**Predicting Stock Market Prices From Sentiment Analysis of the Brexit**

“Brexit” is the term used to refer to the UK leaving the European Union. Last June 23, a referendum was held to decide whether the UK should stay or leave in the EU. “Leave” won, 52% to 48%.

Among the many things affected by this event was the stock market. As investors believed that “stay” would prevail, share prices in the FTSE 100 increased on June 23, hours before the vote. Following the unexpected “leave” announcement, the FTSE 100 had slid 5.6%, as of June 27. Global stock market prices similarly followed.

Social media provides a real-time representation of public mood and sentiment at any given time. Since financial decisions can be driven by emotion and mood, it may be worthwhile to observe how changes in sentiment affect prices in the stock market.

The goal of this project is to analyze how sentiment regarding the Brexit affected stock market prices, and to create a predictive model using sentiment analysis.

**Data:**

Various news reports (http://www.wsj.com/articles/stock-market-winners-and-losers-after-brexit-vote-1467055543) (http://www.marketwatch.com/story/these-stocks-are-biggest-post-brexit-winners-and-losers-in-the-us-2016-07-15) have shown that some stocks were more affected by the Brexit, either gaining or losing. Several of these stocks will be chosen for this project.

Intraday prices, with interval of 3 minutes, will be obtained from Google Finance.

To get sentiment from social media, posts will be scraped from Twitter Search, using the keyword “brexit”, in a 2-week period starting from June 17 to July 1.

**Methodology:**

Data will be divided into intervals of 15 minutes, with each interval being one data point.

To create features, two different methodologies will be used. First, built-in functions in TextBlob will be used to measure sentiment and polarity. The other measure will be based on the Profile of Mood States Questionnaire (POMS). For each tweet, the score of 65 different words and their synonyms from Sentiwordnet and the thesaurus, will be mapped to six different mood states (confusion, tension, anger, depression, vigor, fatigue).

These features can further be weighed with other data from the tweets, such as the location, the name of the poster, the number of favourites and retweets, and the number of tweets that day.

To see whether sentiment correlates with the prices, p-values will be computed using Granger causality analysis.

Finally, different classifiers will be constructed, using linear and logistic regression, SVM, and neural networks.

**Deliverables:**

Code, paper, slide deck