



Loan Eligibility Criteria using Machine Learning

Rajitha Boddepalli

UG student, GMR Institute of Technology

ABSTRACT

With the development of the banking sector, number of people applying for the bank loans are also increasing. But, the bank has its limited funds. So, the system has to decide to whom the loan has to be granted and to whom the loan has to be rejected. A bank's profit or loss mainly depend on loan recovery from the candidates. So, it is important to predict whether the client is able to repay the loan or not based on some parameters like candidate's historical data. But, it is difficult to predict it manually for each customer. So, we have machines to solve this problem. The application of a candidate can be approved or rejected by using machine learning algorithms like Logistic regression, Random forest classification, Support vector machine classifier etc.. It is possible to predict whether the candidate is eligible or not for a loan using machine learning models trained on the historical data set.

Key Words : Loan, Machine learning, Profit, Prediction

1.INTRODUCTION

A loan is a financial arrangement whereby a sum of money is lent to another party in return for the principal or value of the loan being returned at a later time. In many cases, the borrower must pay additional fees like interest or finance charges on top of the principal amount since the lender increased the principal value by adding those items.

THE LOAN PROCESS :

Here is the loan application procedure. One applies for a loan from a bank, when they need money. The borrower could be asked for particular information such as the loan's purpose, their financial background, their Social Security Number (SSN), and other things. A person's debt-to-income (DTI) ratio is taken into consideration by the lender when determining whether a loan can be repaid. The bank decides whether to accept or reject the application based on the applicant's creditworthiness. If the loan application is rejected, the bank must state why.

Components Of A Loan:

Principal: This is the original amount of money that is being borrowed from the bank.

Loan Term: The amount of time that the borrower has to repay the loan.

Interest Rate: The annual percentage rate, which is typically used to describe the pace of rise in the amount owing (APR).

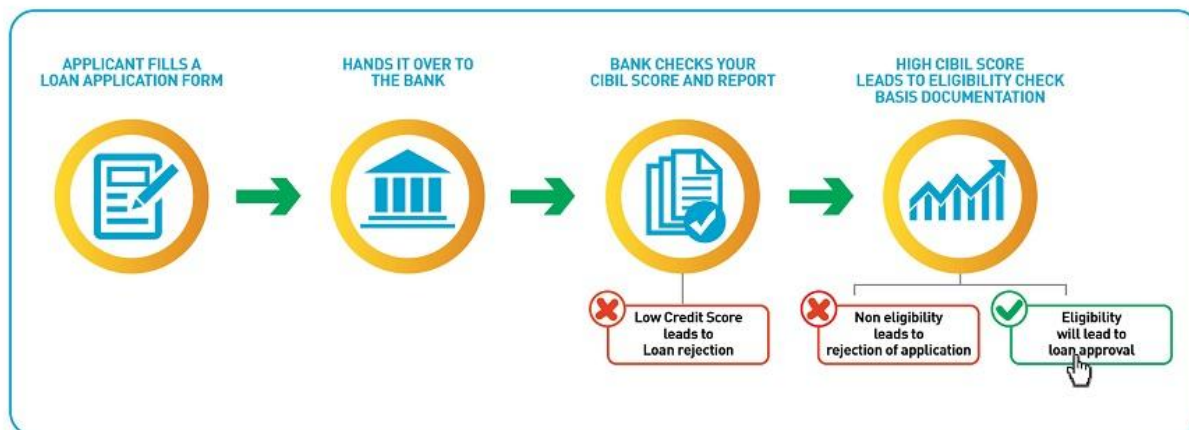
Loan Payments: The sum of money that must be paid each week or month to fulfil the loan's requirements. according to the loan's amount, period, and interest rate.

Prospective borrowers must show their capacity to repay the loan amount and their financial discipline in order to be approved for a loan. When determining whether a certain borrower is worth the risk, bank considers a number of criteria, including:

INCOME :For larger loans, banks could set a minimum income requirement to make sure borrowers won't have difficulties making payments. Several years of reliable employment may also be necessary, particularly in the case of mortgages.

CREDIT SCORE :Based on their history of borrowing and repaying debt, a person's creditworthiness is represented numerically by their credit score. Missed payments and bankruptcy can have a significant negative impact on a person's credit score.

DEPT-TO-INCOME RATIO :Lenders look at a borrower's credit history in addition to their income to determine how many open loans they have at any given moment. A high level of debt means that the borrower may have trouble repaying their debts.



2.LITERATURE SURVEY

When there is no quick cash available, people turn to loans to cover various needs like buying a home or car, going to school, or dealing with crises. When principle or interest payments are overdue by 90 days, the loan is deemed in default. For a variety of reasons, borrowers may fail on debts. It can be because of unfavourable economic conditions, such as lost income or increased costs, which hinder them from making timely EMI payments. Regardless of the cause, borrowers who are unable to make their loan or EMI payments are referred to as defaulters or loan defaulters. So, banks has to check the preconditions before granting the loan to the borrowers to reduce the number of loan defaulters. This process can be implemented using machine learning algorithms. Some of them are:

Pidikiti Supriya1 ,Myneedi Pavani and some other authors described that data mining is the process of looking at data from many angles and glean knowledge from it. The process of knowledge discovery revolves around it. Classification, clustering, association rule mining, prediction and sequential patterns, neural networks, regression, etc. are some examples of different data mining approaches.

Amruta S. Aphale Prof. Dr. Sandeep and other authors described that that labelled data is appropriate for classification algorithms. Because the dataset utilized in this study is labelled, classification analysis may be performed on it. As a result, they used a variety of categorization techniques. To estimate the creditworthiness of bank clients in terms of their capacity to pay their credit or otherwise within a specified time period, some algorithms are developed in MATLAB R and some derived from the Python scikit-learn package.

Anshika Gupta1 , Vinay Pant2 and other authors described that a prediction is a proclamation of what someone anticipates to occur in the future. A subset of advanced analytics known as predictive analytics analyses current data using a variety of methods from data mining, statistics, modelling, machine learning, and artificial intelligence to produce predictions. They work on predicting how a bank will approve a loan. They put up a model using machine learning technologies like SVM and neural networks.

P. L. Srinivasa Murthy1 , G. Soma Shekar2 , P. Rohith3 , G. Vishnu Vardhan Reddy4 . The primary goal of these authors is to categorize the types of loan applicants. The report categorizes the clients based on specific parameters. Exploratory data analysis is used for classification. Here the author used three algorithms for prediction of loan. They are 1. K Nearest Neighbor 2. Decision Tree 3. Random Forests. A large number of people will seek for loans in the banking industry. Examining a customer's eligibility through paperwork is challenging. For the n persons, the system can deliver precise results.

Mohammad Ahmad Sheikh Amit Kumar Goel and other authors described that for classification issues, the popular and effective machine learning approach known as logistic regression is used. Logistic regression has the benefit of being a predictive analysis. It is used to describe data and to clarify the relationships between a single binary variable and one or more independent nominal, ordinal, or ration level variables. Given that the result is binary and can either be 0 or 1, logistic regression takes into consideration model construction for the prediction.

3.LOAN ELIGIBILITY CRITERIA

Many prediction techniques were proposed by authors and eligibility criteria process are as follows:

3.1 Machine learning techniques used :

Neural Networks- Neural networks, also known as artificial neural networks (ANNs) or simulated neural networks (SNNs). A node layer of an artificial neural network (ANN) consists of an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, is connected to others and has a weight and threshold that go along with it. Any node whose output exceeds the defined threshold value is activated and begins providing data to the network's uppermost layer. Otherwise, no data is sent to the network's next tier.

Discriminant Analysis - In supervised machine learning, the approach of linear discriminant analysis (LDA) is frequently used in classification tasks to minimize dimensionality. It is employed for portraying class distinctions. In plain English, we may state that it is utilized to convey to lower dimensions the characteristics of a group in higher dimensions.

Naïve Bayes - The definition of naive bayes in the context of machine learning is a probabilistic model in the supervised learning genre that is

employed in a variety of use cases, mostly classification but also applicable to regression. Naive Bayes is a straightforward yet incredibly effective algorithm. This is the go-to method for any algorithm where one wants to either react to a request fast or make some calculations in order to deliver some simple yet effective insights from the data due to its lower complexity.

K-Nearest Neighbor- The k-nearest neighbor algorithm, sometimes referred to as KNN or k-NN, is a supervised learning classifier that employs proximity to produce classifications or predictions about the grouping of a single data point. Although it may be applied to classification or regression issues, it is commonly employed as a classification method since it relies on the idea that comparable points can be discovered close to one another.

Linear Regression- In regression, the variables that best match the provided data points are shown on a graph. In relation to the data, the machine learning model can provide predictions. The regression line "shows a line or curve that passes through all the data points on a target-predictor graph in such a way that the vertical distance between the data points and the regression line is smallest," in the words of the naive. It is mostly used for forecasting, time series modelling, prediction, and establishing the causal connection between variables.

Ensemble learning/method- A machine learning technique called ensemble techniques combines many base models to create a single, ideal prediction model. Let's take a step back and consider the ultimate purpose of machine learning and model development to better grasp this term. As I get into particular instances and the benefits of using Ensemble approaches, this will make more sense.

Decision Trees - One of the common and effective methods for predicting and categorizing data or an event is the usage of decision trees. Although it has a tree-like form, it is similar to a flowchart. The internal nodes in the trees stand in for a test or a question about an attribute; each branch represents a potential answer to the question; and the terminal node, also known as the leaf node, designates a class label.

3.2. Datasets used in papers:

In this study, we used data on Korean companies ranging from 2009 to 2015, provided by the DouzoneBizon ICT Group, which services enterprise resource planning (IERP) and accounting service tools. As for the composition ratio, corporations account for 61.9% and private enterprises account for 38.1%. The number of data increased from 81 in 2009 to 196,611 in 2015, which is a result of the increase in the number of customers using the DouzoneBizon ERP service.

The bank loan prediction system dataset is taken from kaggle competition which belong to different age group and gender of the applicants. There are thirteen attributes in the data set, such as education, married status, income, assets, etc.

Data has been collected from the Kaggle one of the most data source providers for the learning purpose and hence the data is collected from the Kaggle, which had two data sets one for the training and another testing. The training dataset is used to train the model in which datasets is further divided into two parts such as 80:20 or 70:30 the major datasets is used for the train the model and the minor dataset is used for the test the model and hence the accuracy of our developed model is calculated.

4. ANALYSIS :

Reference	Approach	Accuracy
1	Four machine learning Algorithms K-nearest neighbour, SVM, Decision Trees, Gradient Boosting	Decision Tree with highest accuracy(81%)
2	Seven machine learning algorithms Neural Networks, Naïve Bayes, Discriminant Analysis, K-Nearest Neighbor, Ensemble learning/method, Decision Trees, Linear Regression	Except Nearest Centroid and Gaussian Naïve Bayes ,all other achieved an accuracy learning/method, rate between 76% to 80%
3	Two machine learning algorithms Random Forest, Logistic Regression	Both show significant result with 96.7% and 96.3% accuracy
4	Three level machine learning algorithms K Nearest Neighbor ,Decision Tree ,Random Forests	Achieved highest accuracy with Random Forest(98.2%)
5	Three level machine learning algorithms SVM, KNN, Logistic Regression	Achieved highest accuracy with KNN(81.2%)

5. CONCLUSION :

In order to eliminate human interference and boost productivity, the rapidly expanding IT sector of today has to develop new technology and upgrade existing technology. Anyone looking to apply for a loan or the banking system will utilize this model. It will be very beneficial for managing banks. It is abundantly obvious from the data analysis that it lessens any fraud committed at the time of loan acceptance. Everyone values their time highly, thus by doing this, not only the bank but also the applicant's wait time will be shortened. Although it appears that it won't handle some particular circumstances when a single parameter is sufficient for the choice, it is pretty effective.

6.ABBREVIATIONS

ANNs - Artificial Neural Networks

K-NN - k-nearest neighbor

LDA - Linear Discriminant Analysis

SVM -Support Vector Machine

REFERENCES:

- 1.Supriya, P. Usha, M. Pavani, NagarapuSaisushma, Namburi Vimala Kumari and Khullar Vikas. "Loan Prediction by using Machine Learning Models." International Journal of Engineering and Techniques - Volume 5 Issue 2, Mar-Apr 2019.
- 2.M.S.Park, H. Son, C.Hyun and H. J. Hwang," Explainability of Machine Learning Models for Bankruptcy Prediction," in IEEE Access, vol. 9, pp. 124887-124899, 2021.
- 3.Amruta S. Aphale , Dr. Sandeep R. Shinde, 2020, Predict Loan Approval in Banking System Machine Learning Approach for Cooperative Banks Loan Approval, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09,Issue 08 (August 2020).
- 4.Nureni , Azeez A . and OluwadunsinEadekola "LOAN APPROVAL PREDICTION BASED ON MACHINE LEARNING APPROACH." FUDMA JOURNAL OF SCIENCES(2021)
- 5.P. L. Srinivasa Murthy¹, G. Soma Shekar², P. Rohith³, G. Vishnu Vardhan Reddy⁴ Journal of Innovation in Information Technology Vol. 4(1), Jan-Jun 2020@ ISSN: 2581-723X 21
- 6.D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," Electronics, vol. 8, no. 8, p. 832, Jul. 2019.
7. J. Kim, "Determinants of corporate bankruptcy: Evidence from chaebol and non-chaebol firms in Korea," Asian Econ. J., vol. 34, no. 3, pp. 275–300, 2020.
- 8.S. Joshi, R. Ramesh, and S. Tahsildar, "A bankruptcy prediction model using random forest," in Proc. 2nd Int. Conf. Intell. Comput. Control Syst. (ICICCS), Jun. 2018, pp. 1–6.
- 9.X. Wang and W. Cao, "Non-iterative approaches in training feedforward neural networks and their applications," Soft Comput., vol. 23, pp. 3473–3476, Apr. 2018.
- 10.Nikhil Madane, Siddharth Nanda,"Loan Prediction using Decision tree", Journal of the Gujrat Research History,Volume 21 Issue 14s, December 2019.