International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Predicting Loan Eligibility: A Machine Learning Approach

Prof.Karthik

1. K Ranjith 2. P Tejawini 3. S R Ravindra 4. N Renuka 5. A Revanth 6. SK Rezwan Ali

Artificial Intelligence and Machine Learning

Department of Computer Science and Engineering

Malla Reddy University,Hyderabad,Telangana,India

Abstract: The goal of this ML project is to build a model that accurately predicts how much loan a user is eligible for based on their personal and employment details, including their marital status, education, number of dependents, and employment status. This project will use machine learning techniques to build a linear regression model that can analyze and identify patterns in data, to make accurate predictions.

Ones the data is cleaned, a linear regression model will be built using the personal and employment details of users as input features and their corresponding loan amounts as output. The model will be trained on a subset of the dataset and then tested on a separate validation set to assess its accuracy

This ML project has the potential to make the loan application process more efficient and accurate. By automating the loan eligibility process, it can help reduce the workload for loan officers and improve the accuracy of loan decisions

The project has several benefits, including making the loan application process more efficient, reducing the risk of default for lenders, and increasing transparency and fairness in the loan application process.

The project also has several challenges, including collecting and analyzing accurate and relevant data, building a model that is both accurate and interpretable, and creating a user interface that is easy to use and provides useful feedback to the user.

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

1.INTRODUCTION

Prediction of modernized loan approval system based on machine learning approach is a loan approval system from where we can know whether the loan will pass or not. In this system, we take some data from the user like his monthly income, marriage status, loan amount, loan duration, etc. Then the bank will decide according to its parameters whether the client will get the loan or not.

So there is a classification system, in this system, a training set is employed to make the model and the classifier may classify the data items into their appropriate class. A test dataset is created that trains the data and gives the appropriate result that, is the client potential and can repay the loan.

Prediction of a modernized loan approval system is incredibly helpful for banks and also the clients. This system checks the candidate on his priority basis. Customer cansubmit his application directly to the bank so the bank will do the whole process, no third party or stockholder will interfere in it. And finally, the bank

will decide that the candidate is deserving or not on its priority basis. The only object of this research paper is that the deserving candidate gets straight forward and quick results

II. MACHINE LEARNING ALGORITHMS

Logistic regression- Logistic regression is a statistical model used to predict the probability of a binary outcome based on one or more independent variables. It is a type of regression analysis commonly used when the dependent variable is categorical (e.g., yes/no, true/false) rather than continuous. The model applies the logistic function, also known as the sigmoid function, to transform the linear regression equation into a range between 0 and 1, representing the probability of the outcome. This makes logisticregression suitable for classification tasks, where the goal is to assign an observation to one of two classes based on its features. It is a type of regression analysis commonly used when the dependent variable is categorical rather than continuous.

Random Forest– Random forests is a classification algorithm which builds big number of Decision tree, whose prediction is more accurate than any of individual decision tree.

III.PROBLEM STATEMENT

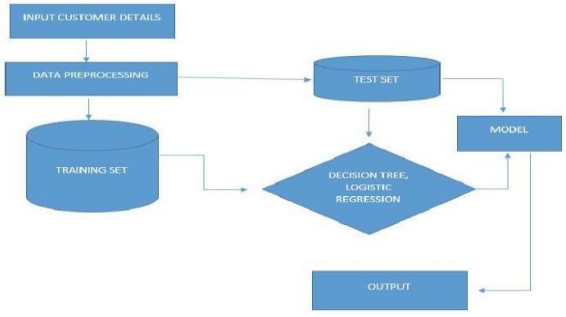
There is a major problem that many people not able to back the loans to banks. And banks are going in losses.

Banks received many applications for loan approval day by day and not everyone gets approved. Most of the banks have their own credit score and risk assessment techniques so as to check that the loan is approved or not. Why this loan problem arises this question will get resolved in just a few minutes. The main reason to get a loan is to fulfill the needs of something. For a businessman he/she wants to increase the business or if that company is at loss to get over from that he/she needs a loan. In middle-class people wants to fulfill their needs so they want a loan. So, the main thing of this to fulfill the needs of someone or for something.

Again the question arises that what are the problems that are forming in providing the loans. The answer to this question that not everybody can loan because if he/she is not able to return then who is providing the loan he/she or the company or the bank that is providing the loan will get in the loss. So, first who is providing the loan they have to verify or set some criteria that who is taking the loan is able to return or not. Like in banks like we have a credit card facility but not everybody gets a credit card. For that, a credit score is there to check whether eligible or not. For credit score one should have a good credit score then he/she be able to get a loan. Some criteria like a source of income should be there for getting a credit card. Banks provide loans on behalf of one who is taking the loan he/she should provide some documents and verify. Like some company not able to provide the loans then banks get in loss and they called it NBFC's.

During this project data processing algorithms are going to study loanapproved data might help in predicting the like defaulters thereby helping the banks for

creating better decisions within the future.



IV. REQUIREMENTS

- Excel
- Jupyter notebook
- Data set
- Numpy
- Pandas
- Logistic regression
- Machine learning algorithms

V. LOAN PREDICTION

DATA ANALYSIS

The question arises that on what basis we analyze that we should provide the loan or not.We have two target variables on that basis we provide the loan to our customer.We have to check all the formalities like income proof, address proof, id proof, etc. Then we provide the loan that the customer is eligible to return or not.In the middle class, there is a major need for loans as parents need for their child's education, for business also.

In some cases, people suddenly undergo financial crises, while some try and scam money from banks.

So, the reason we have to check all the things because banks are not undergoing an NPA loan.

Better the customer, chances of loan to be back are high.

Background verification should be high so that we can expect a return of the loan at the perfect time.

So, we analysis on several bases and these are called our target variable

- Data set

TABLE I. MAIN DATA SET

Variable Name	Description	Type
ID	UniqueLoan Id	Integer
Gender	Male/ Female	Character
Married	Applicant	married (Y/N) Character

TABLE II. DATA SET

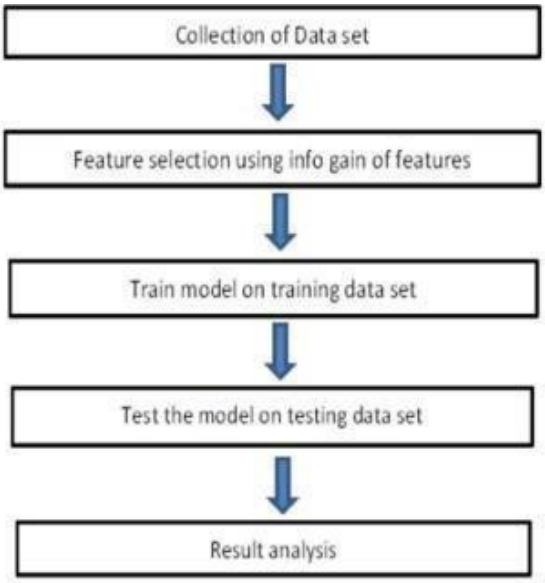
Variable Name	Description	Type
Dependents	Total Number of dependents	Integer
Education	Graduate/ Under Graduate	String
Self_Employed	Employed/Unemployed	(Y/N) Character
Applicant Income	Monthly Income of Applicant	Integer
CoApplicant Income	Coapplicants Income	Integer
Loan Amount	Amount in thousands	Integer
Loan sTerm	Term of loan in months	Integer
Credit_History	credit history guidelines	Integer
Property Loaction	Urban/ Semi Urban/ Rural	String
Loan Status	Loan Approved(Y/N)	Character

VI. LOAN PREDICTION METHODOLOGY

This proposed model will characterize the behavior of customers on the Basis of their record.

These records is taken from the customers, and create a data set. With the help of These data sets and training machine learning
model, we predict that the customer's loan will passed or not.

This Machine algorithms predict the possibility of a customer would be able to repay the loan or not



VII. EXISTING AND PROPOSED SYSTEM :

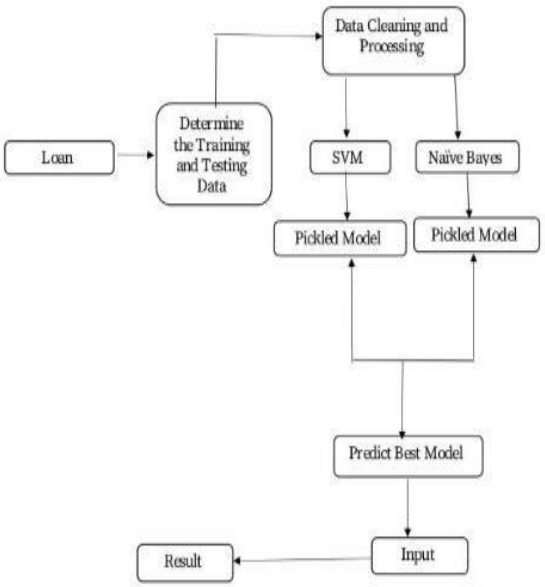
EXISTING SYSTEMS : Existing systems for predicting loan eligibility include traditional credit scoring models, which rely on factors such as credit history, debt-to-income ratio, and payment history. These models are often rulebased and have limited flexibility to adapt to changing market conditions or individual borrower characteristics.

PROPOSED SYSTEMS : Proposed systems for predicting loan eligibility using a machine learning approach aim to overcome these limitations by incorporating more diverse data sources and leveraging the power of complex algorithms to identify patterns and trends in the data.

In conclusion, predicting loan eligibility is a crucial challenge in the banking and financial industry. Machine learning algorithms provide a promising solution to this problem, with logistic regression being one of the most effective techniques for predicting the likelihood of loan approval based on input variables.

.

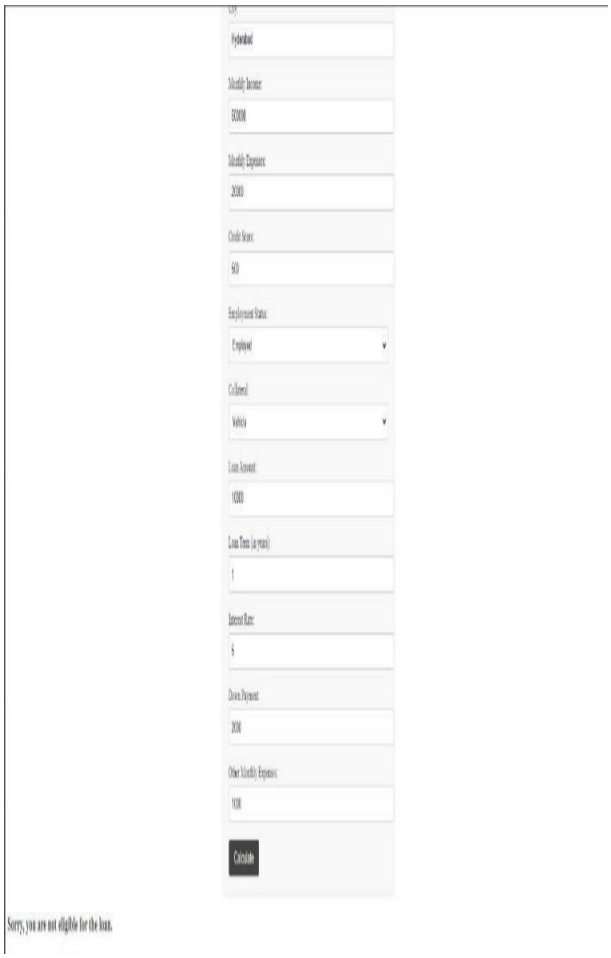
VIII.ARCHITECTURE DIAGRAM FOR PROPOSED METHOD

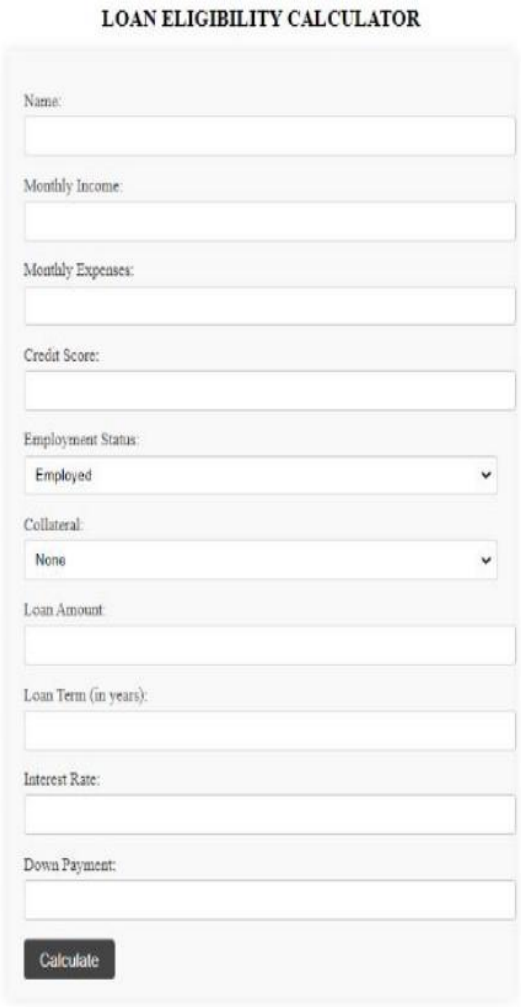


IX. CONCLUSION

In conclusion, predicting loan eligibility is a crucial challenge in the banking and financial industry. Machine learning algorithms provide a promising solution to this problem, with logistic regression being one of the most effective techniques for predicting the likelihood of loan approval based on input variables. However, accurate predictions rely on the quality of the data pre-processing techniques such as encoding and scaling, which play a critical role in preparing data for model training. Additionally, cross-validation can be used to estimate the accuracy of the model, ensuing its effectiveness in real world scenarios.

X. RESULTS


Hyderabad
50000
2000
650
Employed
Vehicle
1000
1
5
200
100
Calculate
Sorry, you are not eligible for the loan.



XI.REFERENCES:

[1] Loan Prediction Using Ensemble Technique, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016

[2] https://www.projectpro.io/articl e/loan-prediction-usingmachine-learning-projectsource-code/632

[3] https://www.kaggle.com/datase ts/altruistdelhite04/loanprediction-problem-datasets

[4] https://mail.iosrjournals.org/ios r-jce/papers/Vol18issue3/Version1/O1803017981.pdf