Survey of Loan Prediction System Using Machine Learning Techniques

Umar Shaikh[1], Manoj Surywanshi[2], Saurabh Patil[3], Abhishek Lovhade[4], Prof. Mrs. Priya Khune[5]

Students, Department of Computer Engineering [1,2,3,4]

Guide, Department of Computer Engineering[5]

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

Corresponding Author: umarmshaikh115@gmail.com

Abstract: *In today's world, due to advancement in technologies day by day the quality of life and ease of doing things improved. Using different technologies, banking sector is also improving rapidly. By use of latest tools, organizations can reduce their repitative task. Banking sector always requires accurate tasks. Banks has different sources of incomes but major source of banks income is loan given to customers and other small and large businesses. So it is very important for any bank or financial organization to choose good applications among several applications for loan approval. For approving loan of the particular customer banks need to follow different processes to ensure that customer is eligible for loan or not. Bank considers different parameters to decide their final decision about loan approval. They need to verify customer details one by one and then need to decide whether the loan should be given or not. But using these prediction system, one can check whether he/she is eligible for loan or not. Both customer and bank officials can use this to check eligibility of a particular application. In this system we are using different machine learning algorithms and techniques for predictions. In this system, our model is trained on past loan dataset and then tested on the test data to check accuracy. Using this machine learning model, it becomes easy to predict loan approval chances of a particular application. This model uses mainly two machine learning algorithms, support vector machine (SVM) and random forest (RF). Using machine learning algorithms and techniques, it performs different processes on dataset and finds final prediction. If there is huge number of customers who were unable to return loan before stipulated time then banks will undergo a huge financial crisis. So it is important for banks to find out that the particular customer is defaulter (Bad) or non defaulter (Good).*

Keywords: Machine Learning, Loan Dataset, Support Vector Machine, Random Forest, Decision Tree, Website, Automation, Prediction, train, test.

I. INTRODUCTION

Core business of every bank is loan distribution. Most of the banks income directly comes from interest of given loan. It is one of the most profitable business of the bank. So every bank gives different schemes to different customers to increase loan amount that eventually increases interest amount. Bank focuses on different types of customers to give loan for various purposes. Different customers take loan for various tasks like home loan, car loan, business loan, etc.

In recent years requirement of loan is increasing due to increase in income of middle class people. So in this situation it is very important for banks to select good customers from all applicants that can repay loan on time. This can increase banks profit and lead to more loan approval. In this process of loan sanctioning bank employees have to check each application and check whether the particular applicant is eligible for loan or not. This is long and hectic process.

In this survey paper, we have understood the real time problems of the bank and other financial institutions so that we can work on them to reduce that problem by using different machine learning tools and techniques. Using machine learning algorithms such as Support Vector Machine (SVM) and Random Forest (RF) we can minimize the number of problems and decrease time required for processing a particular application. We can prioritize some of the applications over other to focus on important customers.

1.1 Motivation

Today number of applications for credit from banks increasing day by day. It is becoming difficult for banks to manage all this applications manually. So here to solve this problem and to increase efficiency of banks and to ease the process of sanction of loans we are developing a machine learning model. This model helps banks and customers to check eligibility of a individual for loan. We will train this model on previous dataset of loans. This model learns from that previous history and finds out good and bad applications basically defaulters and non defaulters. Using these it will be easy for individual to check his eligibility for loan.

1.2 Objectives

- To predict whether assigning the loan to particular person will be safe or not.
- To train the dataset using various machine learning techniques.
- To test the trained model using various inputs.
- Use of machine learning algorithms like support vector machine and random forest for training of model.

II. METHODOLOGY

To conduct this survey, we conducted a thorough literature search using various academic databases such as IEEE Xplore, ACM Digital Library, and Google Scholar. We used a combination of keywords such as" credit history,"" SVM,"" loan prediction,"" forecasting,"" machine learning," and" timeseries analysis" to identify applicable exploration papers published between 2002 and 2022. We originally recaptured an aggregate of 20 papers, which we also screened to count duplicates and inapplicable papers. We further narrowed down the list to 7 papers that met our additional criteria, which were papers that concentrated on loan prediction for customers using past record dataset. Next, we conducted a detailed analysis of the named papers, including their exploration questions, datasets used, SVM models applied, and evaluation criteria used. We organized our analysis into several orders, including SVM and RF-grounded approaches for predicting loaneligibility, datasets used in loan prediction, evaluation criteria for assessing the performance of machine learning models, and crucial challenges and unborn directions of SVM and RF- grounded inloan prediction. Eventually, we summarized our findings and handed insights into the state of the art in SVM and RF- grounded approaches for predicting loan eligibility. We also discussed the strengths and weaknesses of machine learning models in this area and linked areas for unborn exploration.

2.1 Algorithms Used

- **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.
- **Random Forest:** 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.
- **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.
- **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. ADVANTAGES

- Eligible customers will receive credit in short time.
- Customers can check their eligibility without physically visiting bank.
- Applicant can understand eligibility criteria of getting loan.
- Paper work will be reduced.
- Risk of loss will be reduced.

IV. LIMITATIONS

- Sometimes due to lack of training data, model may show wrong predictions.
- It requires a huge amount of data to train the model.
- Without appropriate and sufficient data model may outperform.
- High cost of predictive model.
- Huge amount of time and resources required to train and test the model.
- Highly trained workforce required to develop model.

V. SYSTEM ARCHITECTURE
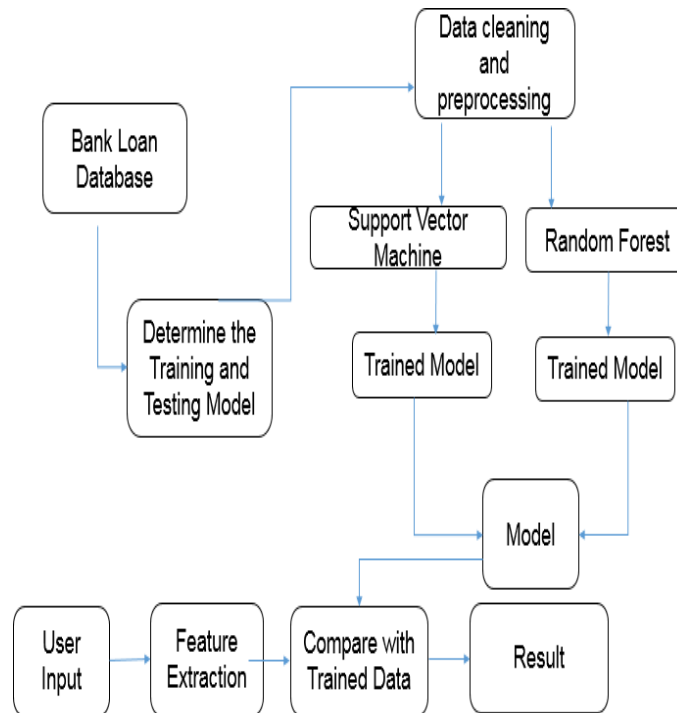
Fig : Block Diagram

VI. CONCLUSION

In this survey, we analysed that to reduce complexities and time required for loan sanctioning of a individual will be reduced by using our proposed model. This will help customers to check their eligibility. It is easy to use. Customer with basic computer knowledge can use it. Customers have to enter his or her basic details and then model will predict whether he or she can get loan or not. It also helps bank officials to see history of predictions and number of peoples willing to get loans. According to that banks may change their criteria for loan eligibility. Overall it is a very efficient and easy to use thing that save time and resources of both individuals and officials.

VII. FUTURE SCOPE

- Loan sanction time period will be reduced.
- Overall operations are automated and human interactions will be reduced.
- Eligible customers will receive credits in short time without hectic paper work.

REFERENCES

[1]. Babu Ram and A. Rama Satish, "Improved of K-Nearest Neighbor Techniques in Credit Scoring", International Journal For Development of Computer Science & Technology, vol. 1, 2013.

[2]. Aida KricheneAbdelmoula, "Bank credit risk analysis with k-nearest neighbor classifier: Case of Tunisian banks", Accounting and Management Information Systems, vol. 14.1, pp. 79, 2015.

[3]. G. Arutjothi and C. Senthamarai, "Comparison of Feature Selection Methods for Credit Risk Assessment", International Journal of Computer Science, vol. 5, no. 1, 2017.

[4]. Anshika Gupta and Vinay Pant, "Bank Loan Prediction System using Machine Learning", IEEE 2020.

[5]. J. Tejaswini, T. M. Kavya, R. D. N. Ramya, P. S. Triveni and V. R. Maddumala, "Accurate loan approval prediction based on machine learning approach", Journal of Engineering Science, vol. 11, no. 4, pp. 523-532, 2020

[6]. B. Dushimimana, Y. Wambui, T. Lubega and P. E. McSharry, "Use of machine learning techniques to create a credit score model for airtime loans", Journal of Risk and Financial Management, vol. 13, no. 8, pp. 180, 2020.

[7]. Vishal Singh and Ayushman Yadav , "Prediction of Modernized Loan Approval System Based on Machine Learning Approach" IEEE, 2021.