VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A6a- Time Series Analysis

RAKSHITH.H

V01107367

Date of Submission: 22-07-2024

CONTENTS

Introduction

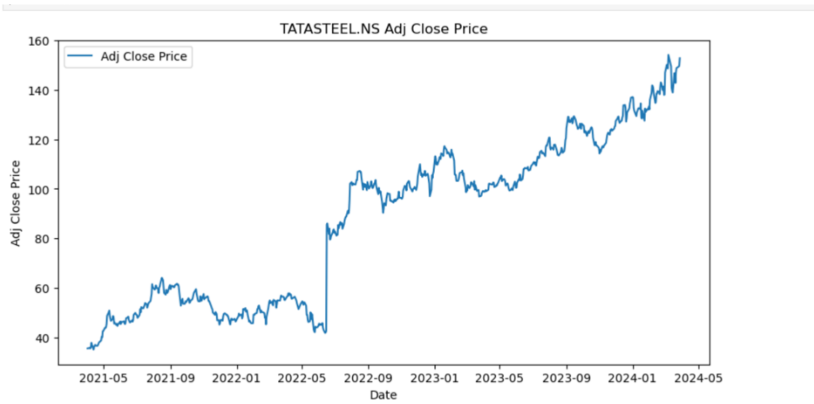
This study focuses on analysing and forecasting Tata steel stock price trends from January 2020 to July 2024. The dataset, "TISC Historical Data," includes key attributes such as Date, Price, Open, High, Low, Volume, and Change %. This comprehensive time series data allows for a detailed examination of Tata steel stock performance, encompassing daily fluctuations and broader market trends. Through this analysis, we aim to apply various time series forecasting techniques to predict future stock prices and gain insights into the underlying patterns and dynamics driving Tata steel's stock movements.

The analysis involves a thorough data cleaning process, identifying and handling outliers, and interpolating missing values to ensure data integrity. Subsequently, we decompose the time series into its components using both additive and multiplicative models to uncover seasonal and trend patterns. Additionally, we implement univariate forecasting using conventional statistical models such as Holt-Winters, ARIMA, and SARIMA to predict future stock prices. Furthermore, we explore multivariate forecasting using machine learning models, including Neural Networks (LSTM), Decision Trees, and Random Forests. This comprehensive approach provides a robust framework for understanding Tata Steel stock price behaviour and offers valuable predictive insights for investors and stakeholders.

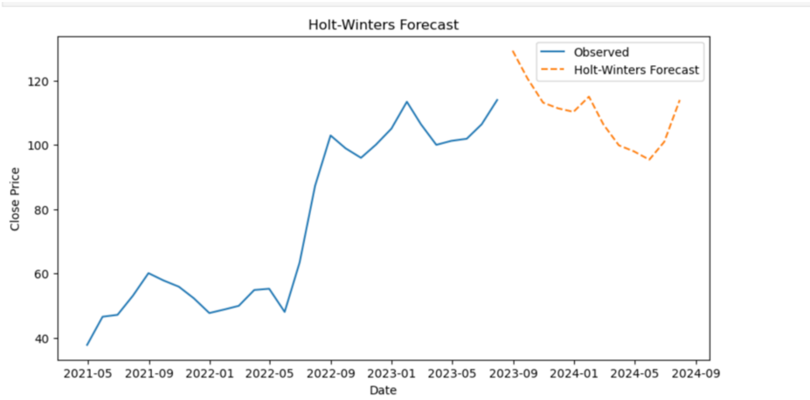
Objectives

The primary objectives of this task are:

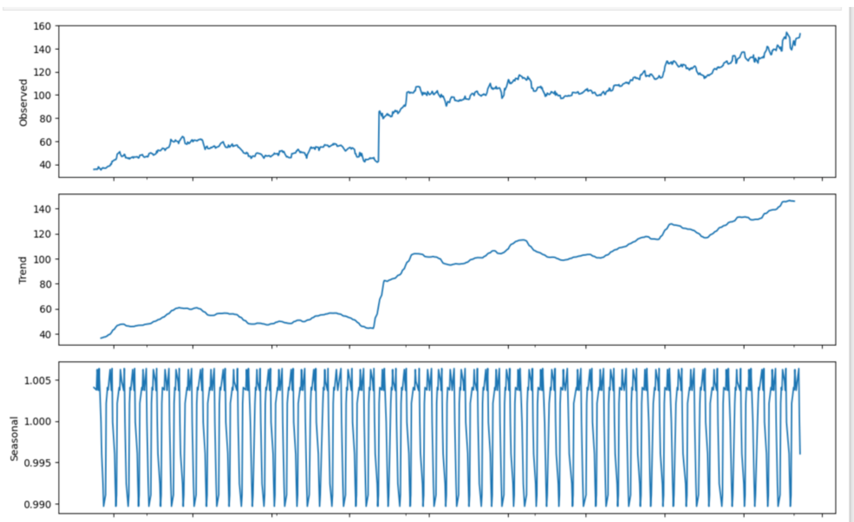
1. To clean and pre-process the Tata Steel stock data from January 2020 to July 2024, addressing missing values and outliers.
2. To visualize the Tata Steel stock price over time using line plots.
3. To decompose the time series data into its components using both additive and multiplicative models.
4. To implement univariate forecasting models, including Holt-Winters, ARIMA, and SARIMA, and evaluate their performance.
5. To perform multivariate forecasting using machine learning models such as Neural Networks (LSTM), Decision Trees, and Random Forests.
6. To compare the results of different forecasting models and provide insights into the predictive accuracy for Tata Steel stock prices.



The graph depicts the adjusted close price of Tata Steel (TATASTEEL.NS) from May 2021 to May 2024. The price shows a general upward trend with occasional fluctuations. A significant jump occurs around mid-2022, followed by continued growth with intermittent dips. The price peaks at its highest point by early 2024.



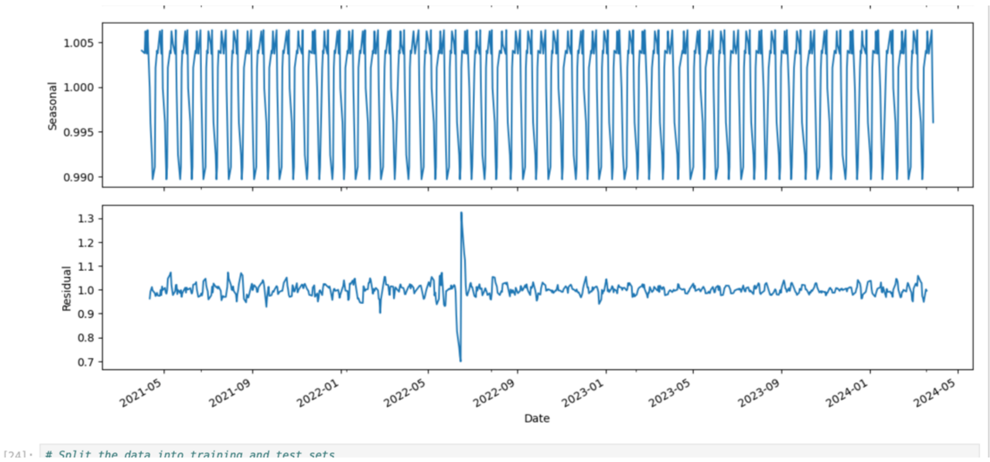
The graph shows the historical closing prices of Tata Steel (TATASTEEL.NS) from May 2021 to September 2023 and a forecast using the Holt-Winters method from October 2023 to September 2024. The observed data shows a significant increase around mid-2022, followed by fluctuations. The Holt-Winters forecast predicts a decline in prices until early 2024, with a subsequent rise starting around mid-2024. The forecasted prices show more variability compared to the observed data.



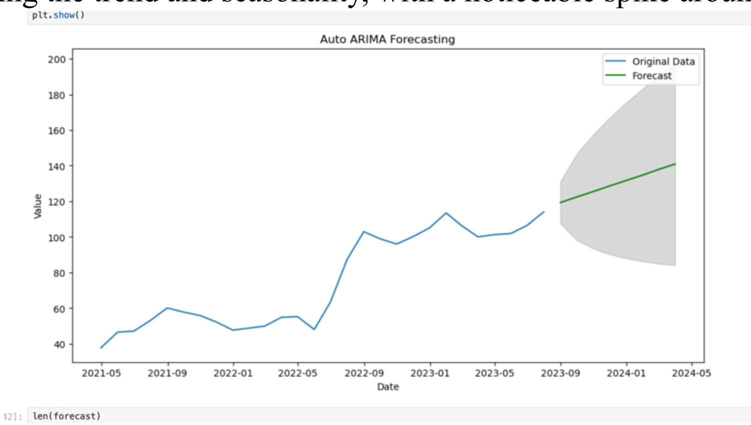
The graph displays the observed data, trend, and seasonal components of Tata Steel's adjusted closing prices.

1. The top panel shows the observed prices with an upward trend and periodic fluctuations.
2. The middle panel illustrates the overall trend, highlighting a significant increase around mid-2022 with continued growth.
3. The bottom panel shows the seasonal component, indicating regular, cyclical patterns in the data.

This decomposition helps in understanding the underlying trends and seasonal effects in the stock price.



The top graph shows the seasonal component of the time series, indicating a regular repeating pattern over time. The bottom graph depicts the residuals, showing the remaining variations after removing the trend and seasonality, with a noticeable spike around mid-2022.



The plot shows the original time series data along with the forecast generated by an Auto ARIMA model. The forecast extends into 2024, with a confidence interval shaded in gray, indicating the expected range of values.

```
# Print the model summary
print(arima_model.summary())
```

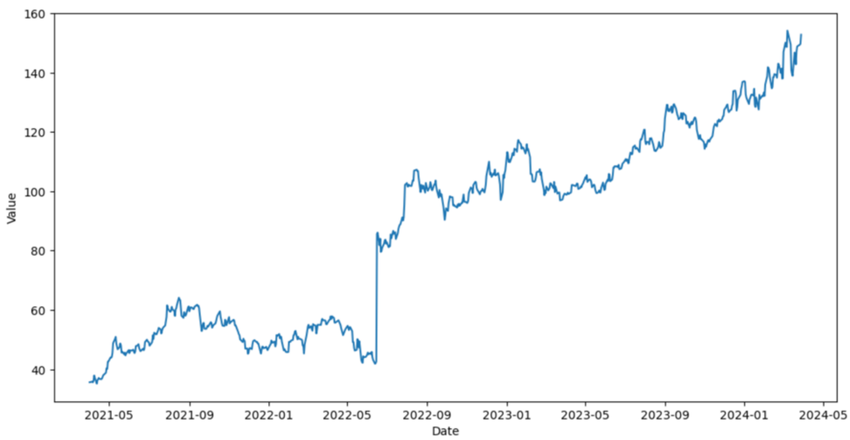
```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                   28
Model:               SARIMAX(0, 1, 1)   Log Likelihood                 -87.140
Date:                Mon, 22 Jul 2024   AIC                            180.281
Time:                        21:12:52   BIC                            184.168
Sample:                    04-30-2021   HQIC                           181.437
                         - 07-31-2023
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      3.0927      2.065      1.498      0.134      -0.954       7.140
ma.L1          0.7919      0.140      5.641      0.000       0.517       1.067
sigma2        35.8896      8.240      4.355      0.000      19.739      52.040
===================================================================================
Ljung-Box (L1) (Q):                   0.34   Jarque-Bera (JB):                 1.54
Prob(Q):                              0.56   Prob(JB):                         0.46
Heteroskedasticity (H):               1.27   Skew:                             0.31
Prob(H) (two-sided):                  0.73   Kurtosis:                         3.99
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

```
print(arima_model.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  740
Model:               SARIMAX(0, 1, 0)   Log Likelihood               -1671.522
Date:                Mon, 22 Jul 2024   AIC                           3347.044
Time:                        21:13:21   BIC                           3356.254
Sample:                             0   HQIC                          3350.595
                                - 740
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.1585      0.116      1.372      0.170      -0.068       0.385
sigma2         5.3971      0.043    125.660      0.000       5.313       5.481
===================================================================================
Ljung-Box (L1) (Q):                   0.19   Jarque-Bera (JB):            741393.19
Prob(Q):                              0.66   Prob(JB):                         0.00
Heteroskedasticity (H):               2.21   Skew:                             8.41
Prob(H) (two-sided):                  0.00   Kurtosis:                       157.25
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

```
[48]: # Generate in-sample predictions
fitted_values = arima_model.predict_in_sample()
```

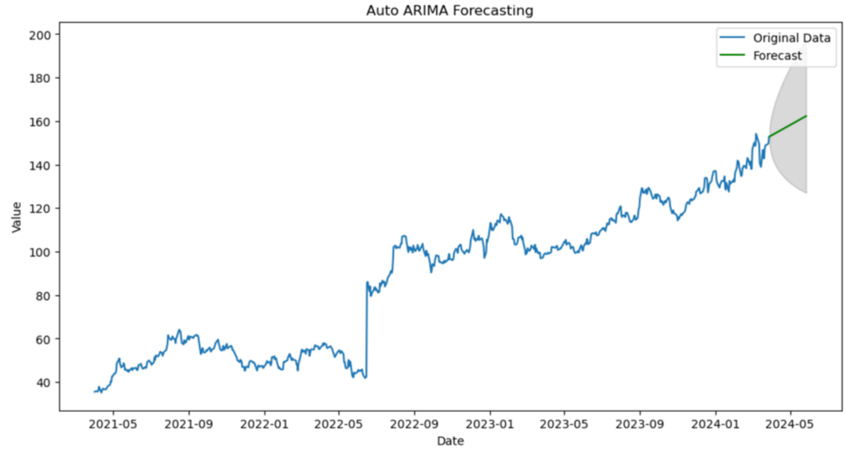
```
5]: # Plot the original data, fitted values, and forecast
plt.figure(figsize=(12, 6))
plt.plot(daily_data['Adj Close'])
plt.xlabel('Date')
plt.ylabel('Value')
plt.show()
```



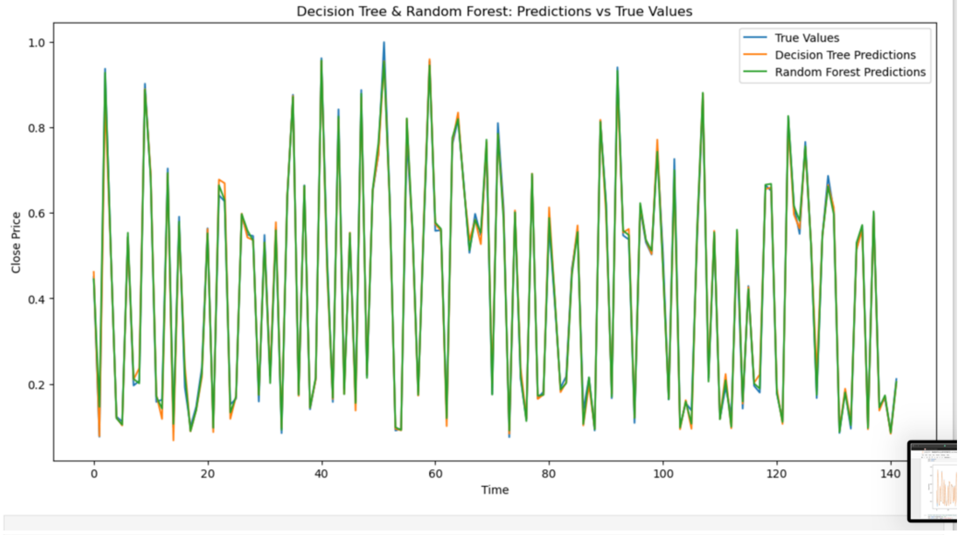

```
6]: # Fit auto_arima model
arima_model = auto_arima(daily_data['Adj Close'],
```

The plot displays the adjusted close values of a time series from 2021 to 2024. There is a noticeable jump in mid-2022, followed by a steady upward trend.

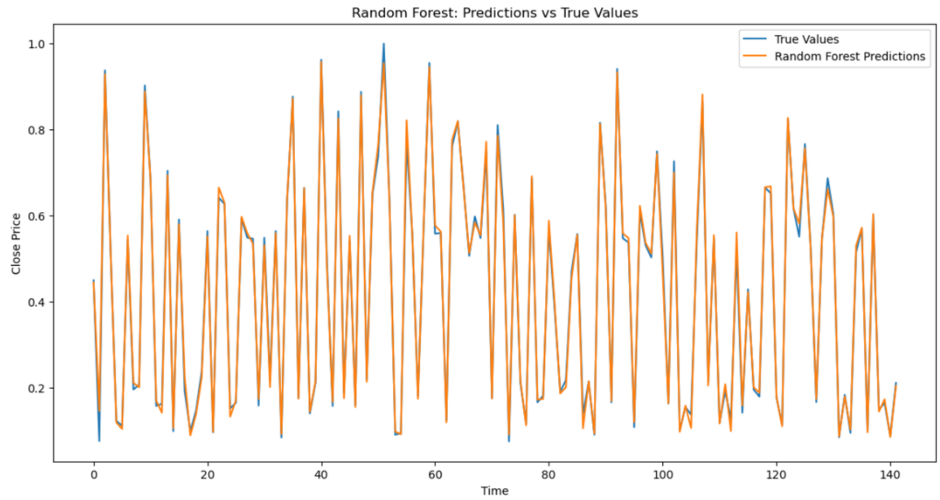
```
plt.ylabel('Value')
plt.title('Auto ARIMA Forecasting')
plt.show()
```

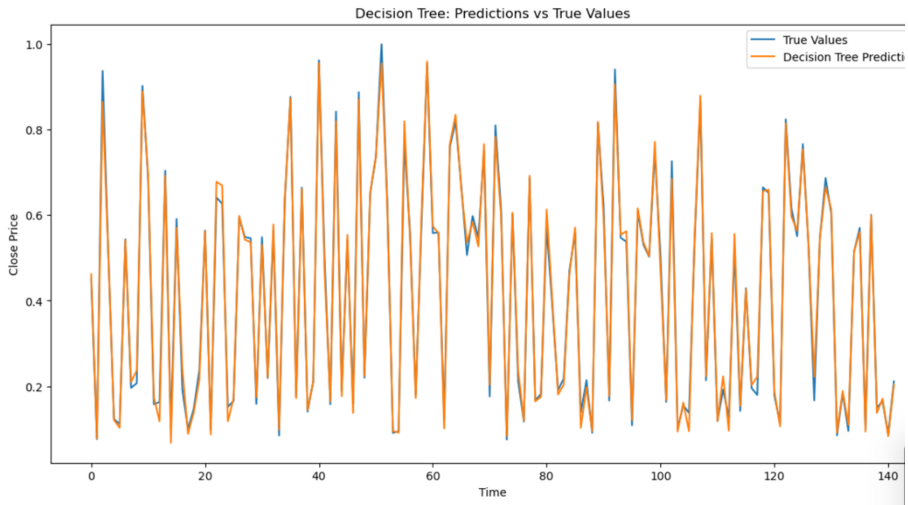


2. Multivariate Forecasting - Machine Learning Models

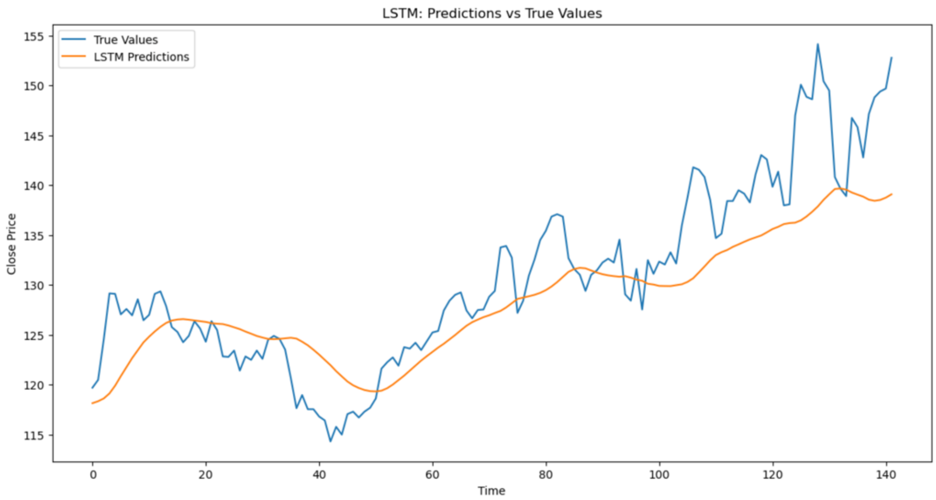


[ ]:






```
[81]: print("\nRandom Forest Predictions vs True Values:")
      for i in range(10):
          print(f"Prediction: {y_pred_rf[i]}, True Value: {y_test[i]}")
```



Tree Based Models

```
[66]: model.summary()
```

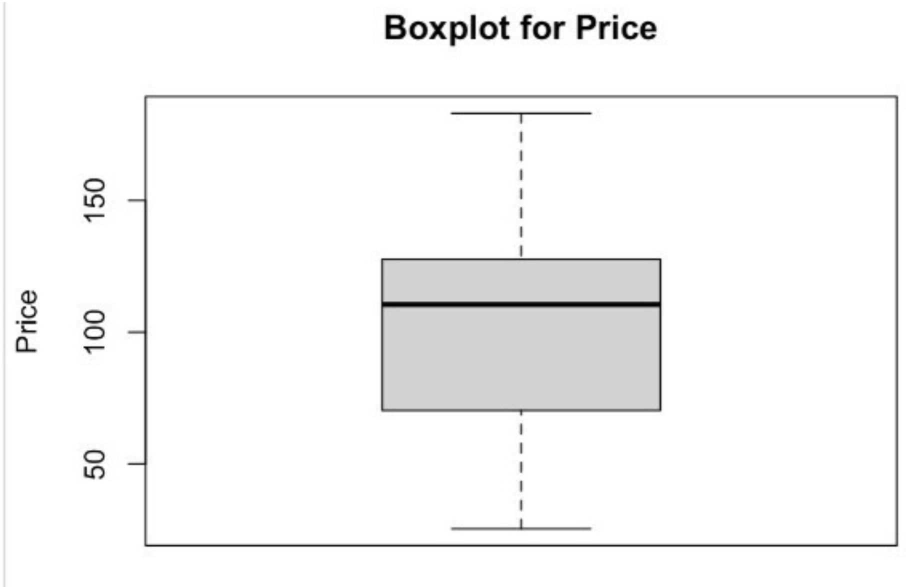
Model: "sequential"

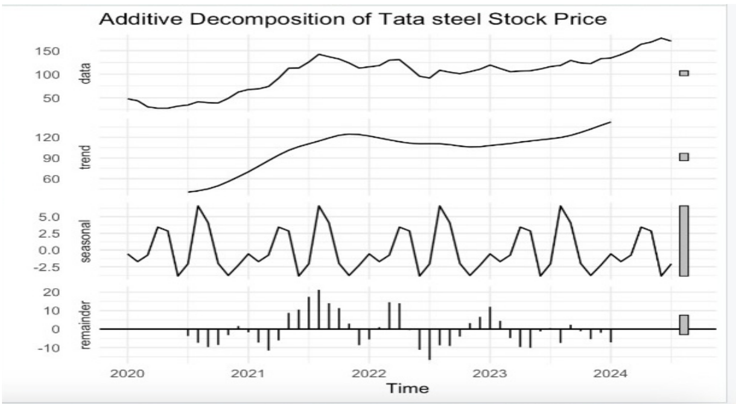
Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 30, 50)	11,400
dropout (Dropout)	(None, 30, 50)	0
lstm_1 (LSTM)	(None, 50)	20,200
dropout_1 (Dropout)	(None, 50)	0
dense (Dense)	(None, 1)	51

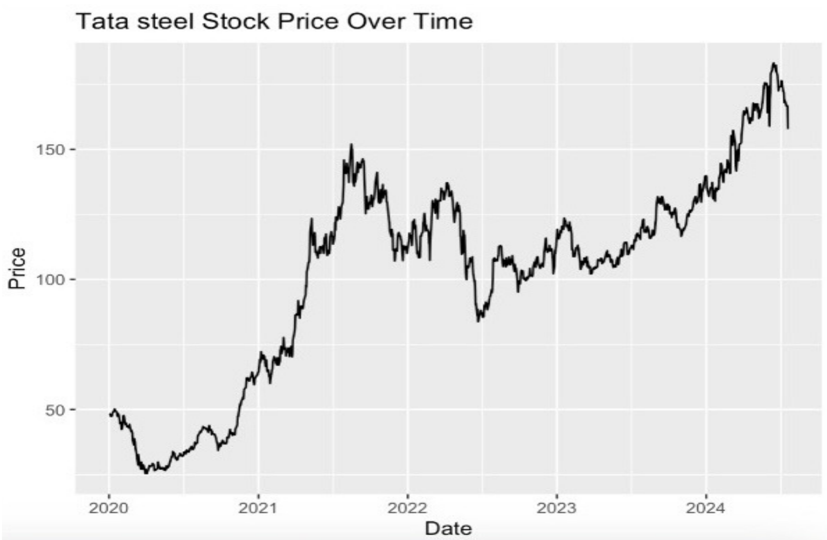
Total params: 31,651 (123.64 KB)
Trainable params: 31,651 (123.64 KB)
Non-trainable params: 0 (0.00 B)

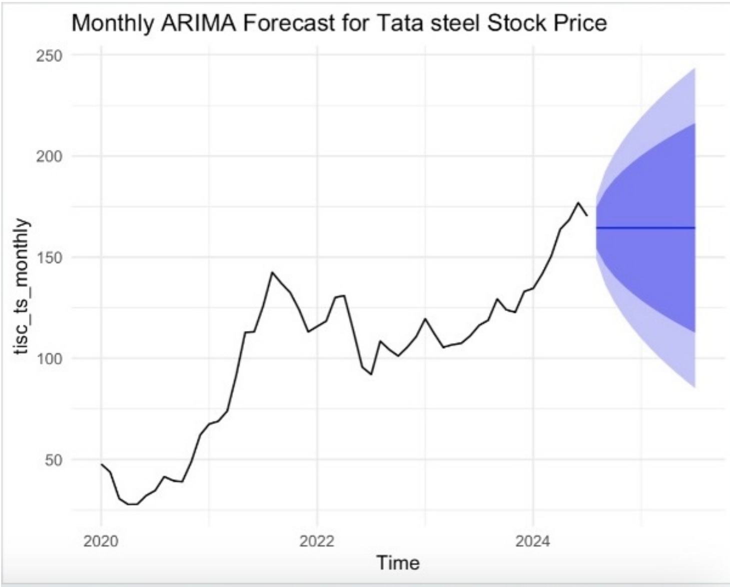
```
[67]: # Compile the model
      model.compile(optimizer='adam', loss='mean_squared_error')

      # Train the model
      history = model.fit(X_train, y_train, epochs=20, batch_size=32, validation_data=(X_test, y_test), shuffle=False)

      # Evaluate the model
      loss = model.evaluate(X_test, y_test)
      print(f"Test Loss: {loss}")
```

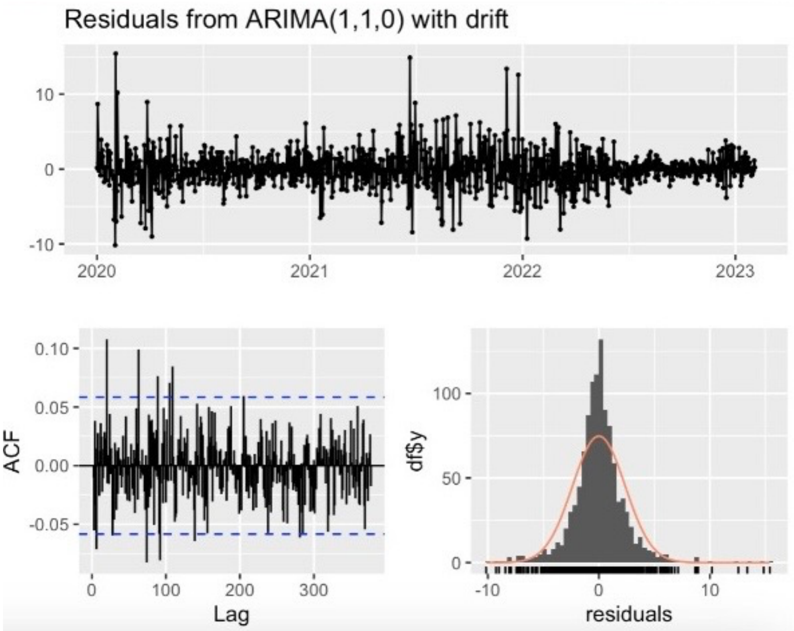
```
Epoch 1/20
18/18 ━━━━━━━━━━━━━━━━━━━━ 2s 26ms/step - loss: 0.0564 - val_loss: 0.0190
Epoch 2/20
18/18 ━━━━━━━━━━━━━━━━━━━━ 0s 14ms/step - loss: 0.1642 - val_loss: 0.1540
Epoch 3/20
18/18 ━━━━━━━━━━━━━━━━━━━━ 0s 14ms/step - loss: 0.0047 - val_loss: 0.0078
```

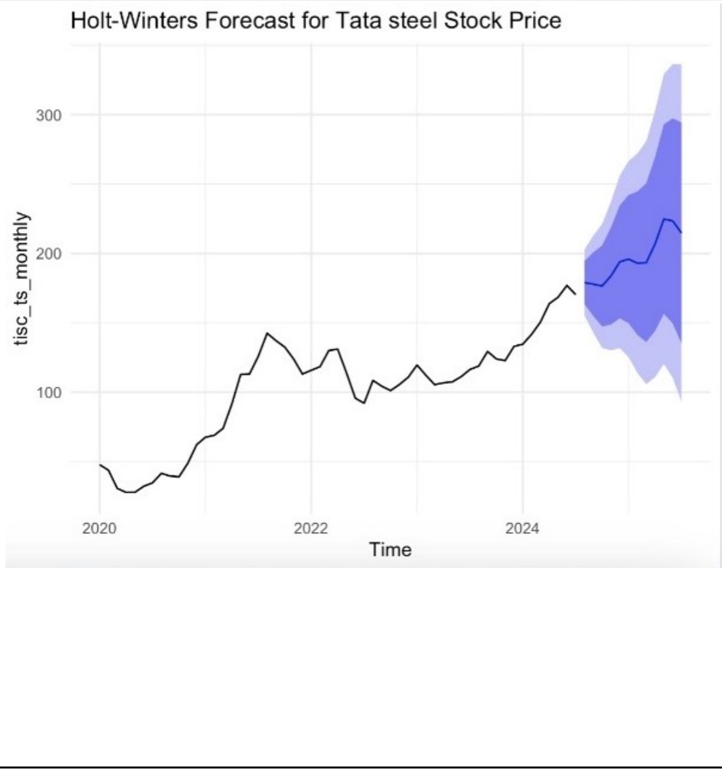

Boxplot for Price
Price
50
100
150

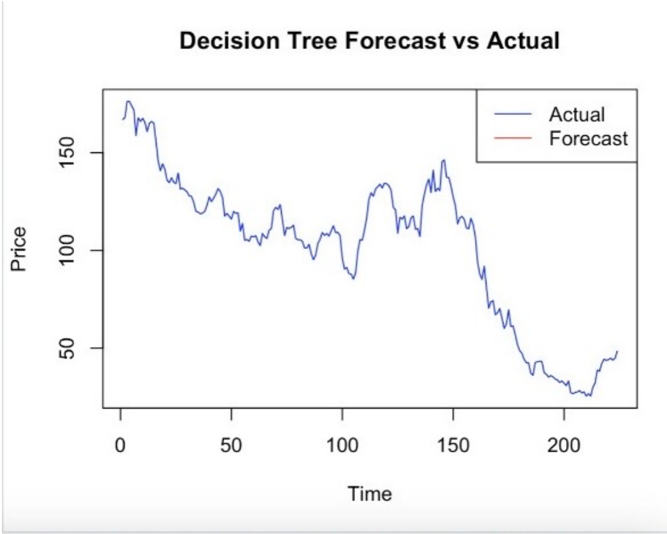

Additive Decomposition of Tata steel Stock Price
data
trend
seasonal
remainder
150
100
50
120
90
60
5.0
2.5
0.0
-2.5
20
10
0
-10
2020
2021
2022
2023
2024
Time

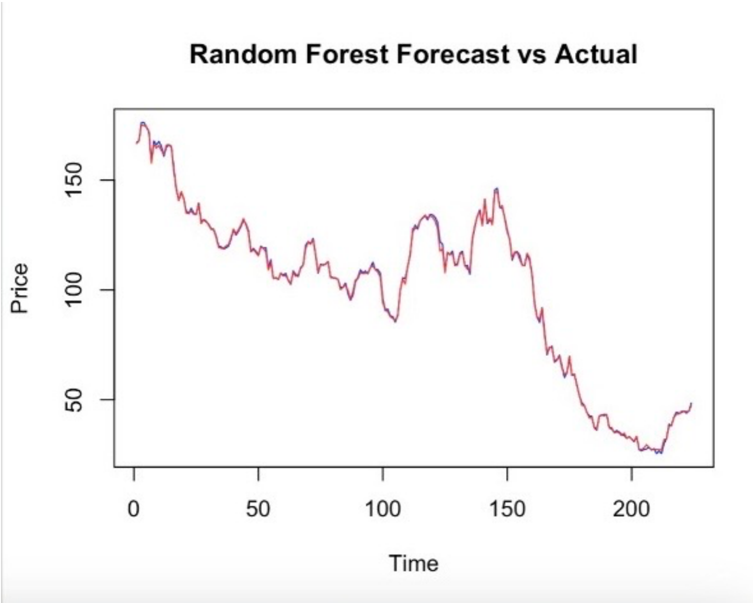

Tata steel Stock Price Over Time
Price
150
100
50
2020
2021
2022
2023
2024
Date

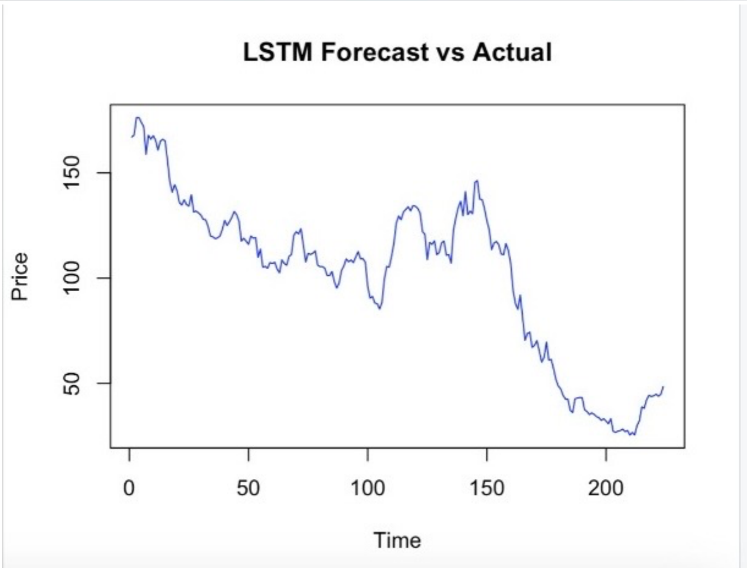

Monthly ARIMA Forecast for Tata steel Stock Price
250
200
150
100
50
tisc_ts_monthly
2020
2022
2024
Time


Residuals from ARIMA(1,1,0) with drift
10
0
-10
2020
2021
2022
2023
0.10
0.05
0.00
-0.05
ACF
0
100
200
300
Lag
100
50
0
df$y
-10
0
10
residuals


Holt-Winters Forecast for Tata steel Stock Price
300
200
100
tisc_ts_monthly
2020
2022
2024
Time


Decision Tree Forecast vs Actual
Actual
Forecast
Price
150
100
50
0
50
100
150
200
Time


Random Forest Forecast vs Actual
Price
150
100
50
0
50
100
150
200
Time


LSTM Forecast vs Actual
Price
150
100
50
0
50
100
150
200
Time