LOAN APPROVAL PREDICTION USING MACHINE LEARNING

Mrs. R.L. Indu lekha, M.E.,

Assistant Professor, Department of CSE, Adhiyamaan College of Engineering, TamilNadu, India

Prasath G, Pugazhendhi S, Surendar U

UG Scholar, Department of CSE, Adhiyamaan College of Engineering, TamilNadu, India

Abstract: Both human existence and quality of life have benefited from technological advances. Every day we strive to create something fresh and unique. We have machines that support our lives and make them relatively complete in the banking sector. Candidates will obtain evidence / guarantee before the loan amount is approved. The system's decision on whether to approve the application is based on the candidate's historical data. Every day, many people in the banking sector seek credit, but bank reserves are limited. In this scenario, it is very useful to make correct predictions using the class function algorithm. Examples are logistic regression, random forest classifiers, support vector machine classifiers, and other algorithms. A bank's profit or loss depends on the size of the loan and whether the customer or the customer repays the loan. The most important aspect of banking is the collection of loans. The improvement process is very important in the banking industry. Using the candidate's historical data, we built a machine learning model that uses multiple classification techniques. The main goal of this task is to use a machine learning model trained with historical data to predict whether a new application will be loaned.

Keywords:- Machine learning, Data, Loan, Training, Testing, Prediction

I. INTRODUCTION

Banks have various goods to sell in our banking system, but their major source of income is their credit lines,this allows to profit the interest on loan. The profit or loss of a bank is mostly determined by loans, i.e. whether consumers repay the loan or default. The bank can lower its Non-Performing Assets by forecasting loan defaulters. As a result, research into this phenomenon is critical. Previous research in this age has revealed that there are numerous techniques for studying the subject of loan default control. However, because accurate predictions are critical for profit maximisation, It is critical to investigate the nature of the various methods and compare them. As a result, using a machine learning method, the ideal customers to target for loan granting may be simply identified by assessing their risk of defaulting on the loan. This model concludes that banks need to not only target wealthy consumers for loan approval, but also consider other characteristics of customers that play a key role in credit selection and loan sanctioning.

II. LITERATURE REVIEW

[1] A.Vaidya has proposed Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval.This article describes logistic regression and its mathematical representation. This white paper is based on logistic regression as a machine learning tool for updating predictive and probabilistic approaches to specific loan approval prediction problems. This document uses logistic regression as a tool to specifically explain whether a loan is approved for a set of applicant records.

[2] Chandra Blessie E, R Rekha proposed a study of machine learning algorithms to predict the loan sanctions process. This paper proposes a credit prediction model using the NB and support vector machine (SVM) methods. Naive Bayes, an independent speculative approach, incorporates probability theory into data classification. SVMs, on the other hand, use statistical learning models to classify predictions. To evaluate the proposed method, a dataset from the UCI repository with 21 attributes was adopted. The experiments concluded that the integration of NB and SVM, rather than the individual performance of the classifiers (NB and SVM), provides efficient credit prediction.

[3] Kumar Arun, Garg Ishan, and Kaur Sanmeet have proposed a paper on predicting loan approvals based on a machine learning approach. The main purpose of this paper is to predict whether allocating a loan to a particular person is safe. A linear discriminant analysis technique for approving or rejecting loan applications, or for short, loan forecasting issues. This paper is divided into four sections. (I) Data acquisition (ii) Comparison of machine learning models using the collected data (iii) Training Most promising model system (iv) Test.

[4] Guilder and Ozrem have proposed a category logistic principal component regression for credit approval. While the main focus of this article is to address the issue of multicollinearity between categorical explanatory predicting binary responses using variables, logistic regression is relatively easy to develop and provides the most accurate predictive analytics. Because it provides, it is most commonly used for credit prediction. One reason for this is that other algorithms are generally not good at making predictions from unnormalized data. However, the independent variable for which the prediction is made does not have to be normally distributed, so the nonlinear effect and exponentiation terms are easily handled by logistic regression.

III. PROPOSED SYSTEM

We collect the dataset required to train our model from Kaggle. We split the dataset into 80% for training and 20% for testing and evaluating our model.Next we pre-process the dataset by imputing missing values and performing outlier treatment.We pass this dataset and train our model using SVM, Logistic Regression, Random Forest.Finally we perform Stacking of above algorithms to predict loan will be approved or not.Bank with interest rates for availing loan will be suggested.
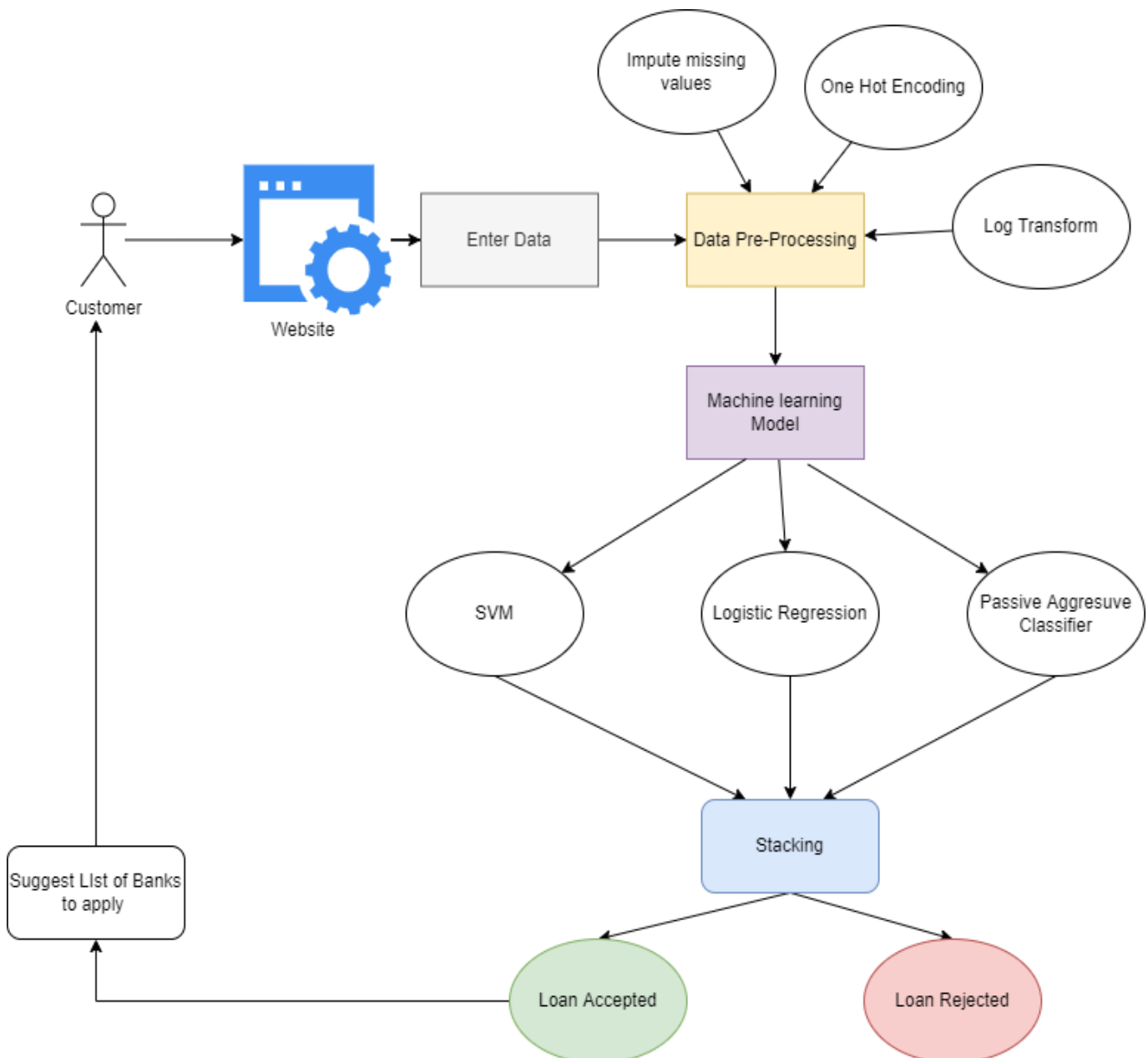

Fig 1: Architectural Diagram

IV. METHODOLOGY

Data Collection
- The dataset is collected from kaggle loan-prediction dataset.
- The dataset is split into train.csv and test.csv with 367 rows and 614 rows respectively.
- The dataset contains 12 features.

Data Preprocessing:

Converting categorical data into numerical data through one-hot encoding.Using imputation the missed values in the dataset are cleaned.Using log transform the skewed data is transformed to normal form

Model Training
- **SVM Algorithm**

  SVM is one of the supervised Learning algorithm, which is used for the Classification and Regression problems.
- **Logistic Regression**
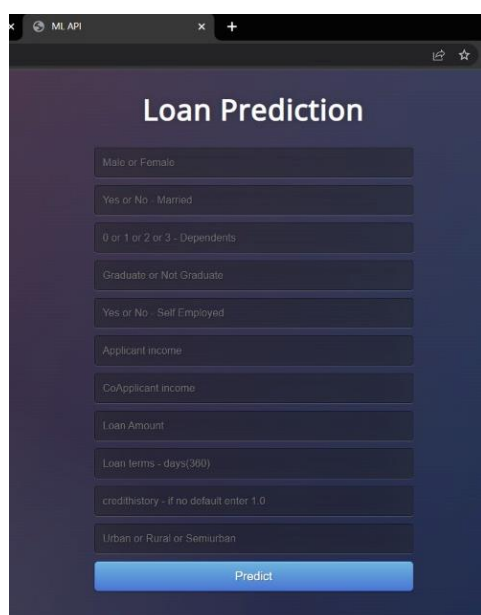
  Logistic regression are most used for loan prediction as they are easier comparatively develop and provide the most accurate predictive analysis
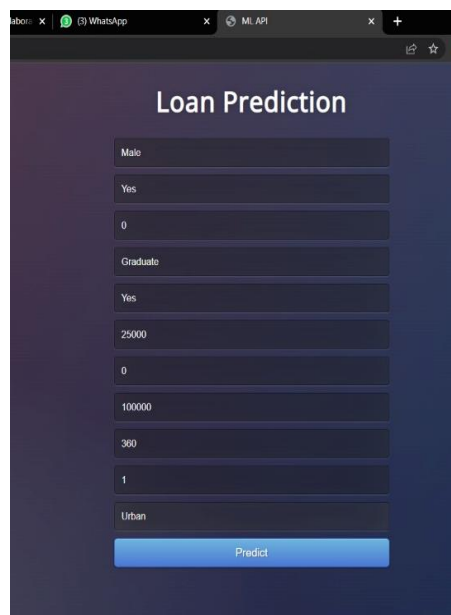- **Passive Aggressive Classifier:**

  Passive Aggressive Classifier belongs to the category of online learning algorithms in machine learning. It works by responding as passive for correct classifications and responding as aggressive for any miscalculation

V. RESULT AND DISCUSSION

Fig 2: Entry Page

Fig 3: Input Details

Fig 4: Output for Accepted

Fig 5: Input Details

Fig 6: Output for Declined

VI. CONCLUSION AND FUTURE SCOPE

It can be concluded with confidence that the Support vector machine is extremely efficient and gives a better result when compared to other models. It works correctly and fulfils all requirements of bankers. This system properly and accurately calculate the result. It predicts the loan is approve or reject to loan applicant or customer very accurately.

VII. REFERENCES

1. Abakarim, Y., Lahby, M., & Attioui, A. (2018, November). Towards an efficient real-time approach to loan credit approval using deep learning. In *2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC)* (pp. 306-313). IEEE.

2. Aphale, A. S., & Shinde, S. R. (2019). Predict Loan Approval in Banking System Machine Learning Approach for Cooperative Banks Loan Approval. *International Journal of Engineering Trends and Applications (IJETA)*, *9*(8).

3. Arun, K., Ishan, G., & Sanmeet, K. (2016). Loan approval prediction based on machine learning approach. *IOSR J. Comput. Eng*, *18*(3), 18-21.

4. Arutjothi, G., & Senthamarai, C. (2017, December). Prediction of loan status in commercial bank using machine learning classifier. In *2017 International Conference on Intelligent Sustainable Systems (ICISS)* (pp. 416-419). IEEE.

5. Bracke, P., Datta, A., Jung, C., & Sen, S. (2019). Machine learning explainability in finance: an application to default risk analysis.

6. Dhokley, W., Shaikh, U., Ansari, S., & Ansari, N. (2018). Machine learning approach to predict farmer's loan/credit repaybility using weather prediction and credit history. In *IC-CSOD-2018 Conference* (p. 283).

7. ElMasry, M. H. A. M. T. (2019). *Machine learning approach for credit score analysis: a case study of predicting mortgage loan defaults* (Doctoral dissertation).

8. Eweoya, I. O., Adebiyi, A. A., Azeta, A. A., & Azeta, A. E. (2019, August). Fraud prediction in bank loan administration using decision tree. In *Journal of Physics: Conference Series* (Vol. 1299, No. 1, p. 012037). IOP Publishing.

9. Gautam, K., Singh, A. P., Tyagi, K., & Kumar, M. S. (2020). Loan Prediction using Decision Tree and Random Forest. *International Research Journal of Engineering and Technology (IRJET)*, *7*(08).

10. Ghildiyal, B., Garg, S., & Raturi, V. (2019). Analyze of Different Algorithms of Machine Learning for Loan Approval. In *Smart Trends in Computing and Communications* (pp. 719-727). Springer, Singapore.

11. Karthiban, R., Ambika, M., & Kannammal, K. E. (2019, January). A review on machine learning classification technique for bank loan approval. In *2019 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-6). IEEE.

12. Kemalbay, G., & Korkmazoğlu, Ö. B. (2014). Categorical principal component logistic regression: a case study for housing loan approval. *Procedia-Social and Behavioral Sciences*, *109*, 730-736.

13. Nikam, S. R., Student, U. G., COE, S., Yeola, R. C., Kadam, A. S., Shelke, G. V., ... & Chandgude, A. S. (2019). Prediction for Loan Approval Using Machine Learning Algorithm.

14. Nikhil Madane, Siddharth Nanda,"Loan Prediction using Decision tree"(2019,December), Journal of the Gujrat Research History,21(7).

15. Obare, D. M., Njoroge, G. G., & Muraya, M. M. (2019). Analysis of Individual Loan Defaults Using Logit under Supervised Machine Learning Approach. *Asian Journal of Probability and Statistics*, *3*(4), 1-12.

16. Priya, K. U., Pushpa, S., Kalaivani, K., & Sartiha, A. (2018). Exploratory analysis on prediction of loan privilege for customers using random forest. *International Journal of Engineering Technology*, *7*(2.21), 339-341.

17. Semiu, A., & Gilal, A. A. R. (2019). A boosted decision tree model for predicting loan default in P2P lending communities. *Studies*, *24*, 13.

18. Sheikh, M. A., Goel, A. K., & Kumar, T. (2020, July). An approach for prediction of loan approval using machine learning algorithm. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 490-494). IEEE.

19. Soni, P. M., & Paul, V. (2019, March). A novel optimized classifier for the loan repayment capability prediction system. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 23-28). IEEE.

20. Supriya, P., Pavani, M., Saisushma, N., Kumari, N. V., & Vikas, K. (2019). Loan prediction by using machine learning models. *International Journal of Engineering and Techniques*, *5*(22), 144-148.

21. Tariq, H. I., Sohail, A., Aslam, U., & Batcha, N. K. (2019). Loan default prediction model using sample, explore, modify, model, and assess (SEMMA). *Journal of Computational and Theoretical Nanoscience*, *16*(8), 3489-3503.

22. Thomas, T. C., Sridhar, J. P., Chandrashekar, M. J., Upadhyaya, M., & Aurelia, S. Developing a website for a bank's Machine Learning-Based Loan Prediction System.

23. Vaidya, A. (2017, July). Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval. In *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.

24. Wu, M., Huang, Y., & Duan, J. (2019, July). Investigations on classification methods for loan application based on machine learning. In *2019 International Conference on Machine Learning and Cybernetics (ICMLC)* (pp. 1-6). IEEE.