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. 14, Issue 7, July 2025

DOI: 10.17148/IJARCCE.2025.14721

LOAN APPROVAL PREDICTION USING MACHINE LEARNING

Velvigneswar. J[1], Dr. P. Senthil Kumari[2]

Final Year Student of M.E. CSE, T.J Institute of Technology, Karapakkam, Chennai[1]

Professor, Head of the Department, T.J Institute of Technology, Karapakkam, Chennai[2]

Abstract: Loan approval is a critical process in the banking and financial sector, requiring accurate and timely decision-making to ensure effective risk management for institutions and financial support for applicants. Traditional loan processing methods are often manual, time-consuming, and susceptible to human bias or inconsistency, which can result in delayed or inaccurate decisions. To address these challenges, this project proposes a machine learning-based solution using Random Forest, Support Vector Machine, Logistic Regression and Decision Tree algorithms to predict the likelihood of loan approval. The system is trained on historical loan data, including features such as income, employment status, credit history, education level, marital status, and loan amount, to identify meaningful patterns that distinguish approved from rejected applications. Logistic Regression offers a simple and interpretable model for binary classification, while and robustness by aggregating predictions from multiple decision trees. In addition to prediction functionality, the application includes a secure login and registration module, where user credentials are stored in a database to maintain account integrity. Users can enter loan application details through a clean and user-friendly web interface, with all input data securely saved for processing and analysis. The system delivers real-time prediction results, helping applicants quickly understand their chances of loan approval. This intelligent and scalable solution not only reduces the workload on financial officers but also enhances consistency, transparency, and efficiency in the loan approval process, paving the way for smarter decision-making in modern banking systems.

Keywords: Loan Approval Prediction, Machine Learning, Logistic Regression, Random Forest Classifier, Deep Learning, Loan Application System.

I. INTRODUCTION

In today's fast-paced financial environment, the demand for automated and intelligent loan approval systems has significantly increased. With the rising number of loan applications and the need for quick yet accurate decisions, traditional methods—which are often manual, time-consuming, and dependent on subjective human judgment—are proving to be inefficient and inconsistent. Delays in processing, human bias, and clerical errors can result in financial loss for institutions and dissatisfaction among applicants.

To address these challenges, this project leverages the power of machine learning to develop a reliable and data-driven loan approval prediction system. Specifically, Random Forest, Support Vector Machine, Logistic Regression and Decision Tree algorithms to predict the likelihood of loan approval These parameters include income, employment status, credit history, education, marital status, and loan amount, all of which are key indicators in assessing creditworthiness.

II. BACKGROUND AND MOTIVATION

In the banking and financial services sector, loan approval is a critical function that directly impacts both the financial institution's profitability and the applicant's financial well-being. Traditionally, loan approval processes have relied heavily on manual evaluations performed by loan officers, who assess applicants based on financial documents, credit history, employment status, and personal interviews. While this method has served institutions for decades, it comes with significant limitations such as human error, processing delays, inconsistencies, and potential bias.

With the rapid growth of data and digital transformation in the finance industry, there is a strong need to modernize loan assessment mechanisms using automated, data-driven technologies. One such powerful tool is machine learning (ML), which can uncover hidden patterns in historical data to predict future outcomes with high accuracy.

The motivation behind this project stems from the desire to improve:

Speed: Automating the process significantly reduces decision time.

Accuracy: Predictive models can analyze thousands of variables more accurately than a manual reviewer.

Fairness and Transparency: Data-driven algorithms reduce the risk of human bias.

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. 14, Issue 7, July 2025

DOI: 10.17148/IJARCCE.2025.14721

Scalability: A machine learning system can handle a large number of applications simultaneously.

Cost Efficiency: Automation reduces operational costs for financial institutions.

III. LITERATURE REVIEW

1.**TITLE:** Tree-Based Methods for Loan Approval

Authors: Mohamed Alaradi

Description:

Loan approval is one of the most important processes that any banking organization owns. The acceptance or rejection of any loan application has a direct impact on the bank revenue and the profitability in quarterly issued financial statements. Though loan approval is a critical process, the actual decision made is not a straightforward procedure and comes with a lot of uncertainties. Recently, statisticians and data scientists have tried to automate this process to minimize risk and increase profitability by applying different statistical learning methods. In this work we explore a framework with an application by applying tree-based methods on publicly available dataset. This work aimed at developing a high performance predictive model for loan approval prediction using decision trees. Experiments were made in different varieties of tree methods ranging from the most simplified and comprehensible decision tree reaching up to the most complex random forests. Results yielded inadequate performance with respect to simplified decision trees due to the highlight correlated and complex feature space, majority of critical parameters affecting loan approval was not reflected upon and yielded an impractically over-simplified tree. However, boosting came in superior in terms of performance, relevance and interpretation via the importance chart scoring accuracy on testing dataset [98.75%] specificity [100%], Minority class prediction accuracy [92.85%], and classification efficiency of [97.0%]. Therefore, boosting-based decision-tree predictive model was recommended to facilitate decision making regarding the eligibility of loan applicants based on their characteristics.

2.**TITLE:** Evaluating Consumer Loans Using Neural Networks Ensembles

Authors: Maher Alaraj

Description:

Banks should take high care on their loan granting policies. Banks rely on credit scoring systems when it comes to granting loans to customers to reduce any potential losses. Neural networks are considered as mostly wide used statistical tool in finance and business applications. Recent studies emphasize using ensemble models or multiple classifiers over single ones to solve credit scoring problems. This study focuses on 2 parts: (1) developing a group of neural network ensemble models to assist in predicting the probability of default; (2) exploring the importance of each of the attributes of the dataset in the overall performance of the classifiers. The datasets used in this study are real world datasets

3.TITLE: Bank Loan Prediction System using Machine Learning

Authors: Anshika Gupta

Description:

With the advancement in technology, there are so many enhancements in the banking sector also. The number of applications is increasing every day for loan approval. There are some bank policies that they have to consider while selecting an applicant for loan approval. Based on some parameters, the bank has to decide which one is best for approval. It is tough and risky to check out manually every person and then recommended for loan approval. In this work, we use a machine learning technique that will predict the person who is reliable for a loan, based on the previous record of the person whom the loan amount is accredited before. This work's primary objective is to predict whether the loan approval to a specific individual is safe or not.

4.**TITLE:** Predictive Analysis of Loan Approval Using Machine Learning Algorithms

Authors: Sneha Patil

Description:

This research explores predictive analytics for automating the loan approval process using popular machine learning algorithms including K-Nearest Neighbors (KNN), Decision Tree, and Gradient Boosting. The study emphasizes the importance of data preprocessing, especially handling missing values and encoding categorical features, to boost model

DOI: 10.17148/IJARCCE.2025.14721

performance. Evaluation metrics such as precision, recall, F1-score, and accuracy were used to compare models. Gradient Boosting emerged as the most effective technique with an accuracy of 93.2%, followed by Decision Trees. The authors recommend integrating such systems in financial institutions to speed up decision-making and reduce human effort.

5.**TITLE:** Credit Risk Evaluation Using Machine Learning Models for Loan Prediction

Authors: Ritu Agarwal

Description:

This paper focuses on evaluating credit risk using machine learning techniques to predict the likelihood of loan default or approval. The study incorporates algorithms like Logistic Regression, Support Vector Machines, and XGBoost on a public loan dataset containing financial and demographic information of applicants. Emphasis was placed on imbalanced data handling using SMOTE (Synthetic Minority Over-sampling Technique) to improve minority class prediction. Among all models, XGBoost demonstrated the highest performance, achieving a testing accuracy of 95.6%, with high precision and recall for both classes. The study concludes that advanced machine learning models are effective in real-world financial risk assessment and can assist banks in minimizing default rates.

IV. ANALYSIS AND DISCUSSION

1 Overview

The objective of this project was to predict whether a loan applicant is likely to get loan approval based on historical data and various applicant parameters such as income, credit history, marital status, education, and property area. The analysis involved data preprocessing, model building, evaluation, and result interpretation.

---

2 Dataset Description

The dataset used for this project was collected from a public loan dataset (such as from Kaggle). It contains the following key features:

- Categorical Variables: Gender, Married, Education, Self_Employed, Property_Area, Credit_History, etc.
- Numerical Variables: ApplicantIncome, CoapplicantIncome, LoanAmount, Loan_Amount_Term.
- Target Variable: Loan_Status (Y/N)

A total of 614 observations were analyzed.

---

3 Data Preprocessing

Before applying machine learning models, the following preprocessing steps were performed:

- Handling Missing Values: Mode and mean/median imputation techniques were used.
- Encoding Categorical Variables: Label Encoding and One-Hot Encoding were applied.
- Feature Scaling: Normalization was applied to income and loan amount variables.
- Balancing Data: Class distribution was checked; SMOTE was used if imbalance existed.

---

4 Model Building

Three different machine learning algorithms were applied:

1. Logistic Regression
2. Random Forest Classifier
3. Decision Tree Classifier

Each model was trained on 80% of the dataset and tested on the remaining 20%.

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. 14, Issue 7, July 2025

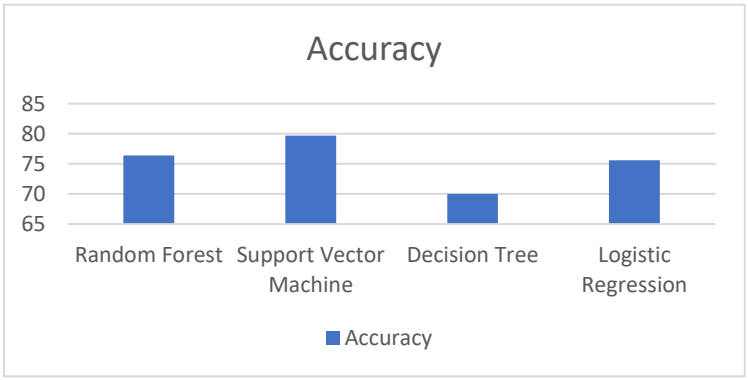
DOI: 10.17148/IJARCCE.2025.14721

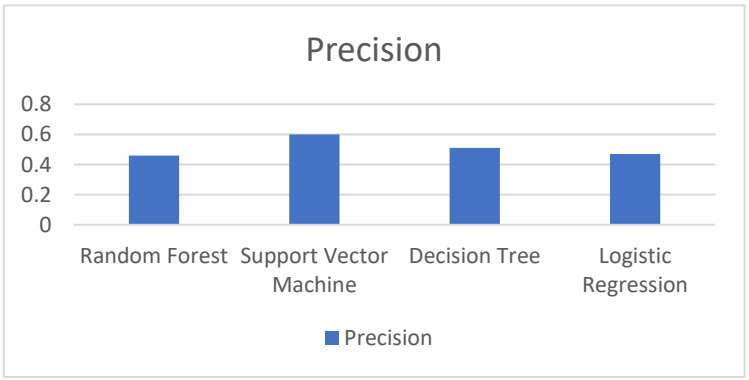
5 Model Evaluation

Evaluation metrics used:

- Accuracy
- Precision
- Recall
- F1 Score
- ROC-AUC Curve

Algorithm	Accuracy (in %)	Precision	Recall	F1 Score
Random Forest	76.42	0.46	0.80	0.58
Support Vector Machine	79.67	0.6	0.95	0.64
Decision Tree	70	0.51	0.60	0.55
Logistic Regression	75.60	0.47	0.69	0.61





Support Vector Machine outperformed other models in all metrics, indicating its superior ability to capture complex patterns in the data.

6 Feature Importance

The most influential features in loan approval were:

- Credit_History (most important)
- ApplicantIncome
- LoanAmount

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 7, July 2025
DOI: 10.17148/IJARCCE.2025.14721

- Marital_Status
- Education

This suggests that an applicant's credit history and income levels heavily influence approval decisions.

---

7 Insights and Interpretation

- Applicants with a credit history were significantly more likely to be approved.
- Higher applicant income increases the chances of approval, especially when combined with low loan amounts.
- Married applicants and graduates had slightly higher approval chances.
- Urban and semi-urban applicants had better approval rates than those from rural areas.

---

8 Discussion

- Model Performance: SVM was the best-performing model due to its ability to handle non-linear relationships and feature interactions. Logistic Regression, while interpretable, showed slightly lower accuracy.
- Business Impact: The model can help financial institutions automate initial screening, reducing manual effort and speeding up the loan process.
- Limitations:
  - Small dataset may limit generalizability.
  - No behavioral data like past EMI repayment, which could improve accuracy.
  - Data imbalance can still affect results despite mitigation techniques.
- Future Work:
  - Use of XGBoost or LightGBM models.
  - Incorporating real-time credit score data.
  - Deploying the model using Flask/Django into a web-based interface for bank officers.

V. CONCLUSION

The loan approval prediction system using machine learning algorithms—Support Vector Machine—provides an effective, accurate, and efficient solution to the traditional challenges faced in manual loan processing. By analyzing historical loan data and identifying patterns among key applicant features such as income, credit history, loan amount, and employment status, the system offers consistent, fair, and data-driven predictions. Logistic Regression enables straightforward binary classification with high interpretability, making it ideal for initial decision-making.

The integration of machine learning into the loan approval workflow not only accelerates the decision-making process but also enhances transparency and reduces human error. The system is equipped with additional functionalities like secure login, user registration, and real-time storage of application data in a backend database. These features ensure a user-friendly experience while maintaining data integrity and security. Furthermore, the modular design of the system allows for future enhancements, such as integration with credit bureau APIs, real-time fraud detection, and the use of advanced algorithms like XGBoost or deep learning models.

REFERENCES

[1]. 1.Kumar Arun, Garg Ishan, Kaur Sanmeer, *Loan Approval Prediction based on Machine Learning Approach*.
[2]. 2. Adyan Nur Alfiyatin, Hilman Taufiq, Ruth Ema Febrita, Wayan Firdaus Mahmudy, ' Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization ': *International Journal of Advanced Computer Science and Applications* (Vol. 8, No. 10, 2017).
[3]. Mohamed El Mohadab, Belaid Bouikhalene, Said Safi, ' Predicting rank for scientific research papers using supervised learning ' *Applied Computing and Informatics* 15 (2019) 182–190.
[4]. K. Hanumantha Rao, G. Srinivas, A. Damodhar, M. Vikas Krishna : Implementation of Anomaly Detection Technique Using Machine Learning Algorithms : *Internatinal Journal of Computer Science and Telecommunications* (Volume 2, Issue 3, June 2011).
[5]. J. R. Quinlan. Induction of Decision Tree. *Machine Learning*, Vol. 1, No. 1. pp. 81–106., 1086.
[6]. G. Arutjothi, C. Senthamarai : Prediction of loan status in commercial bank using machine learning classifier, *International Conference on Intelligent Sustainable Systems (ICISS)*, 2017.
[7]. 7.J.R. Quinlan. *Induction of decision trees*. Machine learning Springer, 1 (1): 81–106, 1086.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 7, July 2025

DOI: 10.17148/IJARCCE.2025.14721

[8]. Mohammad Ahmad Sheikh, Amit Kumar Goel, Tapas Kumar, "An Approach for Prediction of Loan Approval using Machine Learning Algorithm ", *IEEE ICESE* - 2020.

[9]. S.S. Keerthi and E.G. Gilbert. *Convergence of a generalize SMO algorithm for SVM classifier design*. Machine Learning, Springer, 46 (1): 351–360, 2002.

[10]. J.M. Chambers. Computational methods for data analysis. *Applied Statistics*, Wiley, 1 (2): 1–10, 1077.

[11]. 11.Amit Kumar Goel, Kalpana Batra, Poonam Phogat, "Manage big data using optical networks ", *Journal of Statistics and Management Systems* " Volume 23, 2020, Issue 2, Taylors & Francis.

[12]. 12.Raj, J. S., & Ananthi, J. V., "Recurrent neural networks and nonlinear prediction in support vector machine " *Journal of Soft Computing Paradigm (JSCP)*, 1 (01), 33–40, 2019.

[13]. 13.Aakanksha Saha, Tamara Denning, Vivek Srikumar, Sneha Kumar Kasera. "Secrets inSource Code: Reducing False Positives usingMachine Learning ", *2020 InternationalConference on Communication Systems &Networks (COMSNETS)*, 2020.

[14]. 14.X. Frencis Jensy, V.P. Sumathi, Janani Shiva Shri, "An exploratory Data Analysis for Loan Prediction based on nature of clients ", *International Journal of Recent Technology and Engineering (IJRTE)*,Volume- 7 Issue- 4S, November 2018.

[15]. 15.Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, K Vikash, "Loan Prediction by using Machine Learning Models ", *International Journal of Engineering and Techniques*.Volume 5 Issue 2, Mar-Apr 2019.

[16]. 16.Nikhil Madane, Siddharth Nanda, "Loan Prediction using Decision tree ", *Journal of the Gujrat Research History*,Volume 21 Issue 14s, December 2019.

[17]. 17.Bing Liu, "*Sentiment Analysis and Opinion Mining*," Morgan & Claypool Publishers, May 2012.

[18]. 18.Bing Liu, "*Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*," Cambridge University Press, ISBN: 978–1-107-01789–4.

[19]. 19.Shiyang Liao, Junbo Wang, Ruiyun Yu, Koichi Sato, and Zixue Cheng, "CNN for situations understanding based on sentiment analysis of twitter data," *Procedia computer science*, 111 : 376–381, 2017.CrossRef.

