ASSIGNMENT 3: DEEP LEARNING – a1894603

RNN’s FOR STOCK PRICE PREDICTION

**Abstract**

*The integration of artificial intelligence algorithms in the financial domain has emerged as a captivating and extensively researched topic for stock price prediction. This study employs LSTM and ARIMA models to delve into the intriguing realm of forecasting stock prices. A comprehensive comparison of forecasting accuracy is conducted using key statistical indicators, namely MSE, MAE and RMSE. The models are constructed utilizing data obtained from google finance. The predictive outcomes reveal that the LSTM algorithm outperforms the alternative ARIMA, exhibiting smaller values for MSE, MAE, and RMSE. The findings presented in this paper offer valuable insights for investors in the capital market seeking to enhance their ability to forecast future stock prices of Google.*

***Keywords:*** *stock prices, prediction, LSTM, ARIMA.*

**1.INTRODUCTION**

Presently, big data has become a focal point of interest, with global investors exploring diverse big data-based approaches to forecast stock price fluctuations. Traditional models are challenged by the expanding stock market, prompting the adoption of neural networks, whose formidable information processing capabilities find increasing applications in real stock price prediction.

Methods for predicting stock prices primarily encompass time series prediction, technical index analysis, and artificial intelligence. Time series information, encompassing techniques like moving average (MA), auto regressive moving average (ARMA), and auto regressive integrated moving average (ARIMA). The MA model, a straightforward prediction model, calculates the average value of past variables as the prediction for the next cycle. This helps smooth out noise by averaging, though the subjectivity in setting model parameters makes it challenging to analyze and accurately predict real-world scenarios. The stock price prediction primarily focuses on qualitative rather than quantitative aspects, emphasizing the forecast of the future directional changes in stock prices. Deep learning algorithms, such as recurrent neural network (RNN), long-term and short-term memory network (LSTM), constitute the main modelling methods. Different neural network types are suited for analyzing distinct data types, with RNN being particularly applicable for time series data like stock prices. While RNN can learn sequence relationships, prompting scholars to enhance it using LSTM. The gating mechanism in each LSTM unit enables the adjustment of long-term memory based on current input and historical information.

This study introduces two predictive models, namely LSTM and ARIMA, and applies them to predict Google stock close prices. Recognizing the numerous factors influencing prediction results, the paper emphasizes optimizing data selection and processing, parameter adjustment, and overall framework design. Additionally, the study presents future predictions, conducting in-depth investigations by calculating three widely used indicators – MSE, MAE, and RMSE – to assess the forecasting performance of the two selected models.

**2.LITERATURE REVIEW**

Time series analysis and dynamic modelling represent a compelling research domain with extensive applications across business, economics, finance, and computer science. The primary objective of time series analysis is to investigate the sequential observations of time series data, construct a model that elucidates the data structure, and subsequently make predictions about future values. Given the pivotal role of time series forecasting in various applied sciences, the development of effective model is crucial for enhancing forecasting accuracy. Numerous time series forecasting models have emerged in the literature to address this need.

Traditionally, econometric approaches employ ARIMA models for time series forecasting. ARIMA has long been a standard method for time series forecasting. Despite its widespread use in modelling economic and financial time series, ARIMA models do have notable limitations. For example, simple ARIMA models struggle to capture nonlinear relationships between variables. Additionally, these models assume a constant standard deviation in errors, which may not hold true in practice. Integrating an ARIMA model with a Generalized Auto-regressive Conditional Heteroskedasticity (GARCH) model allows for the relaxation of this assumption. However, optimizing a GARCH model and its parameters presents challenges and complexities. While ARIMA finds applications in modelling short and long-run effects of economic parameters, it is imperative to acknowledge the need for more flexible and robust modelling techniques to overcome the limitations associated with traditional approaches.

In recent times, advancements in deep learning have emerged as effective solutions to tackle challenges associated with forecasting models. One notable technique is Long Short-Term Memory (LSTM), a specialized form of Recurrent Neural Network (RNN) initially introduced by Hochreiter and Schmidhuber. While relatively new, deep learning-based approaches have garnered significant attention from researchers.

Researchers such as Krauss et al. have explored diverse forecasting models, including deep learning, gradient-boosted trees, and random forests, to model constituents of the S&P 500. Interestingly, their findings indicated that, contrary to expectations, deep learning-based models performed less effectively than gradient-boosted trees and random forests. Krauss et al. also emphasized the inherent difficulty in training neural networks, a key component of deep learning algorithms.

Lee and Yoo proposed an RNN-based method for predicting stock returns. Their approach involved constructing portfolios by adjusting threshold levels of return within the internal layers of RNN. Similarly, Fischera et al. conducted work on financial market prediction using an RNN-based approach.

**3.ALGORITHM**

**3.1 ARIMA**

ARIMA is a statistical method used for time series forecasting. It combines three components: Auto Regressive (AR), Integrated and Moving Average (MA). AR component captures the relationship between an observation and its lagged values. Integrated component involves differencing the time series to make it stationary.

MA component represents the correlation between a current observation and residual error from a moving average model applied to lagged observations.

**A diagram of mathematical equations

Description automatically generated with medium confidence**

**3.2 MLP**

MLP is a type of artificial neural network with multiple layers. It consists of an input layer, one or more hidden layers, and an output layer. Neurons in each layer are connected to neurons in the next layer by weights. Activation functions introduce non-linearity to the model.

**3.3 RNN**

RNN is a type of neural network designed for sequential data. It has connections that form a directed cycle, allowing information persistence over time. Each neuron in an RNN layer receives input not just from the previous layer but also from its own output at the previous time step.

**3.4 LSTM**

LSTM is a type of RNN designed to overcome the vanishing gradient problem. It introduces memory cells and gates (input, forget, output) to selectively remember or forget information. It can capture long-term dependencies in sequential data.

**Formulas:**

* Forward pass:

*A math equations and symbols

Description automatically generated with medium confidence*

Where:

* ft, it, ot are the forget, input, and output gates,
* **W** matrices are weight matrices
* b terms are biases,

These formulas represent the forward pass through the networks. During training, the backward pass (backpropagation) is used to adjust the weights and biases to minimize the prediction error. The specific architectures and hyperparameters are determined through experimentation and tuning.

**4. EXPERIMENT ANALYSIS**

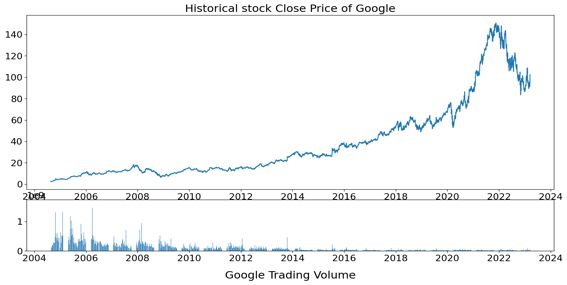
**4.1 Data Loading and Explorations**

The Google Stock Price data in this article loaded from ‘Google.csv’ which contains data from 2004- 2023. The data contains High, Low, Open, Close, Volume etc as shown in Table 1.

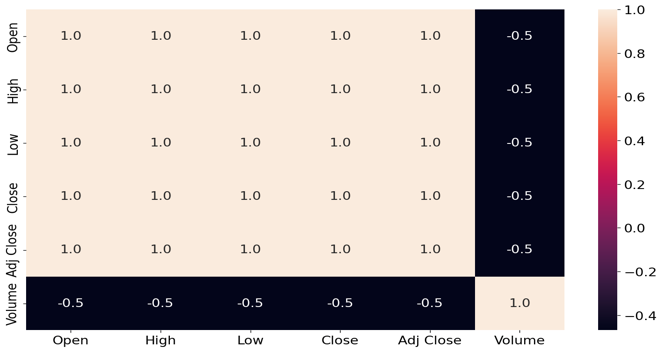
**Table 1:** describing the dataset.



Plotted a few data points to visualize the historical stock close prices and trading volumes.



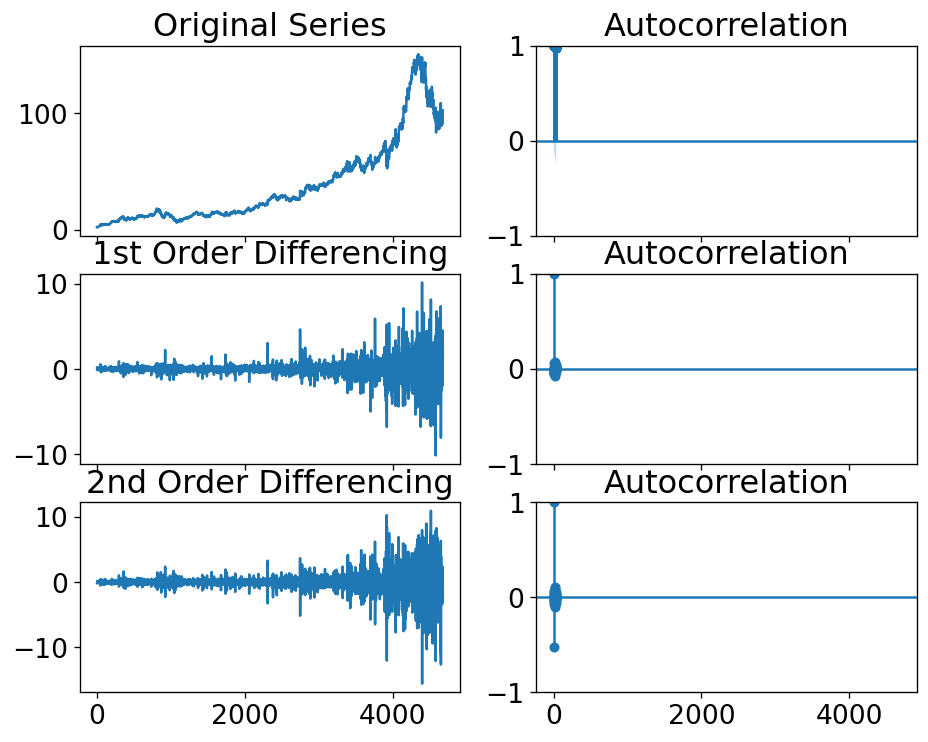
And then produced correlation among features using a heatmap.



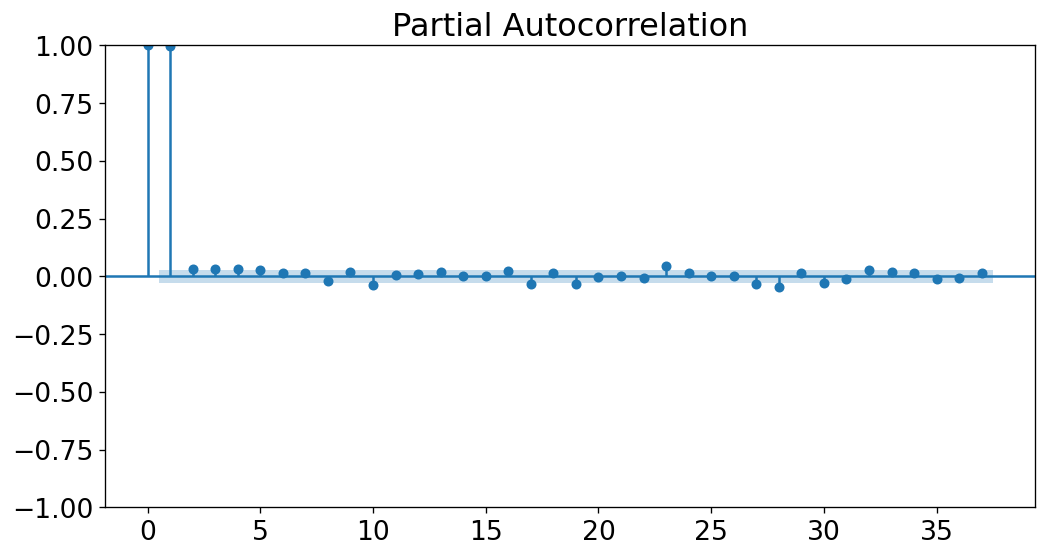
**4.2 Time Series Prediction Models**

**a. ARIMA Models**

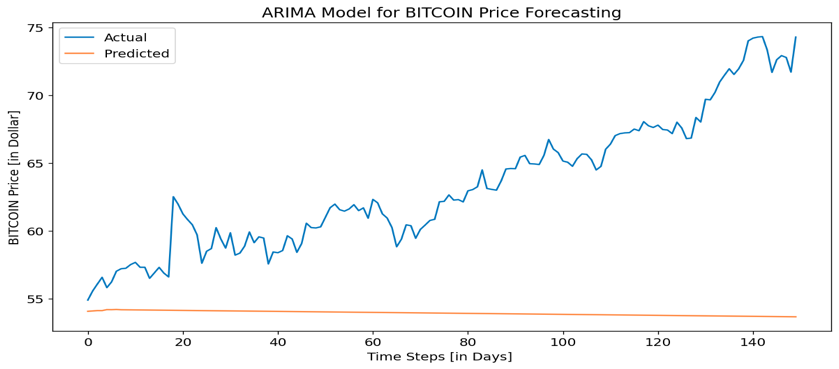
In this Model we explored Auto Regressive Integrated Moving Average (ARIMA) for univariate time series forecasting and conducting exploratory data analysis (EDA) by observing original series, 1st differencing, and 2nd differencing.



To explain more by determining the order of the ARIMA model using the Partial Autocorrelation Function (PACF) plot.



Later we built and trained the ARIMA model on the training data. Then it evaluates the model and forecasted stock prices for the test data and plotted the actual vs. predicted results.



**b. Artificial Neural Network (ANN)**

**-** selected relevant features (Close and Volume) for the prediction.

- Explored a Multi-Layer Perceptron (MLP) model using Keras for univariate time series forecasting.

- Splits the data into training and testing sets.

- Conducted feature engineering and reshaped the data for model input.

- Built, trained, and evaluated the MLP model.

- Visualized the actual vs. predicted results.

A graph showing a line graph

Description automatically generated with medium confidence

**c. Recurrent Neural Network (RNN)**

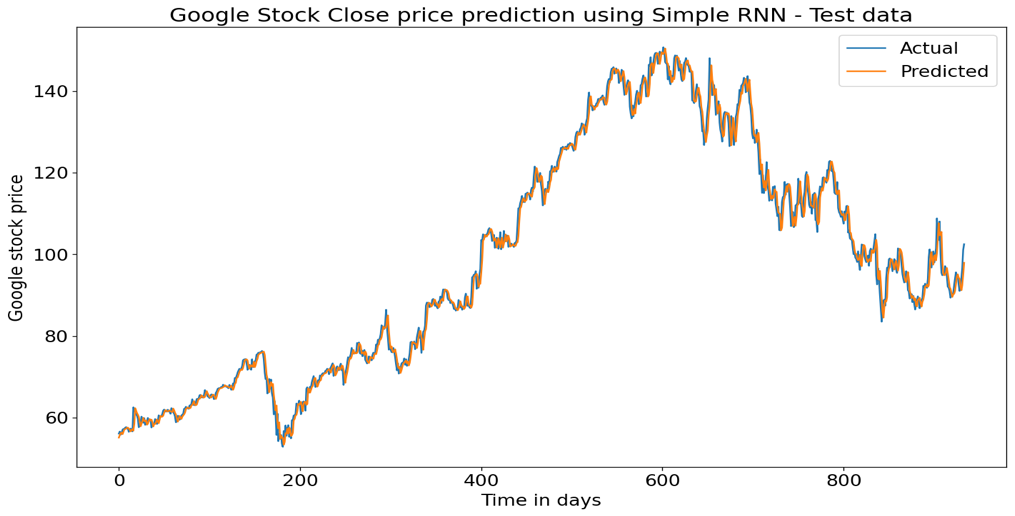
- Applied a Simple Recurrent Neural Network (SimpleRNN) for time series forecasting.

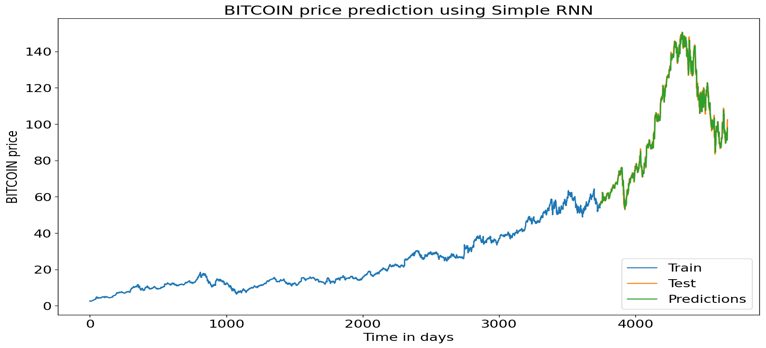
- Determined the lag value from the PACF plot

- Prepared the data for input by creating sequences.

- Built, trained, and evaluated the SimpleRNN model.

- Plotted the actual vs. predicted results.





**d. Model Evaluation**

- Employed various evaluation metrics such as R-squared, Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Normalized RMSE (NRMSE), Weighted Mean Absolute Percentage Error (WMAPE) etc.

- Applied these metrics to assess the performance of each model on the test data.

**e. Long Short-Term Memory (LSTM)**

**-** Utilized LSTM, a type of RNN, for time series for forecasting.

- Chose the lag value and reshaped the data for input.

- Built, trained, and evaluated the LSTM model.

-Plotted the actual vs. predicted results.

A graph showing a line graph

Description automatically generated with medium confidence

**5. CONCLUSION**

In conclusion, the experiment involved the application of various time series prediction models to forecast Google stock prices. The ARIMA model, artificial neural network (MLP), recurrent neural network (SimpleRNN), and long short-term memory (LSTM) were explored. Each model was analyzed, trained, and evaluated on the dataset. The evaluation metrics provided insights into the accuracy and reliability of each model in predicting stock prices. The experiment demonstrated the versatility of different models and highlighted the significance of careful model selection and parameter tuning in achieving accurate prediction.

**6. REFERENCES**

Hochreiter, S & Schmidhuber, J 1997, ‘Long Short-Term Memory’, *Neural Computation*, vol. 9, no. 8, pp. 1735–1780.

Krauss, C, Do, XA & Huck, N 2017, ‘Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500’, *European Journal of Operational Research*, vol. 259, no. 2, pp. 689–702.

Sang Il Lee & Yoo, SJ 2019, ‘Multimodal Deep Learning for Finance: Integrating and Forecasting International Stock Markets’, *IDEAS Working Paper Series from RePEc*.

Schmidhuber, J 2015, ‘Deep learning in neural networks: An overview’, *Neural Networks*, vol. 61, pp. 85–117.