EECS 731 – Results

10/19/2020

Brian Clark

2928899

This was an interesting assignment, and I enjoyed being able to plot the anomalies that were found within the dataset. I began by reading in the data, displaying the information about the dataset, and some calculations that are available via Pandas. I next provided a general graphical representation of the data that was to be used in ‘Feature Engineering’. Next I displayed the data over differing periods of time along with the average margin for the number of Tweets concerning that corporation on given days.

After displaying where I believed the majority of anomalies are in another graphical representation, I then began applying the feature engineering models to the data. I chose to use Isolation Forest, and Local Outliers Factor. I applied the same procedure to two different datasets, Google and Amazon, and received similar results from both sets of data. It appears to me that LOF found as many or more anomalies when compared to Isolation Forest. However, both models found the same anomalies and were able to be plotted successfully using Matplotlib.