Patrick Deniel Pomer

Dec 130 HW2

At first I was trying to visualize the dataset by creating graphs of which independent variables greatly affect the churn rate of the customers. But I had a lot of errors trying to do so therefore I just used trial and error to which variable I should drop until I come up with a high accuracy. I split the test data into train_train and train_test. I sorted out which variables were not that important for my model so I ended up dropping, gender, seniorCitizen, and Dependents. Most of the services that a customer applied for were included in the model. This process was repeated several times until I was able to create a Model with a training accuracy of 0.797 and a test accuracy of 0.818. After coming up with a satisfactory model, I fitted the model to my test data. After fitting the model, there were 2 customerIDs that have no data. I was not sure how to fix it so I dropped it instead.

To summarize the results, I counted the number of No and Yes through the function value_counts() and I got to conclude that 77.71% will not churn and 22.29% will churn. This prediction is more accurate than the first homework given that I was actually able to try and drop variables that weren't important compared to what I did with the first homework which I just assumed which variables were important based on logic and research.