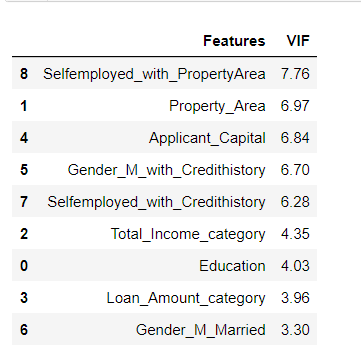
# Loan Eligibility Prediction: Consumer Creditworthiness Analysis

Milestone 4

# Feature Selection:

1. After the feature engineering the relevant independent features are selected as those which have association with the target variable (Hypothesis tests).
2. These set of features are checked to see if they have multicollinearity between them (VIF). The threshold is set 10 and the features which have VIF <10 are only selected.

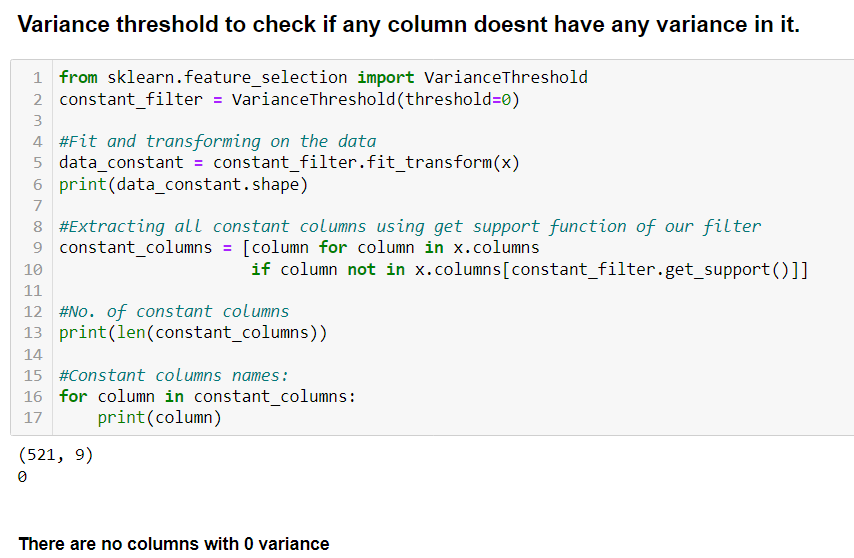
Output:



1. Variance Threshold:

All the columns are further checked to see if they have zero variance.

Output:



# Train and test split:

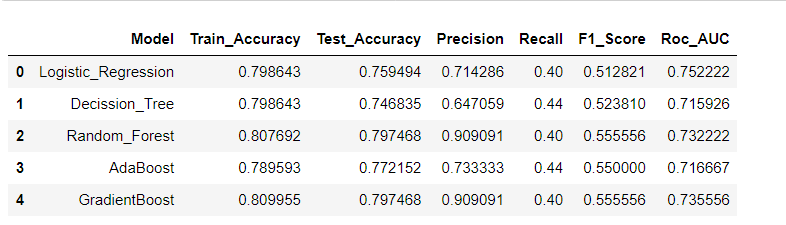
1. The data is split using train and test split with testing size as 15% and using stratified split.

# Classification using various classifiers:

# 1.The below mentioned classifiers are used for training and prediction.

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. Adaptive Boosting
5. Gradient Boost
6. These classifiers are analyzed using various performance metrics like
7. Accuracy
8. Precision
9. Recall
10. F1score
11. ROC\_AUC\_score

# Summary:



# Top 3 Models:

1. From the summary we can choose the top 3 models considering the f1score,roc\_auc and accuracy
2. Gradient Boost has similar performance as Random Forest
3. Adaptive Boost
4. Logistic Regression

# Visualising the performance of the models using the ROC\_AUC and PR curves

