Loan Eligibility Prediction using Machine Learning Algorithms

Sachin Magar[1], N.S.Nikam[2], Nilesh Taksale[3], Suprem Hajare[4]

E&TC Department, SKNCOE, SPPU

Pune, India

sachinmagar7350@gmail.com[1]

namrata.nikam_skncoe@sinhgad.edu[2]

taksalenilesh1504@gmail.com[3]

supremhajare100@gmail.com[4]

Abstract— **Is Bank for banking institutions, loan acceptance is a critical step. The loan applications were approved or rejected by the system. In a bank's financial accounts, loan recovery is a major contributing aspect. It is quite difficult to predict if the customer will repay the loan. Many researchers have been working on loan approval prediction algorithms in recent years. Machine Learning (ML) techniques are extremely beneficial in the System for predicting outcomes for enormous amounts of data. Two machine learning techniques, Support Vector Machine (SVM) and Random Forest (RF) are used in this work to predict client loan approval**.

Keywords — Loan, Machine Learning, Training, Testing, Prediction

I. INTRODUCTION

Almost every bank's fundamental operation is the distribution of loans. The profit earned from the loans disbursed by the banks accounts for the majority of the bank's assets. The primary goal in the banking sector is to place their funds in safe hands. Many banks and financial institutions now grant loans after a lengthy process of verification and validation, but there is no guarantee that the chosen candidate is the most deserving of all applicants. We can forecast whether a given applicant is safe or not using this approach, and the entire feature validation process is automated using machine learning techniques. The problem of this model is that it gives varied weights to each component, however in reality, a loan might sometimes be accepted solely on the basis of a single strong factor, which is not conceivable with this approach. Loan Prediction is extremely beneficial to both bank employees and applicants. The purpose of this paper is to provide a quick, straightforward, and efficient method of selecting qualified applicants. It may provide the bank with unique benefits. The Loan Prediction System can determine the weight of each characteristic involved in loan processing automatically, and the same features are processed with respect to their associated weight on new test data. The applicant can be given a deadline to determine whether or not his or her loan will be approved. The Loan Prediction System allows you to jump to a specific application and review it on a priority basis. This paper is intended just for the management of a bank or finance firm; the entire prediction process is conducted in private, and no stakeholders will be able to influence the outcome. The results for a specific Loan Id can be sent to various bank departments for appropriate action on the application. This assists all other departments with other formalities.

In today's environment, there are numerous dangers associated with bank loans; therefore, before authorizing a loan, banks should do a risk and assessment analysis of the individual. In a market economy, banks are extremely important. The competence of the industry to assess credit risk determines whether or not a firm succeeds. Before granting credit to borrowers, the bank determines if the borrower is good or bad (defaulter) (non-defaulter). Predicting borrower status, such as whether a future borrower will default or not, is a difficult assignment for any organization or bank. Predicting loan defaulters is essentially a binary classification problem. The size of the loan is determined by the customer's credit history. The problem is to classify

borrowers as defaulter or non-defaulter. However, developing such a model is a very challenging task due to the increasing demands for loans.

II. LITERATURE SURVEY

Sr.no	Title	Author	Year	Methodology
1	Prediction of Modernized Loan Approval System Based on Ma- chine Learning Approach	Vishal Singh and Ayushman Yadav	2021	They used historical data of candidates were used to building a machine learning model using logistic regression and random forest classifier
2	Improvement of personal loans granting methods in banks using machine learning methods and approaches in Palestine	Mohammad J. Hamayel and Mohammad Moreb	2021	Proposed a study on three Machine-learning algorithms [1], Decision Tree (DT), Logistic Regression (LR), and Random Forest (RF), by using real data collected from Quds Bank with variables that cover credit restriction and regulator instructions.
3	Prediction of Defaulters using Machine Learning on Azure ML	Social Distancing and Face Mask Detection using Deep Learn- ing Models: A SurveyAb- hishek Shivanna and Dharma P Agarwal	2020	Proposed a system that used different algorithms including Deep Support Vector Machine (DSVM), Boosted Decision Tree (BDT), Aver- aged Perceptron (AP), and Bayes Point Machine (BPM) to build various models, in an attempt to better predict de- faulters.
4	Bank Loan Pre- diction System using Machine Learning	Anshika Gupta and Vinay Pant	2020	Random Forest, Logistic Regression techniques were used
5	Prediction of loan status in commercial bank using machine learning classifier	G. Arutjothi and C. Senthamarai	2017	Used the combination of Min-Max normalization and K-Nearest Neighbor (K-NN) classifier.
6	Loan Approval Prediction Based On Machine Learning	Deepak Ishwar Gouda, Ashok Kumar, Anil Manjunatha Madivala, Dilip Kumar ,	2020	The main objective was to predict whether assigning the loan to a particular person will be safe or not. This paper is divided into various four sections
7	Loan Approval Prediction using Machine Learning Algorithms Approach	Nitesh Pandey, Raman and Gupta, Sagar Uniyal, Vishal KumarVishal Singh and Ayush- man Yadav	2021	Machine Learning techniques are very crucial and useful in the prediction of these types of data. In this research paper, four algorithms of classification-based machine learning that are Logistic Regression, Decision tree, Support Vector Machine, and Random forest are applied and among them, the Support Vector Machine algorithm is the most accurate to predict loan approval with large accuracy.

| 8	Accuracy Prediction for Loan Risk Using Machine Learning Models	Anchal Goyal, Ranpreet Kaur	2016	the accuracy of several mods-els in R language and evaluate it to establish the finest model to forecast the finance status for an organization. We did the experiment five times on the same data set and found the experimental results that show the Tree Model for Genetic Algorithm is the best model for forecasting the finance for customers.
9	Loan Prediction by using Machine Learning Models	Pidikiti Supriya, Myneedi Pavani , Nagarapu Saisushma Namburi Vimala Kumari, K Vikas	2019	The proposed system was Divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of systems on most promising model (iv) Testing. In this paper we predict the loan data by using some machine learning algorithms. They are classification, logic regression, Decision Tree, and gradient boosting.
10	Loan Credibility Prediction System Based on Decision Tree Algorithm	Sivasree M S, Rekha Sunny	2015	They introduce an effective prediction model for the bankers that help them predict the credible customers who have applied for loans. Decision Tree Induction Data Mining Algorithm is applied to predict the attributes relevant for credibility. A prototype of the model is described in this paper which can be used by organizations in making the right decision to approve or reject the loan request of the customers.

III. PROPOSED SYSTEM
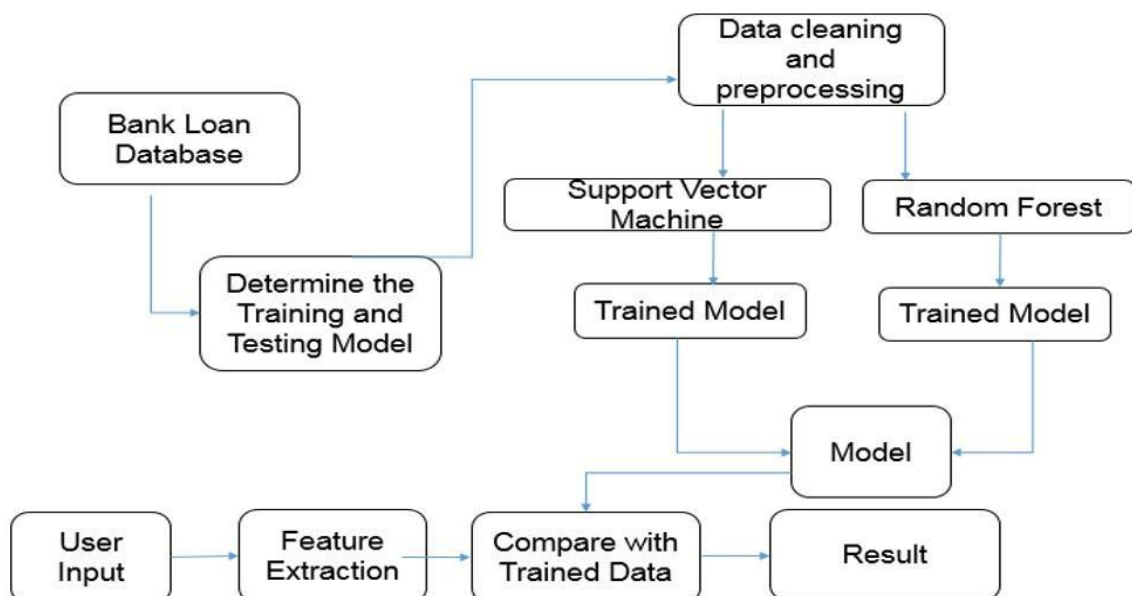
Fig.1 Block Diagram

Step 1: Gather data: This paper's dataset comes from the kaggle open-source bank dataset.

Step 2: Prepare the input data: This step was completed by the dataset's original owners. In Table, I, the dataset's composition is shown.

 Step 3: Analyze the input data to determine the relationships between various features. The basic features and the complete dataset are plotted. The dataset is then divided into two-thirds for training and one-third for testing the algorithms. Furthermore, each class in the whole dataset is represented in about the right proportion in both the training and testing datasets in order to create a representative sample.

• Step 4: Train the algorithm: Different data sets are used to train the various categorization algorithms. The Kaggle training dataset has been downloaded. Kaggle is an open-source platform with a massive dataset that can be analyzed

• Step 5: Put the algorithm to the test: Different algorithms are employed to estimate the algorithm's effectiveness on the test dataset. The accuracy, precision, recall, specificity, and F-measure are all used to evaluate the classification algorithms' performance (F1-measure). The confusion matrix entities are calculated using the Python scikit-learn programmed using input values as entities. The formula for the various evaluating metrics is shown in III, with their definitions. A 'positive' instance refers to no (signifying there will not be a default in the payment of the loan) whereas the 'negative' instance refers to yes (signifying there will be a default in the payment of the loan).

Support Vector Machine

A support vector machine (SVM) is a supervised machine learning algorithm that can be used for both classification and regression purposes. Svms are mostly used in classification problems. SVM is founded on the idea of finding a hyperplane that best divides a dataset into two classes. Support vectors are the data points nearest to the hyperplane, the points of a data set that, if deleted, would alter the position of the dividing hyper plane. Because of this, they can be considered the critical elements of a data set. The distance between the hyperplane and the nearest data point from either set is known as the margin. The aim is to choose a hyperplane with the greatest possible margin between the hyperplane and any point within the training set, giving a higher chance of new data being classified correctly.

Random Forest Algorithm

•   Step-1: Select random K data points from the training set.

   • Step-2: Build the decision trees associated with the selected data points (Subsets).

•   Step-3: Choose the number N for decision trees that you want to build.

•   Step-4: Repeat Step 1 2.

IV. RESULT

1.

Fig.2 login window

#	Gender	Married	Dependents	Education	Self employed	A-income	CA-income	Amount	Term	History	Area	Prediction
1	female	yes	yes	Graduate	yes	270000	00	100000	2	800	2	Yes
2	male	yes	yes	Graduate	yes	27000	00	100000	1	800	0	Yes
3	male	no	no	non graduate	no	00	00	30000	2	400	2	No
4	male	no	no	non graduate	no	00	0	30000	365	400	1	No
5	male	no	yes	graduate	no	10000	00	100000	365	500	0	No
6	female	no	yes	graduate	yes	30000	00	500000	3600	760	0	No
7	female	yes	yes	graduate	yes	300000	00	500000	365	760	1	Yes

Fig.3 Final Prediction

2. CONCLUSION

The proposed system for Bank loan credibility prediction may help the organizations in making the right decision to approve or reject the loan request of the customers. This can definitely help the banking industry to open up efficient de- livery channels

and the huge financial losses. In this proposed system support vector machine and random forest, algorithms are being used for the prediction. The incorporation of other techniques may outperform.

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 the commercial bank using machine learning classifier", IEEE 2017.

[6] Nitesh Pandey, Ramanand Gupta, Sagar Uniyal, Vishal Kumar, "Loan Approval Prediction using Machine Learning Algorithms Approach", International Journal Of Innovative Research in Technology, IJIRT 2021.

[7] Anchal Goyal, Manpreet Kaur, "Accuracy Prediction for Loan Risk Using Machine Learning Models", International Journal of Computer Science Trends and Technology(IJCST), 2016.

[8] Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma Namburi Vimala Kumari, K Vikas, "Loan Prediction by using Machine Learning Models", International Journal of Engineering and Techniques, 2019.

[9] Deepak Ishwar Gouda, Ashok Kumar, Anil Manjunatha Madivala, Dilip Kumar, "Loan Approval Prediction Based On Machine Learning", International Research Journal of Engineering and Technology IRJET], 2021.

[10] Sivasree M. S, Rekha Sunny, "Loan Credibility Prediction System Based on Decision Tree Algorithm", International Journal of Engineering Research and Technology, 2015.