Stockalyzer

Yi Man | Hannie Ngo

Project Goals

- To apply deep learning techniques (Recurrent Neural Network algorithm) to predict the future stock closing value of a given stock across a given period of time.
    - Using **Long Short Term Memory network** improved training accuracy and avoided overfitting issue.
    - Achieved Mean Squared Error rating of just **0.0318 RMSE**

- To build financial analysis for a user-specified group of stocks based on daily/monthly/quarterly returns by sampling and calculating volatility of a given stock over a specific time frame.

Tools

- Python/Jupyter
- Quandl
- Tensorflow/Keras
- MatplotLib
- Numpy
- Pandas
- Sklearn
- Github

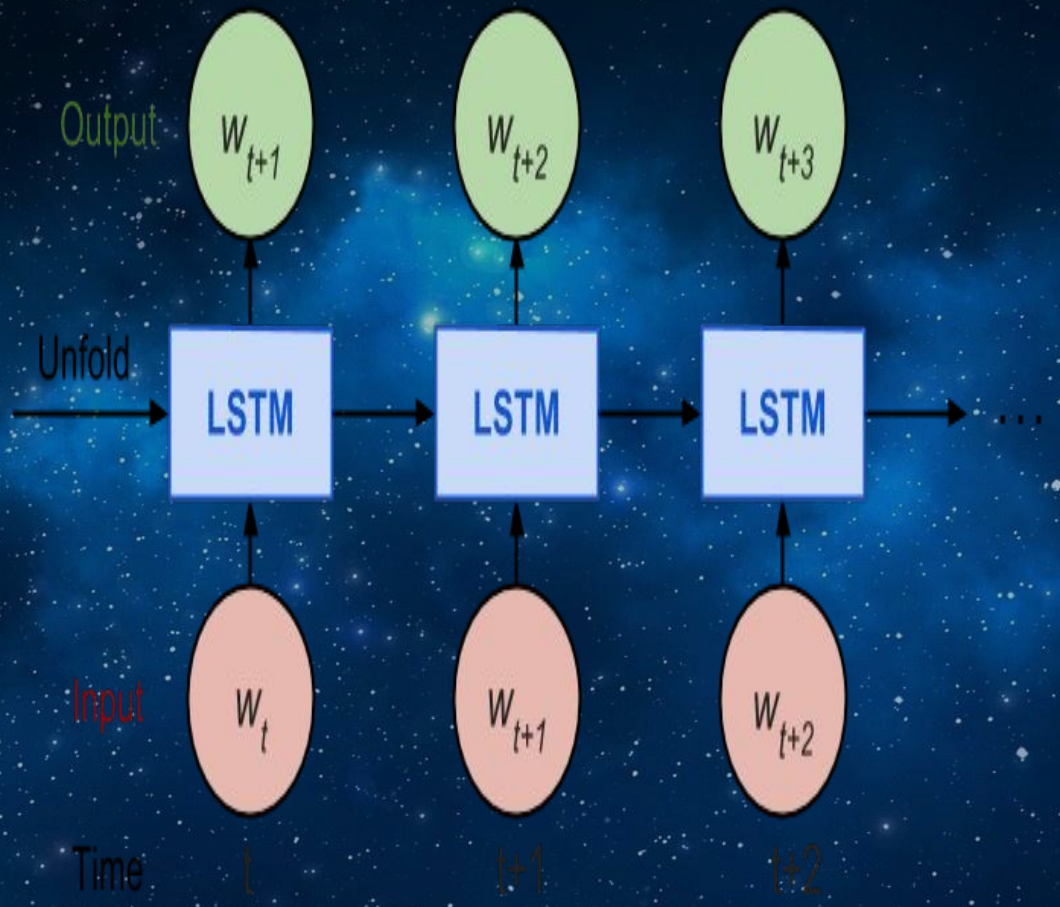
Preparation Work

- **Data Collection**
  Fetch individual stocks in S&P 500 for maximum of 4000 days and save to a folder
- **Data Cleaning**
  Remove non-applicable columns: Volume, Ex-Dividend, Split Ratio, Adj. OHLC
  Split and reshape original data into training/testing batches (75/25)
- **Data Normalization**
  Use standard normalization from sklearn with mean 0 and standard deviation 1
- **Data Transformation**
  Construct samples to calculate monthly/quarterly percentage change in daily returns
- **Data Integration**
  Concatenate data from different symbols of stocks in a specified time frame for comparison
- **Model Construction**
  Construct LSTM model with no hyperparameters being initiated

Recurrent Neural Network (RNN)

RNN is a type of neural network with self-loop in its layers, which enables it to use the previous state of neurons to learn the current state given the new input → good at handling sequential data.

Long Short Term Memory cell (LSTM) is specifically designed to help RNN better memorize the long-term context.



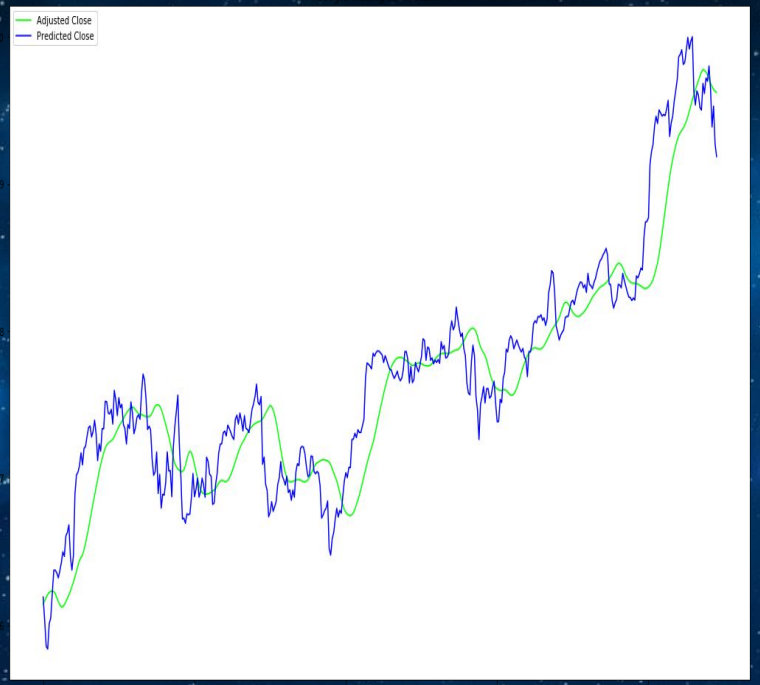
Recurrent Neural Network

Basic LSTM model (3 layers)

Number of epochs: 10

0.2607 RMSE

Improved LSTM model with additional dropout layers



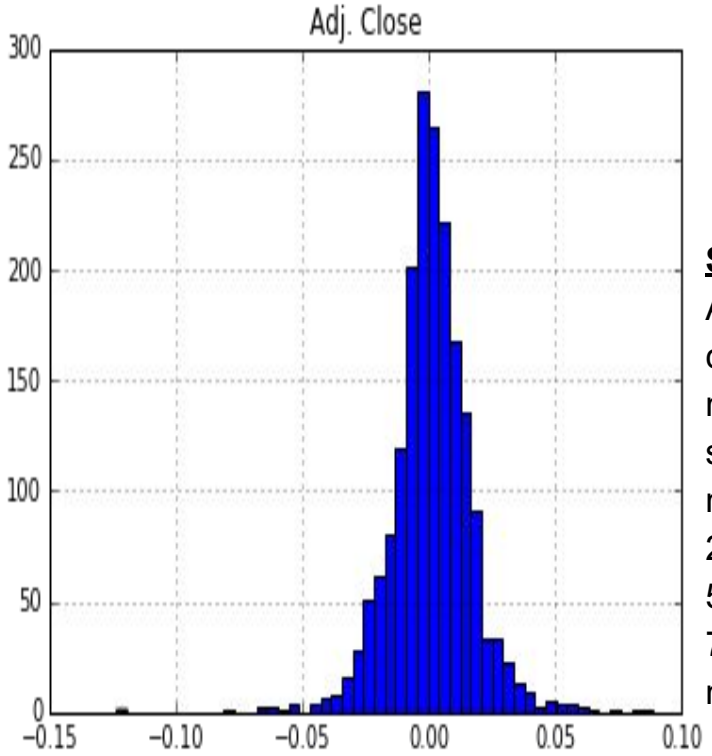
Number of epochs: 10

0.0318 RMSE

Predicted stock value

```
Last Day Value: 0.5643
Next Day Value: 0.5612
Value Difference: -0.5501%
```

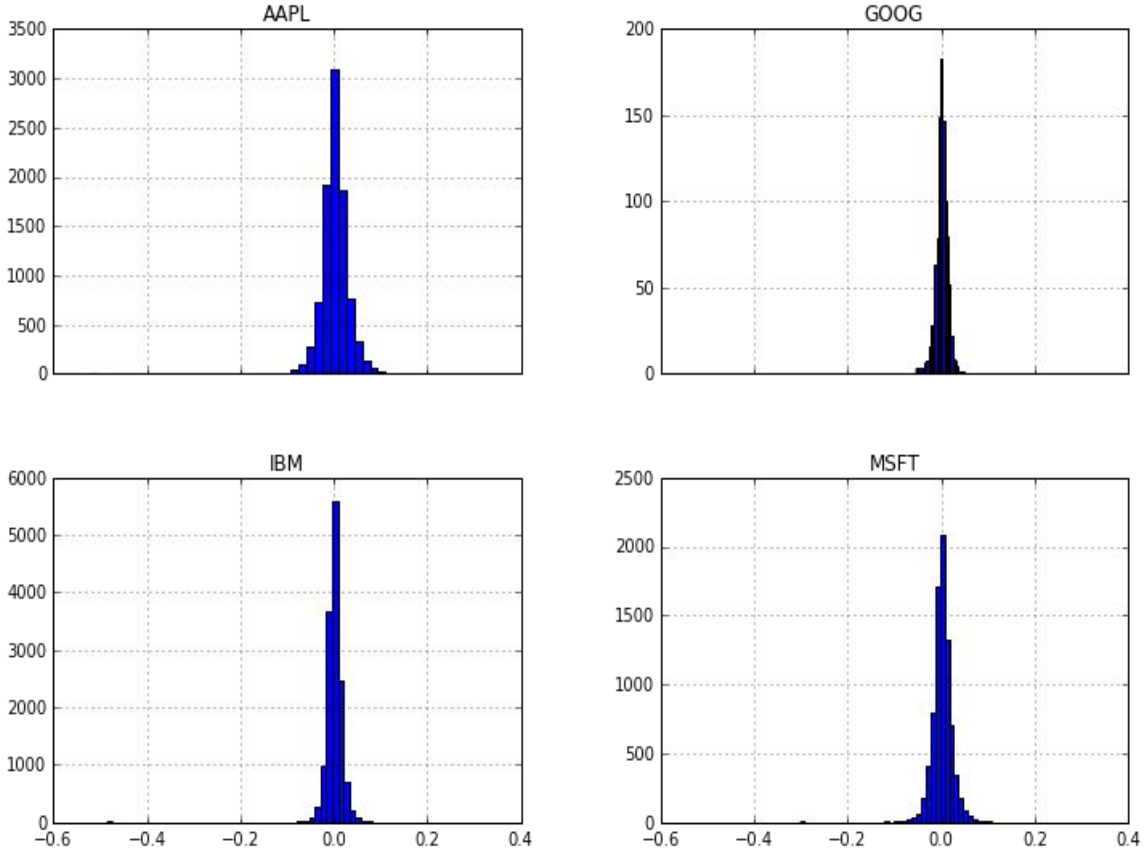
Stock Volatility Calculation & Comparison



Summary

Adj. Close	
count	1881.000000
mean	0.000940
std	0.015737
min	-0.123549
25%	-0.006818
50%	0.000547
75%	0.009411
max	0.088741

The distribution looks very symmetrical and normally distributed: the daily changes center around the bin 0.00



Change in cumulative daily rate of return among four different stocks

Application of Knowledge

Investment firms, hedge funds, and daily stock traders can:

- Predict stock prices with machine learning model in a specific time frame
- Project financial analysis and comparison within a specified set of stocks based on historical datasets
- Better decide which stock may have less risk when examining the stock's volatility in the overall market

Q&A