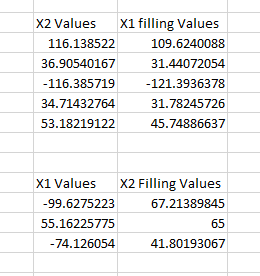
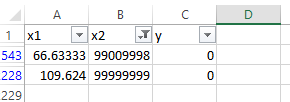
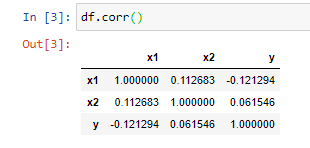
1. Before Loading the data in Jupiter, I filled the empty cells in excel. I found that there are almost 5 empty cells in X1. So, instead of removing or aggregating those rows, I found their relative terms in cell X2 and Filled it with that number using VLOOKUP function in excel, this led me take upon nearest possible values that can be filled for missing values. I repeated the same process for X2 missing variables.



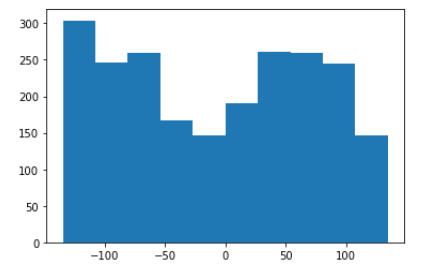
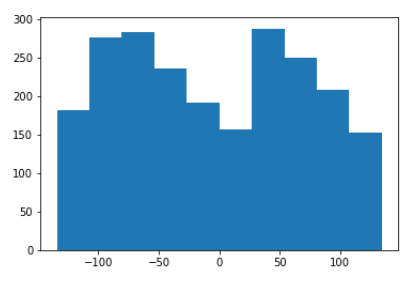
I also Found that there are some outliers in the data. So, modified them also based on the possible numbers. Ex: - 99009998 is changed to 99. Since, all the remaining values also, fall between that range only.



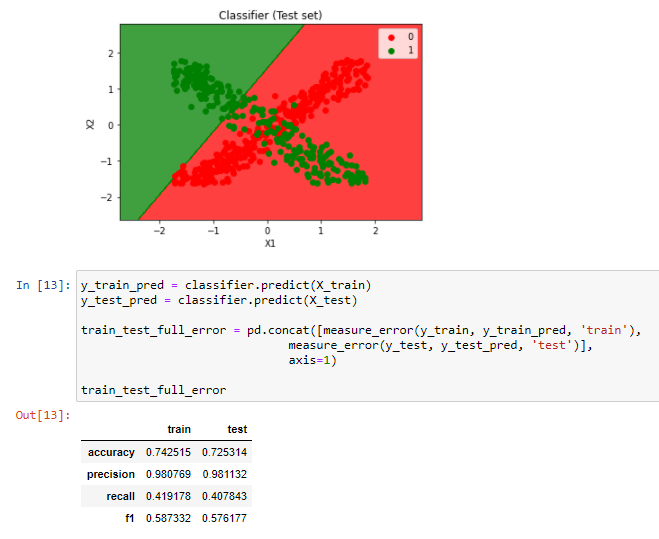
1. There is a negative correlation between Y (the predicted feature) and the independent features (X1, X2). But the accuracy may not be affected since, there is not more dependent variables present.



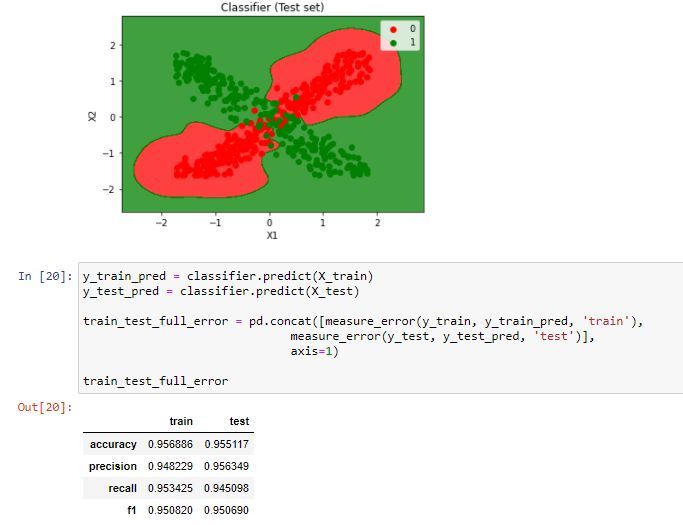
1. The values both the independent variables seem to be in same range. But better we can normalize them for more accuracy.



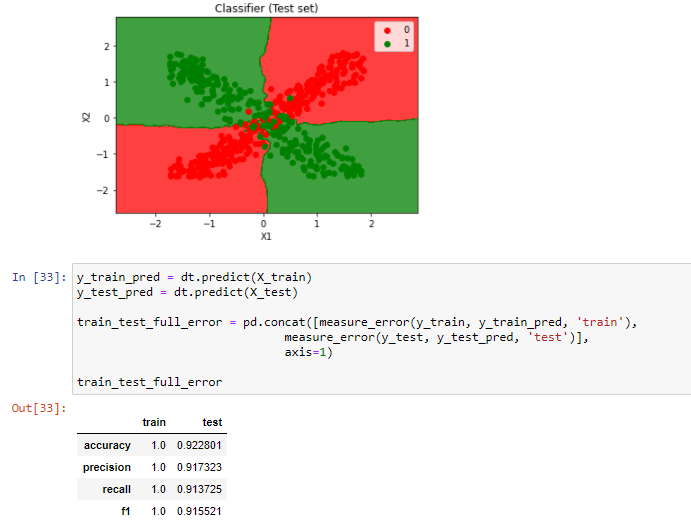
1. Logistic Regression with best model accuracy.



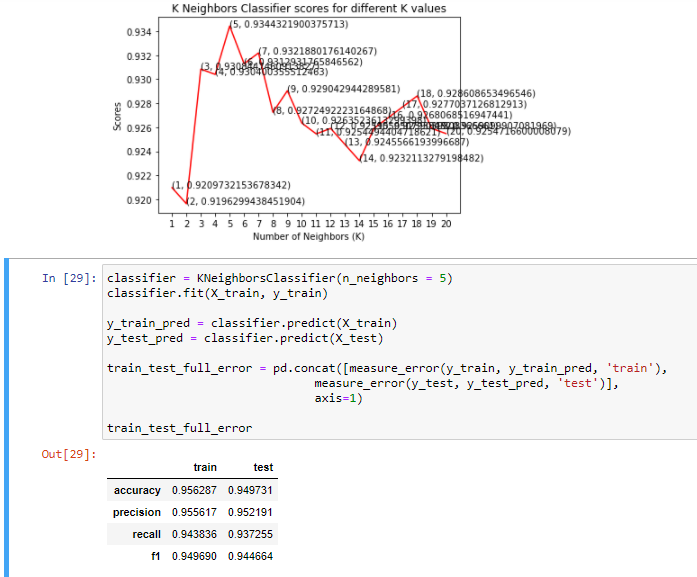
1. SVM with best model (rbf) and accuracy



1. Decision Tree with best model accuracy



1. KNN best model and accuracy



***Observation***

***We can see that we have different accuracies for different models. I would like to take SVM rbf kernel for my model building because this model is not overfitting like decision tree and also giving a max accuracy on the test set compared to remaining models.***

* **SAI GOWTHAM BABU AMBURI**