CS591 Network and Markets **Project Progress Report**short line

**Social Media Impact on Stock Market & Prices**

Scarleth Estevez

Nathan Gallaway

Sang-Joon Lee

Dec 4 2016

# 

# **Table of Contents**

[Table of Contents](#_qezgdvhmuoh)

[Section A. Project Progress Report](#_gexkng1roy73)

[Initial Project Title](#_fo5gjfmzihd)

[Challenges](#_v2vex5pios83)

[Change of Scope](#_2mo120o20k7s)

[Revised Project Title](#_pnc9pd3mcelk)

[Overview](#_x6i4bybgn0ig)

[Methodology](#_lgfr0jlyomb9)

[Data Scraping and Dataset](#_c1v7n0ph6q9p)

[Correlation of volume of Positive and Negative news articles with Stock Market Prices](#_5uzjtub6oxy2)

[Intermediate Results & Analysis](#_rx1kva2ptfiw)

[Remaining Work](#_hkyski6p5zl2)

[Related Work](#_2vode8bsbra)

# 

# 

# **Section A. Project Progress Report**

## **Initial Project Title**

Reliability and Spread of News in Social Media

## 

Our initial project goal was to study on measuring the news articles reliability based on our calculations of where the article originated, how many similar articles have been written across different source, and how many biased vocabulary identified in the article effect the spread of the article throughout social media. We quickly realized that this was very ambitious and challenging goal as described in list below:

### Challenges

1. Challenges of mapping new articles to articles
   * 1. The data scraped form webhose.io (the tool we used to scrap data) contained information on the number of shares of an article, however, it did not provide the links of who has shared the article. Without these information, we realized that it is very difficult to measure spread of the news which was the main analysis and characteristics that we relied upon.
   1. Challenges of scrapping Twitter Data
      1. We also looked into scrapping the twitter using the Twitter API. However, the twitter API does not allow data to be scrapped for twits that are more than 7 days old. This became another roadblock to gathering wide range of data source on different news event that we can focus for analysis and test our hypothesis.
   2. Correctly identifying criteria for measuring what is a reliable news
      1. With loss of two data sources, it was difficult to identify a reasonable criteria for measuring the reliability of news.

## **Change of Scope**

Due to reasons stated above, we decided to change the direction and narrow down the scope of our project. The goal of our project now it to take the new and articles from different source on the web by companies that are listed on stock exchange, more specifically on Dow Jones. The goal is to relate the articles with different sentiment and measure its effect on stock market prices.

## **Revised Project Title**

Social Media Impact on Stock Market Prices

## 

## **Overview**

We are researching on effect of news article in social media of companies listed in stock market and whether or not the sentiment of articles and spread throughout the network is correlated with the performance of stock market prices.

## **Methodology**

We are getting our data from webhose.io which is an api we can use that data scraps news articles and returns it to us in json format. We will pick initially 10 stocks from various sectors from fortune 500 companies and search for news articles based on its sentiments, positive, negative and neutral sentiments. These data will be saved into JSON file for analysis. Initially the number of articles from each articles to see if there are any affect on them, then we will see factor other factors to its amplification of sentiment by its popularity (spread) of sharing which expect to see impact on news to stock market.

## **Data Scraping and Dataset**

We scraped news article from 10 different stock market prices from different sectors. They are: Facebook, Google, Starbucks, Visa, Walmart, Bershire Hathaway, Ford Motor, Chevron, J.P. Morgan Chase.

We have scraped news articles from each company and stock market price for the last 30 days for each companies, to perform analysis.

If time allows, based on outcome of analysis, we will attempt to expand our research into different companies from fortune 500 of companies to test validate our hypothesis and anslysis.

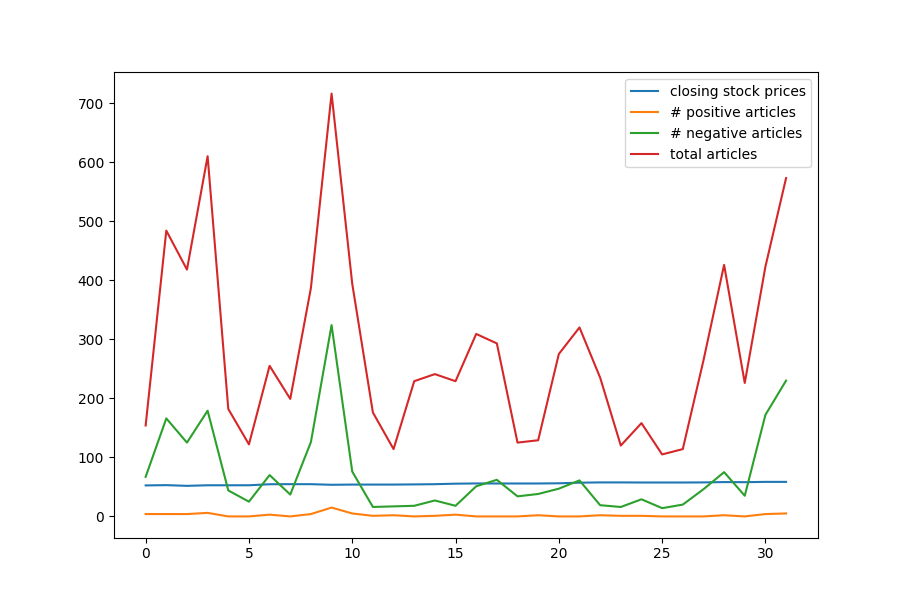
**Measuring Correlation Between News/Article Sentiment and Stock Market Prices**

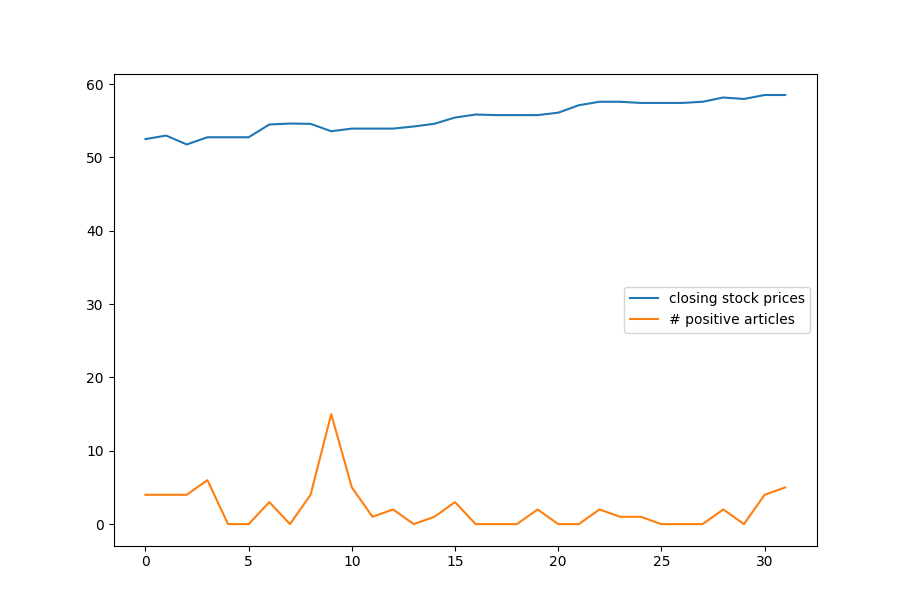
## Correlation of volume of Positive and Negative news articles with Stock Market Prices

1. Calculate Correlate Positive and Negative new articles on series of data over last 30 days
2. Calculate and perform Granger Causality test on positive and negative news articles with stock market prices [1] In the paper, it uses twitter data for determining correlation between twitter sentiment and stock market prices. Our hypothesis is that this holds true for news articles.

## **Intermediate Results & Analysis**

Below is example of Starbucks Company Positive & Negative News article and stock market prices:





## **Remaining Work**

Due to change in scope, we have significant amount ahead of us prior to presentation date. We goal is complete analysis of data for targeted amount of company stocks and plan for validating our work on other companies.

## **Related Work**

[1] John Bollen, Huina Mao, Xiao-Jun Zeng (Oct 2010), Twitter mood predicts the stock market.

[2] C Castillo, M Mendoza, Barbara Poblete (2011), Information credibility on Twitter, Proceedings of the 20th international conference on World Wide Web., page 675-684