Predict If a Customer Would Signing a Loan

Sabitha B [1], Pochaboina Saraswathi Yadav [2], Katakam Shivani [2], Satyasi Anusha [2]

[1]Assistant Professor, Sreyas institute of engineering and Technology, JNTUH, India.

[2]Student, Sreyas institute of engineering and Technology, JNTUH, India.

Corresponding Author: Sabitha.b@sreyas.ac.in

Abstract: - As the loan is one of the most important products of banking and financial companies who have interests in personal loans face a lot of competition from their competitors. Hence to increase profitability and maintain a healthy lending market the company requires predicting customer churn and all the banks are trying to figure out effective business strategies to persuade customers to apply for their loans. Machine learning algorithms have a pretty good performance on this purpose, which are widely used by banking. In this project, there are lots of people applying for bank loans but the banks have limited assets which they have to grant to limited people. This project is done by classifying the previous records of the people which contains various data such as employee id, age, pay schedule, income, years, and months employed, amount requested to predict whether a customer would sign for a loan or not based on their financial history.

Key Words: — *Bathing soaps, Consumer Preference, Factors, Fast Moving Consumer Goods (FMCG).*

I. INTRODUCTION

Distribution of loans is the core business part of almost every bank. The main portion of the bank's assets directly comes from the profit earned from the loans distributed by the banks. So, this type of prediction is performed by lenders to determine the possibility a borrower may default on his debt obligations. Financial history, health status, and risk score depict how a person meets this debt obligation, which establishes the financial character of a person. Financial history counts generally count for 35% of a person's risk score.

Lending loans also consider the number of available liabilities to determine the probability of a customer's default. Also, it sees other factors such as age, income, employment status, debts owed, accounts, length of financial history, and capability to repay debt.

1.1 Motivation

The loan is one of the most important products of banking. All the banks are trying to figure out effective business strategies to persuade customers to apply their loans. However, there are some customers who behave negatively after their applications are approved.

To prevent this situation, banks have to find some methods to predict customers' behaviors. Machine learning algorithms have a pretty good performance on this purpose, which are widely-used in banking. Here, I will work on loan behaviors prediction using machine learning models.

1.2 Objective

The objective of the problem is to pick out which customer will be able to pay the debt and which customer is likely not to be able to pay the debts. Clearly, we have to create a classification model here. We have to use algorithms like logistic regression, decision tree or random forest. We need to create a model that is accurate and the error percentage should be less.

1.3 Scope

This model can be used to compare various machine learning algorithm generated prediction models and the model which will give higher accuracy will be chosen as the prediction model. This paper work can be extended to higher level in future. Predictive model for loans that uses machine learning algorithms, where the results from each graph of the paper can

be taken as individual criteria for the machine learning algorithm.

1.4 Outline

Machine Learning Machine learning is an arena of computer science that involves the learning of pattern identification and computational learning theory in AI. Machine learning generally refers to the changes in systems that carry out tasks linked with artificial intelligence (AI). Such tasks include recognition, analysis, planning, robot control, forecasting, etc. It explores the study and construction of algorithms that can make predictions on data.

II. LITERATURE SURVEY

We started our literature survey with more general systematic literature reviews that focus on the application of machine learning in the general field of Banking Risk Management. A major portion of risk management is the approval of loans to promising candidates. The author pointed out how Artificial neural networks and Logistic regression are most used for loan prediction. Many others also used CART (classifications and regression trees) to predict the customers Analyzing every research and papers we developed a proposed a model using machine learning algorithms like naive Bayes, decision tree and logistic regressions.

III. IMPLEMENTATION

3.1 Module Description

Loan Dataset: Loan Dataset is very useful in our system for prediction of more accurate results. Using the loan Dataset, the system will automatically predict which customer's loan it should approve and which to reject. System will accept loan application forms as an input. Justified format of application form should be given as an input to get processed.

Determine the training and testing data: Typically, Here the system separates a data set into a training set and testing set, most of the data used for training, and a smaller portion of data is used for testing. after a system has been processed by using the training set, it makes the prediction against the test set.

Data cleaning and processing: In Data cleaning the system detects and corrects corrupt or inaccurate records from the database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying or detecting the dirty or coarse data.

In Data processing the system convert data from a given form to a much more usable and desired form i.e., making it more meaningful.
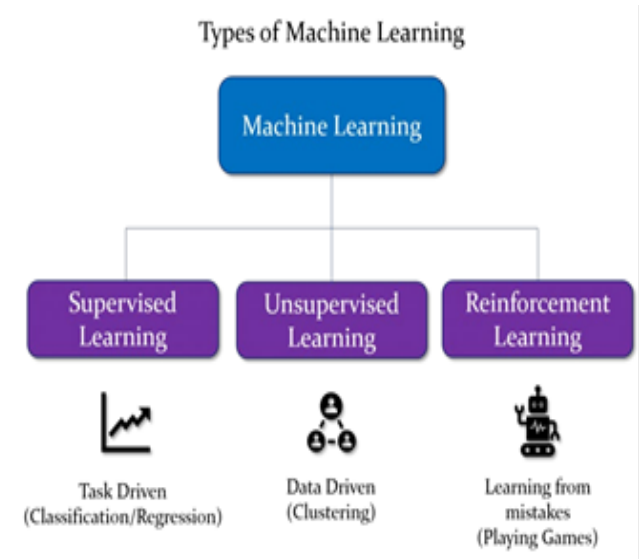

Fig.1. Machine Learning Types

Logistic Regression: This is a classification algorithm which uses a logistic function to predict binary outcome (True/False, 0/1, Yes/No) given an independent variable. The aim of this model is to find a relationship between features and probability of particular outcome. The logistic function used is a logit function which is a log of odds in the favor of the event. Logit function develops a shaped curve with the probability estimate similar to a step function.

Decision Tree: This is a supervised machine learning algorithm mostly used for classification problems. All features should be discretized in this model, so that the population can be split into two or more homogeneous sets or subsets. This model uses a different algorithm to split a node into two or more sub-nodes. With the creation of more sub-nodes, homogeneity and purity of the nodes increases with respect to the dependent variable.

System Architecture: A system architecture diagram would be used to show the relationship between different components. Usually, they are created for systems which include hardware and software and these are represented in the diagram to show the interaction between them. However, it can also be created for web applications.
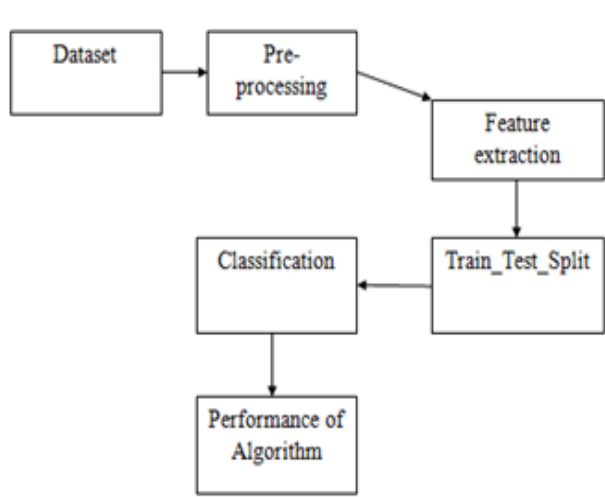
Fig.2. System architecture
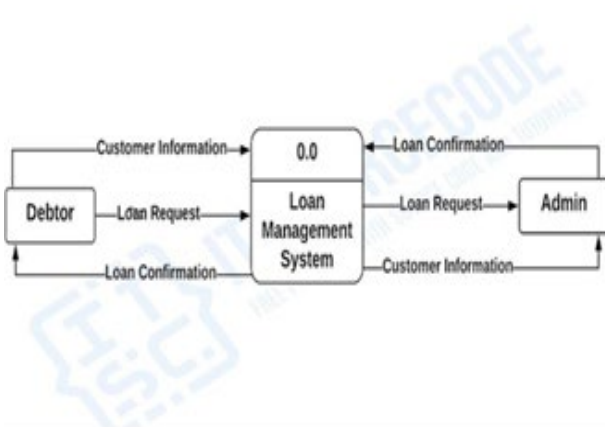
3.2 System design

Fig.3. Flow Diagram

Loan lending has been an important part of daily lives for organizations and individuals alike.
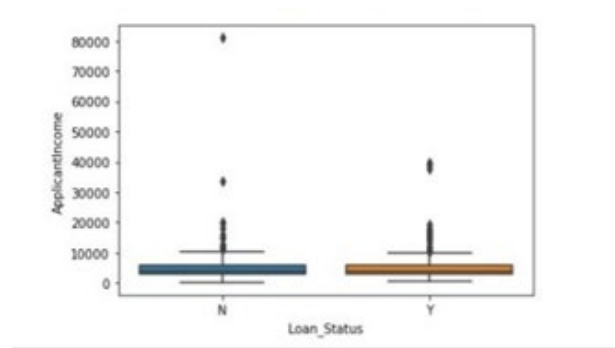
IV. RESULTS

Fig.4. Loan Prediction

V. CONCLUSION

From the Exploratory Data Analysis (EDA), we could generate insight from the data. How each of the features relates to the target? Also, it can be seen from the evaluation of three models that Random Forest performed better than Logistic Regression and Decision Tree.

From a proper analysis of positive points and constraints on the component, it can be safely concluded that the product is a highly efficient component. This application is working properly and meeting to all Banker requirements. This component can be easily plugged in many other systems.

Future Scope:

The system is trained on old training dataset so future software can be made such that new testing data should also take part in training data after some fixed time.

In our banking system, banks have many products to sell but main source of income any banks is on its credit line. By predicting the loan defaulters, the bank can reduce its non-performing assets. This makes the study of this phenomenon very important.

Acknowledgment:

REFERENCES

[1]. A. Goyal and R. Kaur, "A survey on Ensemble Model for Loan Prediction", International Journal of Engineering Trends and Applications (IJETA), vol. 3(1), pp. 32-37, 2016.
[2]. A. J. Hamid and T. M. Ahmed, "Developing Prediction Model of Loan Risk in Banks using Data Mining".
[3]. G. Shaath, "Credit Risk Analysis and Prediction Modelling of Bank Loans Using R".

[4]. A. Goyal and R. Kaur, "Accuracy Prediction for Loan Risk Using Machine Learning Models".