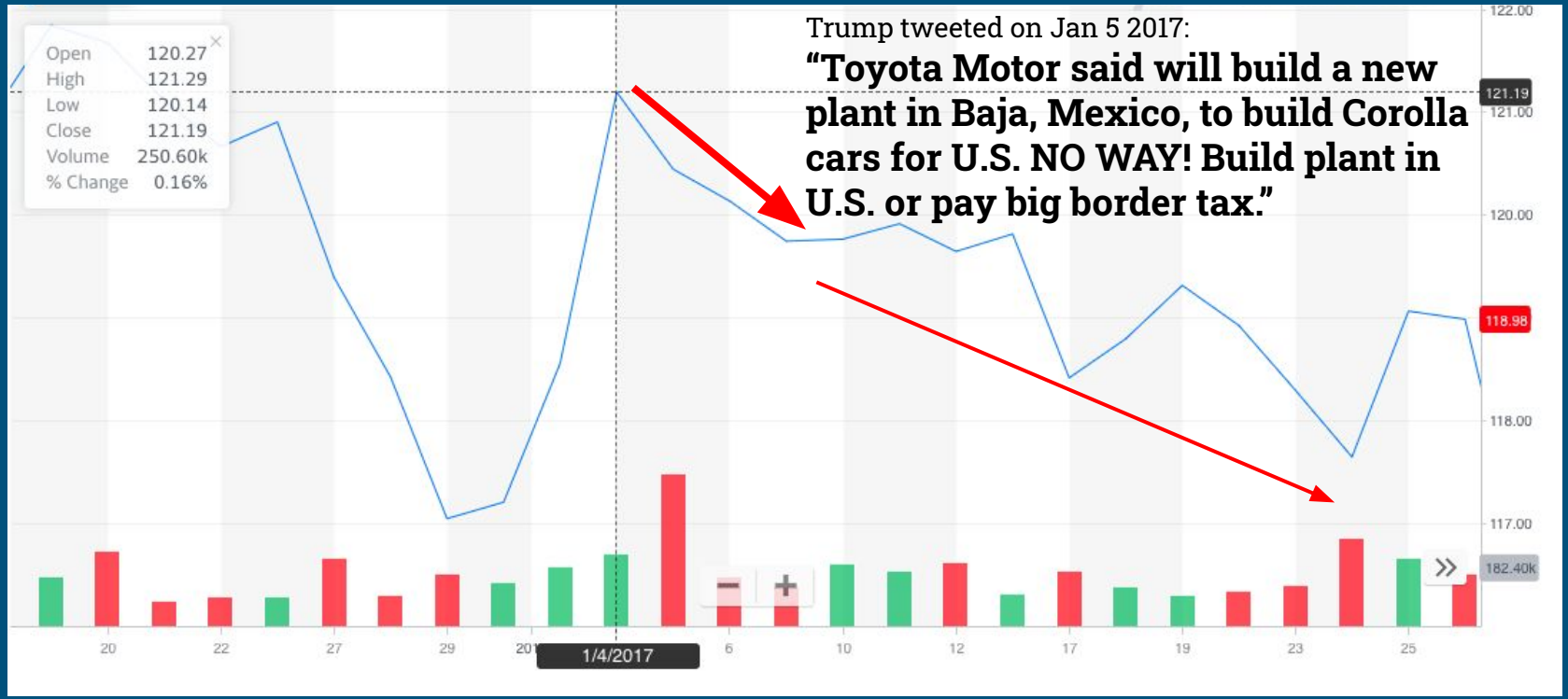
Stock Price Predictor using Trump's tweets

(Twitter Sentiment Analysis)

Audrey, DSI 13

Toyota’s share price plummets after Trump’s tweet



Problem statement

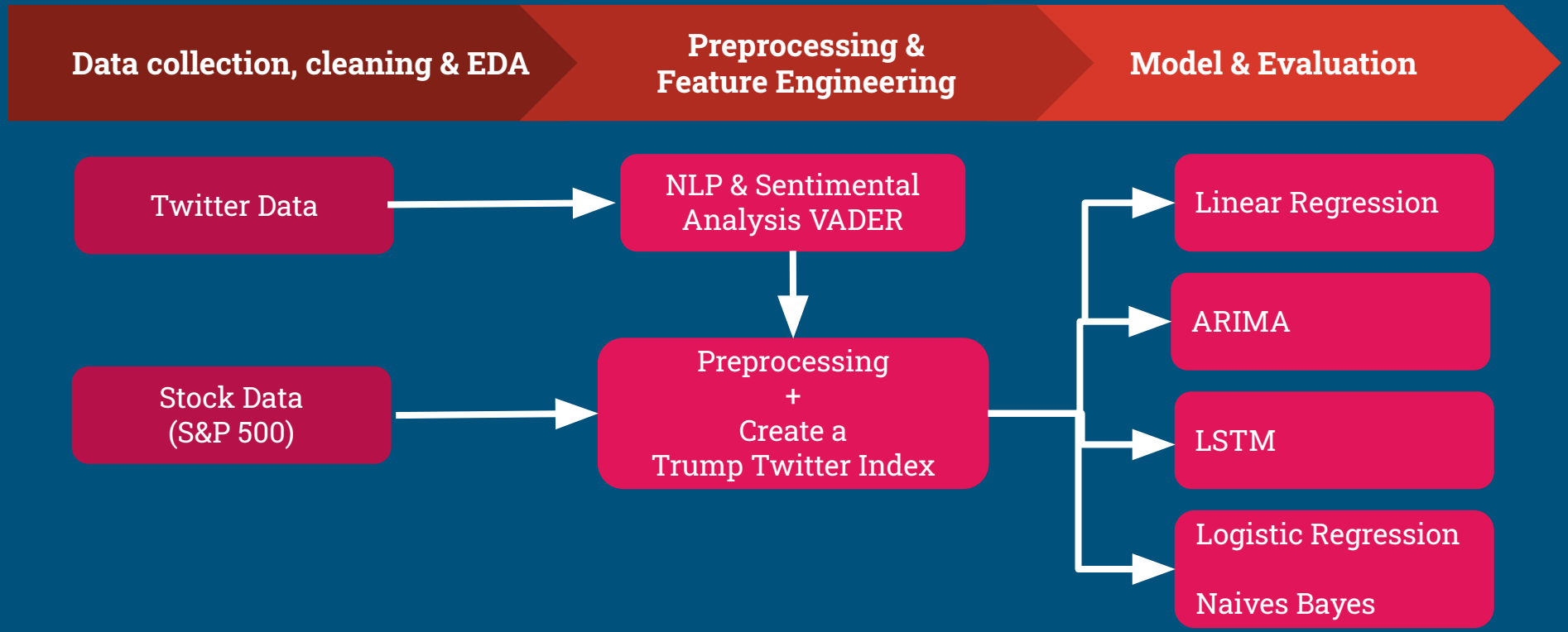
To analyse the effectiveness of using U.S. President Trump’s tweets as a signal for predicting changes in S&P 500 Index prices in the short term

Data Collection

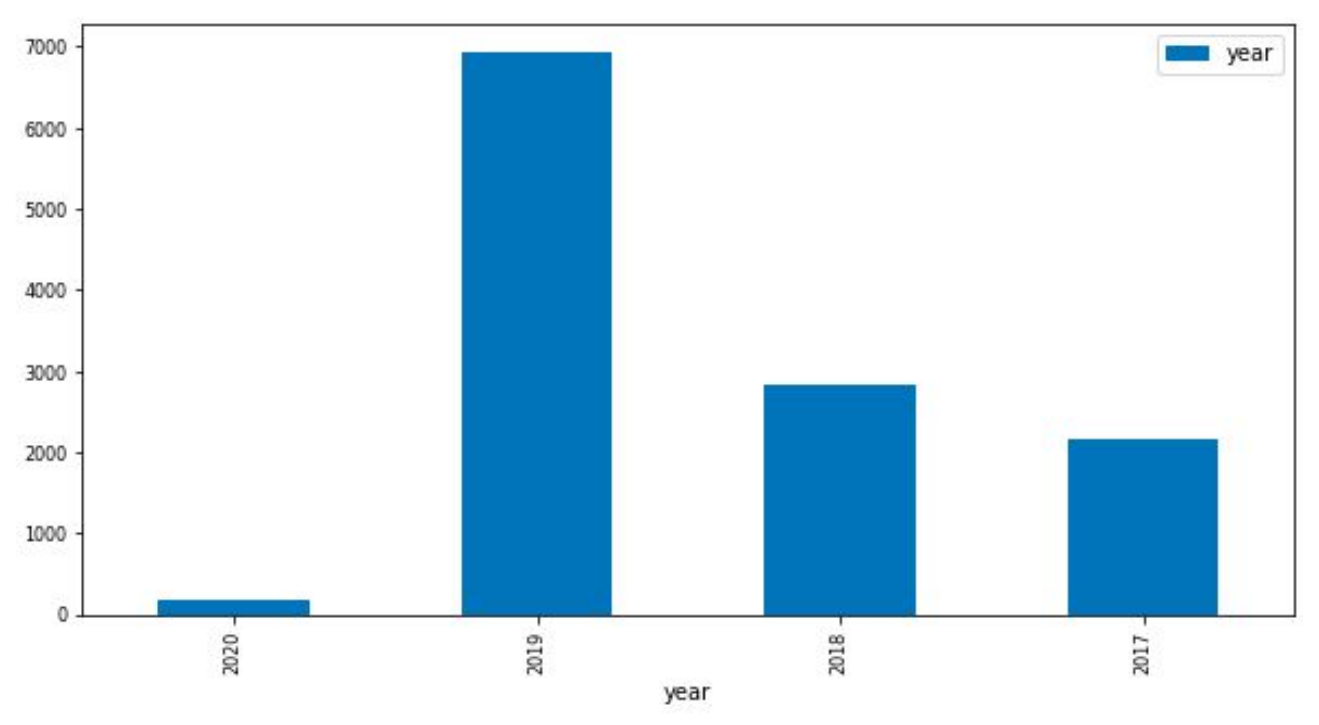
1. President Trump's tweets
   - Period: Jan 20 2017 to Jan 20, 2020
   - 13,020 tweets collected
   - Columns: source, text, created_at, retweet_count, favorite_count, is_retweet
   - Data source: Trump Twitter Archive

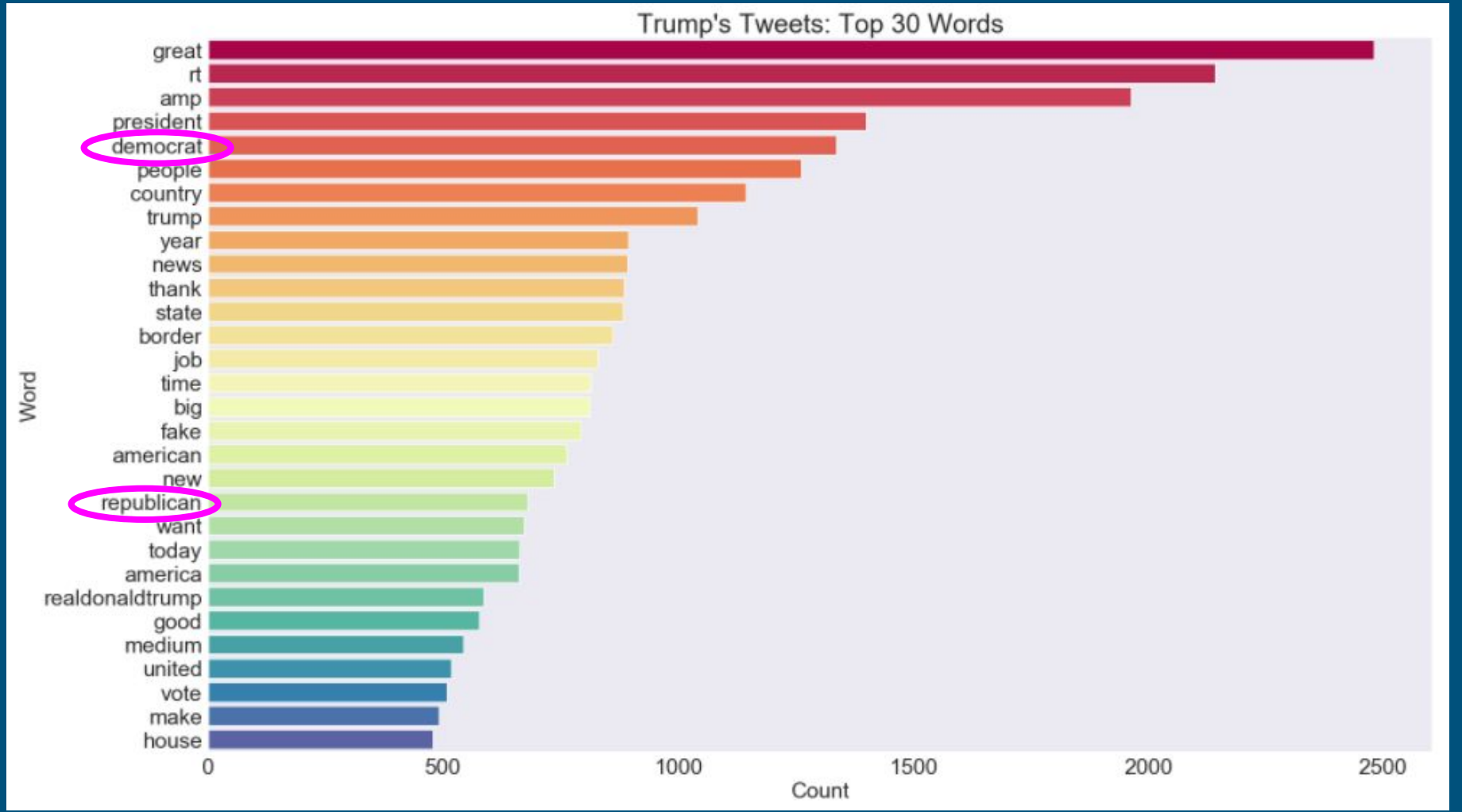
2. S&P 500 Index
   - Same period as above
   - 755 trading days
   - Columns: High, Low, Open, Close, Volume, Adj Close
   - Data source: Yahoo! Finance

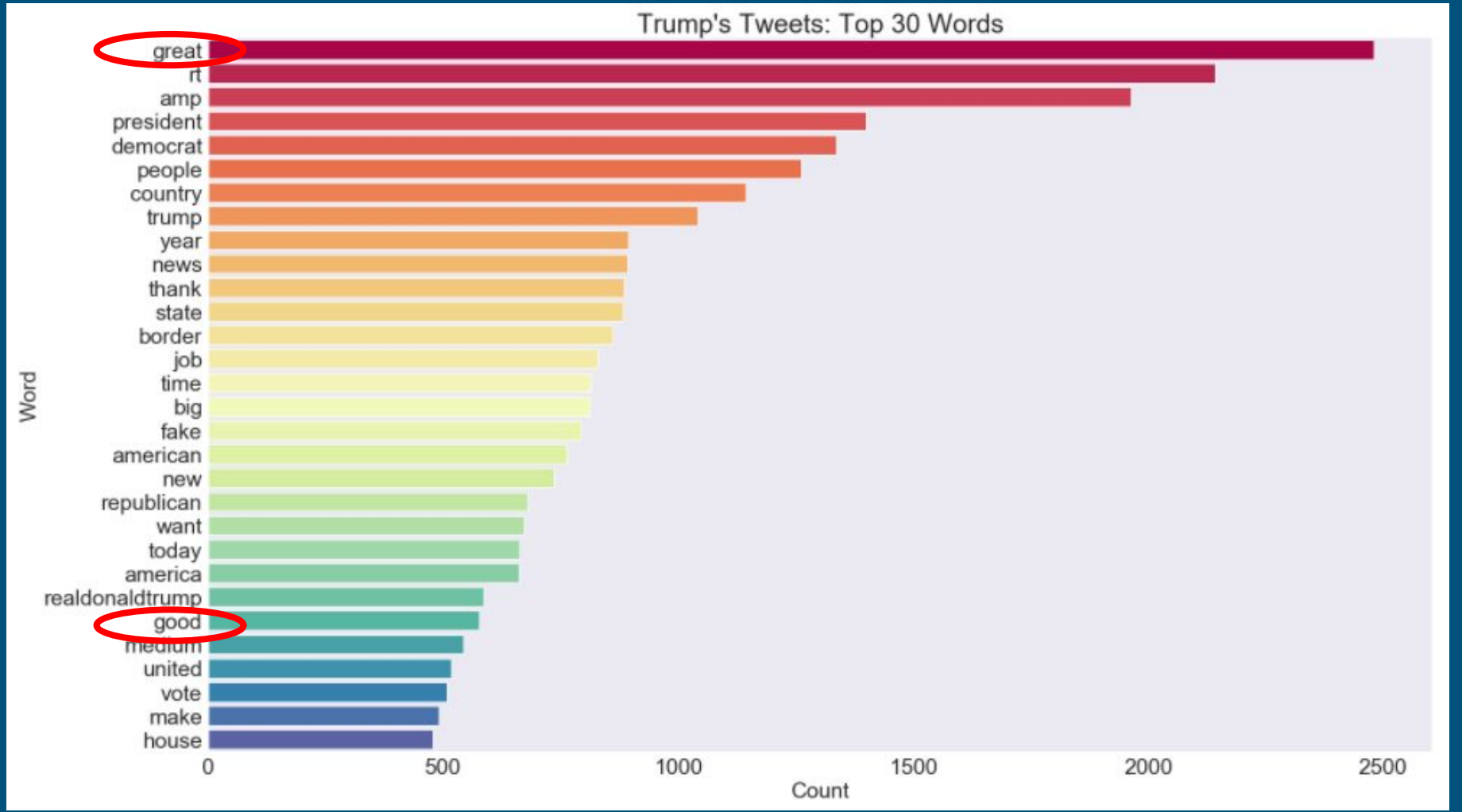
Project Framework



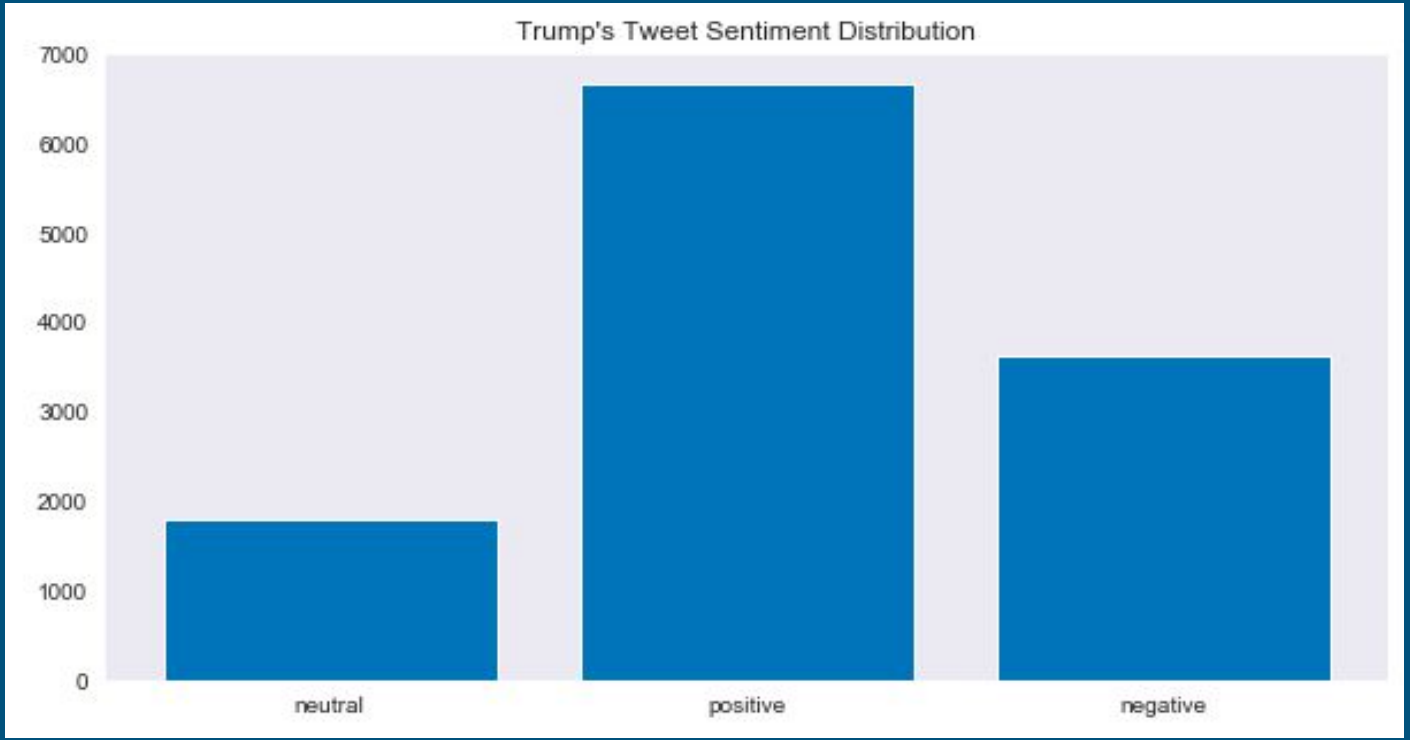
EDA - Tweets per year




Trump's Tweets: Top 30 Words
great
rt
amp
president
democrat
people
country
trump
year
news
thank
state
border
job
time
big
fake
american
new
republican
want
today
america
realdonaldtrump
good
medium
united
vote
make
house
Word
0
500
1000
1500
2000
2500
Count


Trump's Tweets: Top 30 Words
great
rt
amp
president
democrat
people
country
trump
year
news
thank
state
border
job
time
big
fake
american
new
republican
want
today
america
realdonaldtrump
good
medium
united
vote
make
house
Word
0
500
1000
1500
2000
2500
Count

Sentiment Analysis - using VADER



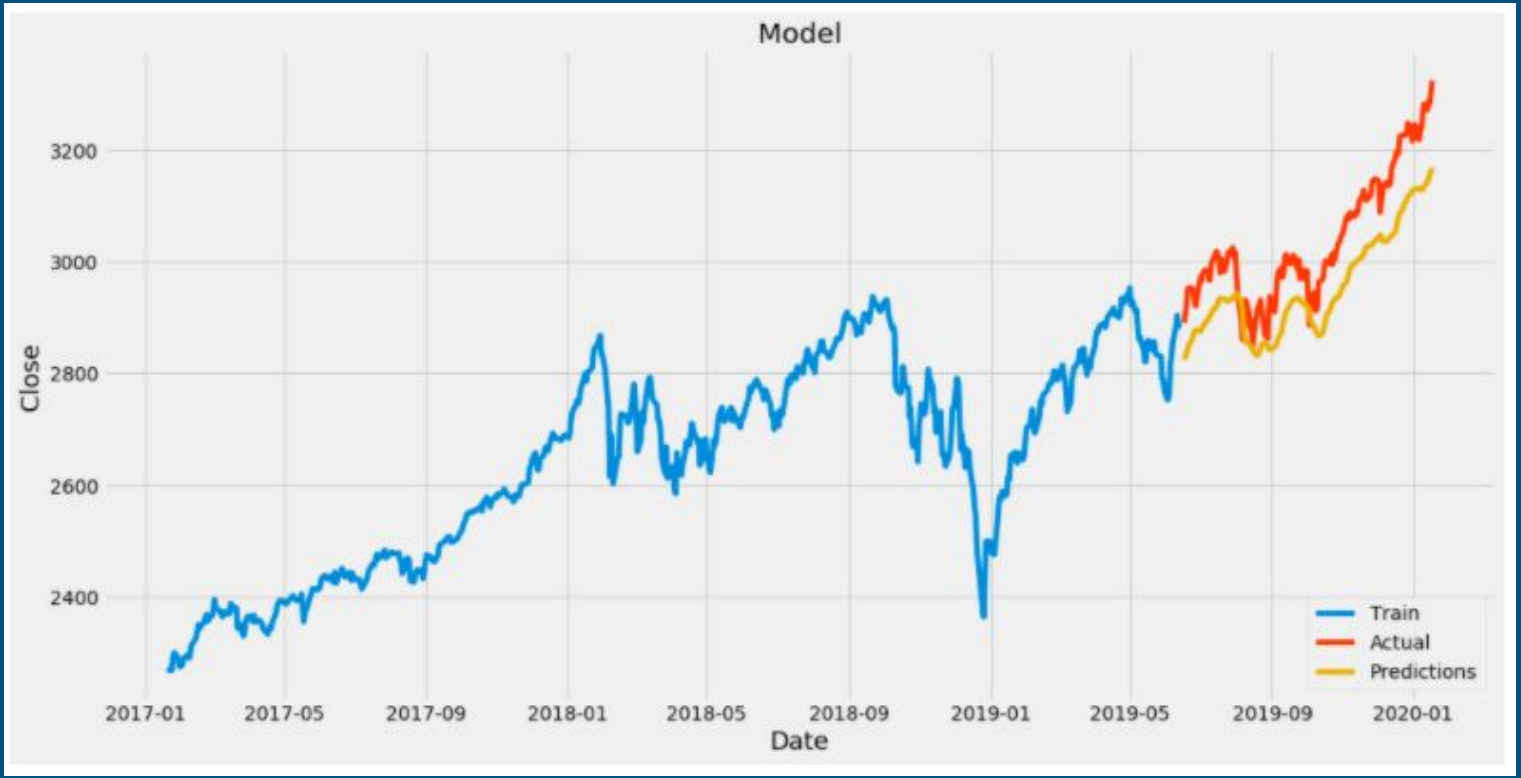
Create a Trump Twitter Index

Date	High	Low	Open	Close	Volume	Return	vader compound	Trump Twitter Index	Close_TTI
2017-01-20	2276.959961	2265.010010	2269.959961	2271.310059	3524970000	0.003366	2.2815	0.135696	2269.965939
2017-01-23	2271.780029	2257.020020	2267.780029	2265.199951	3152710000	-0.002690	0.4939	-0.056192	2267.777551
2017-01-24	2284.629883	2266.679932	2267.879883	2280.070068	3810960000	0.006565	1.6901	0.908453	2267.919940
2017-01-25	2299.550049	2288.879883	2288.879883	2298.370117	3846020000	0.008026	0.7527	0.312087	2288.893518
2017-01-26	2300.989990	2294.080078					0.4878	0.041381	2298.631683

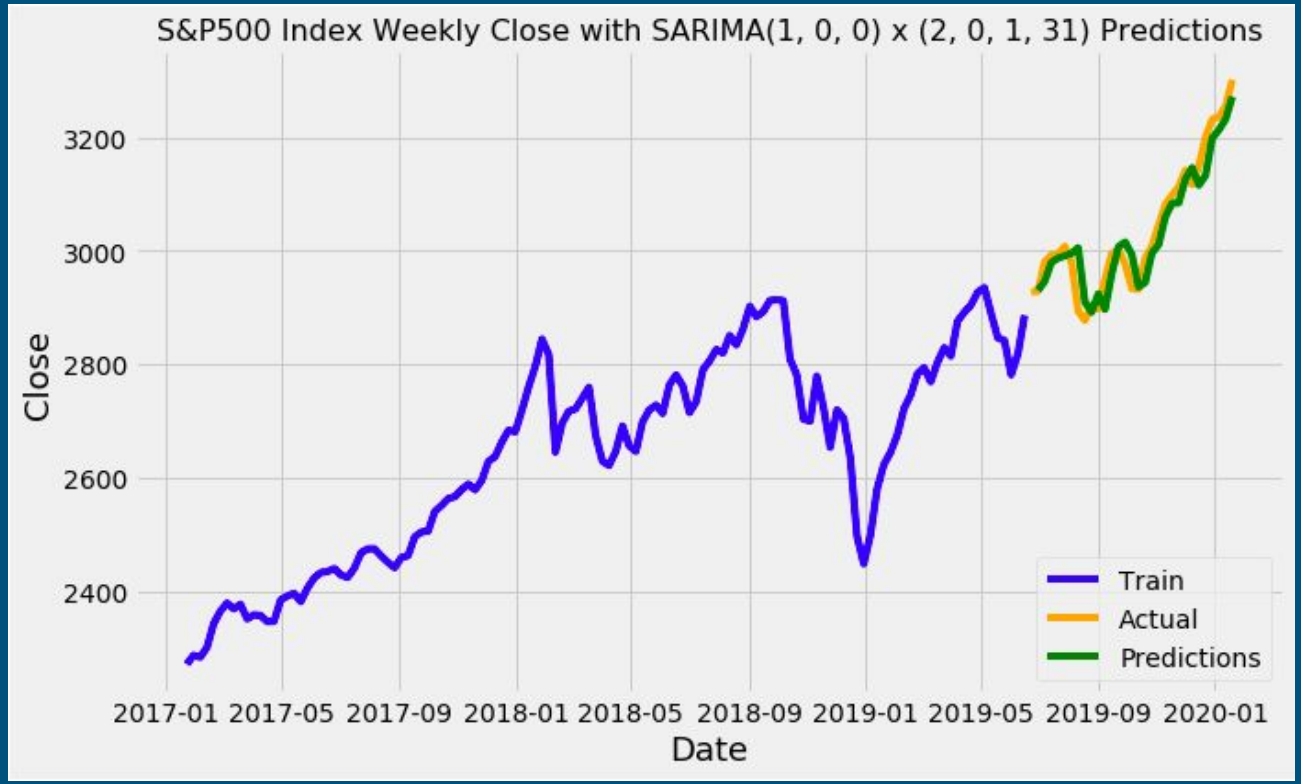
Model - RMSE summary

Model	$RMSE$
Baseline: Linear Regression	108.84
Auto Regressive Inteegration Moving Average (ARIMA)	36.61
Long Short-Term Memory (LSTM)	37.73

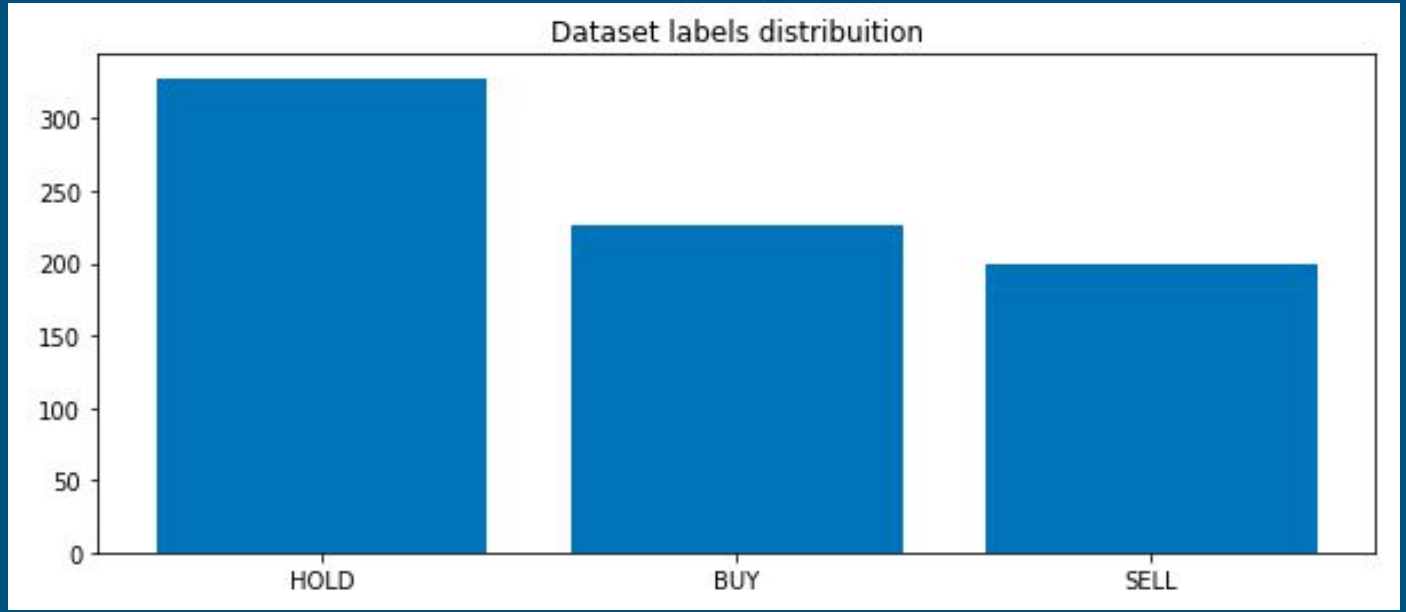
Model - LSTM


Model
Close
3200
3000
2800
2600
2400
2017-01
2017-05
2017-09
2018-01
2018-05
2018-09
2019-01
2019-05
2019-09
2020-01
Date
Train
Actual
Predictions

Model - S-ARIMA



Model - Classification - Labels



Model - Classification - Accuracy

Accuracy	Baseline	Naïve Bayes	Logistic Regression
Score	0.37	**0.53**	0.43

Model - Classification - Precision

Precision	Baseline	Naïve Bayes
Sell	0.23	**0.46**
Buy	0.36	**0.39**
Hold	0.45	**0.59**

Model - Unseen data

Mar 25, 2020 10:45:36 AM

Today is National #MedalofHonorDay. Join me in a #MomentofHonor to remember those who have earned our Nation's highest award for valor in combat. The Nation is eternally grateful to these patriots who so valiantly fought for America's values and the American way of life! https://t.co/T5JAobQ7K6 [Twitter for iPhone] [RT 18761] [❤ 74468] link

Mar 25, 2020 09:54:47 AM

Congratulations to Prime Minister Abe of Japan, and the IOC, on their very wise decision to present the Olympics in 2021. It will be a great success, and I look forward to being there! [Twitter for iPhone] [RT 26609] [❤ 176907] link

Source: http://www.trumptwitterarchive.com/archive

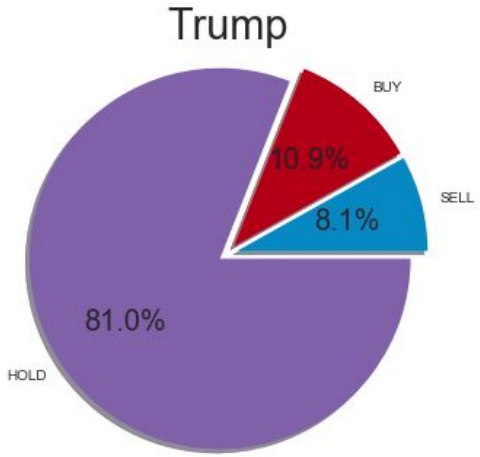
Model - Using classifier to make a prediction

```
# Predict the class for latest Trump's Tweet posted
class_label = model_GNB.predict(qr_df)[0]

prediction(class_label)
```

```
The Classifier has determined this stock to be a...
HOLD
```

```
# Visualise the Prediction probabilities for latest Trump's Tweet post
pie_stock(model_GNB.predict_proba(qr_df))
```



**Disclaimer: The value of your investments can go down as well as up and you may get back less than you originally invested. This does not offer any advice, so it's important you understand the risks, if you're unsure please consult a suitably qualified financial adviser. Tax treatment depends on your individual circumstances and rules may change.*

Limitations

- Twitter does not represent the global population
- Sample size of ~10,000 tweets may not be large enough
- VADER Sentiment Analysis tool does not accurately gauge the sentiment of sarcasm
- Sentimental Analysis is an inherently a subjective process
- Price volatility affected by minutes and seconds
- Several individual factors that affect stock price movement

Conclusion & Further work

- Perform sentiment analysis on specific feelings/attitude
  - Happy, sadness, anger etc
- Adding granularity
  - Predict prices in the next seconds or minutes
- More features
  - Interest rate, inflation, GDP growth rate, etc
- MORE DATA, MORE DATA, MORE DATA!


Questions?
Thank you :)
CNN PHOTO ILLUSTRATION