**Project Proposal: Using Machine Learning to Improve Customer Service in Finance**

I have worked for one the largest financial firms straight out of college. For its employees, it is very much an industry where its workers are striving to be the alpha. There is a lot of competition amongst the various financial firms, but also among it's employees. The finance industry is fueled by sales people working in many different financial firms from banks to mortgage lenders. Thus, every firm is trying to push sales to increase commission and bonuses. There is a fine line that needs to be balanced in terms of customer service and sales. This project will focus on the customer service side of the finance industry. The problem I want to solve is how to increase customer service given the vast amount of data on customer complaints. From experience, I can tell you that there are a lot of complaints from the service of a firm to the products themselves. The clients for this project can be all financial firms who want to improve their responsiveness of their customer service.

The data is coming directly form the database of the Consumer Financial Protection Bureau, or CFPB. The data the project will be working can easily be downloaded and queried here https://data.world/cfpb/consumer-complaints. As of 5/20/2019, the data set is comprised of 1,284,185 individual rows of complaints, updated frequently. I plan to conduct visual and statistical analysis on the entire data set which includes 18 columns. From the data set, I only plan to use the complaints where the customers give a written response. The reason being, I want to use machine learning and natural language processing techniques on the written customer complaints as well as the categorical columns. Taking only the data with written customer complaints, leaves 383,840 rows of data.

To decide whether a complaint is handle in an ‘appropriate’ manner I will use whether the complaint was handled in a timely manner or not as a predictor. I use appropriate in quotes because the term ‘appropriate’ can be ambiguous but I assume it means that the complaint was handled as efficiently as it could. Not responding in time could mean the complaint was not resolved, difficult to handle, or maybe ignored due to lack of concern about the customer’s complaint. The goal will be a machine learning model that can correctly identify these non-appropriate response times based on the features of data and the customer complaint. The results could be extremely valuable by giving insights on which complaints are the most difficult to handle. Being able to quickly identify the difficult complaints can be leveraged to improved customer service and satisfaction which will lead to higher customer retention.

Along with a model that can be used for predicting the complaints, I will provide Jupyter notebooks, mile stone reports, a slide presentation, along with visual and statistical analysis.