Lab 8 Summary – Decision Trees

In this Lab we deal with Decision Trees, by applying them to our dataset, calculating their efficiency in predicting the target values, and studying how altering the training and test datasets results in changes in the model’s efficiency.

We first split the dataset into 80% training and 20% test data. We then ran a Decision Tree Classifier on the training data and used it to predict the target values. Based on the results obtained, we ran evaluation metrics on it and obtained the following :

Accuracy = 0.7377049180327869

Precision Score = 0.7419354838709677

Recall Score = 0.8394160583941606

ROC Score = 0.7292318387208898

Specificity Score 0.6190476190476191

Sensitivity Score 0.7471264367816092

Diagram

Description automatically generated with medium confidence

We then moved a few instances from training set to test set to make it 50:50. Upon running the Decision Tree Classifier on it, we obtained the following evaluation metrics :

Accuracy = 0.7302631578947368

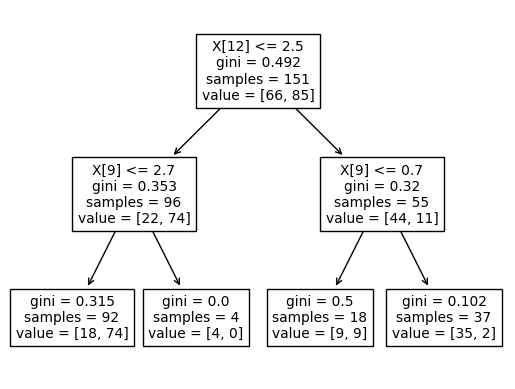
Precision Score = 0.6774193548387096

Recall Score = 0.7411764705882353

ROC Score = 0.6433155080213904

Specificity Score 0.5454545454545454

Sensitivity Score 0.6206896551724138



Lastly, we shifted more instances to the test dataset, making it 80% test and 20% training data. Upon running the Decision Tree Classifier on it, we obtained the following evaluation metrics:

Accuracy = 0.7407407407407407

Precision Score = 0.64

Recall Score = 0.5333333333333333

ROC Score = 0.6166666666666666

Specificity Score 0.7

Sensitivity Score 0.6

A picture containing timeline

Description automatically generated

Upon introspection of the values obtained, we observe that the Precision, Recall and ROC Score all decrease with a decrease in the training dataset volume. This shows that the more the data we use for training, the more efficiently the model performs.