**The Effect of VC Tweets on the Stock Market**

Vishal Kalyanasundaram

**1. Abstract**

The goal of this project was to detect a relationship between features of Tweets posted by prominent venture capitalists and the stock market. A linear SVM was trained on a large set of Tweets, which was used to predict whether the market went up or down, or if it experienced a larger than average change.

**2. Analysis**

Tweet data was collected using Tweepy for Python. The venture capitalists chosen were Ben Horowitz (‘@bhorowitz’), Elon Musk (‘@elonmusk’), John Doerr (‘@johndoerr’), and Reid Hoffman (‘@reidhoffman’). These users were chosen because of the size of their Twitter following (around 250k each), and similarities in their Twitter activity; the users had similar numbers of Tweets with similar focus on startups, Silicon Valley, and the stock market. Tweepy allows up to 3200 tweets to be collected per user; that subset was then reduced to a range between January 1st, 2010, and December 31st, 2014. The features kept were the time created, the text of the tweet, the number of favorites, and the number of retweets.

Stock market data was collected through Yahoo Finance. The data used were from the S&P and the NASDAQ between January 1st, 2010, and December 31st, 2014. The features used were whether the market went up or down from open to close, and whether the market experienced more than a .07% change up or down per day, which was found to be the average daily movement over this time period.

The model made additional use of sentiment scores through the Python Natural Language Toolkit. The raw Tweet text was input to the text processor, which generated Positive, Negative, and Neutral sentiment scores, as well as a label for the overall sentiment if it was Positive, Negative, or Neutral.

A scikit Linear SVM was trained on the set of features for each venture capitalist: Favorites, Retweets, Positive, Negative, Neutral, and Overall Sentiment. A C value of .1 was chosen through cross validation.

**3. Results**

The overall training data accuracy results were as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Market** | **Prediction Type** | **Accuracy** |
| Ben Horowitz | NASDAQ | Up/Down | 0.560687960688 |
| Ben Horowitz | NASDAQ | Jump | 0.733660933661 |
| Ben Horowitz | S&P | Up/Down | 0.54898911353 |
| Ben Horowitz | S&P | Jump | 0.764644893727 |
| Reid Hoffman | NASDAQ | Up/Down | 0.517189835575 |
| Reid Hoffman | NASDAQ | Jump | 0.662182361734 |
| Reid Hoffman | S&P | Up/Down | 0.530642750374 |
| Reid Hoffman | S&P | Jump | 0.674140508221 |
| John Doerrr | NASDAQ | Up/Down | 0.545977011494 |
| John Doerrr | NASDAQ | Jump | 0.532567049808 |
| John Doerrr | S&P | Up/Down | 0.524904214559 |
| John Doerrr | S&P | Jump | 0.63601532567 |
| Elon Musk | NASDAQ | Up/Down | 0.503276539974 |
| Elon Musk | NASDAQ | Jump | 0.462647444299 |
| Elon Musk | S&P | Up/Down | 0.584144645341 |
| Elon Musk | S&P | Jump | 0.73157162726 |

Overall the accuracy rate is slightly higher than a 50/50 guess especially for the jump predictions. However, more in-depth analysis must be performed to measure a correlation between the tweets and jumps in the market.