Profit-generating juggernaut: Predicting stock prices using machine learning

CHRIS ANDREW MENDOZA

APPLIED PHYSICS 157 TX-1
MAY 21, 2024

Background

Typical methods utilized to profit from the stock Market:

Fundamental Analysis

- Financial Statements
- Financial Ratios
- Economic Performance

Technical Analysis

- Candlestick Charts
- Trends/Patterns
- Price Movement

Both methods are widely used resulting in less α (less competitive edge)

Background

The stock market is first and foremost a competitive environment.

- Institutions, especially **hedge funds**, will spend a lot of money on anything just to get a competitive advantage over others
  - News (via Bloomberg Terminal)
  - Satellite Imaging
  - Advanced scientific approaches in analyzing market movement
  - Lobbying politicians

- There is a need to implement every method at your disposal to generete α

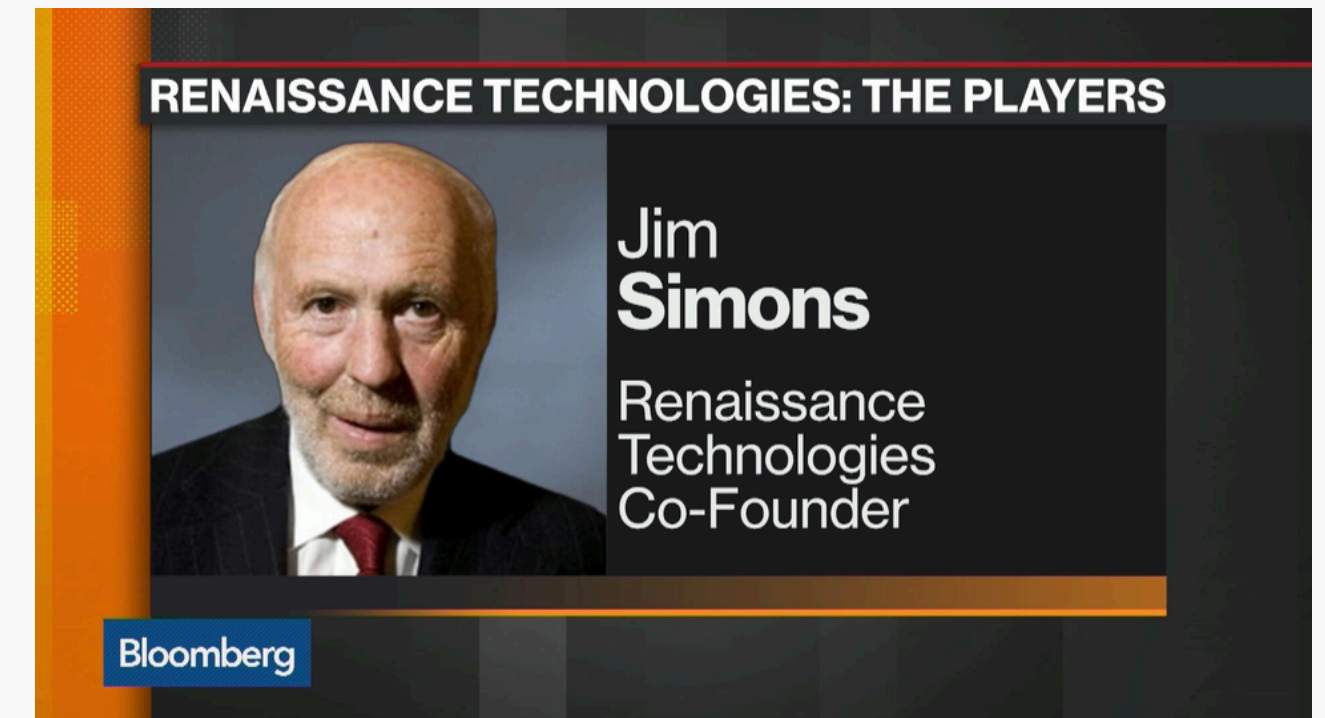- **Machine learning** is capable of detecting patterns a human brain cannot comprehend

Figure 1. Jim SImons of RenTech. Source: Bloomberg News.

Methodology:

- Model used: Long Short-Term Memory (**LSTM**), a type of RNNs Recurrent Neural Network that can detain long-term dependencies in sequential data.

- Utilize the scikit-learn and keras libraries

- Make use of a single stock data set containing all its historical price data (obtained through **Nasdaq**)

- Setting the Target Variable and Selecting the Features
  - **Target Variable**: Close/Last Price
  - **Features**:
    - Open
    - High
    - Low
    - Volume

- Creating a Training Set and a Test Set: **90% / 10 % split**

Methodology:

- Data Processing For LSTM
  - LSTM needs that the data to be provided in the 3D form
  - Transform the training and test data to NumPy arrays
  - Restructure them to match the format (Number of Samples, 1, Number of Features)

- Building the LSTM Model
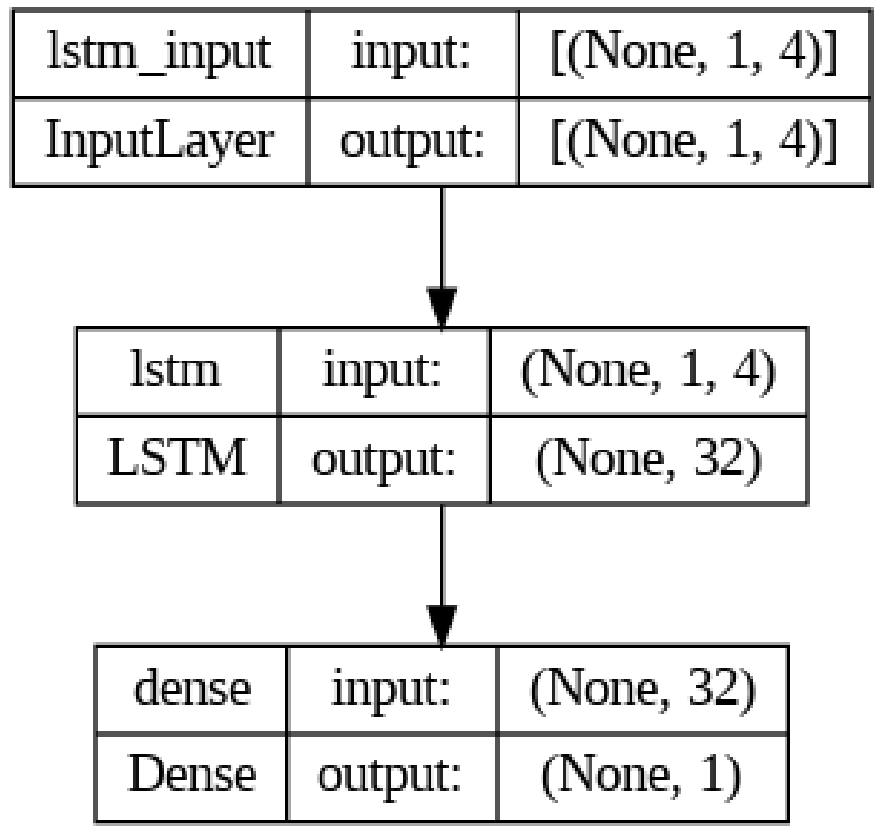 - Sequential Keras model with one LSTM layer

| lstm_input | input: | [(None, 1, 4)] |
| InputLayer | output: | [(None, 1, 4)] |

| lstm | input: | (None, 1, 4) |
| LSTM | output: | (None, 32) |

| dense | input: | (None, 32) |
| Dense | output: | (None, 1) |

Figure 2. LSTM model used for stock price prediction.

Methodology:

- Training the model
    - Use fit function to train the LSTM model for 100 epochs

- Comparing Predicted vs True Close Value
    - Plot the predicted and true values side by side

Results
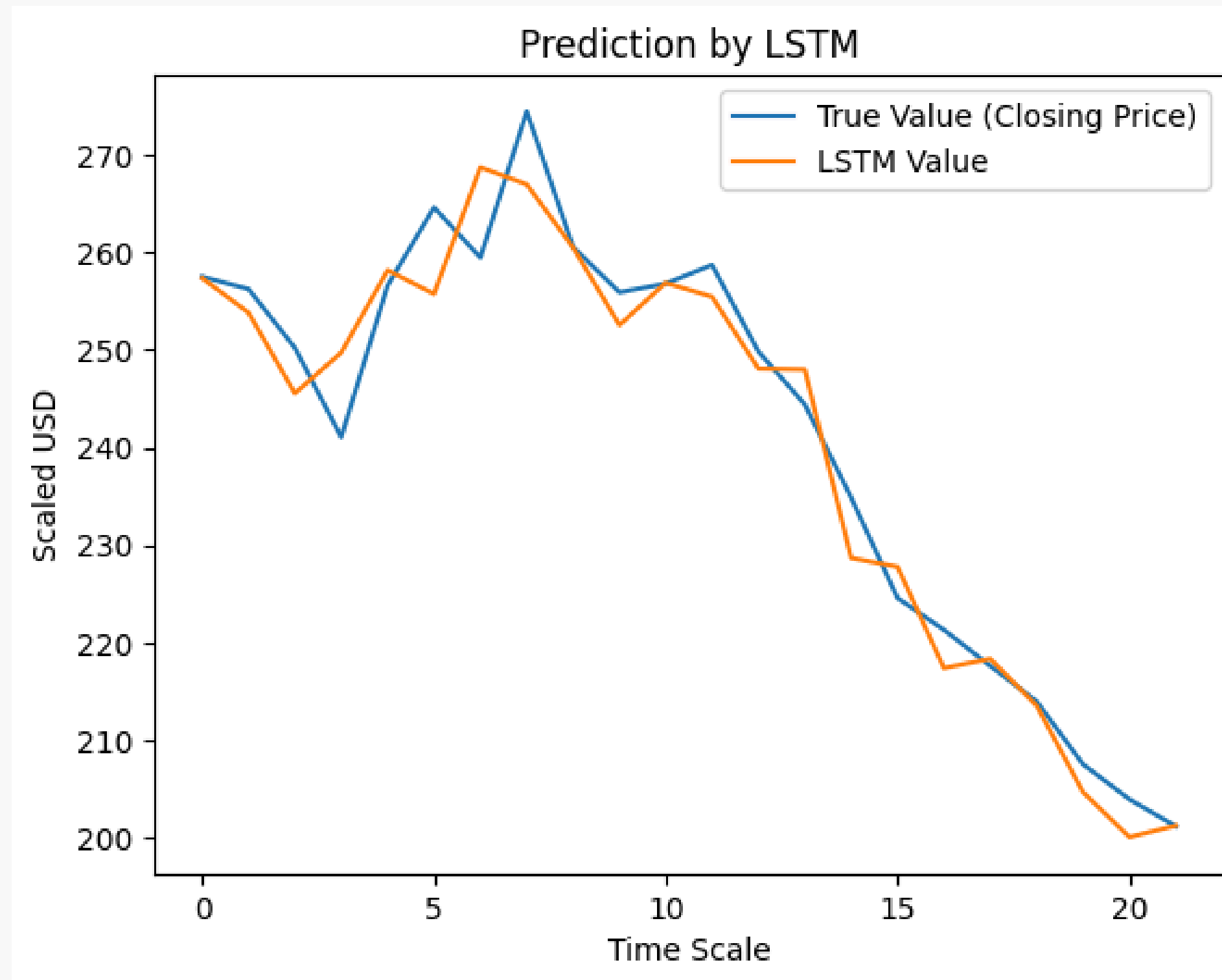
RMSE (after 100 epochs): $10.2159

Insights and limitations

- The model underfits the test set

- Need to increase dataset as the one utilized is only for 1 fiscal year

- Additonal LSTM layers will improve the performance of the model (only 1 layer was used in this project)

Conclusion

- The project was able to utilize the historical Open, High, Low, and Volume data for a specific stock to predict its Close/Last Price.

- Performance of the model is still suboptimal but may be improved with the further addition of multiple LSTM layers and a larger dataset.

- As Financial models rely on base case, average case, and best case scenarios, it is will be beneficial to also utilize projected data values in training the model in order to mitigate risk.

Thank you for listening!