

# Deep learning for relative stock performance prediction

Ruben Contesti, Sebastian Hurubaru, Jun Yan



## Motivation

- Stock movement prediction is always important.
- Over recent years several Deep learning methods have been applied.
- Unstructured data are also significantly correlated to the stock market

In this work we combine the Tweet data and some traditional finance data to predict the relative performance of Apple, using a deep neural network structure based on wavelet transform, stacked auto-encoder and LSTM/GRU.

## Dataset

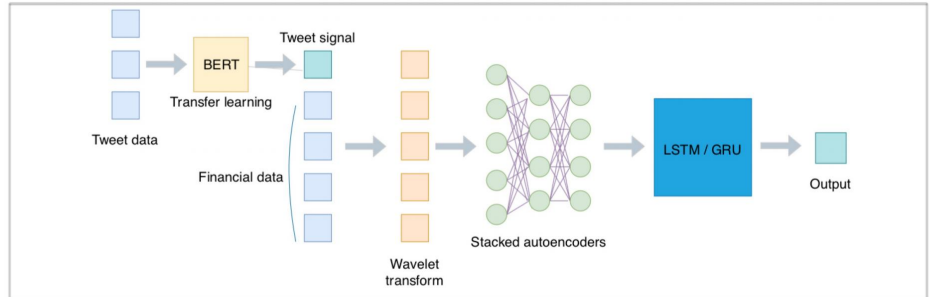
We have gathered a data set including daily data from 2009 to 2018. The data set containing 20 features in three categories:

- Historical trading data for Apple (APPL) and SPY, including the Open, High, Low, Close, Adj Close prices, relative performance and Volume.
- Macroeconomic indicators: interest rate and exchange rate. Effective Federal Funds Rate (DFF) and the US dollar index (DX-Y.NYB).
- Tweet data: Top tweets truncated to a 100 tweets a day.

Goal: predict the relative performance, defined as:

$$y_i = \frac{APPL_{i+1}/APPL_i}{SPY_{i+1}/SPY_i} - 1.$$

## Method



## Results

| Model                   | RMSE     | Theil U  | MAPE     | Kaggle score | Profitability |
|-------------------------|----------|----------|----------|--------------|---------------|
| LSTM                    | 0.016772 | 0.01274  | 1.677205 | 0.343304     | 10.18         |
| Wavelet LSTM            | 0.016558 | 0.012582 | 1.655797 | 0.301474     | 12.57         |
| SAE Wavelet LSTM        | 0.016128 | 0.01226  | 1.612825 | 0.347782     | 17.04         |
| Tweets SAE Wavelet LSTM | 0.016002 | 0.012164 | 1.600169 | 0.347783     | 17.06         |

| Model                  | RMSE     | Theil U  | MAPE     | Kaggle score | Profitability |
|------------------------|----------|----------|----------|--------------|---------------|
| GRU                    | 0.016241 | 0.012372 | 1.624109 | 0.299805     | 16.90         |
| Wavelet GRU            | 0.018551 | 0.014131 | 1.855082 | 0.3453428    | 11.23         |
| SAE Wavelet GRU        | 0.016203 | 0.012315 | 1.620265 | 0.344048     | 17.03         |
| Tweets SAE Wavelet GRU | 0.016249 | 0.01235  | 1.624864 | 0.344049     | 17.06         |

## Conclusion

In general the best results are a combination of using wavelet transformation, Stacked Auto Encoders, Tweets sentiment and GRU/LSTM, with LSTM generating slightly better results than GRU.

Future work:

- Weigh each tweet by the importance of the tweeter account.
- Analyzing other unstructured data, such as news data.