Data Mining Assignment 2

Training and Testing Data:

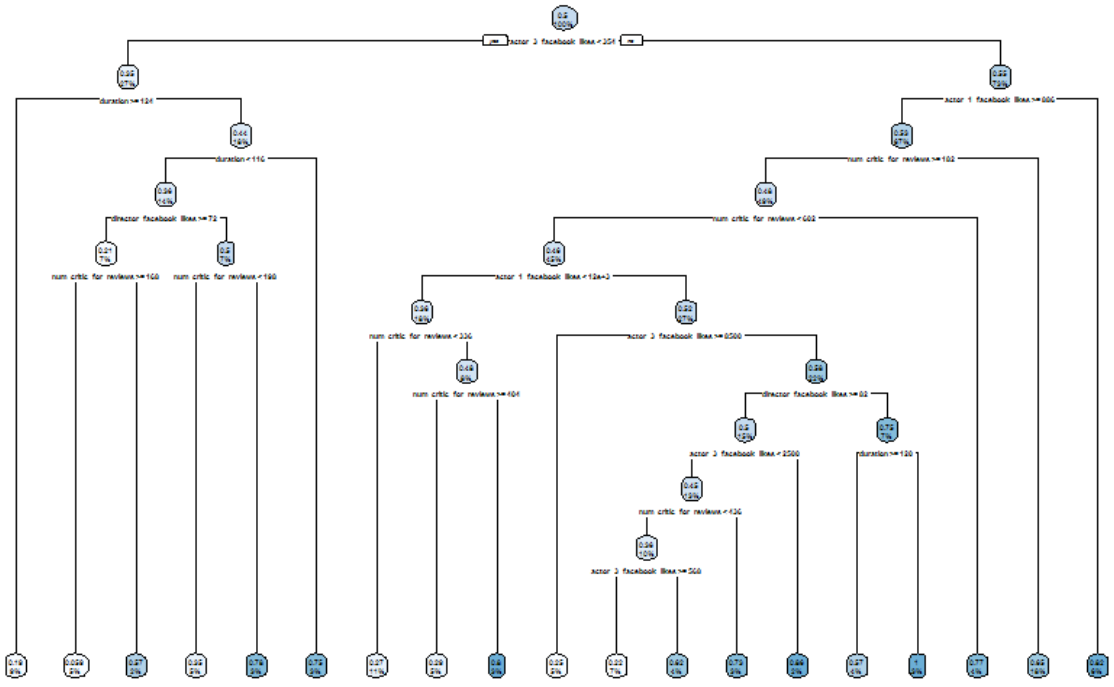
We have divided our dataset into two parts :

1.Training Data

2.Testing Data

Using Stratified Sampling with ratio of 0.7

The following is the decision tree obtained:



After that we have compared the result obtained from ID3 algorithm with the testing data. Then we obtained a confusion matrix by which we can identify the efficiency of the algorithm.

```
Confusion Matrix and Statistics

          Reference
Prediction  0  1
         0 45 31
         1 39 35

               Accuracy : 0.5333
                 95% CI : (0.4502, 0.6151)
    No Information Rate : 0.56
    P-Value [Acc > NIR] : 0.7708

                  Kappa : 0.0652
 Mcnemar's Test P-Value : 0.4028

            Sensitivity : 0.5357
            Specificity : 0.5303
         Pos Pred Value : 0.5921
         Neg Pred Value : 0.4730
             Prevalence : 0.5600
         Detection Rate : 0.3000
   Detection Prevalence : 0.5067
      Balanced Accuracy : 0.5330

       'Positive' Class : 0
```

Team Members:

1.Nadimpalli Sri Harsha(14UCS074)

2.Ankit Gupta(14UCC006)

3.Lovish Jain(14UCC018)