IT1244
Predicting Stock Closing Price

Group 29: Joash Lim, Maverick Loh, Leong Wynthia, Tan Jun Yi Brandon
National University of Singapore

Introduction

Predicting the US stock market is extremely challenging due to stock prices' dependence on many variables. Traditional methods to predict stock prices include technical analysis and fundamental analysis which may not always capture the full spectrum of market dynamics. Models like **ARIMA (Autoregressive Integrated Moving Average)** have been standard in financial time series analysis. However, their linear nature and assumptions about data stationarity can limit their predictive power in the volatile financial market. AI models, through deep learning techniques like **LSTM (Long Short-Term Memory)** and **GRU (Gated Recurrent Unit)**, can model the stock market more accurately by learning sequentially from complex patterns in historical data (Liu 2023; Zhang 2023). These methods can process large amounts of data more efficiently, assisting investors in making informed decisions. AI and machine learning also excel in handling non-linear relationships of financial data more effectively (Hendahewa et al. 2021). However, due to our dataset only spanning 2 financial years, LSTM would be prone to overfitting due to the large number of parameters. In our investigations, we explore and compare the use of machine learning models based on the **Auto Arima - SARIMAX (Seasonal Arima with Exogenous Variables)** and **LSTM** models to predict closing prices of stock traded in the US stock exchange.

Dataset

The main dataset contains time series data of stock price, trade volume, news events and news sentiment for **S&P 500** companies during the period *30/9/20 to 30/6/22* but upon further inspection we neglect the data corresponding to the 30/9/20 as it included empty values in news sentiments. As such, our cleaned data spanned from *01/10/20 to 30/6/22*. The data set comprises 495 securities belonging to 11 distinct sectors and 120 distinct sub-industries. The data is largely consistent with each security having 441 days of data apart from 2 securities, namely Constellation Energy and Organon & Co. which have only 113 and 285 days of data respectively (Table 1a-b).

Exploratory Data Analysis

Selecting Securities and Sectors of Interests

We chose GICS Sectors which are major players in the overall economy, based on their aggregated market capitalization index weights determining the key sectors to be Information Technology, Consumer Discretionary, Communication Services, Health Care, and Industrials as shown in Figure A. We selected our top companies of each security by their aggregated index weights during the duration of our time series data set. The 5 companies of interest are Apple, Tesla, Meta, Moderna and Boeing. We narrowed our scope as the largest stocks usually greatly impact the long-term performance of the index.

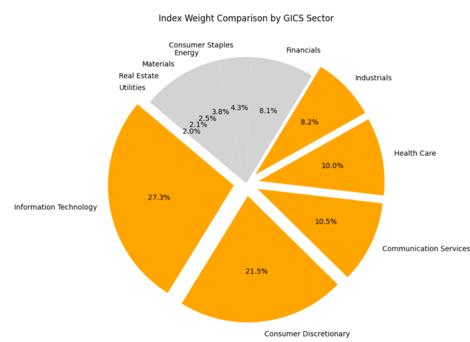

Figure A: Index Weight Comparison by GICS Sector.

Closing Prices of Each Security

We observe from the closing price of stocks over time that they display varying trends and levels of volatility. Tesla shows a sharp rise and fall, indicating high volatility which is typical for tech stocks while Moderna's stock exhibits extreme volatility with a significant peak, which may relate to events of COVID 19 (Figure 1a-e).

Rolling Mean and Standard deviation

We observe in stocks like Apple and Meta, the rolling mean remains consistent while trends are visible for Tesla and Moderna, which suggest non-stationarity in the time series, indicating the presence of cycles that could be seasonal. The rolling standard deviation for these stocks is relatively low and stable, indicating that the volatility of the stocks does not vary widely over time (Table 2a-e).

Feature Transformation

Engle's ARCH Test

Engle's ARCH (Autoregressive Conditional Heteroskedasticity) test is designed to test if the variance of time series residuals depends on their past values. This step is crucial for ARIMA and LSTM models, which perform better with input data that has consistent variance and minimal skewness. We obtained a p value of < 0.05 when applied to the entire S&P 500 data therefore log transformation is justified to normalise the distribution of data. In addition, we also standardised our data to further reduce the skewness of the features.

Methods for Feature Selection

For each security, we have 20 features we can use to predict the closing prices. Several methods are used to select features that seem to influence the stock's closing price. The methods implemented are ANOVA, Principal Component Analysis and Random Forest.

ANOVA Test

ANOVA (Analysis of Variance) is a statistical method that uses the F test to check if the means of two or more groups are significantly different from each other.

- H0: Means of all groups are equal.
- H1: At least one of the groups is different.

If there is equal variance between groups, it means this feature has little impact on predicting the target and will not be considered. After performing the test on the dataset, all features have significant differences.

Principal Component Analysis

Principal Component Analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The resulting vectors (principal components) are an uncorrelated orthogonal basis set. Results are shown in Figure 3.

Random Forest

Random Forest is an ensemble learning method used for both classification and regression tasks, which operates by constructing multiple decision trees during training time and outputting mean prediction of the individual trees. We used Random Forest for our feature selection process where each tree in the Random Forest is trained on a random subset of the data and a random subset of features. At each split in the tree, the algorithm will choose the feature that maximises the decrease in impurity, a visualisation is shown in Figure 4. In regression tasks, impurity is measured by variance. The average decrease in impurity across all trees is calculated for each feature, giving a measure of the overall importance of the feature. We used a threshold of 0.95 of cumulative importance to identify features that are instrumental in the model's predictions. Results are shown in Figure 5.

Conclusion for Feature Selection

After conducting our feature selection methods and data inspection, we have removed New Products, Layoffs, Dividends, Product Recalls, and Personnel Changes as well as Positive and Negative Sentiments and Stock Rumors due to a disproportionate number of "0" in the data (Figure 1a-b). After omitting the features, we are left with Volume, All News Volume, News Volume, Analyst Comments, Stocks and Adverse Events.

Methods for Model Building

We would be comparing the accuracy of prediction for Auto - Arima, LSTM and GRU for our prediction model study. A classical method like ARIMA would not be able to develop a general training model for each of the sectors identified due to the volatility of the average standard deviation. Therefore, we would be training ARIMA for each company only. Whereas for the deep learning models, we would train over data of the top 5 sectors before applying the model to the representative companies.

ARIMA

ARIMA (AutoRegressive Integrated Moving Average) models are particularly useful for financial time series data, which often exhibit volatile and non-stationary behaviour.

Stationarity

Stationarity is a concept in time series analysis that refers to the statistical properties of a series being constant over time, where mean, variance are consistent. Hence the data should not display seasonal effects. Stock data of *Apple, Tesla, Meta, Moderna* and *Boeing* have all demonstrated non-stationarity after applying decomposition (Figure 7a-e). ARIMA requires the data to be stationary to focus on developing forecasts based on the observed patterns.

ADF Test (Augmented Dickey-Fuller Test)

Augmented Dickey-Fuller test is a type of statistical test called a unit root test.

- *Null Hypothesis (HO)*: Series is non-stationary, or series has a unit root.
- *Alternate Hypothesis (HA)*: Series is stationary, or series has no unit root.

After performing the ADF Test on the five companies, we realise the p-value of all to be more than 0.05 indicating non-stationarity.

	Apple	Tesla	Meta	Moderna	Boeing
P-value	0.3380	0.2099	0.8807	0.4288	0.6622

Table A: P-Value for ADF Test.

ARIMA model is made up of 3 components, namely the AR, I and MA components:
In terms of y, the general forecasting equation is:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$$

AutoRegressive (AR) (p) is a regression model that captures the relationship between an observation y and several lagged observations until p-th time in the past.

$$\hat{y}_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t$$

Integrated (I) (d) represents the differencing of observations to make the time series stationary as observed in apple's stock data (Figure 6):

In general, a d th-order difference can be written as

$$y'_t = (1 - B)^d y_t$$

Moving Average (MA) models a regression model on the dependency between the observations and the residual error of lagged observations:

$$\hat{y}_t = c + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

However, ARIMA on its own is merely a statistical model that requires its own fine-tuning of parameters. To qualify as a machine learning model, we used the auto-arima package where an AIC minimising function is used to choose the model's best parameters.

AIC

Akaike Information Criterion (AIC) is a widely used measure for model selection. In the context of Auto ARIMA, the AIC is used to compare different ARIMA models with varying orders and seasonal components, selecting the one with a minimised AIC value. The AIC is defined as: AIC = 2k - 2ln(L) where k is the number of model parameters and L is the likelihood of the model.

SARIMAX

After running our Auto-Arima model, the Seasonal ARIMA with eXogenous variables model is selected for due to the seasonal behaviours exhibited by the stock data as well as the presence of news features. The seasonal part of the model is defined by four additional parameters: (P, D, Q, s), where "P" is the number of seasonal autoregressive terms, "D" is the number of seasonal differences, "Q" is the number of seasonal moving average terms, and "s" is the length of the seasonal cycle.

LSTM

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) architecture widely used in predicting stock market prices due to their ability to model sequential and time-dependent nature of financial time series data. They have a more complex structure than regular RNNs, consisting of three gating mechanisms— input, output, and forget gates. This allows them to capture long-term dependencies and patterns in sequential data, making them particularly suited for tasks like time series forecasting to a higher accuracy than more primitive models such as ARIMA.

GRU

Gated Recurrent Unit (GRU) is a type of Recurrent neural Network (RNN) architecture with a simpler but improved architecture compared to previous RNN models. Unlike LSTM, which consists of three gating mechanisms, GRU contains two: Update and reset gate.

- Update gate: Determines which information from the previous hidden state and current input to keep.
- Reset gate: Determines which information to discard.

The final hidden state combines the information retained by the update gate and the current input.

Fine-Tuning Hyperparameters

Hyperparameter tuning is an important part of deep neural networks, affecting accuracy, complexity, and time efficiency of the model. Our main hyperparameters are the type of optimizer, number of past days as input and batch size. We experimented with different combinations of the hyperparameters to obtain a balance between the three.

Errors to minimise

The typical errors we aim to minimise for Auto ARIMA, LSTM and GRU models are the Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) of the predicted and the true stock closing prices.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|y(i) - \hat{y}(i)\|^2}{N}}, \quad \text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{| y_i - \hat{y}_i |}{y_i} \ . \ 100\%$$

Results And Discussions

We analyse the average values of the key metrics used to evaluate our models for all the 5 companies, and have placed them in Table B. The individual values of the key metrics have been placed in Table 2.

	Auto - ARIMA	**LSTM**	**GRU**
RMSE	52.218	13.683	16.820
MAPE	0.146	0.0724	0.09691
Training Time (in seconds)	66.037	146.2	131.2

Table B: Evaluation of Auto-ARIMA, LSTM and GRU, using RMSE, MAPE, and Training Time.

From Figure 8-10, the Actual vs Predicted Closing Price for Auto-ARIMA, LSTM and GRU have been plotted to analyse the reliability of the models. It is seen that deep-learning models LSTM and GRU performed with greater success at predicting the stock prices.

Conclusion

The LSTM model demonstrates superior performance with the lowest RMSE of 13.683, followed by the GRU model with an RMSE of 16.820. Auto-ARIMA has the highest RMSE at 52.218, indicating it has the least accurate predictions among the three models. The LSTM model has a MAPE of 0.0724, suggesting it has the highest accuracy in forecasting. The GRU model comes next with a MAPE of 0.09691, while the Auto-ARIMA model has a MAPE of 0.155. When it comes to efficiency, Auto-ARIMA is the clear winner with a significantly lower training time of 66.037 seconds. LSTM and GRU models require considerably more time, at 146.2 and 131.2 seconds respectively. Hence, while they offer better accuracy, it is at expense of increased computational resources and time.

The LSTM model appears to be the most accurate in forecasting, with the lowest RMSE and MAPE values suggesting that LSTM is preferable for applications where accuracy is of great importance and computational resources are not a limiting factor. The GRU model also performs similarly well to LSTM. Auto-ARIMA, while being the fastest model to train, exhibits significantly lower forecasting accuracy, making it suitable for scenarios where speed is critical and a rough estimation is needed. We conclude that the LSTM model is the most suitable for investors to accurately predict the stock market while ARIMA can be used for short term estimated forecasting.

The reason why ARIMA performs worst is due to the fundamental linear relationship between past and future values. Therefore it falls short in capturing complex, non-linear patterns present in financial data that LSTM and GRU, being non-linear models, excel. LSTM and GRU models, which are designed to remember long-term dependencies in the data, makes them more adept at capturing patterns over longer sequences hence they were highly accurate in forecasting (Qu, Z 2020).

In conclusion, LSTM performs the best in forecasting the stock price in the US stock exchange, followed by GRU, and lastly, ARIMA.

References

Liu, L. 2023. A Comparative Examination of Stock Market Prediction: Evaluating Traditional Time Series Analysis Against Deep Learning Approaches. AEMPS Vol. 55: 196-204. DOI: 10.54254/2754-1169/55/20231008.

Zhang, L. 2023. A study of stock price prediction models based on time series analysis: LSTM and CNN. ACE Vol. 16: 118-126. DOI: 10.54254/2755-2721/16/20230879.

Hendahewa, D. N., Perera, A. A. D. S. N. A., De Meraal, T. D. N., Indikadulle, I. K. P. H., Perera, S. D. and Rathnayake, H. M. S. C. 2021. Comparative Analysis between Supervised Machine Learning and Time Series Models for stock price prediction, *2021 International Conference on Data Analytics for Business and Industry (ICDABI)*: 354-358, DOI: 10.1109/ICDABI53623.2021.9655888.

Qu, Z. 2020. Using LSTM in stock prediction and Quantitative Trading. https://cs230.stanford.edu/projects_winter_2020/reports/32066186.pdf

Appendix

	Date	Close	Volume	News - All News Volume	News - Volume	News - Positive Sentiment	News - Negative Sentiment	News - New Products	News - Layoffs	News - Analyst Comments
count	217811	217811	2.18E+05	217811	217811	217811	217811	217811	217811	217811
mean	2021-08-15 07:15:53	183.536161	4.88E+06	522740.3046	84.613729	4.343073	3.790346	1.317059	0.087773	10.454871
min	2020-09-30 00:00:00	3.85	6.00E+02	317745	0	0	0	0	0	0
25%	2021-03-10 00:00:00	59.5	9.00E+05	494005	11	0	0	0	0	2
50%	2021-08-16 00:00:00	111.910004	1.88E+06	531109	24	0	0	0	0	5
75%	2022-01-21 00:00:00	206.729996	4.38E+06	559155	58	1	1	0	0	10
max	2022-06-30 00:00:00	5959.330078	3.27E+08	669851	9769	4550	1182	1148	823	842
std	NaN	318.819043	1.08E+07	49908.52695	246.733269	28.30533	16.039958	9.156474	2.981767	31.121526

Table 1a: Summary Statistics for Data

	News - Stocks	News - Dividends	News - Corporate Earnings	News - Mergers & Acquisitions	News - Store Openings	News - Product Recalls	News - Adverse Events	News - Personnel Changes	News - Stock Rumors	Market Cap
count	217811	217811	217811	217811	217811	217811	217811	217811	217811	2.18E+05
mean	11.289861	0.476197	3.555895	2.115026	0.081685	0.119746	4.246103	0.80288	0.005032	5.14E+08
min	0	0	0	0	0	0	0	0	0	1.97E+05
25%	2	0	0	0	0	0	0	0	0	1.12E+08
50%	5	0	0	0	0	0	0	0	0	2.01E+08
75%	11	0	3	2	0	0	2	0	0	4.10E+08
max	858	173	772	1388	93	599	1679	1686	12	1.54E+11
std	32.589345	2.356074	15.594759	11.237718	0.972197	3.307637	18.099585	6.866718	0.100832	1.66E+09

Table 1b: Summary Statistics for Data

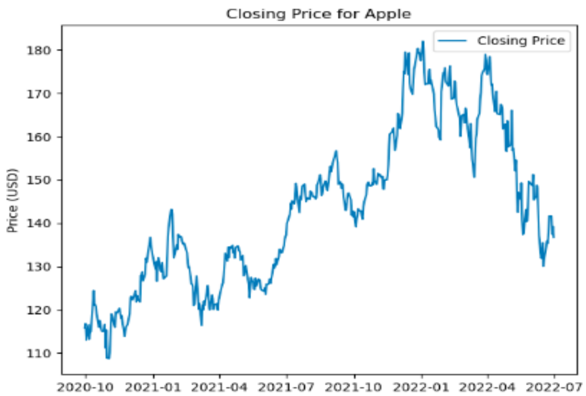

Figure 1a: Apple's Closing Price

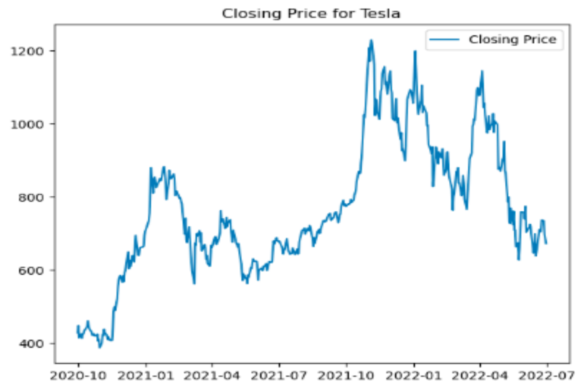

Figure 1b: Tesla's Closing Price

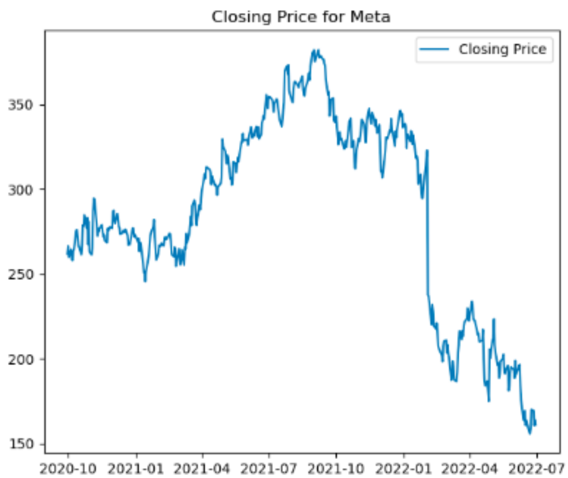

Figure 1c: Meta's Closing Price

Figure 1d: Moderna's Closing Price

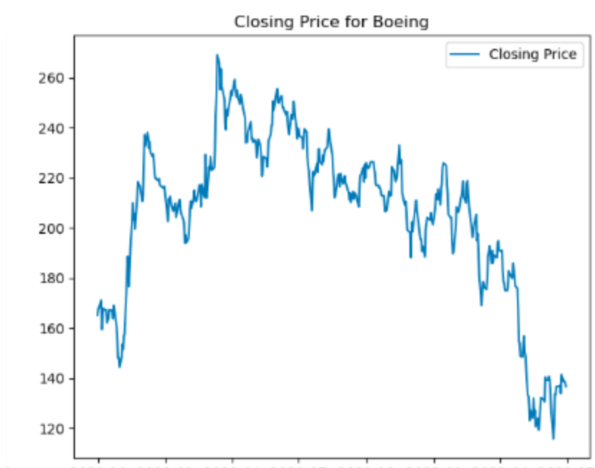

Figure 1e: Boeing's Closing Price

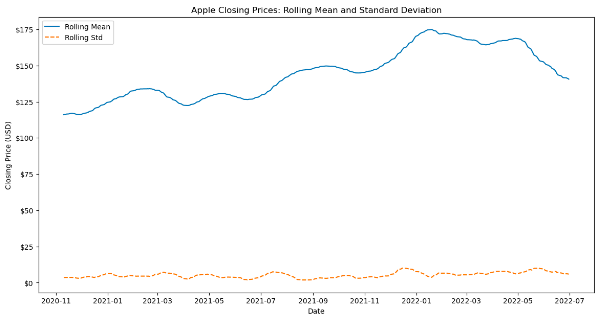

Figure 2a: Apple's Closing Price: Rolling Mean and Standard Deviation

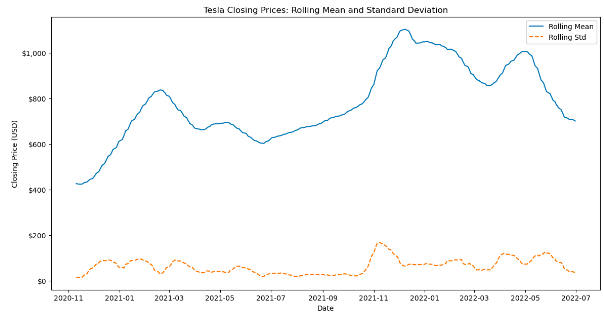

Figure 2b: Tesla's Closing Price: Rolling Mean and Standard Deviation

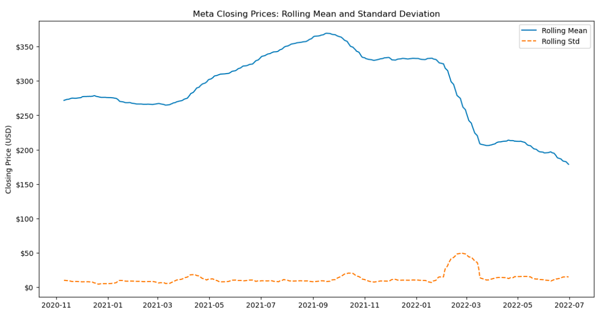

Figure 2c: Meta's Closing Price: Rolling Mean and Standard Deviation

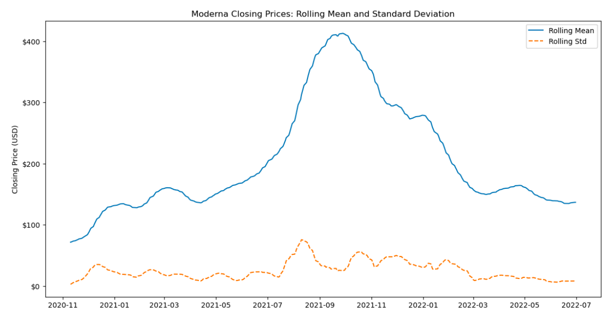

Figure 2d: Moderna's Closing Price: Rolling Mean and Standard Deviation

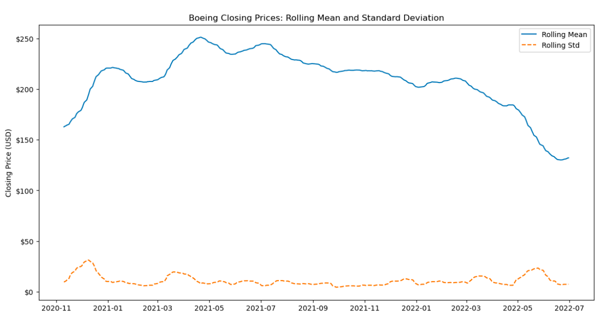

Figure 2e: Boeing's Closing Price: Rolling Mean and Standard Deviation

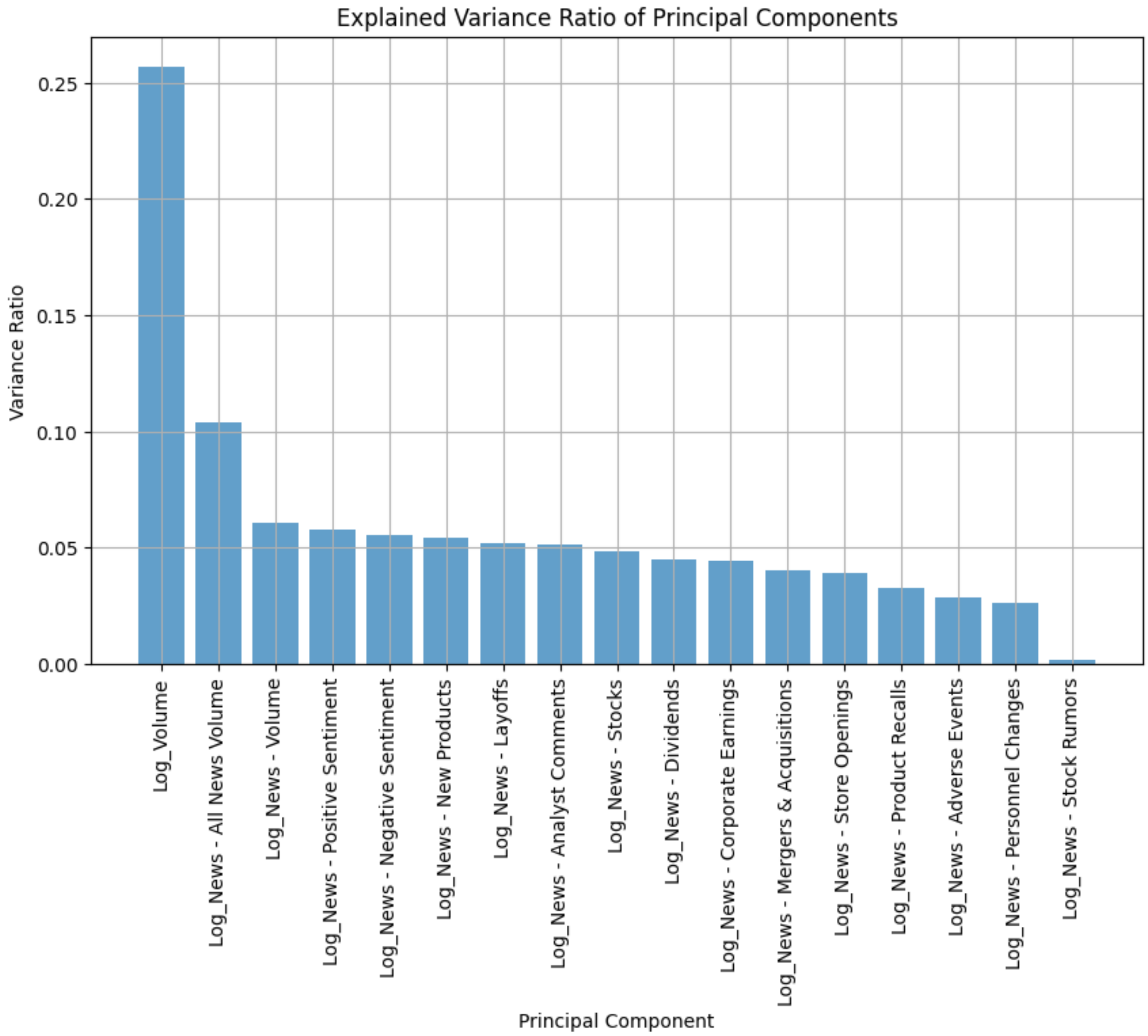

Figure 3: Principal Component Analysis of Log S&P 500 Data

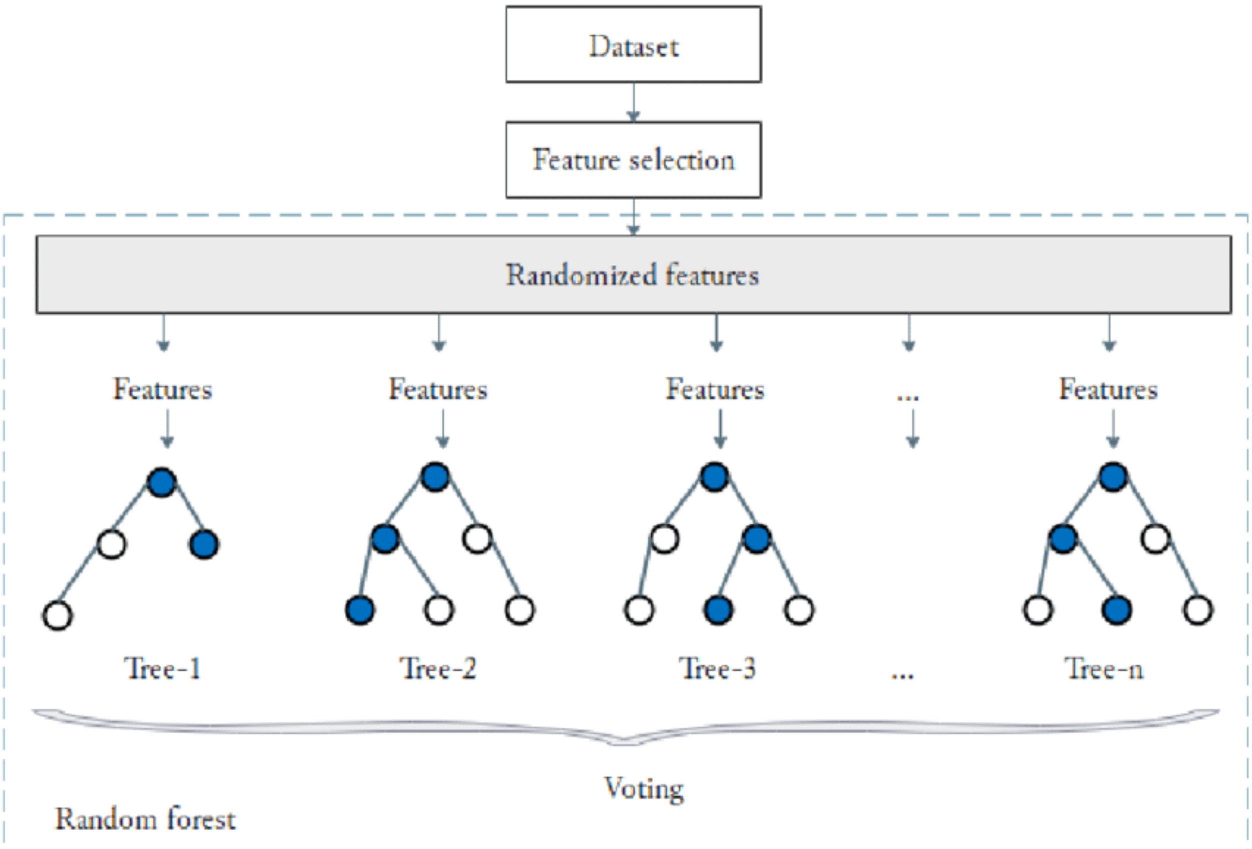

Figure 4: Random Forest Feature Selection

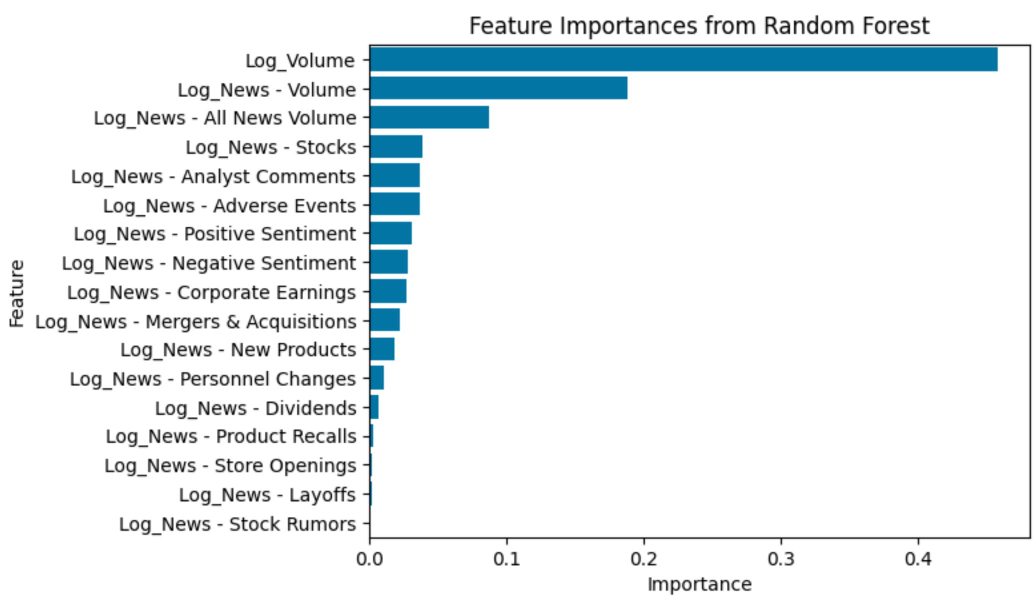

Figure 5: Random Forest Analysis of Log S&P 500 Data

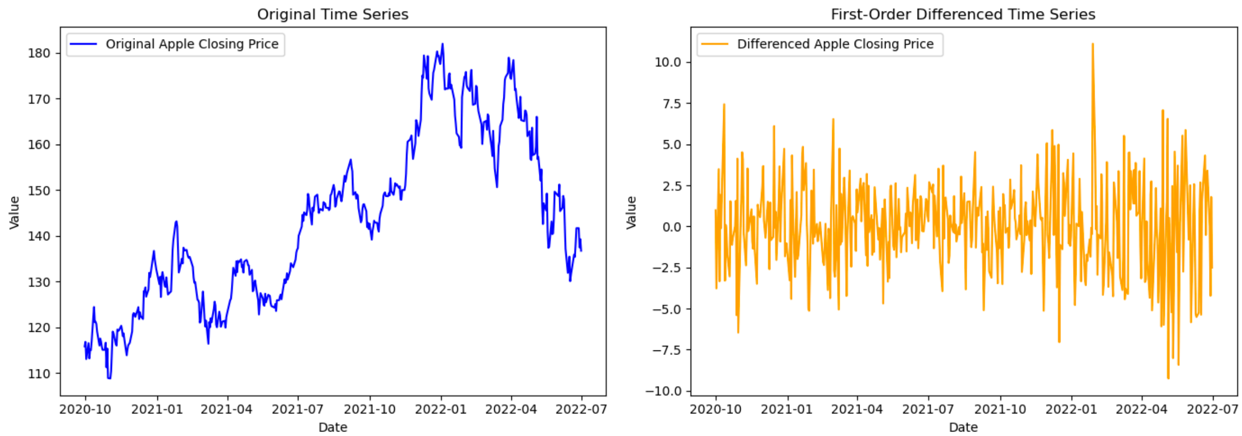

Figure 6: Comparison between Original Time Series and First-Order Differenced Time Series

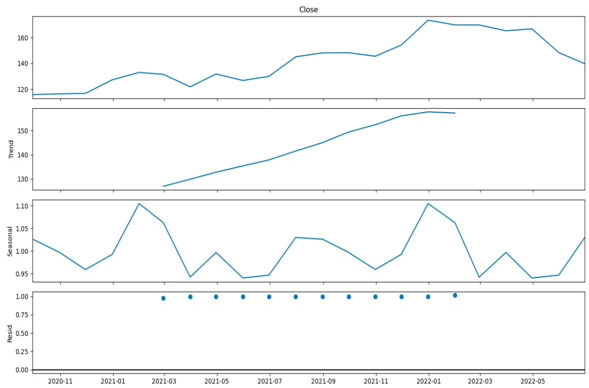

Figure 7a: Apple's Data Decomposition

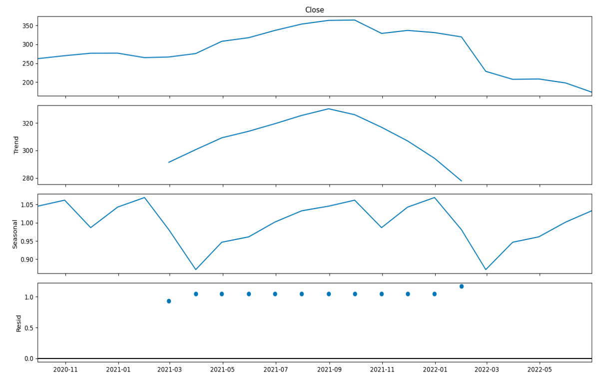

Figure 7b: Tesla's Data Decomposition

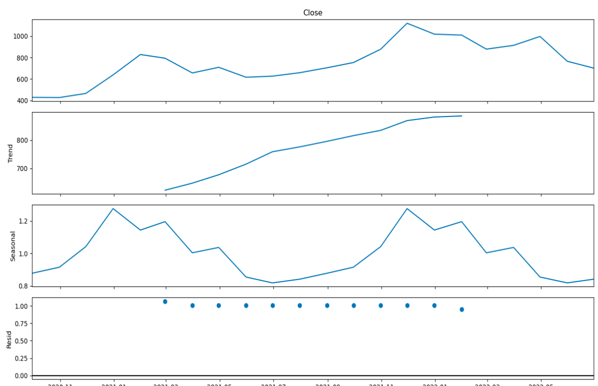

Figure 7c: Meta's Data Decomposition

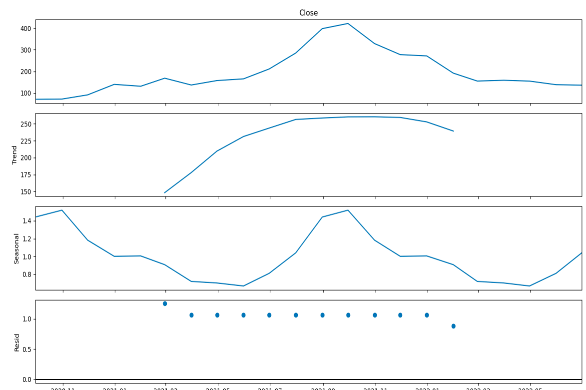

Figure 7d: Moderna's Data Decomposition

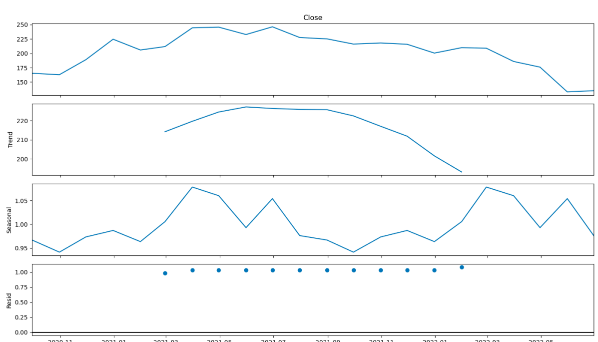

Figure 7e: Boeing's Data Decomposition

	Auto-ARIMA	LSTM	GRU
RMSE	20.27440151342699	4.07428542860713	3.9124929286530183
MAPE	0.1351334995163001	0.061594677713364	0.0596007677098928
Training Time (in seconds)	78.970	202	163

Table 2a: Results from Training Over Information Technology for Apple

	Auto-ARIMA	LSTM	GRU
RMSE	178.54175874102916	36.416892334599176	37.903736064004484
MAPE	0.2392794956909402	0.08582022234832247	0.09495807617068501
Training Time (in seconds)	162.294	143	136

Table 2b: Results from Training Over Consumer Discretionary for Tesla

	Auto-ARIMA	LSTM	GRU
RMSE	35.39737720243397	8.08252047448049	14.184871434180419
MAPE	0.17919024786773696	0.09675944953645463	0.11889096258643943
Training Time (in seconds)	25.552	59	51

Table 2c: Results from Training Over Communication Services for Meta

	Auto-ARIMA	LSTM	GRU
RMSE	10.947269922006555	13.034058237940041	15.65572748048723
MAPE	0.06784721685428068	0.09721222770293472	0.11407051550652934
Training Time (in seconds)	35.505	165	144

Table 2d: Results from Training Over Health Care for Moderna

	Auto-ARIMA	LSTM	GRU
RMSE	15.931457683672976	6.808294657184514	12.442405707702264
MAPE	0.10876744503349052	0.07606892060563412	0.09703197291300442
Training Time (in	27.864	162	162

seconds)			

Table 2e: Results from Training Over Industrials for Boeing

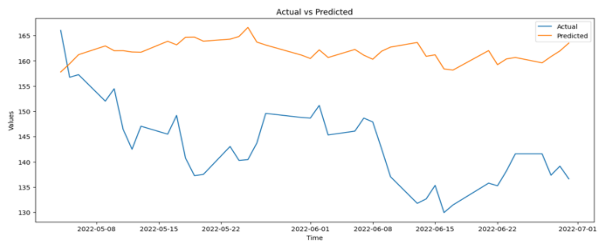

Figure 8a: Apple's Actual vs Predicted Closing Price for ARIMA (Last 40 Days)

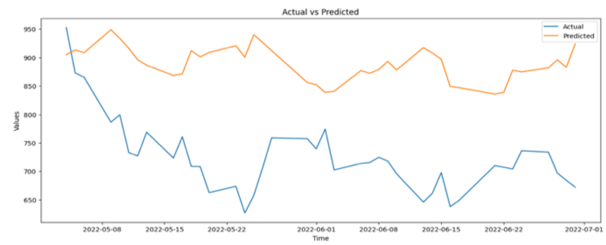

Figure 8b: Actual vs Predicted Closing Price for ARIMA (Last 40 Days)

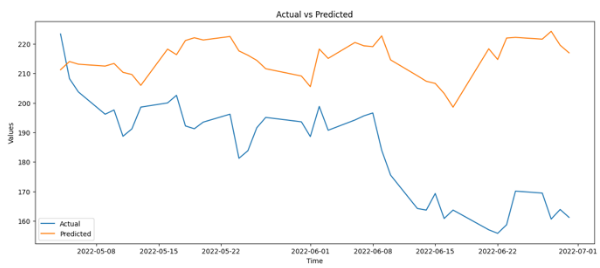

Figure 8c: Meta's Actual vs Predicted Closing Price for ARIIMA (Last 40 Days)

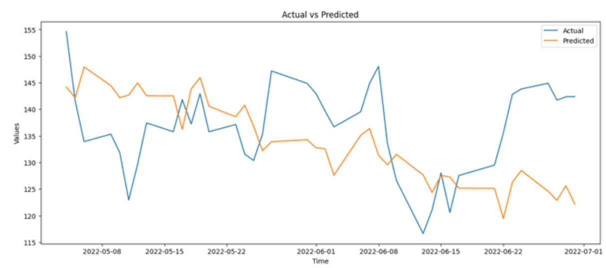

Figure 8d: Moderna's Actual vs Predicted Closing Price for ARIMA (Last 40 Days)

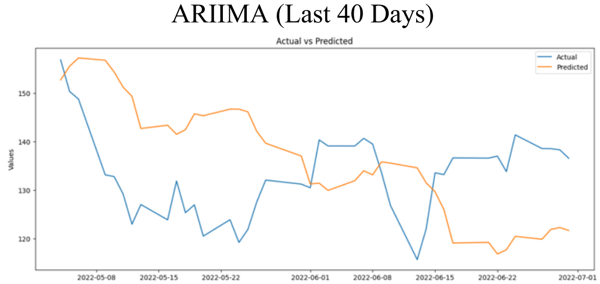

Figure 8e: Boeing's Actual vs Predicted Closing Price for ARIMA (Last 40 Days)

Figure 9a: Apple's Actual vs Predicted Closing Price for LSTM (Last 40 Days)

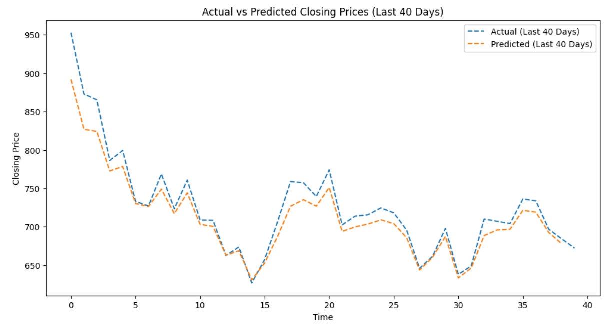

Figure 9b: Tesla's Actual vs Predicted Closing Price for LSTM (Last 40 Days)

Figure 9c: Meta's Actual vs Predicted Closing Price for LSTM (Last 40 Days)

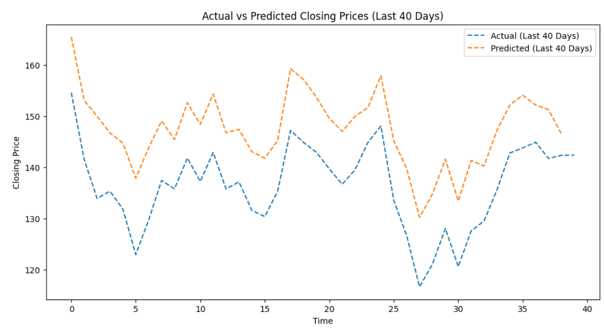

Figure 9d: Moderna's Actual vs Predicted Closing Price LSTM (Last 40 Days)

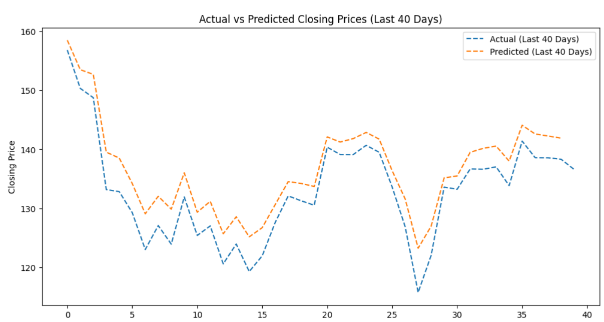

Figure 9e: Boeing's Actual vs Predicted Closing Price for LSTM (Last 40 Days)

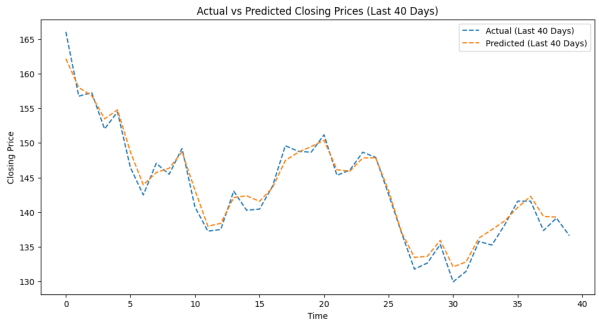

Figure 10a: Apple's Actual vs Predicted Closing Price for GRU (Last 40 Days)

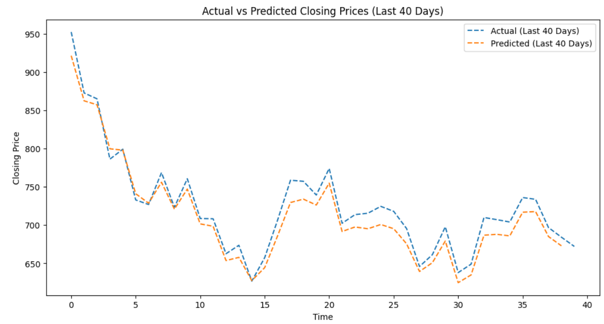

Figure 10b: Tesla's Actual vs Predicted Closing Price for GRU (Last 40 Days)

Figure 10c: Meta's Actual vs Predicted Closing Price for GRU (Last 40 Days)

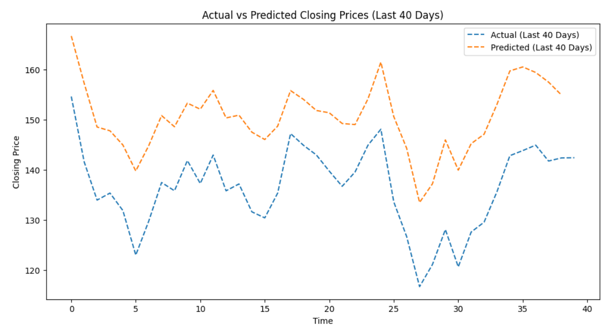

Figure 10d: Moderna's Actual vs Predicted Closing Price for GRU (Last 40 Days)

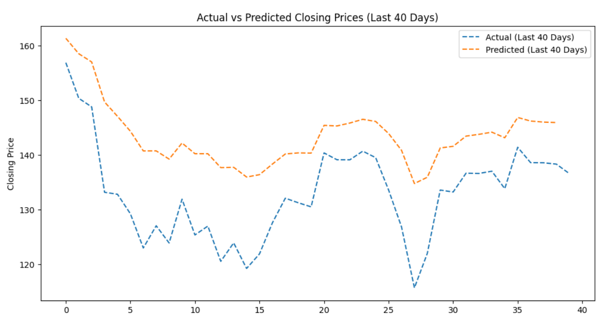

Figure 10e: Boeing's Actual vs Predicted Closing Price for LSTM (Last 40 Days)

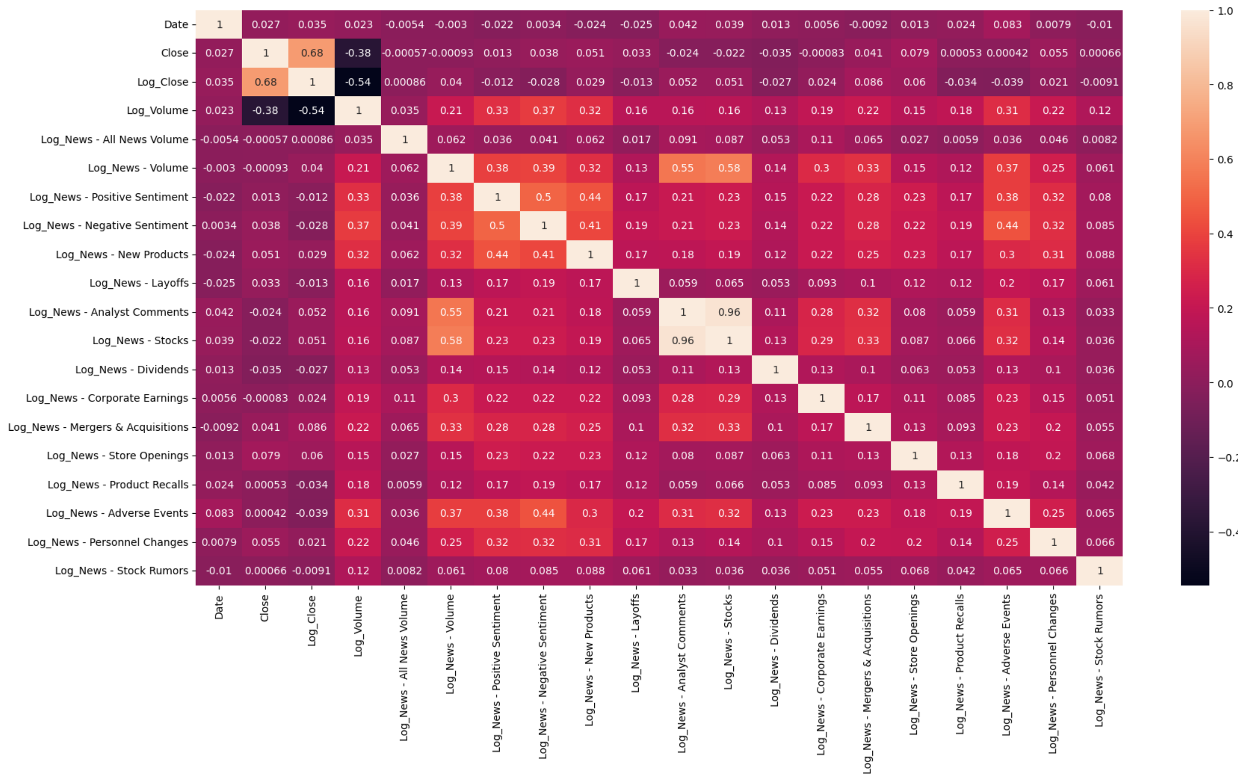

Figure 11: Heat Map for Logged Features