JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ANALYZING FINANCIAL AND BEHAVIORAL FACTORS BEHIND LOAN DEFAULTS

A Machine Learning Framework for Default Risk Prediction and Loan Recommendation

[1]Mr. Jawahar Jonathan K, [2]Dr. G Prakash Babu, [3]Suchithra S*, [4]K H Sumati*, [5]Sharath K S*, [6]Karthik M*

[1]Assistant Professor, [2]Professor, [3]BE, [4]BE, [5]BE, [6]BE
[1–6]Department of CSE
[1–6]Acharya Institute of Technology, Bengaluru, Karnataka
[1]jawahar2393@acharya.ac.in, [2]gprakash2757@acharya.ac.in, [3]Suchitras1218@gmail.com
[4]kh.24.becs@acharya.ac.in, [5]sharaths.24.becs@acharya.ac.in, [6]karthikm.24.becs@acharya.ac.in

Abstract : LoanAssist is an AI-assisted loan decision-support system designed to improve the efficiency and reliability of loan evaluation. The system analyzes an applicant's personal, employment, financial, loan, and property-related information to assess the likelihood of loan default. It uses a Random Forest classifier to predict default probability and classify applicants into different risk levels. At the same time, an Isolation Forest model identifies unusual application patterns that may indicate potential fraud or anomalies. The system also performs integrated property verification by checking ownership details, property value, legal status, registration information, and duplicate property usage. Based on the applicant's financial profile and predicted risk, the system provides personalized recommendations, including a suitable loan amount, suggested interest rate, and EMI. A bank officer dashboard allows authorized users to review the application details, property verification results, prediction outputs, anomaly indicators, and recommendations before making the final approval, rejection, or additional-review decision. Overall, the proposed system combines machine learning, property verification, anomaly detection, and rule-based financial analysis into a unified platform to support more informed and efficient loan decision-making.

Index Terms - LoanAssist, Loan Default Prediction, Credit Risk Assessment, Random Forest, Isolation Forest, Anomaly Detection, Fraud Detection, Property Verification, Loan Recommendation, Risk Classification, Machine Learning, Feature Engineering, Financial Risk Analysis, Decision Support System, Loan Approval, EMI Calculation.

I. INTRODUCTION

The loan approval process is an important financial activity for banks and lending institutions, as incorrect assessment of an applicant's repayment capability can lead to financial losses and increase the default risk. Traditional loan processing systems usually focus on collecting application details, checking eligibility criteria, and manual verification by bank officers. Property verification, fraud identification, and risk assessment can be handled as separate processes. Hence there is a need for an integrated system that can assist financial institutions in analyzing different risk factors for efficient loan decision-making.

Our proposed system is an AI-assisted loan decision-support system that analyzes the applicant's financial, employment, loan, and property-related information. The system uses a Random Forest classifier to estimate the probability of loan default and classify the applicant into Low, Medium, or High risk categories. The system also performs property verification using property records to check ownership matching, property value, registration information, legal status, and duplicate property usage. The Isolation Forest is used to identify unusual application patterns that may indicate potential anomalies or fraud.

Based on the prediction and verification results, the system provides a risk-aware loan recommendation including a suitable loan amount, suggested interest rate, and EMI. The bank officer can view the application details, property verification results, AI predictions, anomaly indicators, and recommendations through the dashboard before making the final decision. The system therefore combines machine learning, property verification, anomaly detection, and rule-based financial analysis into a unified platform while keeping the final loan approval or rejection decision with the bank officer.

II. BACKGROUND AND RELATED WORK

A. Machine Learning for Credit Risk and Loan Default Prediction

Traditional assessment of a loan involves the use of the applicant's information and factors that indicate the risk of the loan. Machine learning models have been increasingly used for credit-risk assessment and loan-default prediction [4], [5]. Classification models can be trained on past applicant and loan details to recognize patterns that predict default or repayment behavior, and earlier works have tested algorithms such as Random Forest, Decision Tree, Support Vector Machine, and Logistic Regression in

identifying applicants with higher default probabilities. These studies show that the financial and behavioral characteristics of an applicant can be analyzed to assess credit risk.

Jain et al. [1] use borrower characteristics, including credit histories, the purpose of the loan, and the level of debt, with the Random Forest algorithm and other machine learning algorithms to predict whether a loan will default. Dulgerov [2] evaluates different algorithms, including ensemble methods and deep learning, for forecasting loan defaults. Gashi [3] presents an approach that applies data cleaning through SMOTE and various machine learning models to the problem of anticipating loan defaults. Such methodologies can help bank officers assess the default risk of an applicant before the final approval of a loan. There remains a need for a system that integrates prediction with other risk perspectives and presents the results in a manner that is easy to interpret, which would help institutions assessing loan applications.

B. Property Verification and Financial Risk

When a loan is secured against property, verifying the associated property becomes important. Existing approaches have focused primarily on applicant financial and credit information [1], [3], while property-related verification has been less explored for the purpose of loan-default prediction. The proposed system addresses this issue by associating the property ID provided by the applicant with the registered property information and performing automated verification of ownership, legal status, duplicate usage, and other property-related details. These property-related risk factors can then be incorporated with the applicant's financial information in the loan assessment process [1], [5].

C. Fraud and Anomaly Detection in Loan Applications

Loan applications could have hidden patterns or irregularities that signal potential risks and therefore need additional checks. Machine learning-based anomaly detection can single out unusual applications for additional manual review before they are approved. Other studies have used Isolation Forest and other isolation-based techniques to detect financial anomalies and fraudulent patterns; however, most of these works focus on either default prediction or fraud detection. The proposed system uses the Isolation Forest algorithm, which isolates outliers by measuring an anomaly score that contributes to the assessment of a loan application's risk of being fraudulent. It is then up to the domain expert to determine whether this score is high enough for the application to be labeled as such [1], [3].

D. Integrated Loan Decision Support and Recommendation

While credit-risk assessment, fraud detection, and property verification can each be beneficial individually, the proposed system combines them with loan recommendation in a single decision-support system. The system pre-processes the application data and extracts features such as Debt-to-Income (DTI), Loan-to-Income, Property-to-Loan, Legal Risk, and Ownership Risk. The Random Forest model predicts the default risk, the Isolation Forest model produces an anomaly score, and rule-based algorithms verify property ownership and legal compliance [1], [3]. In addition, the system recommends a loan amount, interest rate, and EMI.

The decision to approve or reject the application remains with the bank officer, who reviews the applicant's financial information, property verification result, default risk, and anomaly score before making the final decision. Thus, the proposed system combines machine-learning-based default prediction, anomaly detection, property verification, and loan recommendation to help the bank officer assess the application [1], [2], [3].

III. Existing System and Limitations

The existing loan processing systems focus on gathering applications, eligibility checks, credit information reviews, and manual verification of the applicant's details. Machine learning approaches have been increasingly used for credit-risk and loan-default prediction based on financial and behavioral information [1], [2]. However, these approaches generally focus on individual risk-model prediction or comparison, whereas property verification, anomaly detection, and loan recommendation can be handled as separate steps [1], [3].

The conventional assessment also requires manual verification of property ownership, legality, value, and duplication, which can add to the processing time. Similarly, unusual patterns in an application may require manual investigation. Machine learning approaches such as anomaly detection can assist in detecting suspicious patterns, whereas property-related information can provide additional risk factors for the loan consideration [2], [3].

Therefore, there is a need for a system that unifies loan-default prediction, property verification, anomaly detection, and personalized loan recommendation to provide decision support to the bank officer in one system, while the officer retains responsibility for the final approval or rejection. The proposed system uses Random Forest for default-risk prediction, Isolation Forest for anomaly detection, and rule-based property verification and recommendation logic to provide the financial and property-related information, together with anomaly information, to the officer who takes the final decision [1], [2], [3].

IV. Proposed System and Methodology

The proposed system, LoanAssist, takes the applicant's personal, employment, financial, loan, and property details as input through a Streamlit-based application interface. Applicants register and log in to the system and then submit their loan applications. The application data is stored in a MySQL database (managed through XAMPP) and preprocessed for model training and prediction. The applicant's financial and behavioral factors, such as the source and stability of income, credit rating, outstanding debts, employment status, defaults, and requested loan amount, among others, are considered together with the related property information.

The system preprocesses the input data by encoding the input variables, scaling, and applying feature engineering techniques to generate financial and property features for model training and prediction. The Random Forest classification model is then applied to predict the likelihood of the applicant's default, and the applicant is assigned a risk level of Low, Medium, or High. The system also generates the DTI, loan-to-income ratio, property-to-loan ratio, and other features to enhance the prediction results. The default prediction results are displayed on the bank officer dashboard alongside the applicant's information.

The system uses the property information for property verification. The property registry is queried to confirm whether the property belongs to the applicant as indicated in the application form and to verify its nature, legality, and value. In addition, the system checks whether the same property has been applied for another loan in the banking system. Property verification is

performed using rule-based algorithms rather than a machine learning model, and the verification results are displayed on the bank officer dashboard alongside the default prediction results to support the officer's decision.

The system also detects anomalies and possible fraud in the loan application data. An Isolation Forest model is used to identify unusual and isolated data points in the application data that may indicate anomalies or fraud. The results are presented on the bank officer dashboard as flags that alert the officer to investigate further if an application has been submitted fraudulently to encumber the bank's resources. The Isolation Forest model is used for anomaly detection only; the risk level is determined by the Random Forest model.

Finally, the loan recommendation component recommends a loan amount and interest rate, and calculates the monthly installment (EMI) using a mathematical formula to form the recommended loan package. The recommendation depends on the applicant's financial status, the default prediction results, and the property verification results. The bank officer dashboard displays all the results required for making the final decision on the loan application. The recommendation is advisory only; the bank officer makes the final approval or rejection decision. The overall architecture of the proposed system is shown in Fig. 1.

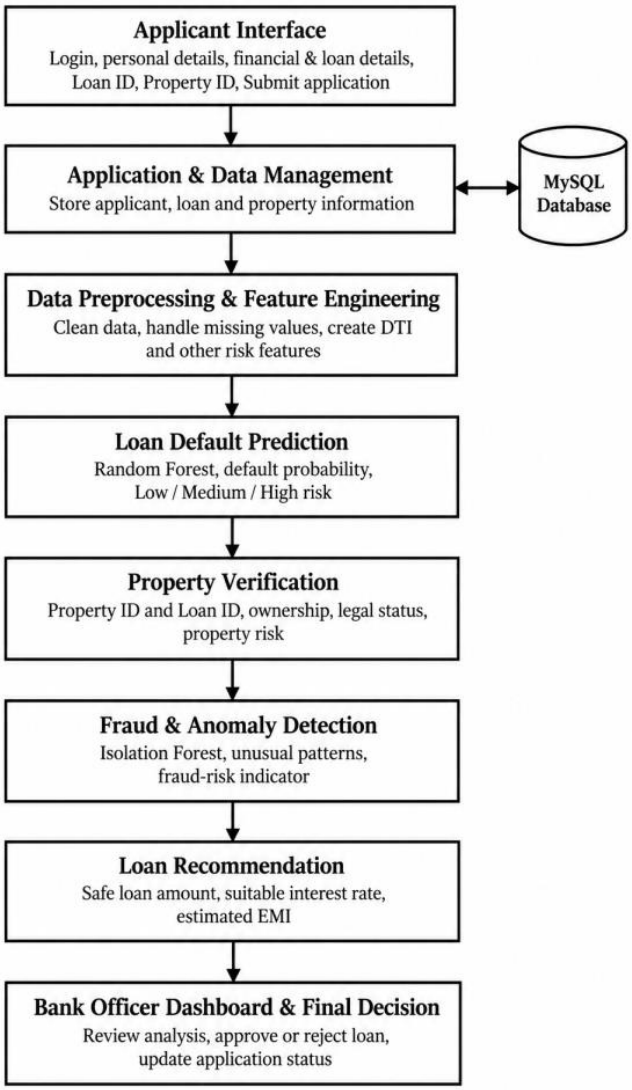


Fig. 1. Proposed system architecture.

A. Risk Scoring and Loan Recommendation

The proposed system produces three separate outputs for each application, each from a distinct component. The Random Forest model predicts the probability of default and assigns the applicant to a Low, Medium, or High risk level. The Isolation Forest model highlights anomalies in the application through an anomaly score and flag. The rule-based property verification component examines ownership, legal status, property value, and duplicate use of the property. These outputs are presented together on the bank officer dashboard so that the officer can assess the application as a whole. Based on the predicted risk level, the applicant's financial details, and the property verification results, the system recommends the amount of money to be provided as a loan, the corresponding interest rate, and the EMI for the officer's consideration. The end-to-end processing workflow, from application entry to the officer's decision, is shown in Fig. 2.

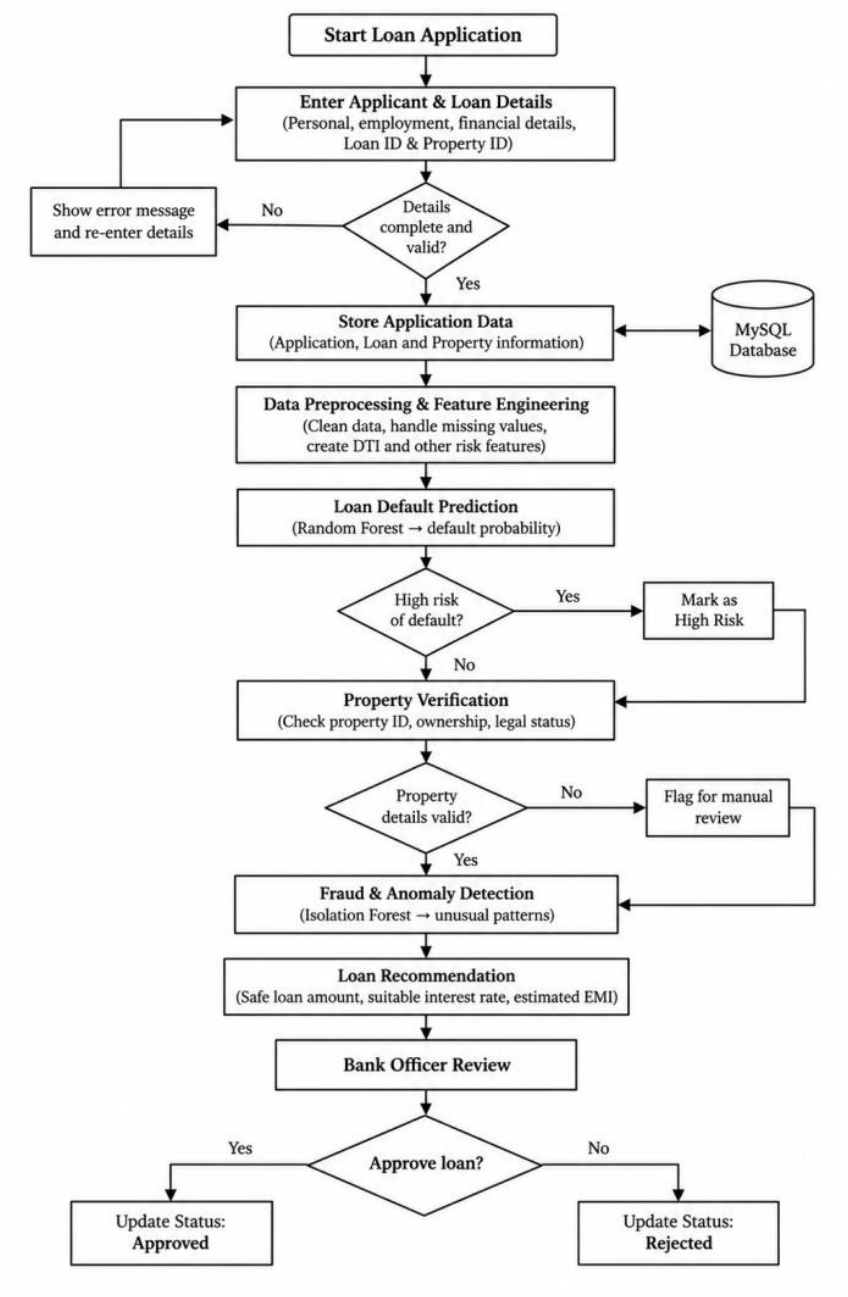


Fig. 2. Loan application processing workflow, from application entry to the bank officer's approve or reject decision.

B. Security and Privacy Measures

Access to the system is provided through user login, with separate interfaces for applicants and bank officers. The application data, the analysis results, and the bank officer's final decisions are stored in the MySQL database. In addition, the Isolation Forest component flags unusual application patterns for manual review by the bank officer.

C. Deployment and Monitoring

The system is deployed as a web application using the Streamlit framework, with MySQL (through XAMPP) used for data storage. It offers applicants and bank officers an interface where they can submit loan applications, check property verification results, examine default-risk prediction findings, and review loan recommendations. The bank officer dashboard enables officers to assess the applications, analyze the prediction and anomaly results, and record the final decision. The system stores the applications, the recorded decisions, and the analysis results to support the continuous improvement of the loan risk assessment process. Figure 3 summarizes the eight stages of the proposed methodology, from stakeholder identification to deployment and monitoring.

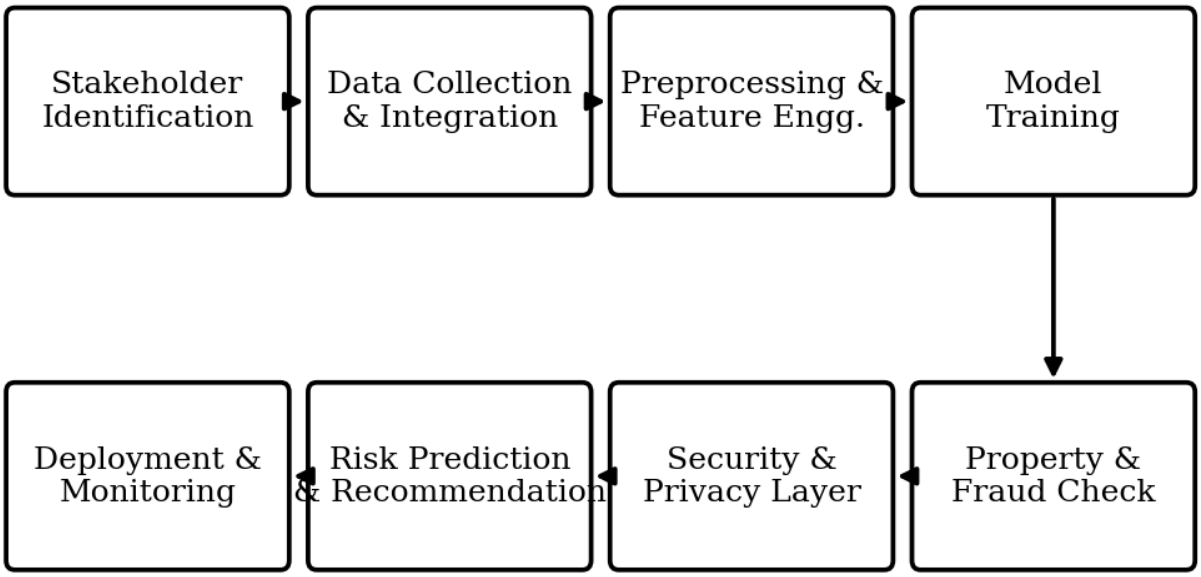


Fig. 3. Proposed methodology workflow.

The software and hardware requirements of the proposed system are listed in Table 1.

Table 1. Software and hardware requirements

Category	Specification
Programming language	Python 3.10+
Web framework (user interface)	Streamlit
Database	MySQL (through XAMPP)
Machine learning models	Random Forest and Isolation Forest (Scikit-learn)
Core libraries	Pandas, NumPy, Scikit-learn, Joblib, Plotly
Development tools	VS Code
Minimum hardware	Intel i5-class CPU, 8 GB RAM, 256 GB storage

V. Results and Discussion

The proposed loan decision-support system was successfully implemented to capture the applicant's details, process the data, verify the property details, and assess the loan risk. The system employed the Random Forest model to predict the likelihood of default and classify the applicant into Low, Medium, or High Risk. The Isolation Forest model was also used to detect any anomalies in the application.

The property verification module was successfully used to verify the ownership, legality, property information, and duplicate property usage. The system then recommended an appropriate loan amount, interest rate, and EMI based on the risk and financial information.

The results were displayed on the bank officer dashboard so that the officer could evaluate the application, prediction, property verification, and recommendation before taking the final decision. Therefore, the system provides a simple and integrated solution to the loan assessment and decision-making process.

The proposed model achieved an accuracy of 84.14%, indicating that 84.14% of the loan applications were correctly classified as default or non-default. The test cases used to evaluate the system and their observed outcomes are summarized in Table 2.

Table 2. Results and evaluation

Test Case	Condition	Expected	Observed
Application	Details entered correctly	Application submitted	Submitted
Missing Details	Some details are missing	Error shown	Error shown
Default Prediction	Application is checked	Risk level shown	Risk shown
Property Check	Property details entered	Property verified	Verified
Fraud Check	Unusual details found	Fraud warning shown	Warning shown
Loan Recommendation	Analysis completed	Loan details suggested	Recommendation shown
Approve Loan	Officer approves	Status becomes Approved	Approved
Reject Loan	Officer rejects	Status becomes Rejected	Rejected

VI. Future Work

In the future, the LoanAssist system could be improved by implementing more complex machine learning models to predict loan-default probabilities and assess loan risk. The system could also benefit from using more extensive and diverse datasets for training the machine learning models, making the system more accurate for different categories of applicants. Additionally, the property verification feature could be expanded by implementing real-time property information and document verification. The fraud detection feature could be improved by analyzing more application and transaction-related patterns. Finally, there is an opportunity to add explainable artificial intelligence to make the reasoning behind suggested risk levels or recommendations transparent, and to test the system on actual loan examples to assess its effectiveness.

VII. Conclusion

The proposed LoanAssist decision-support system is an integrated platform that utilizes machine learning, property verification, anomaly detection, and loan recommendation for processing loan applications. The system employs the Random Forest algorithm for predicting loan-default risk and the Isolation Forest algorithm for detecting anomalous patterns in applications. In addition, it verifies property ownership and related information and recommends an appropriate loan amount, interest rate, and monthly installment suited to the financial and risk information provided by the applicant. This information assists the bank officer in considering multiple factors associated with a loan application and in making an informed final decision. Thus, the proposed loan decision-support system presents a simple and efficient framework for decision-making that can be further enhanced through the use of larger datasets, advanced machine learning techniques, real-time verification, and explainable algorithms.

References

[1] Y. K. Jain et al., "Effective Machine Learning-Based Predictive Analytics for Loan Default Prediction in Banking Sector," IEEE CCIS, 2024, doi: 10.1109/CCIS63231.2024.10931843.

[2] E. E. Dulgerov, "Comparing Different Algorithms when Predicting Loan Defaults," IEEE TELECOM, 2025, doi: 10.1109/TELECOM66943.2025.11304098.

[3] P. Gashi, "Loan Default Prediction Model," 2023, doi: 10.13140/RG.2.2.22985.01126.

[4] P. Isaac, "Prediction of Credit Risk using Machine Learning Models," 2022.

[5] A. Soomro et al., "Loan Default Prediction Using Machine Learning Algorithms: A Systematic Literature Review 2020–2023," 2024, doi: 10.57239/PJLSS-2024-22.2.00469.