Tyler Burleson
CSCI-4047-901
Exercise 6

Report:

Below are two Confusion Matrices, the blue one is created from a decision tree (DT) and the black one is from the Naive Bayes (NB) model. Both of these models created predictions using the same data set containing 32560 records with 80% of it being used to train each model and 20% to test it. If you compare both these models, the decision tree had better results in this comparison. While both of these were close, DT won in overall accuracy of calculating the "Label".

DT had an accuracy of 84.32% and NB had an accuracy of 82.83%. These were calculated using the True positive/ negative, and the False positive/ negative. DT had a sensitivity (True Positive) of 94.46% while NB had a sensitivity of 93.13%. Next, DT had a specificity (True Negative) of 51.34% and NB had a rate of 49.31%. In each of these fields, the Decision Tree out performed the Naive Bayes model, meaning that the Decision Tree is the better solution in this study.

```
Confusion Matrix and Statistics

          <=50K   >50K
  <=50K   4705    745
  >50K     276    786

              Accuracy : 0.8432
                95% CI : (0.8342, 0.852)
    No Information Rate : 0.7649
    P-Value [Acc > NIR] : < 2.2e-16

                 Kappa : 0.5123

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.9446
            Specificity : 0.5134
         Pos Pred Value : 0.8633
         Neg Pred Value : 0.7401
             Prevalence : 0.7649
         Detection Rate : 0.7225
   Detection Prevalence : 0.8369
      Balanced Accuracy : 0.7290

       'Positive' Class :  <=50K
```

A.   [Decision Tree confusion Matrix]

Tyler Burleson
CSCI-4047-901
Exercise 6

```
Confusion Matrix and Statistics

              <=50K   >50K
    <=50K      4639    776
    >50K        342    755

                 Accuracy : 0.8283
                   95% CI : (0.8189, 0.8374)
      No Information Rate : 0.7649
      P-Value [Acc > NIR] : < 2.2e-16

                    Kappa : 0.4707

   Mcnemar's Test P-Value : < 2.2e-16

              Sensitivity : 0.9313
              Specificity : 0.4931
           Pos Pred Value : 0.8567
           Neg Pred Value : 0.6882
               Prevalence : 0.7649
           Detection Rate : 0.7124
     Detection Prevalence : 0.8315
        Balanced Accuracy : 0.7122

         'Positive' Class : <=50K
```

B.   [Naive Bayes Confusion
    Matrix]