

SMART LENDER - APPLICANT CREDIBILITY

PREDICTION FOR LOAN APPROVAL

Abstract

Banks are making major part of profits through loan. Loan approval is a very important process for banking organizations. It is very difficult to predict the possibility of payment of loan by the customers because there is an increasing rate of loan defaults and the banking authorities are finding it more difficult to correctly access loan request and tackle the risks of people defaulting on loans. In the recent years, many researchers have worked on prediction of loan approval systems. Machine learning technique is very useful in predicting outcomes for large amount of data. In this paper, four algorithms are used such as Random Forest algorithm, Decision Tree algorithm, Naive Bayes algorithm, Logistic Regression algorithm to predict the loan approval of customers. All the four algorithms are going to be used on the same dataset and going to find the algorithm with maximum accuracy to deploy the model. Henceforth, we develop bank loan prediction system using machine learning techniques, so that the system automatically selects the eligible candidates to approve the loan.

INTRODUCTION

A loan is the major source of income for the banking sector of financial risk for banks. Large portions of a bank's assets directly come from the interest earned on loans given. The activity of lending loans carry great risks including the inability of borrower to pay back the loan by the stipulated time. It is referred as "credit risk". A candidate's worthiness for loan approval or rejection was based on a numerical score called "credit score". Therefore, the goal of this paper is to discuss the application of different Machine Learning approach which accurately identifies whom to lend loan to and help banks identify the loan defaulters for much-reduced credit risk.

LITERATURE SURVEY

Bank data management on loan approval processes has great room for improvements of information quality and data problems prevention especially with regards to fair lending and fair pricing practices. They first reviewed briefly typical data collection protocols deployed at many financial institutions for loan approval and loan pricing. Federal regulations mandate portions of these data protocols. While discussing the data capture and analysis for fair lending, they illustrated some initial key steps currently needed for improving information quality to all parties involved.

with the enhancement in the banking sector, lots of people apply for bank loans but the bank has its

limited assets which it grants to only limited people , so finding out to whom the loan can be granted is a typical process for the banks. So, in this paper , they tried to reduce this risk by selecting the safe person so as to save lots of bank efforts and assets. It was done by mining the previous records of the people to whom the loan was granted before and on the basis of these records the machine was trained using the machine learning model which gave the most accurate result. The main goal of this paper is to predict if loan assignment to a specific person

will be safe or not. This paper has into four sections (i) Collection of data (ii) Comparing the machine learning models on collected data (iii) Training the system on most promising model (iv) Testing the system.

Conclusion

The analysis starts from data cleaning and processing missing value, exploratory analysis and finally model building and evaluation of the model. The best accuracy on public test set is when we get higher accuracy score and other performance metrics which will be found out. This paper can help to predict the approval of bank loan or not for a candidate.