Twitter sentiment and Stock price

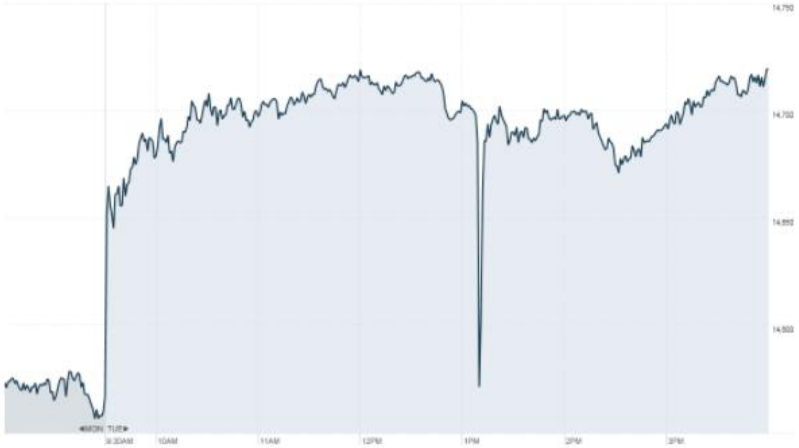
Xinpeng Shen, Xiaochen Jin

High speed trading fueled Twitter flash crash

by Maureen Farrell @maureenmfarrell

April 24, 2013: 10:05 AM ET

Recommend



The Dow dropped more than 140 points Tuesday.

“One scary -- and false -- tweet, and the Dow quickly plunged 140 points, or roughly 1%. Many are pointing fingers at high speed trading by computers for the swift decline.”

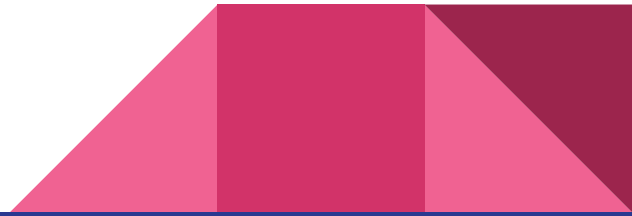
-- http://money.cnn.com/2013/04/24/investing/twitter-flash-crash/

Background

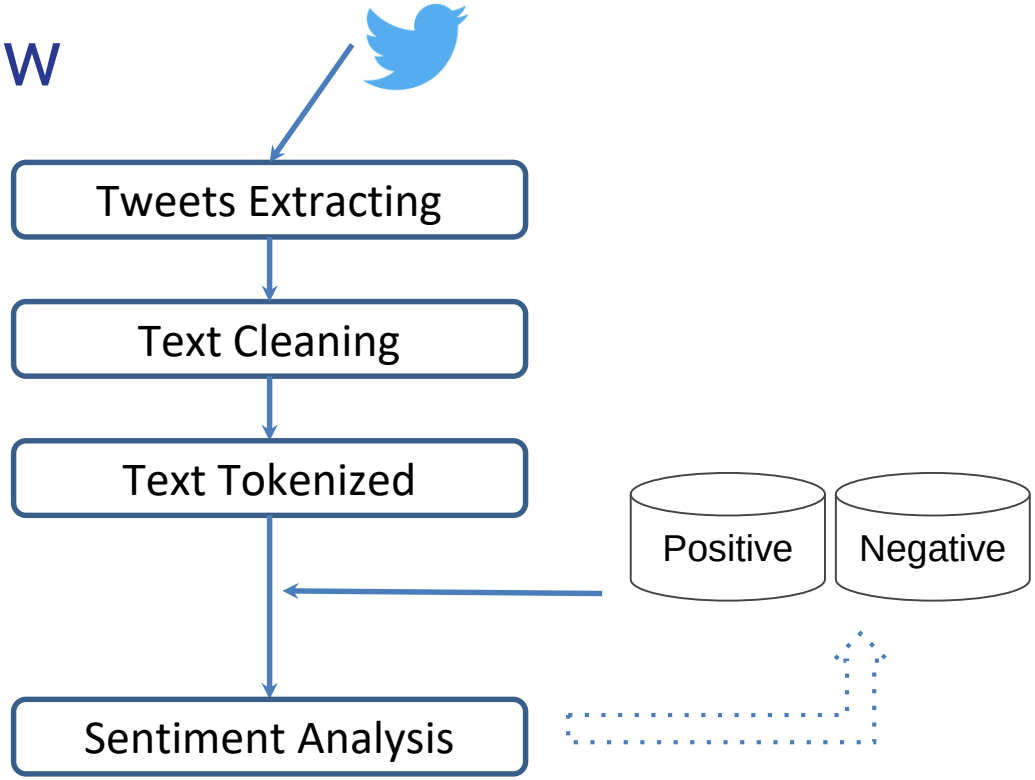
- It’s known that news have significant impact on High-Speed computer trading -- what can we extract from historical streaming tweets

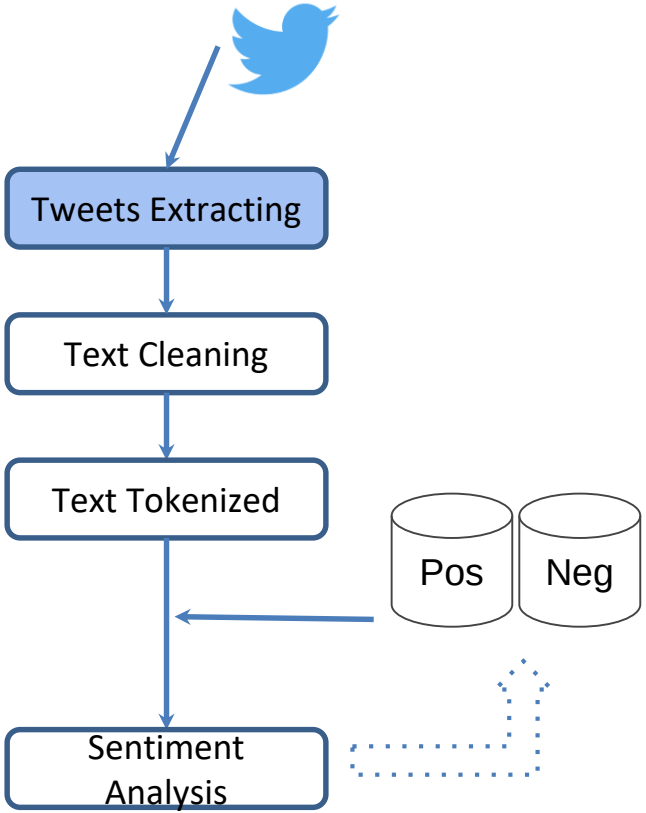
- Text mining and NLP(natural language processing) have been a hot topic in data science world

- Tweets are “open data”

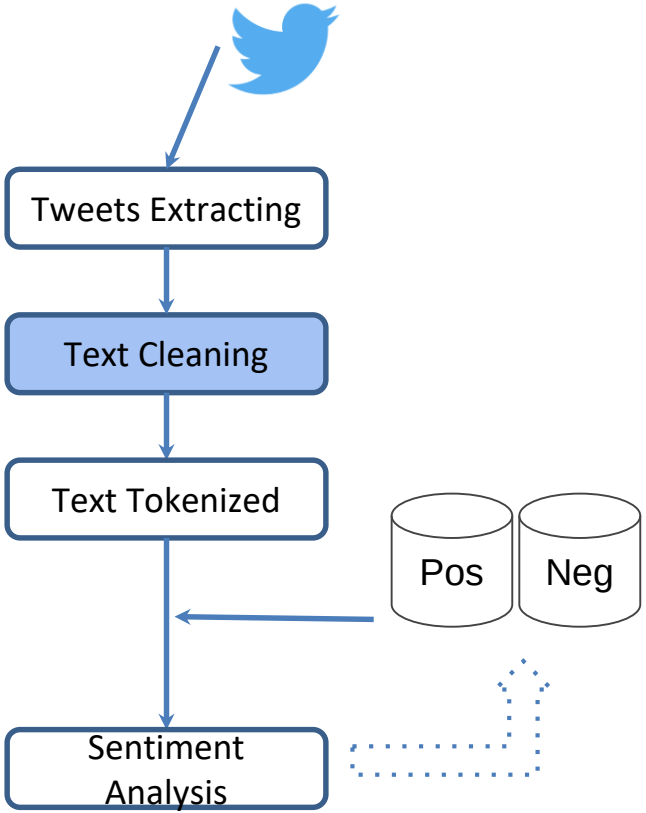


Workflow

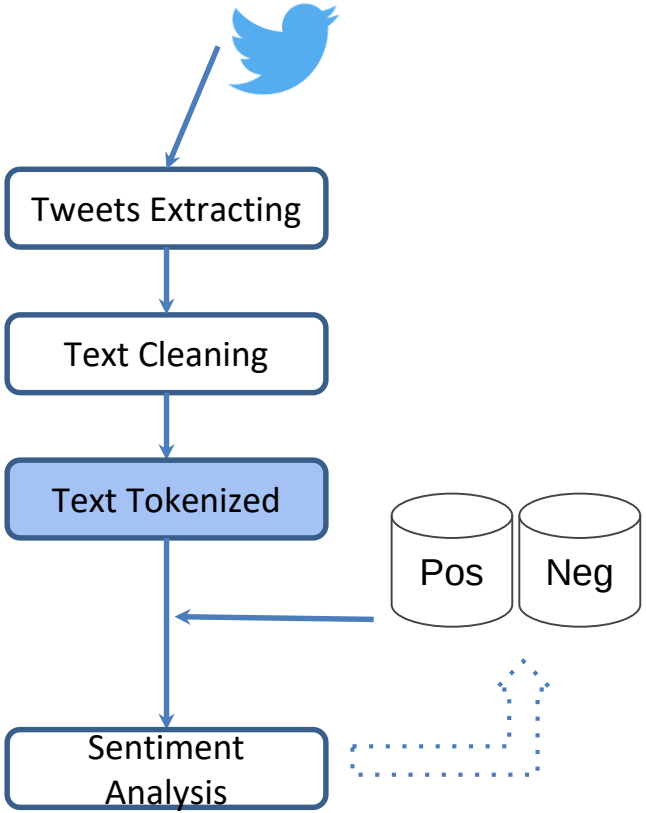

Tweets Extracting
Text Cleaning
Text Tokenized
Positive
Negative
Sentiment Analysis



- Set up connection to Twitter Application Management Portal
- Extracting tweets by searching keywords.
  “$AAPL”, “$SPY”, “$FB”
- Limitations:
  - Historical tweets as far as 10 days
  - First 140 words of each tweets



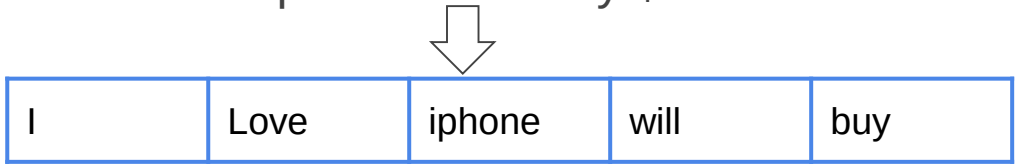
- Remove Punctuation and Special symbols. (#, !, @, http://) using Regular Expression

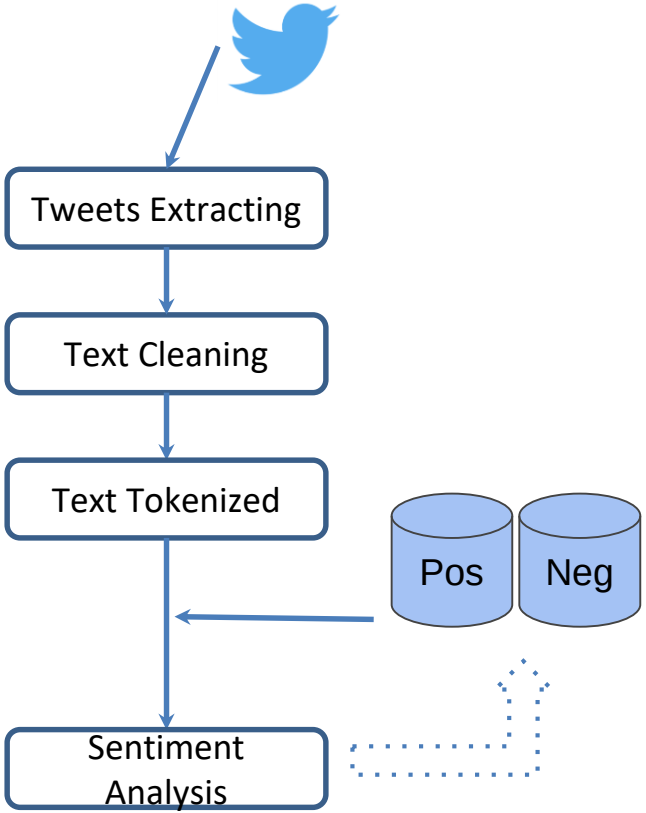


- Splitting sentence into words

E.g.

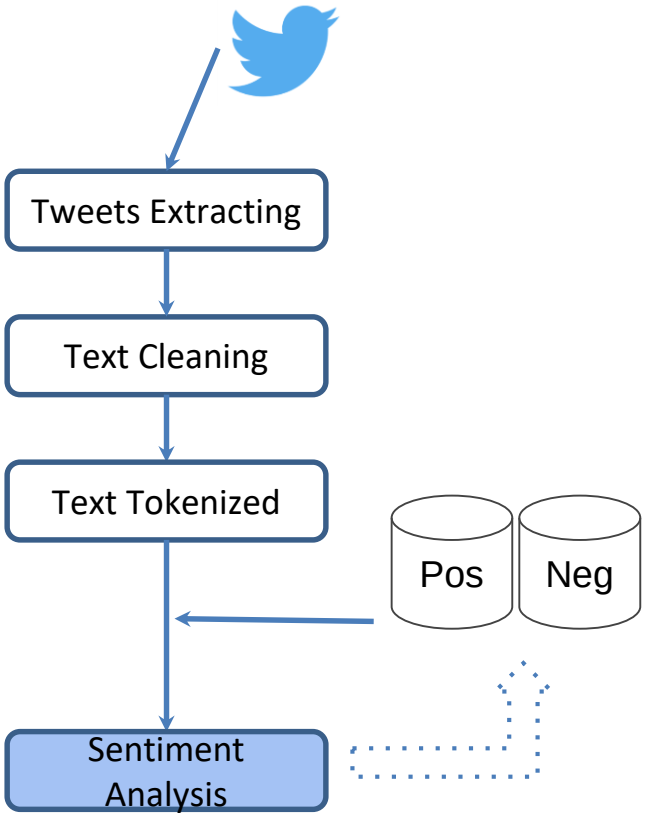
I love iphone! I will buy $AAPL





good	trash
nice	weak
amaze	bad
...	...

- Existing Positive and Negative word dictionary. Developed by other researchers



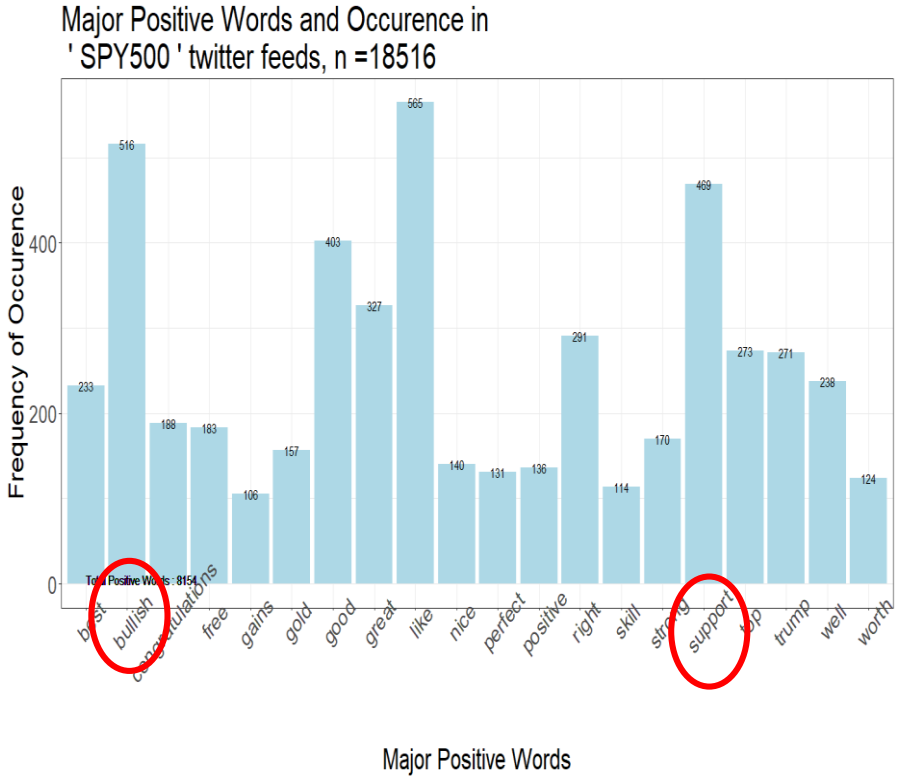
- Comparing our tokens with positive and negative word dictionary
- Generating summary statistics and analysis results(will show later)
- Revise the word dictionary by adding positive or negative word

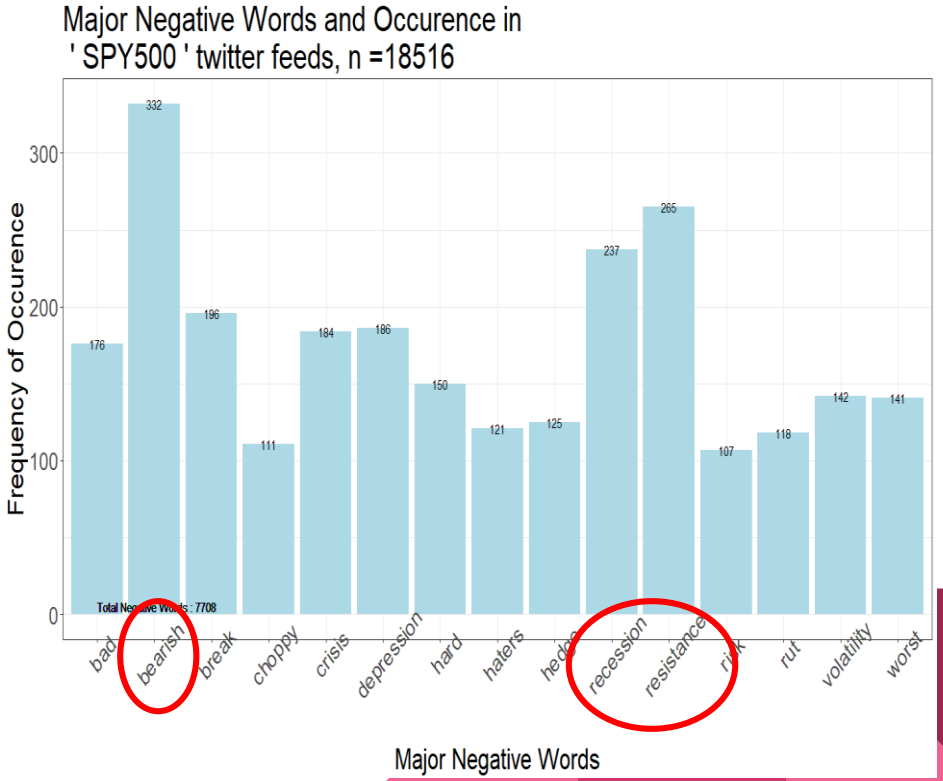
E.g. “The stock is really cheap”

Results I

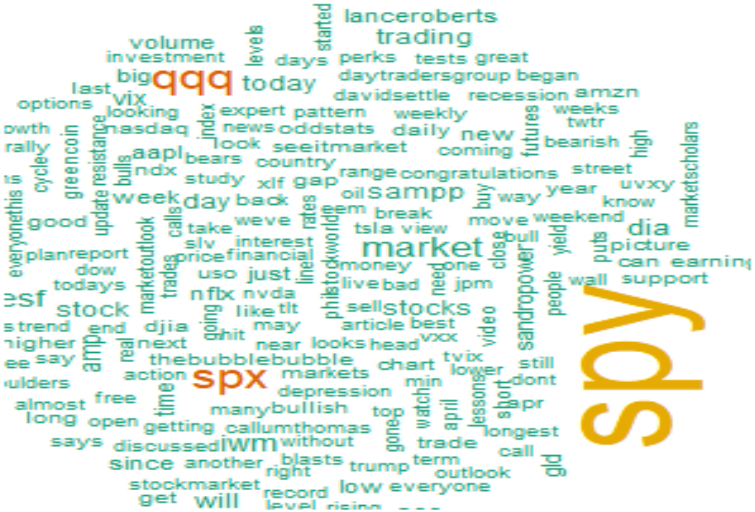
- Visualize the frequency of positive and negative words for each stock
  - Frequency plot
  - Word cloud

Positive and Negative occurrence in Tweets about SPY500

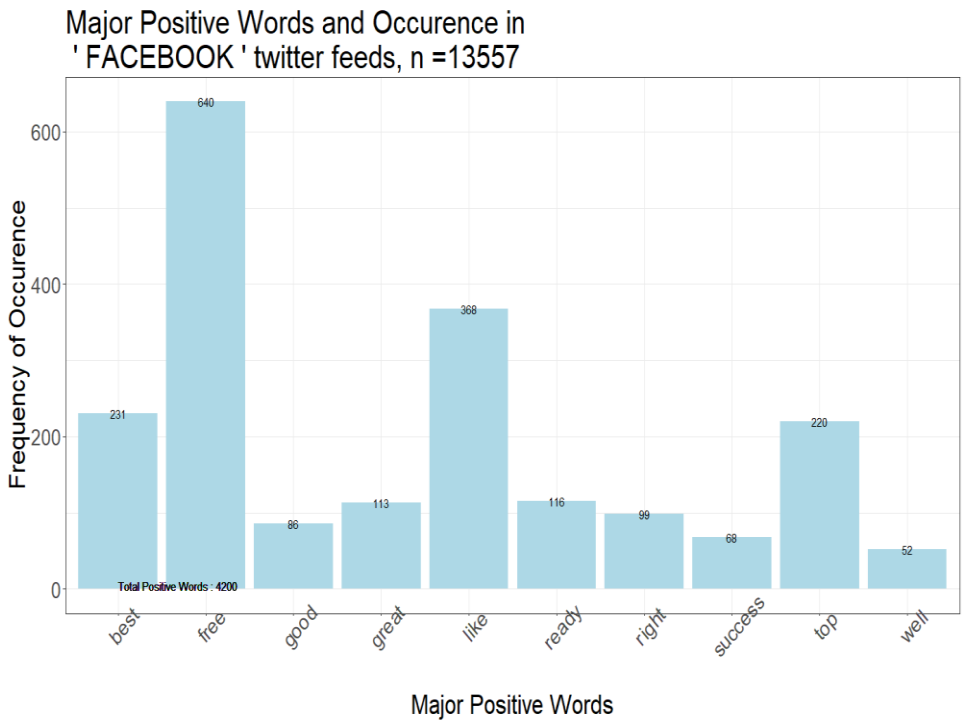


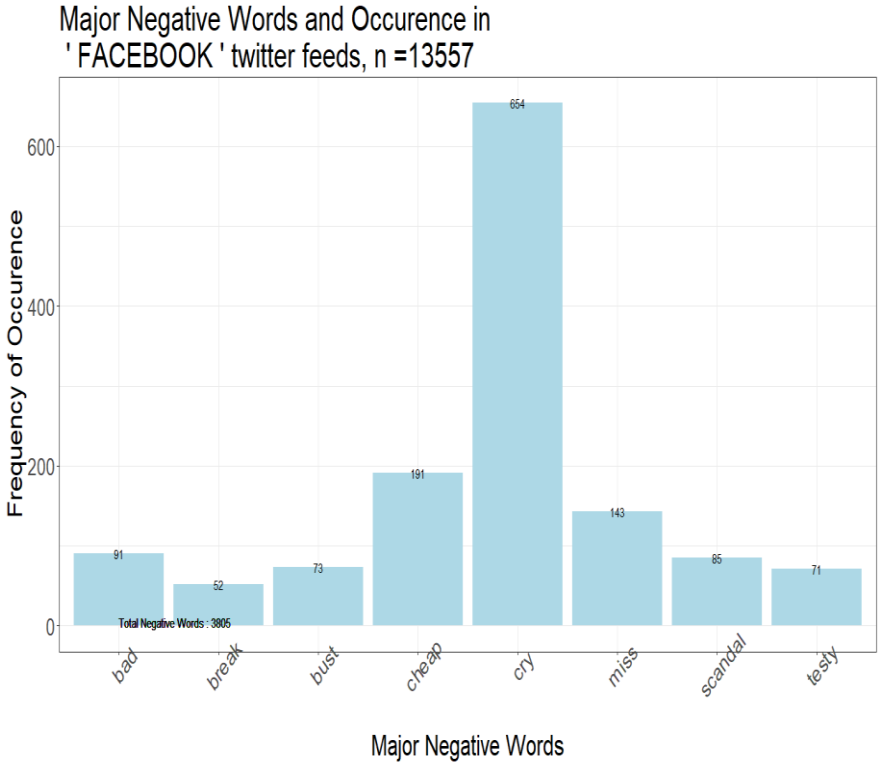


Word cloud of SPY

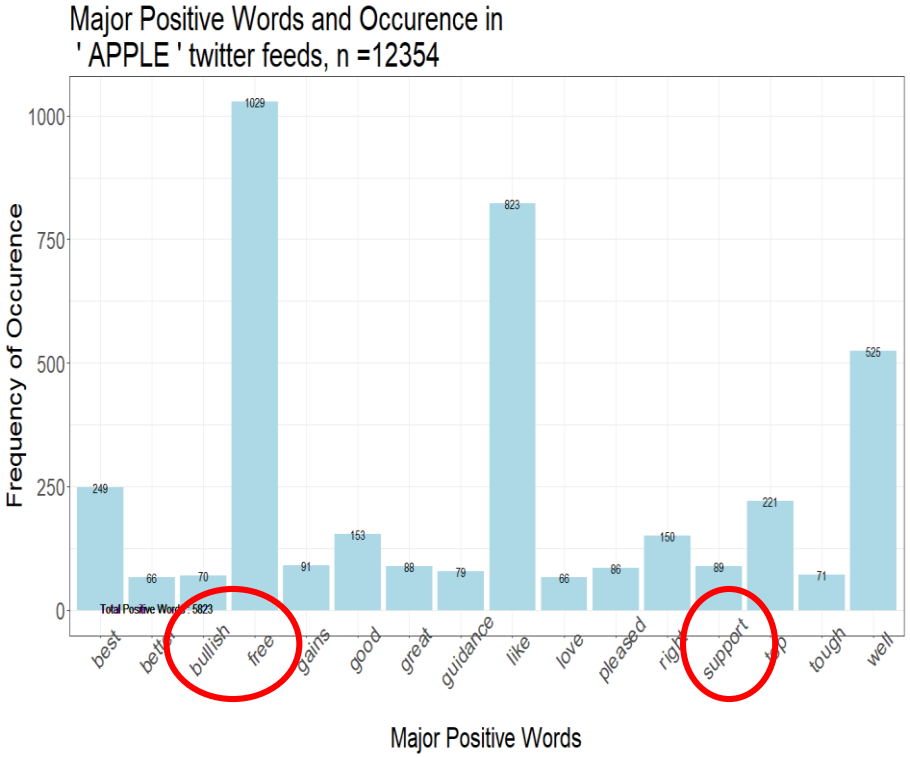


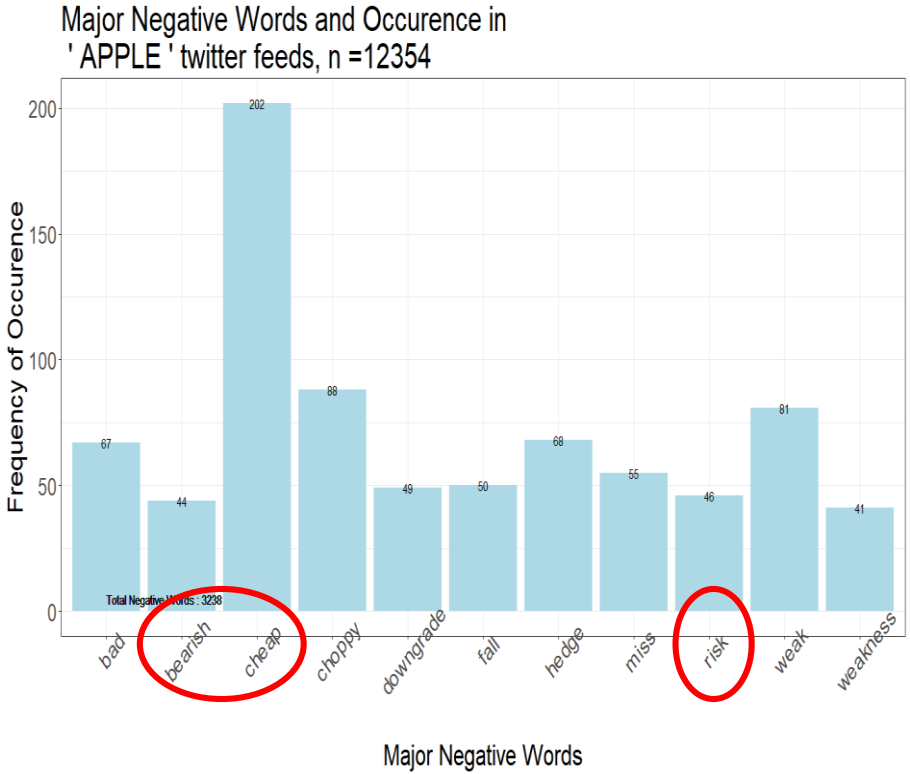
Positive and Negative Occurrence in FB tweets





Positive and Negative Occurrence in Apple tweets





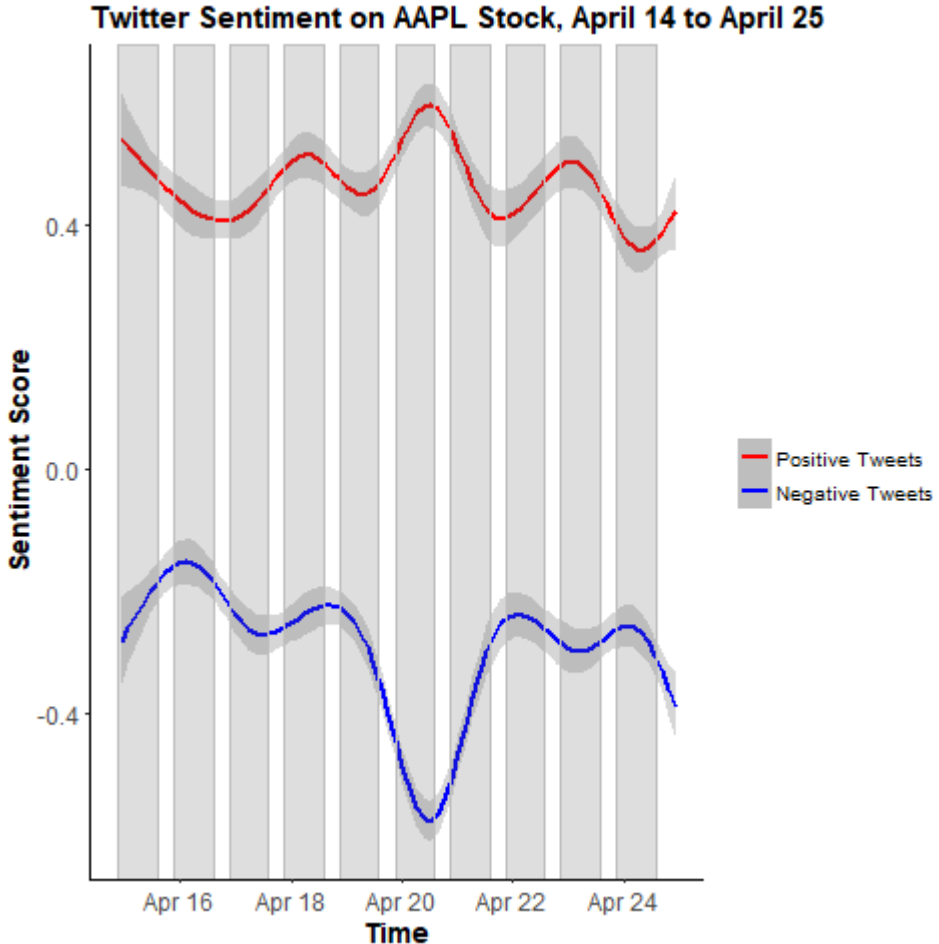
Word Clouds of Apple vs Facebook

Results II

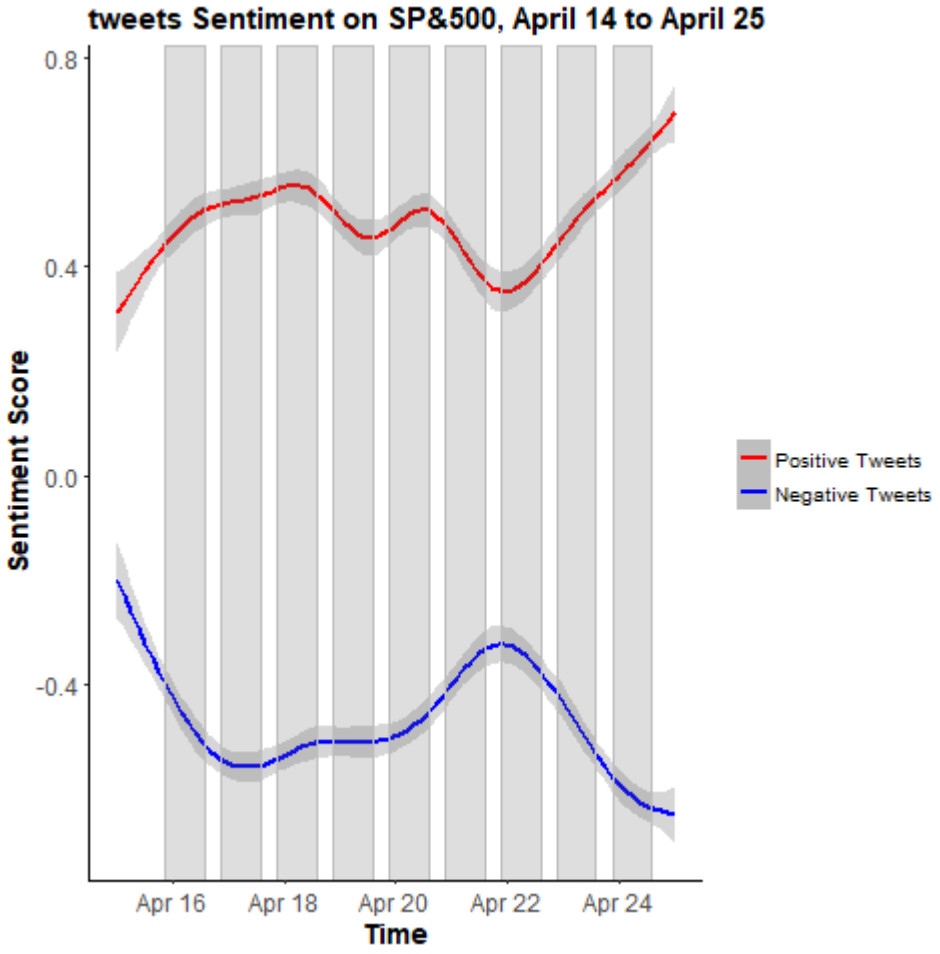
Time series

- Time series of Twitter Sentiment
- Time series of Stock price

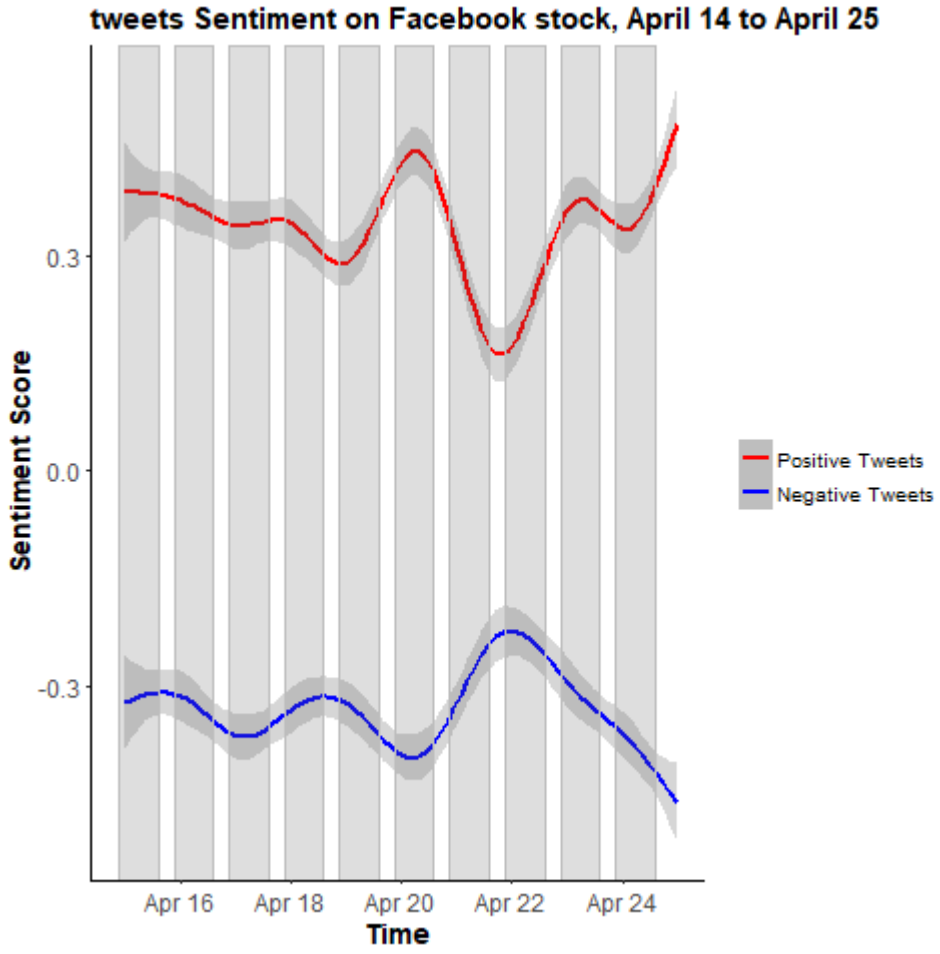
12354 Tweets



18516 Tweets



13557 Tweets



Discussion of Results

- Strong visual correlation between sentiment score and stock market price
- Due to limitation of data point, we can’t fit time series model, but the sentiment might be a good variable
- Causal relation is not very clear, need intraday stock price movement

Further Direction

As we learning and exploring the Text mining filed, we realized there are a lot to do.

1. Keep updating our Positive and Negative word dictionary
2. Labeling the tweets by hand, make the sentiment analysis into a supervised learning problem. → Creating our text mining model, special word dictionary
3. Using new technique Word2vec, topic modeling(Latent Dirichlet allocation)
4. Adding new text source, like news

Reference

1.Twitter Sentiment Analysis Tutorial, https://rpubs.com/abNY2015/90345

2. News Sentiment Analysis Using R to Predict Stock Market Trends. Anurag Nagar and Michael Hahsler, Southern Methodist University, http://past.rinfinance.com/agenda/2012/talk/Nagar+Hahsler.pdf

3. Short Introduction to tm.plugin.webminin, Mario Annau https://cran.r-project.org/web/packages/tm.plugin.webmining/vignettes/ShortIntro.pdf

4. Ingo Feinerer. Introduction to the tm Package Text Mining in R. https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf

Thank you & any questions?