Loan Prediction System Using SVM, DT and RF

Atharva K. Ingale[1], Laksh R. Bhamare[2], Rakhamaji G. Nagapure[3], Rutik K. Pimpale[4], Govind Pole[5]
[1,2,3,4,5]Department of Computer Engineering, MESCOE, Pune, India

Abstract— Technology has brought significant advancements to the banking industry. With loan applications flooding in every day, it has become more challenging to approve loans. Banks must adhere to strict policies when selecting a candidate for loan approval, considering several criteria to find the most suitable candidate. Manually checking each application for loan approval is arduous and risky. To overcome this, we employ machine learning to predict the loan candidate's creditworthiness based on their prior performance. Loans generate a substantial proportion of bank profits, but identifying legitimate applicants who will repay the loan is difficult. Therefore, we are developing a machine learning-based loan prediction system that will autonomously select qualified applicants, reducing loan processing time significantly. We are utilizing three machine learning techniques, Random Forest Classifier, Decision Tree and Support Vector Machine (SVM), to predict the loan data. Logistic Regression estimates the parameters of a logistic model, while SVM is a popular Supervised Learning algorithm primarily used for Classification problems in Machine Learning.

Keywords— Machine Learning, Supervised learning, Support vector machine, Decision Tree, Random Forest, Dataset, ML algorithms and ML models.

I. INTRODUCTION

Nowadays people prefer to apply for loans online because of the daily data growth brought on by the banking industry's digitalization. As a typical tool for information analysis, artificial intelligence (AI) is receiving more and more attention. People from diverse industries are using AI calculations to solve problems based on their knowledge of the relevant industries. When approving loans, banks are having a lot of trouble. The bank workers oversee a large number of applications every day, which is difficult and increases the likelihood of mistakes.

The majority of banks make money through loans; however, it may be dangerous to choose eligible clients from the pool of applicants. A bank might suffer a severe loss due to one error. We have used SVM, Decision Tree and Random Forest algorithm in the system. Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. A decision tree is a flowchart-like structure that uses if-else conditions to make decisions and predict outcomes based on input features. Random forest classifier is an ensemble learning method that combines multiple decision trees to make predictions and improve accuracy while reducing overfitting. However, primarily, it is used for Classification problems in Machine Learning.

II. MOTIVATION

For a banking firm, loan acceptance is a crucial step in the process. Applications for loans are approved or rejected using the system. Loan recovery rates are a key determining factor in a bank's financial results. The likelihood that the customer will pay back the loan is extremely difficult to forecast. We can forecast a loan applicant's acceptance using machine learning.

Some banks that have defunct due to bad loans are Allahabad Bank, PNB Bank, Bank of Always, DNFL etc. On 13 July, 2019, Allahabad Bank claimed they were defrauded by Bhushan Power & Steel (BPSL), with unpaid dues amounting to Rs 1,775 crore. Further, the Punjab National Bank (PNB) also reported fraud by BPSL amounting to Rs 3,800 crore. In 2017, the company also defaulted on nearly Rs 48,000 crore worth of loans.

III. LITERATURE SURVEY

A. Gupta , V. Pant , S. Kumar and P. K. Bansal, ”Bank Loan Prediction System using Machine Learning”, IEEE 2020. This paper states that, People prefer to apply for loans online because of the daily data growth brought on by the banking industry's digitization. As a typical tool for information analysis, artificial intelligence (AI) is receiving more and more attention. People from diverse industries are using AI calculations to solve problems based on their

knowledge of the relevant industries. When approving loans, banks are having a lot of trouble. The bank workers oversee a large number of applications every day, which is difficult and increases the likelihood of mistakes. The majority of banks make money through loans, however it can be dangerous to select eligible Almost all banks' main line of operation is loan distribution. Out of all candidates, this project seeks to provide a loan to a deserving applicant. A quick, impartial procedure that saves the bank time involves prioritising each applicant for review. The timely completion of all other client formalities by the bank authorities benefits the customers. The nicest thing is that it works well for both applicants and banks. With the help of this mechanism, specific applications might be granted priority consideration. There are some features for the prediction like- 'Gender', 'Married', 'Dependents', 'Education', 'Self_ Employed', 'ApplicantIncome', 'CoapplicantIncome', 'LoanAmount', 'Loan_Amount_Term', 'Credit_History', 'Property_Area', 'Loan_Status'.

A. S. Kadam, S. R. Nikam, A. A. Aher, G. V. Shelke and A. S. Chandgude, "Prediction for Loan Approval using Machine Learning Algorithm", IRJET 2021. This paper states that, Both applicants and bank employees benefit greatly from loan prediction. The purpose of this paper is to offer a quick, straightforward, and immediate method of selecting the qualified applicants. Dealing with all loans is Dream Housing Finance Company. All urban, semi-urban, and rural areas are served by them. After the company or bank confirms the consumer's eligibility for a loan, the customer submits their loan application. Based on the information customers supply on the application form, the company or bank wants to automate the loan eligibility procedure (in real time). Gender, marital status, education, the number of dependents, income, loan amount, credit history, and other information are included. In order to complete this project, a set of banks' prior clients' data were used. In order to obtain reliable results, the machine learning model is trained on that record. Predicting loan safety is the primary goal of this study. The SVM and Nave Bayes algorithm are used to forecast loan safety. The data is first cleaned to remove any missing values from the data collection.

H.V. Ramachandra, G. Balaraju, R. Divyashree, H. Patil, "Design and Simulation of Loan Approval Prediction Model using AWS Platform", IEEE, 2021. The paper states that, Numerous study fields are changing as a result of the effect of new computer techniques, the expansion of enormous datasets, the availability of more and more powerful and accessible computing resources, and more. Machine learning algorithms control and direct a significant portion of our life. Today, almost all fields use these techniques to create models that have a strong degree of confidence when projecting the future. The majority of such models and algorithms are cloud-based. Data is being gathered and amassed at enormous and uncontrollable speeds. Businesses track a variety of information, including credit card transactions, GPS trails, website clicks, and social media interactions. However, most businesses find it nearly impossible to digest all of this data and put it to any kind of useful use. Machine learning models were previously simply out of the price range of the majority of organisations.

S. Bhattad, S. Bawane, S. Agrawal, U. Ramteke, and Dr. P. B. Ambhore, "Loan Prediction using Machine Learning Algorithms", IJCST, 2021. This paper states that, The main business of practically all banks is the distribution of loans. The majority of a bank's assets are directly attributable to the profits made from the loans that the bank disbursed. Even if many banks and financial institutions today accept loans after a lengthy process of verification and validation, there is still no guarantee that the chosen applicant is the most deserving candidate among all applicants. Through this method, we can determine whether a specific application is secure or not, and the entire feature validation process is automated using machine learning. Both applicants and bank employees benefit greatly from loan prediction. The purpose of this paper is to offer a quick, straightforward, and immediate method of selecting the qualified applicants. It might offer the bank unique benefits. The Loan Prediction System can determine the weight of every characteristic involved in loan processing automatically, and on new test data, the same features are processed in accordance with their associated weight. It is possible to set a deadline for the applicant to determine whether or not their loan will be approved.

J. Tejaswini,T. Mohana Kavya , R. D. Naga Ramya , P. S. Triveni, V. R. Maddumala. "ACCURATE LOAN APPROVAL PREDICTION BASED ON MACHINE LEARNING APPROACH", JES, 2020.

The paper stated that, Banks all throughout the world frequently examine bank credit risk. Risk level computation uses a range of methodologies since credit risk rating is so important. Additionally, one of the primary roles of the banking industry is to manage credit risk. Distribution of the loans is the core business part of almost every banks. The main portion the bank's asset is directly coming from the profit earned from the loans distributed by the banks. The prime objective in banking environment is to invest their assets in safe hands where it is. Both applicants and bank employees benefit greatly from loan prediction. The purpose of this paper is to offer a quick, easy, and direct method of selecting the qualified applicants. It may offer the bank particular benefits. The Loan Prediction System can determine the weight of each characteristic involved in loan processing automatically, and on fresh test data, the same features are processed in accordance with their associated weights. It is possible to set a deadline for the applicant to determine whether or not their loan can be approved. Jumping to a specific application in the Loan Prediction System enables priority-based review of that application. The controlling authority of the bank or finance firm is the only audience for this paper, and the entire prediction procedure is carried out in confidence without the involvement of any other parties. Results for a certain Loan Id can be sent to other bank departments, allowing them to respond to applications in the most effective way possible. This facilitates the completion of other formalities by all other departments.

IV. METHODOLOGY

Data Collection: Collect relevant data from various sources that are needed for the project. This data can be structured, semi-structured, or unstructured.

Data Pre-processing: Pre-process the collected data to clean it, transform it, and prepare it for use in the ML model.

Feature selection: Selection of useful features from the pre-processed data that will help the ML model to learn patterns and make predictions.

Model Selection: Select the appropriate ML model that suits the problem and the data. This can include supervised, unsupervised, or reinforcement learning algorithms.

Model Training: Train the selected ML model using the pre-processed data.

Model Evaluation: Evaluate the trained ML model on a validation dataset to measure its performance.

Model Deployment: Deploy the trained and fine-tuned ML model into production for real-world use.

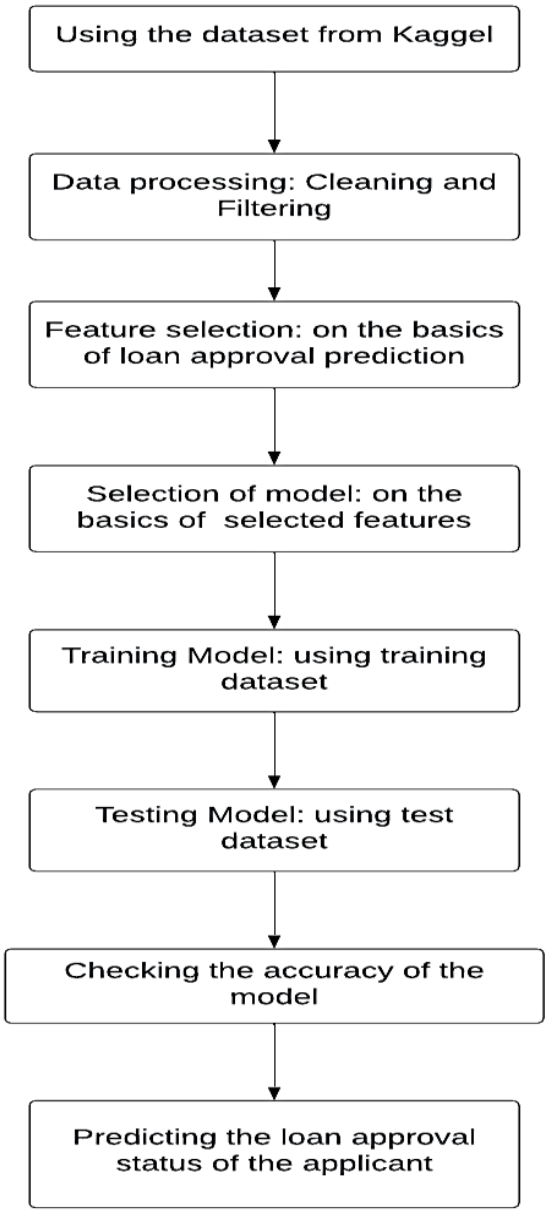


Figure 1: Methodology

V. USE CASE DAIGRAM

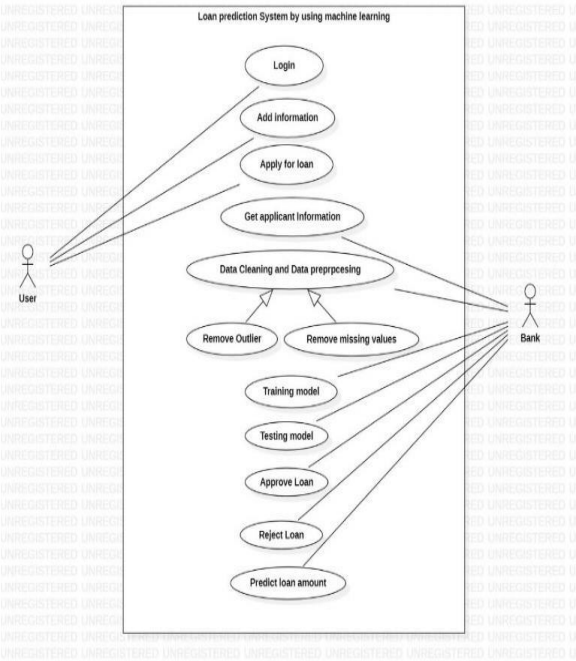
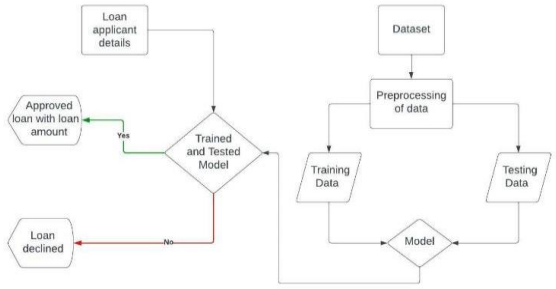


Figure 2: Use Case diagram

In above use case diagram, the use cases are login, add information, apply for loan, Get applicant information, Training model, Testing model, approve loan, reject loan, predict loan amount and data cleaning and data processing. Data cleaning and data processing is generalized from two sub-class or sub-use cases that are remove outlier and remove missing values.

VI. SYSTEM ARCHITECTURE



The system architecture consists of data pre-processing, data splitting (i.e., training and testing dataset), fine-tuned model and loan applicant details. The loan applicant details are given to the fine-tuned model and the model predicts weather the loan applicant should get the loan or not.

VII. IMPLEMENTATION

The Loan Prediction System was implemented using Python programming language and the Scikit-Learn library. The system is developed using a Spyder (Python IDE), which allows for easy experimentation and testing of different parameters.

The system's implementation involved several steps, including data pre-processing, data splitting, algorithm training, and prediction. The system's user interface is developed using the Tkinter (Python GUI), allowing users to input loan application data and receive a prediction of the likelihood of loan repayment.

Figure 3: Welcome window

This is the main pain page of the project. When clicked on “start” the login and registration page appear.

Figure 4: Registration form

The above image shows the registration form which has a fields like name, address, email, phone no., username and password. The password field is provided with the validations like length of password must be 8 characters, one special character needed etc. The registration details of the user is saved in database using SQLite.

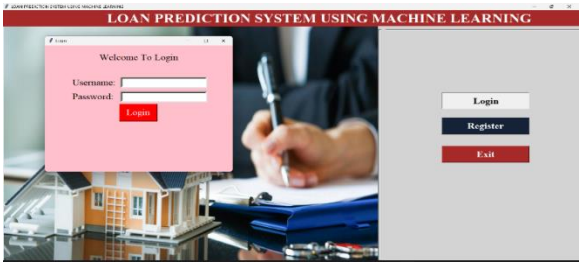


Figure 5: Login window

The above figure is of login window. The username and password is matched with the database. If matched, the user is logged in otherwise access is denied.

Figure 6: The main window

The main window consists of buttons for "Data preprocessing", "Model Train SVM", "Model Train RF", "Model Train DT" and "loan prediction". The model train buttons retrieve all the evaluation matrices of the model. The loan prediction button further opens check loan window.

A classification report is a performance evaluation metric in machine learning. It is used to show the accuracy, precision, recall, F1 Score, and support of the trained classification model.

	precision	recall	f1-score	support
0	0.69	0.33	0.45	33
1	0.79	0.94	0.86	90
accuracy			0.78	123
macro avg	0.74	0.64	0.66	123
weighted avg	0.77	0.78	0.75	123

Accracy : 78.04878048780488%
Model saved as Loan_Prediction_MODEL.joblib

Figure 7: Classification report of SVM

	precision	recall	f1-score	support
0	0.54	0.49	0.51	39
1	0.77	0.81	0.79	84
accuracy			0.71	123
macro avg	0.66	0.65	0.65	123
weighted avg	0.70	0.71	0.70	123

Accuracy : 70.73170731707317%
Model saved as DT_MODEL.joblib

Figure 8: Classification report of decision tree

	precision	recall	f1-score	support
0	0.61	0.51	0.56	39
1	0.79	0.85	0.82	84
accuracy			0.74	123
macro avg	0.70	0.68	0.69	123
weighted avg	0.73	0.74	0.73	123

Accuracy : 73.98373983739837%
Model saved as RF_MODEL.joblib

Figure 9: Classification report of random forest

VII. RESULT

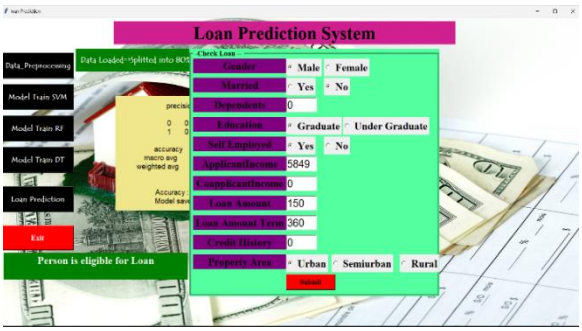


Figure 10: Check loan window with eligible message

The check loan window consists of a form which is to be filled to check if the user is eligible for loan of not. The form consists of fields like gender, married, dependents, education, income, co-applicant income, loan amount, loan term and property area.

From the above given details the model predicts weather the user will get loan or not.

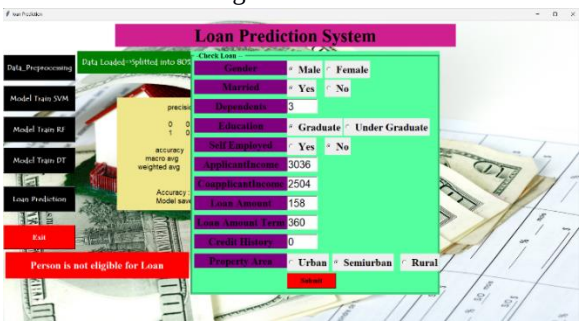


Figure 11: Check loan window with eligible message

VIII. SYSTEM ENVIRONMENT

The following are the hardware and software specifications used in this application:

A. Hardware Configuration

- Processor: Intel Core i3(Recommended) or higher
- RAM: 8GB(Recommended) or higher

- SSD Capacity: 512 GB(Recommended) or higher

B. Software Configuration

- Operating System: Windows
- IDE: Spyder
- Anaconda Navigator

C. Programming Language- Python

IX. OTHER SPECIFICATAIONS

A. Advantages:

1.By reducing risk and the amount of defaulters, the prediction model benefits both the applicant and the bank.

2.The weight of each feature (like age, income, no. of dependent) can be determined automatically by the Loan Prediction System.

3.Compared to current methodology of loan approval, loan prediction using ML is fairly simple.

4.It replaces the old time consuming process of loan.

B. Limitations:

1.This module's limitation is that it emphasises distinct weights for each aspect, however in reality loans can occasionally be authorised just based on one powerful factor, which is not possible using this approach.

2.This model limitation is that parameter estimation requires a large sample

3.The model only estimate of independent factors necessitates independent variables.

C. Applications:

This project's application are:

1. Banking institutions and organizations.
2. Investment and broking companies.
3. Non-Banking Financial Company (NBFC).
4. High Net worth Individuals (HNI).
5. Private lender.

X. CONCLUSION

We can conclude that, in order to eliminate human interference and boost productivity, the rapidly expanding IT sector of today needs to develop new technology and upgrade existing technology. Anyone looking to apply for a loan or the banking system will use this model. It will be very beneficial for managing banks. It is abundantly obvious from the data analysis that it lessens all 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 unusual situations when a single parameter is sufficient for the choice, it is highly effective and trustworthy in some situations.

XI. FUTURE SCOPE

This prediction module can be enhanced and integrated more in the future. The system is trained using prior training data, but it is feasible to alter the software in the future so that it may accept new testing data as well as training data and predict as necessary.

ACKNOWLEDGEMENT

We would like to take this opportunity to thank Prof. G. S. Pole, our internal guide, for providing us with all the support and direction we required. We are incredibly appreciative of his kind assistance. His insightful advice was incredibly beneficial.

We are very grateful to Dr. (Mrs.) N. F. Shaikh, Head, Computer Engineering Department, MES College of Engineering, for her invaluable help and advice.

We would like to extend a special thank you to the non-teaching staff for offering a variety of tools for our Project, including a lab with all necessary software platforms and an ongoing Internet connection.

REFERENCES

[1] A. Gupta , V. Pant , S. Kumar and P. K. Bansal, "Bank Loan Prediction System using Machine Learning", IEEE 2020.

[2] A. S. Kadam, S. R. Nikam, A. A. Aher, G. V. Shelke and A. S. Chandgude, "Prediction for Loan Approval using Machine Learning Algorithm", IRJET 2021.

[3] H.V. Ramachandra, G. Balaraju, R. Divyashree, H. Patil, "Design and Simulation of Loan Approval Prediction Model using AWS Platform", IEEE, 2021.

[4] S. Bhattad, S. Bawane, S. Agrawal, U. Ramteke, and Dr. P. B. Ambhore, "Loan Prediction using Machine Learning Algorithms", IJCST, 2021.

[5] J. Tejaswini,T. Mohana Kavya , R. D. Naga Ramya , P. S. Triveni, V. R. Maddumala. "Accurate Loan Approval Prediction Based On Machine Learning Approach", JES, 2020.

[6] A. Goyal and R. Kaur, "Accuracy Prediction for Loan Risk Using Machine Learning Models", IJCST, 2016.
[7] A. Khan, E. Bhadola, A. Kumar and N. Singh, "Loan Approval Prediction Model A Comparative Analysis", AAMS, 2021.
[8] Xingyun Lia,Daji Ergub,Di Zhangb,Dafeng Qiub,Ying Caib,Bo Mab, "Prediction of loan default based on multi-model fusion", ITQM 2020 & 2021.
[9] Anchal Goyal, Ranpreet Kaur, "Accuracy Prediction for Loan Risk Using Machine Learning Models", IJCST 2016.
[10] H. Faris, B. Al-Shboul and N. Ghatasheh, "A Genetic Programming Based Framework for Churn Prediction in Telecommunication Industry", Sep 2014, 253-362.
[11] G. Arutjothi, C. Senthamarai, "Prediction of loan status in commercial bank using machine learning classifier", International Conference on Intelligent Sustainable Systems (ICISS), 2017.
[12] https://machinelearningmastery.com/types-ofclassification-in-machine-learning/
[13] S.S. Keerthi and E.G. Gilbert, "Convergence of a generalize SMO algorithm for SVM classifier design. Machine Learning", Springer,46(1):351–360, 2002.
[14] J.M. Chambers. "Computational methods for data analysis. Applied Statistics", Wiley, 1(2):1–10, 1077.
[15] Mohamed El Mohadab, Belaid Bouikhalene, Said Safi, "Predicting rank for scientific research papers using supervised learning", Applied Computing and Informatics 15 (2019) 182–190.