Stock Market Prediction

William Knoff, Ian Curd, Hesham Elshafey, Omar Khalil

Our Idea:

Use a recurrent neural network (RNN) to predict whether a stock will go up or down the next day using S&P 500 stocks 2020 - 2022

Our Goal:

Explore the feasibility of beating the market

Literature

Goyal, A., Choudhary, A., Malik, D., Baliyan, M.S., Rani, S. (2023). **Implementing and Analysis of RNN LSTM Model for Stock Market Prediction**. In: Tiwari, S., Trivedi, M.C., Kolhe, M.L., Singh, B.K. (eds) Advances in Data and Information Sciences. Lecture Notes in Networks and Systems, vol 522. Springer, Singapore. https://doi.org/10.1007/978-981-19-5292-0_22

Sujatha, R., Abhyankar, V., Gehlot, A., Gupta, P., Subramaniam, S. (2021). **Stock Market Trend Prediction Using Regression Model, RNNs, and Sentiment Analysis**. In: Komanapalli, V.L.N., Sivakumaran, N., Hampannavar, S. (eds) Advances in Automation, Signal Processing, Instrumentation, and Control. i-CASIC 2020. Lecture Notes in Electrical Engineering, vol 700. Springer, Singapore. https://doi.org/10.1007/978-981-15-8221-9_27

Houssein, E.H., Dirar, M., Hussain, K., Mohamed, W.M. (2021). **Artificial Neural Networks for Stock Market Prediction: A Comprehensive Review**. In: Oliva, D., Houssein, E.H., Hinojosa, S. (eds) Metaheuristics in Machine Learning: Theory and Applications. Studies in Computational Intelligence, vol 967. Springer, Cham. https://doi.org/10.1007/978-3-030-70542-8_17

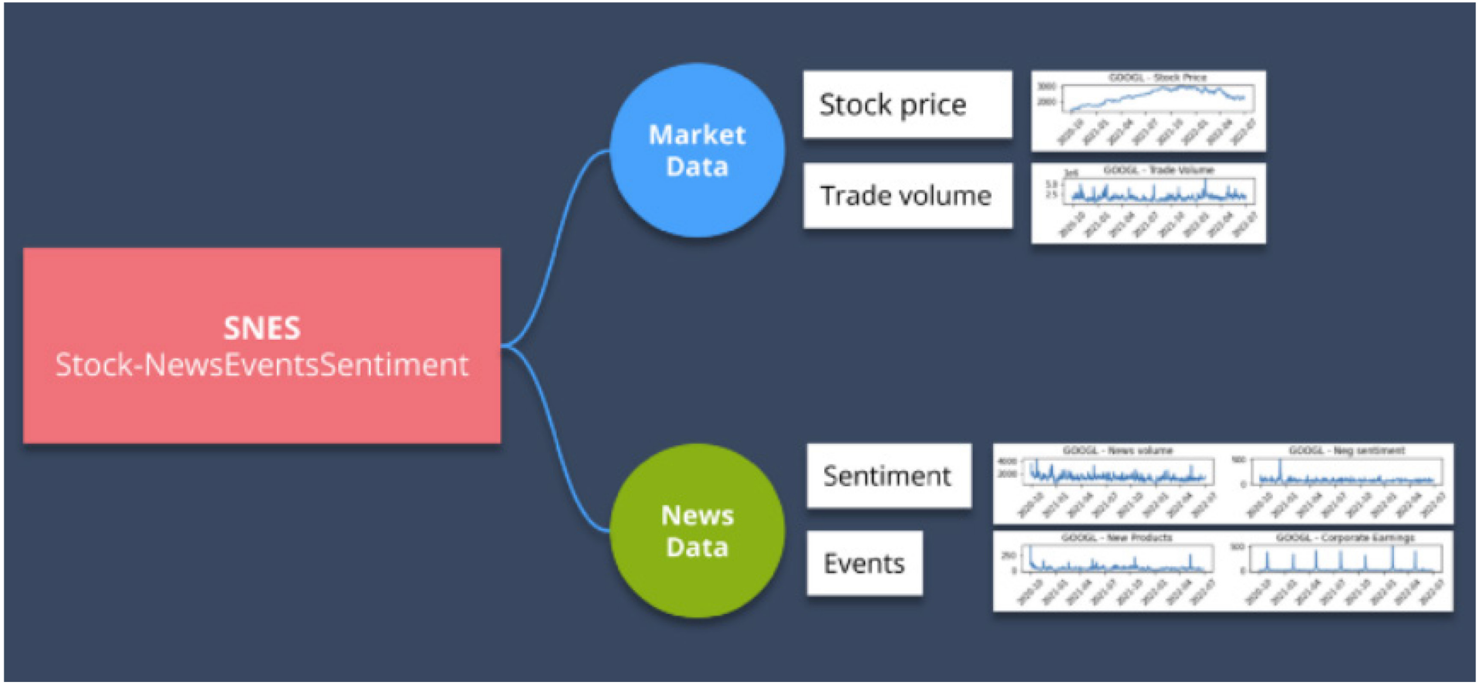
Our Old Dataset

- Date - in format: mm-dd-yyyy
- Open - price of the stock at market open
- High - Highest price reached in the day
- Low - Lowest price reached in the day
- Close - price of the stock at market close
- Volume - Number of shares traded
- Name - the stock's ticker name

** All prices in $ USD

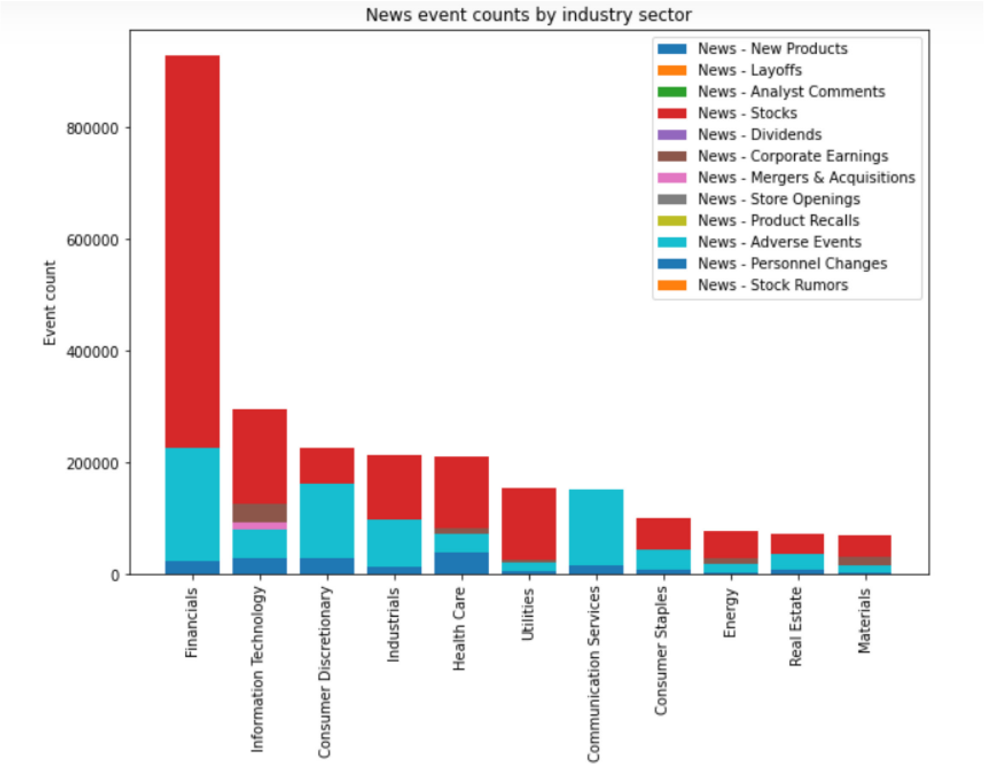
	date	open	high	low	close	volume	Name
0	2/8/2013	15.07	15.12	14.63	14.75	8407500	AAL
1	2/11/2013	14.89	15.01	14.26	14.46	8882000	AAL
2	2/12/2013	14.45	14.51	14.10	14.27	8126000	AAL
3	2/13/2013	14.30	14.94	14.25	14.66	10259500	AAL
4	2/14/2013	14.94	14.96	13.16	13.99	31879900	AAL

Our New Dataset:
Stock- News Events Sentiment (SNES) 1.0

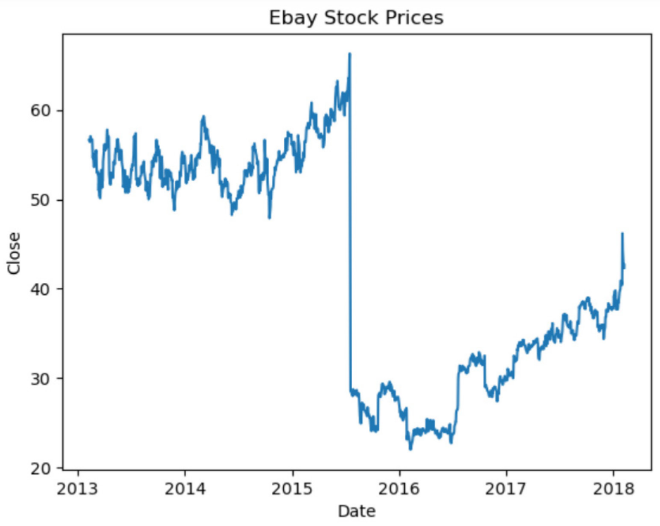


Sentiment Data

Reported Events	Reported Financials	Reported Opinions
Personnel Changes	Mergers & Acquisitions	
Product Recalls	Dividends	Analyst Comments
New Products		
Store Openings	Corporate Earnings	Stock Rumours
Adverse Events	Layoffs	



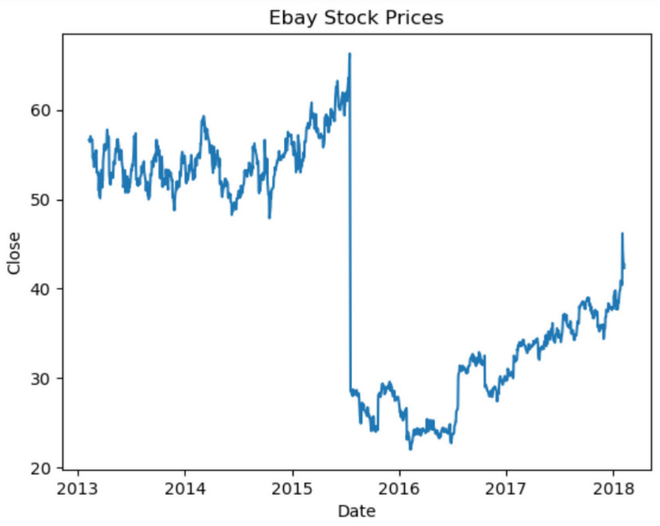
Exploratory Data Analysis: Stock Split Issue



2.375x Stock Split in 2015

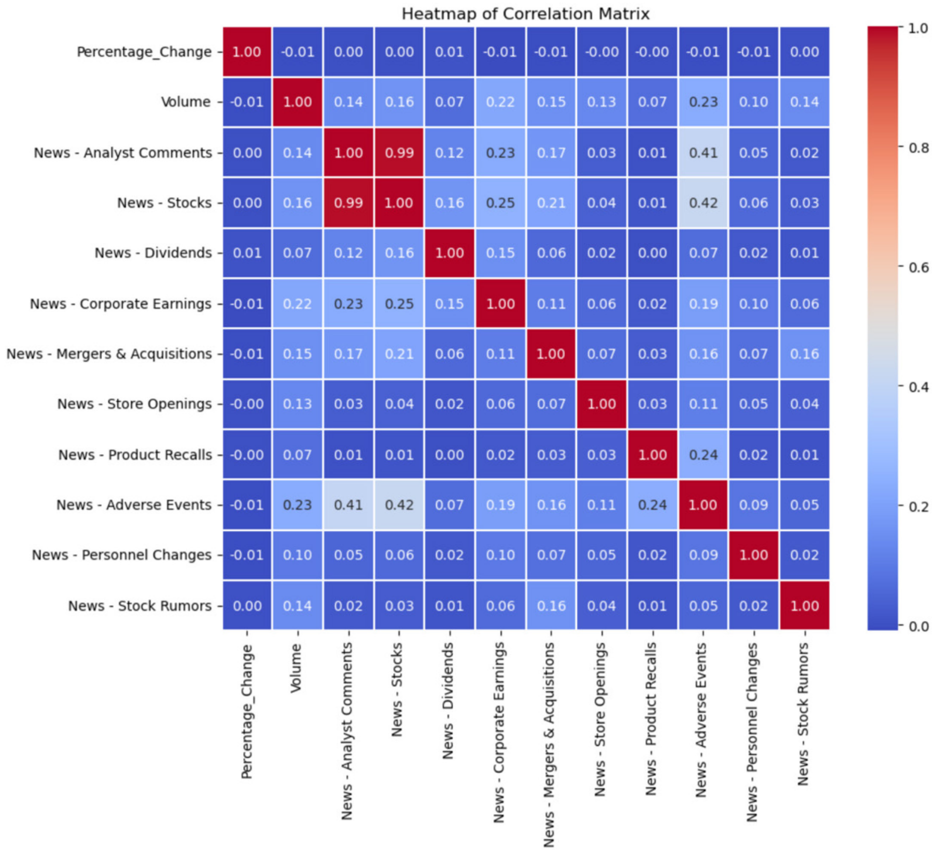
- An issue of new shares in a company to existing shareholders in proportion to their current holdings.
- Although the market value of the company remains the same, the price of the stock itself sharply changes by the ratio it is split
- Would reduce accuracy

Exploratory Data Analysis: Stock Split Issue

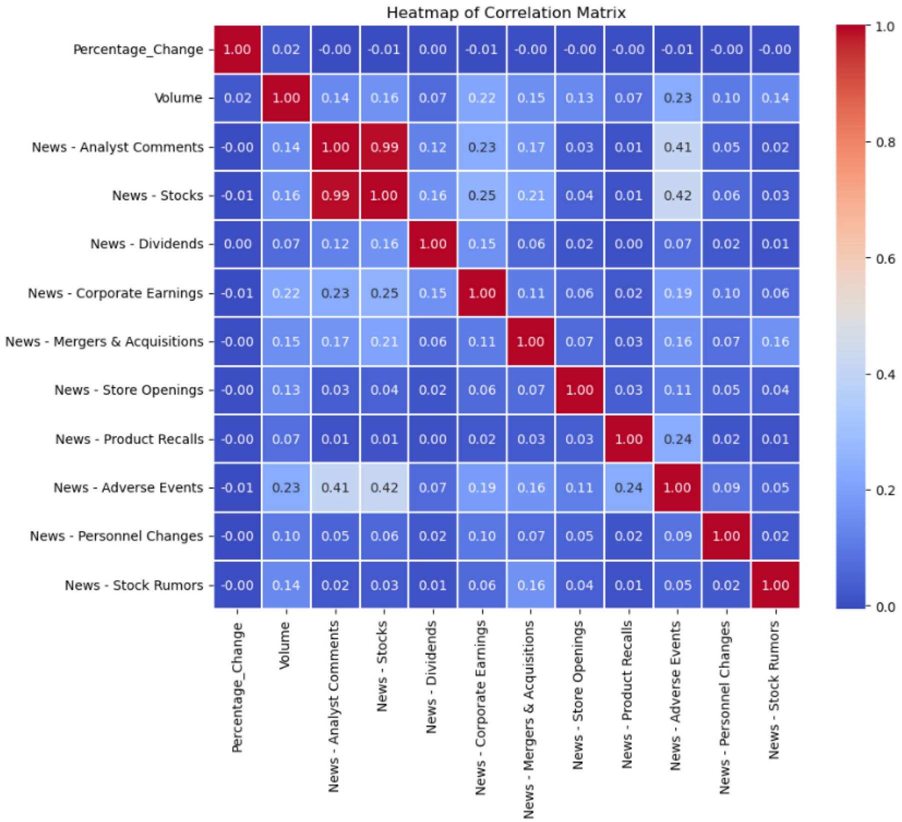


2.375x Stock Split in 2015

- Old dataset had splits
- After checking new one, we concluded there were no stock splits between Oct 2020 - July 2022
- Our online research reaffirmed our findings of no stock splits



Note the lack of correlation between day-to-day Percentage_change and other features



Even when staggering the news with stock changes accumulated a week later

Data Preprocessing:

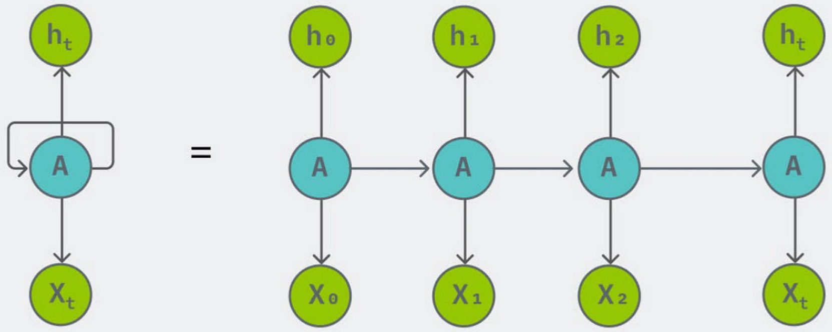
- Feature Normalization using from sklearn MinMaxScaler
- Switching dates to Pandas DateTime type

```
 #   Column  Non-Null Count   Dtype
---  ------  --------------   -----
 0   date    619040 non-null  datetime64[ns]
 1   open    619029 non-null  float64
 2   high    619032 non-null  float64
 3   low     619032 non-null  float64
 4   close   619040 non-null  float64
 5   volume  619040 non-null  int64
 6   Name    619040 non-null  object
```

DateTime Functionality:

- Parsing dates
- Date arithmetic
- Resampling
- Shifting and lagging
- Rolling windows
- Time Zones

Recurrent Neural Network



- Best-suited for sequential data such as time series data, voice recognition, natural language processing, and more.
- Remember past inputs due to internal memory, allowing them to capture short-term dependencies.
- RNNs can suffer from vanishing gradients leading to slow learning or convergence issues.
- Takes in previous trends of gradients in order to predict what future ones will be

Long Short -Term Memory (LSTM) networks

- Type of RNN even more adept at processing time series data
- LSTMs incorporate memory cells and gating mechanisms to better capture long-term dependencies
- Gating Mechanisms
  - Forget Gate: Determines what information to discard from the memory cell.
  - Input Gate: Controls the flow of new information into the memory cell.
  - Output Gate: Regulates the output from the memory cell.
- LSTMs use these gates to selectively update the memory cell, preventing gradients from vanishing.

Models

Single Input

p

Layer (type)	Output Shape	Param #
lstm_16 (LSTM)	(None, 35, 64)	16,896
lstm_17 (LSTM)	(None, 64)	33,024
dense_16 (Dense)	(None, 32)	2,080
dropout_8 (Dropout)	(None, 32)	0
dense_17 (Dense)	(None, 1)	33

Total params: 52,033 (203.25 KB)
Trainable params: 52,033 (203.25 KB)
Non-trainable params: 0 (0.00 B)

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 1259, 50)	10,400
lstm_1 (LSTM)	(None, 50)	20,200
dense (Dense)	(None, 25)	1,275
dropout (Dropout)	(None, 25)	0
dense_1 (Dense)	(None, 1)	26

Total params: 31,901 (124.61 KB)

Trainable params: 31,901 (124.61 KB)

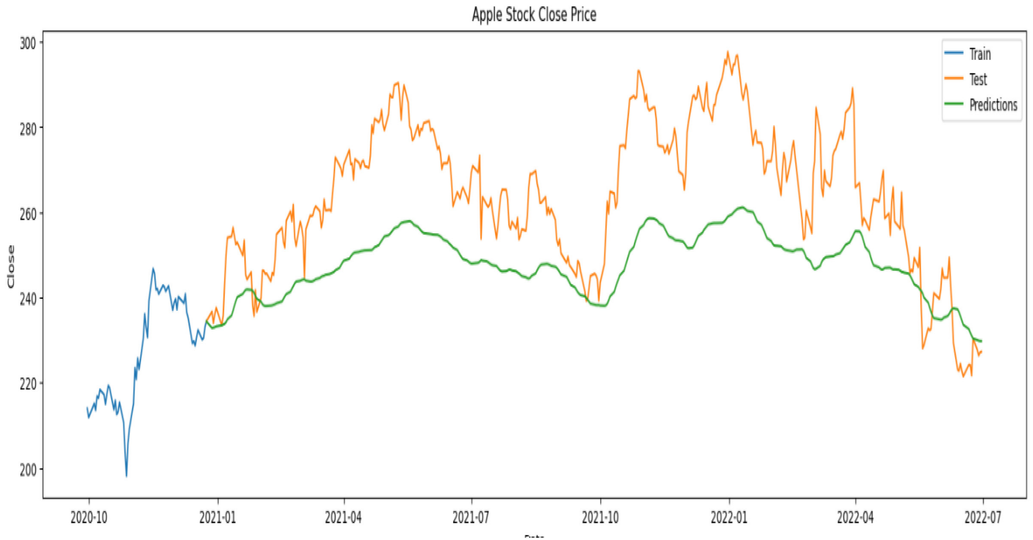
Non-trainable params: 0 (0.00 B)

Initial Results - Predicting Prices of AAPL

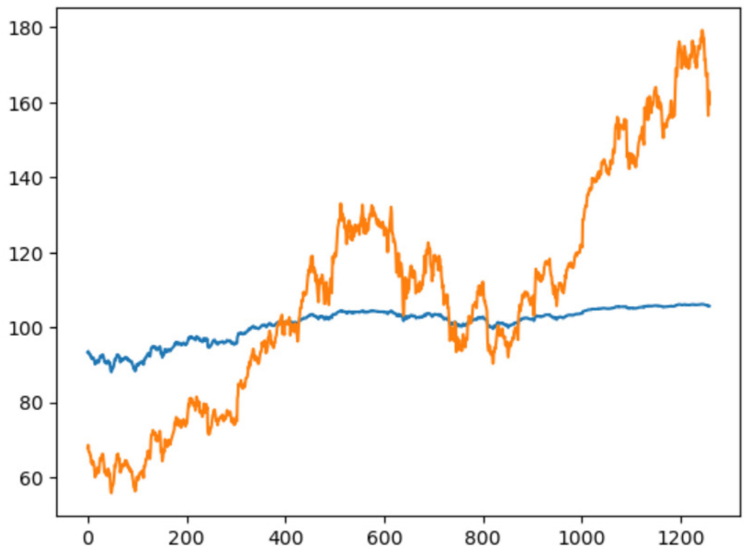
Single Input

All Stock Input

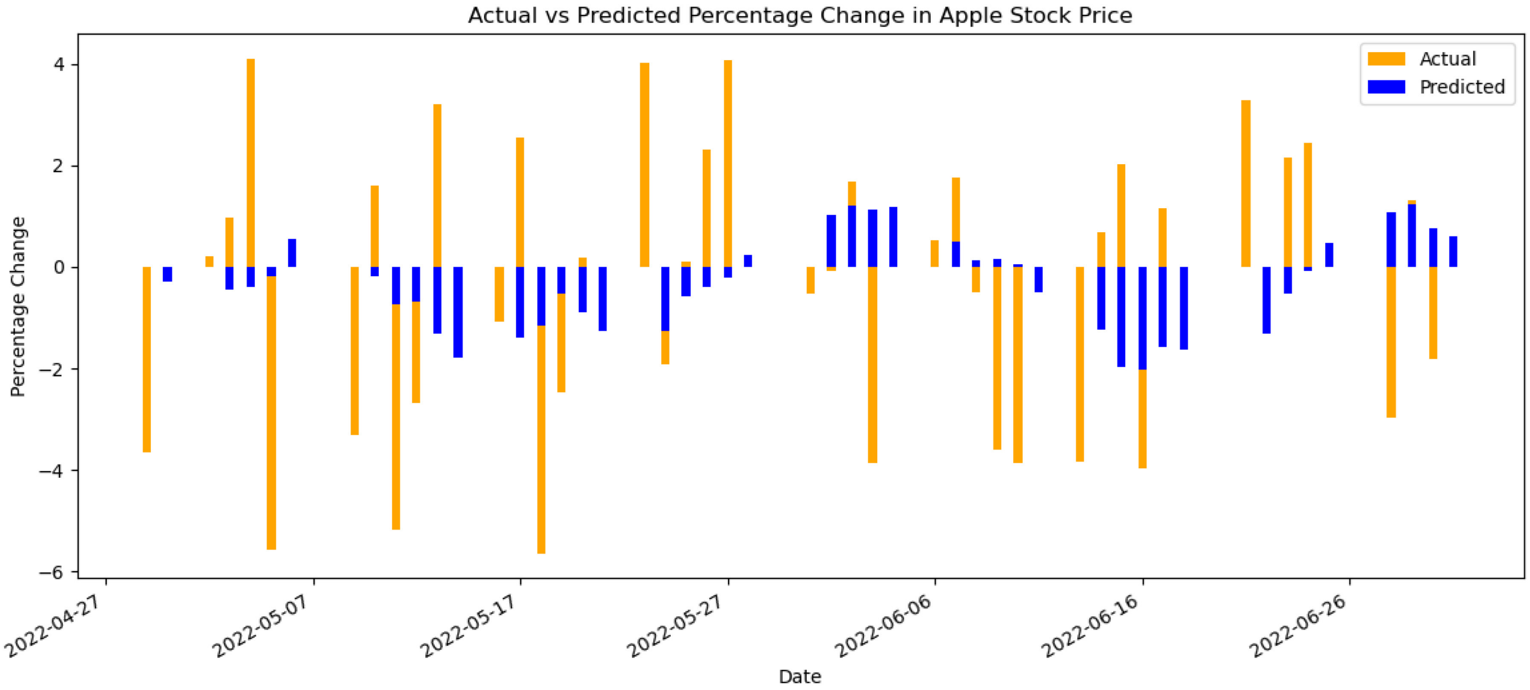
MSE = 389
RMSE = 19.7



MSE = 1782
RMSE = 42



Evaluating Success



Evaluating Success

```
Number of correct predictions:  195
Number of incorrect predictions:  184
----------------------------------------------------------------
Average error:  0.5092199813145148 %
----------------------------------------------------------------
Initial balance: $100
Final balance: $ 97.09
Percent change from initial to final balance:  -2.91 %
Annualized percent change:  -2.81 %
```

Is the stock market actually predictable?

The subtle and amusing reasons that was proven to alter stock market prices.

Are you a physically attractive CEO? Go on TV!

When attractive CEOs of companies appear on television, the stock price of their companies rise but being quoted in a newspaper, without a photo, has no effect (Halford & Hsu, 2014).

Don’t trade without your morning coffee

A company stock price may rise immediately due to investors mistaken it with another company name. Mistaken identities costs million dollars per year (https://www.nysscpa.org/news/publications/the-trusted-professional/article/study-mistaken-identity-in-stock-names-costs-1-million-annually-081319)

If you are a CEO, better say nothing.

Irrelevant comments by a CEO significantly affected stock prices of their companies both on the short and long run. Elon Musk comment on how competitive are Chinese EV companies floored Tesla stock price on that day.

Want to predict better than AI/ML? Follow this scheme:

1. Send an email to 1024 people informing half of them that company X stock is increasing tomorrow and the other half that is decreasing.
2. Notice how the price actually changed and exclude the group you gave wrong prediction.
3. For the new group, repeat steps 1 and 2 until you end up with a single person on the mailing list.
4. This 1 person witnessed you predicting a stock price correctly 10 times in a row!

Conclusion

- Our model learns well how to predict stock prices BASED ON PAST DATA.
- With different models, datasets, and architectures, we conclude that stock market is unpredictable given past information only.
- Real-time events affects the market unpredictability.
- If the market was predictable, it wouldn’t have been an open market!
- A model like ours can serve as a guide about the stock past performance, not a method to guarantee profit.