Shawn Ong
Dec 130 - X

Telecom Churn

In order to see which profile of subscribers are most likely to churn, I first plotted each feature to see the distribution of those who churn and those who do not. From what I noticed, a huge chunk of churners do not avail of Online Security, Device Protection, and Tech Support.

In order to predict which subscribers in the test data would churn, I first created a numerical score for each of those three features. If the user is subscribed to said feature, the score would be 1 and if he is not, the score would be 0. The scores for the three features are then added up. If the result is less than 2, then I predict that the subscriber would churn meaning that they are only availing of only one of the said services. In order to test my theory, I performed it on the train dataset to gauge the accuracy. At first it yielded an accuracy of less than half, which would be worse than a coin flip. I then noticed that some subscribers do not avail of internet services therefore they would not need to avail of the services. Taking this into consideration, I would give the score of 1 if no internet services were availed. Doing this increased the accuracy to about 66% for the train data. Since this is a more acceptable accuracy rate, I then applied this method to the test data. In order to increase the accuracy further, perhaps more features should be used.

In further experimentation, the subscription to Fiber Optic internet is taken into consideration as a lot of the churners subscribed to fiber optic. In taking the fiber optic subscription into consideration, subscribers who subscribed to less than 2 of the services and are subscribed to fiber optic are marked to be the ones to churn. In testing this in the train data, the accuracy increased to about 73%. Thus this is applied to the test data.