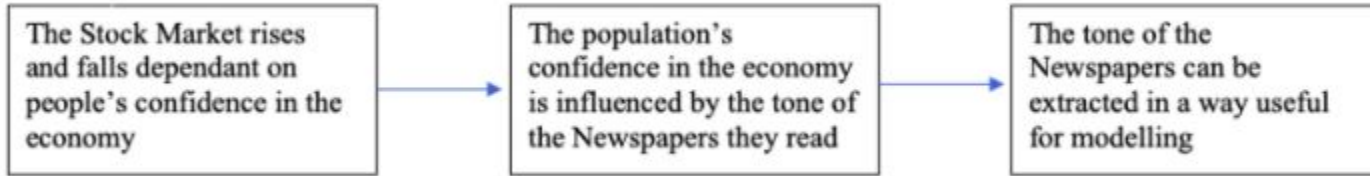


# Improving Stock Market Predictions using Natural Language Processing

Flatiron Capstone Project by Tom Ribaroff

# Outline



My Plan:

- 1 - Obtain Financial Data
- 2 - Obtain Newspaper Data
- 3 - Analyse the Newspaper Data
- 4 - Create a Baseline Models
- 5 - Add in the Newspaper Data and see if it improves our Models

# Data Collection

	Open	High	Low	Close	Volume
Date					
2010-01-04	5412.88	5500.34	5410.82	5500.34	7.392485e+08
2010-01-05	5500.34	5536.38	5480.71	5522.50	1.073713e+09
2010-01-06	5522.50	5536.48	5497.65	5530.04	9.274415e+08
2010-01-07	5530.04	5551.66	5499.80	5526.72	1.109908e+09
2010-01-08	5526.72	5549.25	5494.79	5534.24	9.713993e+08

- The Financial Data was easily accessible, thanks to Yahoo Finance
- The Newspaper information was harder. I used the APIs (official portals to request information) the newspapers developed.
- I wrote functions to filter through all the responses and give me the data I needed

```
1 daily_guardian_headlines("2020-05-23", api_key = api_key)

executed in 440ms, finished 10:19:45 2020-06-18

['New York Stock Exchange traders hope to prove they are still a big noise',
 'Jaguar Land Rover in talks over 'f1bn-plus' state loan',
 'Why the Treasury's panic over debt, when borrowing costs next to nothing?',
 'Derby was once ahead of its time, but reliance on making leaves it exposed',
 'Lockdown laptops: affordable options for those stuck at home',
 'Email leak reveals identities of sexual abuse victims in Northern Ireland',
 'Five appear in court charged with murder of teenager Aya Hachem',
 'I was told to stop Julian Assange if he tried to flee': on the beat with the UK's volunteer police']
```

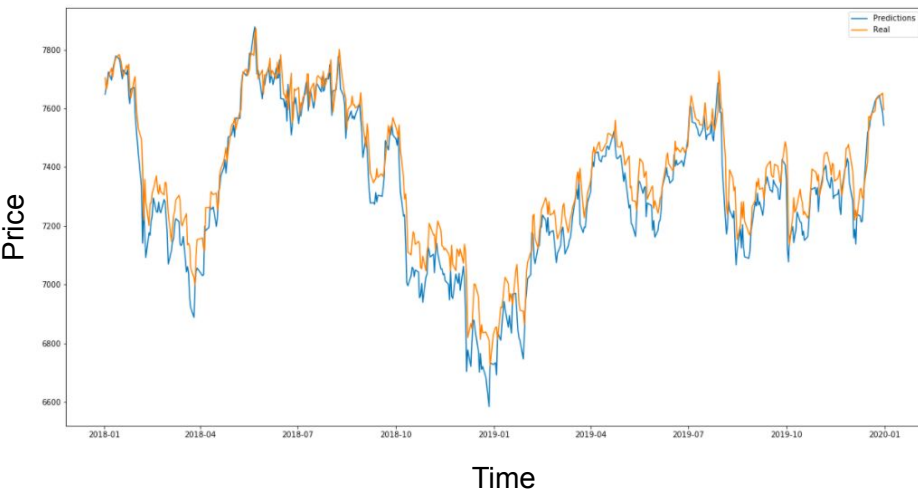
# Data Explorations

- LSTM Models can only take numerical values as input, so the Stock Prices are easy to use, but we had to find a way to put our Newspaper Headlines in numerical form
- TextBlob is a tool which takes in phrases we give it, and returns a number measuring the 'polarity' of the phrase. -1 very bad, 1 very good.
- We ran all of our headlines through TextBlob, and saved the polarity score of the news that day for use in our later modelling.

Guardian		Guardian_score_polarity
Dates		
2010-01-01	['Green technology to be harnessed by top firm...	0.136364
2010-01-02	['Wi-Fi won't travel through your house? It mi...	0.006250
2010-01-03	['Place your bets: which stocks and shares sho...	0.100000
2010-01-04	['How solid is Northern Rock without governmen...	0.136364
2010-01-05	['Fears of higher interest rates as leading bo...	0.114583

# Baseline Modelling

Prediction Part of our Baseline FTSE Model

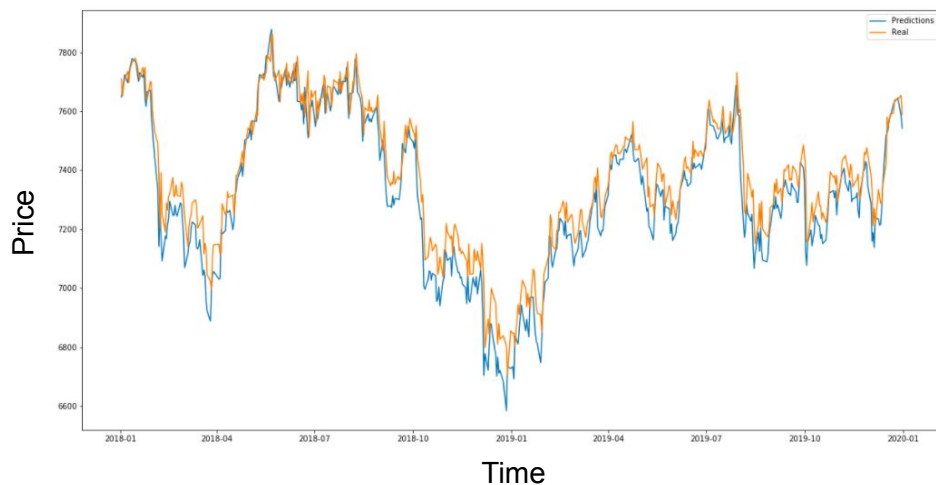


- I first ran the modelling using just the past price history as data for the model to be trained on.
- The green is what actually happened, and the orange is what my model predicted would happen.
- This demonstrates how powerful LSTM modelling is!

# Modelling with the Newspaper Data

- Next step was to incorporate the Newspaper Data to see the effect on the Modelling
- I repeated the same conditions for Modelling so we could have a more genuine comparison
- Our new Model looks to be doing some great prediction work, but we need to have a closer look at the MSE scores to be sure

Prediction Part of our New FTSE Model



# Conclusions

Before Headline data was included: loss: 4.2421e-04 - val\_loss: 7.4619e-04

After Headline data was included: loss: 3.9567e-04 - val\_loss: 7.3004e-04

- Our FTSE Model has improved, if only a little bit! Unfortunately there weren't tangible improvements to our S&P Model.
- These results are evidence that this method could improve future models, and therefore merit further investigation
- LSTM Models need rerunning as you get slightly different results each time, so this is evidence that a small investment in cloud computing power would be worth it to help establish how helpful this method really is

# Future Work

- 1 - Improve Quality of Newspaper Data
- 2 - Improve Quality of Modelling
- 3 - Experiment with different kinds of Data
- 4 - Make predictions for specific stocks, rather than the whole FTSE/S&P
- 5 - Scrape Data from Different Newspapers
- 6 - Make predictions for Stock Market Crashes, rather than just in times of stability