Capstone Project with Moelis & Co.

DSI Spring 2019

Presented By
Minsuk (mk3474@columbia.edu)
Paul (hp2496@columbia.edu)
Yijia (yc3425@columbia.edu)
Jolie (jt2998@columbia.edu)
Gary (pb2632@columbia.edu)

Future Profit Warning Classification using Earnings Call Transcript

Future Profit Warning Classification using Earnings Call Transcript

Profit Warnings Example

1) Apple CEO Tim Cook issued the company’s first profit warning in 16 years on January 3 2019, pointing to China’s slowing economy for the company’s revenue shortfall

1) FedEx Corporation stated during the Q3 2019 earnings call on March 19 2019 that they are “now forecasting fiscal 2019 adjusted earnings $15.10 to $15.90 per diluted share” from previous EPS $15.50 to $16.60

Data Source

RTT News Profit Warning Calendar

Date	Symbol	Co. Name	Curr. Est	New Range
04/30/19	LLY	Eli Lilly And Co.	$25.1 - $25.6 Bln	$22.0 - $22.5 Bln
04/25/19	MMM	3M Co.	10.45 - 10.90	9.25 - 9.75
04/23/19	WAT	Waters Corp.	9.20 - 9.45	9.05 - 9.25
04/18/19	DHR	Danaher Corp.	4.75 - 4.85	4.72 - 4.80
04/02/19	USNA	USANA Health Sciences	5.25 - 5.55	5.00 - 5.35
04/02/19	USNA	USANA Health Sciences	$1.25 - $1.30 billion	$1.21 - $1.26 billion
03/26/19	CCL, CUK, CCL.L	Carnival Corp.	4.50 - 4.80	4.35 - 4.55
03/19/19	FDX	FedEx Corporation	15.50 - 16.60	15.10 - 15.90
03/06/19	DCI	Donaldson Company Inc.	2.31 - 2.45	2.27 - 2.41
02/06/19	LLY	Eli Lilly And Co.	5.90 - 6.10	5.55 - 5.65
01/31/19	ABC	AmerisourceBergen Corp.	6.65 - 6.95	6.65 - 6.85
01/23/19	TEL	TE Connectivity Ltd.	$13.9 - $14.3 Bln	$13.45 - $13.85 Bln
01/23/19	TEL	TE Connectivity Ltd.	5.60 - 5.80	5.35 - 5.55

Seeking Alpha Earnings Call Transcript

Pfizer Inc. (PFE) CEO Albert Bourla on Q1 2019 Results - Earnings Call Transcript

Apr. 30, 2019 2:11 PM ET | 1 comment | 3 Likes | About: Pfizer Inc. (PFE)

Q1: 04-30-19 Earnings Summary

Press Release | Slides

EPS of $0.85 beats by $0.10 | Revenue of $13.12B (1.64% Y/Y) beats by $106.48M

Earning Call Audio Subscribers Only

0:00 / 1:19:18

Pfizer Inc. (NYSE:PFE) Q1 2019 Earnings Conference Call April 30, 2019 10:00 AM ET

Company Participants

Chuck Triano - SVP, IR

Albert Bourla - CEO

Frank D'Amelio - CFO

Data

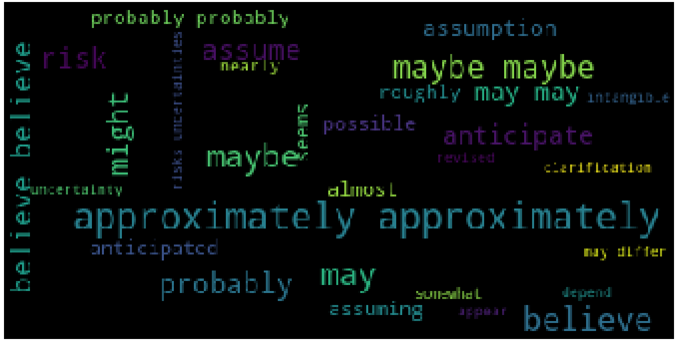
- S&P500 Healthcare (A, ALGN, AMGN, BDX, CAH, CERN, etc..)
- 40 transcripts from future profit warning stocks and 40 transcripts from healthy companies

Profit Warning Dat	Symbol	Company Name	Target	Transcript YR	Transcript Quarter
10/24/2018	LH	Laboratory Corporation Of America Holdings	1	2018	Q1
8/7/2018	XRAY	Dentsply Sirona Inc	1	2018	Q1
7/25/2018	VAR	Varian Medical Systems	1	2018	Q1
5/3/2018	CAH	Cardinal Health Inc	1	2018	Q1
10/23/2018	DGX	Quest Diagnostics	1	2018	Q1
10/23/2018	WAT	Waters Corp.	1	2018	Q1
NA	ABT	Abbott Laboratories	0	2018	Q1
NA	ANTM	Anthem Inc.	0	2018	Q1
NA	BAX	Baxter International Inc.	0	2018	Q1
NA	CELG	Celgene Corp.	0	2018	Q1
NA	HSIC	Henry Schein	0	2018	Q1
NA	IDXX	IDEXX Laboratories	0	2018	Q1
NA	MTD	Mettler Toledo	0	2018	Q1
NA	SYK	Stryker Corp.	0	2018	Q1
NA	WCG	WellCare	0	2018	Q1
NA	ZBH	Zimmer Biomet Holdings	0	2018	Q1
NA	ABMD	ABIOMED Inc	0	2019	Q1
NA	MDT	Medtronic plc	0	2019	Q1

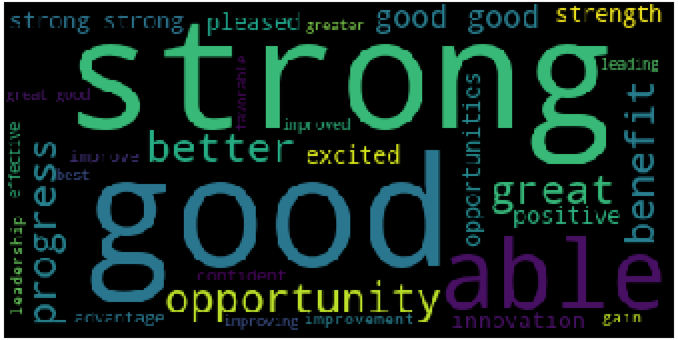
Exploratory Data Analysis (Wordcloud)

Profit Warning Wordcloud (Positive Words)

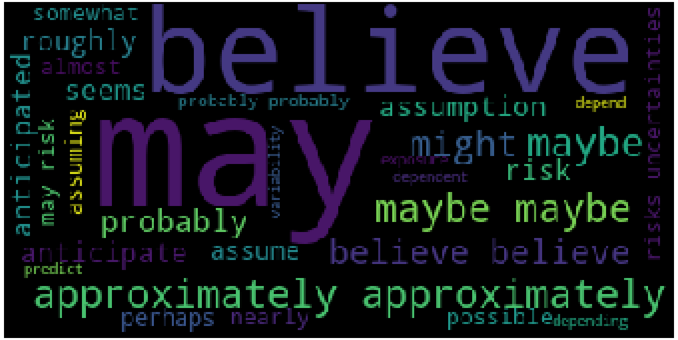
Profit Warning Wordcloud (Uncertainty Words)



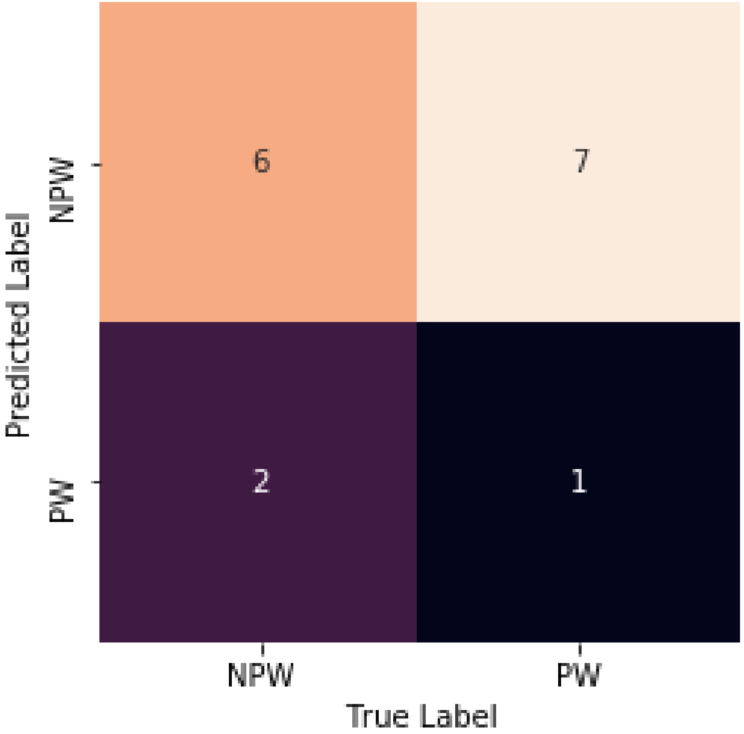
Non Profit Warning Wordcloud (Positive Words)



Non profit Warning Wordcloud (Uncertainty Words)



First Approach: Multinomial Bayes Classifier



- The train accuracy is: 0.5938
- The test accuracy is: 0.4375

More Features (NLTK VADER Library)

- VADER Library incorporates word-order sensitivity relationship so it captures more than what bag-of-words model.
- Scores are ratios for proportions of text fall in each category like positive, negative, and neutral
- V_neg : Negative words portion
- V_neg_MD: Negative words portion from Management Part
- V_neg_QA: Negative words from Q&A Part

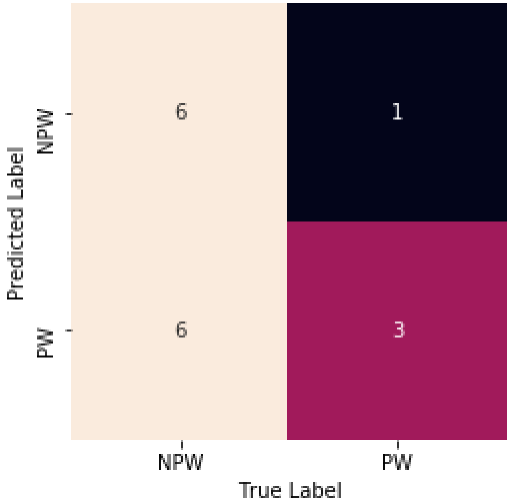
	v_neg			v_neg_MD			v_neg_QA		
	mean	median	std	mean	median	std	mean	median	std
Price Warning									
0	0.131713	0.129222	0.020261	0.133579	0.130318	0.026960	0.129459	0.122796	0.024919
1	0.131444	0.129328	0.025961	0.131925	0.128934	0.020026	0.128935	0.123881	0.040484

	v_pos			v_pos_MD			v_pos_QA		
	mean	median	std	mean	median	std	mean	median	std
Price Warning									
0	0.291604	0.294847	0.043364	0.226861	0.230769	0.023669	0.327504	0.329477	0.056278
1	0.285036	0.289488	0.032298	0.224997	0.222796	0.014370	0.318936	0.318482	0.045623

Logistic Regression & Random Forest

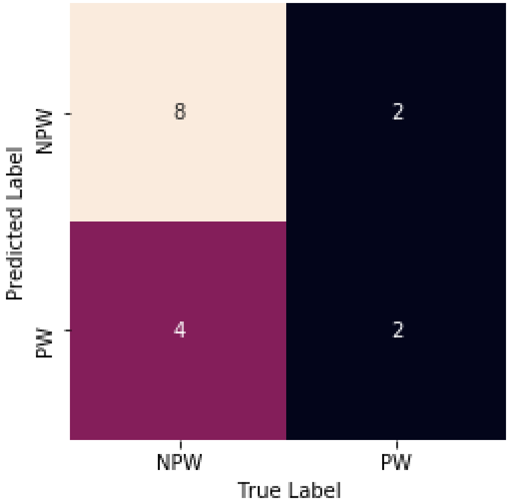
Logistic Regression

Accuracy: 56.25%

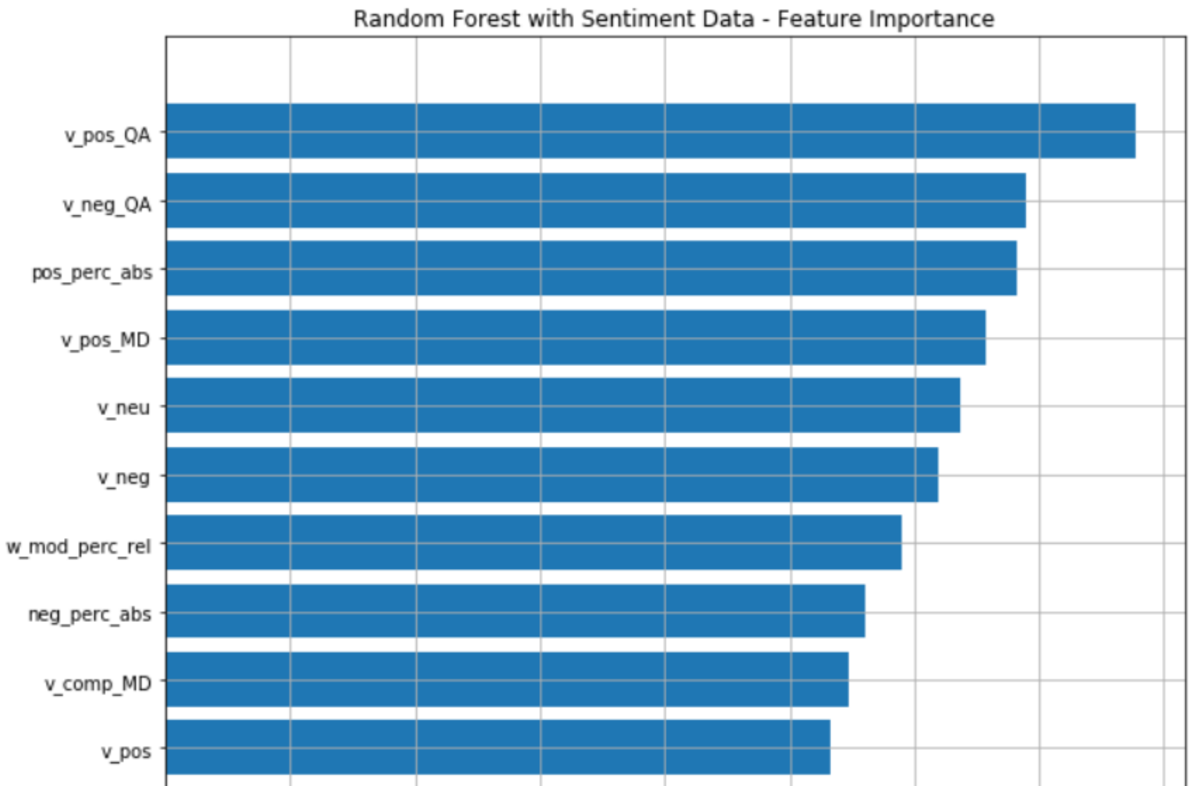


Random Forest

Accuracy: 62.5%



Feature Importance (Random Forest)



Future Works

- Different Industries (Industrials, Consumer)
- More samples
- Different Features by using other libraries or complexity/readability analysis
- Other machine learning models including XGBoost and neural networks

Stock Predictions Using News Sentiments

Project Overview

Hypothesis

- News articles have impact on stock market
- We are assuming news article sentiment is a significant signal for predicting a company’s stock return

Goal

- Use daily average news article sentiment score to predict the future daily stock return
- We are using Apple Inc’s news and stock price to test our hypothesis

News Data

Data Sources

- ProQuest Historical Newspapers
- News sites include: New York time, Wall Street Journal, CNN, Forbes, etc.

Date Range

- April 21st, 2016 - March 15th, 2019

Notes:

- *News Relevance is calculated by applying TF-IDF on key words of Apple Inc.*
- *TF-IDF refers to term frequency-inverse document frequency.*
- *Tf-idf weighting are often used by ranking a document's relevance given a word.*

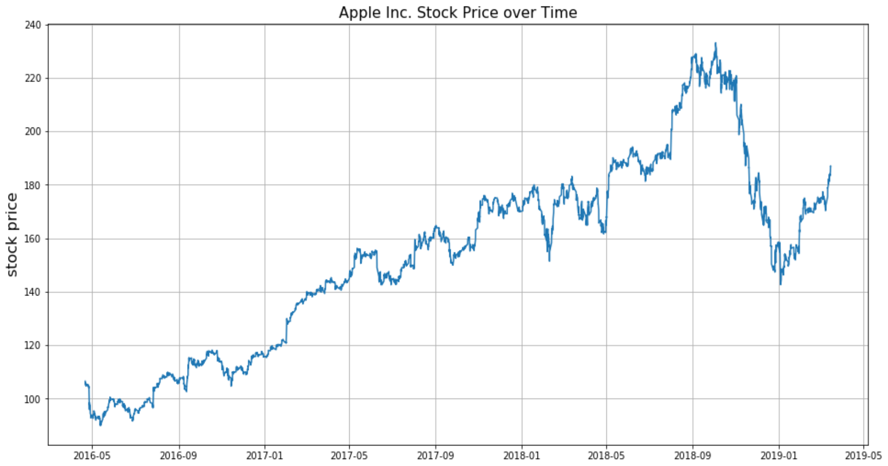
Data Preprocessing

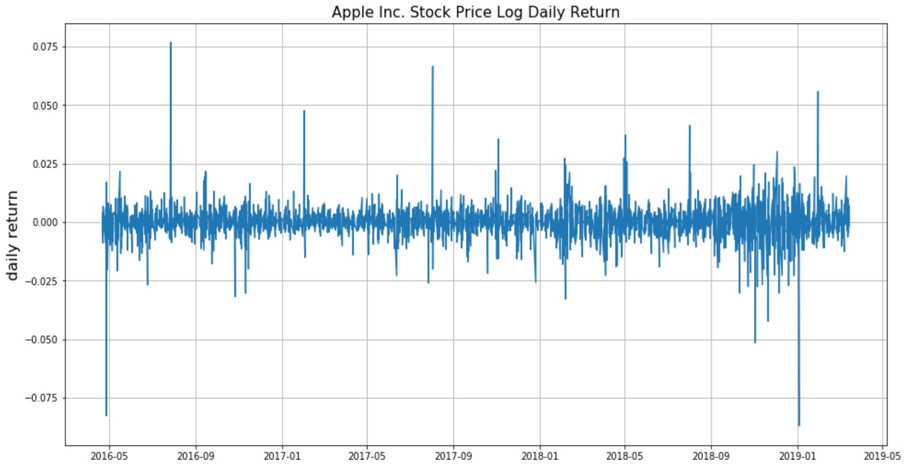
- Data Cleaning
- Data Processing
    - News article text tokenization
    - Word embedding
- Feature Engineering
    - Content Sentiment Score
    - Headline Sentiment Score
    - News Relevance(1)
    - Weighted Content Sentiment Score
    - Weighted Headline Sentiment Score
    - Competitors' stock returns (Microsoft, HP, Google)

1) Qaiser, Shahzad, and Ramsha Ali. "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents." *International Journal of Computer Applications*, vol. 181, no. 1, 2018, pp. 25–29., doi:10.5120/ijca2018917395.

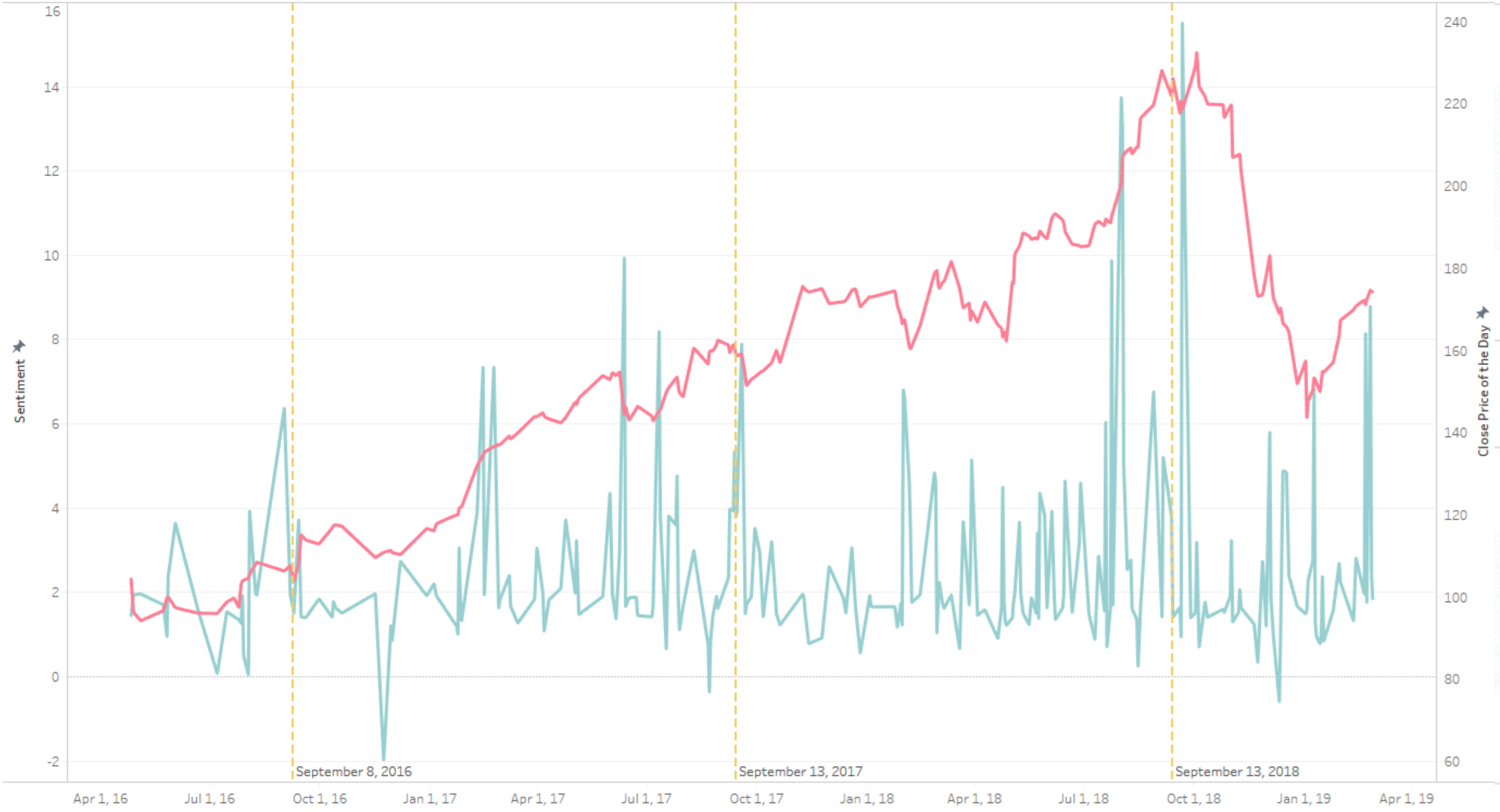
Apple Stock Price Data

- **Data Resource**
  - Yahoo Finance
- **Variables**
  - Date and time
  - Open, close, high, low price
- **Date Range**
  - April 21st, 2016 to March 15th, 2019
- **Augmented Dickey Fuller Test (ADF)**
  - Goal: check stationarity of time series
  - Conclusion: time series of log daily return is stationary

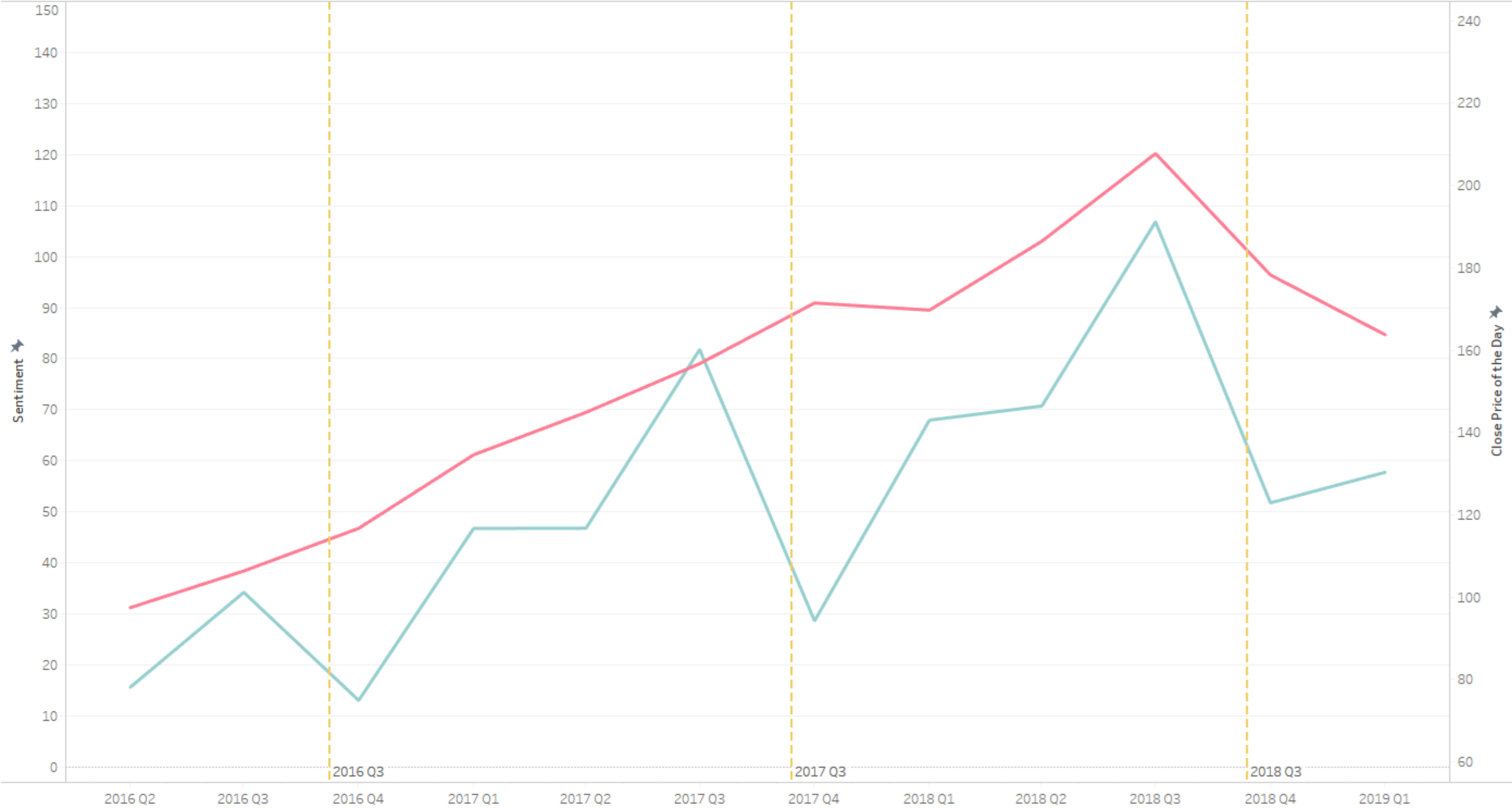




Apple Market Price and News Sentiment Score of the Day



Apple Market Price and News Sentiment Score By Quarter



Model Performance

Regression Task		
Models	RMSE (% of Return)	Correct Signs (% Correct)
Recurrent Neural Network (LSTM)	1.41	50.98
Gaussian Process Regression	1.40	55.33
Vector Autoregression (VAR)	1.39	55.56
ARMA Model	1.40	53.46

Classification Task	
Models	Accuracy (% Correct)
Gaussian Process Classification	55.34
Support Vector Machine	54.77
XGBoost	50.55

Model Improvement/ Further Research

- Can we manipulate our data differently?
    - Can we decrease the data frequency of to weekly or monthly?
        - If so, we need more data, and most of them are not publicly available
    - Can we use a more state-of-the-art model to extract some trend out of the very volatile data?
        - We tried LOESS, and it failed to extract the trend.
- Does the same effect occur when we test on other stocks?
- If we can find more up-to-date news, we can even test our hypothesis on hourly prices instead.

Question Session

Thank You