**INFO 7390 – Machine Learning Research Project Proposal**

Dhruv Patel, Keyur Donga

10/05/2017

**Stock Price Prediction of a pre-defined set of companies using sentiment analysis**

**Description:**

Apart from the fluctuations in stock price of a company due to normal trading, the stock prices of a company could fluctuate due to various factors like a new product launch or say acquisition of a stake or the whole company of a perceived competitor or internal news leaks and much more. Every event that takes place within a company is reacted upon by its shareholders and traders. People post their reactions of these events on social networking sites like Facebook, Twitter and also write about it on blogs and in newspaper articles too. This project aims at future prediction of stock prices of a pre-defined set of companies using sentiment analysis from Twitter, newsrooms and other blogs. The correct result may provide valuable information to the user to invest in that stock and if not then take actions accordingly.

**Background:**

Predicting the exact possible stock price of a company is nearly impossible as there are a lot of factors that are involved which sets the price of a stock at that particular moment and this fact can be better understood by carefully studying the historical data of an organization. However, using machine learning algorithms, we can try our best to predict a stock price range and at least come to conclusion that whether a particular stock will be profitable for an investor or not. Here the background idea is the connection between a statistical and a non-statistical entity. Here product launches, news, rumors are all non-statistical entities but in a way, they affect the prices of stocks which a statistical entity. So this project is an attempt to predict the stock prices, which are affected by this pre-existing connection, using machine learning.

**Data Sources:**

1. Queried twitter datasets downloaded from Twitter using its API.
2. News articles of companies related to our stock prediction taken from major news houses/networks like “The Wall Street Journal”, “The New York Times”, etc.
3. Financial and economic websites such as CNBC, Financial Times, Bloomberg and many more.

**Machine Learning Algorithms:**

1. Linear Regression
2. Logistic regression
3. Sentiment Analysis

**Project Evaluation:**

The algorithm will note the fluctuation of the stock prices and also note the respective sentiments that led to the increase or decrease in the price. A graph of correct and incorrect sentiments will be observed along with changes of prices in stock. The aim is to compare a particular sentiment to a stock price movement. If the change in a stock price acts accordingly to the sentiment then the sentiment is correct otherwise it is not. Thus based on the percentage of correct and incorrect sentiments of a particular company we can predict its future stock price.

**References:**

1. <https://blog.twitter.com/official/en_us/topics/insights/2016/twitter-data-and-the-financial-markets.html>
2. <http://cs229.stanford.edu/proj2011/ChenLazer-SentimentAnalysisOfTwitterFeedsForThePredictionOfStockMarketMovement.pdf>
3. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.375.4517&rep=rep1&type=pdf>
4. <https://www.quora.com/Can-Twitter-sentiment-analysis-guide-stock-market-investment>
5. <https://pdfs.semanticscholar.org/ff65/ab316def7128fac8998178568e25ec1dad8f.pdf>
6. <http://www.knowledgetaiwan.org/ojs/index.php/ijbi/article/viewFile/387/101>