**Final Project Report on Netrality Data Sets**

The Project assigned by our Professor Jared Mroz is titled “Predicting Potential Prospect Customers”. This is the continuation of the earlier Project and the final phase of our Project. The previous week we have dived into the other learnt Regression Models which are: - (1) K Neighbors Regressor, (2) Random Forest Regressor, (3) Extra Trees Regressor, (4) Gradient Boosting Regressor, (5) XGBRegressor, (6) Cat Boost Regressor and (7) LGBMRegressor. Where we created different Data Frames for model to learn.

Since, the model – Random Forest Regressor which we selected from the previous week is like a tentative ideal model, even then we have predicted the potential customers, but it wasn’t up to the mark when checked for the metric scores.

In this final phase of the week, with lots of enthusiasm and hard work we tried a new method yet an effective one for predicting the potential customers. The method which we used was the “Clustering Method”. In this clustering method we used only the Current Customers and Prospect Customers csv files. Pranay created a new column named “Binary” for both the csv files. “Binary” – 0’s for ‘Current Customers’ and “Binary” – 1’s for ‘Prospect Customers’. We then merged the two Data Frames with the new column – Binary.

We then selected only features and skipped giving the targets as the Clustering methods accepts only the input Data. The features are: 'Revenue (in 000s USD)', 'Est. Marketing Department Budget (in 000s USD)', 'Est. Finance Department Budget (in 000s USD)', 'Est. IT Department Budget (in 000s USD)', 'Est. HR Department Budget (in 000s USD)', 'Number of Locations', and 'Recent Funding Amount (in 000s USD)'.

We then created the K-Means, Spectral and Birch Clustering models and also created a column named ClusterIDs which will be having the number of clustered sets for every row of the new DataFrame created.

Pranay then evaluated every clustered set, in order to identify the strong bonds/relationships established between the Prospect and Current Customers Data. (where we ensured the ratio for ‘Prospect : Current’ = 1:1/1:2/2:1)

After deep evaluations of the clusters, we were able to acquire 15 Potential Prospect Company IDs, which were strongly related to the Current Customers Data File.

To conclude we can say that we as a team understood how to analyze Data effectively by applying Machine Learning Algorithms, and with our hard-work and confidence we were able to Predict the Potential Prospect Customers.