

Review

Not peer-reviewed version

Enhancing Loan Approval Accuracy Through Machine Learning and Behavioral Data Analysis

Amina Shahzad and [Noor Ul Amin](#) *

Posted Date: 11 November 2025

doi: 10.20944/preprints202511.0712.v1

Keywords: customer conversion; predictive modeling; marketing optimization; retail banking; loan business



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Enhancing Loan Approval Accuracy Through Machine Learning and Behavioral Data Analysis

Amina Shahzad ¹ and Noor Ul Amin ²

¹ Department of Software Engineering, University of Sialkot, Sialkot, Pakistan
² School of Computer Science, Taylor's University, Subang Jaya, Malaysia
* Correspondence: aminashahzadkhan978@gmail.com

Abstract

This research presents a machine learning (ML)-based system aimed at enhancing the process of personal loan approval in the banking sector. It addresses the limitations of traditional methods that rely heavily on manual verification and rule-based decision-making. The proposed system utilizes a dataset of 5,009 customers, incorporating demographic, financial, and behavioral features. The methodology includes data cleaning, feature selection, and the application of classification algorithms, as well as ensemble methods. Among the evaluated models, gradient boosting achieved an accuracy of 91.4%, while decision trees outperformed it with an accuracy of 97.92%, validating the system's robustness and scalability for large-scale applications.

Keywords: customer conversion; predictive modeling; marketing optimization; retail banking; loan business

1. Introduction

Advanced technology in combination with data-driven solutions has been one of the major drivers of rapid transformation in the banking industry[1]. Some of the key challenges that financial institutions are facing involve effective forecasting and personal loan approval processing, which directly impact revenue and customer satisfaction. Personal loans are one of the important financial products, catering to various customer needs such as education, medical emergencies, and home improvement services[2]. Sometimes, mistaken and delayed decisions in loan approval result in significant losses for banks and considerable dissatisfaction for customers[3].

Some of the traditional methods for determining whether to grant loans involve checking people's applications based on fixed rules. Such methods are not always capable of correctly assessing the complexity of customer details [18]. They are limited because they rely on small quantities of data, are not updated frequently, and may be subject to human error or bias.

This presents a significant opportunity for machine learning to transform how personal loans are analyzed, enabling more accurate and efficient predictions. Personal loan modeling involves analyzing various types of information, such as personal data, financial metrics, and behavioral patterns. Factors such as income level, credit score, employment status, and debt-to-income ratio play a crucial role in determining loan eligibility. However, most of these are difficult to analyze on a granular level using conventional methods.

Machine learning tools excel at addressing these challenges due to their ability to process large volumes of data and uncover deeply embedded patterns to make accurate predictions. Recent research has demonstrated the power of ML in financial analytics by presenting various algorithms for predicting loan approvals and defaults. In addition to benefiting banks, integrating multiple datasets can make predictive models more robust and scalable, ensuring applicability across diverse consumer groups [19].

The machine learning system in this work enhances how banks assess loan eligibility, using features such as customer age, income, credit card usage, account balances, and loan history offering a comprehensive insight into customer profiles. In this paper, the objective is to develop two improved models to accurately estimate whether an individual qualifies for a loan[20].

2. Literature Review

The increasing reliance on data-driven decision-making in financial institutions has accelerated research into machine learning (ML) techniques for personal loan prediction. This section reviews key contributions in this domain, focusing on algorithmic performance, model robustness, customer behavior analysis, and application frameworks.

2.1. Traditional and Machine Learning-Based Loan Prediction Models

Initial studies explored the feasibility of using classical machine learning algorithms like logistic regression, decision trees, SVM, and ensemble methods to improve loan approval accuracy. Kale et al. [1] and Mamun et al. [2] demonstrated the effectiveness of basic models such as logistic regression and random forest in identifying loan eligibility, particularly when trained on well-preprocessed datasets. Similarly, Shinde et al. [3] used a range of ML techniques, including ensemble methods, to enhance prediction accuracy and reduce risk exposure[29,30].

Supriya et al. [4] highlighted the impact of preprocessing steps like feature selection and data normalization in improving model outcomes, reporting an accuracy of 81.1% using logistic regression. Pimcharee and Surinta [5] further supported these findings, emphasizing the value of data mining techniques in uncovering hidden patterns that guide credit decisions[31–33].

2.2. Model Optimization and Federated Learning Approaches

Recent research has shifted toward optimizing prediction frameworks using advanced algorithms. Anannya et al. [6] introduced a federated learning-based model that selects eligible personal loan applicants while ensuring data privacy. Their work illustrates the evolving landscape of ML applications in finance, where distributed learning can help institutions balance performance and compliance.

2.3. Behavioral and Psychological Influences on Loan Decisions

Ismail et al. [7] examined personal loan borrowing behavior using empirical data, highlighting the influence of demographic and psychographic factors. Theoretical models such as the Theory of Reasoned Action [8] were used alongside marketing strategies like Recency, Frequency, and Monetary (RFM) segmentation to model borrower behavior and enhance personalization of financial products. Duggirala and Gandhi [8] combined behavioral indicators with ML algorithms to build personalized models, while Khedr et al. [9][10] proposed predictive approaches that integrate psychological, financial, and behavioral variables to prevent defaults.

2.1. Algorithm Comparison and Model Robustness

Multiple studies have compared algorithmic performance for loan approval. Sum et al. [11] demonstrated the superiority of ensemble techniques such as Random Forest and Gradient Boosting over traditional models. Neural networks were particularly effective in learning complex relationships from large datasets. Variable selection also plays a critical role factors like income, payment method, and age, despite low Information Value (IV), were found to significantly affect model accuracy [12].

2.4. Customer Satisfaction and Risk Management

Customer satisfaction is a crucial metric in the loan approval process. Research by Divya et al. [12] and Parasuraman and Hanif [13] emphasized service quality and demographic considerations

as strong predictors of borrowing behavior. These insights align with the broader aim of ML-enhanced systems: not only to assess financial risk but also to improve customer retention and trust.

2.5. Comparative Applications of Machine Learning in Other Domains

While the primary focus remains on financial analytics, several references offer valuable insights into the broader application of ML and deep learning. For instance, Rehman et al. [14] and Ali et al. [15] explored feature selection and algorithmic performance in healthcare diagnostics, providing transferable techniques applicable to credit scoring models. Similarly, Mir et al. [16] and Nawaz et al. [17] demonstrated effective data preprocessing and classification strategies, which can be adapted for personal loan prediction. Kok et al. [18] and Gouda et al. [19], presented robust classification architectures in cybersecurity and medical imaging that could inspire similar enhancements in financial applications. Furthermore, Lim et al. [20] used reinforcement learning in dynamic systems, a method with potential for future loan risk modeling.

2.6. Toward Web-Based and Scalable Systems

The transition to web-based loan prediction platforms has also attracted attention. Hossain et al. [21] and Shahid et al. [22] discussed the challenges and opportunities in deploying real-time ML systems, including issues like data privacy, transparency, and scalability. These considerations are essential for building secure and accessible loan approval systems

3. Proposed Methodology

The bank’s personal loan programs use statistical and machine learning techniques to assess the creditworthiness of individuals applying for personal loans. These programs evaluate a variety of data points including credit scores, income, business history, existing credit accounts, and other financial attributes that influence repayment capability[23–25]. A point-based analysis is performed on the bank’s internal loan models to enhance explanation and risk assessment, enabling banks to make informed loan approval decisions. This also supports the development of credit scoring models that improve evaluation accuracy[26–28].

3.1. Proposed Framework Proposed

The dataset used contains 5,009 entries with 14 attributes, each representing a unique customer profile, including demographic, economic, and behavioral characteristics. Key continuous variables include age, work experience, annual income, credit card spending, loan amount, and mortgage value. In addition, the dataset features categorical binary attributes such as education level, postal code, securities account, CD account, online banking usage, and credit card ownership that offer contextual insights into customer behavior and financial product ownership [22].

The target variable, Accepted Personal Loan, makes this dataset highly suitable for classification and prediction problems.

4. Results

The proposed machine learning framework significantly improves the accuracy, efficiency, and reliability of personal loan eligibility prediction. It leverages comprehensive data preprocessing, feature engineering, and robust model training strategies on a dataset of 5,009 customer profiles. These profiles encompass diverse attributes, including demographic details, financial indicators, and behavioral patterns.

The performance of several machine learning models was evaluated. The Decision Tree algorithm achieved the highest accuracy at 97.92%, followed by Gradient Boosting with 91.4%, Random Forest with 90.36%, Logistic Regression at 88.5%, Naïve Bayes at 88.08%, and KNN at 86.51%. These results are summarized in Table 1 and Figure 1.

This high-performance framework effectively captures complex patterns within customer data, such as income level, credit history, and spending behavior, which are often difficult to analyze using

traditional methods. The system’s ability to generalize across various customer profiles demonstrates its scalability and robustness.

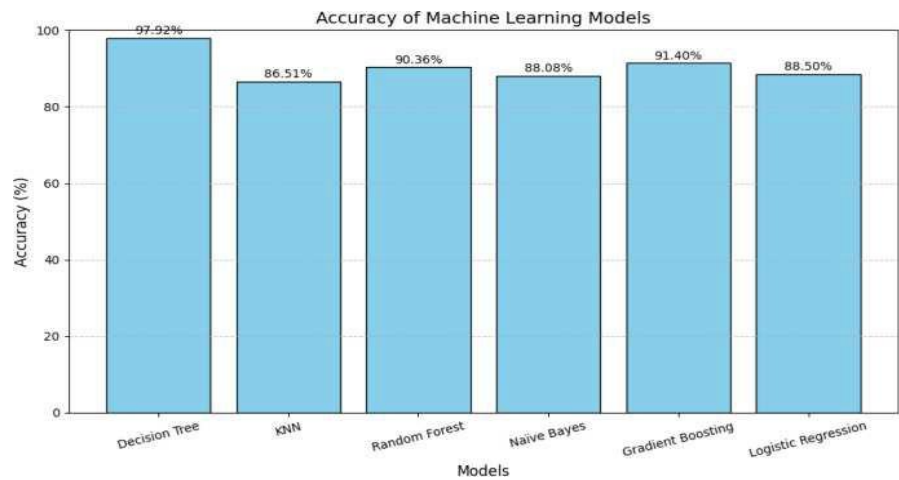


Figure 1. Comparison of ML Models.

Table 1. Model Accuracy Comparison.

Models	Accuracy
Decision Tree	97.92%
KNN	86.51%
Random Forest	90.36%
Naïve Bayes	88.08%.
Gradient Boosting	91.4%
Logistic Regression	88.05%

5. Conclusion

The proposed machine learning system for personal loan modeling represents a transformative advancement in banking decision-making. By leveraging a comprehensive dataset that captures customers’ demographic, financial, and behavioral information, the system accurately predicts loan eligibility with high precision. Key advantages of the framework include effective data preprocessing, feature selection, and algorithm fine-tuning, making it both scalable and adaptable to real-world banking applications. It enhances operational efficiency, supports risk-based pricing strategies, reduces default rates, and improves customer satisfaction.

References

1. P. S. R. Kale, S. R. Pawar, T. M. Behare, S. H. Hingwe, and S. A. Khuje, “Issue 2,” JETIR, vol. 11, no. 2, 2024. [Online]. Available: www.jetir.org
2. M. Al Mamun, A. Farjana, and M. Mamun, “Predicting Bank Loan Eligibility Using Machine Learning Models and Comparison Analysis,” unpublished.
3. Y. Shinde, I. Patil, A. Kotian, A. Shinde, and R. Gulwani, “Loan Prediction System Using Machine Learning,” ITM Web of Conferences, vol. 44, p. 03019, 2022. doi: 10.1051/itmconf/20224403019
4. P. Supriya et al., “Loan Prediction by using Machine Learning Models,” International Journal of Engineering and Techniques, vol. 5. [Online]. Available: www.ijetjournal.org
5. K. Pimcharee and O. Surinta, “Data Mining Approaches in Personal Loan Approval,” unpublished.

6. M. Anannya, M. S. Khatun, M. B. Hosen, S. Ahmed, M. F. Hossain, and M. S. Kaiser, "Eligible Personal Loan Applicant Selection using Federated Machine Learning Algorithm," *International Journal of Advanced Computer Science and Applications*, [Online]. Available: www.ijacsa.thesai.org
7. S. Ismail et al., "Determinants of Personal Loans Borrowing: An Empirical Study," *2014 IEEE Int. Conf. on Industrial Engineering and Engineering Management*, pp. 1–5, Dec. 2014. doi: 10.1109/IEEM.2014.7058739
8. Duggirala and R. Gandhi, "Bank Loan Personal Modelling using Classification Algorithms of Machine Learning," [Online]. Available: <https://www.researchgate.net/publication/355546061>
9. M. H. Khedr, N. A. Azim, and A. M. Ammar, "A New Prediction Approach for Preventing Default Customers from Applying Personal Loans Using Machine Learning," *Int. J. of Computer Science and Mobile Computing*, vol. 10, no. 12, pp. 71–82, Dec. 2021. doi: 10.47760/ijcsmc.2021.v10i12.009
10. R. M. Sum, W. Ismail, Z. H. Abdullah, N. F. M. N. Shah, and R. Hendradi, "A New Efficient Credit Scoring Model for Personal Loan Using Data Mining Technique for Sustainability Management," *Journal of Sustainability*, 2024.
11. D. Divya, D. Aggrawal, and A. Anand, "Analysing Customer Satisfaction Towards Personal Loans: Evidence from Banking Industry," *Journal of Graphic Era University*, pp. 1–22, Jun. 2024. doi: 10.13052/jgeu0975-1416.12110
12. P. Dutta, "A Study on Machine Learning Algorithm for Enhancement of Loan Prediction," [Online]. Available: www.irjmets.com
13. T. M. Ali et al., "A Sequential Machine Learning-cum-Attention Mechanism for Effective Segmentation of Brain Tumor," *Frontiers in Oncology*, vol. 12, Jun. 2022. doi: 10.3389/fonc.2022.873268
14. S. Mir et al., "A Novel Approach for the Effective Prediction of Cardiovascular Disease Using Applied AI Techniques," *ESC Heart Failure*, Jul. 2024. doi: 10.1002/ehf2.14942
15. Ahmed, Q. W., Garg, S., Rai, A., Ramachandran, M., Jhanjhi, N. Z., Masud, M., & Baz, M. (2022). Ai-based resource allocation techniques in wireless sensor internet of things networks in energy efficiency with data optimization. *Electronics*, 11(13), 2071.
16. Khan, N. A., Jhanjhi, N. Z., Brohi, S. N., Almazroi, A. A., & Almazroi, A. A. (2022). A secure communication protocol for unmanned aerial vehicles. *CMC-Computers Materials & Continua*, 70(1), 601-618.
17. Muzafar, S., & Jhanjhi, N. Z. (2020). Success stories of ICT implementation in Saudi Arabia. In *Employing Recent Technologies for Improved Digital Governance* (pp. 151-163). IGI Global Scientific Publishing.
18. Jabeen, T., Jabeen, I., Ashraf, H., Jhanjhi, N. Z., Yassine, A., & Hossain, M. S. (2023). An intelligent healthcare system using IoT in wireless sensor network. *Sensors*, 23(11), 5055.
19. Shah, I. A., Jhanjhi, N. Z., & Laraib, A. (2023). Cybersecurity and blockchain usage in contemporary business. In *Handbook of Research on Cybersecurity Issues and Challenges for Business and FinTech Applications* (pp. 49-64). IGI Global.
20. Alagic, A., Zivic, N., Kadusic, E., Hamzic, D., Hadzajlic, N., Dizdarevic, M., & Selmanovic, E. (2024). Machine learning for an enhanced credit risk analysis: A comparative study of loan approval prediction models integrating mental health data. *Machine Learning and Knowledge Extraction*, 6(1), 53-77.
21. Ndayisenga, T. (2021). Bank loan approval prediction using machine learning techniques (Doctoral dissertation).
22. Hanif, M., Ashraf, H., Jalil, Z., Jhanjhi, N. Z., Humayun, M., Saeed, S., & Almuhaideb, A. M. (2022). AI-based wormhole attack detection techniques in wireless sensor networks. *Electronics*, 11(15), 2324.
23. Shah, I. A., Jhanjhi, N. Z., Amsaad, F., & Razaque, A. (2022). The role of cutting-edge technologies in industry 4.0. In *Cyber Security Applications for Industry 4.0* (pp. 97-109). Chapman and Hall/CRC.
24. Humayun, M., Almufareh, M. F., & Jhanjhi, N. Z. (2022). Autonomous traffic system for emergency vehicles. *Electronics*, 11(4), 510.
25. Esther, D. (2023). AI-Driven Behavioral Analytics for Loan Approval Decisions.
26. Salaheldin, Y., Abdelhady, S., Abdallah, R., Fawzy, A. M., & Mohamed, M. A. (2025). A Data Driven Model for predicting Loan Approval Using Machine Learning Approaches. *ERU Research Journal*, 4(1), 2271-2289.
27. Muzammal, S. M., Murugesan, R. K., Jhanjhi, N. Z., & Jung, L. T. (2020, October). SMTrust: Proposing trust-based secure routing protocol for RPL attacks for IoT applications. In *2020 International Conference on Computational Intelligence (ICCI)* (pp. 305-310). IEEE.

28. Brohi, S. N., Jhanjhi, N. Z., Brohi, N. N., & Brohi, M. N. (2023). Key applications of state-of-the-art technologies to mitigate and eliminate COVID-19. Authorea Preprints.
29. Panda, A. R., Roy, S., Mohapatra, S., Sahoo, S., Mishra, M. K., & Gourisaria, M. K. (2024, August). Analyzing Loan Approvals Through Supervised Machine Learning Techniques. In *2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 1169-1174). IEEE.
30. Khalil, M. I., Humayun, M., Jhanjhi, N. Z., Talib, M. N., & Tabbakh, T. A. (2021). Multi-class segmentation of organ at risk from abdominal ct images: A deep learning approach. In *Intelligent Computing and Innovation on Data Science: Proceedings of ICTIDS 2021* (pp. 425-434). Singapore: Springer Nature Singapore.
31. Humayun, M., Jhanjhi, N. Z., Niazi, M., Amsaad, F., & Masood, I. (2022). Securing drug distribution systems from tampering using blockchain. *Electronics*, 11(8), 1195.
32. Dhummad, S. (2025). The Imperative of Exploratory Data Analysis in Machine Learning. *Scholars Journal of Engineering and Technology*, 13.
33. JingXuan, C., Tayyab, M., Muzammal, S. M., Jhanjhi, N. Z., Ray, S. K., & Ashfaq, F. (2024, November). Integrating AI with Robotic Process Automation (RPA): Advancing Intelligent Automation Systems. In *2024 IEEE 29th Asia Pacific Conference on Communications (APCC)* (pp. 259-265). IEEE.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.