**LOAN APPROVAL PREDICTION BASED ON MACHINE LEARNING APPROACH**

**ABSTRACT:**

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 give the most accurate result. The 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.

**System Architecture:**

Loan Dataset

Pre-Processing

Training of System

Testing

Result as a graph

And

Accuracy Result

**EXISTING:**

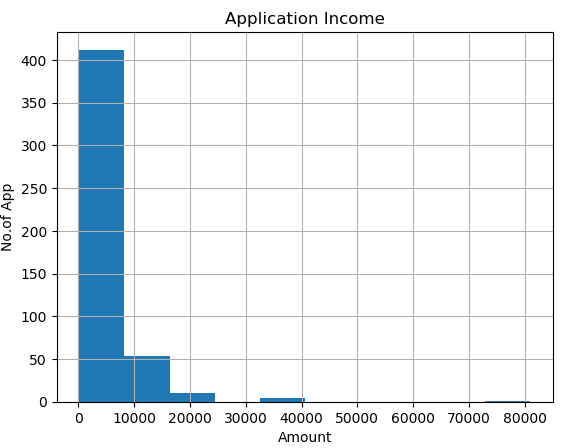
Assessing the risk, which is involved in a loan application, is one of the most important concerns of the banks for survival in the highly competitive market and for profitability. These banks receive number of loan applications from their customers and other people on daily basis. Not everyone gets approved. Most of the banks use their own credit scoring and risk assessment techniques in order to analyze the loan application and to make decisions on credit approval. In spite of this, there are many cases happening every year, where people do not repay the loan amounts or they default, due to which these financial institutions suffer huge amount of losses.

**PROPOSED:**

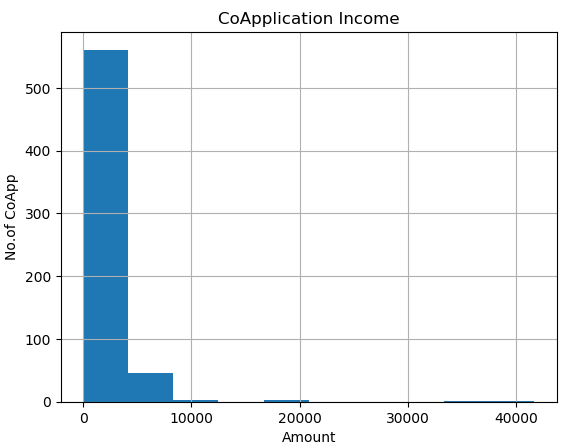
The primary goal of this project is to extract patterns from a common loan approved dataset, and then build a model based on these extracted patterns, in order to predict the likely loan defaulters by using classification data mining algorithms. The historical data of the customers like their age, income, loan amount, employment length etc. will be used in order to do the analysis. Later on, some analysis will also be done to find the most relevant attributes, i.e. The factors that affect the prediction result the most. Using different type of Machin Learning algorithm and Predicting an Accuracy Result and Plotting a graph.

**Exploratory Data Analysis (EDA)**

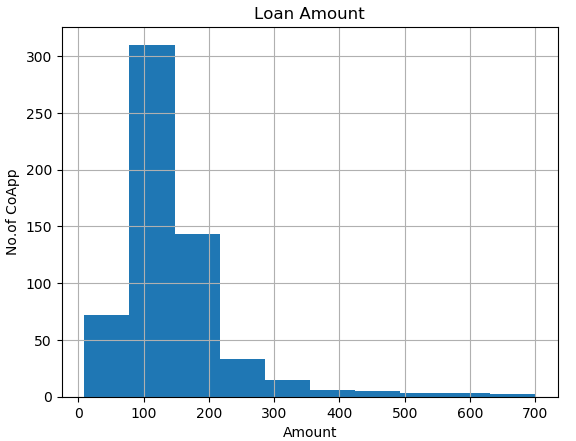
1. **Application Income**



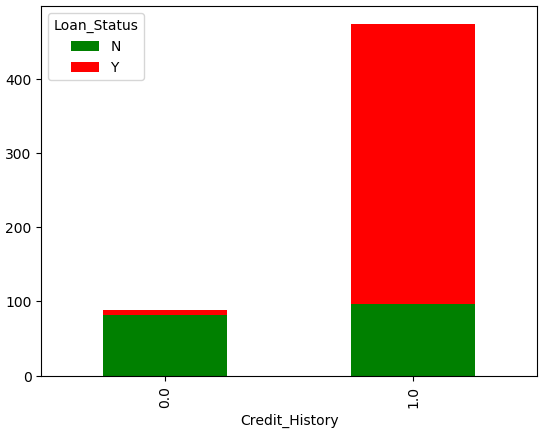
1. **Co-Application Income**



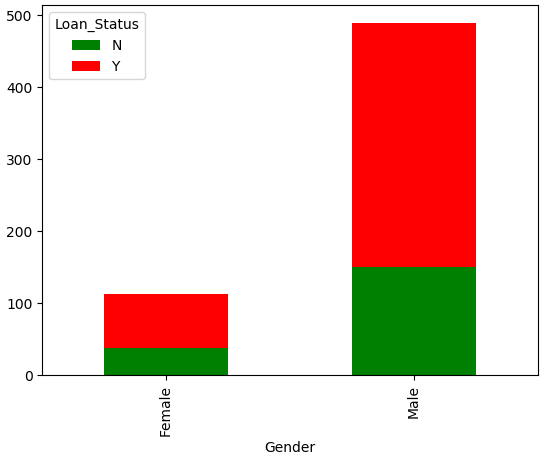
1. **Loan Amount**



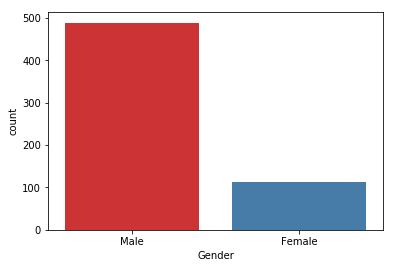
1. **Credit History**



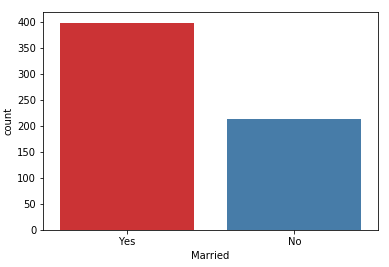
1. **Gender Loan Status**



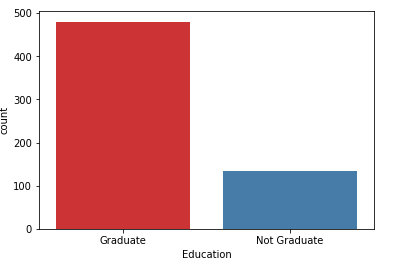
1. **Graph for Gender**



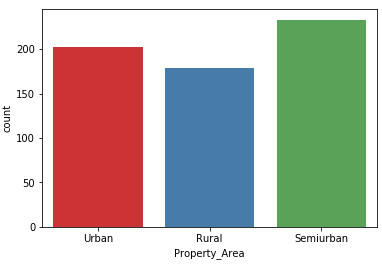
1. **Graph for Relation Status**



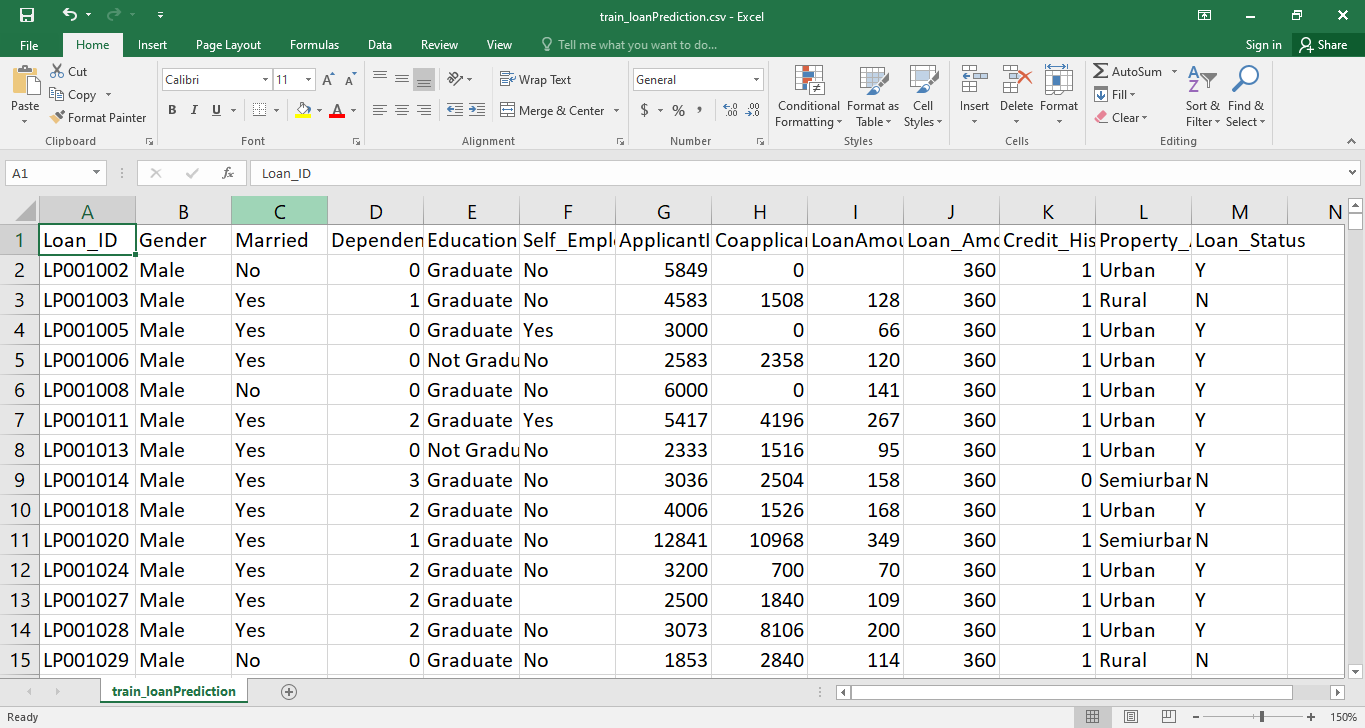
1. **Graph for Graduate and Not graduate**



1. **Graph for Rural or Urban or Semi urban**

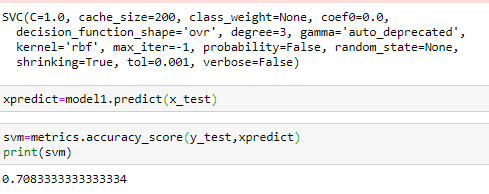


**ANALYZE AND PREPARE THE DATASET**

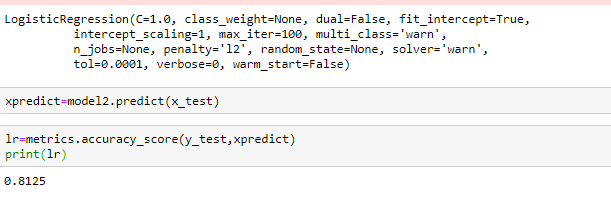


**ALGORITHMS IMPLEMENTATION**

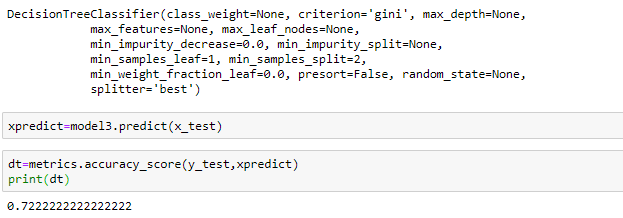
**1.SVM**



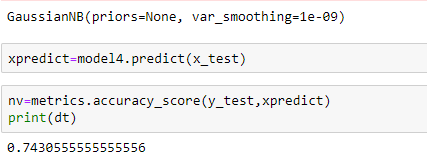
### **2.** Logistic Regression



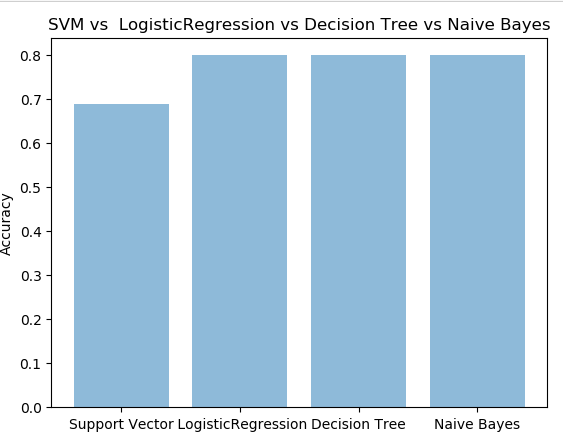
3.Decision Tree



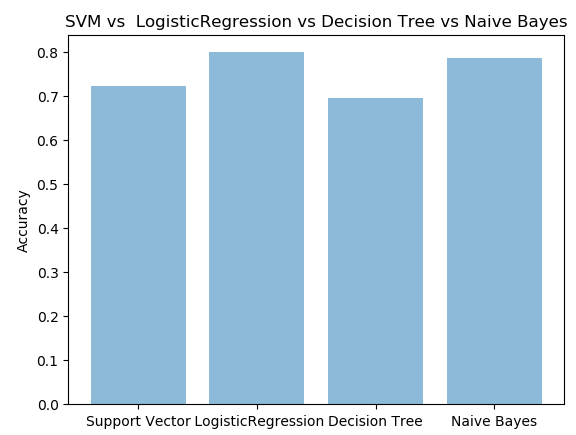
4. Naive Bayes



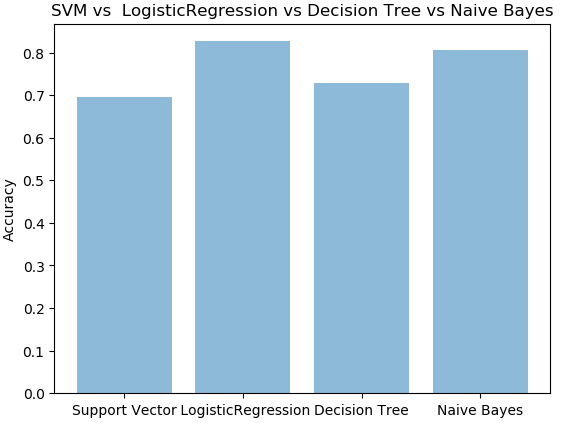
* **PREDICT THE RESULT RUN 1**



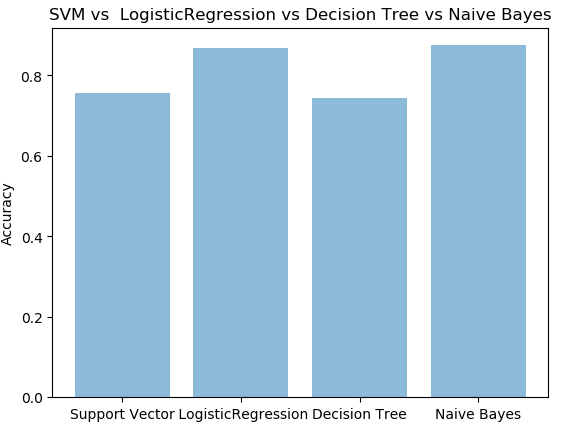
* **PREDICT THE RESULT RUN 2**



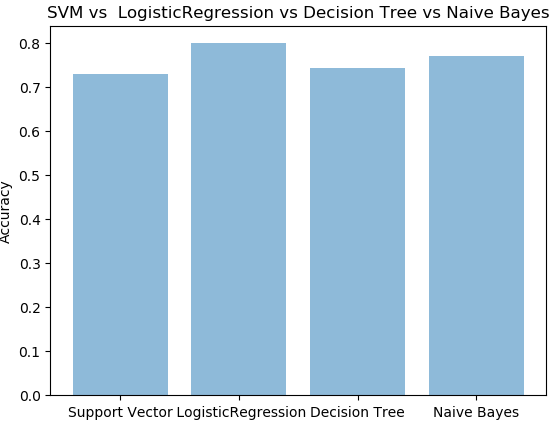
* **PREDICT THE RESULT RUN 3**



* **PREDICT THE RESULT RUN 4**



* **PREDICT THE RESULT RUN 5**



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Models / No. of runs** | **RUN1** | **RUN2** | **RUN3** | **RUN4** | **RUN5** |
| **Support Vector** | **0.68** | **0.72** | **0.69** | **0.75** | **0.72** |
| **Logistic Regression** | **0.79** | **0.79** | **0.82** | **0.86** | **0.79** |
| **Decision Tree** | **0.69** | **0.69** | **0.72** | **0.74** | **0.74** |
| **Naive Bayes** | **0.80** | **0.69** | **0.72** | **0.74** | **0.74** |

**BEST ALGORITHM: Logistic Regression**