

# Bank Loan Prediction using Machine Learning Algorithms and Web Development

Umar Shaikh, Manoj Surywanshi, Saurabh Patil, Abhishek Lovhade, Prof. Priya Khune

Department of Computer Engineering

Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

**Abstract:** Technology has improved human existence and quality of life. With the advancement of technology, the banking sector is also seeing many improvements. The banking industry is always in need of more accurate predictive modeling systems for many problems. A bank's primary concern is to approve or deny a loan to an applicant. The number of loan approval applications is increasing daily. There are some bank policies that should be considered when selecting a loan approval applicant. Based on several parameters, the bank has to decide the best parameters for approval. Manually screening and recommending each individual for credit approval is difficult and risky. Predicting loan defaults is a difficult task for the banking industry. Many lenders suffer large losses because they do not have accurate models to predict defaults. The main purpose of this work is to predict whether it is safe to approve a loan for a particular person. The goal of these systems is to accurately predict a borrower's ability to repay a loan on time or make a credit card payment. Researchers have taken different approaches to solving this problem, and it remains an active area of research. Data mining and machine learning are emerging tools widely used by financial institutions to predict payment defaults. These tools can effectively mine large datasets that are not possible with traditional methods. Various machine learning algorithms are available, including support vector machines (SVM), random forests (RF), and decision trees. Web development allows you to display forecasts in graphs, images, histograms, etc. on your website. Bank employees and applicants can use a simple website to check if a loan should be approved by entering the appropriate data.

**Keywords:** Machine Learning, Customer Dataset, Graphical Representation, SVM, Random Forest, Web Development.

## I. INTRODUCTION

Selling loans is the core business of almost every bank. Most of the bank's wealth comes directly from the profits of loans issued by the bank. The main goal in the banking environment is to invest assets safely in a safe place. Today, many banks/financial companies approve loans after a regressive screening and verification process, but it is very difficult to decide whether the selected applicant is a legitimate applicant among all applicants or not.

In general, the banking industry's successes and failures are based on its credit risk. If the loan amount is not properly collected, the bank will incur a loss. Therefore, a bank's profit correlates with credit risk. Credit risk is a significant issue and a complex task to manage and assess. Credit evaluation tasks can be divided into two groups application rating and behavior evaluation. Application scoring divides loan applicants into good and bad risk groups. The task of behavioral evaluation is to classify existing customers based on their payment behavior and personal information.

This system can predict whether a particular applicant is safe or not and the entire process of verifying characteristics is automated using machine learning algorithms. In this model each factor is weighted differently, but in practice, sometimes loans are only approved based on a single strong factor, which is not possible with this system. Credit forecasts are very useful for both bank employees and applicants. The goal of this system is to provide a fast, immediate and easy way to select suitable applicants. It can give banks special advantages. You can set a deadline for the applicant to be considered for loan approval. The credit prediction system allows you to jump to specific applications for priority review. The system is dedicated to banking/financial management authorities and the entire forecasting process is private and cannot be altered by any stakeholder. Result for a specific credit id can be sent to various departments of the bank to take appropriate action upon request. This helps all other departments to complete other necessary procedures.

There are several functions for prediction, such as Property, Loan Status, Previous Financial History, Job Role, Salary, Area, Education.

## II. LITERATURE SURVEY

They used historical data from candidates. It was used to build machine learning models using various classification algorithms. They aimed to predict whether a new applicant would approve a loan using a machine learning model trained on historical datasets. [1]

Proposed research is on three machine learning algorithms, decision tree (DT), logistic regression (LR), and random forest (RF). It uses realworld data collected by Quds Bank, with variables representing credit constraints and covering regulatory directives. The algorithm was implemented to predict customer loan commitments and the output prediction accuracy was tested. [2]

Different models can be created using different algorithms such as Deep Support Vector Machine (DSVM), Boosted Decision Tree (BDT), Averaged Perceptron (AP), Bayes Point Machine (BPM) to better detect defaults Proposal of a system to make it possible Predict. [3]

Model can use machine learning techniques to predict who can be trusted for a loan based on previous records of people whose loan amounts have been previously qualified. The main purpose of this work is to predict whether it is safe to approve a loan for a particular person. [4]

Proposed system was able to create a credit scoring model for credit data. They developed a scoring model for financial loans using various machine learning techniques. This is a machine learning based analytical model for credit data. They used a combination of min-max normalization and a K-Nearest Neighbor (K-NN) classifier. [5]

### ALGORITHM SUSED :

#### (1) SUPPORT VECTOR MACHINE:

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both classification and regression purposes. SVMs are mainly used for classification problems. SVM is based on the idea of finding a hyperplane that best divides the dataset into two classes. The support vector is the data point closest to the hyperplane and the point in the data set whose removal changes the position of the splitting hyperplane. For this reason, they can be considered important elements of the dataset. The distance between the hyperplane and the closest data point in each set is called the margin. The goal is to choose a hyperplane with the largest possible margin between the hyperplane and any point in the training set. This increases the chances that new data will be classified correctly.

#### (2) RANDOM FOREST ALGORITHM:

Random forest is an ensemble classifier that collects the results of many decision trees by majority vote. Ensemble learning combines the results of multiple classifiers so that it can make a single decision on behalf of the community. Each decision tree in the forest is constructed by selecting different samples from the original dataset using the bootstrapping technique. The decisions made by many different individual trees are then put to a vote and the class with the most votes is presented as the committee's class. In the RF method, trees are created by his CART algorithm (classification tree and regression tree) and a combined method of boot-bagging. The dataset is divided into training data and testing data. From the training data set, tree-forming samples (inside the pocket) and non-tree-forming data (outside the pocket) are selected as a bootstrapping method (resampling and sampling). 1/3 of the training set is split into non-tree-forming data and 2/3 into tree-forming data.

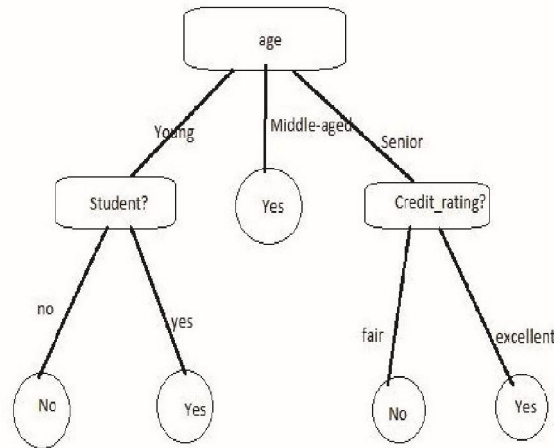
#### (3) XGBOOST:

XGBoost is an open source decision tree based software library. It implements machine learning algorithms using the gradient reinforcement framework. It nearly works on all systems such as linux, windows and macOS.

#### (4) DECISION TREES:

A decision tree is a type of supervised machine learning used to categorize or make predictions based on how a previous set of questions were answered. The model is a form of supervised learning, meaning that the model is trained

and tested on a set of data that contains the desired categorization. decision tree divides the data set into smaller pieces and predict all odds.



**Fig 1 : Decision Tree**

**(5) LOGISTIC REGRESSION:**

This is a set of classification rules used to relate observations to a discrete set of statements. Logistic regression, like other methods of regression analysis, is predictive analysis. Logistic regression is basically used to define the relationship between a dependent binary variable and a nominal or other independent variable. Today, logistic regression is used in many research fields such as medicine, machine learning, and social sciences.

**III. DATA ANALYSIS TO PREDICT LOANS**

The question arises as to what criteria should be used to determine whether to give a credit or not. We have two goals in making loans available to our customers. All formalities must be verified, including proof of income, proof of address and identification before approving loans. Everything needs to be checked because if the bank won't get his loan repayment on time then they may get in trouble.

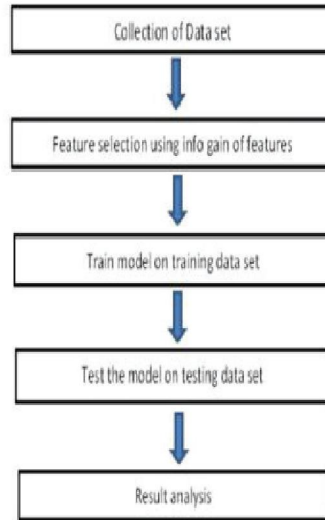
Credit is very much needed in the middle class because parents need credit not only for their children's education but also for their business. In some cases, someone can suddenly find themselves in a financial crisis, while others try to swindle their money out of the bank.

Better the customers, more the chances of getting your loan back. Background checking should be high so that you can expect a perfectly timed loan repayment. Therefore, we analyze on several bases and these are called target variables.

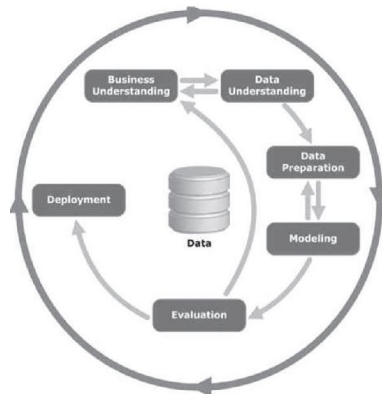
**IV. METHODOLOGY**

This model can characterize customer behavior based on records. These records are taken over and created by the customer record. With the help of these records and trained machine learning models we can predict whether a customer's loan will go through repayment. These machine learning algorithms predict whether the customer can repay the loan on time or not.

This system contains different steps for developing and using machine learning model. In this process, we collect a specific dataset with huge number of past records. We insert data for preprocessing and after these we use preprocessed data for training of model. In this training phase model extracts features, patterns and insights from preprocessed dataset. Using test data we test model to ensure that model is predicting according to expectations. After successful testing of model we can deploy model for real time use.

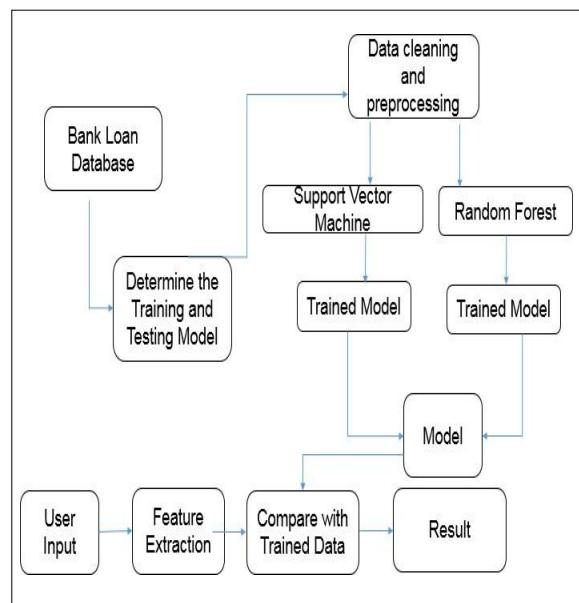


**Fig3 :Methodology**



**Fig 2 : Process Diagram**

**V. SYSTEM ARCHITECTURE**



**Fig 4. System Architecture of proposed methodology**

#### **VI. ADVANTAGES OF THE PROPOSED SYSTEM:**

In this paper, we will discuss the benefits of credit forecasting. The system predicts whether people who apply for loans will be able to repay or not. If the customer is able to repay, we predict yes to qualify for the loan and if the candidate fails, we predict no, that the client will not be eligible. The advantage of this system is simple, we need to set the conditions, algorithm will evaluate the details to learn the eligibility criteria for loan to predict whether a client is eligible or not. This system can be built to accommodate a variety of systems.

User input such as salary, address, loan amount, and loan term predicts whether a bank will approve an application or not. This research paper will assist banks to understand possibility of loss and the amount of credit can be increased or decreased.

#### **VII. RESULTS AND DESCUSSION**

The result of this system is loan prediction of individual customer. This system predict that whether the individual customer is eligible for loan or not. It uses previous loan dataset of bank to train and test model so that it can predict accurately. In this system we have used web interface for user interaction. In this user needs to enter basic details like salary, loan amount, credit history, educational details, etc. On the basis of all this details our system predict whether individual customer is eligible or not. Admin can see all predictions carried out by individual customers.

#### **VIII. CONCLUSION**

In today's rapidly growing IT industry, new technologies must be discovered and old ones should be updated to reduce human intervention and increase work efficiency. Introducing a bank loan reliability prediction system that helps organizations make the right decisions to approve or deny customer loan applications. In this model, we developed a loan default detection technique that identifies loan defaults. In the situation where the customer is in new and don't have any past records the algorithm here may not predict the accurate corresponding outcome.

This research paper can find out whether a customer is potential and whether he will repay the loan or not and the accuracy is very good. This will definitely help the banking industry develop efficient delivery routes and avoid huge financial losses. It is very useful for bank management. From the analysis of the data, it is clear that all fraud incidents can be decreased during loan approvals by using these system. Time is very valuable to everyone as it cuts down waiting times for applicants as well as banks. In this proposed system, Support Vector Machine (SVM) and Random Forest (RF) algorithms are used for predictions. Decision trees and logistic regression improve the accuracy of your predictions, giving you better, more accurate results. You can go beyond these by incorporating other technologies.

Applicants and employees can easily use this prediction system on websites built with HTML, CSS, Javascript, and other web development concepts. This website will be easy to use and easy to interact with. The user should need to have only basic computer skills. It works in real time with the provided dataset.

#### **IX. FUTURE SCOPE:**

- The credit sanction period will be shortened.
- The entire process is automated and human error is eliminated.
- Eligible applicants will receive approved loans immediately.

#### **REFERENCES**

- [1] Vishal Singh and Ayushman Yadav, "Prediction of Modernized Loan Approval System Based on Machine Learning Approach" IEEE, 2021
- [2] Mohammad J. Hamayel and Mohammad More, "Improvement of personal loans granting methods in banks using machine learning methods and approaches in Palestine", IEEE, 2021
- [3] Abhishek Shivanna and Dharma P Agarwal, "Prediction of Defaulters using Machine Learning on Azure ML", International Research Journal Of Engineering And Technology, 2021
- [4] Anshika Gupta and Vinay Pant, "Bank Loan Prediction System using Machine Learning", IEEE 2020
- [5] G. Arutjothi and C. Senthamarai, "Prediction of loan status in commercial bank using machine learning classifier", IEEE 2017