International Journal of Innovative Research in Engineering
Volume 4, Issue 3 (May-June 2023), PP: 295-298.
www.theijire.com ***ISSN No: 2582-8746***

Smart Lender Applicant Credibility Prediction for Loan Approval

Chandru V[1], Poovarasan D[2], Tamilan M[3], Abilesh R[4], Umameshwari M[5]
[1,2,3,4,5]Computer Science and Engineering, The Kavery Engineering College/ Anna University, Tamilnadu, India.

How to cite this paper:
Chandru V[1], Poovarasan D[2], Tamilan M[3], Abilesh R[4], Umameshwari M[5], "Smart Lender Applicant Credibility Prediction for Loan Approval", IJIRE-V4I03-295-298.

https://www.doi.org/10.59256/ijire.2023040392

Abstract: *In our banking system, banks have many products to sell but main source of income of any banks is on its credit line. So they can earn from interest of those loans which they credits. A bank's profit or a loss depends to a large extent on loans i.e. whether the customers are paying back the loan or defaulting. By predicting the loan defaulters, the bank can reduce its Non-performing Assets. This makes the study of this phenomenon very important. Previous research in this era has shown that there are so many methods to study the problem of controlling loan default. But as the right predictions are very important for the maximization of profits, it is essential to study the nature of the different methods and their comparison. A very important approach in predictive analytics is used to study the problem of predicting loan defaulters (i) Collection of Data, (ii) Data Cleaning and (iii) Performance Evaluation. Experimental tests found that the Naïve Bayes model has better performance than other models interms of loan forecasting.*

Key Word: *Support Vector Machine(SVM), Machine learning, Loan Prediction.*

I.INTRODUCTION

One of the most important factors which affect our country's economy and financial condition is the credit system governed by the banks. The process of bank credit risk evaluation is recognized at banks across the globe. "As we know credit risk evaluation is very crucial, there is a variety of techniques are used for risk level calculation. In addition, credit risk is one of the main functions of the banking community .The prediction of credit defaulters is one of the difficult tasks for any bank. But by forecasting the loan defaulters, the banks definitely may reduce their loss by reducing non-profit assets, so that recovery of approved loans can take place without any loss and it can play as the contributing parameter of the bank statement. This makes the study of this loan approval prediction important. Machine Learning techniques are very crucial and useful in the prediction of these types of data. We will be using classification algorithms such as Decision tree, Random Forest, KNN, and xgboost. We will train and test the data with these algorithms. From this best model is selected and saved in pkl format. Flask integration and IBM deployment is also be done.

II.LITERATURE SURVEY

" Loan Approval Prediction based on Machine Learning Approach" Author- Kumar Arun, Garg Ishan, Kaur Sanmeet Year- 2018The main objective of this paper is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing.

"Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process" Author- E. Chandra Blessie, R. Rekha - Year- 2019 Extending credits to corporates and individuals for the smooth functioning of growing economies like India is inevitable. As increasing number of customers apply for loans in the banks and non- banking financial ompanies (NBFC), it is really challenging for banks and NBFCs with limited capital to device a standard resolution and safe procedure to lend money to its borrowers for their financial needs. In addition, in recent times NBFC inventories have suffered a significant downfall in terms of the stock price. It has contributed to a contagion that has also spread to other financial stocks, adversely affecting the benchmark in recent times.n this paper, an attempt is made to condense the risk involved in selecting the suitable person who could repay the loan on time thereby keeping the bank's nonperforming assets (NPA) on the hold. This is achieved by feeding the past records ofthe customer who acquired loans from the bank into a trained machine learning model which could yield an accurate result. Theprime focus of the paper is to determine whether or not it will be safe to allocate the loan to a particular person. This paper has thefollowing sections (i) Collection of Data, (ii) Data Cleaning and (iii) Performance Evaluation. Experimental tests found that the Naïve Bayes model has better performance Evaluation. Experimental tests found that the Naïve Bayes model has betterperformance than other models in terms of loan forecasting.

"Loan Prediction using machine learning model" Year-2019whether or not it will be safe to allocate the loan to a particularperson. This paper has the following sections (i) Collection of Data, (ii) Data Cleaning and (iii) Performance Evaluation.Experimental tests found that the Naïve Bayes model has better performance than other models in terms of loan forecasting.With the enhancement in the banking sector lots of people are applying for bankloans but the bank has its

limited assets which it has togrant to limited people only,so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this project we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result The main objective of this project is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of systemon most promising model (iv) Testing. In this paper we are predict the loan data by using some machine learning algorithms they are classification, logic regression, Decision Tree and gradient boosting.

"Loan Prediction using Decision Tree and Random Forest"Author- Kshitiz Gautam, Arun Pratap Singh, Keshav Tyagi, Mr. Suresh Kumar Year-2020. In India the number of people or organization applying for loan gets increasd every year. The bank have to put in a lot of work to analyse or predict whether the customer can pay back the loan amount or not (defaulter or non-defaulter) in the given time. The aim of this paper is to find the nature or background or credibility of client that is applying for the loan. We use exploratory data analysis technique to deal with problem of approving or rejecting the loan request or in short loan prediction. The main focus of this paper is to determine whether the loan given to a particular person or an organization shall be approved or not.

III.PROBLEM STATEMENT

To design and implement the system using machine learning and data mining to predict the probability of the user to get loan or not from bank to improve the accuracy and to minimize the frauds. Banks, Housing Finance Companies and some NBFC deal in various types of loans like housing loan, personal loan, business loan etc in all over the part of countries. These companies have existence in Rural, Semi Urban and Urban areas. After applying loan by customer these companies validates the eligibility of customers to get the loan or not. This paper provides a solution to automate this process by employing machine learning algorithm. So the customer will fill an online loan application form. This form consist details like Sex, Marital Status, Qualification, Details of Dependents, Annual Income, Amount of Loan, Credit History of Applicant and others. To automate this process by using machine learning algorithm, First the algorithm will identify those segments of the customers who are eligible to get loan amounts so bank can focus on these customers.

IV.PROPOSED SYSTEM

S.NO	PARAMETER	DESCRIPTION
1	Problem Statement (Problem to be solved)	The prediction of credit defaulters is one of the difficult tasks for any bank. But by forecasting the loan defaulters, the banks definitely may reduce their loss by reducing their non-profit assets, so that recovery of approved loans can take place without any loss and it can play as the contributing parameter of the bank statement. But manually assessing the credibility of applicants is a time consuming process and incorrect manya times.
2	Idea / Solution description	A Machine learning model must be developed to predict the credit defaulters. This model must be trained on previous Loan approval data and their manual credibility checked data. This can be then used to predict the applicant's credibility automatically.
3	Novelty / Uniqueness	In this model, the previous manually checked credibility is taken as training data. Once trained it will take Data on Loan history, Financial status and stability, Family status and Co- applicant Credibility as inputs and will provide a Booleanvalue output for credibility.
4	Social Impact / Customer Satisfaction	This model mostly predicts the credibility of a loan applicant accurately, automatically in less time compared to conventional manual checking. This socially helps banks to identify credible loan applicants thus also reduces the loss factor of the Lender (usually Bank). It also speeds up the loan sanctioning process, thus helping the applicants too.
5	Business Model (Revenue Model)	A model without human intervention reduces capital investment for the man power and it saves time consumed in this manual process. It will also be accurate than the manual credibility checking process, thus preventing money landing on fraudulent hands.
6	Scalability of the Solution	This model can be used with any number of Loan Applicant data and the same algorithm can be used in all the banks or all lenders. With proper organisation and pre-processing of the data about the loan applicant the above proposed solution is completely scalable.

V.REQUIREMENT ANALYSIS

Functional requirement

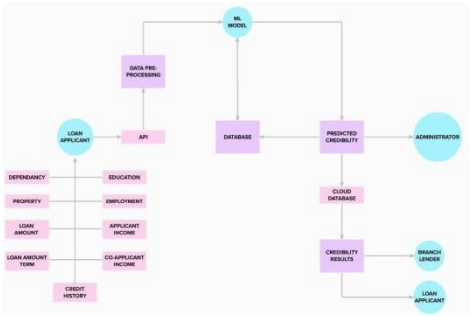
S.NO	FUNCTIONAL REQUIREMENT (EPIC)	SUB REQUIREMENT(STORY / SUB-TASK)
1	User Registration	Registration through Form Registration through GmailRegistration through LinkedIN.
2	User Confirmation	Confirmation via Email In-Person Confirmation
3	User Requirements	Knowledge on how to Input Data Basic idea on using a Machine Learning Model.
4	User Infrastructure	A system with suitable CPU and GPU to support training anddeployment of ML model.
5	Final Result Visualization	Generated results will be visible to the admin through a Webserver.

Non-Functional Requirement

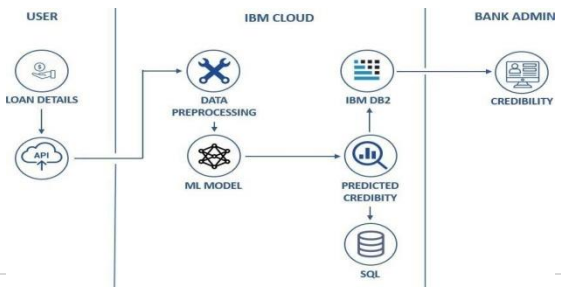
S.NO	NON-FUNCTIONAL REQUIREMENT	DESCRIPTION
1	Usability	It can be used by all Lenders (i.e., Banks) and can be trained specifically with respect to the Location's Financial Stability.
2	Security	Storage and Transfer is secure via Encryption Methodologies.
3	Reliability	The Predicted Credibility is highly reliable than Manual Identification.
4	Performance	The Performance relies on Input Dataset.
5	Availability	Can be made available as a Software.
6	Scalability	Can be scaled for more branches of the same Lender and thedataset can be shared.

VI.PROJECT DESIGN

Data Flow Diagram:



Solution & Technical Architecture:



Model Summary:

```
In [33]: model_details

Out[33]: {'entity': {'hybrid_pipeline_software_specs': [],
  'label_column': 'Loan_Status',
  'schemas': {'input': [{'fields': [{'name': 'Gender', 'type': 'int64'},
     {'name': 'Married', 'type': 'int64'},
     {'name': 'Dependents', 'type': 'int64'},
     {'name': 'Education', 'type': 'int64'},
     {'name': 'Self_Employed', 'type': 'int64'},
     {'name': 'ApplicantIncome', 'type': 'int64'},
     {'name': 'CoapplicantIncome', 'type': 'int64'},
     {'name': 'LoanAmount', 'type': 'int64'},
     {'name': 'Loan_Amount_Term', 'type': 'int64'},
     {'name': 'Credit_History', 'type': 'int64'},
     {'name': 'Property_Area', 'type': 'int64'}],
    'id': '1',
    'type': 'struct'}],
   'output': []},
  'software_spec': {'id': '12b83a17-24d8-5082-900f-0ab31fbfd3cb',
   'name': 'runtime-22.1-py3.9'},
  'type': 'scikit-learn_1.0'},
 'metadata': {'created_at': '2022-11-16T10:48:43.132Z',
  'id': '03542d22-55b9-4830-af6f-c000da875e4e',
  'modified_at': '2022-11-16T10:48:46.959Z',
  'name': 'Model_building_SL_223_IBM',
  'owner': 'IBMid-6620042VBA',
  'resource_key': 'cdb1c157-cfd2-4271-a4a5-9f28198439ca',
  'space_id': '99b1a4a9-7cc5-4852-9388-f8907fe20de7'},
 'system': {'warnings': []}}
```

Accuracy:

Accuracy	0.9851000180244446
Loss	0.0481402396762371

VII.CONCLUSION

The main aim of my research is to develop a new feature selection algorithm. This feature will enable the banks to predict accurately if the customer can repay the loan on time or not. Kudo and Sklansky applied a two-stage feature selection, where a filter method preceded a wrapper algorithm.This combination was shown to select better feature subsets than if the wrapper model is applied directly. Algorithm-based methods rank the features as a pre- processing step prior to the learning algorithm, and select those features with high ranking scores. Algorithm based methods score the features using the learning algorithm that will ultimately be employed. Random Forest algorithm is a proposed hybrid feature selection algorithm that comprises the features of both wrapper model and fisher score concept. LCPS produces better accuracy while it was executed by proposed algorithm. The proposed model experimented on two different data sets such as original dataset and standard data set with improved accuracy thantraditional random forest algorithm.

VIII.FUTURE SCOPE

The further works that can be done in this project is to include few more features in model training to study the effect on the prediction. A long history of data (dataset of more than 3 years) can be used for training for ncreased accuracy. The application can be upgraded such that the input values are fetched directly from the application file and then fed to the model rather than the user entering it manually. A login systems for banks can be developed, so that each bank can have its own login hence making their applicants data more secure.

References

[1] *B. P. Lohani, M. Trivedi, R. J. Singh, V. Bibhu, S. Ranjan and P. K. Kushwaha, "Machine Learning Based Model for Prediction of Loan Approval," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), 2022, pp. 465-470, doi: 10.1109/ICIEM54221.2022.9853160.*

[2] *Soni PM, Varghese Paul, "Algorithm For the Loan Credibility Prediction System", International Journal of Recent Technology and Engineering (IJRTE), Volume-8, Issue-1S4, June 2019.*

[3] *Okfalisa, R. Fitriani and Y. Vitriani, "The Comparison of Linear Regression Method and K- Nearest Neighbors in Scholarship Recipient," 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2018, pp. 194-199, doi: 10.1109/SNPD.2018.8441068.*

[4] *Kathe Rutika Pramod, Panhale Sakshi Dattatray, Avhad Pooja Prakash, "An Approach For Prediction Of Loan Approval using Machine Learning Algorithm",International Journal of CreativeResearch Thoughts (IJCRT), Volume 9, Issue 6, June 2021.*

[5] *Ambika, & Biradar, Santosh. (2021). Survey on Prediction of Loan Approval Using Machine Learning Techniques. International Journal of Advanced Research in Science, Communication and Technology. 449-454. 10.48175/IJARSCT-1165.*

[6] *Yash Divate, Prashant Rana, Pratik Chavan, "Loan Approval Prediction Using Machine Learning", International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 05, May 2021.*

[7] *Anant Shinde, Yash Patil, Ishan Kotian, Abhinav Shinde, Reshma Gulwani, "Loan Prediction System Using Machine Learning", International Conference on Automation, Computing and Communication 2022 (ICACC-2022), Volume 44, May 2022.*

[8] *Q. Du, N. Li, S. Yang, D. Sun and W. Liu, "Integrating KNN and Gradient Boosting Decision Tree for Recommendation," 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021, pp. 2042-2049, doi: 10.1109/IAEAC50856.2021.9390647.*