Anthony Marcolongo

D212

2/2/2021

Analyzing Customer Churn with k-means Clustering

The goal of this project was to determine if the k-means clustering method could be applied to a dataset and predict if a customer was likely to churn or not. The dataset that was used for this project was made up of 10,000 observations and 50 variables. The dataset was cleaned and prepared for use with k-means clustering. This method requires that the dataset be normalized, and that required the use of the scale function in R. Once the data set was cleaned and normalized, the data was split into a test and train set. The sets were then applied to the k-means algorithm. One limitation of k-means clustering is that the value of k has to be input into the model. This was accomplished by using the function fviz\_nbclust in R. This function returned a value of k = 2. With the value of k in hand, the test and train datasets were used with the k-means algorithm to determine which variables have the largest effect on customer churn.

The variables that were shown to have the greatest effect on customer churn were tenure and data usage. The longer a customer was with their provider, the less likely they were to switch. The higher a customer’s data usage was, the more likely they were to switch. This can be explained by the fact that customers that use more data are more likely to incur overage fees. These extra fees can be prohibitively expensive and to a lot of customers they feel like a punishment. Most customers react negatively to this, and it causes those customers to look for new providers.

After looking at the results of the k-means clustering method, the next step was to calculate the accuracy of the model. The model is shown to be 77% accurate, which when you consider the value of k has to be input into the model, and the fact that the dataset only represents a small portion of customers, is really impressive. The recommended course of action based off the results would be for the company to offer a customer loyalty program. This would make customers feel like their time with the company is appreciated, and it would also help to increase customer tenure, which in turn would drive down churn. The second recommendation would be for the company to offer an unlimited data plan. This would stop customers from getting hit with overage fees, which in turn would stop customers from feeling like they are being punished and switching providers.