

Stock market modeling and forecasting

Team: *Suman Aich, Nafis Fuad, Siu Cheung Lam, Xiaoyu Wang*

Objective: We perform stock market analysis using data from multiple stocks (including index funds and companies). Our approach is based on both statistical modeling and LSTM neural networks. For statistical modeling we use autocorrelation plots to examine trends in data and root mean squared error (RMSE) as our key performance indicator. Using the LSTM neural network we design a regression model for forecasting and a classifier to predict whether to buy, hold or sell stocks at any given day. Finally, we explore the LSTM regression model's ability to generalize to multiple stocks, as well as its usage for multi-day forecasting.

Stakeholders: Short-term investment companies such as Fidelity Investments, Interactive Brokers, Charles Schwab, etc. Long-term investors.

Statistical modeling: We perform statistical modeling based on Gaussian rolling averages, naive forecasts, simple exponential smoothing and ARIMA. We evaluate their performance using RMSE. We use historical stock data and split it into training and test datasets with the test data being the final day. We also preprocess the data using min-max scaler because data from the 1980s and 2020s have drastic differences in mean and variance. We find that naive forecasts, simple exponential smoothing and ARIMA perform equally good and far better than Gaussian rolling average based on RMSE values and we also confirm that by plotting the training, test and predicted data from all the models for the last month.

LSTM modeling: We utilize a recurrent neural network with LSTM cell for predictive modeling of the financial market data. The historical stock data is segmented into rolling time windows of 41 days, with the first 40 days

used for training, and the last day as the training target. We perform the same data preprocessing as done for statistical modeling. MSE is used as the metric to evaluate the model effectiveness. Our results demonstrate reasonable effectiveness of the LSTM regressor in modeling the market data. However, the regressor shows similar performance to that of a 1 day lag predictor which predicts the next day's price with the current day's price. To further explore the power of a LSTM neural network we conduct four subsequent experiments:

- (1) A LSTM regressor trained with quantile loss function. This allows us to not only forecast the stock price but also the confidence intervals.
- (2) A LSTM binary classifier which predicts the movement of the stock price between two consecutive days. We used a train data set where the movement of the stock price is classified as up or down, with ~51% of the dataset labeled by "up". The classifier is able to achieve ~75% accuracy in predicting the market movement, much higher than a random guess.
- (3) A LSTM regressor using multiple stocks as training features. This model achieved a slightly better MSE metric than that trained with a single stock, and performs better when generalizing to other stocks not included in the training set.
- (4) A LSTM regressor for multi day forecasting. We demonstrate that although LSTM is good for one day forecasts, it severely underestimates the market volatility associated with multi day stock movements.

Outlook: Given the effectiveness of the LSTM binary classifier in predicting stock price movements, we plan to evaluate a multiclass classifier which forecasts the percentage change of the stock price between two consecutive days. We shall also test classification accuracy by adding market indicators such as sentiments, macroeconomic data, Relative Strength Index, etc.