International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0466
e ISSN: 2584-2854
Volume: 02
Issue: 10 October 2024
Page No: 3160-3165

Study on Predictive Modeling for Loan Approval

Gargi Fadtare[1], Antarisksha Dhanure[2], Gayatri Gunjal[3], Krutika Kolpate[4], Nilesh Bhelkar[5]
[1,2,3,4]UG Scholar, Dept. of AIDS, Rajiv Gandhi Institute of Tech., Andheri, Maharashtra, India.
[5]Assistant Professor, Dept. of AIDS, Rajiv Gandhi Institute of Tech., Andheri, Maharashtra, India.
Emails: *gargifadtare02@gmail.com[1], antarikshad16@gmail.com[2], gunjalgayatri04@gmail.com[3], krutikask22@gmail.com[4], nilesh.bhelkar@mctrgit.ac.in[5]*

Abstract
In the financial business, authorizing a loan is a critical decision-making step that often necessitates a human review of applicants' financial information and credit history. However, as massive datasets become more readily available and computing technology advances, machine learning approaches will become more powerful tools for improving the precision and effectiveness of loan approval processes. This study investigates the use of machine learning techniques combined with predictive modeling to automate and optimize loan approval processes. Multiple algorithms analyze crucial features such as job status, credit score, income, and loan amount to determine if a loan will be accepted or rejected. These approaches include logistic regression, decision trees, random forests, and neural networks. The goal of the research is to create an automated loan approval system.

Keywords: *Loan Prediction, Machine Learning, Hybrid Models, XGBoost, Random Forest Trees, Logistic Regression, Neural Networks.*

1. Introduction

The approval of bank loans was for decades based on prejudice, documents, and a stereotyped system where people made decisions without incorporating analytical models. This manual system is characterized by the following: low productivity in terms of paperwork and information processing, including decision-making that is not consistent and is not transparent. While the financial sector is undergoing the process of digitalization, the banks are experiencing growing challenges for providing the increased volume of loan applications in detailed spheres such as the student loans, the business loans, and the small finance loans. This remains the case even today, as more intricate financial environments have been accompanied by the need for faster, better-quality decisions, calls for the implementation of automated loan approval with the use of data science and machine learning. Current screening systems in the provision of bank loan approval rely on history and standard predisposing factors/models. These systems involve, among others factors, income, history of credit, loan amount, type of credit needed among others. However, the modern models are most often of a more traditional nature and are based on a set of rules, which do not always completely reflect all the individual circumstances of each case. In addition, opacity of those models poses major problems, In particular, it may be difficult for an applicant to know why his/her application was accepted or rejected. The machine learning of the recent years offers new data-oriented ways that can be adopted to improving the speed of credit approvals, as well as their reliability. It can analyze large data sets and identify those patterns that cannot be easily dismantled by human evaluators. This has promoted process rationalisation whereby major operations such as credit scoring, risk rating and loan grading among other have been automated to improve on the decision making process. Models, we will be able to show the gaps to be faced in NVI and its improvement, especially for interpretability and automation. This research will also elaborate on how an enhanced loan approval system incorporating the best ML algorithms can assist the banks in handling application efficiently and fairly centered on understanding the benefits of incorporating XAI into machine learning for increasing loan decision's transparency and fairness. We also look at the

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0466
e ISSN: 2584-2854
Volume: 02
Issue: 10 October 2024
Page No: 3160-3165

difficulties that banks encounter in addressing the influx of applications as well as how refined feature engineering including both basic and complex features, usage of hybrid models, as well as state of affairs in the bank loan approval systems, their current approaches, and drawbacks. Our work is external data integration can improve the efficiency and stability of the systems. In this research, we investigate the current loan approval rationality and the modern rationality proposed by Machine Learning

2. Literature Survey

The machine learning methods which can be applied in the prediction of the probability of approval granted by the bank for the loan as well as to identify which icons are worthy of approval include; Random Forest, DT, NB, Logistic Regression. In order to anticipate loan eligibility and give users a quick loan status check, this study [1] presents an intelligent loan helper that uses machine learning techniques and achieved 88% accuracy. In [2] for banks and consumers alike, machine learning techniques—Random Forest in particular—have the ability to anticipate loan acceptance and affordability with high accuracy, streamlining the approval process. [3] By offering a more accurate and efficient substitute for conventional techniques, machine learning algorithms can completely transform the loan approval process by lowering errors and enhancing decision-making. In [4] Banks can use machine learning algorithms to assess a customer's creditworthiness, which helps with loan request approval or rejection. The author implement SVM, DT and XGboost ML algorithms and achieved 81% accuracy. The author [5] used the Random Forest algorithm, machine learning models can increase the efficiency, speed, and accuracy of bank loan approval processes. In [6] machine learning models trained on historical data can forecast whether a new loan application will be approved or denied, thus streamlining the loan approval process in the banking industry. The author [7] examining loan payback likelihood and client credit history, Support Vector Machine (SVM) and Random Forest (RF) algorithms forecast loan acceptance for customers with high accuracy [8]. By efficiently using machine learning algorithms, enhancing accuracy and performance is still a challenge [9]. To avoid spending considerable amount of time on manually going through the applications to approve, loan banks have been developed to help in prioritizing them. It will also improve satisfaction of customers since the applications are processed within the agreed time. In general, this project is intended to enhance the efficiency of loan issuance and increase the probability of lending success for both banks and applicants; data mining techniques such as clustering, association, and classification can determine whether a loan application will be approved or rejected, helping the banks makes their decision without risk of loss.[10] Other methods of machine learning involve the ability of the random forest method, which can identify probable defaulters from customers' loan-approved history with a higher precision compared with other methods [11]. Based on the needs of the banking industry to assess loan approval risk, this study proposes an adaptive machine learning model for automatic loan approval that utilizes four categorization algorithms [12]. While as by comparing the results of the model with other machine learning algorithms then it has found that Logistic Regression has the highest accuracy of 92% and F1 Score of 96% for successfully classifying bank loan eligibility [13]. It tries to help financial institutions reduce occurrences of loan defaults by improving loan approval forecast accuracy. [14] This is because machine learning technology provides a solution to effectively forecast loan eligibility, hence minimizing on the use of labor and enhancing the efficiency in decision making in the loan approval process. [15] For this reason, machine learning can help with loan eligibility prediction[16]. To improve the accuracy of bank loan approval forecasts, the following algorithms employ a huge amount of data to predict loan acceptance for consumers. [17]. Now for candidates, the logistic regression approach reduces risk and allows banks to handling loans more effectively. The model incorporates the individual algorithm, such as NB, DT, KNN, Random Forest Tree (RFT), and ensemble learning approaches [18]. The logistic regression approach lowers risk and enables banks to process loans more quickly. The

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0466
e ISSN: 2584-2854
Volume: 02
Issue: 10 October 2024
Page No: 3160-3165

model combines individual algorithms like Naive Bayes, Decision Tree, and K-Nearest Neighbors (KNN) with ensemble learning approaches like Random Forest Trees (RFT)[19]. [20] Despite the fact that the model does not incorporate the hyper parameter tuning and is not capable of solving the real-world problems, it employs individual classifiers NB, DT and KNN along with number of models encompassing ensemble learning techniques such as Random Forest Trees.(RFT).[21] In addition to it, the model comprises of Logistic Regression, Support Vector Machine, and Naive Bayes along with a few more[22]. However, the accuracy of these approaches has been variable particularly in ill or real environments. Random Forests and Decision Trees are used by the model and it has the disadvantage of using old data. Further, the feature weights are fixed, and hence this method is less useful for cases where the variance between two datasets is large. [23] The literature review details for a few papers is presented in the Table 1.

Table 1 Literature Review Details

Author	Title (Year) & Ref.	Methods	Accuracy
R, S., L, V., B, S., & Manikandan, M	Bank Loan Approval Prediction Using Data Science Technique (ML) (2022)[1]	• Logistic Regression • Decision Trees • Random forest • Naive bayes	88%
Shinde, A.	Intelligent Loan Assistant using Machine Learning and Data Science (2022) [2]	Supervised learning methods	89%
Diwate, Y., Rana, P., & Chavan, P	Loan Approval Prediction Using Machine Learning (2023) [4]	Support Vector Machine, XGboost, Decision Tree	81%
Tumuluru, P., Burra, L., Loukya, M., Bhavana, S., CSaiBaba, H., & Sunanda, N.	Comparative Analysis of Customer Loan Approval Prediction using Machine Learning Algorithms (2022) [11]	Random Forest, Support Vector Machine, K-Nearest Neighbor	90%
Shinde, A., Patil, Y., Kotian, I., Shinde, A., & Gulwani, R.	Loan Prediction System Using Machine Learning (2022) [15]	Feature Engineering Feature Engineering	78%
Meshref, H.	Predicting Loan Approval of Bank Direct Marketing Data Using Ensemble Machine Learning Algorithms (2020) [20]	Ensemble learning(bagging), Logistic Regression &Decision Trees	86%
Viswanatha v., and Ramachandra Ac.	Prediction of Loan Approval in Banks using Machine Learning Approach (2023) [21]	Ensemble learning(RFT), Naive Bayes,Decision Tree,K-Nearest Neighbors (KNN)	82%
Ketki Kinkar, Rahul Sannat, Nitin Pise	Analysis of Loan Availability using Machine Learning Techniques (2021) [22]	Logistic Regression,Support Vector Machine,Naive Bayes	76%
Kshitiz Gautam, Arun Pratap Singh, Keshav Tyagi, Suresh Kumar	Loan Prediction using Decision Tree and Random Forest (2020) [23]	Decision Trees and random forest	84%

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0466
e ISSN: 2584-2854
Volume: 02
Issue: 10 October 2024
Page No: 3160-3165

3. Algorithms

3.1 Logistic Regression

Logistic Regression is a linear algorithm of classification that recognizes a probability of a binary outcome (e.g. the approval of a loan) depending on a single predictor or a set of predictors. In particular, as a dependent variable, the identity of the target class is estimated by the function of sigmoid transformation of the independent variables. This is an easy model, has low complexity and compatible with problems that can be separated by linear functions. However, the simple method for deriving the weights makes a naive assumption that each feature's contribution is independent of the others and that the relationship between features and the outcome is linear.

3.2 Support Vector Machine (SVM)

SVM is a strong supervised machine learning technique that can be used for a classifying and a regression technique. It seeks to determine the line that separates the classes of data points that are most far apart to achieve maximum margin of separation. SVM is capable to work with nonlinear data through the use of kernel functions and works very well on a high dimensional space. Its performance is normally better for deciding on complicated decision.

3.3 Random Forest (RF)

Random Forest is a machine learning technique in the family of decision trees and directly builds a number of trees during the training phase and at test time outputs the class which was the mode of the classes or the mean of the predictions of the individual trees. In its essence, it cultivates more accurate trees by averaging the outcome of several trees in order to minimize overfitting. In general, it demonstrates high effectiveness when it is applied to small and big data sets, although it's understandable is rather low if compared with single decision trees, and it may be slow when it is used when working with big data sets.

3.4 Naive Bayes

Naïve Bayes is a probabilistic classifier employing the decisions made by using Bayes' formula and assuming that different features are statistically independent on certain class. It is easy, fast and is perfect for small data set especially in text classification and spam detection. However, this assumption that the features are independent is not always true, which make it less effective to handle datasets that are related like loan approvals. 5. K-Nearest Neighbors (KNN): KNN is an instance base learning algorithm which classify a new data based on K nearest neighbors training data.

3.5 Decision Tree

Decision Tree is a categorized model of classification, constructed in a non-parametric method known as decision trees in which the data keeps on distributing and transforming from one decision node to the other till it finally gets to the last node which is called the leaf node. That makes it easier to understand and visualize and is thus popular to use in predictive modeling. Both numerical and categorical data can be applied in decision trees but risk of overfitting enhances the model's limited capability of meeting new data.

3.6 Extreme Gradient Boosting (XGBoost)

XGBoost is an optimized, distributed version of Tree boosting, an efficient approach to gradient boosting. In its structure, it builds sequential decision trees, positive errors from the previous models are considered. The XGBoost is faster, handles missing values and tackles issues that leads to overfitting dilemma. The algorithm works perfectly fine when the data arrangement is in tabular form and also when the data is imbalanced, which makes it suitable for techniques such as loan approval prediction where a variety of relations among features may exist.

Conclusion

Loan companies grant loans following a stringent screening and validation process. Therefore, they are uncertain about the applicant's ability to repay the loan without any problems. The loan prediction method will enable them to quickly, easily, and efficiently choose the most deserving candidates. It could provide the bank unique benefits. In this post, we have looked at the procedures for developing a loan approval prediction system. This system was built using a variety of analytical techniques, such as data collection, exploratory data analysis, data preprocessing, model building, and model testing. We carefully examined the earlier studies on this subject in our work. There has been use of the most widely used algorithms, such as Random Forest Technique, Decision Trees, and Logistic Regression.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0466
e ISSN: 2584-2854
Volume: 02
Issue: 10 October 2024
Page No: 3160-3165

In any case, the system can be identified as being an advancement in the area of loan processing technology. Owing to its high functionality and flexibility it can be named profitable for the banks as it helps to simplify decision making and increase the effectiveness of operations

Reference

[1]. R, S., L, V., B, S., & Manikandan, M. (2022). Bank Loan Approval Prediction Using Data Science Technique (ML). International Journal for Research in Applied Science and Engineering Technology. https://doi.org/10.22214/ijraset.2022.43665.

[2]. Shinde, A. (2022). Intelligent Loan Assistant using Machine Learning and Data Science. International Journal of Scientific Research in Engineering and Management. https://doi.org/10.55041/0522.12643.

[3]. Rahman, A. T., Purno, M. R. H., & Mim, S. A. (2023, July). Prediction of the Approval of Bank Loans Using Various Machine Learning Algorithms. In 2023 IEEE World Conference on Applied Intelligence and Computing (AIC) (pp. 272-277). IEEE.

[4]. Diwate, Y., Rana, P., & Chavan, P. (2021). Loan approval prediction using machine learning. International Research Journal of Engineering and Technology (IRJET), 8(05), 1741-1745.

[5]. Aphale, A., & Shinde, S. (2020). Predict Loan Approval in Banking System Machine Learning Approach for Cooperative Banks Loan Approval. International journal of engineering research and technology, 9.

[6]. Orji, U. E., Ugwuishiwu, C. H., Nguemaleu, J. C., & Ugwuanyi, P. N. (2022, April). Machine learning models for predicting bank loan eligibility. In 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON) (pp. 1-5). IEEE.

[7]. Singh, V., Yadav, A., Awasthi, R., & Partheeban, G. (2021). Prediction of Modernized Loan Approval System Based on Machine Learning Approach. 2021 International Conference on Intelligent Technologies (CONIT), 1-4. https://doi.org/10.1109/CONIT51480.2021.9498475

[8]. Kokru, J., Ghodke, A., Chavan, P., Chand, S., & Mane, P. (2022). Bank Loan Approval Prediction System Using Machine Learning Algorithms. International Journal of Advanced Research Science, Communication and Technology. https://doi.org/10.48175/ijarsct-2637.

[9]. Karthiban, R., Ambika, M., & Kannammal, K. E. (2019, January). A review on machine learning classification technique for bank loan approval. In 2019 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-6). IEEE.

[10]. N, P., D., Shetty, C., N, R., B, D., &, P. (2022). Predictive Analysis of Loan Data using Machine Learning. 2022 International Conference on Artificial Intelligence and Data Engineering (AIDE), 272-276.https://doi.org/10.1109/AIDE57180.2022.10060781.

[11]. Tumuluru, P., Burra, L. R., Loukya, M., Bhavana, S., CSaiBaba, H. M. H., & Sunanda, N. (2022, February). Comparative Analysis of Customer Loan Approval Prediction using Machine Learning Algorithms. In 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS) (pp. 349-353). IEEE.

[12]. Sharma, V., & Sharma, R. (2022). A Systematic Survey of Automatic Loan Approval System Based on Machine Learning. International Journal of Security and Privacy in Pervasive Computing (IJSPPC), 14(1), 1-25.

[13]. Mamun, M., Farjana, A., & Mamun, M. (2022). Predicting Bank Loan Eligibility Using Machine Learning Models and Comparison Analysis. Proceedings of the International Conference on Industrial Engineering and Operations Management. https://doi.org/10.46254/na07.20220328.

[14]. Saini, P., Bhatnagar, A., & Rani, L. (2023). Loan Approval Prediction using Machine

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0466
e ISSN: 2584-2854
Volume: 02
Issue: 10 October 2024
Page No: 3160-3165

Learning: A Comparative Analysis of Classification Algorithms. 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 1821-1826.

[15]. Shinde, A., Patil, Y., Kotian, I., Shinde, A., & Gulwani, R. (2022). Loan Prediction System Using Machine Learning. ITM Web of Conferences.

[16]. Arunkumar, G., Panchuram, C., Afzal, K., Yadav, N., & Goradiya, U. (2023). Predictive Analysis in Banking using Machine Learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. https://doi.org/10.32628/cseit2390247.

[17]. Bhojane, P., Londhe, A., Puri, L., Shukla, S., & Vavhal, A. (2022). Bank Loan Prediction Using Machine Learning International Journal of Advanced Research in Science, Communication and Technology. https://doi.org/10.48175/ijarsct-3857.

[18]. Rath, G., Das, D., & Acharya, B. (2020). Modern Approach for Loan Sanctioning in Banks Using Machine Learning, 179-188. https://doi.org/10.1007/978-981-15-5243-4_15.

[19]. Bhatt, K., Sharma, P., Verma, M., & Agarwal, K. (2023, April). Loan status prediction in the banking sector using machine learning. In 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN) (pp. 253-259). IEEE.

[20]. Machine, M. D. U. E. (2020). Predicting loan approval of bank direct marketing data using ensemble machine learning algorithms.

[21]. Machine, M. D. U. E. (2020). Predicting loan approval of bank direct marketing data using ensemble machine learning algorithms.

[22]. Dosalwar, S., Kinkar, K., Sannat, R., & Pise, N. (2021). Analysis of loan availability using machine learning techniques. International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), 9(1), 15-20.

[23]. Gautam, K., Singh, A. P., Tyagi, K., & Kumar, M. S. (2020). Loan Prediction using Decision Tree and Random Forest. International Research Journal of Engineering and Technology (IRJET), 7(08), 853-856.