



Predicting the impact of President Trump's tweets on US stock market

Jeremy Tan Wei Ming DSI 17 Singapore



Agenda

1. Problem Statement
2. Data collection & Preprocessing
3. Topic Modelling
4. Sentiment Analysis
5. Modelling & Result
6. Backtesting trading strategy
7. Conclusion



Problem Statement

- Stock market is always unpredictable but full of opportunities
- President Trump has 87.2 million followers on Twitter
- Binary classification model that predict President Trump's trade related tweets on it's impact (positive/ negative) on the S&P 500 index
- Evaluation metrics: Accuracy
- Using these predictions to create a profitable trading strategy





Data Collection & Preprocessing

- 23,357 tweets were collected from 20 Jan 2017 till 08 Oct 2020
- Emojis, url links and non english characters were removed
- Retweets were dropped
- Tweets that have less than 3 word count were treated as noise
- Default stopwords added (thank, will, great, just, etc..)
- S&P 500 financial data collected for same period
- Tweets were labeled on the performance of the index





Topic Modelling

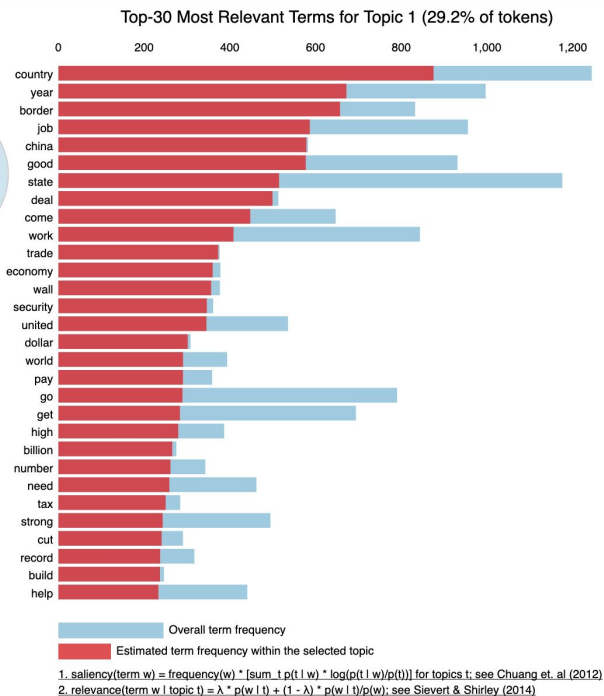
- A statistical modelling for discovering abstract topics
- Latent Dirichlet Allocation (LDA) assumes each topic is a mixture over an underlying set of words, and each document is a mixture of over a set of topic probabilities
- 3 key parameters:
 - N_components - Number of topics
 - Alpha - higher alpha documents are assumed to be made up of more topics and result in more specific topic distribution per document
 - Beta - with high beta, topics are assumed to made of up most of the words and result in a more specific word distribution per topic
- Topic that capture words relevant to trade and economy



N_components= 6

alpha=0.2

beta=0.8





Topic_1: 0.817 Topic_2: 0.015 Topic_3: 0.015 Topic_4: 0.122 Topic_5: 0.015 Topic_6: 0.015

Dominant Topic: Topic_1 (Trade/Economy)

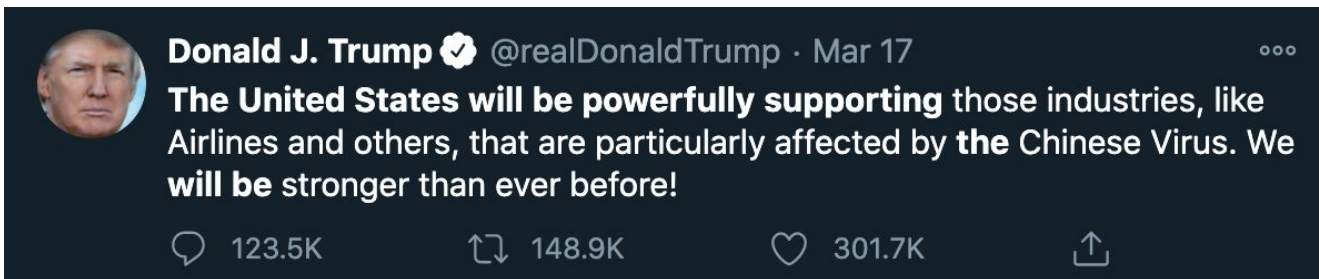


Sentiment Analysis

- Identify and extract opinions with a given text
- VADER (Valence Aware Dictionary and Sentiment Reasoner)
- Returns a positive, neutral, negative, and compound sentiment score
- Compound score range from -1 (extremely negative) to +1 (extremely positive)
- Tweets with substantial positive and negative sentiment were isolated

Positive Sentiment: Compound Score > 0.7

Negative Sentiment: Compound Score < -0.7



Dominant Topic: Trade/Economy

Neg: 0.053

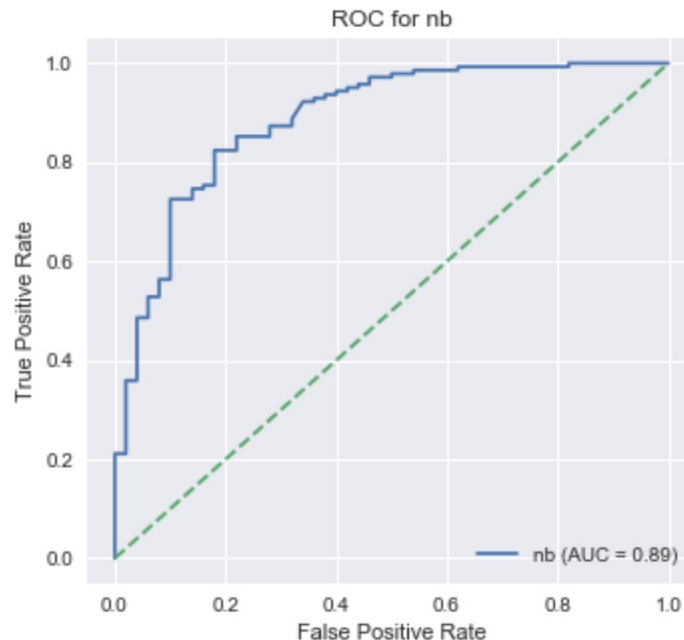
Neu: 0.639

Pos: 0.308

Compound: 0.8481

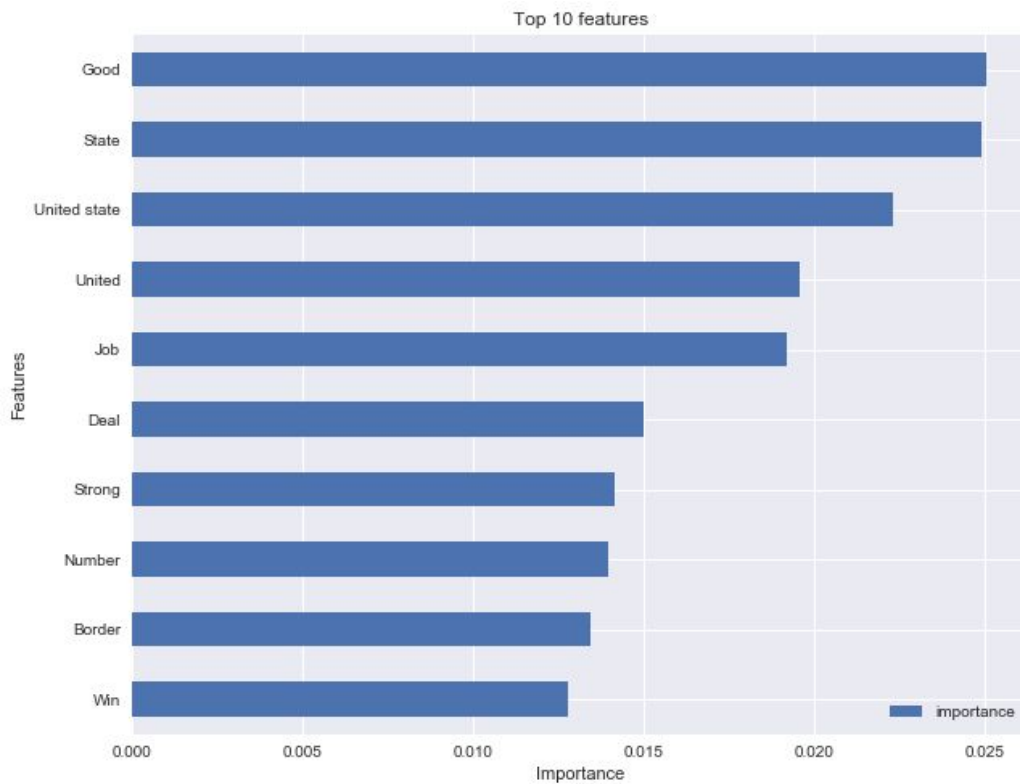
Modelling

	Train	Test
1. Logistic Regression	0.983	0.829
2. MultinomialNB	0.925	0.833
3. Random Forest Classifier	0.956	0.843
4. XGBoost Classifier	0.976	0.849
5. SVC	1.000	0.864





Top word features





Backtesting Trading Strategy

- Focusing only on buy signals (where model predicted 1, ignore if 0)
- Buy at tomorrow's open and sell at tomorrow's closing price
- Benchmark to compare is Buy and Hold strategy

72%

VS

52%

Day Trade vs Buy and Hold





Conclusion

- We are all going to be rich??
- Trending bull market will generate more buy signals
- Trump tweets coincide with FED economic data release
- Backtest is a test on the past
- Gather other data such as real time news articles, other high ranking politicians tweets/press release
- Real time market data
- Trump tweets does have certain relationship with market returns
- Further studies on NLP and market returns



THE END



Donald J. Trump 
@realDonaldTrump

If Biden somehow manages to win this election, you won't see me again, I'll simply leave the country!



Joe Biden 
@JoeBiden

Bi den