Joseph Vince C. Vertulfo

4 BS MIS

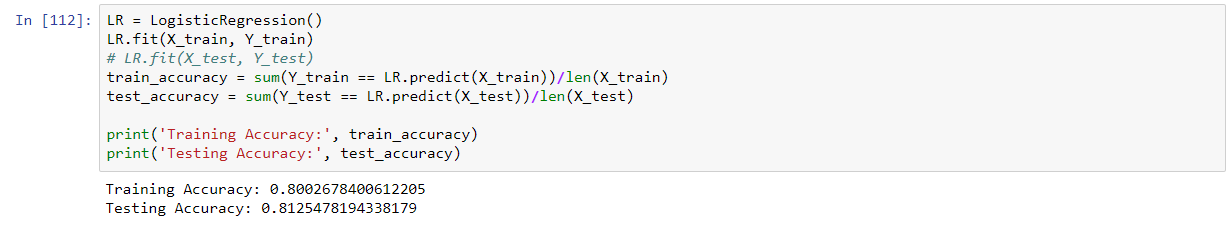
DEC130 Homework

Logistic Regression

Basically, what I did during the whole process was plug and play the values using Sir Ram’s code. Since I’m not new anymore to the Telco dataset because it was used already in our first homework, there were just a few lines of code that I ran to describe the dataset. After studying the train dataset, I ran the plot\_me() function, made by Sir Ram, to see what insight I can get from the visualizations.

After running the plot\_me() function, I dropped the columns/features which I think are not that important in this model and also the independent features. The features I dropped were ‘customerID’, ‘gender’, ‘SeniorCitizen’, ‘Partner’, ‘Dependendts’, ‘PhoneService’, ‘MultipleLines’, and ‘InternetService’. What I’ve noticed also is that there is a huge impact if a customer has not availed of some specific services like streaming service and protection services. If they didn’t avail or they don’t have internet service, they are much likely to churn.

Next, after dropping all the features, I took 20 percent of the train set to test the accuracy of the model I made. The results were good.



Since I’m already satisfied with the model. I proceeded in fitting my model to the actual test data. There, I found out that there are some rows that should be dropped because they contained ‘na’ values. After running ‘.predict’ function, I finally got the results.

