Time Series Forecasting

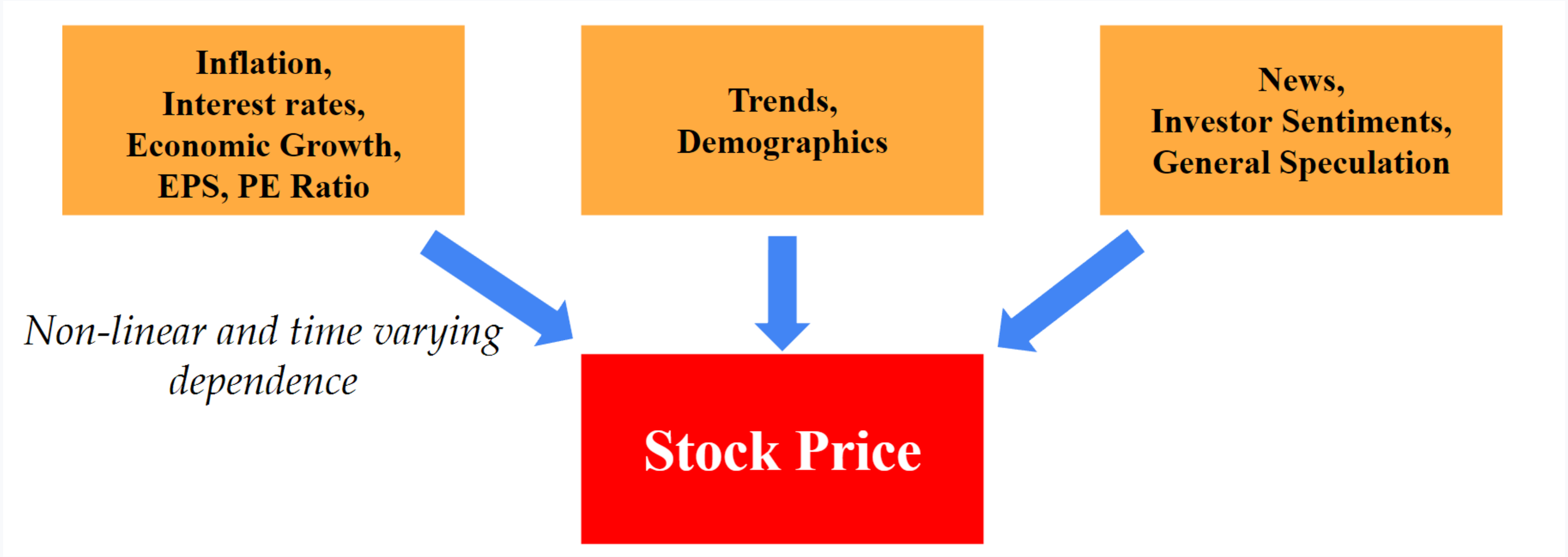
Predicting Closing Price of
S&P 500 Stocks

Group 29
Joash Lim, Maverick Loh,
Leong Wynthia, Tan Jun Yi Brandon

Introduction

Predicting Stock Prices

Predicting Stock Prices is **extremely challenging** due to the stock prices being dependent on a **multitude of factors**. The relationship between these factors are typically **non-linear and varies with time**.



Conventional Methods to Predict Stock Prices

Technical And Fundamental Analysis

- Technical Analysis is the forecasting of price movement using historical data of price trends.
- Fundamental analysis uses ratios like P/E ratio to analyze the strength of the company and by extension the value of the stock.

Statistical Methods: ARIMA

Model easily expandable to account for many variables that might affect stock price.

Downsides Of Conventional Methods

- Reliant on past patterns and fundamental indicators without accounting for **unexpected market movements**.
- In general they **do not capture the complex dynamics** of financial market.
- Limited to **short term prediction**

AI And Machine Learning Methods

LSTM (Long Short-Term Memory)

Liujun Liu (University of California) showed that LSTM model surpasses the conventional ARIMA model in terms of forecasting accuracy of Apple Inc. stock.

GRU (Gated Recurrent Unit)

Chi Chen (China School of Communication and Information Engineering) demonstrated a substantial improvement in prediction accuracy across various industries using a GRU model. GRU uses a more generalized model, more comprehensive training leading to less overfitting.

Our Purpose Of Study

In our investigations, we explore and compare the use of machine learning models based on the **Auto Arima - SARIMAX** (Seasonal Arima with Exogenous Variables), **LSTM** models and **GRU** to predict closing prices of stock traded in the US stock exchange.

Data Set

Description

- **S&P 500** companies
- **495 securities** belonging to **11 distinct sectors** and **120 distinct sub-industries**
- Time series data includes information such as closing price, volume traded and various types of news reported for each day from **September 2020** to **June 2022**.

Inconsistencies

- Each security includes **441 days of data** apart from 2 securities, namely **Constellation Energy** and **Organon & Co.** which have only 113 and 285 days of data respectively
- 30/9/20 to 30/6/22 but upon further inspection we **neglect** the data corresponding to the **30/9/20**

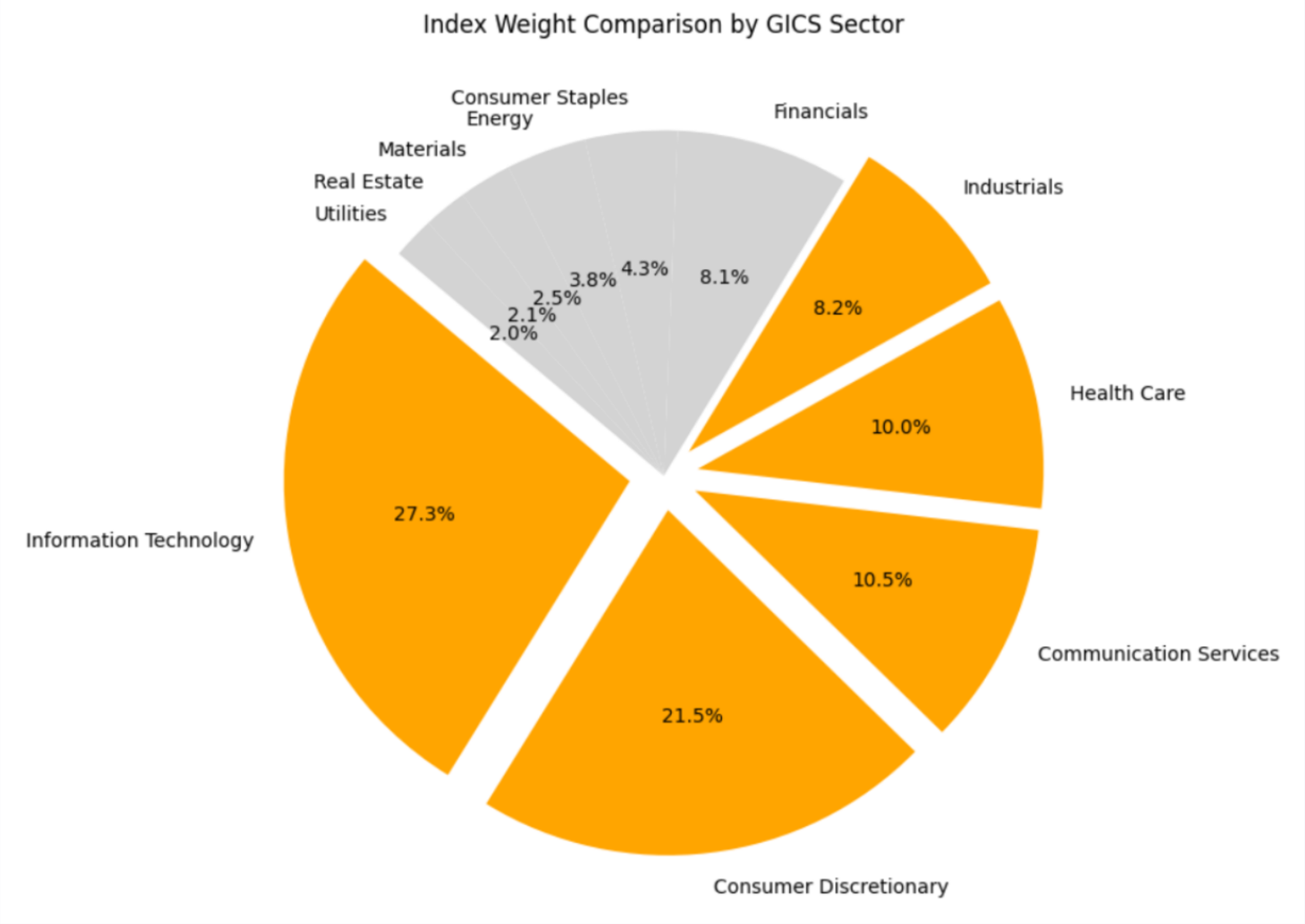
Selecting Securities And Sectors Of Interests

Key Sectors

- Information Technology
- Consumer Discretionary
- Communication Services
- Health Care
- Industrials

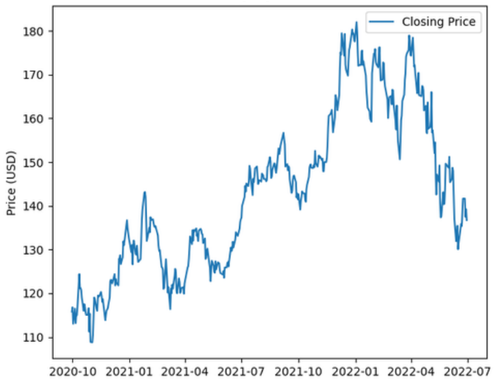
Top 5 Companies

- Apple
- Tesla
- Meta
- Moderna
- Boeing



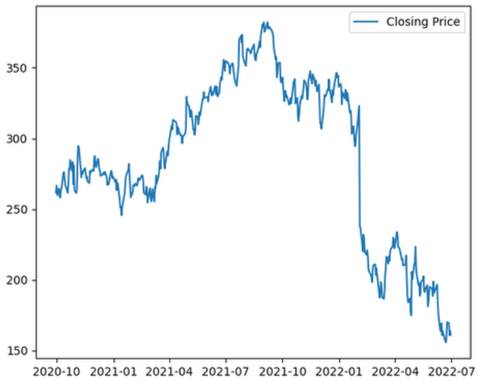
Closing Price Analysis

Apple



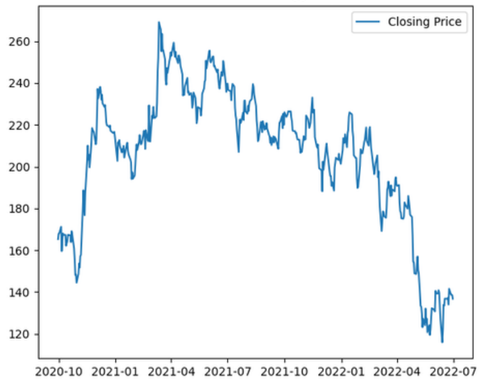
Tesla

Meta



Moderna

Boeing



Finding possible trends and behaviours

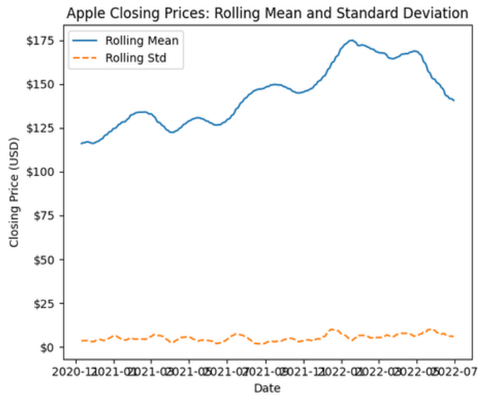
Tesla shows a sharp rise and fall and high volatility, typical for tech stocks

Moderna shows a significant peak, possibly due to COVID 19

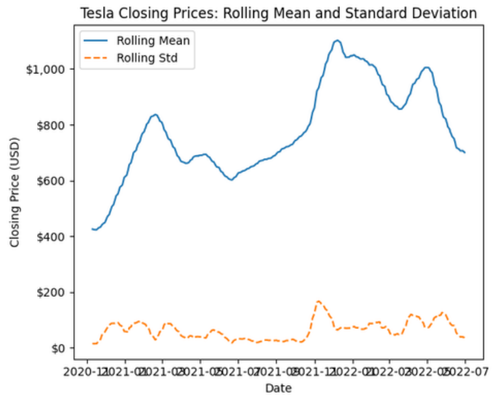
Rolling Mean and Standard Deviation

Average of data points over a 30 Day Window: Analysis of Short Term Fluctuations

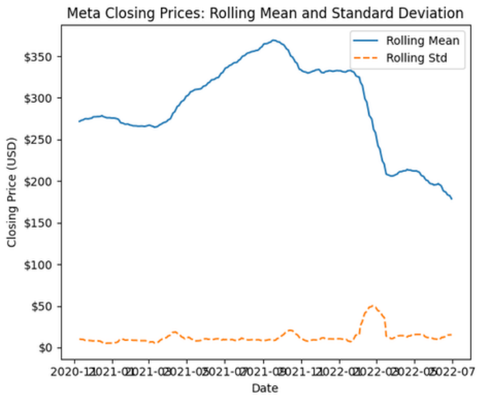
Apple



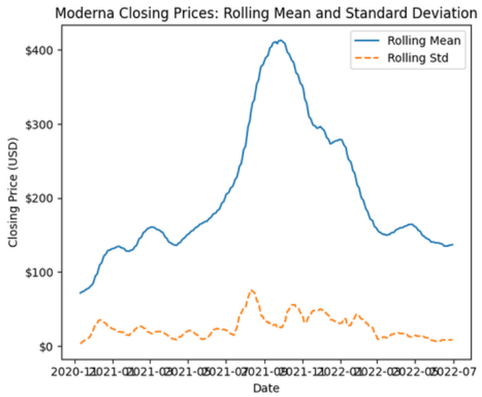
Tesla



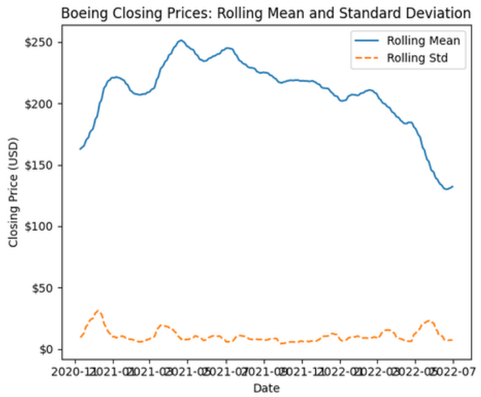
Meta



Moderna



Boeing



Rolling Mean

Consistent for Apple, Meta and Boeing

Visible Changes for Tesla and Moderna

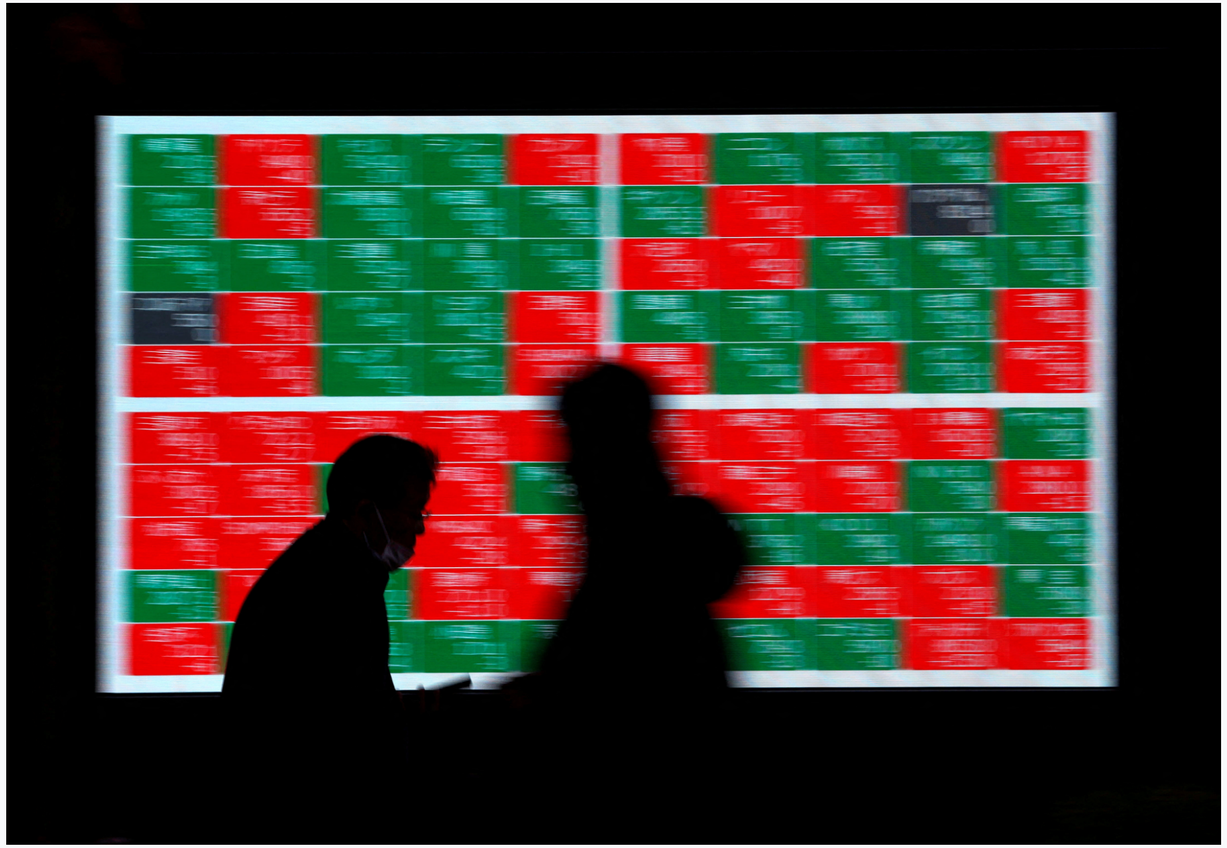
Rolling Standard Deviation

Relatively low and stable Standard Deviation for all stocks

Indication that **volatility** does not vary widely over time

Feature Transformation

ARIMA and LSTM Models perform better with consistent variance and minimal skewness



Engle's ARCH Test

Autoregressive Conditional Heteroskedasticity Test

: To test if variance of time series residuals depend on past values (across the entire dataset)

H0: Variance of error term **not dependent** on past squared error terms

H1: Variance of error term **dependent** on past squared error terms

We obtained p-value = 0 < 0.05, which indicates to us that a transformation of the variables in our dataset are required.

Hence,

To stabilise the variances, we log-transformed our data

Standardisation

To reduce skewness, we also standardised our data

$$x_{scaled} == \frac{x - mean}{sd}$$

Feature Selection

Principal Component Analysis, ANOVA, Random Forest

Components

- Volume
- News – All News Volume
- News – Volume
- News – Positive Sentiment
- News – Negative Sentiment
- News – New Products
- News – Layoffs
- News – Analyst Comments
- 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

Data Inspection

	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

Disproportionate Number of “0”

- News–Positive Sentiments
- News–Negative Sentiments
- News–New Products
- News–Layoffs

Data Inspection

	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

Disproportionate Number of “0”

- News–Dividends
- News–Corporate Earnings
- News–Merger & Acquisitions
- News–Store Openings
- News–Product Recalls
- News–Personnel Changes
- News–Stock Rumors

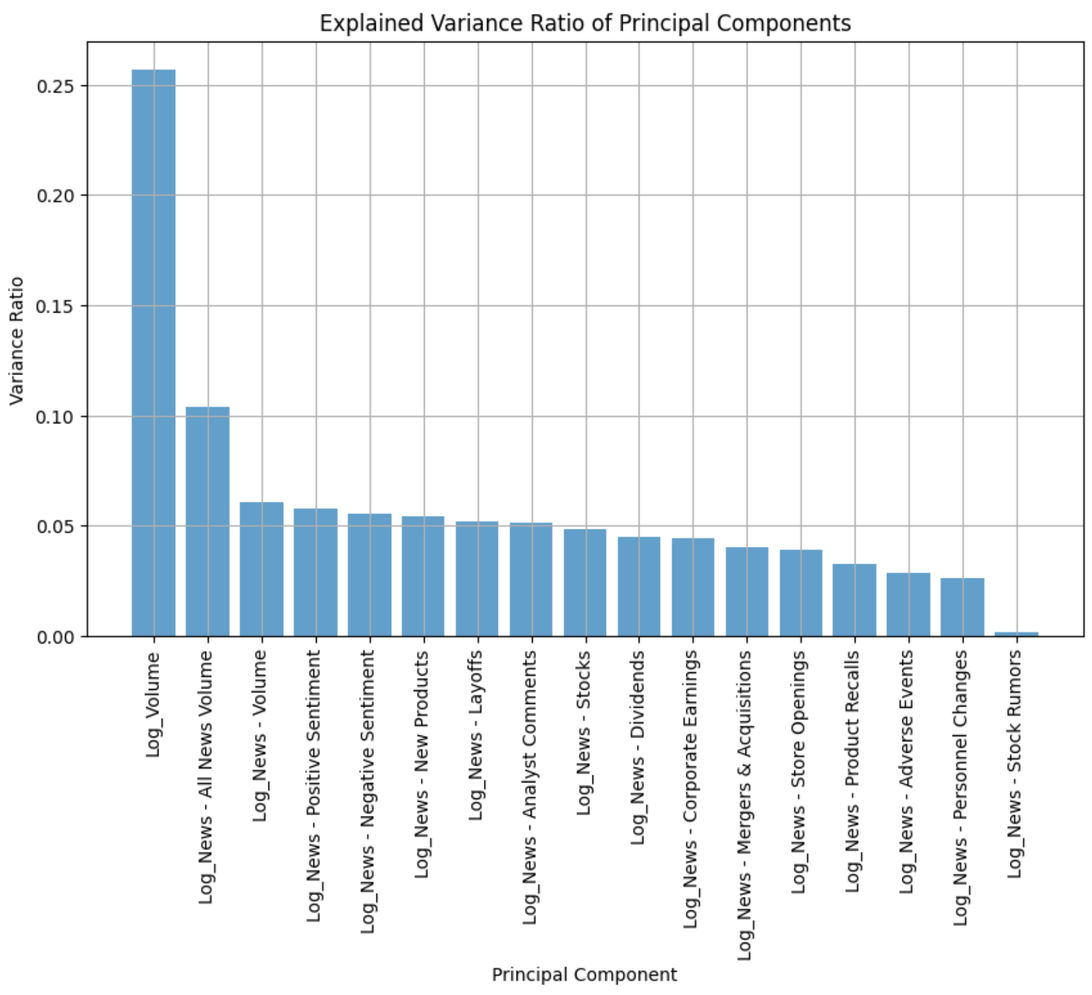
Principal Component Analysis

: uses an orthogonal transformation to convert correlated variables into linearly uncorrelated variables

Steps:

1. Standardization
2. Covariance Matrix
3. Eigenvalue Decomposition.
4. Select Principal Components
5. Dimensionality Reduction

Principal Component Analysis




```
Log_Volume: Variance: 0.2569
Log_News - All News Volume: Variance: 0.1040
Log_News - Volume: Variance: 0.0605
Log_News - Positive Sentiment: Variance: 0.0579
Log_News - Negative Sentiment: Variance: 0.0556
Log_News - New Products: Variance: 0.0543
Log_News - Layoffs: Variance: 0.0521
Log_News - Analyst Comments: Variance: 0.0515
Log_News - Stocks: Variance: 0.0485
Log_News - Dividends: Variance: 0.0452
Log_News - Corporate Earnings: Variance: 0.0445
Log_News - Mergers & Acquisitions: Variance: 0.0403
Log_News - Store Openings: Variance: 0.0390
Log_News - Product Recalls: Variance: 0.0324
Log_News - Adverse Events: Variance: 0.0288
Log_News - Personnel Changes: Variance: 0.0265
Log_News - Stock Rumors: Variance: 0.0020
```

ANOVA Analysis of Variance

: Uses the F test to check if the means of two or more groups are significantly different from each other

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

Equal Variance = Little impact on predicting target

Result: All features have significant differences

ANOVA Analysis of Variance

	sum_sq	df	F	\	PR(>F)
Q("Log_Volume")	44296.734504	1.0	80585.500154		0.000000e+00
Q("Log_News - All News Volume")	9.111676	1.0	16.576142		4.676313e-05
Q("Log_News - Volume")	2.360053	1.0	4.293455		3.826134e-02
Q("Log_News - Positive Sentiment")	73.768854	1.0	134.201766		5.113572e-31
Q("Log_News - Negative Sentiment")	260.785115	1.0	474.425467		5.403570e-105
Q("Log_News - New Products")	1197.955816	1.0	2179.345039		0.000000e+00
Q("Log_News - Layoffs")	35.268342	1.0	64.160868		1.156111e-15
Q("Log_News - Analyst Comments")	25.334250	1.0	46.088572		1.135124e-11
Q("Log_News - Corporate Earnings")	295.517494	1.0	537.611300		1.093364e-118
Q("Log_News - Mergers & Acquisitions")	1723.720617	1.0	3135.826818		0.000000e+00
Q("Log_News - Store Openings")	577.661085	1.0	1050.892532		1.316081e-229
Q("Log_News - Product Recalls")	20.672991	1.0	37.608715		8.670727e-10
Q("Log_News - Adverse Events")	23.209529	1.0	42.223236		8.172201e-11
Q("Log_News - Personnel Changes")	149.506639	1.0	271.985451		4.840317e-61
Q("Log_News - Stock Rumors")	150.341086	1.0	273.503492		2.263258e-61

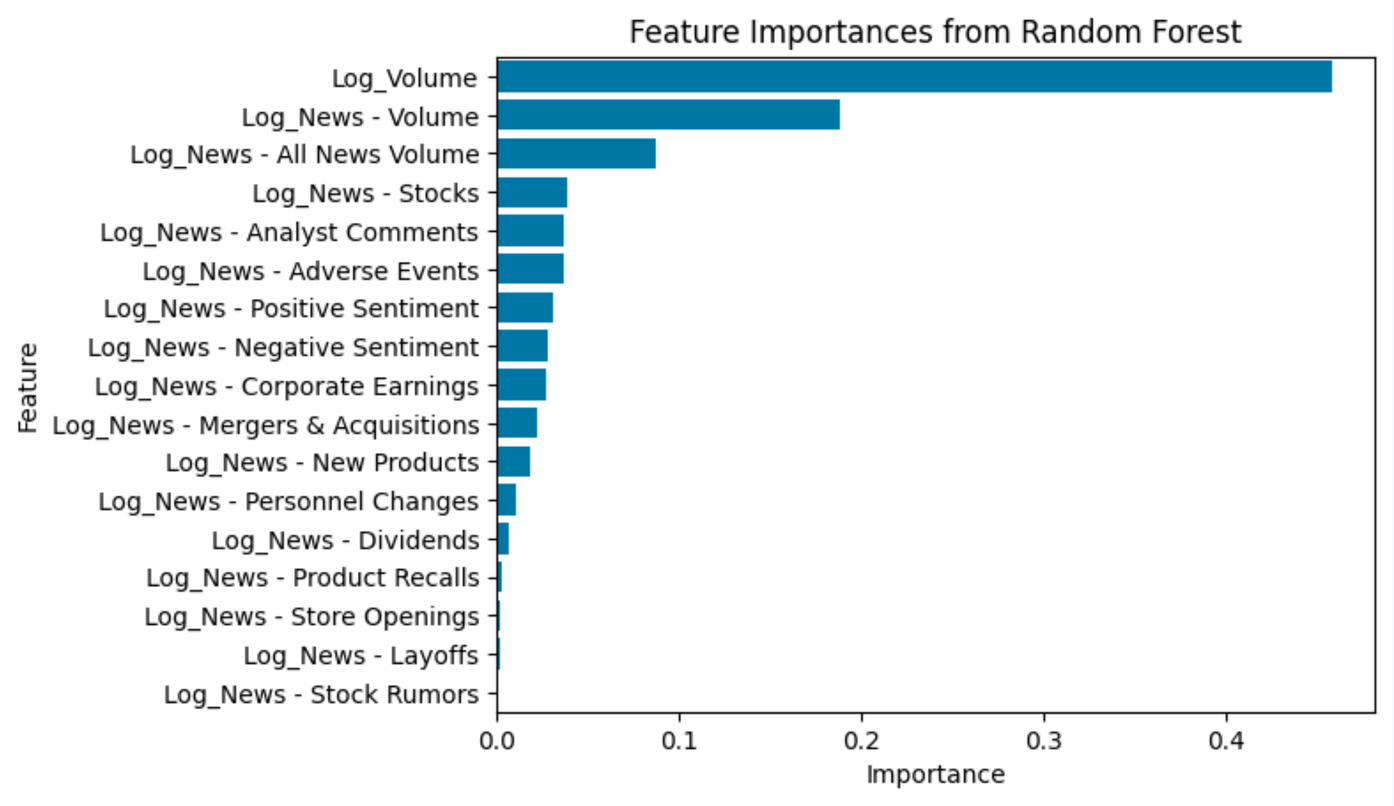
Random Forest

: Ensemble learnng method used for classification and regression tasks

Operates by constructing multiple decision trees

Outputs mean prediction of the individual trees

Trained on a random subset of data and features



Random Forest

Dataset

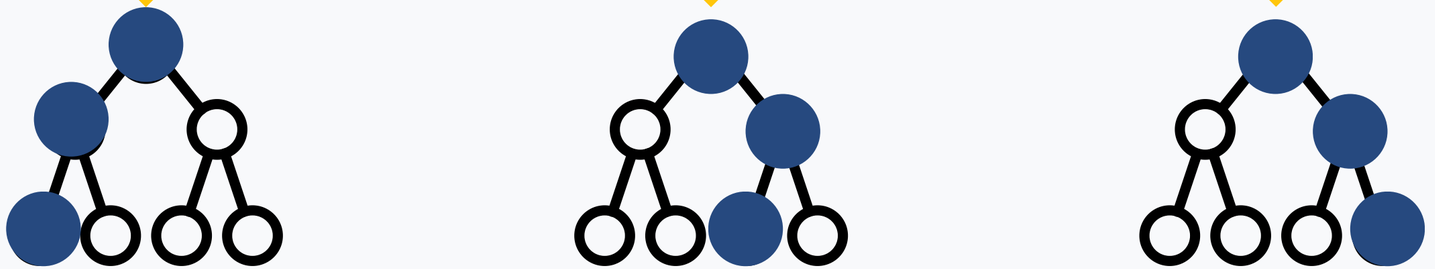
Feature Selection

Randomised Features

Stock Rumors, Volume

All News Volume, Layoffs

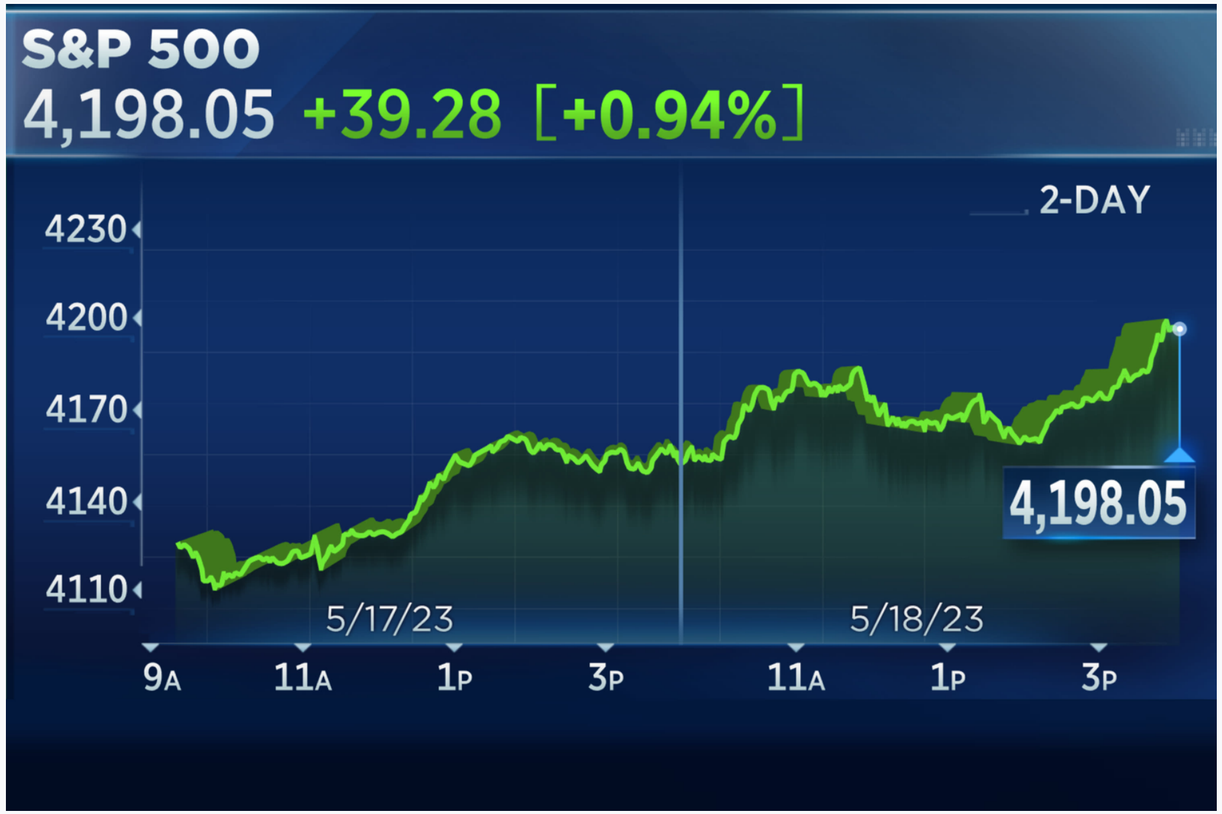
Analyst Comments
News Products



Voting

Features Selected

1. Volume
2. All News Volume
3. News Volume
4. Analyst Comments
5. Stocks
6. Adverse Events.



Model Building: Classical Methods

ARIMA
(AutoRegressive Integrated Moving Average)

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

AutoRegressive (AR) (p)

Regression model of the relationship between observation y and several lagged observations

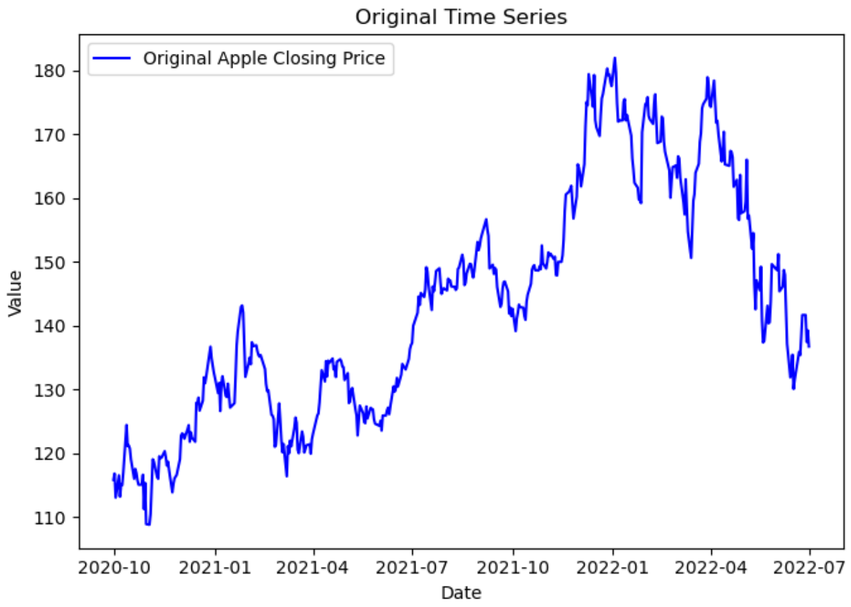
Integrated (I) (d)

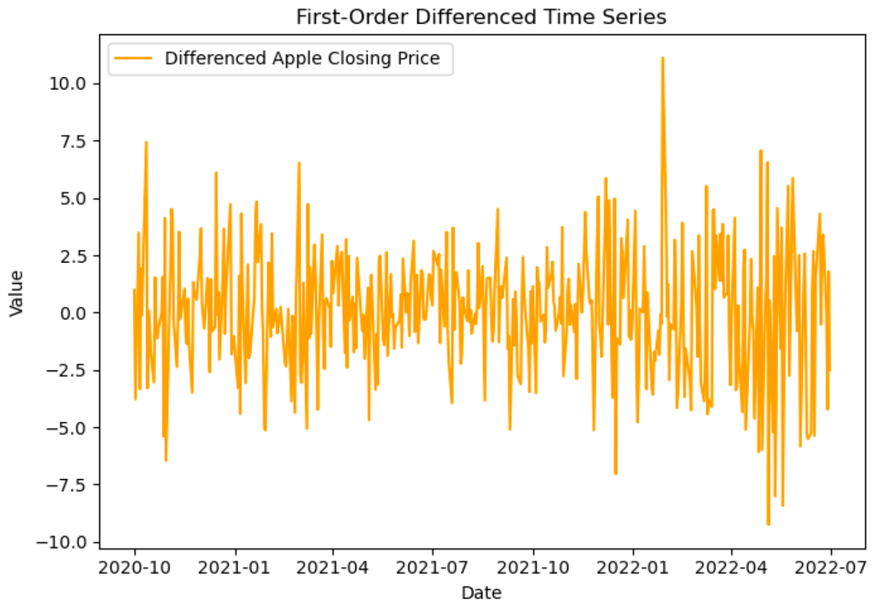
Represents the differencing of observations to make the time series stationary

Moving Average (MA)(q)

Regression model of observation y and the residual error of lagged observations

Stationarity





Consistent Statistical Properties

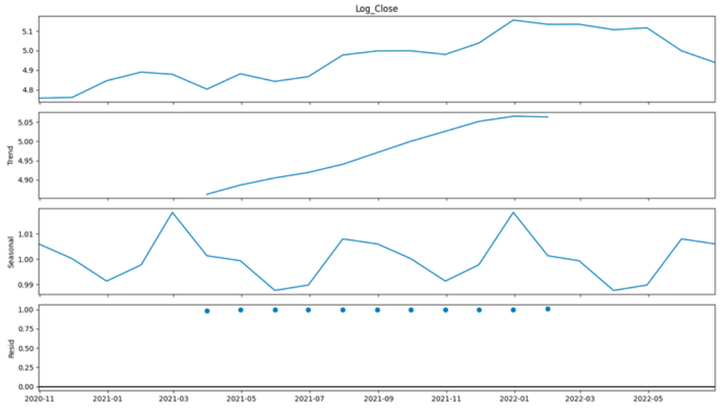
	Apple	Tesla	Meta	Moderna	Boeing
P-value	0.3380	0.2099	0.8807	0.4288	0.6622

ADF TEST(Augmented Dickey-Fuller Test)

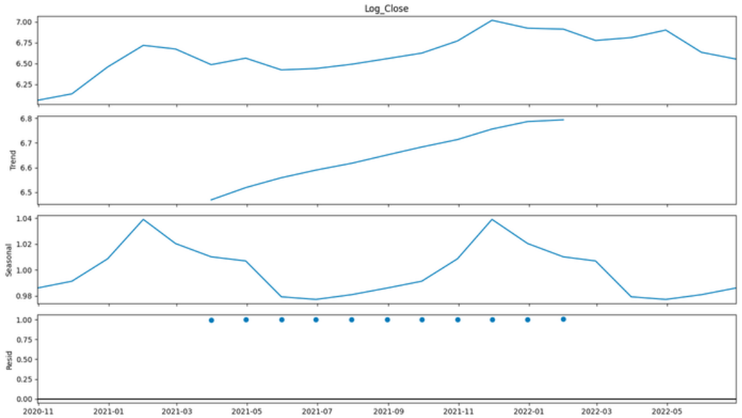
- Null Hypothesis (HO): Series is non-stationary, or series has a unit root.
- Alternate Hypothesis (HA): Series is stationary, or series has no unit root.

Stationarity

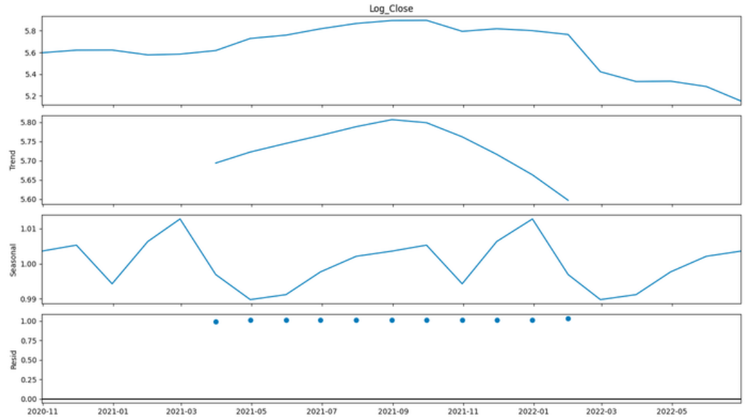
APPLE



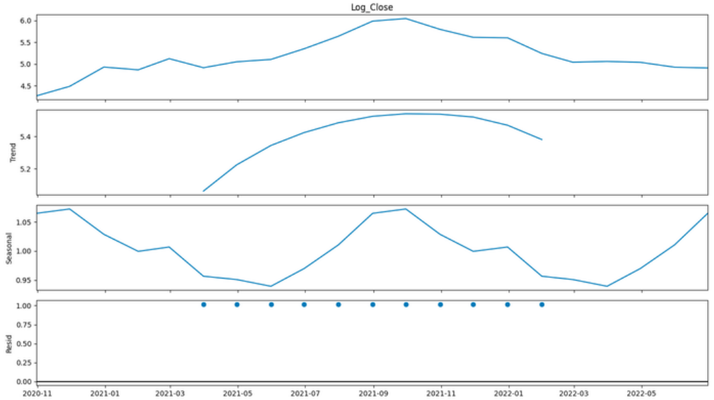
TESLA



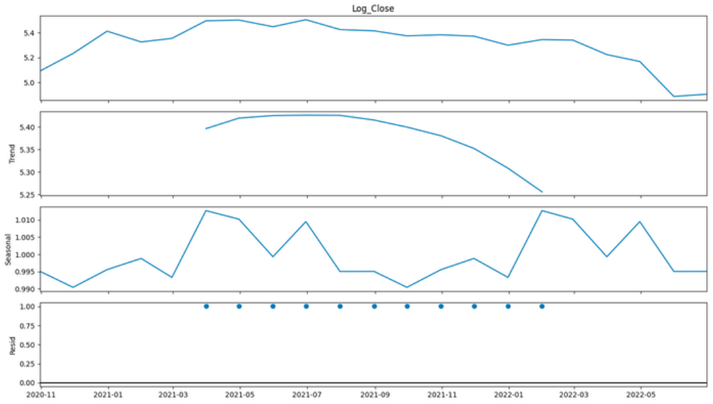
META



MODERNA



BOEING



Seasonal Decomposition

- All 5 stocks displayed seasonality
- Stock data is non stationary

Auto - Arima

Machine Learning Model with “Loss Function”

AIC minimising function is used to choose the model’s best parameters

```
Performing stepwise search to minimize aic
 ARIMA(0,0,0)(0,1,0)[12] intercept   : AIC=442.804, Time=0.16 sec
 ARIMA(1,0,0)(1,1,0)[12] intercept   : AIC=-283.539, Time=2.81 sec
 ARIMA(0,0,1)(0,1,1)[12] intercept   : AIC=96.316, Time=2.37 sec
 ARIMA(0,0,0)(0,1,0)[12]             : AIC=453.443, Time=0.05 sec
 ARIMA(1,0,0)(0,1,0)[12] intercept   : AIC=-180.902, Time=0.37 sec
 ARIMA(1,0,0)(2,1,0)[12] intercept   : AIC=-325.614, Time=5.40 sec
 ARIMA(1,0,0)(2,1,1)[12] intercept   : AIC=inf, Time=8.21 sec
 ARIMA(1,0,0)(1,1,1)[12] intercept   : AIC=inf, Time=2.63 sec
 ARIMA(0,0,0)(2,1,0)[12] intercept   : AIC=426.512, Time=1.63 sec
 ARIMA(2,0,0)(2,1,0)[12] intercept   : AIC=-323.834, Time=5.41 sec
 ARIMA(1,0,1)(2,1,0)[12] intercept   : AIC=-323.857, Time=7.60 sec
 ARIMA(0,0,1)(2,1,0)[12] intercept   : AIC=97.163, Time=5.32 sec
 ARIMA(2,0,1)(2,1,0)[12] intercept   : AIC=-322.501, Time=13.60 sec
```

Akaike Information Criterion (AIC)

Used to compare different ARIMA models

Minimised AIC Values

The AIC is defined as: AIC = 2k - 2ln(L)

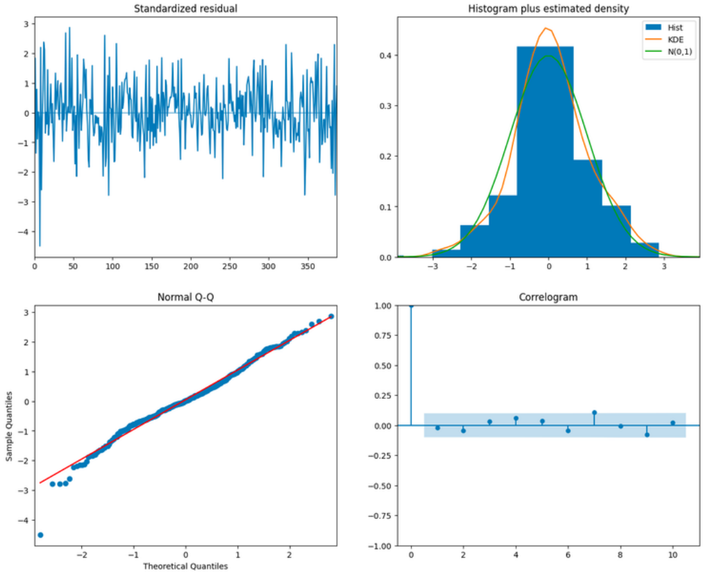
SARIMAX (Seasonal ARIMA with Exogenous Variables)

$$d_t = c + \sum_{n=1}^{p} \alpha_n d_{t-n} + \sum_{n=1}^{q} \theta_n \epsilon_{t-n} + \sum_{n=1}^{r} \beta_n x_{n_t} + \sum_{n=1}^{P} \phi_n d_{t-sn} + \sum_{n=1}^{Q} \eta_n \epsilon_{t-sn} + \epsilon_t$$

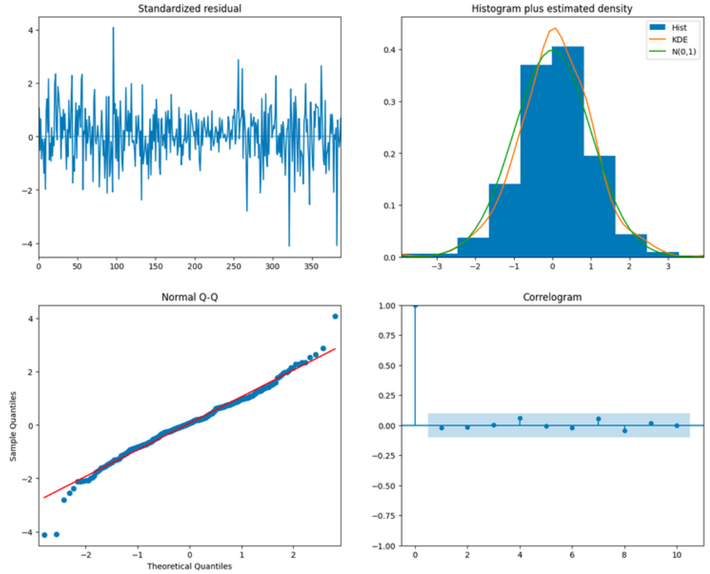
(P) seasonal autoregressive + (D) seasonal differences + (Q) seasonal moving average + (s) seasonal cycle

Model Dianogstics

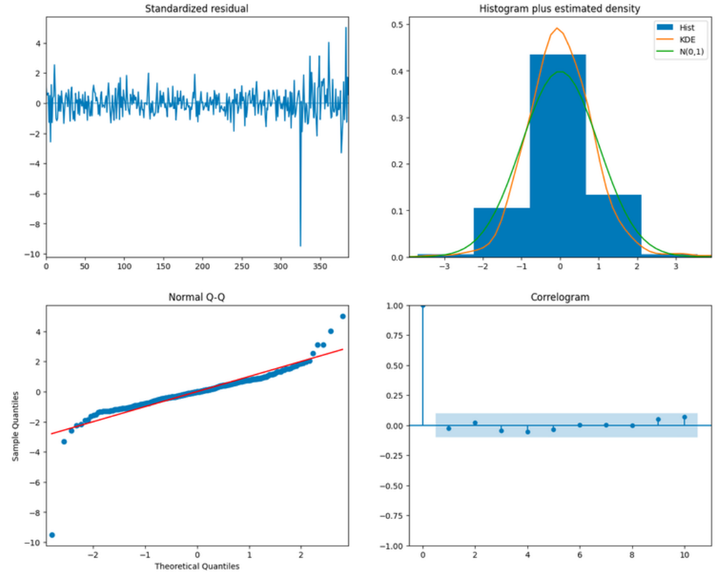
APPLE



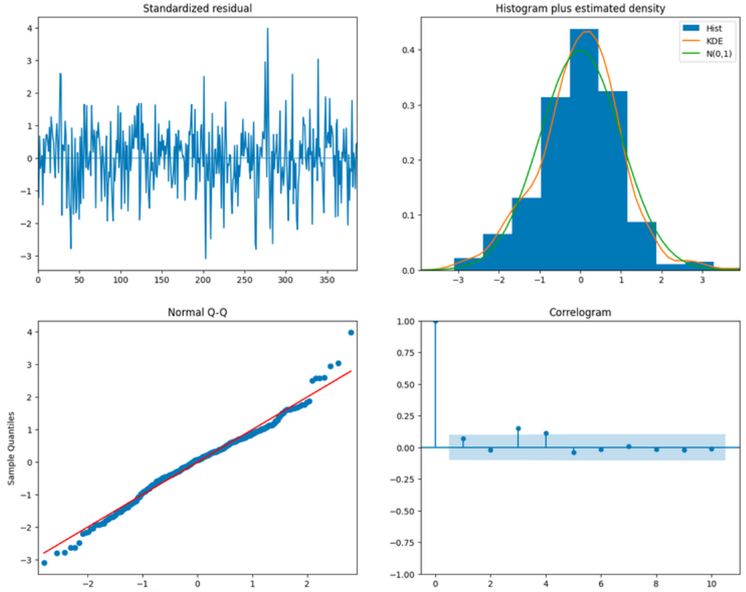
TESLA



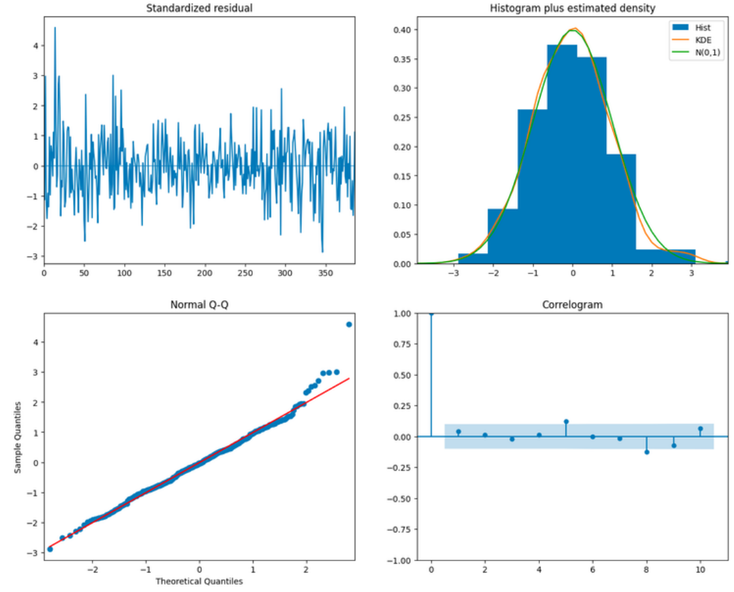
META



MODERNA



BOEING



Advanced Stock Predictions Using Time Series

Model training using Deep learning

Our Models

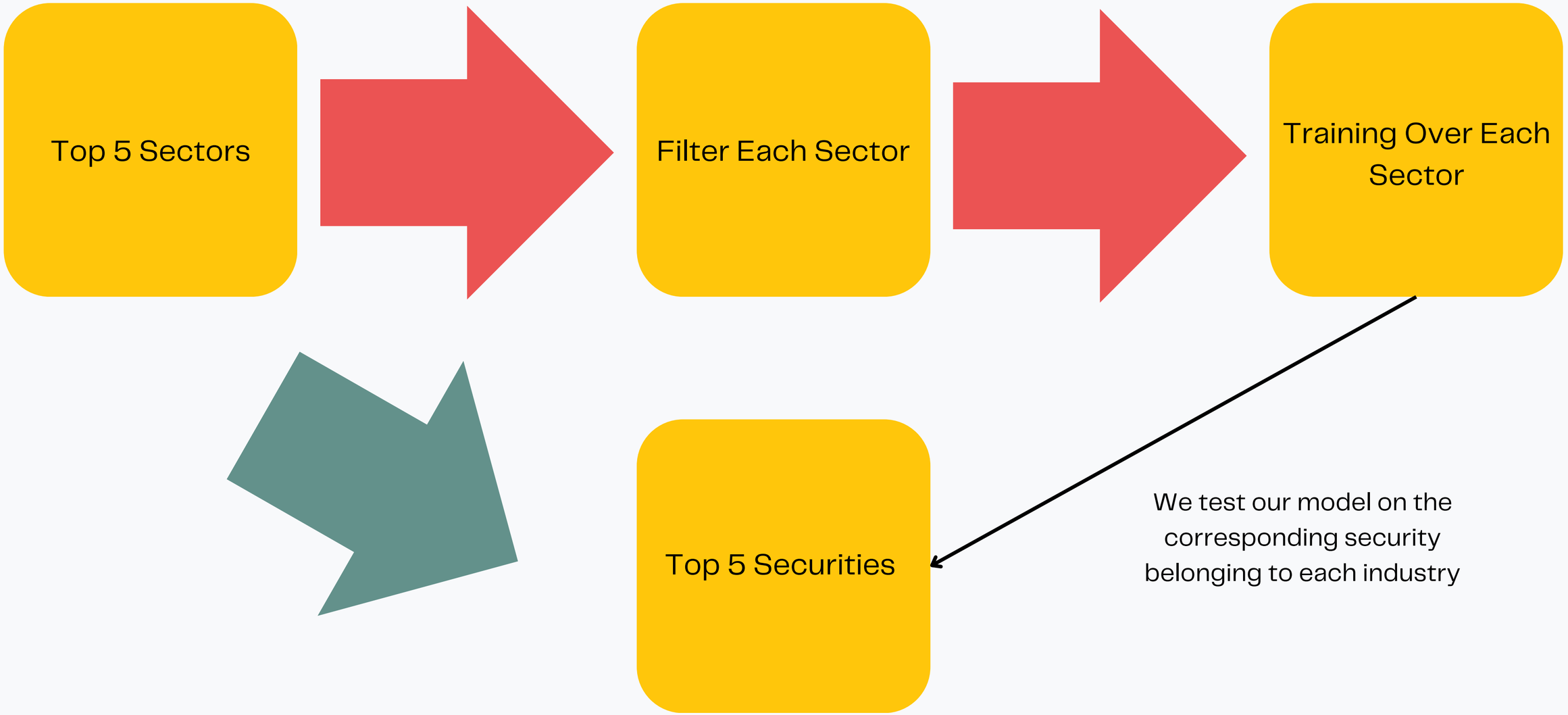
LSTM (Long Short-Term Memory)

- LSTMs are composed of three gates: the input gate, the forget gate, and the output gate, which regulate the flow of information into, out of, and within the memory cell.
- This architecture enables LSTMs to effectively learn and retain information over extended time periods, making them well-suited for tasks such as speech recognition, language modeling, and time series forecasting.

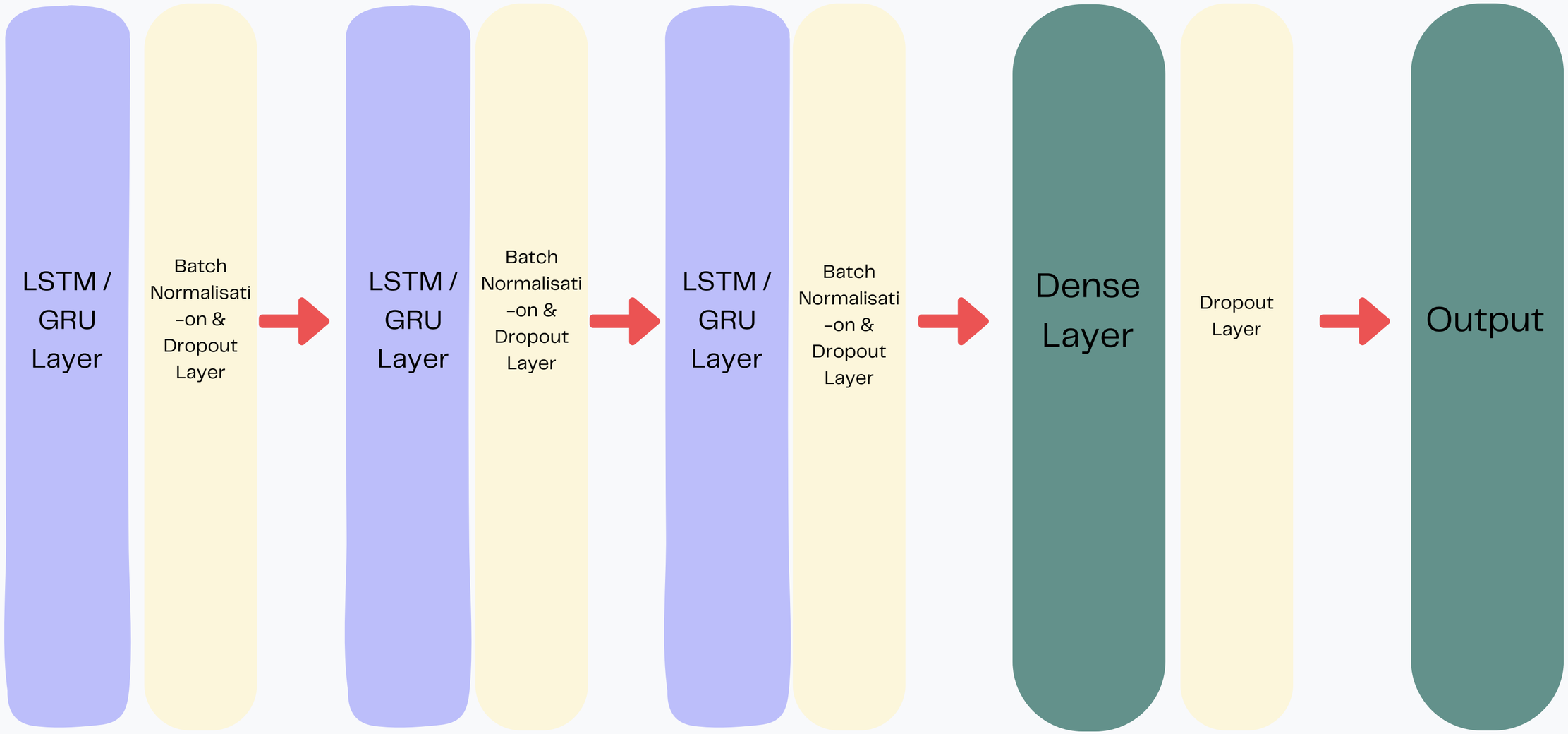
GRU (Gated Recurrent Unit)

- GRUs simplify the architecture of LSTMs by combining the input and forget gates into a single update gate and merging the cell state and hidden state.
- GRUs have fewer parameters compared to LSTMs, making them computationally more efficient and faster to train.

Model Training Overview



Model Architecture



Model Hyper-Parameters

Hyperparameters	Description	Input
Sequence Length	Number of time steps or data points used to create a single input instance for the model	60
Epochs	Number of times the entire dataset is passed forward and backward through the model	3
Learning Rate	Step size at which the model weights are updated during training	0.001
Batch Size	Number of training examples used in each iteration	100
Optimizer	Algorithm used to minimize the loss function by adjusting the model parameters during training	RMSProp
Loss Function	Measures how well the model performs on the training data by comparing the predicted output to the actual target values	Mean Squared Error

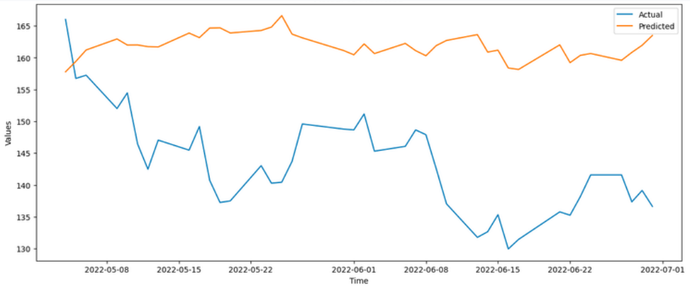
Results - Overall

Average Across 5 Sectors

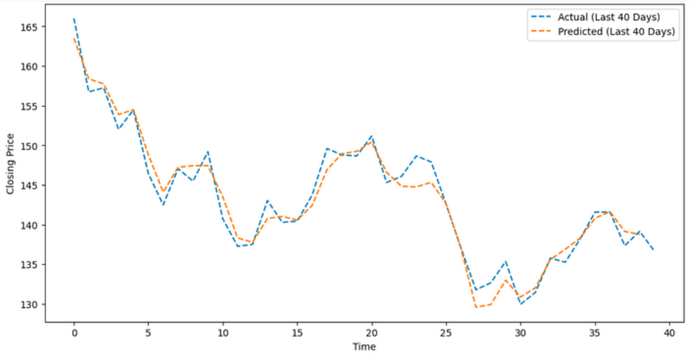
	Auto - ARIMA	**LSTM**	**GRU**
RMSE	52.218	13.683	16.820
MAPE	0.146	0.0724	0.09691
Training Time (in seconds)	66.037	146.2	131.2

Results - Time Series Visualisation (APPLE)

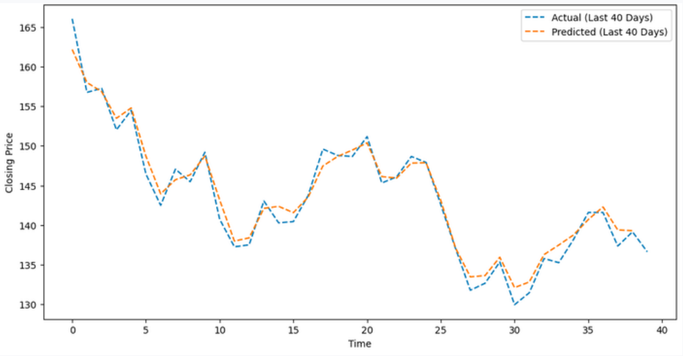
Auto – Arima



LSTM

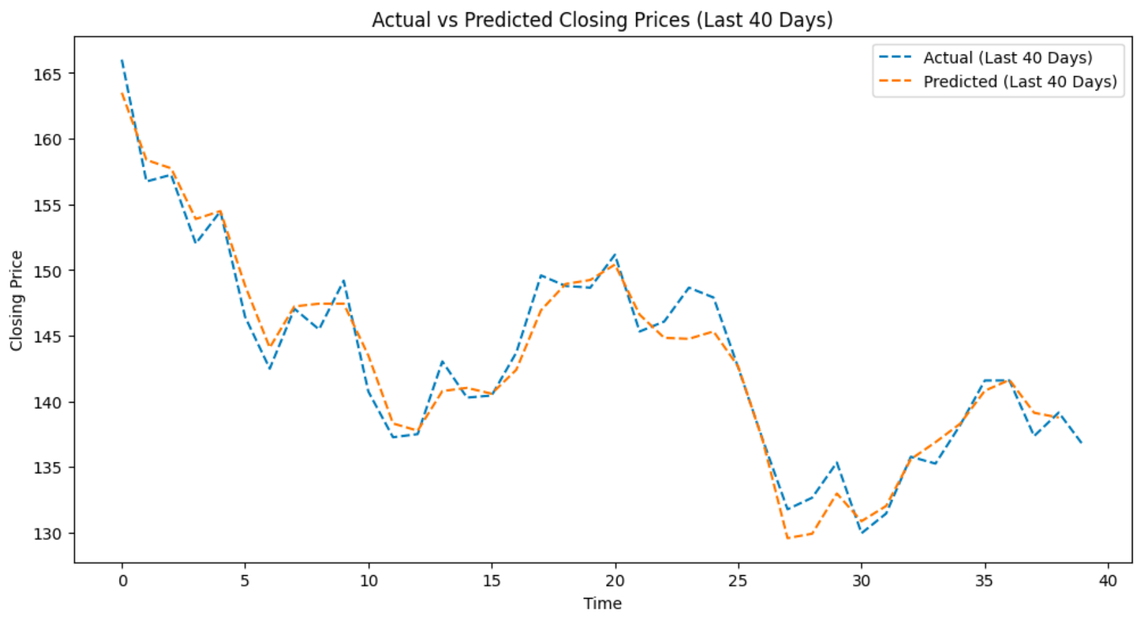


GRU



Limitations

APPLE (LSTM)



MODERNA (LSTM)

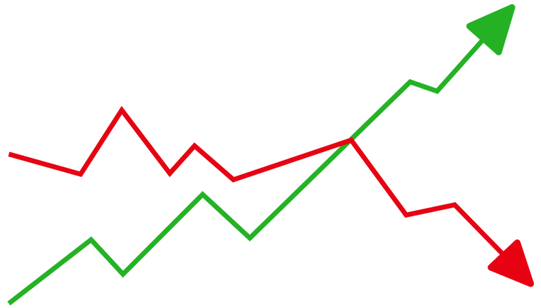
Limitations

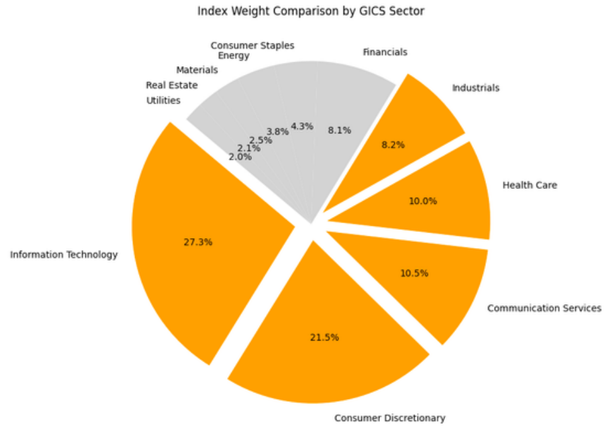
Quality of Data

Unpredictability

Proportion of Data

Example: Many Numerical Features with disproportionate number of “0”s





Thank you!

Project Distribution

JOASH: 27%

MAVERICK: 27%

WYNTHIA: 27%

BRANDON: 19%