

Loan Eligibility Prediction System

1st Chandan Singh Palhania

*University Institute of Engineering
Chandigarh University
Kharar, India*

2th Amit Kumar Jaiswal

*University Institute of Engineering
Chandigarh University
Kharar, India*

3rd Gaurav Kumar

*University Institute of Engineering
Chandigarh University
Kharar, India*

4th Utkarsh Raj

*University Institute of Engineering
Chandigarh University
Kharar, India*

5nd Shekhar Rana

*University Institute of Engineering
Chandigarh University
Kharar, India*

Abstract—In the modern financial system, banks give firms or people looking to buy anything the necessary initial investment. to assess a borrower's creditworthiness and forecast the possibility that they will be granted a loan. For lenders, banks, and financial organisations, a loan eligibility prediction system can be helpful in automating the loan application process and determining the risk of giving money to a certain applicant.

It is a piece of software that uses techniques for data analysis and machine learning. The system includes compiling data on sanctioned loans and loan applications from a variety of sources. The data contains facts on the borrower's income, job history, debt-to-income ratio, loan amount, loan period, and other relevant information. The data is then prepared for use in the machine's training by being cleaned, preprocessed, and transformed.

Then, relevant traits that can influence loan eligibility are identified from the data. This entails creating new factors or changing the ones already in use to predict loan eligibility. Following the division of the data into train and test sets, a machine learning model is selected and trained from different algorithms that are available. The testing set is used to evaluate the model's performance after it has been trained on the training set.

After the method for predicting loan eligibility is created, it can be incorporated into a programme that banks and lenders can use to determine loan eligibility. The loan eligibility decision-making process should be well explained in the application, which should also be easy to use. To make sure the model is reliable and useful over time, the loan eligibility prediction system should be constantly reviewed and updated with fresh data.

In conclusion, a loan eligibility prediction system will be a useful tool for banks, financial institutions, and lenders to automate the application process and determine the risk involved in giving money to a certain borrower. The system entails gathering, pre-processing, and manipulating data; extracting pertinent features; choosing an appropriate machine learning model; training the model; and implementing it in a lending and banking application.

I. INTRODUCTION

Financial organisations employ loan eligibility prediction systems to assess the likelihood that a person or company

will be able to repay a loan. This system assesses a borrower's creditworthiness and determines if they are eligible for a loan using a variety of data points and algorithms.

The method often makes use of previous data to spot trends and create predictive models that can precisely anticipate a borrower's capacity to repay the loan. The information used could include of a employment history, borrower's credit score, debt-to-income ratio, income, and other pertinent elements.

Using the loan eligibility prediction system, financial organisations may improve efficiency, reduce the risk of subprime loans, and automate the loan approval procedure. Additionally, it can help borrowers assess their chances of getting a loan approved and make prudent financial decisions.

The automated loan approval process, increased efficiency, and decreased risk of subprime loans are all benefits that financial institutions can gain by using the loan eligibility prediction system. Also, it can aid borrowers in understanding their possibilities of loan approval and assisting them in making wise financial decisions.

II. LITERATURE REVIEW

G.MURALI KRISHNA, V.MADHAVI[1], When one or more people, organizations, or other entities give money to other people, organizations, or other entities, it is known as a loan. The borrower assumes debt and is responsible for timely principal and interest payments until the loan is repaid. The project's objective is to make sure that every person or organisation applying for a loan is vetted before the loan is approved. Gender, education level, and the number of dependents must all be taken into consideration before the loan can be approved.

This project intends to automate the process in order to save time and energy and improve operational efficiency. The two sets of data that make up the input are the train data set and the test data set. The machine learning model is trained using the train data set, and its accuracy is evaluated. The expected loan eligibility is derived using the test data set.

Kunduru Sai Asritha, Amjan Shaik, Neelam Lahre, Velagapudi Sri Harsha, Bollu Joshua[2], Majority of a bank's profit comes from income from loans, although it carries risk. The job for any bank is the vaticination of credit defaulters; by doing this, bad loans can be avoided. This underscores the need of studying this loan eligibility validation. Machine learning is incredibly important and practical in this field. to keep the depositor's money safe. This paper research about how to look for a safer investment. It is a simple procedure done by collecting data a bank already have about repayment history, and using this data to train machine learning models to predict eligibility for loan. This study's objective study is to identify the machine learning algorithm that predicts loan eligibility with the highest accuracy.

Miraz Al Mamun, Afia Farjana and Muntasir Mamun[3], As people's needs increase, so does the demand for bank loans. Customers submit a lot of loan applications to banks every day, but not all of them are granted. Banks typically spend a lot of time reviewing these loan applications. Most banks use their risk and credit scoring algorithms to evaluate loan applications and determine whether to grant loans. Despite this, some applicants consistently fail to pay their debts, which results in substantial losses for financial institutions. In this study, patterns in a dataset of typical loan applicants are found using machine learning (ML) algorithms, which are then utilised to predict which applicants will be a secure investment. This study seeks to identify the most effective algorithm among K-Nearest Neighbor, Decision Tree, Random Forest, XGBoost, Adaboost, and Lightgbm. It can be used to create high-accuracy models.

U. E. Orji, P. N. Ugwuanyi, C. H. Ugwuishiwu, J. C. N. Nguemaleu[4], One of the time-consuming processes in the banking sector is the approval of loans. Modern technology, including machine learning models, may be used to improve the effectiveness, accuracy, and efficiency of loan approval procedures. Six machine learning techniques are presented in this study to forecast loan eligibility. (Random Forest, Gradient Boost, Decision Tree, Support Vector Machine, K-Nearest Neighbor, and Logistic Regression). On the cloud-based Jupyter Notebook platform provided by Kaggle, the dataset was processed and examined using Python programming tools. The study's findings demonstrated great performance accuracy, with the Random Forest algorithm earning the best, at 95.55 percentage, and Logistic Regression scoring the lowest, at 80 percentage.

Yash Patil, Anant Shinde, Ishan Kotian, Reshma Gulwani, Abhinav Shinde [5], The demand for bank loans normally rises everyday as a result of people's growing requirements. After verifying the applicant's eligibility, which is a challenging and time-consuming procedure, banks frequently handle the applicant's loan. Occasionally, applicants do default, costing banks money. By applying machine learning technologies that use categorization algorithms to predict approved loan applications, the machine learning approach is perfect for minimising human effort and enabling quick decision-making in the loan approval process.

M. Kalyani, C. Naveen Kumar, M. Kavitha, D. Keerthana[6], Every person when need to expand their business relies on banks for loans these institutions sit on large amount of money. In general, banks will offer loans based on if the borrower owns some collateral/ have good income against which loan safety is guaranteed. To predict a customer's loan eligibility, this study employs a range of machine learning techniques. The most modern way of loan approval is a system that uses machine learning to assess data and deliver results based on client profiles. The main goals are to clean the data, choose the most important characteristics, and evaluate the predictions of client loan eligibility made using decision trees, random forests, support vector machines, K-nearest neighbors, and decision trees using AdaBoost machine learning techniques.

C. S. Reddy, A. S. Siddiq and N. Jayapandian[7], A loan is when someone lends money to other persons, groups, or institutions. The borrower assumes a debt and is responsible for interest payments on that obligation. This paper's main objective is to make sure that anyone applying for a loan, whether they are a person, group, or institution, is thoroughly vetted before being approved. Before approving a loan for any person or company, a number of variables must be taken into account. The suggested model aims to automate the procedure, which will improve the process' accuracy while saving time and resources. The Random Forest technique is employed to forecast loan eligibility and train this model. Compared to standard prediction models, this model offers a 28% greater level of accuracy.

P. B. R, A. K, A. Kumar, B. Rao, P. S. K and S. A. P[8], Lending accounts for the majority of bank revenue. The process of obtaining a house loan and determining eligibility requires time from the non-banking finance companies, Housing Finance Companies and the applicant. If we start the entire process manually its going take a large amount time to complete. Even online applications can be challenging to complete because there are so many details to consider, such as education, dependents, and if they are self-employed or are employed in a multi national cooperation, the loan amount, term of the loan, credit history, property. All of this serves to determine loan safety and create eligibility. Machine learning might be quite useful to solve this issue and automate the process of determining a customer's legitimacy. However, long-standing issues don't care about outliers or machine learning models, which lowers accuracy and decreases the model's efficiency as a result. despite the fact that a lot of individuals are applying for loans. It's challenging to find a reliable borrower who will repay the loan. It could be challenging to select a genuine applicant if the procedure is manual. As a result, we are developing a machine learning loan prediction system that will choose the best applicants on its own. Both the applicant and the bank staff will gain from this. There will be a considerable reduction in the loan sanctioning timeframe.

G. N. Partheeban, A. Yadav, V. Singh, R. Awasthi[9], In the banking industry, the candidate receives documentation before

the loan amount is approved. The candidate's prior credit history affects whether or not the application is approved. Algorithms are used by banks to expedite the loan approval process. such as random forest classifier, logistic regression, etc. The amount of interest-bearing loans that a bank receives determines how much money it makes. Loan recovery is the banking industry's most important responsibility as the money they own is actually the depositor's money. Maintaining a lower non performing asset rate is crucial to the banks to avoid closure. So creating a machine learning model, candidates' credit histories are taken into account to perform a person loan safety test. The purpose of this project is to develop a machine learning model, train it, and then use it to forecast whether or not a fresh loan application would be granted.

Amar S. Chandgude, Ashwini S. Kadam, Gayatri V. Shelke, Shraddha R. Nikam, Ankita A. Aher[10], A bank's credit line is its main revenue stream in the financial system. Customers who pay back loans bring in profit to the bank. The bank can lower its non-performing assets by foreseeing loan defaulters. This demonstrates how crucial it is to do this study. There are many different approaches to investigate the subject of loan default, as shown by prior research from this era. However, learning about and comparing the various approaches is critical since making accurate estimates is crucial for optimising income. The issue of forecasting loan defaulters is investigated using a substantial predictive analytics technique. The Naive Bayes model for loan forecasting is the main topic of this paper.

Akash Dagar[11], The focus of this research is on using machine learning to assess loan eligibility. Many specialists have been concentrating on predicting loan approval in recent years. For creating predictions, machine learning algorithms are highly helpful. In this study, the chance that a person would be approved for a loan is predicted using four machine learning algorithms: Support Vector Machine, Logistic Regression, XGBoost, and Random Forest. We discovered that Logistic Regression offers more accuracy than other models based on the dataset used.

Ketki Kinkar, Sharayu Doswalwar, Rahul Sannat, Dr Nitin Pise[12], Banks are the financial institutions that offer a wide range of financial credit lines. The profit from the interest payment is the main motivation here. Customers' loan repayment or default decisions have an impact on a bank's profit or loss. It is crucial to evaluate and compare the various strategies since precise projections are required to optimise income. An essential predictive analytics technique for locating loan defaulters is the logistic regression model. The various performance metrics were calculated using logistic regression models in order to both measure and forecast. Performance indicators like sensitivity and specificity are used to compare the models.

G. Vishnu Vardhan Reddy, G. Soma Shekar, P. L. Srinivasa Murthy, P. Rohith[13], As the financial sector expands, more individuals are asking for loans in banks. These loans can't be accepted for any of them. Banks' primary goal is to invest their cash in borrowers who can repay their loans. There is still no

guarantee that the consumer is a secure investment, despite the fact that many banks today offer loans after careful validation and verification procedures. Therefore, it is essential to apply machine learning techniques in the banking industry to choose customers who can repay loans on time. The categorization of the data in this study is accomplished using the random forest approach. The random forests technique builds a model from a training dataset, which is then used to test data to get the desired outcome.

A. K. Goel, M. A. Sheikh and T. Kumar[14], The modern banking system provides capital to business/individual to expand their businesses. So it become important for banks to assess this risk before hand. The logistic regression models were used to generate the various performance metrics. Based on performance metrics like sensitivity and specificity, the models are contrasted. The final outcomes have shown that the model produces a range of outcomes. The model is only slightly superior because it takes into account variables (such as credit amount, credit duration, age, credit history, etc.) in addition to checking account information that must be taken into account when precisely calculating the probability of loan default. So, using a logistic regression technique, the right consumers may be selected for loan approval.

Mr. Abhiroop Sarkar[15], A loan can be considered a debt taken on by a person or a company in order to expand or purchase something. Banks are the principal lenders of loans. The individual who takes out the loan is responsible for paying interest on the principle sum. This study intends to investigate how factors including a person's credit history, marital status, and gender affect whether or not they are approved for a loan. This study will help select the best algorithm that can automate the process and make it simpler for both the banking staff and the client to obtain the eligibility analysis straight away. Manually evaluating a loan forecast takes a lot of time and money.

III. METHODOLOGY

In this section we are going to answer mainly two questions based on the research papers RQ1: - How machine learning models helping modern banking system to be more efficient using the data banks have collected over the course of time.

RQ2: - Which machine learning algorithms are being used for this purpose and why?

A. Use of machine learning in banks

Historically, approving loans was a manual task done by experienced banker, by screening the applicant's profile. And at that time number of people applying for loan were very less. But as time has passed, more people/organizations are taking for purchasing something new or expanding business. Manually checking each and every profile is not possible. Machines can do this task more efficiently without getting tired. As banks operating for quite some time has collected huge piles of data, this data can be feed into the machine learning models to predict accurate results. The following are the ways in which machine learning algorithms can be used:-

- 1) Automated Credit Scoring: Based on a borrower's credit history and financial behavior, a borrower's creditworthiness is evaluated using machine learning algorithms. Machine learning algorithms may construct precise credit scoring models and find trends by examining vast amounts of historical data, which can aid banks in making better lending choices.
- 2) Risk assessment: To evaluate the risk involved with each loan application, machine learning algorithms can be deployed. This entails examining a number of elements, including the how borrower uses credit, income, employment status, and other pertinent data. Banks can make better lending decisions and lower the chance of default by using risk assessment models that can analyse these elements.
- 3) Fraud detection: By seeing suspicious patterns or abnormalities in the data, machine learning models may be used to find fraud in loan applications. Banks can lessen the possibility of financial loss and safeguard their image by identifying and eliminating fraudulent loan applications.
- 4) Customer segmentation: Based on their behaviour and preferences, clients can be divided into groups using machine learning algorithms. Banks can customise their loan offerings and enhance the consumer experience by segmenting their customer base, which will increase client loyalty and retention.
- 5) Loan Monitoring: Machine learning models can be used to track the success of loans and spot potential pitfalls or gains. Banks can react rapidly to market developments and take well-informed decisions that can increase their profitability by examining loan data in real-time.

B. Search strategy and Study selection

To find relevant studies on this topic, we searched for papers having keywords 'regression', 'classification' and phrases like 'loan eligibility prediction using machine learning', that have main focus on creating models, feature creation and selection, analysis of machine learning algorithms based on regression or classification used for loan eligibility system as these kind of algorithms are best fitted for this job. One more criteria for selection of research paper was on which data set the study has been done. Most of the paper used here for research have used common data set available over Kaggle. Research papers not focusing on machine learning, using methods that no longer relevant are rejected.

IV. RESULT

Machine Learning algorithms can help banks grant loans more easily as it will reduce the time for loan approval plus will ensure the safety of loan. The model will it self take all the risk assessment based on the data a bank already have. It increase the opportunities for the bank as well as the borrower. Which gives answer to our research question one RQ1:- How machine learning models helping modern banking system to be more efficient using the data banks have collected over

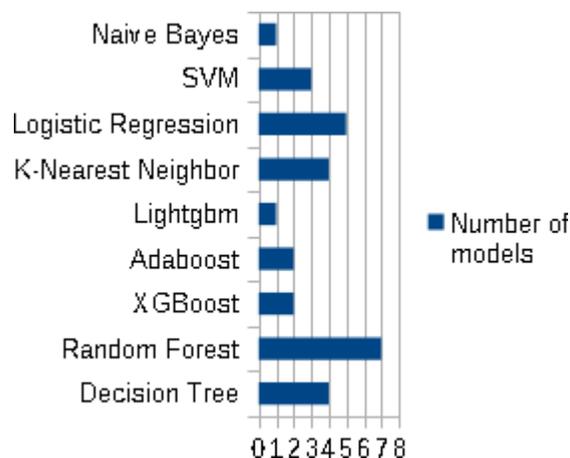


Fig. 1. Various algorithms mentioned in the research papers

the course of time. To answer research question number two RQ2:- Which machine learning algorithms are being used for this purpose and why?. We have the following graph

The majority of scholarly publications use the random forest method. Both classification and regression applications can use the Random Forest supervised machine learning method. It is a method of ensemble learning that combines different decision trees to improve prediction accuracy.

Building a lot of decision trees, each trained on a different sample of data and employing a different subset of the available attributes, is the principle behind a Random Forest. The final projection is then made using the aggregated forecasts of every single tree.

In order to prevent over fitting and improve the model's generalisation performance, each decision tree is trained on a random subset of the characteristics that are available, giving the Random Forest its random component.

The technique can manage missing values and noisy data, and it is especially helpful for processing high-dimensional data. Also, it can offer feature importance rankings, which can be used to find the data set's most crucial features.

In conclusion, Random Forest is a strong and adaptable machine learning algorithm that has a wide range of applications, including in industries like engineering, medicine, and finance.

V. CONCLUSION

The banking sector is just one of many areas that have seen major changes because to technology. Powerful machines have made it feasible to automate processes like determining loan eligibility, which has greatly enhanced the quality of service in the banking industry. Also, this automation has made it safer to offer loans, which ultimately results in cost savings for the depositors. Financial institutions may now evaluate and accept loan applications more quickly, accurately, and efficiently thanks to technology, which lowers the risk of default and possible money losses. Automation has also made it possible for operations to be more streamlined and effective, freeing up resources and time for financial institutions to

concentrate on more activities that add value for their clients. A more effective and dependable banking system has resulted from the integration of powerful machines into the financial sector, which has also substantially improved the quality of service and safety of depositors' funds.

REFERENCES

- [1] P. B. R, A. K, A. Kumar, B. Rao, P. S. K and S. A. P, "An Approach to Predict Loan Eligibility using Machine Learning", 2022 International Conference on Artificial Intelligence and Data Engineering (AIDE), Karkala, India, 2022, pp. 23-28, doi: 10.1109/AIDE57180.2022.10059881.
- [2] U. E. Orji, C. H. Ugwuishiwi, J. C. N. Nguemaleu and P. N. Ugwuanyi, "Machine Learning Models for Predicting Bank Loan Eligibility," 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON), Lagos, Nigeria, 2022, pp. 1-5, doi: 10.1109/NIGERCON54645.2022.9803172.
- [3] Anant Shinde, Yash Patil, Ishan Kotian, Abhinav Shinde and Reshma Gulwani, "Loan Prediction System Using Machine Learning", ITM Web Conf. Volume 44, 2022 International Conference on Automation, Computing and Communication 2022 (ICACC-2022)
- [4] C. S. Reddy, A. S. Siddiq and N. Jayapandian, "Machine Learning based Loan Eligibility Prediction using Random Forest Model," 2022 7th International Conference on Communication and Electronics Systems (ICES), Coimbatore, India, 2022, pp. 1073-1079, doi: 10.1109/ICES54183.2022.9835875.
- [5] C. Naveen Kumar, D. Keerthana, M. Kavitha and M. Kalyani, "Customer Loan Eligibility Prediction using Machine Learning Algorithms in Banking Sector," 2022 7th International Conference on Communication and Electronics Systems (ICES), Coimbatore, India, 2022, pp. 1007-1012, doi: 10.1109/ICES54183.2022.9835725.
- [6] Miraz Al Mamun, Afia Farjana and Muntasir Mamun, "Predicting Bank Loan Eligibility Using Machine Learning Models and Comparison Analysis", Proceedings of the 7th North American International Conference on Industrial Engineering and Operations Management, Orlando, Florida, USA, June 12-14, 2022
- [7] Amjan Shaik, Kunduru Sai Asritha, Neelam Lahre, Bollu Joshua, Velagapudi Sri Harsha, "Customer Loan Eligibility Prediction using Machine Learning". Journal of Algebraic Statistics, Vol. 13 No. 3 (2022)
- [8] G.MURALI KRISHNA, V.MADHAVI, "PREDICTION OF LOAN ELIGIBILITY OF THE CUSTOMER". Journal of Engineering Sciences Vol 13 Issue 2022
- [9] Ashwini S. Kadam, Shraddha R. Nikam, Ankita A. Aher, Gayatri V. Shelke, Amar S. Chandgude, "Prediction for Loan Approval using Machine Learning Algorithm". International Research Journal of Engineering and Technology (IRJET) Volume: 08 Issue: 04, 2021
- [10] Akash Dagar, "A Comparative Study on Loan Eligibility". International Journal of Scientific Research & Engineering Trends, Volume 7, Issue 3, 2021.
- [11] V. Singh, A. Yadav, R. Awasthi and G. N. Partheeban, "Prediction of Modernized Loan Approval System Based on Machine Learning Approach," 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 2021, pp. 1-4, doi: 10.1109/CONIT51480.2021.9498475.
- [12] Sharayu Dosalwar, Ketki Kinkar, Rahul Sannat, Dr Nitin Pise, "Analysis of Loan Availability using Machine Learning Techniques". International Journal of Advanced Research in Science, Communication and Technology (IJARST) Volume 9, Issue 1, 2021.
- [13] Mr. Abhiroop Sarkar, "MACHINE LEARNING TECHNIQUES FOR RECOGNIZING THE LOAN ELIGIBILITY". International Research Journal of Modernization in Engineering Technology and Science Volume:03/Issue:12/2021
- [14] P. L. Srinivasa Murthy, G. Soma Shekar, P. Rohith, G. Vishnu Vardhan Reddy, "Loan Approval Prediction System Using Machine Learning". Journal of Innovation in Information Technology, Vol. 4(1), 2020
- [15] M. A. Sheikh, A. K. Goel and T. Kumar, "An Approach for Prediction of Loan Approval using Machine Learning Algorithm," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 490-494, doi: 10.1109/ICESC48915.2020.9155614.
- [16] M. A. Sheikh, A. K. Goel and T. Kumar, "An Approach for Prediction of Loan Approval using Machine Learning Algorithm", 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 490-494, 2020.
- [17] A. Gupta, V. Pant, S. Kumar and P. K. Bansal, "Bank Loan Prediction System using Machine Learning", 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), pp. 423-426, 2020.
- [18] Z. Ereiz, "Predicting Default Loans Using Machine Learning (OptiML)", 2019 27th Telecommunications Forum (TELFOR), pp. 1-4, 2019.
- [19] Anthony Bellotti, Damiano Brigo, Paolo Gambetti and Frederic Vrins, "Forecasting recovery rates on non-performing loans with machine learning", International Journal of Forecasting, vol. 37, no. 1, pp. 428-444, 2021, ISSN 0169-2070.
- [20] Netty Setiawan and Diana Suharjo, "A Comparison of Prediction Methods for Credit Default on Peer to Peer Lending using Machine Learning", Procedia Computer Science, vol. 157, no. 2019, pp. 38-45, ISSN 1877-0509.
- [21] E. Chandra Blessie and R. Rekha, "Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process", International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 9, no. 1, November 2019.
- [22] Kumar Arun, Garg Ishan and Sanmeet Kaur, "Loan Approval Prediction based on Machine Learning Approach", IOSR Journal of Computer Engineering (IOSR-JCE), vol. 18, no. 3, pp. 79-81, May-Jun. 2016, ISSN 2278-0661.
- [23] Anchal Goyal and Ranpreet Kaur, "A survey on Ensemble Model for Loan Prediction", International Journal of Engineering Trends and Applications (IJETA), vol. 3, no. 1, Jan-Feb 2016.
- [24] Kumar Arun, Garg Ishan and Kaur Sanmeet, "Loan Approval Prediction based on Machine Learning Approach", IOSR Journal of Computer Engineering (IOSR-JCE), vol. 18, no. 3, pp. 79-81, May-Jun. 2016.
- [25] Shrishti Srivastava, Ayush Garg, Arpit Sehgal, Ashok kumar – Analysis and comparison of Loan Sanction Prediction Model using Python, International journal of computer science engineering and information technology research(IJCSEITR), Vol and issue 2, 2018