Stock News Prediction

By: Vincent Welsh

Outline

- Business Definition
- Data
- Methodology
- Results
- Conclusions
  - For a live dashboard an XGBOOST model is the best fit for implementation.
  - Add news sentiment to stock predictions.
  - Stock volatility adds noise to prediction models. To mitigate this, lower the test size.
- Further Work
- Questions

Business Definition

Business Problem: Stock prices are a a very important metric because they represent the value of publicly traded companies in the eyes of society. There is also a strong financial incentive to have a higher stock value for current shareholders because stocks can be liquidated into currency. Stock prices fluctuate based on the supply and demand of stocks associated with a particular company. If stock price can be predicted it could bring these business values:

- Increased confidence of return when investing in a company's stock

- Increase the confidence of current shareholders regarding stock's direction which can influence their decision to buy/sell your stock

- Determine key metrics that drive a stock price up or down

- Forecasting stock crashes will lead to the minimization of capital loss

Data

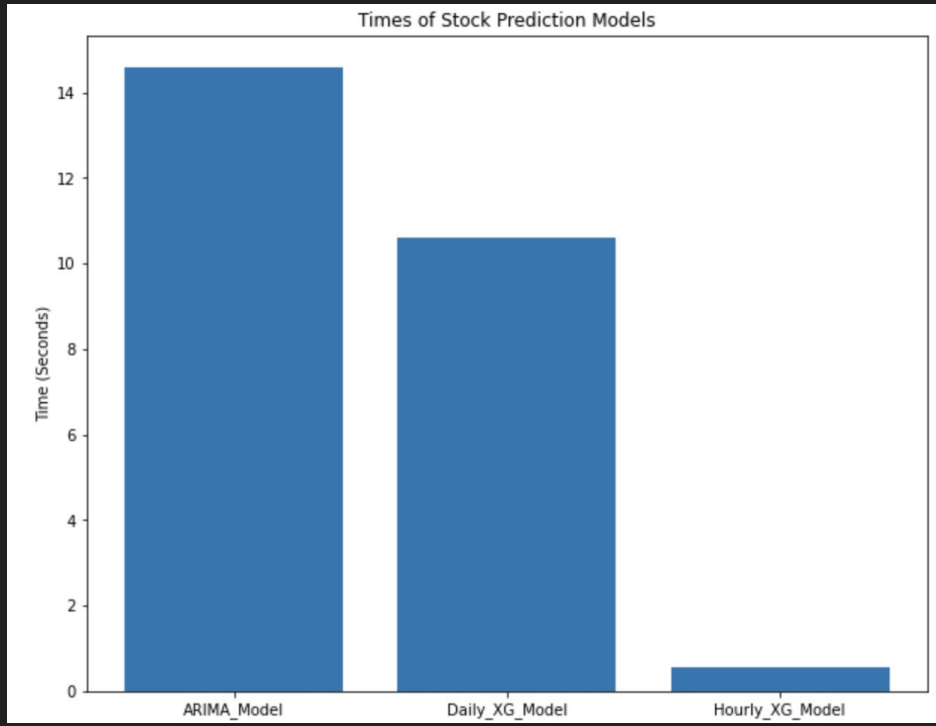
- The data for this dataset comes from https://fmpcloud.io/
- The data included in the dashboard are
  - Annual Revenue
  - Earnings Per Share
  - Insider Trades
  - Recent News
  - Stock Price Over Time
  - Recent News + Sentiment Scores

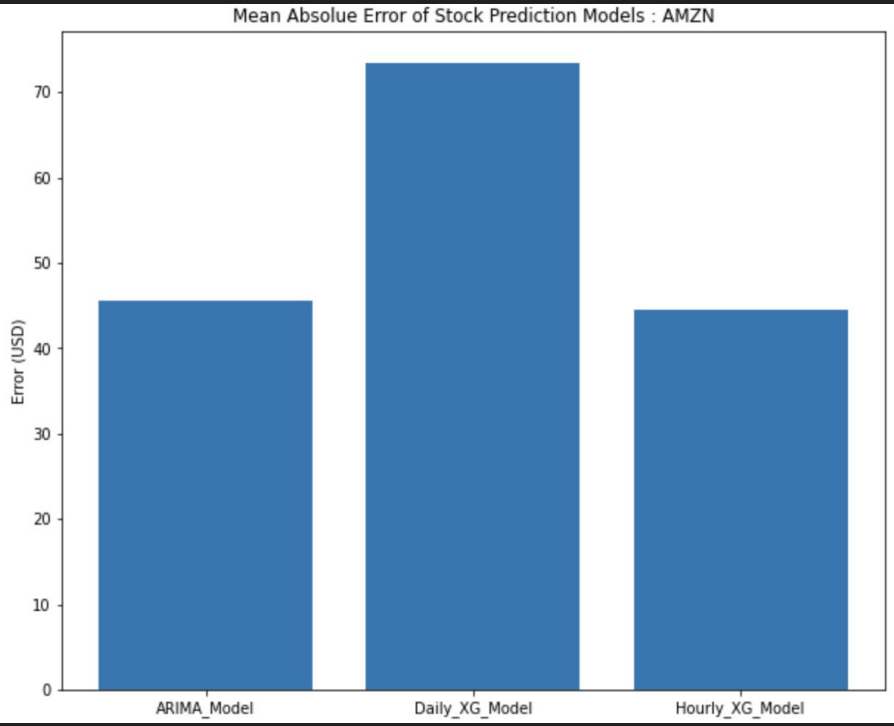
Methodology

- General Model Daily Prediction
  - XGBOOST
  - Contains 3 day previous values for:
    - High Price
    - Close Price
    - Change Over Time Decimal
    - Volume
  - Previous day news sentiment

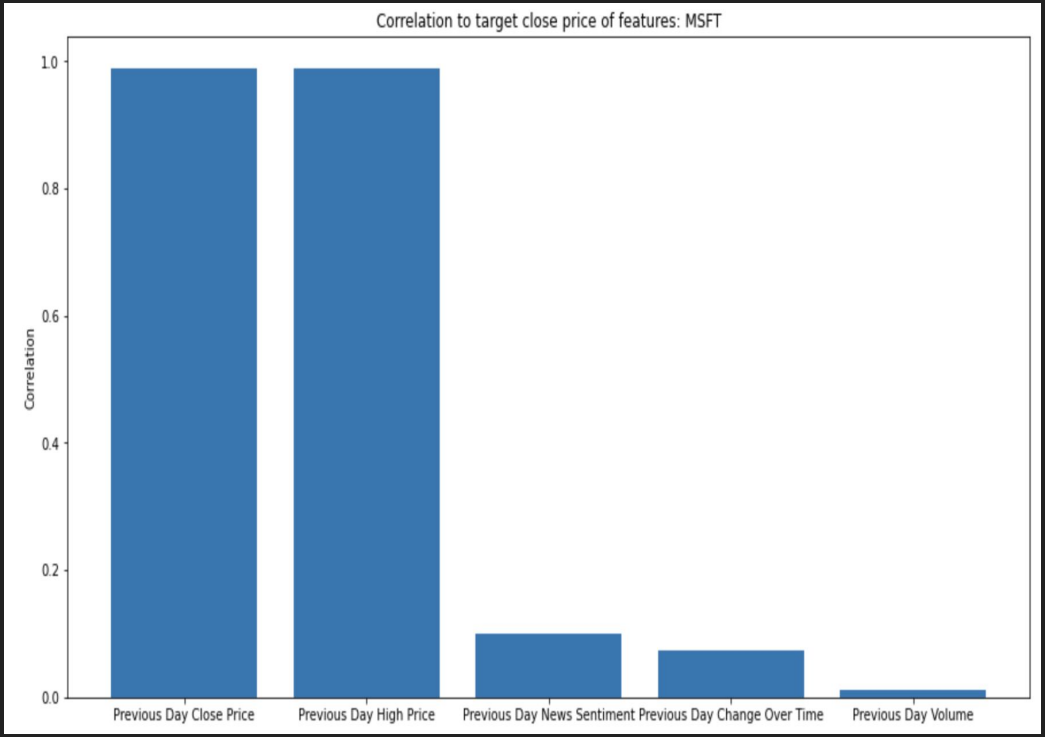
- General Model Hourly Prediction
  - XGBOOST
  - Contains 3 hourly previous values for:
    - High Price
    - Close Price
    - Volume
    - Low Price

Results: Model Comparison



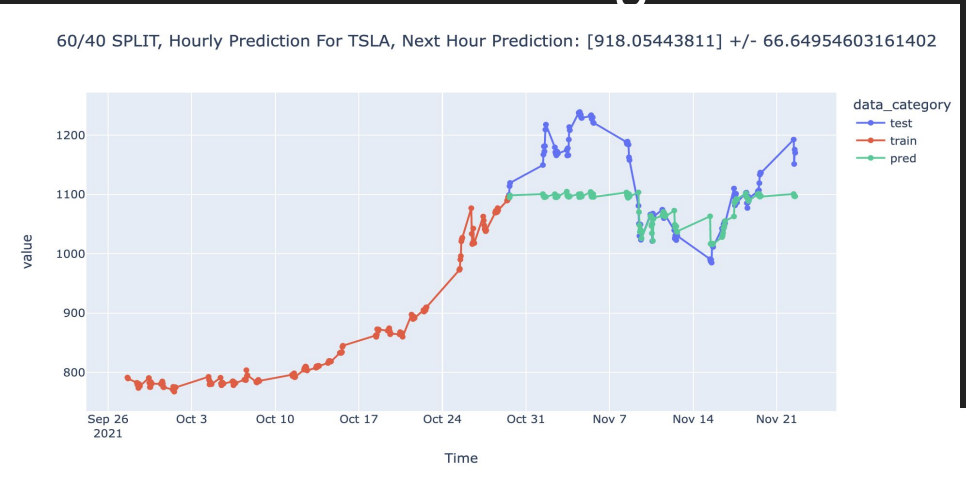


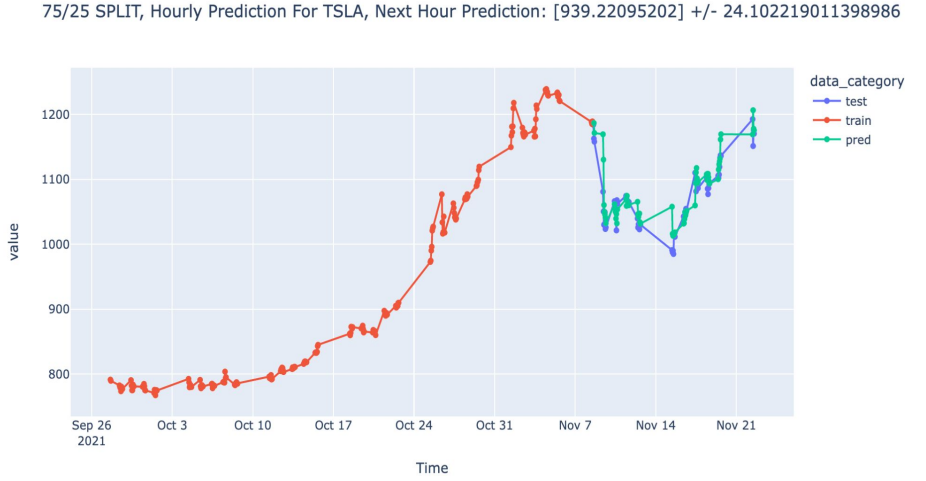
Results: News Sentiment



	importance	feature_name
0	0.472796	Close_1day
1	0.293361	Close_3day
2	0.148589	Close_2day
3	0.049592	High_1day
4	0.012810	High_2day
5	0.004726	Cot_3day
6	0.004395	High_3day
7	0.003072	Vol_3day
8	0.002869	Cot_1day
9	0.002842	Vol_1day
10	0.001985	Vol_2day
11	0.001796	News_Sentiment_1day
12	0.001170	Cot_2day

Results: Overfitting





Conclusions

- For a live dashboard an XGBOOST model is the best fit for implementation.
  - Faster than ARIMA/NN
- Add news sentiment to stock predictions.
- Stock volatility adds noise to prediction models. To mitigate this, lower the test size.

Further Work

- Provide with more time, I would :
  - Investigate if/how insider training affects model performance
  - Implement analyst ratings to the prediction model
  - Investigate actively raising stocks to determine causers of such movement

Questions?

About the author: Strongly self-motivated data science professional with 3 years of programming experience. Recent graduate of a full time data science bootcamp and professional experience as a language analyst for 3 years. Passionate about data, STEM, business intelligence and am eager to learn and grow.

LinkedIn: https://www.linkedin.com/feed/

Github: https://github.com/Eucalyptusss

Email: jvincentwelsh99@hotmail.com