Stock Selection and Comparative Analysis of Predictive Models for S&P 500 Future Price Forecasting

University of Hertfordshire UH

By - Tom Thomas (22008590)
Supervised by – Dr. Stephen Kane

AIM : To analyse historical S&P 500 stock data and evaluate predictive models for accurate stock price forecasting. Through dataset analysis, literature review, and model evaluation, the project seeks to identify effective approaches for informing investment decisions in financial markets.

INTRODUCTION

In the dynamic landscape of financial markets, effective stock selection and accurate price forecasting are essential for investors seeking to optimize their portfolios and maximize returns. This poster focuses on the utilization of a single predictive model, LSTM (Long Short-Term Memory), to forecast future prices within the renowned S&P 500 index. A comparative analysis of LSTM's performance in capturing the intricacies of S&P 500 stock movements are done here. By evaluating LSTM's predictive capabilities, the aim is to provide insights into its effectiveness compared to other methodologies.

OBJECTIVE

- Examine a comprehensive review of the literature to determine the most suitable model for stock price prediction.
- Utilize techniques such as hiding some of the actual prices and predicting the hidden values, including those already known, to evaluate the model's predictive capabilities.
- Conducting additional review and analysis to select two more additional models for comparison against the LSTM model.

DATASET

- The dataset is obtained from Yahoo Finance, provides historical stock data for various securities, including those listed in the S&P 500 index.
- The sector information is extracted from the S&P 500 companies data available on Wikipedia, specifically from the page listing the S&P 500 companies.
- The historical stock data covers a period of 5 years until February 27, 2024. Specifically, 5 stocks are selected.
- The data used in this analysis is publicly available and used solely for research and analytical purposes, adhering to ethical standards.
- Figure 1, depicts a line graph illustrating the price variations of the selected top stocks over the period of time.

METHODOLOGY

- Scraping S&P 500 companies' data from Wikipedia and formatting it.
- Utilizing Yahoo Finance API to fetch historical stock data for the S&P 500.
- Computing dollar volume and daily returns for individual stocks for organizing sector information alongside data processing.
- Determining monthly average dollar volume and ranking stocks accordingly to select top-ranked stocks within each sector.
- Analysing the correlation matrix among the selected top stocks by choosing 5 least correlated stocks for final selection.
- Aligning the index to business days, filling missing values with forward filling.
- Splits dataset into features and target, then divides into training (80%) and test sets (20%).
- Scales the input data using MinMaxScaler.
- Trains an LSTM model with epochs = 100.

LITERATURE REVIEW

AUTHORS (YEAR)	DATA	ML TECHNIQUES USED	FINDINGS
Md. Arif et al., (2020)	yahoo finance	LSTM, and BI- LSTM	BI- LSTM model generates lower RMSE compared to LSTM model.
Gao, Y et al., (2021)	NetEase Finance and Economics website	LSTM and GRU Models	LSTM and GRU models effectively predict stock prices .
Ji, H et al., (2022).	yahoo finance	SVM and LSTM	SVM performs well in short-term predictions, LSTM excels in long-term forecasts.
Chaajer, P et al.,(2021)	Eleven stocks, from the Brazilian stock series of 2016	LSTM, MLP, SVM.	LSTM stood out of all the models.
Venikar et al., (2022)	Tiingo	LSTM	LSTM is suitable for forecasting the stocks.

Table 1

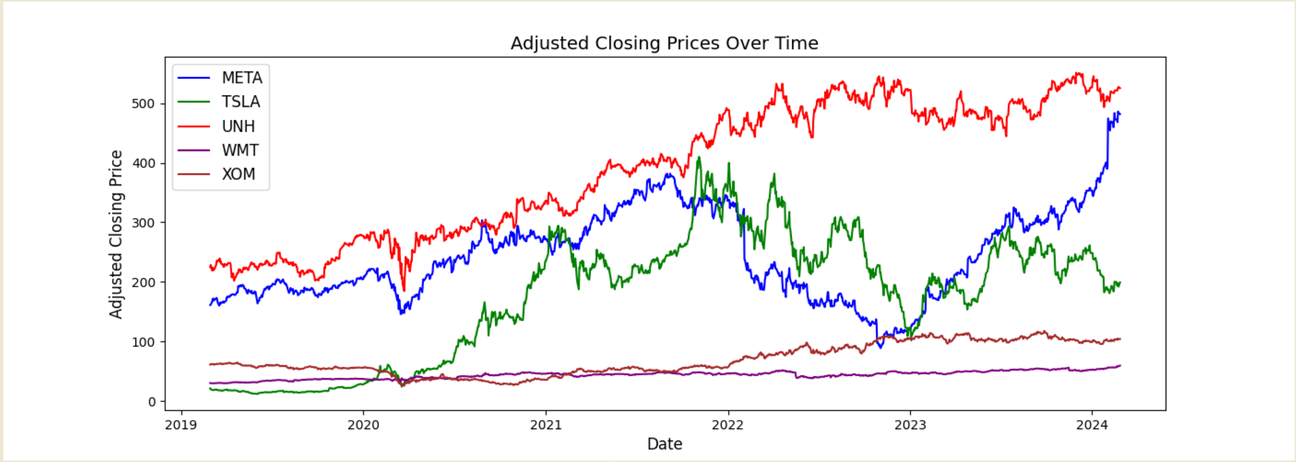

Figure 1

WHAT'S NEXT..?

- Implementing two additional models based on the literature reviews.
- Train models with different epochs and tuning.
- Get conclusions on the most suitable model for stock prediction.

RESULTS SO FAR…

The Long Short-Term Memory (LSTM) model, employed for stock price forecasting, achieved an RMSE value of 9.56. The dataset was divided into an 80% training set, comprising 257 data points, and a 20% testing set, consisting of 48 data points. This division facilitated comprehensive model training on historical data while enabling rigorous evaluation of unseen test data.

Figure 2, displays a line graph presenting predicted values for the next 60 days. With 30 days of data hidden, the model forecasted the subsequent 30 days. Impressively, the model accurately predicted the hidden 30-day values, underscoring its potential to reliably forecast future stock prices.

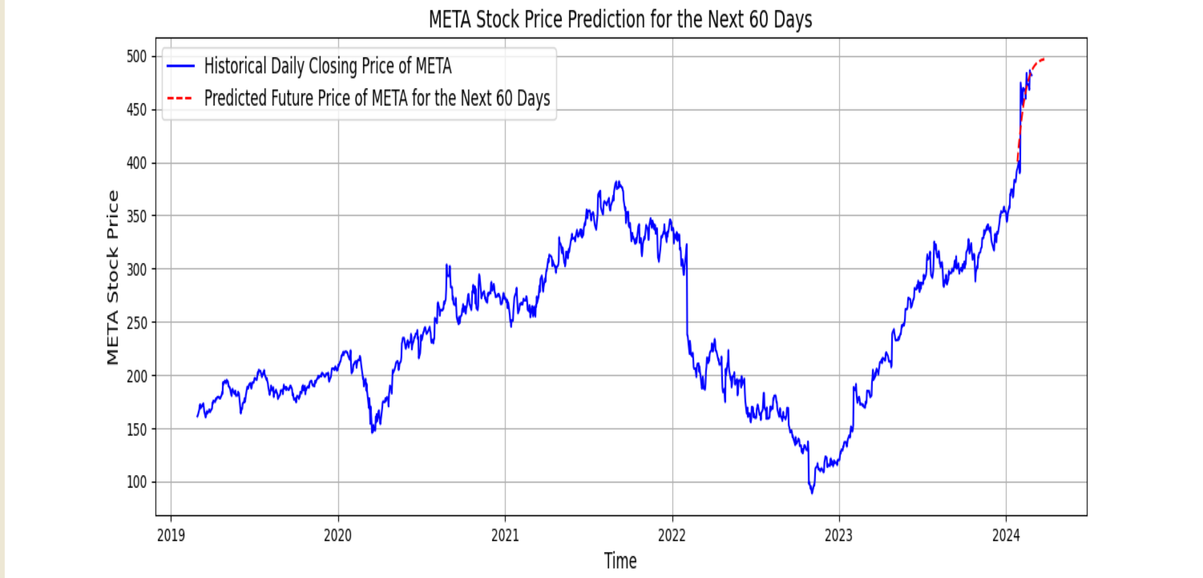

Figure 2

REFERENCES:

I. Chaajer, P., Shah, M. and Kshirsagar, A. (2021). The applications of artificial neural networks, support vector machines, and long-short term memory for stock market prediction. Decision Analytics Journal, 2, p.100015. doi:https://doi.org/10.1016/j.dajour.2021.100015.
II. Gao, Y., Wang, R. and Zhou, E. (2021). Stock Prediction Based on Optimized LSTM and GRU Models. Scientific Programming, 2021, pp.1–8. doi:https://doi.org/10.1155/2021/4055281.
III. Istiake Sunny, Md.A., Shahriar Maswood, M.M. and G. Alharbi, A. (2020). *Deep Learning-Based Stock Price* Prediction Using LSTM and Bi-Directional LSTM Model | IEEE Conference Publication | IEEE Xplore. [online] ieeexplore.ieee.org. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9257950.
IV. Ji, H. (2022). Stock price prediction based on SVM, LSTM, ARIMA. BCP Business & Management, 35, pp.267–272. doi:https://doi.org/10.54691/bcpbm.v35i.3302.
V. Venikar, I., Joshi, J., Jalnekar, H. and Raut, S. (2022). Stock Market Prediction Using LSTM. International Journal for Research in Applied Science and Engineering Technology, 10(12), pp.920–924. doi:https://doi.org/10.22214/ijraset.2022.47967.
VI. Wikipedia Contributors (2019). List of S&P 500 companies. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/List_of_S%26P_500_companies.
VII. Yahoo Finance (2023). Yahoo Finance – stock market live, quotes, business & finance news. [online] uk.finance.yahoo.com. Available at: https://uk.finance.yahoo.com.