Data Set Explanation

The Balance Scale Weight and Distance Database has 625 numbers of instances. The datasets is divided into two parts:

- Training Datasets ~ 80%
- Testing Datasets ~ 20%

This data set was generated to model psychological experimental results.

It is a multivariate (multiple variables) data consisting of 625 instances i.e.
- 49 Balanced
- 288 Left
- 288 Right

 And have 5 attributes (4 Numeric and 1 Class Name).

Attribute Information:

- Class Name: 3   (L,B,R)
- Left-Weight: 5   (1-5)
- Left-Distance: 5   (1-5)
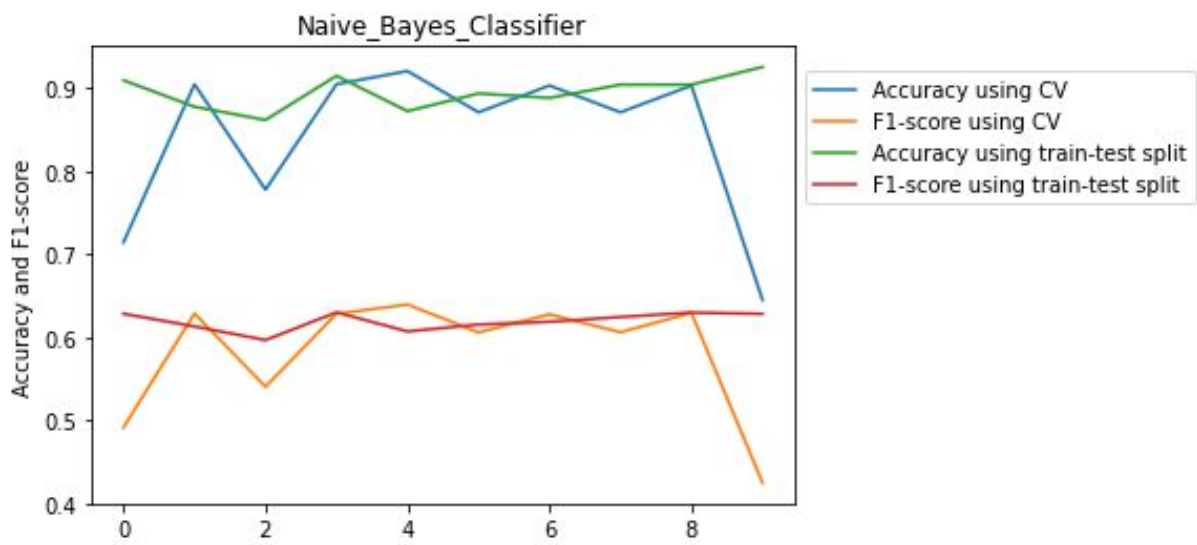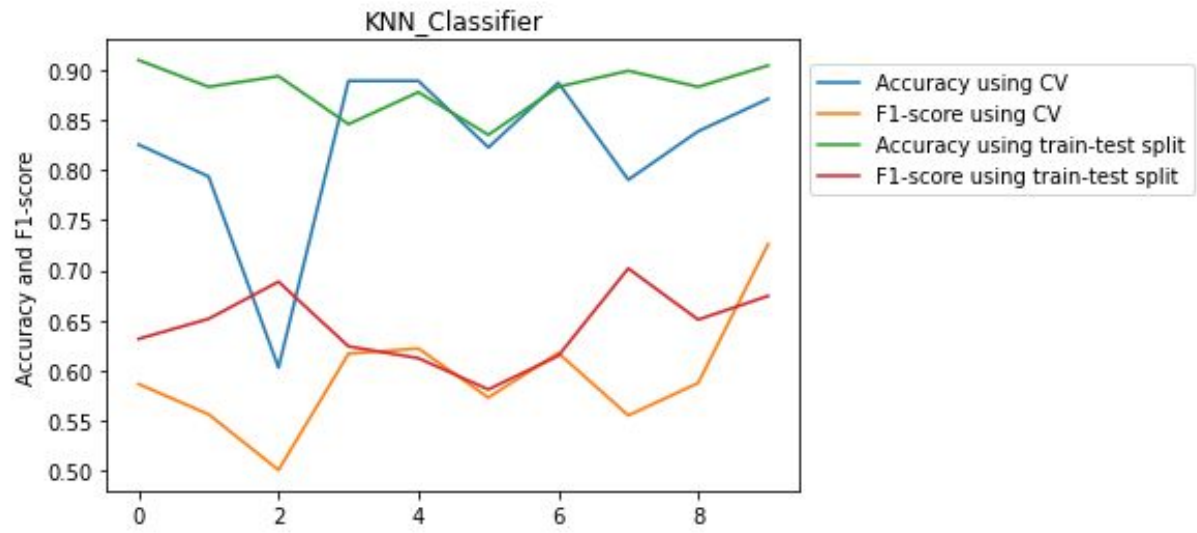- Right-Weight: 5   (1-5)
- Right-Distance: 5   (1-5)

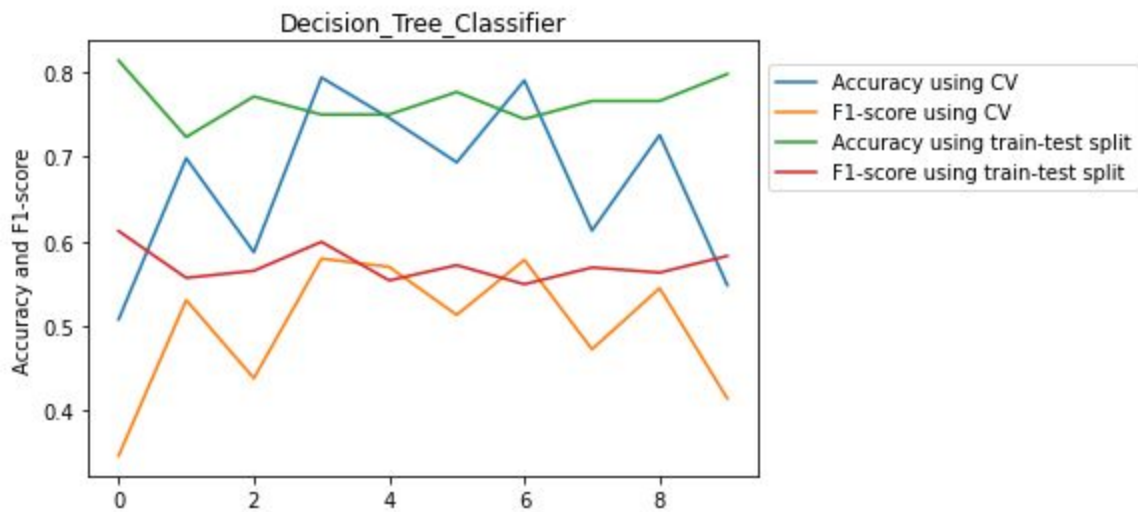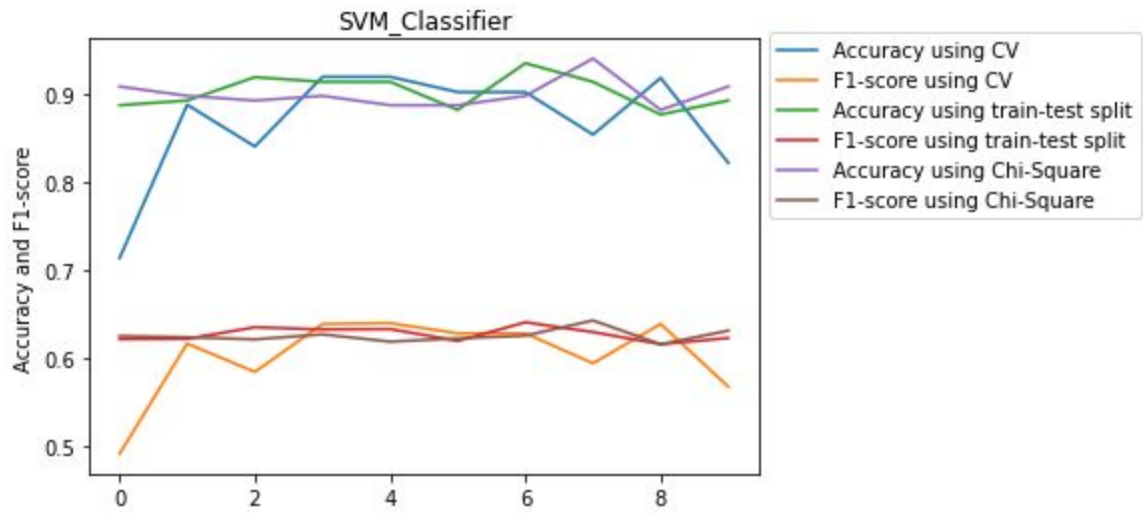The way to find the class is to selecting the greater value from

max((Left-Distance * Left-Weight), (Right-Distance * Right-Weight))

If both are equal, then it is considered to be balanced.

Class Distribution:

- L ~ 46.08%,
- B ~ 07.84%,
- R ~ 46.08%.

KNN_Classifier

- Accuracy using CV
- F1-score using CV
- Accuracy using train-test split
- F1-score using train-test split

Naive_Bayes_Classifier

- Accuracy using CV
- F1-score using CV
- Accuracy using train-test split
- F1-score using train-test split

SVM_Classifier

Decision_Tree_Classifier

Hence by visualizing all the graphs, it is clear that accuracy is greater than F1-Score due to less balanced data.