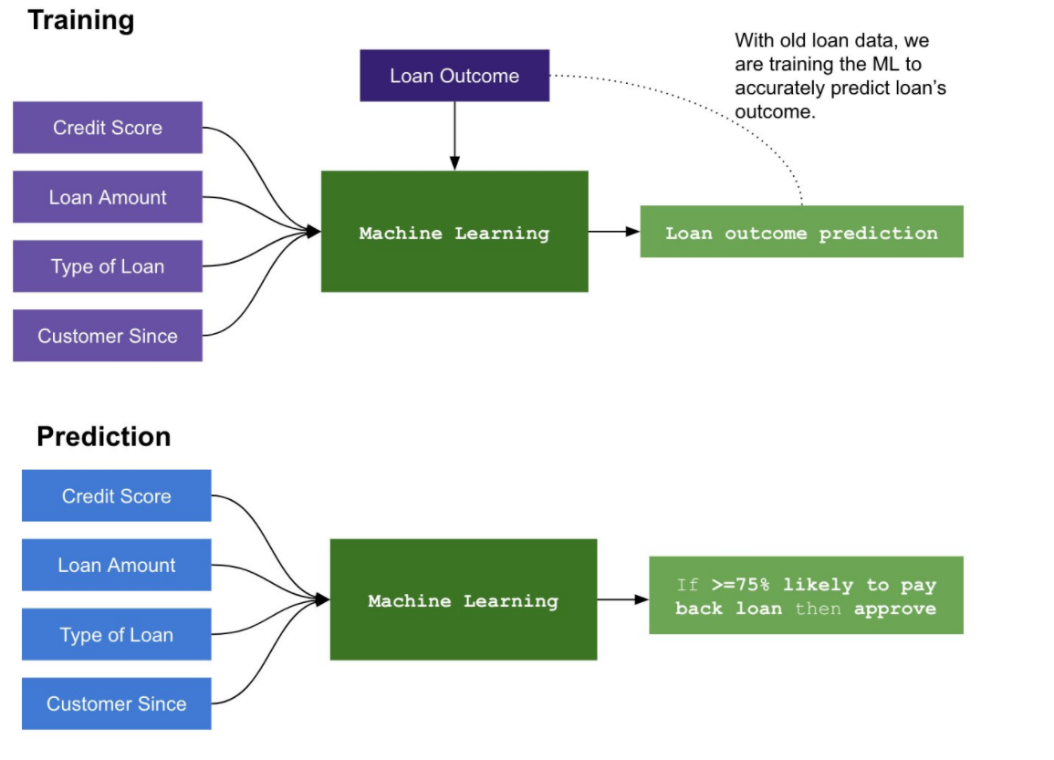


**Abstract: -** Loan taking from bank and financial institute has become a very common phenomenon. Everyday a large number of people are applied and make application for loan. But all application of loan is not approved by bank because most of people are not eligible for loan it depends on people credit score and transections history to bank or any other financial company. Every year we have seen about a large of cases in which people do not repay loan amount to the banks due to which they suffer huge losses. The bank and financial institute with making a loan approval is immense. This project is to gather loan data from multiple data sources and use various machine learning algorithms on this data to extract important information. So that the model can be used by the organisations in making the right decision to approve or reject the loan request of the customers. In this paper we have apply various machine algorithm like logistic (LR), Decision Tree (DT) and Random Forest (RF) are applied to predict the loan approval of customers.



1. **Introduction: -**

Almost every bank's core business is loan distribution. All most all of the bank's assets and profit are directly derived from the loans distributed by the bank. Bank’s plays a vital role in market economy and their success growth. The primary goal in the banking industry is to place their loan in safe hands. The success or failure of organization largely depends on the industry’s fundamental and ability to evaluate credit risk. Before giving the credit loan to borrowers, bank decides whether the borrower is bad (defaulter) or good (non-defaulter), means bank check his credit score and transection history. The prediction of borrower status i.e. in future borrower will be defaulter or non-defaulter is a challenging task for any organization or bank. Basically, the loan defaulter prediction is a binary classification problem Loan amount; costumers history governs his credit ability for receiving loan. The problem is to classify borrower as defaulter or non-defaulter. However, developing such a model is a very complicated task due to increasing in demands for loans. This work includes the construction of different machine learning models Credit Risk assessment is a crucial issue faced by Banks nowadays which helps them to evaluate if a loan applicant can be a defaulter or non-defaulter. The machine learning models and algorithm provide the help to predict the genuine borrower and any organisation for loan approval.

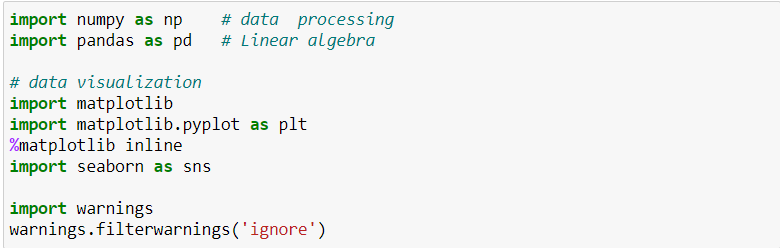
Problem definition

Here our aim is to automate the loan eligibility process on the basis of the data provided by the customer while filling up the application. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others. We have used all factor of the customer in machine learning model to predict or evaluate the perfect customer which is eligible for loan.

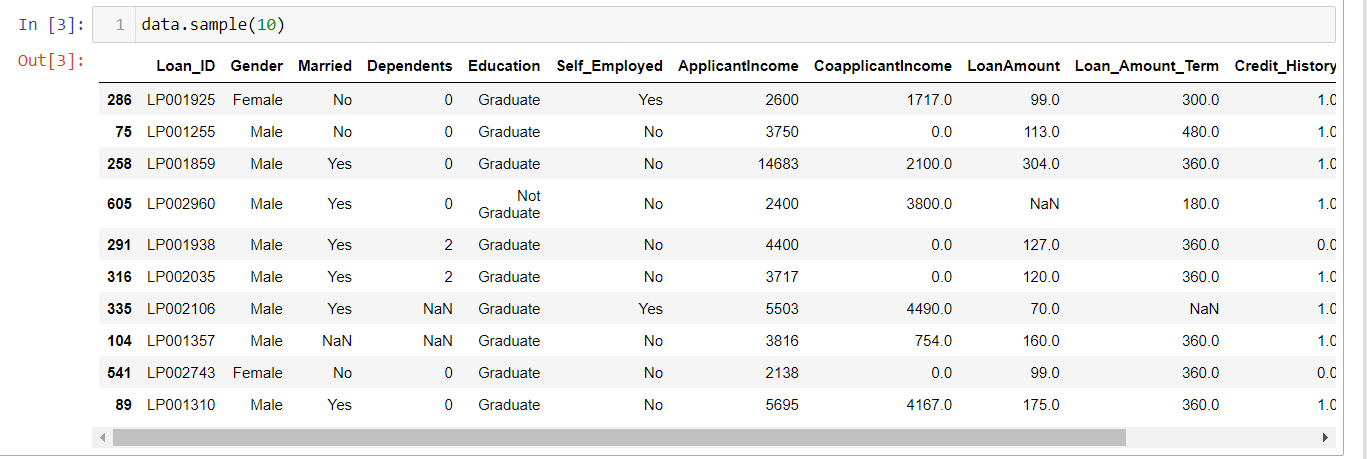
Data Analysis

In this section we will analyse and visualize our data using tools available in python for analysis like pandas’ data frame for data processing, NumPy for linear algebra and Matplotlib, Seaborn for visualization. Let’s start with analysis by import all the require library for data analysis.

Importing all the required libraries:



Now our data set is loaded into our data variable, let’s check the ten sample for our data.



Dataset have total 614 entries.  
It is having 614 Rows and Total 13 columns.

### **Dataset Description:**

Loan ID :- Unique Loan ID  
Gender :- Male/ Female  
Married :- Applicant married (Y/N)  
Dependents :- Number of dependents  
Education :- Applicant Education (Graduate/ Under Graduate)  
Self Employed :- Self employed (Y/N)  
Applicant Income :- Applicant income  
Co-applicant Income :- Co-applicant income  
Loan Amount :- Loan amount in thousands  
Loan Amount Term :- Term of loan in months  
Credit History :- credit history meets guidelines  
Property Area :- Urban/ Semi Urban/ Rural  
Loan Status :- Loan approved (Y/N)

The above all factor we have used in machine learning model to predict the eligibility of customer for loan. From sample we can see we have columns like Gender, Married, Education, Applicant income, Co-Applicant income, Loan Amount and Loan Status etc. Loan Status is our target column and all other columns are feature columns. We have only two types of values in Loan Status ‘Yes’ and ‘No’, which indicates it is classification problem. We can see Loan ID is unique identification number which we can drop during analysis.

We can see that in below columns

