NEW YORK TIMES ARTICLES AND SHARE PRICE RETURN

STAT 418 Final Project

Ah Sung Yang

May 7, 2019

DATA SOURCE

- Obtained articles from **New York Times Article Search API**

- Searched all the articles on Apple Inc. published from Jan. 1, 2014 to Apr. 30, 2019

- Besides the query, also used the filter query field to ensure articles have Apple Inc in their organizations field

- Extracted the following information from the JSON file

  - abstract, date, desk, document type, headline, id, locations, organizations, snippet, source, subjects, type of material, url, word count

- For the share price data, used the python package called yahoo_financial to download daily share prices from Yahoo Finance API

abstract	date	desk	doctype	headline	id	locations	organization	snippet	source	subjects	ty
b'Apple\xe2\x80\x99s iCloud data service works...	2014-01-07	Business	b'article'	b'Taking Along iCloud Calendars'	5c88530a40d36ff070b32a3a	[]	[Apple Inc, iCloud, Samsung Group, Google Inc]	b'Apple\xe2\x80\x99s iCloud data service works...	The New York Times	[Android (Operating System), Smartphones]	Questi
b'In March, Apple and Samsung Electronics are ...	2014-01-09		b'article'	b'Apple and Samsung Chiefs to Meet a Mediator ...	5c883a5240d36ff070ad279b	[]	[Apple Inc, Samsung Electronics Co]	b'In March, Apple and Samsung Electronics are ...	The New York Times	[Inventions and Patents, Smartphones, Suits an...	Nev
b'Last year was the first in which personal co...	2014-01-10		b'article'	b'For PC Makers, the Good News on 2013 Is That...	5c883a5240d36ff070ad27d7	[]	[Apple Inc, Dell Inc, Gartner Inc, Hewlett-Pac...	b'Last year was the first in which personal co...	The New York Times	[Computers and the Internet, Desktop Computers]	Nev
b'A summary of differences in the latest Windo...	2014-01-10	Business	b'article'	b'Changes in Windows 8.1, With Skype Replacing...	5c88530a40d36ff070b32a3c	[]	[Apple Inc, Microsoft Corporation, Skype Techn...	b'A summary of differences in the latest Windo...	The New York Times	[Windows (Operating System), Mobile Applicatio...	Questi
b'To draw young buyers and increase its market...	2014-01-13	Business	b'article'	b'Cost of Cool in India? An iPhone'	5c8873ec40d36ff070b92dac	[India]	[Apple Inc]	b'To draw young buyers and increase its market...	The New York Times	[iPhone, Prices (Fares, Fees and Rates), Smart...	Nev

	doctype	id
0	b'article'	1502
1	b'multimedia'	53
2	b'paidpost'	1

	source	id
4	The New York Times	1538
2	International New York Times	8
3	Reuters	6
1	CNBC	3
0		1

	type	id
5	News	1212
9	Question	206
8	Op-Ed	47
14	Video	24
2	Interactive Feature	23

1. TEXT CLASSIFICATION: NAÏVE BAYES CLASSIFIER

- **NaiveBayesClassifier:** A classifier based on the Naive Bayes algorithm, as implemented in NLTK (Textblob package)

- Combined all articles issued on a single day

- Took log difference to get daily share price returns and convert it into a binary response (positive / negative)

- Inner-joined the two data sets

- Divide it into train / test sets
  - Train set: 700 days (~90% of data)
  - Test set: 85 days (~10% of data)

	date	log_diff	binary_response	abstract	headline	snippet	paragraph
0	2014-01-07	-0.007178	neg	Apple's iCloud data service works best with Ap...	Taking Along iCloud Calendars	Apple's iCloud data service works best with Ap...	Q. Our family life depends on Apple shared cal...
1	2014-01-09	-0.012852	neg	In March, Apple and Samsung Electronics are sc...	Apple and Samsung Chiefs to Meet a Mediator Ah...	In March, Apple and Samsung Electronics are sc...	In March, Apple and Samsung Electronics are sc...
2	2014-01-10	-0.006695	neg	Last year was the first in which personal comp...	For PC Makers, the Good News on 2013 Is That I...	Last year was the first in which personal comp...	The two leading analysis companies tracking th...
3	2014-01-13	0.005221	pos	To draw young buyers and increase its market s...	Cost of Cool in India? An iPhoneDaily Report: ...	To draw young buyers and increase its market s...	BANGALORE, India — After deliberating for mont...
4	2014-01-14	0.019703	pos	This is not the usual start-up acquisition: Th...	Google and Nest: Two Companies in the Business...	This is not the usual start-up acquisition: Th...	Google has announced it is buying Nest Labs, m...

- The highest accuracy rate was <u>**~60%**</u> (when we used abstract)

- The daily share price returns must be difficult to be predicted only based on NY Times article abstracts

1. TEXT CLASSIFICATION: NAÏVE BAYES CLASSIFIER

- The top 20 most informative features used by the classifier

```
Most Informative Features
           contains(we) = True              neg : pos   =      7.6 : 1.0
        contains(worth) = True              pos : neg   =      7.1 : 1.0
          contains(yet) = True              neg : pos   =      6.9 : 1.0
       contains(showed) = True              neg : pos   =      6.1 : 1.0
        contains(women) = True              neg : pos   =      5.4 : 1.0
 contains(improvements) = True              neg : pos   =      5.4 : 1.0
       contains(appeal) = True              neg : pos   =      5.4 : 1.0
       contains(Street) = True              pos : neg   =      5.2 : 1.0
      contains(lawyers) = True              neg : pos   =      4.7 : 1.0
     contains(pressure) = True              neg : pos   =      4.7 : 1.0
         contains(sign) = True              neg : pos   =      4.7 : 1.0
      contains(improve) = True              neg : pos   =      4.7 : 1.0
     contains(internet) = True              neg : pos   =      4.7 : 1.0
        contains(place) = True              neg : pos   =      4.7 : 1.0
    contains(president) = True              pos : neg   =      4.6 : 1.0
        contains(store) = True              pos : neg   =      4.6 : 1.0
         contains(Wall) = True              pos : neg   =      4.6 : 1.0
         contains(file) = True              pos : neg   =      4.0 : 1.0
        contains(third) = True              pos : neg   =      4.0 : 1.0
       contains(turned) = True              pos : neg   =      4.0 : 1.0
```

2. SENTIMENT ANALYSIS USING EXISTING ANALYZERS

- Used the existing sentiment analysis functions to get sentiment scores (i.e. polarity) for each text data

1. **Textblob / Sentiment (Python):**
   - Uses PatternAnalyzer of the Pattern library

2. **Textblob / NaiveBayesAnalyzer (Python):**
   - Uses Naïve Bayes Analyzer trained on a movie review corpus

3. **Sentimentr / sentiment_by (R):**
   - Attempts to take into account valence shifters (i.e., negators, amplifiers (intensifiers), de-amplifiers (downtoners), and adversative conjunctions) while maintaining speed
   - An augmented dictionary lookup

- Created a Shiny app that plots the historical sentiment/polarity scores
- The app allows you to choose:
  - Analyzer
  - Text data (abstract, headline, snippet and paragraph)
  - Frequency of data (daily, weekly and monthly)
- For the sentimentr package, added the top 20 most used positive and negative terms for computing sentiment scores

2. SENTIMENT ANALYSIS USING EXISTING ANALYZERS