



Prediction of Loan Approval Using Machine Learning

Aniket Dalavi¹, Girish Jadhav², Aniket Bande³, Onkar Karande⁴, Moiz Shaikh⁵

¹4th year, BE Computer Engineering, Student SVPMs college of engineering Malegaon, India.,

²4th year, BE Computer Engineering, Student SVPMs college of engineering Malegaon, India,

³ 4th year, BE Computer Engineering, Student SVPMs college of engineering Malegaon, India,

⁴4th year, BE Computer Engineering, Student SVPMs college of engineering Malegaon, India.

ABSTRACT -

Although there are various items that banks in our financial system can market, their primary source of income comes from their credit lines. In order for them to profit on the interest on the loans they credit. The profitability or loss of a bank is mostly determined by the loans it makes. The bank can lower its non-performing assets by foreseeing loan defaulters. This highlights how crucial it is to examine this phenomenon. There are several ways to explore the issue of preventing loan default, according to earlier studies from this age. Kaggle data is gathered for analysis and forecasting. The various performance measures have been estimated using SVM models. On the basis of performance metrics like sensitivity and specificity, the models are contrasted.

Key Words: Machine Learning, SVM, Data pre-processing, Classification, Loan Prediction..

1. INTRODUCTION

Technology has made a lot of improvements, and the banking industry is no exception. Every day, more loan applications are submitted, making loan approval more difficult. When choosing an applicant for loan approval, they must take into account a few bank policies. The bank must select the proposal that has the best chance of being approved based on a few criteria. To manually check out each person before recommending them for loan approval is difficult and dangerous. Based on the prior performance of the person to whom the loan amount was previously accredited, we utilise a machine learning technique in this study to forecast the person who is trustworthy for a loan. The main goal of this effort is to foresee if a certain person will be approved for a loan.

This paper has taken the data of previous customers of various banks to whom on a set of parameters loan were approved. So the machine learning model is trained on that record to get accurate results. Our main objective of this research is to predict the safety of loan [1][3]. To predict loan safety, the logistic regression algorithm is used. First the data is cleaned so as to avoid the missing values in the data set. To train our model data set of 1500 cases and 10 numerical and 8 categorical attributes has been taken. To credit a loan to customer various parameters like CIBIL Score (Credit History), Business Value, Assets of Customer etc has been considered. List of parameters as shown below

Qualification	Categorical
In Service / Business Owner	Categorical
Individual income of Co-Applicant (if any)	Qualitative
Amount of Loan Required	Qualitative
Term for which loan Required	Qualitative
Credit history of Applicant	Qualitative
Area of property	Categorical

2. STATE OF THE ART (LITERATURE SURVEY)

A. A] Prediction of Modernized Loan Approval System Based on Machine Learning

Technology has boosted the existence of human kind the quality of a life they live, Every day we are planning to create something new and different, We have a solution for every other problem we have machines to support our lives and make us somewhat complete in the banking sector candidate gets

proofs/ backup before approval of the loan amount . The application approved or not approved depends upon the historical data of the candidate by the system. Every day lots of people applying for the loan in the banking sector but bank have limited fund. In this case, the right prediction would be very beneficial using some classes-function algorithm. An example the logistic

B. Design and Simulation of Loan Approval Prediction Model using AWS Platform

To determine and to understand the working of loan systems for the cause of Loan Prediction using the demographic information of various factors that combine to form the nature of the approval using algorithms and concepts of Machine Learning and ultimately deploying this model on Cloud Based Platforms. Machine learning being aided by Cloud services are progressively seeing immense growth in the industry as they have benefits of Scalability, Affordability and easy use of models on systems as and when required. Therefore datasets are designed, automated and put under testing and training. The major aim of this project is to predict which of the customers will have their loan paid or not using prominent algorithms like Decision Tree, Logistic Regression and Random Forest. Logistic Regression Confusion matrix analysis is relatively in accordance to Decision Tree and Random Forest algorithm helping us attain an accuracy of 86

C. A federated learning based approach for loan defaults prediction

The number of defaults in bank loans have recently been increasing in the past years. However, the process of sanctioning the loan has still been done manually in many of the banking organizations. Dependency on human intervention and delay in results have been the biggest obstacles in this system. While implementing machine learning models for banking applications, the security of sensitive customer banking data has always been a crucial concern and with strong legislative rules in place, sharing of data with other organizations is not possible. Along with this, the loan dataset is imbalance, there are very few samples of default as compared to repaid loans. Hence, these problems make the default prediction system difficult to learn the patterns of defaults and thus difficult to predict them. Previous machine learning-based approaches to automate the process have been training models on the same organization's data but in today's world, classifying the loan application on the data within the organizations is no longer sufficient and a feasible solution. In this paper, we propose a federated learningbased approach for the prediction of loan applications that are less likely to be repaid which helps in resolving the above mentioned issues by sharing the weight of the model which are aggregated at the central server.

D. Bank Loan Prediction System using Machine Learning

With the advancement in technology, there are so many enhancements in the banking sector also. The number of applications is increasing every day for loan approval. There are some bank policies that they have to consider while selecting an applicant for loan approval. Based on some parameters, the bank has to decide which one is best for approval. It is tough and risky to check out manually every person and then recommended for loan approval. In this work, we use a machine learning technique that will predict the person who is reliable for a loan, based on the previous record of the person whom the loan amount is accredited before. This work's primary objective is to predict whether the loan approval to a specific individual is safe or not.

3. TECHNICAL DETAILS

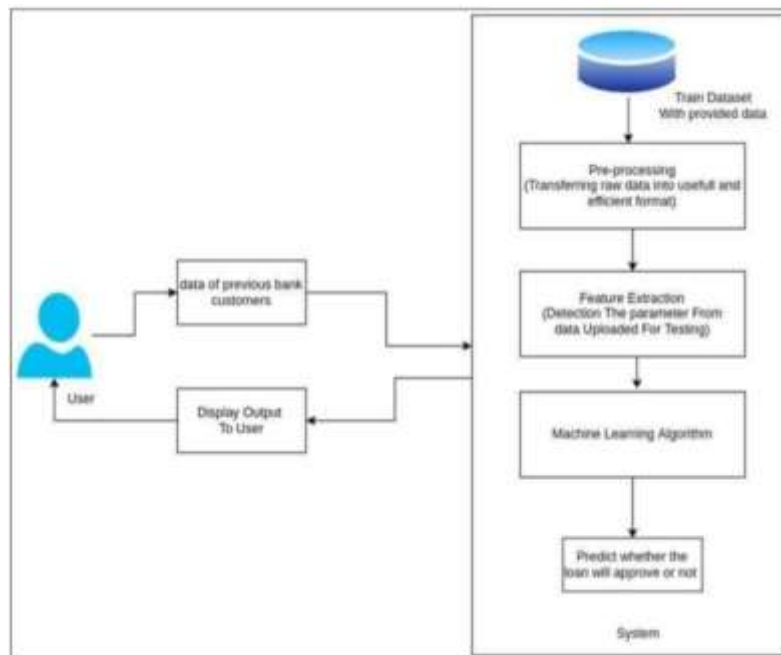
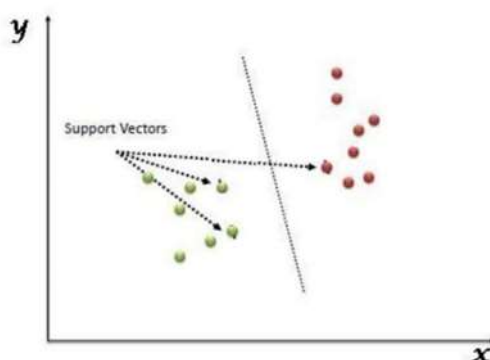


Fig.1.Proposed System

A. Algorithm

SVM (Support Vector Machine):

The algorithm that we shall be using for this purpose, is the Support Vector Machine. Support Vector Machine,(SVM), falls under the "supervised machine learning algorithms" category. It can be used for classification, as well as for regression. In this model, we plot each data item as a unique point in an n-dimension, (where n is actually, the number of features that we have), with the value of each of the features being the value of that particular coordinate. Then, we perform the process of classification by finding the hyper-plane that differentiates the two classes. Take a look at the graph below:



You might be thinking the same question. Why SVM, and not the other commonly used algorithms like Logistic Regression, etc. Well, you see there are no such algorithms, which we can strictly say, are better than the other. It all depends on the type of operations we are performing, and the type of data we are dealing with.

SVM is preferred over other algorithms when:

- 1) The data is not regularly distributed.
- 2) SVM is generally known to not suffer the condition of overfitting.
- 3) Performance of SVM, and its generalization is better on the dataset.
- 4) And, lastly, SVM is known to have the best results for classification types of problems.

B Mathematical Model

Let S be the Whole system $S = (I, P, O)$

I-input

P-procedure O-output Input(I)

I= (Input as Text Dataset of Loan) Where,

Dataset- ζ dataset contain Input as Loan Predication Features like ,Sex ,Married

No-Dependents , Qualification , In Service / Self- Employed ,

AnnualIncomeApplicant,Annual-income- Coapplicant,Amount-Loan,Term,CreditHistory Applicant ,etc.

Procedure (P),

$P = (I, \text{Using } I \text{ System perform operations and calculate the prediction,}$

1. Data Pre-processing

3)Feature Extraction

2. Classification using SVM)

Output(O)-

$O = (O, \text{As per input features Predict Loan})$

4. CONCLUSION

Conclude, That in system Data cleansing and processing, imputation of missing values, experimental analysis of the data set, model construction, model evaluation, and model testing are all steps in the prediction process. The highest level of accuracy on the data set, based on the Training Model. After considering that applicants with the lowest credit scores won't be approved for loans because they have a higher likelihood of not repaying the loan amount, the following conclusions are drawn. It makes sense that loan amount, the following conclusions are drawn. It makes sense that applicants who request smaller loans but have high incomes are more likely to be approved because they are also more likely to repay their loans.

5. REFERENCE

- [1]. Vishal Singh “Prediction of Modernized Loan Approval System Based on Machine Learning Approach”,2021.
- [2]. Vaidya “PREDICTIVE AND PROBABILISTIC APPROACH USING LOGISTIC REGRESSION :APPLICATION TO PREDICTION OF LOAN APPROVAL”,2017.
- [3]. Anshika Gupta “Bank Loan Prediction System using Machine Learning”,2021. ramachandra H V “Design and Simulation of Loan Approval Prediction Model using AWS Platform”,2021.
- [4]. Sujoy Barua “Swindle: Predicting the Probability of Loan Defaults using CatBoost Algorithm”,2021.
- [5]. Kumar Arun, Garg Ishan, Kaur Sanmeet, May- Jun. 2016. Loan Approval Prediction based on Machine Learning Approach, IOSR Journal of Computer Engineering (IOSR-JCE)
- [6]. Wei Li, Shuai Ding, Yi Chen, and Shanlin Yang, Heterogeneous Ensemble for Default Prediction of Peer-to-Peer Lending in China, Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei University of Technology, Hefei 2009, China
- [7]. Short-term prediction of Mortgage default using ensembled machine learning models, Jesse C.Sealand on July 20, 2018.
- [8]. Clustering Loan Applicants based on Risk Percentage using K-Means Clustering Techniques, Dr. K. Kavitha, International Journal of Advanced Research in Computer Science and Software Engineering.
- [9]. Toby Segaran, “Programming Collective Intelligence: Building Smart Web 2.0 Applications.” O’Reilly Media.
- [10]. Drew Conway and John Myles White,” Machine Learning for Hackers: Case Studies and Algorithms to Get you Started,” O’Reilly Media.
- [11]. Trevor Hastie, Robert Tibshirani, and Jerome Friedman,”The Elements of Statistical Learning: Data Mining, Inference, and Prediction,” Springer Kindle
- [12]. PhilHyo Jin Do ,Ho-Jin Choi, “Sentiment analysis of real-life situations using location, people and time as contextual features,” International Conference on Big Data and Smart Computing (BIGCOMP), pp. 39–42. IEEE, 2015.
- [13]. Bing Liu, “Sentiment Analysis and Opinion Mining,” Morgan & Claypool Publishers, May 2012.
- [14]. Bing Liu, “Sentiment Analysis: Mining Opinions, Sentiments, and Emotions,” Cambridge University Press, ISBN:978-1-107-01789-4.
- [15]. Shiyang Liao, Junbo Wang, Ruiyun Yu, Koichi Sato, and Zixue Cheng, “CNN for situations understanding based on sentiment analysis of twitter data,” Procedia computer science, 111:376–381, 2017.CrossRef.
- [16]. K I Rahmani, M.A. Ansari, Amit Kumar Goel, “An Efficient Indexing Algorithm for CBIR,”IEEE- International Conference on Computational Intelligence & Communication Technology ,13-14 Feb 2015.
- [17]. Gurlove Singh, Amit Kumar Goel ,”Face Detection and Recognition System using Digital Image Processing” , 2nd International conference on Innovative Mechanism for Industry Application ICMA 2020, 5-7 March 2020, IEEE Publisher.
- [18]. Amit Kumar Goel, Kalpana Batra, Poonam Phogat,” Manage big data using optical networks”, Journal of Statistics and Management Systems “Volume 23, 2020, Issue 2, Taylors & Francis.