Homework 5 – Due 5/4 at 9 AM Eastern Time

1. Complete the tutorial at https://www.datacamp.com/tutorial/random-forests-classifier-python and answer the following questions in paragraph form:

How is the sample data split and what are some reasons why the test data and training data should not overlap?

What were your first three decision trees and what do they show?

Compare the results of using large max_depth decision trees to a using small max_depth decision trees.

Show your confusion matrix and explain the results, including why you believe you got them.

What does recall tell us in this problem, specifically with regards to bank's direct marketing campaign?

2. Generate a random variable in excel with 100 independent experiments from the binomial distribution and plot the probability mass function. Use Excel's Analysis ToolPak to generate the numbers. https://support.microsoft.com/en-us/office/load-the-analysis-toolpak-in-excel-6a63e598-cd6d-42e3-9317-6b40ba1a66b4