**Capstone Project Proposal**

**Mentor**: Mujtaba Hasan

**Mentee**: Brandon Croarkin

**Date**: 5/6/2016

**Predicting USMTO Submissions**

The goal of this project is to create a predictive model of whether or not companies will submit data to AMT’s USMTO monthly survey. AMT has conducted this survey since 1997 and has about 250k rows of data. This will be coupled with AMT’s other data sources, and the merged dataset will ultimately list detail such as (on a month by month basis)

1. Whether or not the company submitted data for that month
2. What month they submitted for
3. When they submitted the data in reference to when the data period opened
4. How long they have been in the program
5. What values the company submitted
6. Which products those values correspond to
   1. Something to capture market share
7. Geography of company
8. Size of company (employees, sq ft, revenue?)
9. Engagement with AMT Trade Shows

This will be coupled with macroeconomic data such as

1. Industrial production
2. Manufacturing technology production
3. USMTO Total Orders

To come up with a predictive model of if they will submit/not submit in the upcoming month.

**Client**

The Association for Manufacturing Technology (AMT) is the client for this project. AMT represents over 75% (by dollar volume) of manufacturing technology industry in the United States, with members such as Toyoda Machinery, Kennametal, and Mitsubishi Machinery Systems. AMT currently releases a monthly directional forecast of manufacturing technology output based on both subjective/proprietary industry knowledge and objective economic indicators.

**Data**

Association for Manufacturing Technology Proprietary Data, Federal Reserve Economic Database, American Survey of Manufacturers

**Approach**

·        Collect and merge all the data - this will require significant wrangling.

·        Scope the problem down to make it both more tractable, as well as relevant.

·        Model it as a problem to predict the probability of submission for each company. Train the model on an earlier data set and test it on later subsets.

·        Evaluate effectiveness of the model by tracking it against historical data.

·        Tune model based on accuracy to historical data

·        Revise model and test on current data

**Deliverables**

Code, Report, Slide Deck