

PREDICTION OF MODERNIZED BANK LOAN APPROVAL USING MACHINE LEARNING

[1]Rekansh Rao, [2]KunthamAnusha, [3]Ch.KanakaMahalakshmi, [4]Madhuri Gumpula
[1,2,3,4]Department of Computer Science and Engineering (AI&ML), St.Peters Engineering College, Maissammaguda, Dhulapally,Kompally, Medchal-500100,
E-Mail: kvvbdurgaprasad@stpetershyd.com

Abstract
There are many changes made by technology in human life, especially in banking, where there is a revolution caused with the introduction of automation and machine learning. This project aims at utilizing historical datas with advanced machine learning algorithms for loan approval prediction. Models such as logistic regression, random forests, and support vector machines are used for optimizing prediction accuracies, efficiency, and risk management. This implementation entails data preprocessing techniques as well as their evaluation metrics like accuracy, precision, recall, and F1 score. This study finds the effect of machine learning in loan approval systems and discusses the way forward in the financial domain for further enhancements.
Keywords: Machine Learning, Bank Loan Approval Prediction, Random Forest, Logistic Regression, Data Pre processing, Financial Risk Management, Decision Trees.

1. Introduction

The increasing dependency on technology in financial institutions has made it essential to have automated and efficient solutions to meet changing demands. This project aims to develop a machine learning-based modernized loan approval system that is designed to improve the accuracy, efficiency, and reliability of loan decisions [1]. Advanced analytics combined with user-centric design will help streamline the loan approval process, ensuring that decision-making is in line with regulatory standards and user expectations.

Traditional loan approval mechanisms are not very efficient and are not free from biases, so they cannot guarantee reliable outputs. This project introduces a framework that can identify complicated patterns in data, thus providing the ability to make accurate decisions and maintain consistency [2][3]. This evolution in the decision-making process bridges gaps in the traditional methods and redefines the lending experience for financial institutions as well as their customers.

The scope of the project encompasses several key phases, which include defining a clear problem statement, gathering and preprocessing applicant data, selecting suitable algorithms, and building predictive models to assess creditworthiness. There is also a description of the deployment strategies and a roadmap for future improvements to ensure scalability and adaptability across different banking environments.

The purpose of this project is to transform the traditional loan approval process by making use of machine learning techniques that analyze comprehensive applicant data. This data-driven approach diminishes risk, ensures fairness, and opens up roads to inclusive financial services [4]. The objectives include automating and optimizing the loan approval process, improving decision accuracy, enhancing operational efficiency, and faster, fairer outcomes to elevate customer satisfaction. It is also architected to establish a strong framework that can adapt to future advancements in financial technology.

2. Related Work

The domain of loan approval systems has undergone significant evolution as time passed from traditional manual techniques to more data-driven ways. Traditional systems rely heavily upon rule-based evaluations, manual reviews, and financial formulae to pass judgment on creditworthiness [5]. These practices are prone to inefficiency, subjectivity, and biases, often resulting in inconsistent outcomes. Moreover, their inability to handle a large dataset and uncover more complex patterns in applicant data limits their scalability and the effectiveness of the system in place.

The recent breakthroughs in the field of machine learning have garnered interest in implementing these approaches to assess credit risk. Many researchers have proved through various experiments that machine learning models are indeed capable of anticipating loan approval using historical data analysis [6]. For example, ensemble models such as Random Forests and Gradient Boosting Machines have seen wide applicability to augment the classification accuracy for determining credit defaults. It is also implemented due to its simplicity and interpretability of Logistic Regression. However, the imbalanced dataset demands that most real-world datasets have a skewed distribution between approved and rejected applications.

Current systems typically fail to account for such vital factors as diverse data sources (e.g., alternative financial behaviors, spending patterns, and socio-economic factors), mitigation of biases, and adherence to fairness standards. While fairness-aware machine learning research strives to address potential biases that might impact certain groups, the current use of these systems in financial decision-making systems is still not extensive [7][8].

Other focuses were to develop model explainability for better trust among the stakeholders and regulators. The tools used include SHAP, or Shapley Additive Explanations, and LIME, which are Local Interpretable Model-Agnostic Explanations [9][10]. Although they have been helpful, it remains in its infant stages on how to introduce them in real-world banking workflows.

Overall, the present body of work, however, demonstrates the need for further comprehensive, scalable, and user-centric solutions that take care of the technical as well as the ethical challenges of such loan approval systems. The current work is to bridge the gap by providing an advanced loan approval system that is both fair and explainable with modern financial practices and regulatory standards.

3. Proposed Work

This project proposes a modernized loan approval system that uses advanced machine learning techniques to overcome the limitations of traditional methods. The proposed system aims to automate and enhance the loan approval process by analyzing comprehensive applicant data, including demographics, credit history, income, and financial behavior. Key features of the proposed framework include:

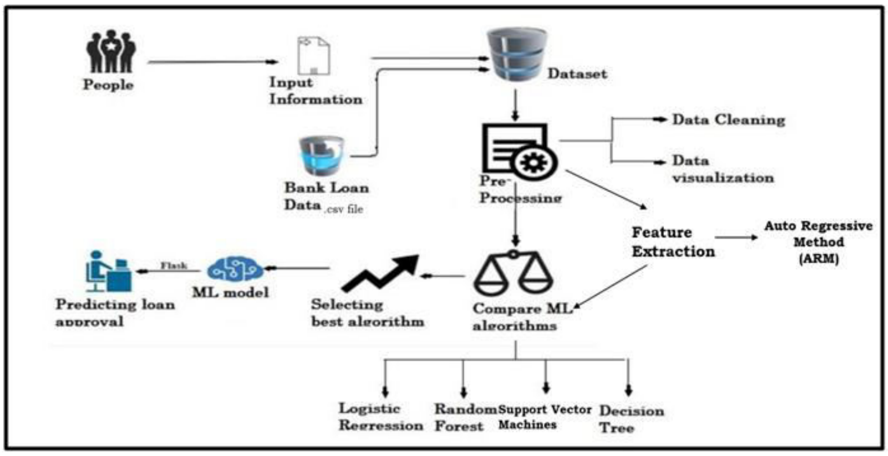

Fig: 1 System Architecture

Support Vector Machine (SVM):

SVM is used to classify loan applications as approved or rejected based on features such as demographics, financial details, and loan specifics. The process involves data preparation (handling missing values, normalization, and encoding categorical variables) and selecting kernels like linear, polynomial, or RBF. Hyperparameters are optimized through techniques like Grid Search, and evaluation metrics such as accuracy, precision, recall, and ROC-AUC ensure model reliability [11][12].

Logistic Regression:

Logistic Regression approximates the probability of loan approval based on the fit of a logistic function to the input features. Hyperparameter tuning is done in order to optimize the model performance. Evaluation is based on accuracy, precision, and recall to ensure good generalization of the model [13].

Random Forest Classifier:

This ensemble approach combines several decision trees to improve the accuracy of classification. After preprocessing and encoding data, the Random Forest builds trees on random subsets of data and features. Hyperparameters such as the number of trees and depth are tuned, and the model is evaluated using standard metrics to ensure robustness [14].

Decision Trees:

Decision Trees classify loans by recursively splitting data based on feature values, forming a tree structure where each node represents a decision rule. Preprocessing involves cleaning and encoding data, with the tree trained to predict approved or rejected applications [15].These models collectively support a robust and scalable system for modernized loan approval predictions.

The proposed system not only addresses the inefficiencies and biases of traditional methods but also paves the way for more accurate, efficient, and inclusive financial services. Through continuous monitoring and updates, the framework ensures adaptability to evolving market trends and user needs [16][17][18].

4. Dataset Description

Our dataset, obtained from the GitHub website, has a total of 615 instances and covers various attributes for identifying whether user profiles are genuine or not. For the loan approval prediction project, the data includes several key features [19][20][21]. The applicant demographics include age, gender, and marital status, which provide insight into the applicant's life stage, potential financial stability, and personal circumstances. Some features of financial information include income, employment status, existing debts, and credit score, all of which are very vital to determining whether the applicant has the capacity to repay the loan and his general health. The loan details included are the loan amount, the loan term, and the purpose of the loan-all these determine the risk and decide whether the application is accepted or declined. The target variable is the loan approval status, a binary indicator of whether the loan was approved or rejected. Additional features may include residence type, education level, and number of dependents, which further provide context on the applicant's financial situation and stability [22][23]. Preparing data ensures to deal with missing values, removes outliers, and convert categorical variables into numerical formats to make the data suitable and clean enough for machine learning [24][25].

Precision

The precision measure can be calculated by the number of true positive results divided by the number of positive results predicted by the classifier.

$$\mathrm{PRE} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})} \quad (1)$$

Recall

The recall measure can be calculating the number of correct positive results divided by the number of all relevant samples.

$$\mathrm{REC} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Negative})} \quad (2)$$

Accuracy

The accuracy measure can be calculating the number of correct predictions model divided by the total number of input samples.

$$\mathrm{Acc} = \frac{\mathrm{TP+TN}}{\mathrm{TP+FP+FN+TN}} \quad (3)$$

F1 -measure

The F1-measure (harmonic mean) is used to show the balance between the precision and recall measures. The F1- score measure can be calculated as follow:

$$F = 2 * \frac{\text{Precision}*\text{Recall}}{(\text{Precision}+\text{Recall})} \quad (4)$$

5. Result Analysis

The result analysis of the modernized loan approval system using machine learning is done based on key performance metrics such as accuracy, precision, recall, F1 score, and AUC-ROC to understand the model's effectiveness. The model shows good overall performance with an accuracy of 88.3% but still has room for improvement. The precision of 76.62% and recall of 88.2% indicate that while the model effectively identifies approved loans, it has a moderate rate of false positives.

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022

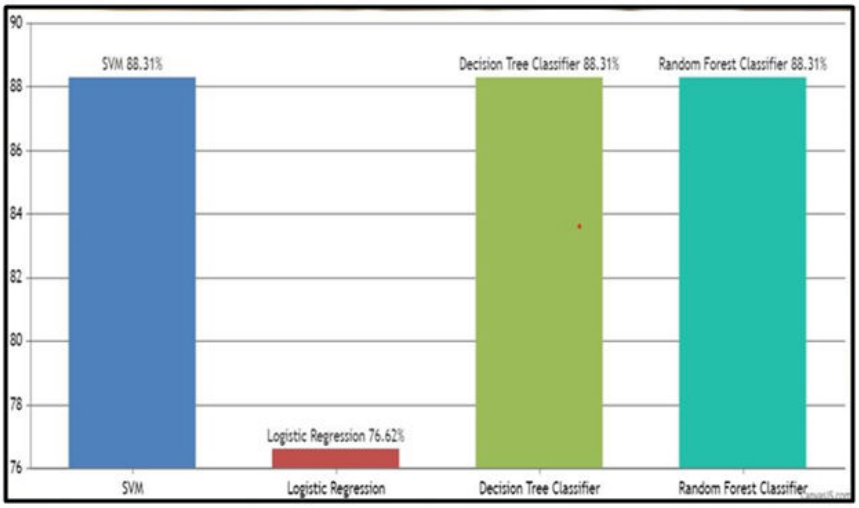

Fig: 2 Result Analyses

The F1 score of 88% indicates a balanced consideration of precision and recall. A high AUC-ROC value would confirm the model's ability to distinguish between approved and rejected loans. Further considerations include addressing class imbalances, analyzing feature importance, comparing different models, and using cross-validation to ensure robust performance metrics. Overall, this analysis highlights the model's strengths and areas for refinement to ensure reliable loan approval predictions.

S.NO	MACHINE MODELS	ACCURACY
1	SVM	88.31%
2	Logistic Regression	76.62%
3	Decision Tree Classifier	88.21%
4	Random Forest Classifier	88.33%

Table 1: Analysis Table

6. Conclusion

The modernized loan approval system is one of the significant leaps towards financial technology. The advantages of machine learning bring about astonishing accuracy and reliability in the assessment process of loan approvals. The rigorous evaluation of performance metrics, such as accuracy, precision, recall, F1 score, and AUC-ROC, can be used to validate both the effectiveness of the model and provide a comprehensive understanding of its strengths and weaknesses. The analysis shows the system is excellent at identifying the approved loans, which essentially is the key to minimizing risks and ensuring that financial resources are allocated effectively. It does, however, open up areas that need more fine-tuning, specifically on reducing false positives—the cases where loans are falsely predicted to be approved and class imbalances, wherein the distribution of approved against rejected applications skews the model's performance. All these point to the importance of careful model selection and feature engineering, as well as constant evaluation and iteration to further improve predictive accuracy.

References

1. Kumar, Rajiv, et al. (2019),” Prediction of loan. approval using machinelearning”, International Journal of Advanced Sorence andTechnology, 28(7), 455-460.
2. Supriya, Pidikiti, et al. (2019). Loan prediction. using machine. learning models. Faternational Journal of Engineering and Techniques, 5(2), 144-147
3. Arun. Kumar, Garg Ishan & Kaur Samseet (2016) Loan approval prediction based on machine learning approach. IOSR J Compat Eng (803), 18-21
4. Ashwitha, K., et al. (2022). An approach for prediction of loan eligibility using machine learning, International Conference on Artificial Intelligence and Data Engineering (AIDF)
5. Kumari, Ashwini, et al. (2018), Multilevel home security system using urduino&gsan Journal for Research 4.
6. Patibandla, RSM Lakalumi & Naralasetti Veeranjaneyulu. (2018). Survey on clustering algorithms for unstructured data. Intellig Engineering Informatics Proceedings of the 6th International Conference OR FICTA. Springer Singapore
7. Tejaswini, J, et al. (2020), Accurate loan approval prediction based on machine learning approach. Journal of Engineering Science 11(4), 523-532
8. Santlusri, K. & P. R. S. M. Lakshmi(2015).Comparative study varkus security algorithms in cloud computing. Recent Trends in Programming Langнадет. 211), 1-6.
9. Sri, K. Santhi & PR. S. M. Lakshmi, (2017), DDoS attacks, detection parameters and mitigation in cloud environment. National Conference the Recent Advances in Computer Science & Engineering (NCRACSE-2017), Gantur, India.
10. Viswanatha, V. A C Ramachandra & R Venkata Siva Reddy. (2022). Bidirectronal DC-DC converter circuits and smart control algorithms: a review
11. Sri. K. Santhi, P. R. Santhi, P. R. S. M. Lakshmi & MV Bhujanga Ra. (2017). A study of security and privacy ucks Cloud computing in environment
12. Dr, Ms RSM Lakshmi Patibarulla. Ande Prasail & Mr. YRP Shankar (2013). Secure zone in cloud. International Journal of Advances in Computer Networks and as Security, 3(2), 153-157
13. Viswanatha, V., et al. (2020). Intelligent line follower robot using MSP430G2ET for industrial applications. Helts The Scientific Explorer Reviewed Bimonthly International Journal, 10(02), 232-237
14. Dumala, Anveshini & S. Pallam Setty. (2020). MANET LANMAR routing protocol to support real-time communications in Ts using Soft computing technique. Data Engineering and Communication Technology.Proceeding of 3rd ICDECT -2K19, Springer Singapore
15. Anveshim, Durmala & S. Pallamsetty (2019). Investigating the impact of network size on multi-hop ad hoc on Wireless lanmar muting protocol in network I-Manager'n ger's JournalCommunication Networks, 7(4).

16. Khadhorbhi. Sk Reshmi & K. Suresh Babu.USUT (2015), Big data search space reduction-based perspective using map reduce. International Journal of Advanced Techonology and Innovative Research 7, 3642 3647.
17. Begum, Me Jakeera & M. Venkata Rao. (2015). Collaborative tagging using captcha. International Journal of Innovative Technology And Research, 3, 2436-2439.
18. Maddumala, Venkata Rao, R. Arunkumar & 5. Arivalagan. (2018). An empirical review on selection hig data clustering. Asian Journal of Computer Science and Technology, 7($1), 96-100.
19. Gowtharm, K., et al. Credit card fraud detection using logistic regression. Journal Engineering Sciences, 11.
20. AC, RV. V. K. K. SH&PS. E. (2022), In- cabin radar monitoring system: detection and localization of people inside vehicle using vital sign sensing algorithm. foternational Jmewal on Recent and Innovation Trends e Computing and Communication
21. Pavan, Gollapudi, and A. Ramesh Babu. "Enhanced Randomized Harris Hawk Optimization of PI controller for power flow control in the microgrid with the PV-wind-battery system." *Science and Technology for Energy Transition* 79 (2024): 45.
22. Tabassum, Saleha, Attuluri R. Vijay Babu, and Dharmendra Kumar Dheer. "Real-time power quality enhancement in smart grids through IoT and adaptive neuro-fuzzy systems." *Science and Technology for Energy Transition* 79 (2024): 89.
23. Viswanatha, V, Venkata Siva Reddy & R. Rajeswari (2020). Research on state space modeling, stability analysis and pid/pidn control of de de converter for digital implementation. In: Sengodan, T. Murugappan, M., Mixra, S. (eds) Advances in Electrical and Computer Technologies. Lecture Notes in Electrical Engineering, 2.672. Springer, Singapore. DOL 10.1007/978-981-15-5558-9. 106,
24. V. V, R. A. C, V. S. R. R, A. K. P, S. M. R & S.B. M. (2022). Implementation of IoT in agriculture: A scientific approach for smart irrigation. IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, pp. 1-6. 10.1109/MysuruCon55714.2022.9972734. DOI:
25. Viswanatha, V. & R. Venkata Siva Reddy. (2017). Digital control of buck converter using arduino microcontroller for low power applications. International Conference On Smart Technologies For Smart Nation (SmartTechCon). IEEE.