. Conf. Financial Engineering*, 2018.

[8] K. Jain and B. B. Gupta, "Machine learning techniques for credit risk assessment: A review," *IEEE Access*, vol. 7, pp. 159–173, 2019.