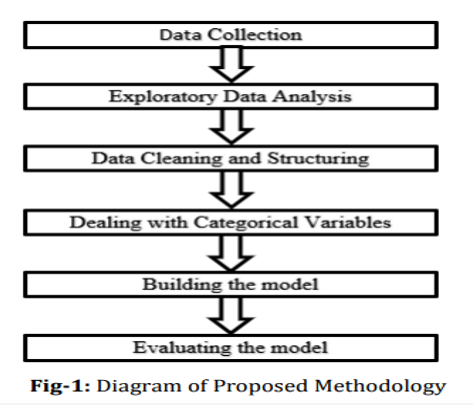
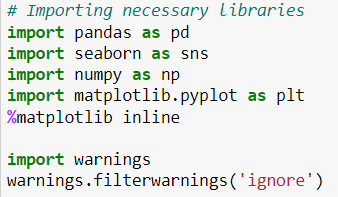
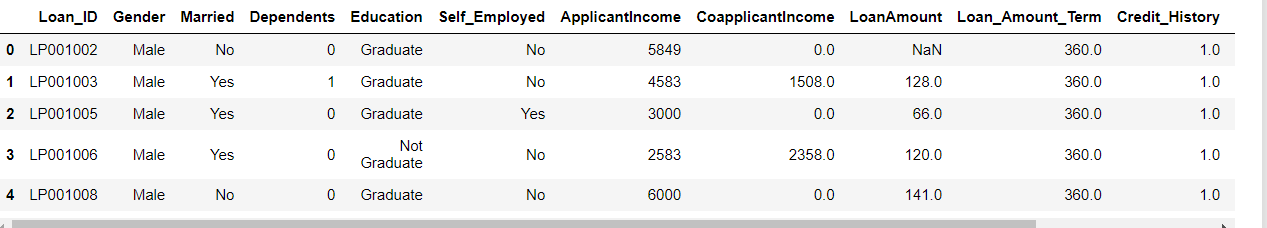
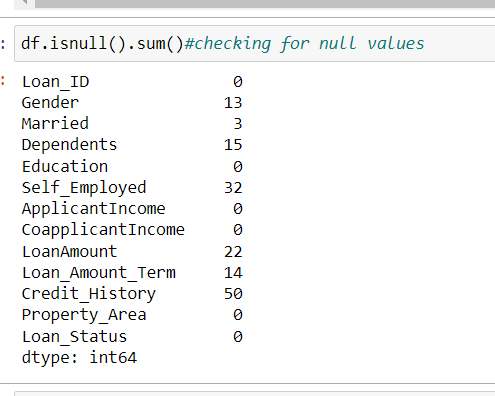
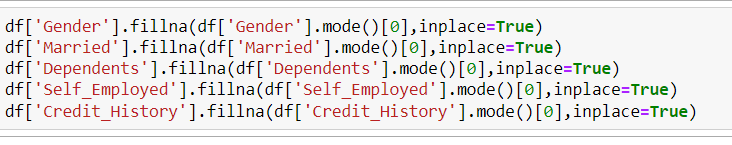
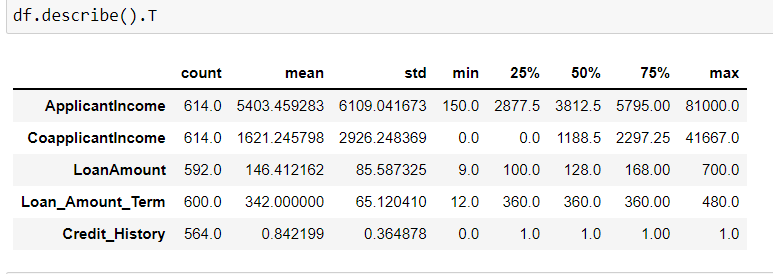
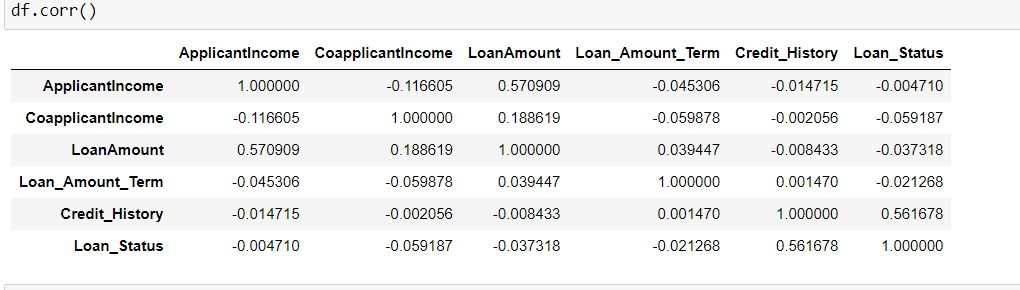
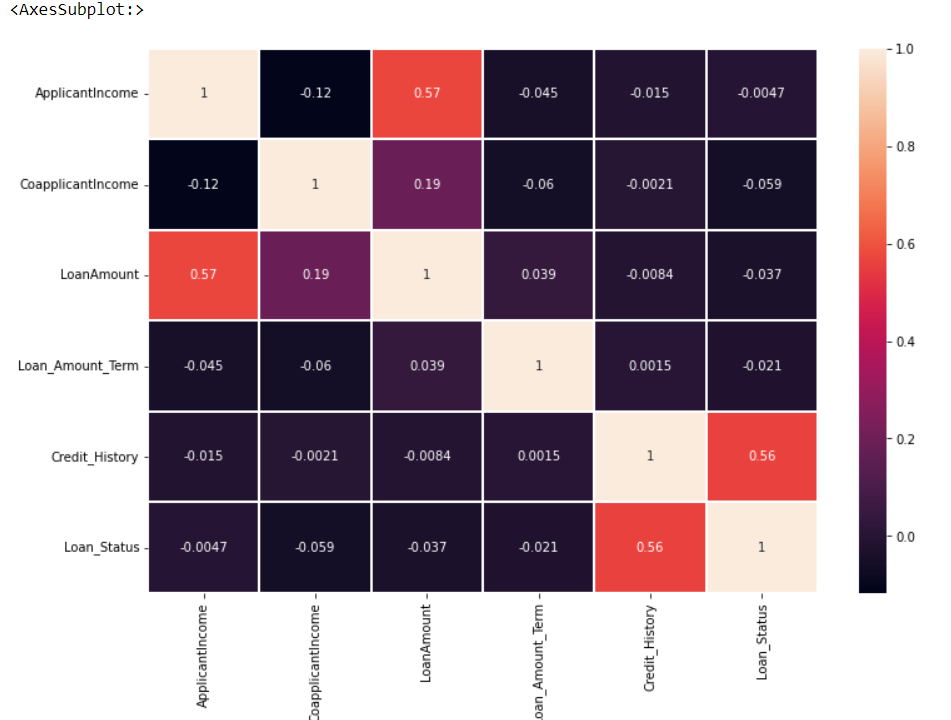
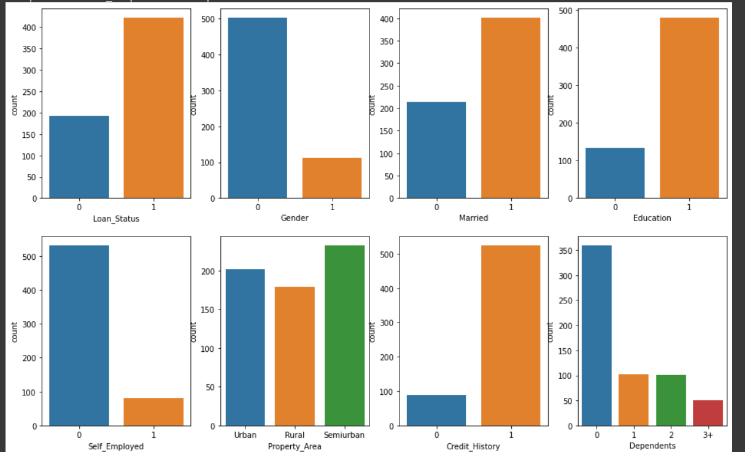
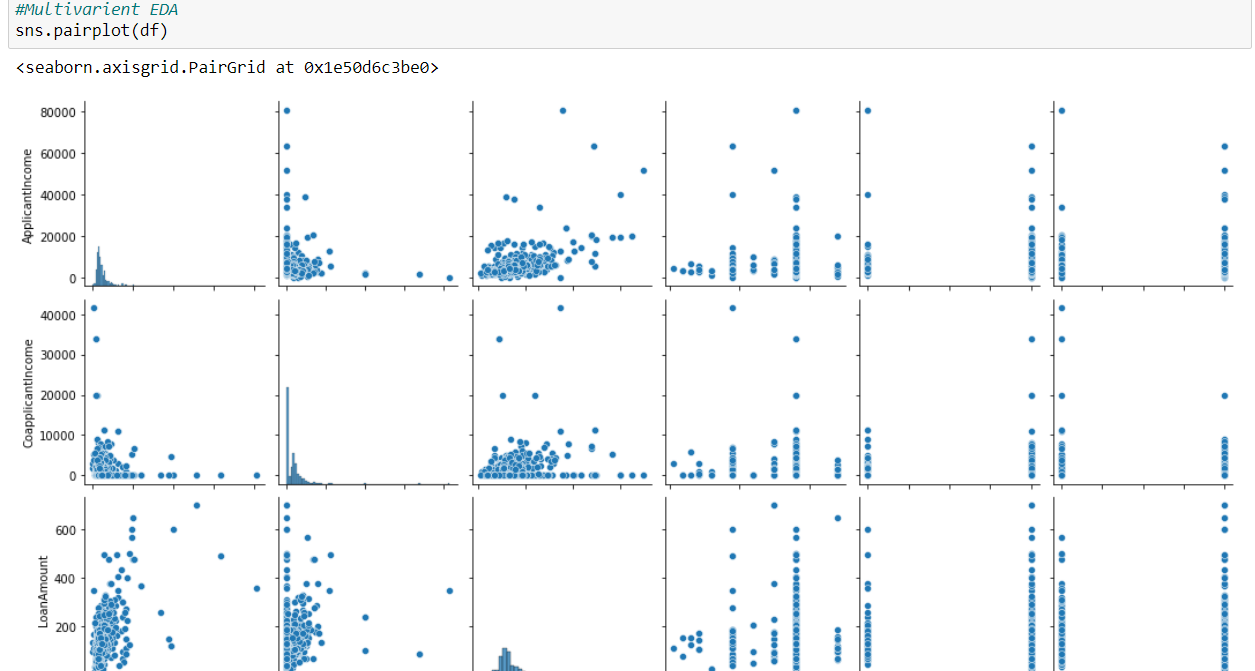
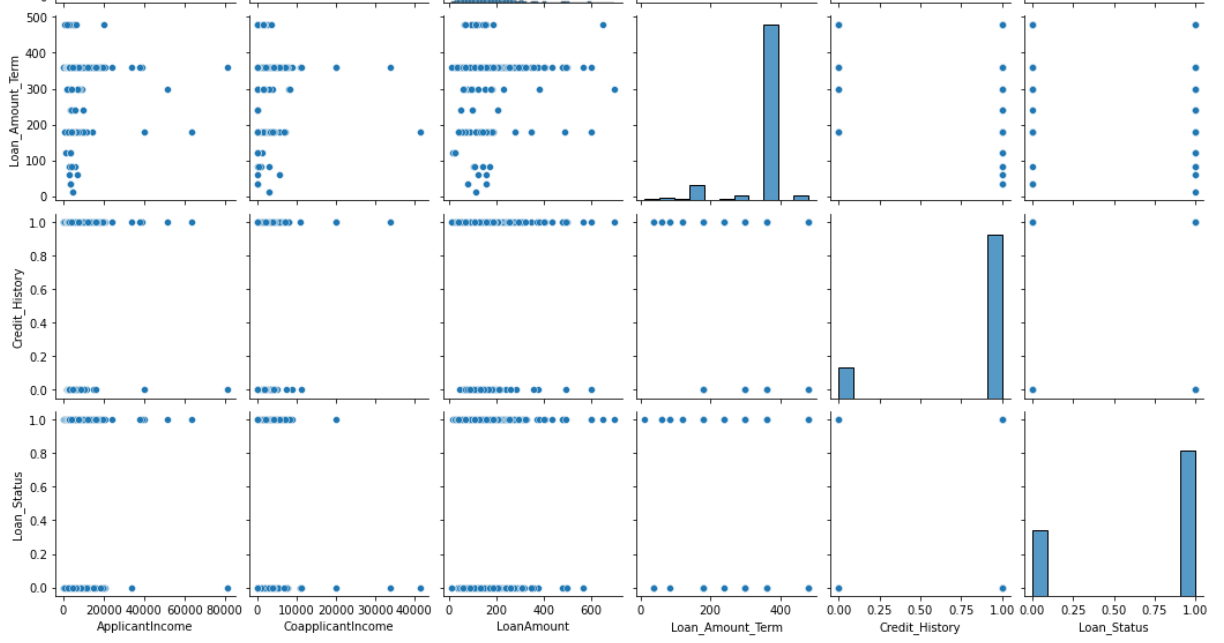
* + Loan Approval Prediction based on Machine Learning Approach
  + **1.Problem Definition.**
  + To design and implement the system using machine learning and data mining to predict the probability of the user to get loan. Banks, Housing Finance Companies and some NBFC deal in various types of loans like housing loan, personal loan, business loan etc in all over the part of countries. After applying loan by customer any financial institute validates the eligibility of customers to get the loan or not...
  + Here is the dataset includes details of applicants who have applied for loan. The dataset includes details like credit history, loan amount, their income, dependents etc. on the basis of the available details any institution will make decision to approve the loan application or not.  To avoid fraud or default the correct analysis of the Applicants Background helps the Bank or any financial institution.
  + The details of the Application are as follows:
  + **Independent Variables:**
  + Loan ID
  + Gender
  + Married
  + Dependents
  + Education
  + Self-employed
  + Applicant Income
  + Co applicant Income
  + Loan Amount
  + Loan Amount Term
  + Credit History
  + Property Area
  + **Dependent Variable (Target Variable):**
  + Loan Status (approved or not)
  + Here we build a model that can predict whether the loan of the applicant will be approved or not on the basis of the details provided in the dataset.
  + **2.Data Analysis.**
  + To deal with the problem, we developed automatic loan prediction using machine learning techniques. We will train the machine with previous dataset. so, machine can analyze and understand the process. Then machine will check for eligible applicant and give us result.
  + **Steps involved to build a suitable model.**
  + 
  + The purpose of the competition is to identify loan eligibility based on the
  + mentioned data. The result of the project was uploaded in order to score
  + the predictions, which was 0.7986, which is a good result
  + **Load Essential Python Libraries**
  + 

## First look at the Dataset

* + 
  + **Shape of the DataSet**
  + The dataset has 614 rows and 13 cloumns , included Target Column.
  + **3.Data Pre-processing**
  + Identify missing values :
  + 
  + **Imputing the missing values:**
  + 
  + **Filled null values with mode**
  + Statistical summary of data
  + 
  + Correlation between the dataset
  + 
  + 
  + **4.Exploratory Data Analysis**
  + Univariate Analysis:
  + 
  + 
  + 

## Bivariate Analysis

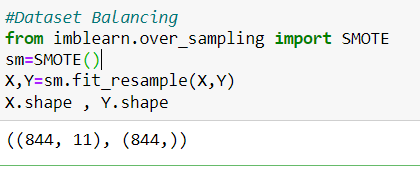
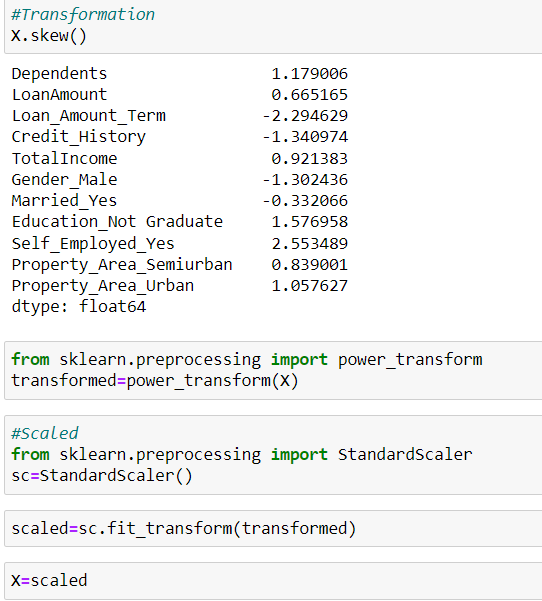
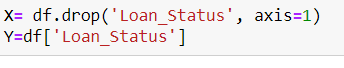
## 

## 

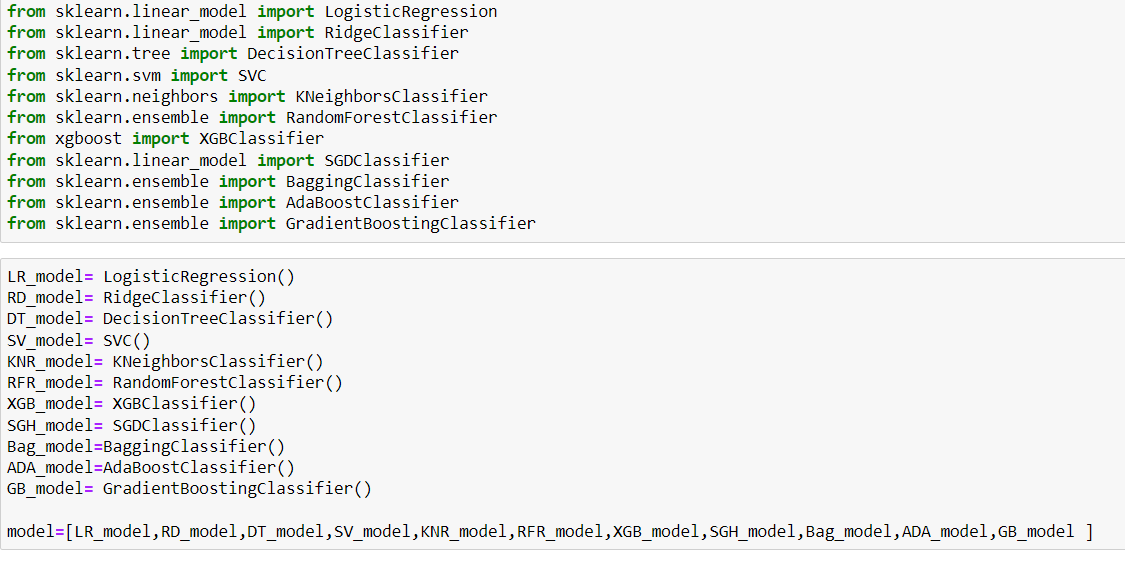
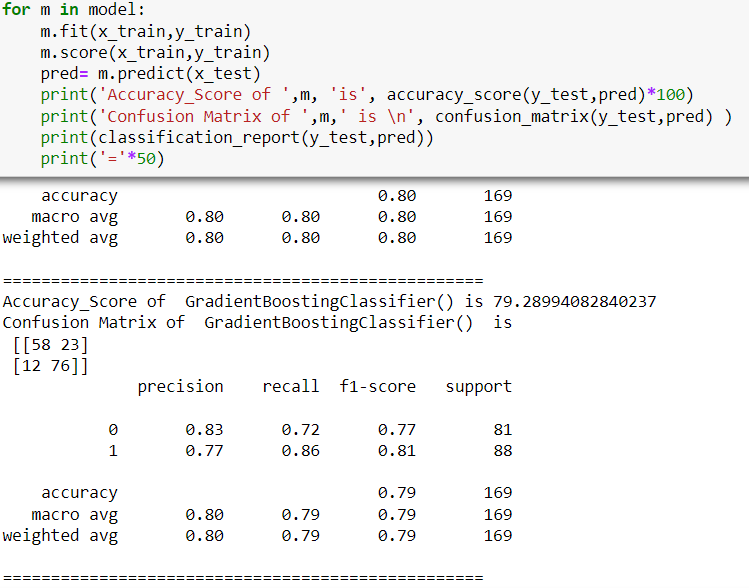
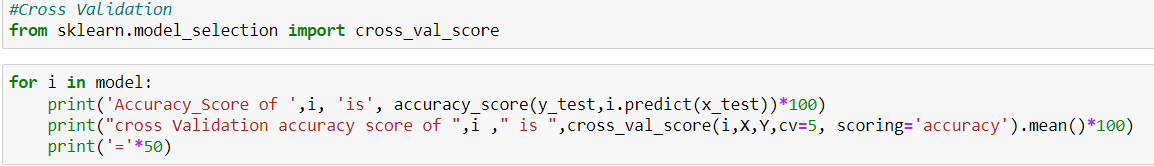
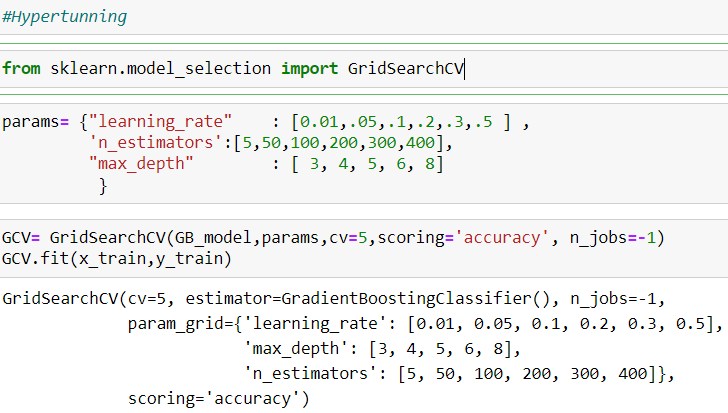
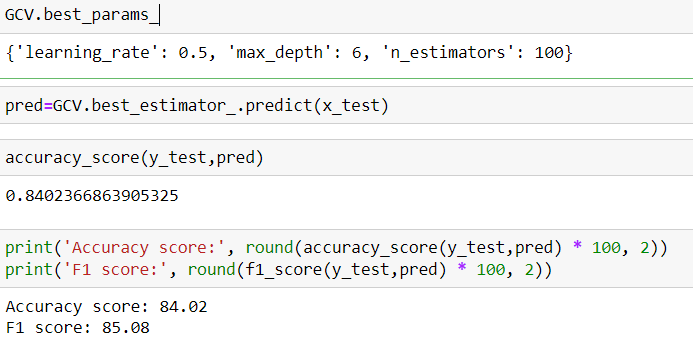
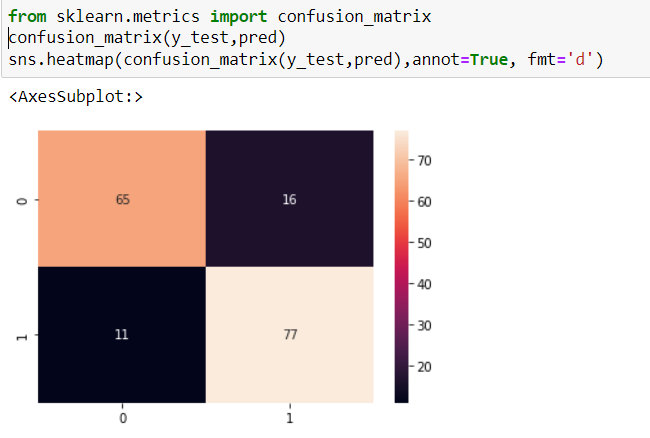
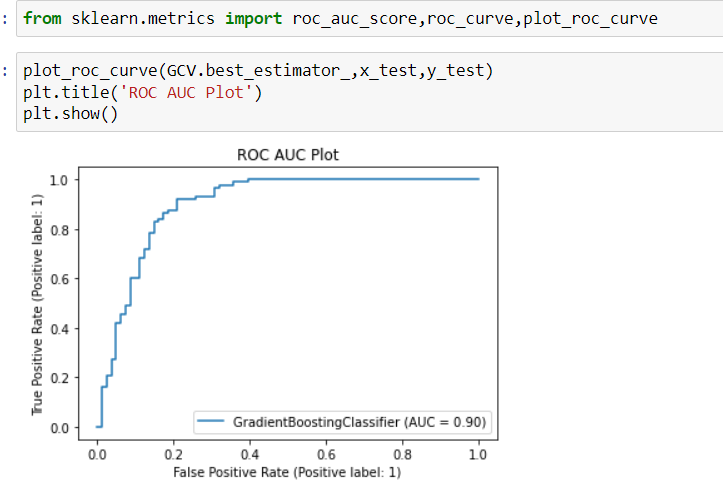
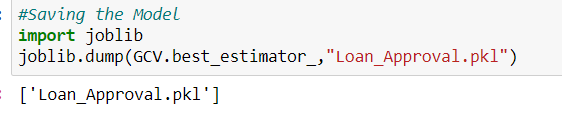
## 

## 

## 

* + This Analysis will be comparing different prediction models and deduce their limitations as well as advantages. Since all the research used different sets of data to infer the accuracy and for cross validation of data, we have used the same data for all the models which will give a clearer view on their performance and lead to a better comparison of the same. On the basis of the results, a modified prediction model will be created to ensure maximum accuracy and performance.
  + Target variable is Loan Status which has just to outcomes “Yes” an
  + “No”, and is not balanced, that is why it had been used a process specific
  + to balance the data. Also, in some cases the data has NaN values for
  + which, depending on the case, they had been imputed.
  + The strategy followed was mainly to transform the categorical variables
  + with zero a one only if there were only two possible values. In other cases,
  + the strategy was to apply one hot encoding to process them. These
  + features were processed easily with very simple transformation. The
  + categorical features processed in the mentioned way were:
  + Gender
  + Married
  + Dependents
  + Education
  + One of the most important problems found in the dataset was that the
  + target feature was unbalanced, that is why a balanced technique had
  + been used, which helped a lot in the accuracy calculation
  + Dataset Balancing
  + 
  + Data Transformation
  + 
  + Self\_Employed
  + In the case of numeric features, the story was di=erent because most of
  + them were skewed and in some cases with lots of NaN values. In this case,
  + there were no clue in the way to process them, that is why the strategy in
  + this case was a little more complicated because it has been used various
  + ways to try to get the best possible results.
  + ApplicantIncome
  + CoapplicantIncome
  + LoanAmount
  + Loan\_Amount\_Term
  + credit\_History
  + One of the most important problems found in the dataset was that the
  + target feature was unbalanced, that is why a balanced technique had
  + been used, which helped a lot in the accuracy calculation
  + **5. Building Machine Learning Models**.
  + Creating X (input variables) and Y (Target Variable) from the dataset
  + 
  + Using train test split on the training data for validation
  + 
  + Find Best Random state at 130.

## Using ML algorithm for training

* + 
  + 
  + Cross Validation
  + 
  + 
  + Find the Best Accuracy Model
  + 
  + Confusion Matrics
  + 
  + 
  + Saving the BEST Model
  + 
  + 6. Concluding Remarks.
  + After the Final Submission of test data, my accuracy score was 79.28%.
  + Amazingly GradientBoostingClassifier worked better than all other models.
  + 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. There have been numbers cases of computer glitches, errors in content and most important weight of features is fixed in automated prediction system, So in the near future the so –called software could be made more secure, reliable and dynamic weight adjustment. In near future this module of prediction can be integrate with the module of automated processing system. the system is trained on old training dataset in future software can be made such that new testing date should also take part in training data after some fix time.