

Loan Status Prediction Using Exploratory Data Analysis

1. INTRODUCTION

1.1 OVERVIEW

In India, the number of people applying for loans gets increased for various reasons in recent years. The bank employees are not able to analyze or predict whether the customer can pay back the amount or not (good customer or bad customer) for the given interest rate. Loan approval is a very important process for banking organizations. A bank's profit or loss depends to a large extent on loans i.e., whether the customers are paying back the loan or defaulting. Recovery of loans is a major contributing parameter in the financial statements of a bank. It is very difficult to predict the possibility of payment of loan by the customer. The aim is to find the nature of the client applying for a personal loan.

1.2 PURPOSE

The result of the analysis shows that short term loans are preferred by the majority of the clients and the clients majorly apply loans for debt consolidation. The results are shown in graphs that help the bankers to

understand the client's behavior. We will prepare the data using jupyter Notebook and use various models to predict the target variable. Machine Learning (ML) techniques are very useful in predicting outcomes for large amount of data. We use Decision tree (DT) machine learning algorithms predict the loan approval of customers.

2. LITERATURE SURVEY

1." 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.

2."Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process" Author- E. Chandra Blessie, R. Rekha - Year- 2019 .
As increasing number of customers apply for loans in the banks and non-banking financial companies (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 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 of the customer who acquired loans from the bank into a trained machine learning model which could yield an accurate result. The prime 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 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 Evaluation. Experimental tests found that the Naïve Bayes model has better performance than other models in terms of loan forecasting.

3. "Loan Prediction using machine learning model" Year2019whether or not it will be safe to allocate the loan to a particular person. 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. 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 system on most promising model (iv) Testing. In this paper we predict the loan data by using some machine learning algorithms they are classification, logic regression, Decision Tree and gradient boosting.

4. “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 increased every year. The bank has 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

PROPOSED SOLUTION

Decision Trees :

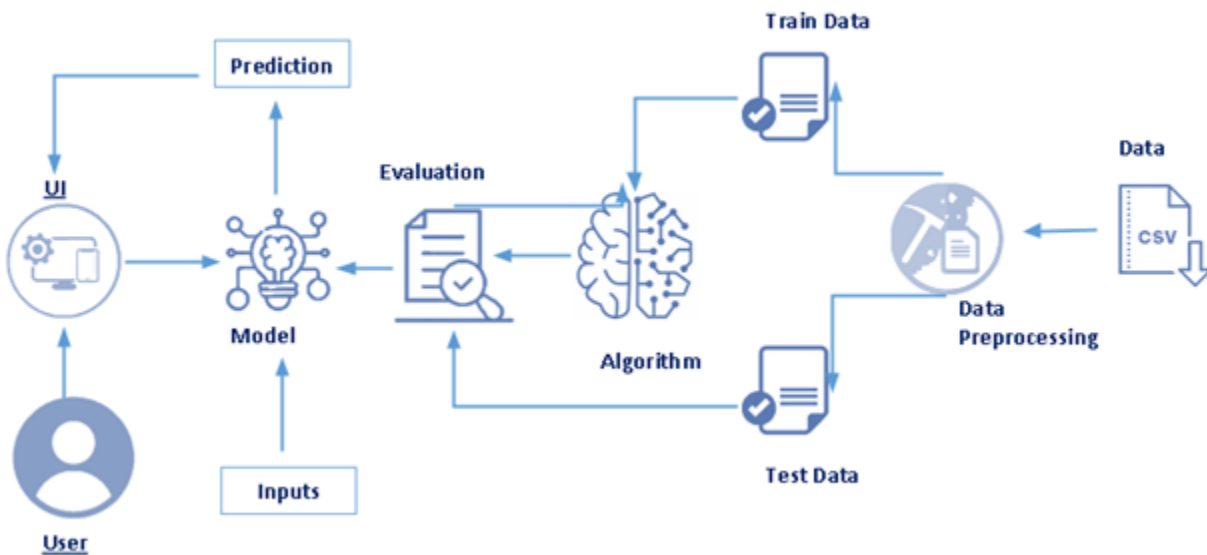
The basic algorithm of decision tree requires all attributes or features should be discretized .The knowledge depicted in decision tree can represented in the form of IF-THEN rules.

One of the most easy and famous classification algorithm is Decision Tree Algorithm .This algorithm helps interpreting and understanding better. Decision tree algorithms is one of the supervised learning algorithms. The decision tree algorithm is capable of solving both, classification and regression problems, which distinguishes it from rest of the supervised learning algorithms.

The main objective of using this algorithm is to predict the value/class of the target variable by learning some decision rules. For predictions using this algorithm, we have to begin with the root node .The value of records attribute and root node are compared .This comparison tells what will be the next node that needs to be followed. Every node acts as a test case for some other attribute, and every edge running down from a node corresponds to a probable answer of that test case. In the starting the entire train dataset is considered as the root of the decision tree. Before building this model the continuous variables are converted into categorical.

THEORITICAL ANALYSIS :

ARCHITECTURE :



REQUIREMENTS :

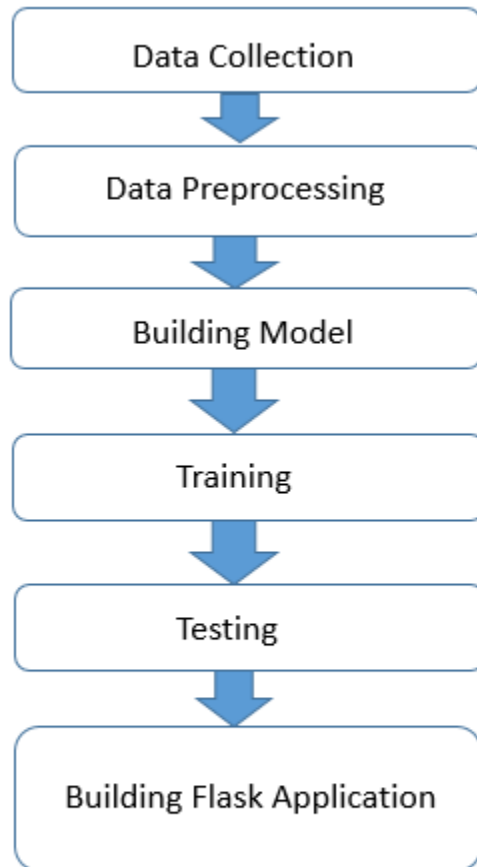
To complete this project you should need the following :

- ✓ Jupyter Notebook for programming, which can be installed by Anaconda IDE.
- ✓ Python packages

EXPERIMENTAL INVESTIGATIONS :

The analysis shows that short term loans are preferred by the majority of the clients and the clients majorly apply loans for debt consolidation. The results are shown in graphs that help the bankers to understand the client's behavior.

CONTROL FLOW OF THE SOLUTION :



RESULT :

In this , we used Decision tree machine learning algorithm to work out a model for loan status prediction using Exploratory data analysis to tell whether the customer is fully paid or charged off .When we give all the details ,based on the dataset it will tell predict. The output can be seen as:

LOAN STATUS PREDICTION

Enter Your Current Loan Amount Small Loan ▼

Enter The Term Type Long Term ▼

Enter Your Credit Score Average ▼

Enter Your Annual Income 400000

Enter Your Years At Work 1

Enter Your Home Ownership Type Own Home ▼

Enter Your Credit History 1

Enter Your Number of Credit Issues 0

Enter If Any Bankruptcies No bankruptcies ▼

Enter Your TaxLiens No Tax Lien ▼

Enter Your Credit Problems No Credit Problem ▼

Enter Your Credit Age Good Credit Age ▼



After giving all the parameters click on the predict button to get the prediction

Predicting Loan Status

A Machine Learning Web App using Flask.

Prediction : **ChargedOff**



ADVANTAGES OF PROPOSED SOLUTION :

1.Easy to read and interpret

One of the advantages of decision trees is that their outputs are easy to read and interpret without requiring statistical knowledge .For example, when using decision trees to present demographic information on customers, they can read and interpret the graphical representation of the data .

2.Easy to prepare

Compared to other, decision trees takes less effort for data preparation. However, users need to have ready information to create new variables with the power to predict the target variable. For complex situations, users can combine decision trees with other methods.

3.Less data cleaning required

There is less data cleaning required once the variables have been created. Cases of missing values and outliers have less significance on the decision trees data.

DISADVANTAGES :

1.Unstable Nature

A small change in the data can result in a major change in the structure of the decision tree, which can convey a different result from what users will get in a normal event. **The resulting change in the outcome can be managed by machine learning algorithms such as boosting and bagging.**

2.Less effective in predicting the outcome of a continuous variable

This is because decision trees tend to lose information when categorizing variables into multiple categories.

Applications:

Marketing:

Businesses can use decision trees to enhance the accuracy of their promotional campaigns by observing the performance of their competitors' products and services. Decision trees can help in audience segmentation and support businesses in producing better-targeted advertisements that have higher conversion rates.

Retention of Customers:

Companies use decision trees for customer retention through analyzing their behaviors and releasing new offers or products to suit those behaviors. By using decision tree models, companies can figure out the satisfaction levels of their customers as well.

Diagnosis of Diseases and Ailments:

Decision trees can help physicians and medical professionals in identifying patients that are at a higher risk of developing serious (or preventable) conditions such as diabetes or dementia. The ability of decision trees to narrow down possibilities according to specific variables is quite helpful in such cases.

Detection of Frauds:

Companies can prevent fraud by using decision trees to identify fraudulent behavior beforehand. It can save companies a lot of resources, including time and money.

CONCLUSION

A system called loan Status prediction system that helps the organization in making right decision to approve or reject the loan request of the customers. Lately people depend on bank loans to meet their wishes. The fee of loan packages will increase with a very rapid speed in current years. Risk is constantly involved in approval of loans. The banking officials are very acutely aware of the price of the mortgage quantity by its customers.

Even after taking lot of precautions and analyzing the mortgage applicant information, the mortgage approval choices are not continually correct. There is need of automation of this system so that loan approval is much less risky and incur less loss for banks. A model is created from a training data. The ML algorithm trained the machine the usage of a fragment of the statistics available and the remaining data is tested.

In this project some ML algorithm Decision Tree is implemented. As it is certainly a very important procedure for a bank to check whether an applicant, applying for a term loan should be approved with the loan or not, this project focuses on making the tedious task mechanized. The required algorithm was implemented successfully and accurate results were generated.

FUTURE SCOPE

Decision tree is a broadly utilized calculation, yet it experiences certain disadvantages. To conquer those disadvantages Ensemble techniques are used and boosting is also used. The future scope of our project is to completely remove human error by making data-sets which are generated by machines instead of humans. In future, the bank employees will be able to analyze or predict whether the customer can pay back the amount or not (good customer or bad customer) for the given interest rate. So that there will be no loss.

BIBILOGRAPHY

[1] Amin, Rafik Khairul and Yuliant Sibaroni, Implementation of decision tree using C4. 5 algorithm in decision making of loan application by debtor.

[2] Kumar Arun, Garg Ishan, Kaur Sanmeet, — Loan Approval Prediction based on Machine Learning Approach, IOSR Journal of Computer Engineering

APPENDIX

SOURCE CODE

<https://github.com/smartinternz02/SI-GuidedProject-4251-1626241363>

Report by Solid 4 team

3rd year CSE

Members:

J.Vamshi Krishna

G.Bharath Chandra

J.Vamshi

P.Rahul