Employing ARIMA And LSTM Deep Learning Models for Time Series Analysis

Name: Friday Odeh Ovibiroro

Student Number: 22034665

Introduction

This report evaluates the efficacy of two commonly utilized models, ARIMA and LSTM, in forecasting sales and share data within time series contexts. ARIMA models are chosen for their adeptness in capturing sequential data patterns and trends, particularly when dealing with autocorrelation and seasonality. Conversely, LSTM networks demonstrate proficiency in sequence prediction due to their capability to comprehend and maintain long-term dependencies within the data, facilitating the capture of intricate temporal patterns and trends. The aim of this report is to construct time series models capable of predicting data for 24 periods and evaluate each model's performance to ascertain their accuracy and effectiveness in prediction.

Data Exploration

After investigating the data, distinct trends have been observed in both Johnson & Johnson and Amazon share prices. Figure 1 depicts the consistent upward trajectory of Johnson & Johnson's share price, interspersed with periodic fluctuations indicating the presence of seasonality. This suggests the suitability of a Seasonal Autoregressive Integrated Moving Average (SARIMA) model for capturing the underlying patterns in the data and making accurate forecasts. Conversely, Figure 2 illustrates the more volatile nature of Amazon's share price, characterized by fluctuating movements in both upward and downward directions over time. This indicates the potential necessity of an ARIMA modeling approach to effectively capture and predict the dynamics of Amazon's share price.

Fig 1

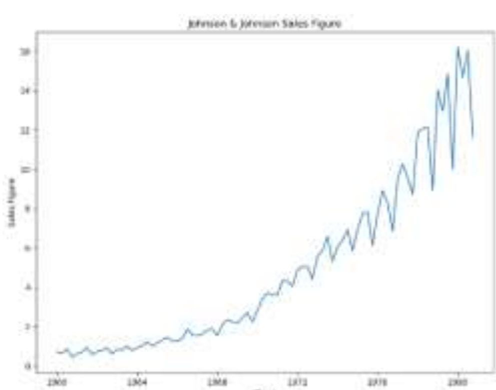
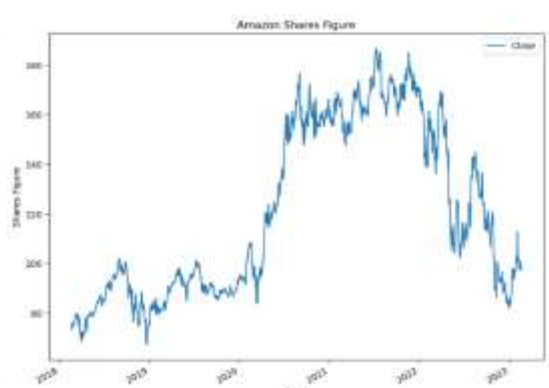


Fig 2



Stationarity Check

The stationarity of both datasets underwent assessment using the Augmented Dickey-Fuller (ADF) statistical test. The findings revealed that Johnson & Johnson's dataset displayed non-stationarity, yielding a p-value of 1, while Amazon's share data also demonstrated non-stationarity with a p-value of 0.45. To rectify this, we implemented log transformation and differencing on both datasets. Following the transformations, the ADF test was rerun. Consequently, Johnson & Johnson's dataset exhibited significant improvement towards stationarity, evidenced by a decreased p-value of 0.0004. Similarly, Amazon's share data confirmed stationarity post-transformation, yielding a p-value of 0.0. This transformation is observable in the visual representation, showcasing a constant mean across both datasets.

Fig 3

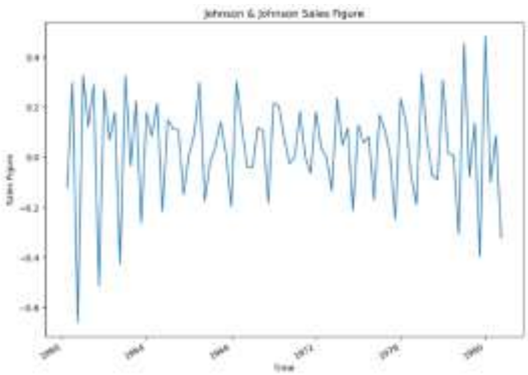
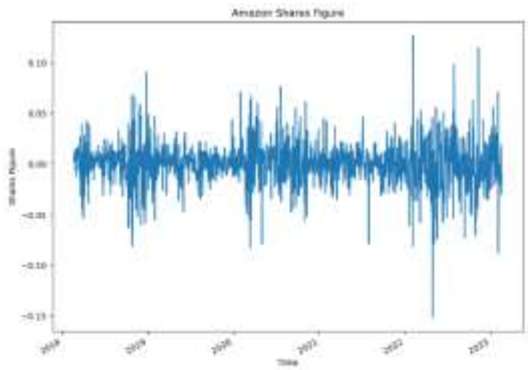


Fig 4



Johnson and Johnson Arima Model

A thorough analysis of the Johnson & Johnson dataset was conducted to identify the optimal model. Two key factors were taken into account: minimizing the AIC value derived from the Auto Arima function and selecting the most suitable model. The selected model, SARIMAX (1, 1, 3) (0, 0, 0, 4), features an autoregressive order of 1, a differencing degree of 1, and a moving average order of 3. Additionally, it includes a seasonal period of 4, signifying quarterly seasonality.

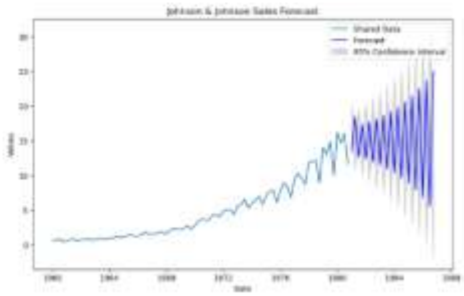
Evaluation Metrics for Johnson & Johnson ARIMA Model Using RMSE And MAE:

The performance of the ARIMA model in analyzing the Johnson and Johnson dataset was primarily assessed using two key metrics: the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE). These metrics provide insight into the extent of differences between the predicted and actual data points. The computed values for RMSE and MAE were 2.25 and 2.01, respectively. These figures indicate the average size of the deviations between the predicted values and the actual observations.

Forecasting Johnson and Johnson Sales Using the Arima model:

Below is a visual representation of sales figures forecast for Johnson & Johnson sales, showcasing past sales in blue and projected future sales in dark blue. Notably, the graph reflects a positive trend, indicating a consistent increase in sales over time. The light-shaded area denotes a 95% confidence interval for the future sales forecast. It's noteworthy that as time advances, this interval widens, implying an increase in the uncertainty associated with predicting future sales.

Fig 5



Johnson & Johnson Lstm Model

The architecture of the neural network model for Johnson & Johnson consists of three LSTM layers organized sequentially, followed by a dense layer. In the first LSTM layer, 30 output units are generated, which are then expanded to 60 units in the second LSTM layer. The third LSTM layer maintains these 30 units without modification. This model is specifically designed to analyze sequences spanning 20-time steps to predict future data points.

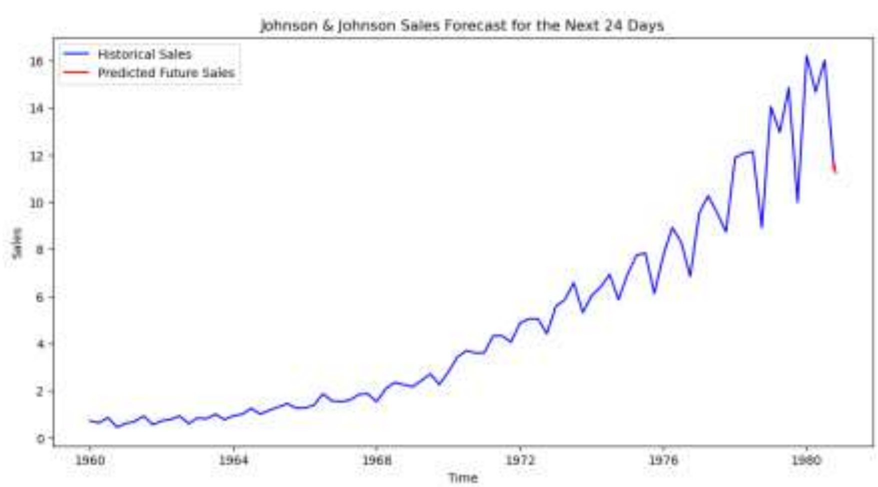
Prediction and Evaluation of Johnson and Johnson model using LSTM

The evaluation of Johnson & Johnson's sales data utilizing the LSTM model reveals consistent Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) values, both standing at 1.07. These metrics indicate that, on average, the model's sales predictions diverge from actual sales figures by 1.07 units.

Johnson and Johnson Sales Forecast with LSTM Model

The model was deployed to predict Amazon sales over a 24-period horizon. The visual depiction of this forecast indicates a downward trend suggesting that sales are expected to decrease over the next 24 days. The prediction may be based on various factors like seasonal effects, market trends, or other external factors that the prediction model has considered.

Fig 6



Arima Model Amazon Shares

In selecting the model for the Amazon dataset, the primary criterion was to minimize the Akaike Information Criterion (AIC) through the application of the Auto Arima function. Following this methodology, the optimal model identified for the Amazon dataset is ARIMA (2, 1, 2). This model is characterized by an autoregressive order of 2, a differencing degree of 1, and a moving average order of 2.

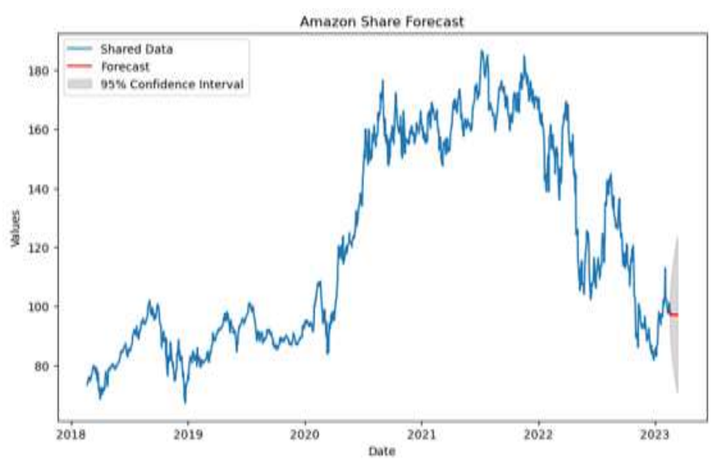
Evaluation Metrics for Amazon ARIMA Model Using RMSE And MAE:

In evaluating the performance of Amazon's model shares, the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) were used. These metrics serve as crucial indicators of the disparity between predicted and actual data points. The calculated RMSE and MAE values, standing at 3.13 and 2.76 respectively, shed light on the average magnitude of discrepancies between the predicted and actual values, thus offering valuable insights into the model's accuracy.

Amazon Shares Forecast Using the Arima model:

The forecast displayed below presents anticipated future values of Amazon shares. Historical share prices are illustrated in blue, while projected values are indicated in red. The gray area represents the 95% confidence interval for the forecast, suggesting a 95% likelihood that forthcoming share values will fall within this range. While the forecast hints at a slight decrease in share price in the near future, the considerable width of the confidence interval underscores the notable uncertainty surrounding this prediction.

Fig 7



Lstm Model for Amazon Shares

The neural network model for Amazon shares employs a series of four LSTM layers, followed by a dense layer. Initially, the first LSTM layer generates 30 output units, which are then maintained at 30 units by the second LSTM layer. The third and fourth LSTM layer preserves this same dimensionality without any increase. Specifically designed to analyze sequences spanning 30-time steps, this model aims to predict future data points effectively.

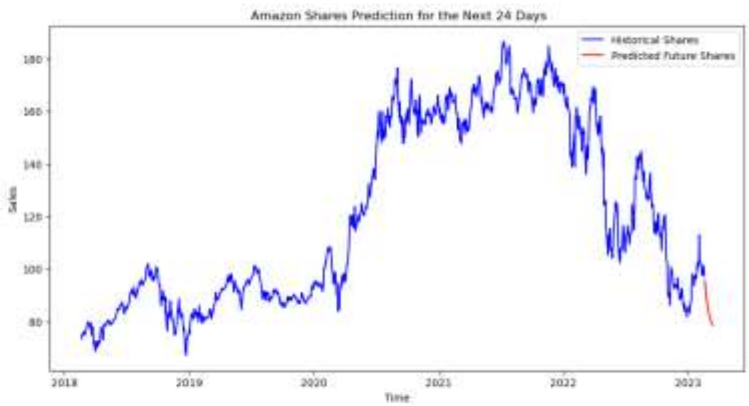
Model Prediction and Evaluation of Amazon model using LSTM

The assessment of the LSTM model on Amazon stock data reveals that both the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are calculated as 3.37 and 2.6, respectively. These metrics provide insight into the typical size of discrepancies between the model's projected sales and the actual sales figures. On average, the model's forecasts differ from the actual values by approximately 3.37 and 2.6 units.

Forecasting Amazon shares With LSTM Model

Following the completion of the prediction and evaluation stages, the model was deployed to forecast Amazon share prices over a 24-period horizon. The plot below illustrates the forecast for future share prices, indicating an upward trend from the last recorded historical data point. However, it is important to note that this prediction offers only a short-term view for the next 24 days, and it suggests a potential downward trend in the immediate future.

Fig 8



Conclusion

In conclusion, the evaluation of the ARIMA model for Johnson & Johnson's dataset highlighted a reasonable performance, with RMSE and MAE values of 2.25 and 2.01, respectively. However, the LSTM model outperformed significantly, with consistently lower RMSE and MAE scores of 1.07, indicating its superior capability in modeling and predicting sales data.

Moving to Amazon's stock data, both ARIMA and LSTM models were evaluated using RMSE and MAE metrics. While the ARIMA model demonstrated satisfactory performance with RMSE and MAE values of 3.13 and 2.76, respectively, the LSTM model showcased improved forecasting ability. With slightly better RMSE and notably lower MAE values of 3.37 and 2.6, respectively, the LSTM model demonstrated closer predictions to actual share prices.

Overall, the results underscore the effectiveness of LSTM models in handling time series data, particularly in capturing intricate patterns and improving forecasting accuracy. However, the enduring relevance of ARIMA models should not be overlooked, especially in scenarios where the data lacks complex temporal dependencies.

References

Chen, J. (2023). Analysis of Bitcoin Price Prediction Using Machine Learning. *Journal of Risk and Financial Management*, 16(1), p.51. doi:https://doi.org/10.3390/jrfm16010051.

Ho, M.K., Darman, H. and Musa, S. (2021). Stock Price Prediction Using ARIMA, Neural Network and LSTM Models. *Journal of Physics: Conference Series*, 1988(1), p.012041. doi:https://doi.org/10.1088/1742-6596/1988/1/012041.

Siami-Namini, S., Tavakoli, N. and Siami Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). doi:https://doi.org/10.1109/icmla.2018.00227.