**Loan Prediction Hackathon**

**Summary:**

|  |  |
| --- | --- |
| **Training Dataset** | 480 |
| **Testing Dataset** | 289 |
| **Unseen Dataset** | 289 |
| **Train Test Split** | 60:40 |
| **Model Selection** | KNeighborsClassifier(), LogisticRegression(), RandomForestClassifier(), DecisionTreeClassifier(), SVC(), GradientBoostingClassifier(), XGBClassifier() |
| **Model Finalized** | GradientBoostingClassifier |
| **Parameter Used** | n\_estimators, random\_state |

**Problem Statement:**

Dream Housing Finance company deals in all kinds of home loans. They have presence across all urban, semi urban and rural areas. Customer first applies for home loan and after that company validates the customer eligibility for loan.

Company wants to automate the loan eligibility process (real time) based on customer detail provided while filling online application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others. To automate this process, they have provided a dataset to identify the customers segments that are eligible for loan amount so that they can specifically target these customers.

**Data Dictionary**

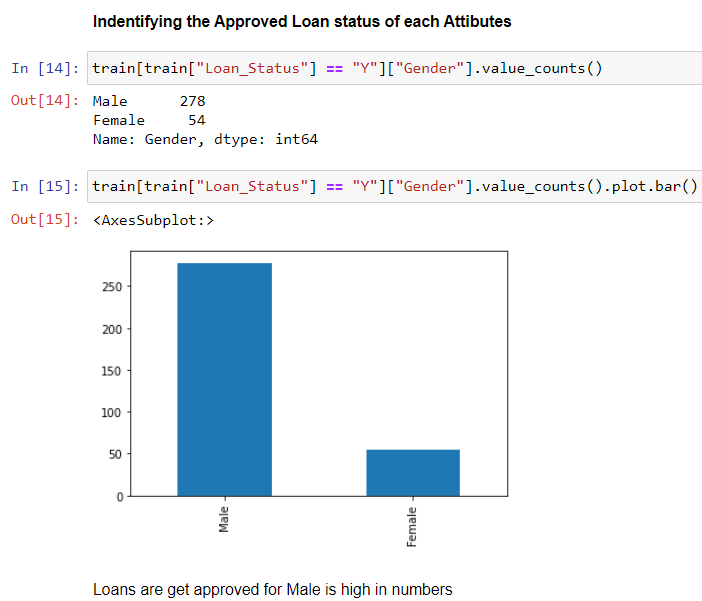
**Train file:**CSVcontaining the customers for whom loan eligibility is known as 'Loan\_Status'

|  |  |
| --- | --- |
| **Variable** | **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) |
| ApplicantIncome | Applicant income |
| CoapplicantIncome | Coapplicant income |
| LoanAmount | 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 | (Target) Loan approved (Y/N) |

**Test file:** CSVcontaining the customer information for whom loan eligibility is to be predicted

|  |  |
| --- | --- |
| **Variable** | **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) |
| ApplicantIncome | Applicant income |
| CoapplicantIncome | Coapplicant income |
| LoanAmount | Loan amount in thousands |
| Loan\_Amount\_Term | Term of loan in months |
| Credit\_History | credit history meets guidelines |
| Property\_Area | Urban/ Semi Urban/ Rural |

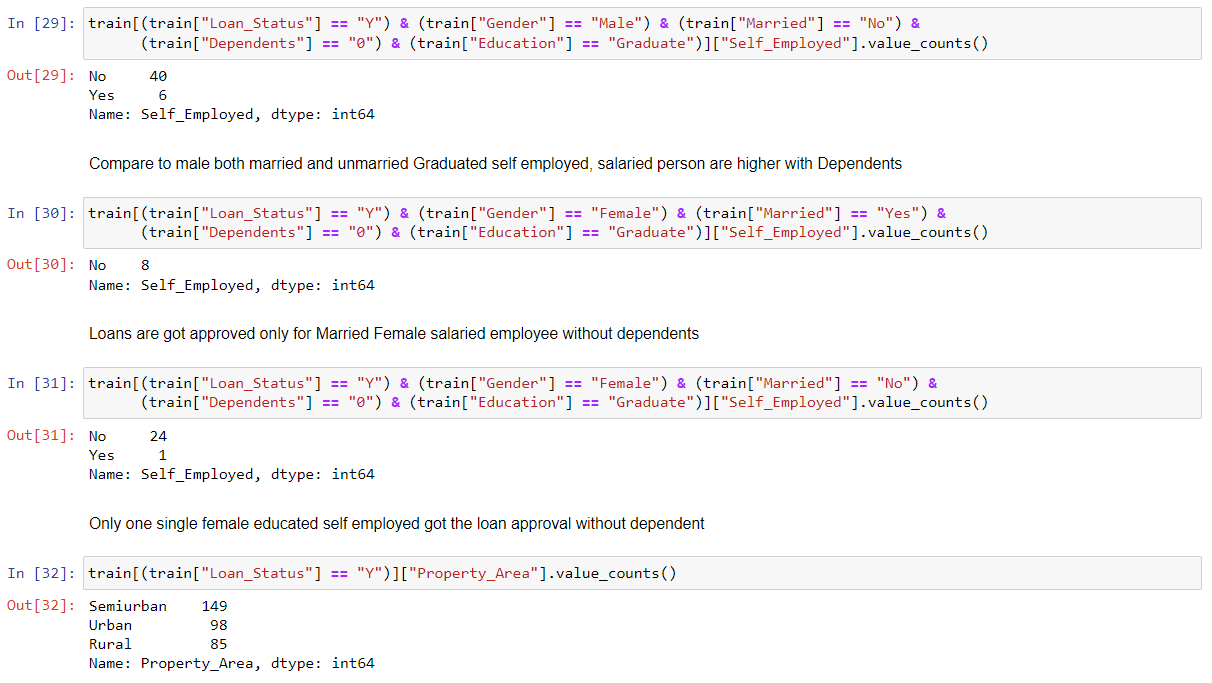
**EDA**

****

****

****

****

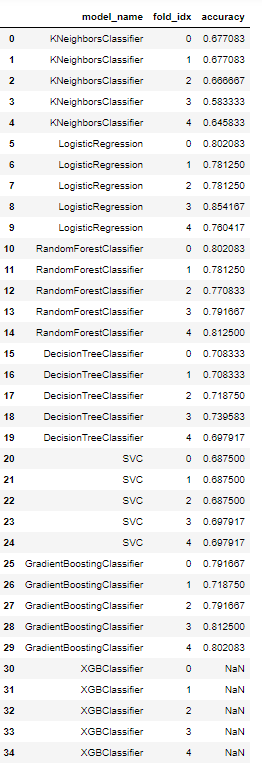
****

**Observation:**

* Loans are get approved for Male is high in numbers
* Married males are higher in number and unmarried Females are higher in number
* As we look loans got approved for without dependents is higher compare to with dependents’ for both married males and Females
* Even for unmarried without dependent is higher than with dependent for both male and Female
* Loan got approved for both Married and Unmarried Males and Female graduate is higher than non-graduate for non-dependent
* Approval for loans is higher for Graduated Married males with the Dependent
* Only for the Graduate Married women are getting Loans with minimal number
* Compare to male both married and unmarried Graduated self-employed, salaried person are higher with Dependents
* Loans are got approved only for Married Female salaried employee without dependents
* Only one single female educated self-employed got the loan approval without dependent
* There are more Outlier on each of the match cases observed

**Model Selection**

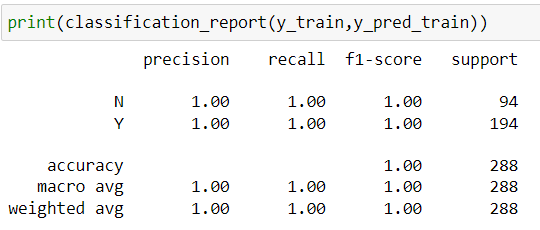
From the below screen shot GradientBoostingClassifier performs well compare to other models. Hence we can choose this model for the final building



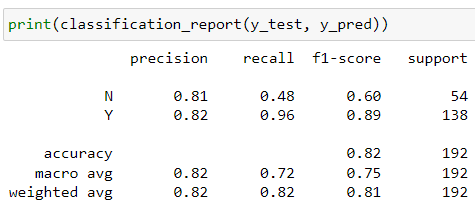
**Models**

**Classification Report:**

**Training Data:**

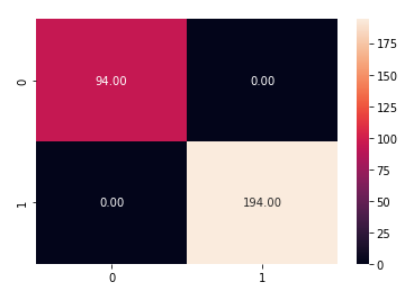
****

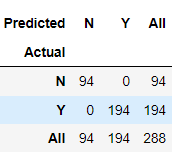
**Test Data:**

****

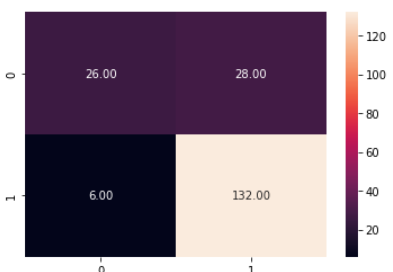
**Confusion Matrix**

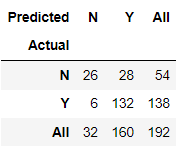
**Training data**

****

****

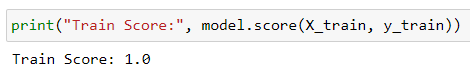
**Test data**

****

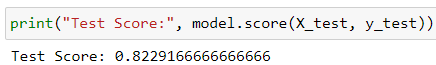
****

**Accuracy Metrics:**

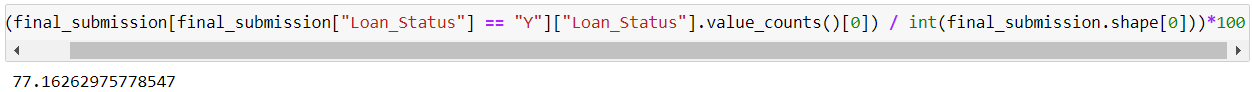
**Train Data:**

****

**Test Data:**

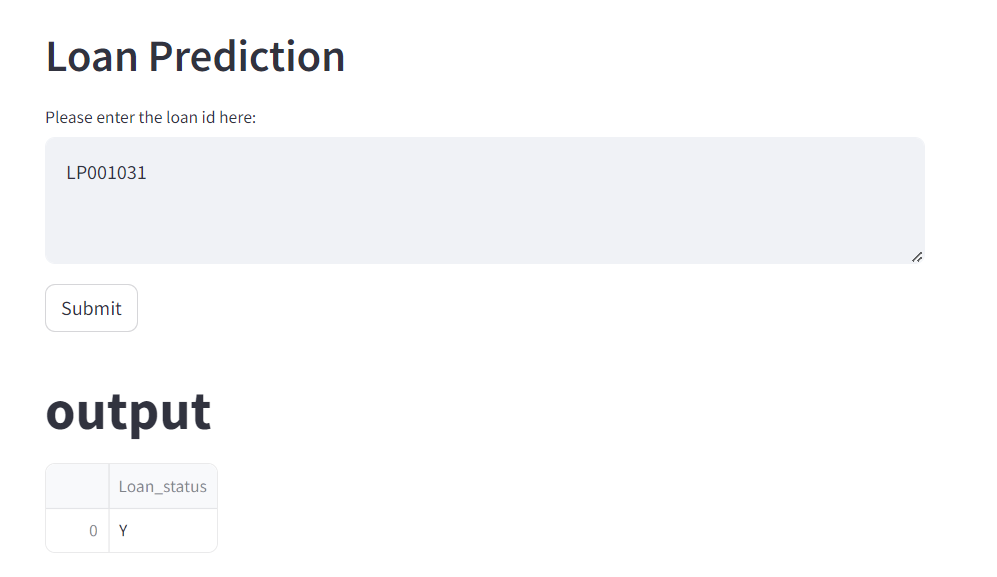
****

**Validation Data:**

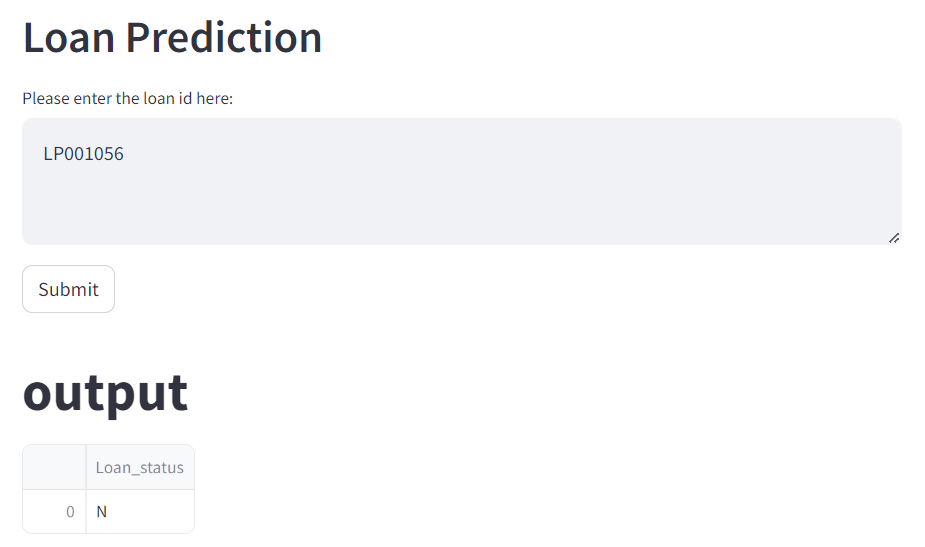
****

**Final Solution for Loan Prediction via Stream lit UI**

**Prediction 1**

****

**Prediction 2**

****