Sabrina Chua So

DEC 130 X

HW 2

To predict the data using the Logistic Regression Model, I first duplicated the Logistic Regression file that was posted in the group thinking that the homework would be as simple as copy and paste. But once I started to copy and paste, I found a lot of error in the process. So, I went to google and tried to understand what the errors meant, and I made the necessary changes to correct these errors and note that this was just for the plot graph. In the end, I ended up dropping more than half the variables to get the most graphs that I could. After getting the graph, I moved on with the test train part of the process and again, I encountered an error. At this point I decided to go on YouTube to find a tutorial that might help me understand what I was doing better. I ended up watching around 3 videos and I helped me understand what I was doing it a bit better, but I still encountered some errors.

I decided to take a break from working on this to clear my head and maybe get a better perspective with some rest. When I went started working again, I decided to not overcomplicate things and keep it simple. So first, I copied all the helper functions from the files that were posted. Then I looked for the codes that I needed to get the graph and to get the test, train data set. I still encountered some problems when it came to the graph, but I decided to consult about that on Saturday because I really did not understand the error. When I was doing the test train data, I realized I had to convert most of my data into dummy variables. So again, I looked at the posted files and played around with it until I got the code that would work. After that, I was looking at the graphs and I was looking at what variables I could drop to for my model. I decided to drop the customer ID because I knew that this could not tell if a person would churn or not. I also decided to drop senior citizen, partners, gender and dependents because I saw that this did not really affect a person’s decision of whether they would churn or not. Based on the graph, I saw that tenure, monthly charges and total charges affected if a person would churn or not. For example, the ones with a smaller tenure would more likely churn, the ones with lower total charges were more likely to churn as well.

So, after getting the dummy variables and dropping the ones that I didn’t need, I then copied the code that was posted in the group for the test train data to get the accuracy of the data. The accuracy that I got was 0.807 and because I was happy with the accuracy level, I loaded the test.csv file to get the churn predictions. I did the exact same process to the test file, except I did not have the graphs and test train code anymore. Then once I got the predictions, I exported it to a csv file.

Compared to getting the predictions manually, this was more difficult for me since this is my first-time coding. However, I think that my predictions are more accurate since my predictions are based on a model which was already tested and I saw that the accuracy was 0.807 which is an accuracy I was happy to get.