IST 707 Applied Machine Learning
HW5: Use a Decision Tree to Solve a Mystery in History
Name: Jainish Savaliya

Section 1: Data preparation

➔ Before the Data Preparation Part I library all the necessary packages like Rweka , caret , rpart, and ggplot. Then I read the CSV file and Removed filename because there is no need of that column and might give and error if I kept it. Now, Since we are splitting the data I have made 2 data sets with just the disputed and other one is the rest. So now the dataset two without disputed is ready to be split into two parts like training and test data sets.

Section 2: Build and tune decision tree models

➔ Now I converted the Author column into factors to build the model & split the data and train the data on train set data & predicted on the test set. Now to check the accuracy I made the confusion matrix with the fed papers data set and test set.
➔ I got the accuracy of roughly 96% which is good to go for the prediction.

Section 3: Prediction

➔ Finally now I made the prediction on the disputed dataset and It gave more number of the observation for Madison and Less number of observation for Hamilton so We can predict that the disputed papers were Written by Madison which is actually the same prediction we made with clustering method.