**Project 2 Proposal – Predicting Investment Returns by Tracking “Insider” Trading and Twitter Activity**

**Team Members:** Jack Galvin & Josh Fram

**Repository Link:** <https://github.com/UC-Berkeley-I-School/Project2_Fram_Galvin>

**Project Summary:** The SEC requires all executives at US listed public companies to disclose any trading activity conducted in the shares of companies they work for. This “Insider” buying and selling activity is updated frequently and publicly available for investors to download from the SEC’s website. Some investors have developed strategies that track this data, driven by the notion that if a company’s CEO is buying stock in their company, it is an indication that they believe it is a good investment. This is ostensibly an important indictor given the CEO is the person who should be most familiar with the company and its prospects. The reverse holds as well – if an executive is reporting sales of their company, it may be a bad signal around the prospects of that company.

In our project, we would like to track this data ourselves and integrate it with stock returns, to see if there is a relationship between “Insider” trading activity and a company’s stock price.

As an additional step, we would like to integrate Twitter activity for the relevant companies, to see if there is any additional insight that can be obtained by combining trading activity with Tweet volumes and sentiment.

**Primary Dataset:** The SEC’s EDGAR Database maintains a running list of all “insider” buys and sells reported via Form 4 submissions. This will be the primary data set. Specifically, our sample set will be isolated to companies that saw Insider activity in the 4th quarter of 2018.

***(I will get the right link / data table sample from Bloomberg before we submit this proposal but for now here is a link to what is available for this:*** [***https://www.gurufocus.com/insider/summary***](https://www.gurufocus.com/insider/summary)

**Secondary Datasets:** We will source daily stock returns for a 1-year period following any insider activity that occurred in our sample period (4Q 2018). This data will be sourced from Bloomberg. We will source Twitter activity from the Twitter API ***(Jack: for Twitter, can you confirm or change this / add data sample / link as needed?)***

Data set joining will rely largely on stock tickers and dates. The relevant columns are as follows:

* SEC Edgar Database: The Direction (Buy / Sell) and Amount of each trade
* Stock Market Data: The Price in the period immediately prior to, and 1 year following, the Insider activity
* Twitter Data: The Volume (and Sentiment, if possible) of Tweets following the disclosure of the Insider activity.

**Report Discussion:** We intend to cover whether or not there is any meaningful correlation between Insider trading and a stocks relative performance. We will also seek to identify if Twitter activity can provide additional insight into whether this Insider buying or selling will be meaningful for a stock’s returns over the following year.