Loan Eligibility Predictor

Pallavi Saindane, Anjali Asrani, Ajay Bathani, Kunal Dongare and
Sneha Indulkar

May 4, 2021

LOAN ELIGIBILITY PREDICTOR

Department of Computer Engineering,

Vivekananda Education Society's Institute of Technology,

Mrs. Pallavi Saindane (CPMN)
Professor,VESIT-Chembur
pallavi.saindane@ves.ac.in

Anjali Asrani (CMPN)
Student,VESIT-Chembur
2017.anjali.asrani@ves.ac.in

Ajay Bhatani (CMPN)
Student,VESIT-Chembur
2017.ajay.bhtani@ves.ac.in

Kunal donagre (CMPN)
Student,VESIT-Chembur
2017.kunal.donagre@ves.ac.in

Sneha Indulkar (CMPN)
Student,VESIT-Chembur
2017.sneha.indulkar@ves.ac.in

Abstract: With the improvements within the banking sector numerous individuals are applying for bank loans yet the bank has its limited resources which it has to grant to a small amount of people only, so finding out to whom the loan can be permitted which will be a safer option for the bank is a typical process. So in this project we try to reduce this risk element behind selecting the safe person so as to save lots of bank endeavor and assets. This is done by using data of the preceding records of the people to whom the loan was granted before and on the basis of these records the machine was trained using the python and ML model which give the most precise result. The main objective of this project is to predict whether assigning the loan to a specific person will be safe or not

1. INTRODUCTION

Nowadays many banks/finance firms give loans after a regress process of authentication and validation but still there is no guarantee whether the chosen applicant is the deserving right applicant out of all applicants. With this model we can predict whether that specific applicant is prudent or not and the whole process of verification of features is self-directed by machine learning technique. Loan Prediction is very helpful for employees of banks as well as for the applicant also. The aim of this system is to provide a quick, immediate and easy way to choose the deserving applicants. It can provide exceptional supremacy to the Bank/Finance firm. To train the model dataset used for the loan prediction model has been provided by Capitalcity Finance.

2. DATA SET

The training data set is now supplied to the machine learning model, on the basis of this data set the model is trained. There are a total of 13 columns available for further processing and 500 plus data values available for training purpose. Every new applicant details filled at the time of application form acts as a test data set. After the operation of testing, model predict whether the new applicant is a fit case for approval of the loan or not based upon the rule and it concludes on the basis of the training data sets .

Dataset Info:

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 615 entries, 0 to 614
Data columns (total 13 columns):
 #   Column             Non-Null Count   Dtype
---  ------             --------------   -----
 0   Loan_ID            615 non-null     object
 1   Gender             602 non-null     object
 2   Married            612 non-null     object
 3   Dependents         600 non-null     object
 4   Education          615 non-null     object
 5   Self_Employed      583 non-null     object
 6   ApplicantIncome    615 non-null     int64
 7   CoapplicantIncome  615 non-null     float64
 8   LoanAmount         593 non-null     float64
 9   Loan_Amount_Term   601 non-null     float64
 10  Credit_History     565 non-null     float64
 11  Property_Area      615 non-null     object
 12  Loan_Status        615 non-null     object
dtypes: float64(4), int64(1), object(8)
memory usage: 62.6+ KB
```

3. LITERATURE SURVEY

[1]Aditi Kacheria, Nidhi Shivakumar, Shreya Sawkar and Archana Gupta,"Loan Sanctioning Prediction System," International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-6 Issue-4, September 2016

This paper proposes a loan sanctioning system which determines whether or not a loan should be given to a person, based on certain attributes. In spite of banks following stringent rules and regulations and conducting meticulous background checks while sanctioning a loan and keeping in mind the probability of the person's ability to return the loan, often such situations are faced where the person is unable to repay the loan that has been given to him. In this paper, the system that we propose for the bankers will help them predict the credible customers who have applied for a loan, thereby improving the chances of their loans being repaid in time. This classification is done using the Naïve Bayesian algorithm. In order to improve the classification accuracy, the quality of the data is improved before classifying it by using K-NN and Binning algorithms. This system uses these algorithms in order to yield a better efficiency so as to reduce the possibility of such a problem. The proposed system additionally facilitates self-confirmation regarding the same for the commoner.

[3]Kwofie, Charles & Owusu-Ansah, Caleb & Boadi, Caleb. (2015). Predicting the Probability of Loan-Default: An Application of Binary Logistic Regression. Research Journal of Mathematics and Statistics. 7. 46-52. 10.19026/rjms.7.2206.

This study examines the performance of logistic regression in predicting probability of default using data from a microfinance company. A logistic regression analysis was conducted to predict default status of loan beneficiaries using 90 sampled beneficiaries for model building and 30 out of sample beneficiaries for prediction. Age, marital status, gender number of years of education, number of years in business and base capital were used as predictors. The predictors that were significant in the model were marital status, number of years in business and base capital. The explained variability in the response variable in the logistic regression was very weak

[4]Kumar Arun, Garg Ishan and Kaur Sanmeet "Loan Approval Prediction based on Machine Learning Approach" IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727 PP 18-21

In this paper the author has explained how with the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant 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 paper 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 gives the most accurate result. The main objective of this paper is to predict whether assigning the loan to a 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

4. PREDICTION METHODOLOGY
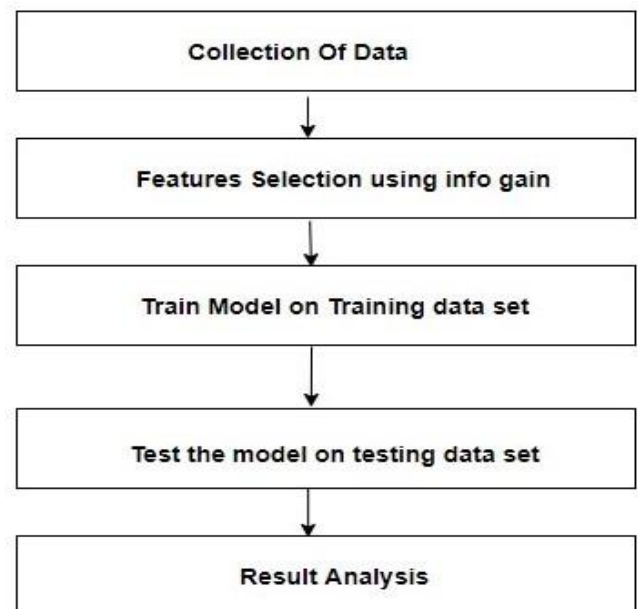
All data we get from a web application that act as input to data processing in that we perform data

completion In the data completion there are often missing values in the dataset due to improper data entry or other data entry problems that can lead to issues in data analysis, thereby affecting the accuracy of the system and after the preprocessing and testing operation and also perform data refining in the data refining data sets often have outliers that are nothing but noisy records present in the data the system will guess whether the new applicant is a fit case for approval of the loan or not based upon the algorithm. it concludes on the basis of the training data set

5. ALGORITHM

Predicting loan eligibility is a classification problem. Logistic regression is a supervised learning algorithm which is mostly used to solve binary "classification" tasks although it contains the word "regression". "Regression"contradictswith "classification" but the focus of logistic regression is on the word "logistic" referring to logistic function which actually does the classification task in the algorithm. Logistic regression is a simple yet very effective classification algorithm so it is commonly used for many binary classification tasks. Customer churn, spam email, website or ad click predictions are some examples of the areas where logistic regression offers a powerful solution. It is even used as an activation function for neural network layers. The basis of logistic regression is the logistic function, also called the sigmoid function, which takes in any real valued number and maps it to a value

6. PROPOSED SYSTEM

The product will begin in a way that it will start by supplying data to the machine learning model, on the basis of this data set the model is trained. Every new candidate details are filled at the time of application form and acts as a test data set. After the testing operation, the system will guess whether the new applicant is a fit case for approval of the loan or not based upon the conclusion it concludes on the basis of the training data sets. The basic data set should have attributes like the loan id, education, self-employed, salary, previous loan history, credit score, loan amount etc. Thus, our prediction model will require the above data to produce accurate result.

.

7. BLOCK DIAGRAM
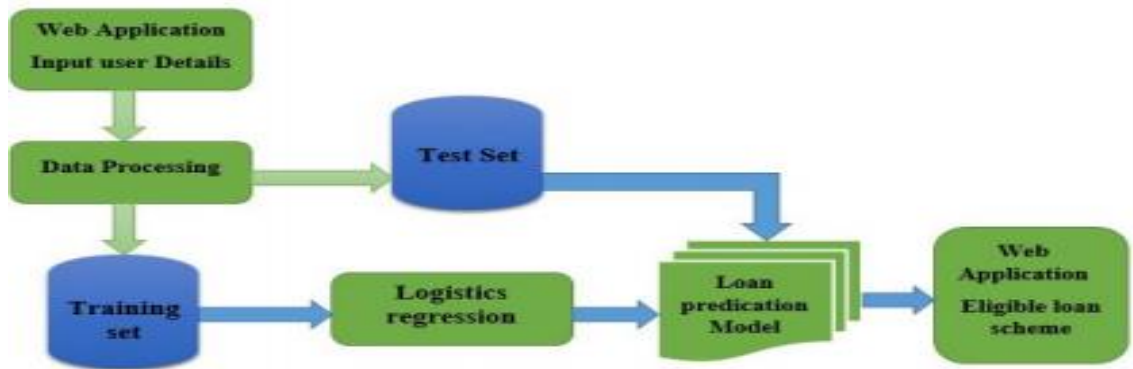
The figures show the flow of the loan prediction system. Firstly we developed a web application that acts as a bridge between the user and our training model. All data we get from a web application that act as input to data processing in that we perform data completion In the data completion there are often missing values in the dataset due to improper data entry or other data entry problems that can lead to issues in data analysis, thereby affecting the accuracy of the system and after the preprocessing and testing operation, the system will guess whether the new applicant is a fit case for approval of the loan or not based upon the algorithm. it concludes on the basis of the training data sets.

8. CONCLUSION

It can be safely concluded from the positive points that the project is a highly efficient component for the banking sector. The product can easily help in identifying the deserving applicants thus producing quick results for the customer by banks. Also it can help the banks in determining whether assigning a loan to a person is safe or not thus answering how risky is the borrower. It is necessary for the banks to identify the risk factor while granting a loan, thus our prediction model can be very useful in such cases.

9. REFERENCES

[1]Aditi Kacheria, Nidhi Shivakumar, Shreya Sawkar, Archana Gupta "Loan Sanctioning Prediction System" International Journal of Soft Computing and Engineering (2016) http://www.ijsce.org/wp-content/uploads/pap ers/v6i4/D2904096416.pdf

[2] Immad Dabbura "Predicting Loan" article Mar 15, 2018

https://towardsdatascience.com/predicting-lo an-repayment-5df4e0023e92

[3] Kwofie, Charles & Owusu-Ansah, Caleb & Boadi, Caleb. (2015). Predicting the Probability of Loan-Default: An Application of Binary Logistic Regression. Research Journal of Mathematics and Statistics. 7. 46-52. 10.19026/rjms.7.2206. https://bit.ly/345awgq

[4]Kumar Arun, Garg Ishan, Kaur Sanmeet "Loan Approval Prediction based on Machine Learning Approach" IOSR Journal of Computer Engineering (IOSR-JCE) https://www.iosrjournals.org/iosr-jce/papers/ conf.15013/Volume%203/4.%2018-21.pdf?id =7557