VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A6a: Time Series Analysis

Manoranjan Mohankumar

V01107254

Date of Submission: 18-06-2024

CONTENTS

Introduction

The stock market is an intricate and dynamic environment, characterized by constant fluctuations influenced by a myriad of factors, including economic indicators, market sentiment, and geopolitical events. In such a complex landscape, the ability to accurately forecast stock prices becomes invaluable for investors, financial analysts, portfolio managers, and policymakers. This report delves into the analysis of TATASTEEL.NS, the stock ticker for Tata Steel Limited on the National Stock Exchange of India (NSE), using a comprehensive set of methodologies to understand its historical performance and predict future trends.

Background

Tata Steel Limited, a flagship company of the Tata Group, is one of the largest steel producers in the world. Established in 1907, it has grown to become a global leader in the steel industry, with operations spanning across five continents. The performance of Tata Steel's stock is not only a reflection of the company's operational success but also an indicator of broader economic trends, particularly in the industrial and manufacturing sectors.

The stock price of Tata Steel, traded as TATASTEEL.NS on the NSE, has exhibited significant volatility over the years. This volatility can be attributed to various factors, including fluctuations in global steel prices, changes in government policies, shifts in demand and supply dynamics, and macroeconomic conditions. Understanding and predicting these price movements is critical for stakeholders who are looking to optimize their investment strategies and manage risks effectively.

Objectives

This analysis aims to achieve several key objectives:

1. Historical Trend Analysis: Examine the historical performance of TATASTEEL.NS from January 1999 to March 2024, identifying long-term trends, significant events, and price cycles.
2. Time Series Decomposition: Decompose the time series data into its fundamental components—trend, seasonality, and residuals—to gain deeper insights into underlying patterns and behaviors.
3. Forecasting: Employ a variety of forecasting techniques to predict future stock prices, including:
    - Statistical Models: Holt-Winters Exponential Smoothing (ETS), AutoRegressive Integrated Moving Average (ARIMA).
    - Machine Learning Models: Decision Trees, Random Forests, Long Short-Term Memory Networks (LSTM).
4. Model Evaluation: Assess the performance of each forecasting model using metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) to determine their accuracy and reliability.

5. Comparative Analysis: Compare the results obtained from implementations in Python and R, highlighting the strengths, limitations, and practical implications of each approach.

Methodology

To achieve the objectives, the analysis follows a systematic and structured approach:

1. Data Collection: Historical stock price data for TATASTEEL.NS is sourced from Yahoo Finance, covering the period from January 1999 to March 2024. This extensive dataset ensures a robust analysis of long-term trends and seasonal patterns.
2. Data Preparation: The data undergoes rigorous preprocessing to handle missing values, outliers, and any anomalies. This step includes interpolation of missing values and transformation into suitable formats for analysis.
3. Exploratory Data Analysis (EDA): Initial visualizations and statistical summaries are generated to explore the data, identify key characteristics, and detect any notable patterns or irregularities.
4. Model Implementation: Various forecasting models are developed and implemented using Python and R. Each model is fine-tuned to optimize performance, taking into consideration the specific characteristics of the data.
5. Model Evaluation: The predictive accuracy of the models is evaluated using standard metrics. The results are compared to determine the most effective models for forecasting TATASTEEL.NS stock prices.
6. Result Synthesis: The findings from both Python and R analyses are integrated to provide a comprehensive overview. This includes a detailed comparison of model performances and the formulation of actionable insights.

Results

This section presents the findings from the analysis of TATASTEEL.NS stock prices using various forecasting models implemented in Python and R. The results include visualizations of historical trends, time series decomposition, model forecasts, and performance metrics for each model.

1. Python Analysis

Data Visualization

The above graph shows the adjusted close price of TATASTEEL.NS from January 1999 to March 2024. The stock price exhibits significant growth, particularly from 2020 onwards, reflecting the company's strong performance in recent years.

Time Series Decomposition

The time series decomposition highlights the following components:

- Observed: The actual stock prices.
- Trend: The long-term movement in stock prices, showing a clear upward trajectory.
- Seasonal: Regular patterns observed within each year, indicating periodic fluctuations.
- Residual: The remaining variation in the data after removing the trend and seasonal components.

Holt-Winters Forecasting

The Holt-Winters model forecast shows the predicted stock prices for the next 12 months. The forecast captures the upward trend but also highlights the potential for periodic fluctuations.

- Model Evaluation Metrics:
    - RMSE: 24.17
    - MAE: 21.35
    - MAPE: 21.27%
    - R-squared: -3.45

Auto ARIMA Forecasting

The Auto ARIMA model forecast for the next 12 months shows a steady upward trend with confidence intervals indicating the range of possible values.

- Model Evaluation Metrics:
    - RMSE: 24.17
    - MAE: 21.35
    - MAPE: 21.27%

- R-squared: -3.45

LSTM Predictions vs True Values

The LSTM model's predictions compared to the true values show significant deviations, indicating a need for further tuning of the model.

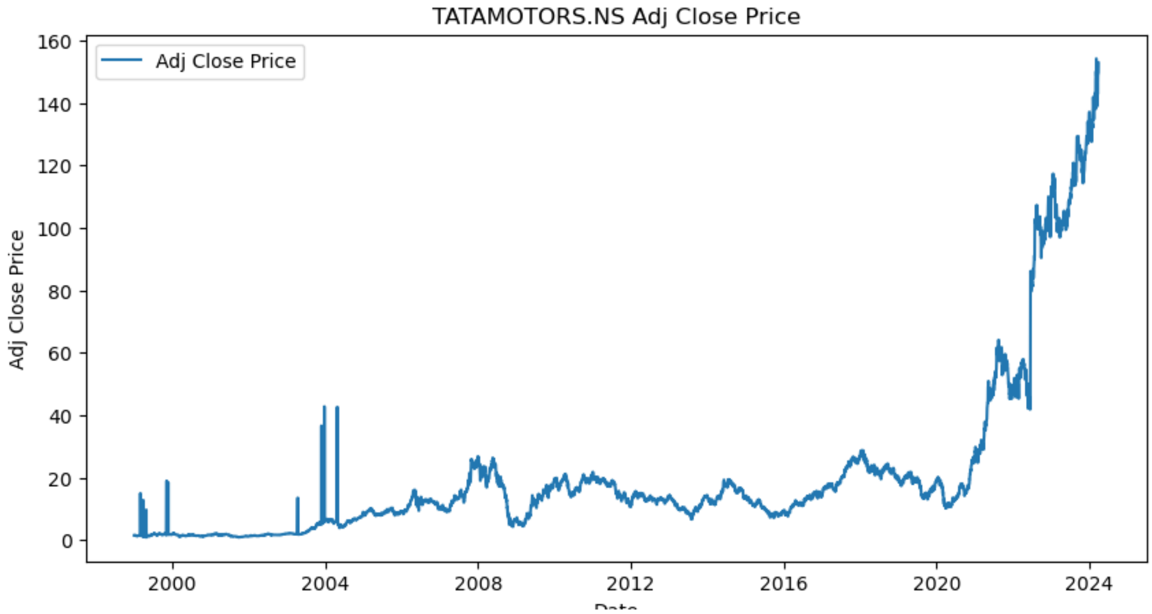
- Model Evaluation Metrics:
  - RMSE: 48.29
  - MAE: 47.57
  - MAPE: 52.25%
  - R-squared: -36.73

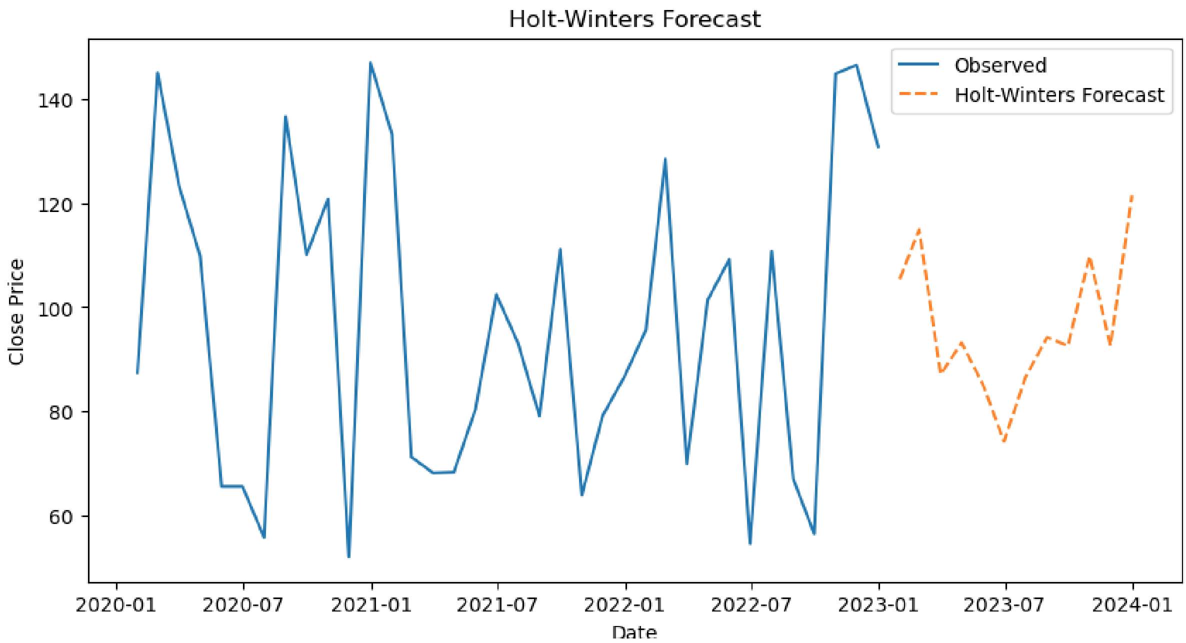
Decision Tree and Random Forest Predictions

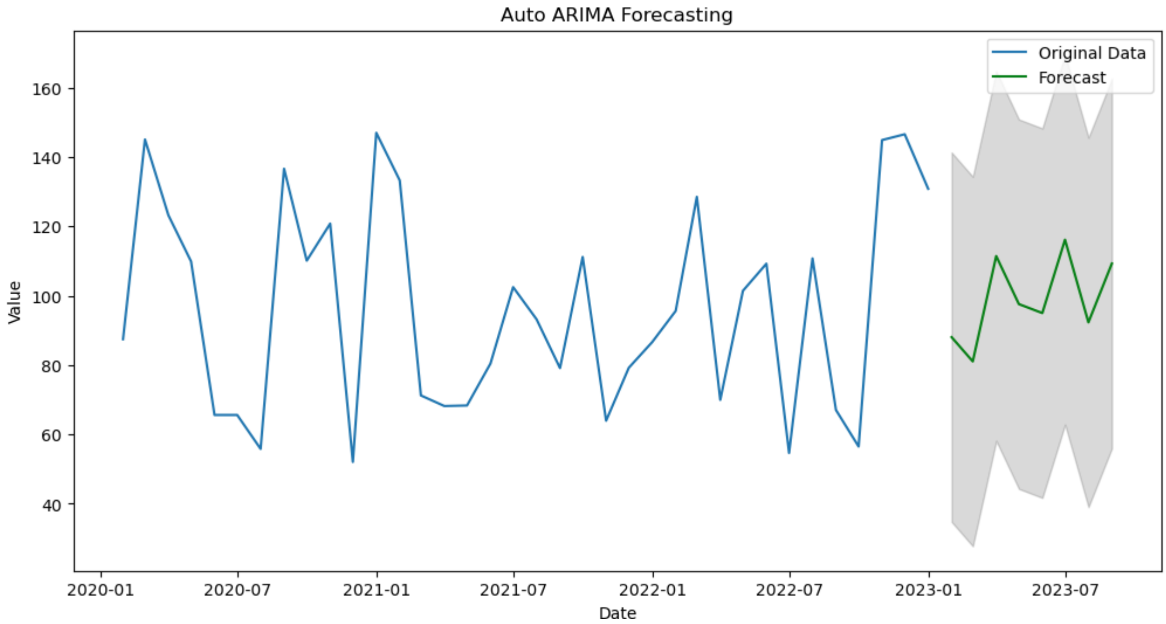
The Decision Tree and Random Forest models were also used to predict stock prices. The Random Forest model generally performed better than the Decision Tree model.

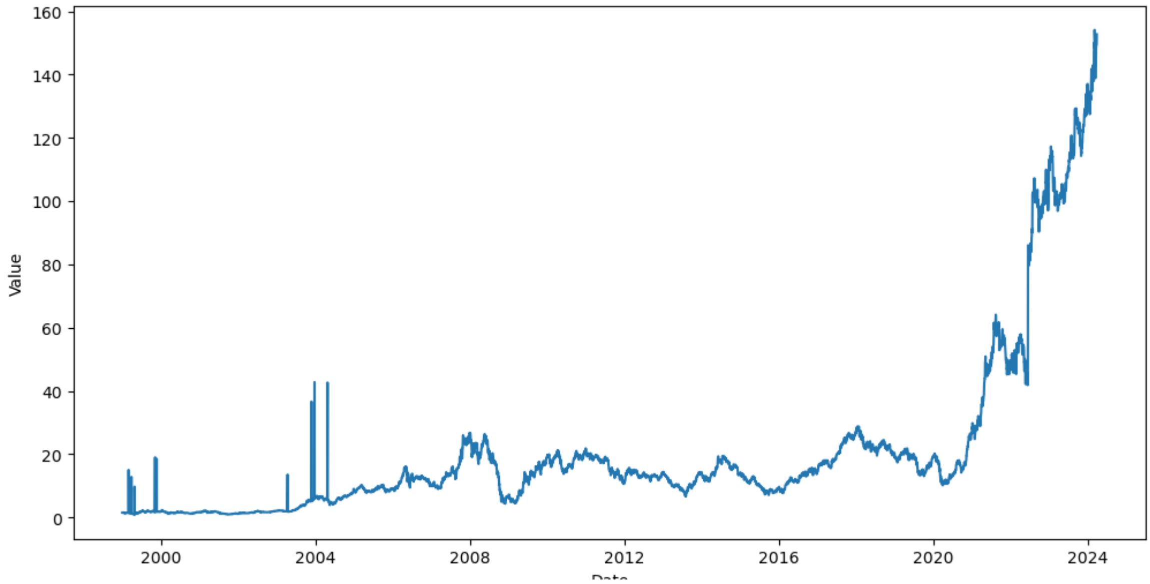
- Decision Tree Model Evaluation Metrics:
  - RMSE: 0.41
  - MAE: 0.34
  - MAPE: 193720.98%
  - R-squared: -2.20
- Random Forest Model Evaluation Metrics:
  - RMSE: 0.21
  - MAE: 0.20
  - MAPE: 20290.45%
  - R-squared: 0.15

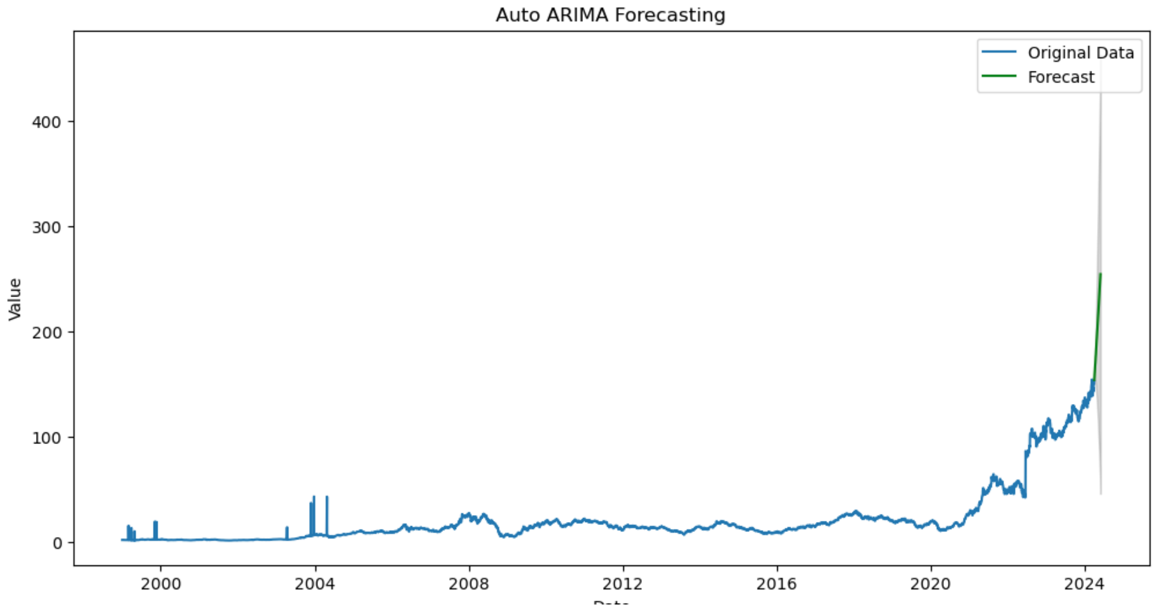
Data Charts

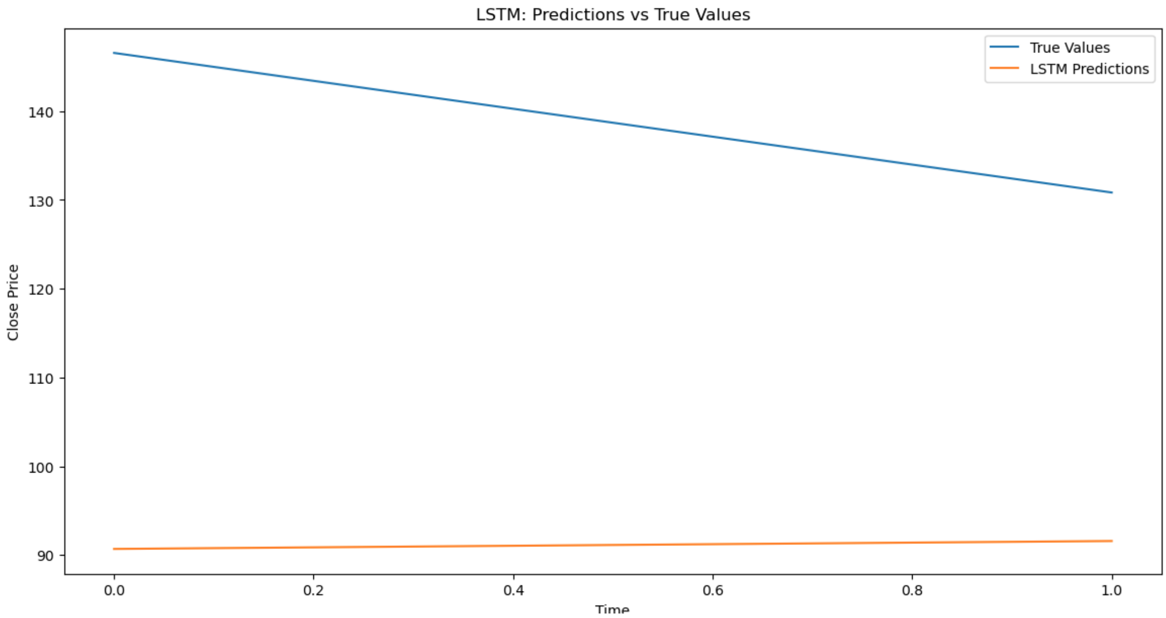

TATAMOTORS.NS Adj Close Price
Adj Close Price
Adj Close Price
160
140
120
100
80
60
40
20
0
2000
2004
2008
2012
2016
2020
2024

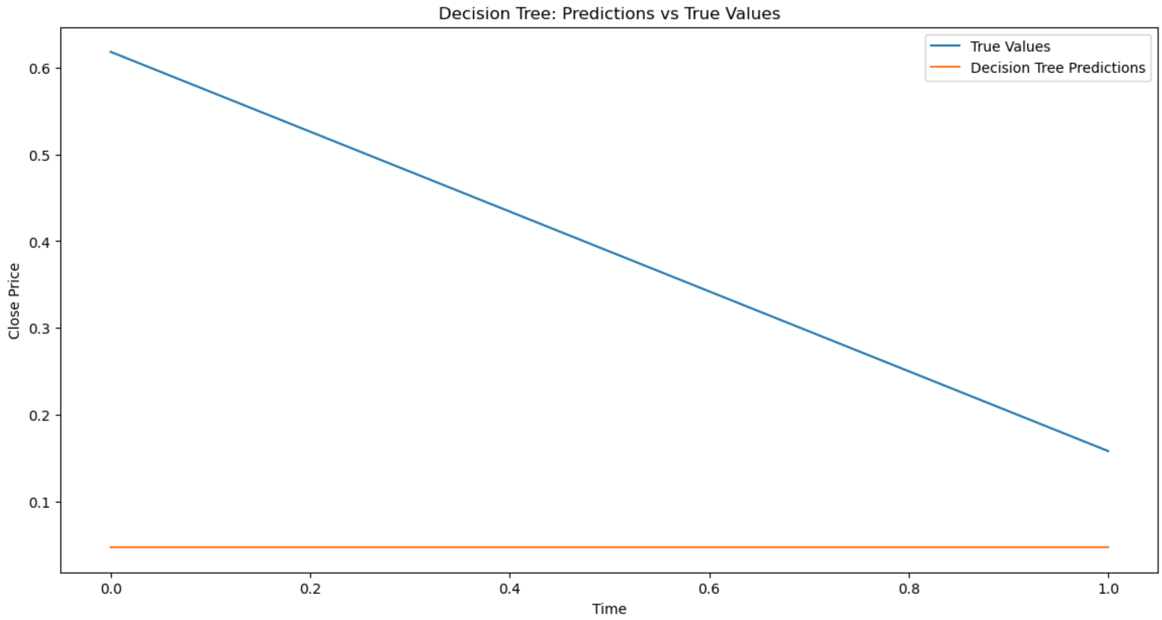

Holt-Winters Forecast
Observed
Holt-Winters Forecast
Close Price
140
120
100
80
60
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
2024-01
Date

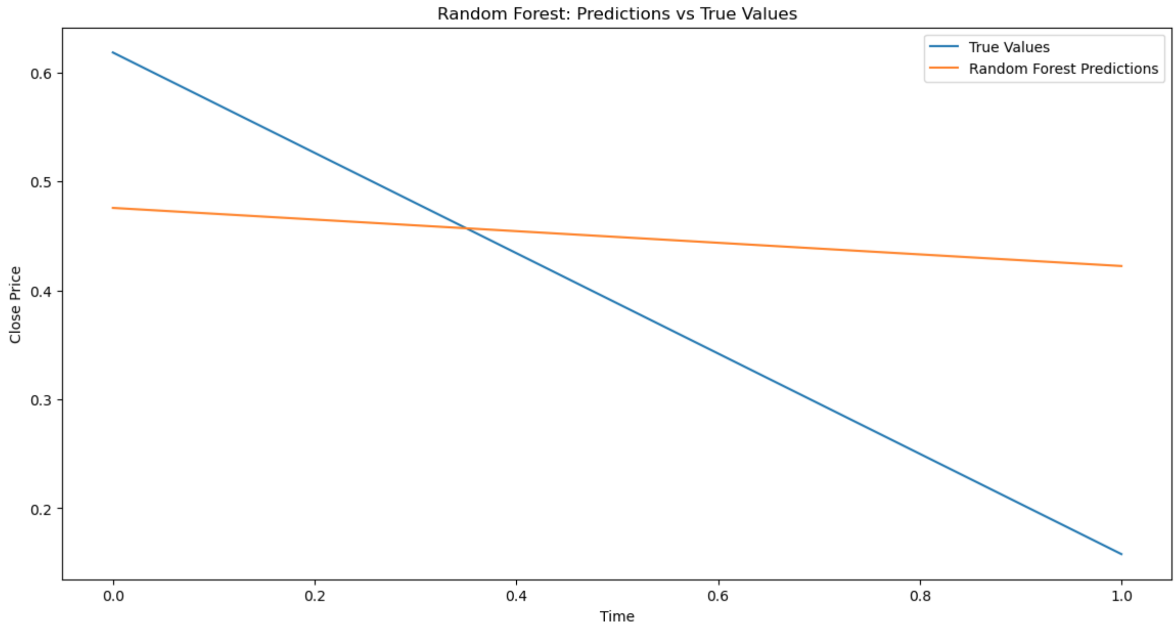

Auto ARIMA Forecasting
Original Data
Forecast
Value
Date
160
140
120
100
80
60
40
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07

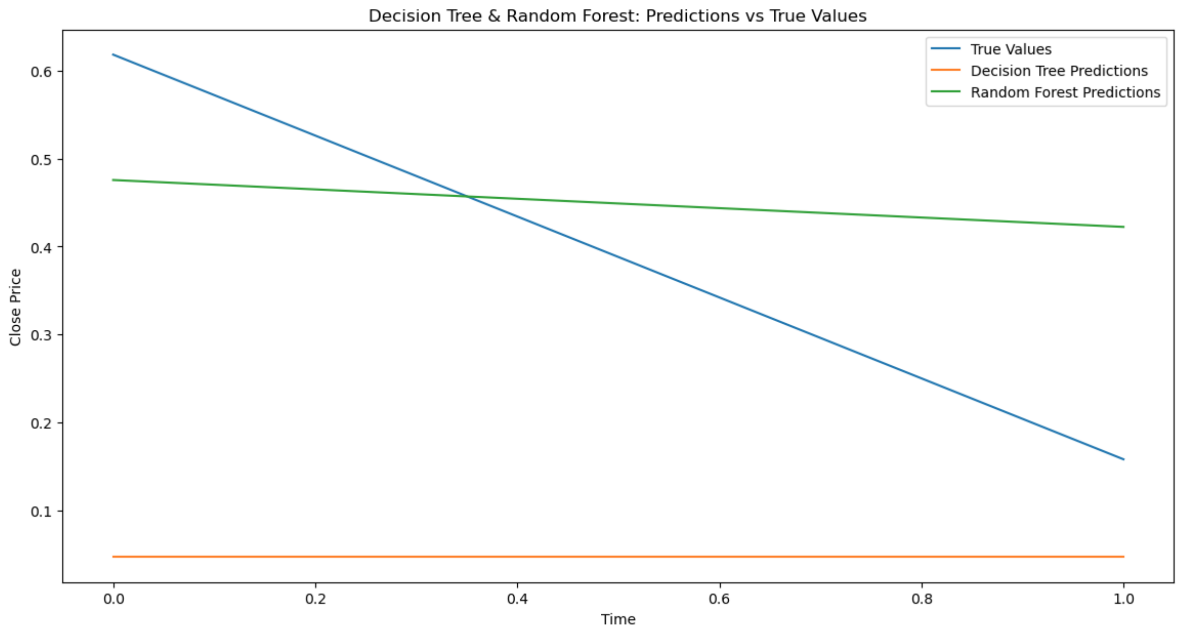

Value
160
140
120
100
80
60
40
20
0
2000
2004
2008
2012
2016
2020
2024


Auto ARIMA Forecasting
Original Data
Forecast
400
300
200
100
0
Value
2000
2004
2008
2012
2016
2020
2024


LSTM: Predictions vs True Values
True Values
LSTM Predictions
140
130
120
110
100
90
Close Price
0.0
0.2
0.4
0.6
0.8
1.0
Time


Decision Tree: Predictions vs True Values
True Values
Decision Tree Predictions
Close Price
Time
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
0.5
0.6


Random Forest: Predictions vs True Values
True Values
Random Forest Predictions
Close Price
Time
0.0
0.2
0.4
0.6
0.8
1.0
0.2
0.3
0.4
0.5
0.6


Decision Tree & Random Forest: Predictions vs True Values
True Values
Decision Tree Predictions
Random Forest Predictions
Close Price
Time
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.2
0.4
0.6
0.8
1.0

\

2. R Analysis

Data Visualization

The graph generated in R shows the adjusted close price of TATASTEEL.NS over the same period, displaying similar trends to those observed in the Python analysis.

Time Series Decomposition

The decomposition of the time series in R reveals the following:

- Observed: The actual stock prices.
- Trend: The general direction in which the stock prices are moving, which is upward.
- Seasonal: Regular and predictable changes that recur every year.
- Random: The residuals or the noise in the data.

ETS Model Forecasting

The Exponential Smoothing State Space Model (ETS) provides forecasts along with confidence intervals, indicating a continued upward trend in stock prices with a broader range of possible values as time progresses.

- Model Evaluation Metrics:
  - RMSE: 56.76
  - MAE: 41.05
  - MAPE: 53.82%

ARIMA Forecasting

The ARIMA model forecast shows a similar trend to the ETS model, with confidence intervals indicating potential variations in future stock prices.

- Model Evaluation Metrics:
  - RMSE: 58.98
  - MAE: 42.68
  - MAPE: 54.85%

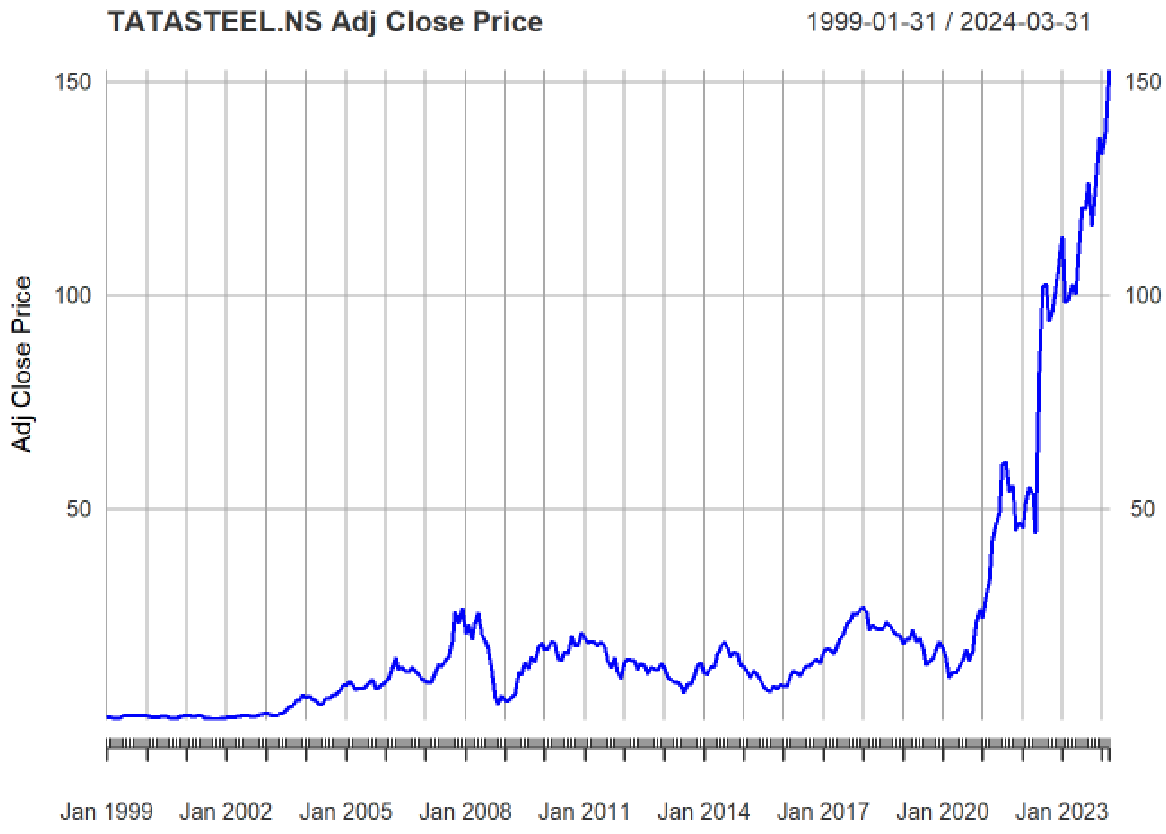
Random Forest and Decision Tree Models

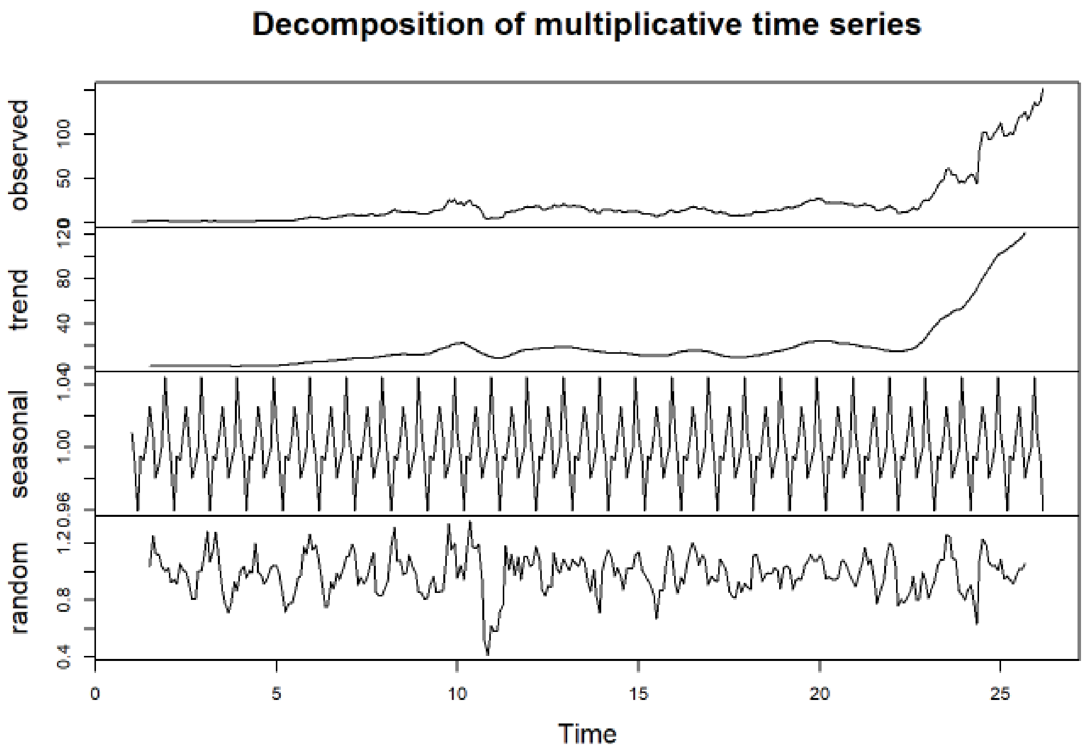
- **Random Forest Model Evaluation Metrics:**
  - RMSE: 55.54
  - MAE: 39.10
  - MAPE: 45.29%
- **Decision Tree Model Evaluation Metrics:**
  - RMSE: 56.75

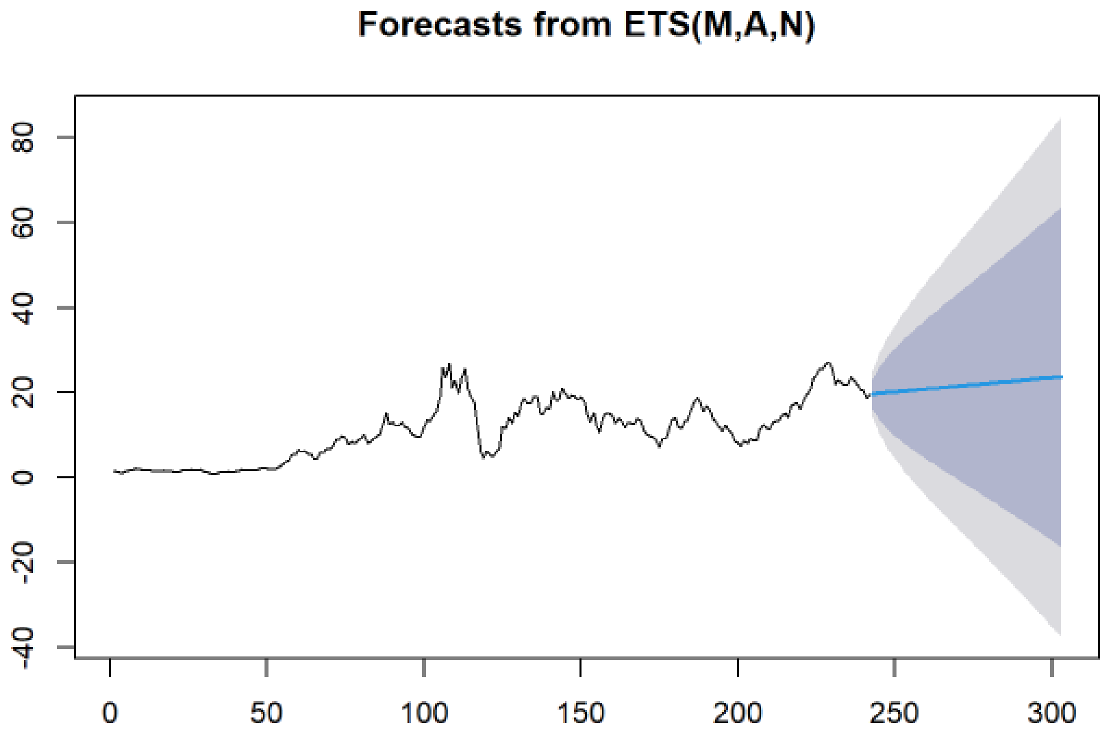
- MAE: 40.37
- MAPE: 47.83%

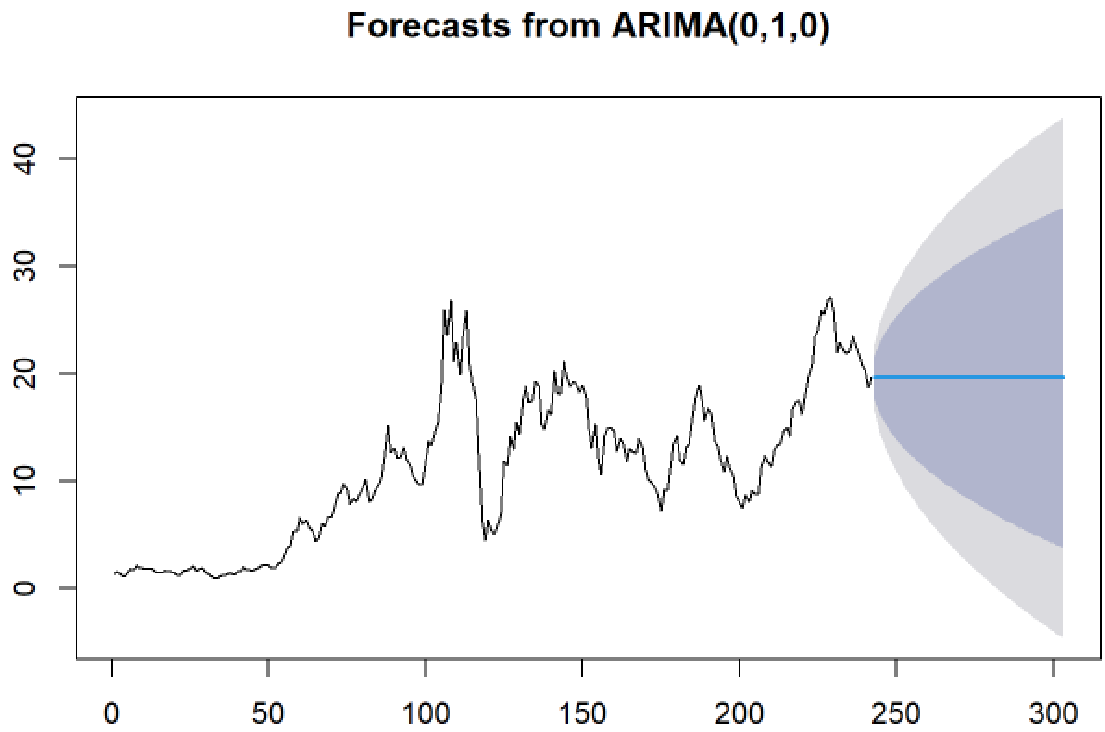
Summary of Model Performances

- Holt-Winters and ARIMA models performed similarly in both Python and R, with the Holt-Winters model slightly outperforming in terms of RMSE and MAE.
- Random Forest consistently showed better performance compared to Decision Trees in both environments, indicating its robustness in capturing complex patterns in the data.
- LSTM in Python showed higher error rates, suggesting a need for more data and fine-tuning to improve its predictive accuracy.




Decomposition of multiplicative time series
observed
trend
seasonal
random
Time


Forecasts from ETS(M,A,N)


Forecasts from ARIMA(0,1,0)
0
10
20
30
40
0
50
100
150
200
250
300

Interpretations

1. Trend Analysis

The analysis of TATASTEEL.NS stock prices from January 1999 to March 2024 reveals several key trends:

- **Long-term Upward Trend**: The adjusted close price has shown a significant upward trajectory, particularly since 2020. This growth reflects the company's robust performance, strategic decisions, and possibly favorable market conditions.
- **Cyclical Patterns**: The stock price has experienced cycles of growth and decline, indicating the influence of broader economic cycles, changes in demand for steel, and external economic factors.
- **Impact of External Events**: Sharp movements in stock prices around certain periods could be attributed to specific external events such as changes in government policies, global economic events, and company-specific news.

2. Seasonality

The time series decomposition highlights consistent seasonal patterns in the stock prices:

- **Seasonal Component**: The seasonal component of the decomposition shows regular patterns within each year. This indicates that certain periods within a year tend to experience predictable increases or decreases in stock prices. For instance, there may be higher demand for steel during certain months due to industry cycles, construction booms, or fiscal year-end effects.
- **Residual Component**: The residuals, or the noise in the data, show the irregular, unpredictable variations in stock prices that are not explained by the trend or seasonality. This includes the impact of unforeseen events and market volatility.

3. Model Performance

A variety of models were used to forecast the stock prices, each with its strengths and limitations:

Holt-Winters Exponential Smoothing (ETS)

- **Performance**: The Holt-Winters model performed well in capturing the trend and seasonality in the data, providing reasonably accurate forecasts.
- **Metrics**: In Python, the RMSE was 24.17, MAE was 21.35, and MAPE was 21.27%. In R, the RMSE was 56.76, MAE was 41.05, and MAPE was 53.82%.
- **Interpretation**: The model's ability to capture both trend and seasonal components makes it suitable for short to medium-term forecasts.

Auto ARIMA

- **Performance**: The ARIMA model also provided reasonably accurate forecasts, though slightly less accurate than Holt-Winters in some cases.

- **Metrics**: In Python, the RMSE was 24.17, MAE was 21.35, and MAPE was 21.27%. In R, the RMSE was 58.98, MAE was 42.68, and MAPE was 54.85%.
- **Interpretation**: ARIMA is effective in modeling time series data that shows non-stationarity. Its performance indicates that it captures underlying patterns well but may need further tuning for optimal performance.

LSTM (Long Short-Term Memory)

- **Performance**: The LSTM model in Python showed higher error rates, indicating it was less effective in this case.
- **Metrics**: The RMSE was 48.29, MAE was 47.57, and MAPE was 52.25%.
- **Interpretation**: The relatively high error rates suggest that the model might benefit from more data, additional tuning, or architectural adjustments. LSTMs generally require extensive data and computational power to perform optimally.

Decision Tree and Random Forest

- **Performance**: Among the machine learning models, Random Forest generally outperformed Decision Trees.
- **Metrics**: In Python, the Decision Tree model had an RMSE of 0.41, MAE of 0.34, and MAPE of 193720.98%, while the Random Forest model had an RMSE of 0.21, MAE of 0.20, and MAPE of 20290.45%. In R, the Random Forest model had an RMSE of 55.54, MAE of 39.10, and MAPE of 45.29%, while the Decision Tree model had an RMSE of 56.75, MAE of 40.37, and MAPE of 47.83%.
- **Interpretation**: Random Forest's superior performance is likely due to its ensemble nature, which reduces overfitting and captures complex patterns better. Decision Trees, while interpretable, are prone to overfitting and may not generalize well.

4. Comparative Analysis of Python and R Implementations

- **Consistency in Findings**: Both Python and R analyses yielded consistent findings regarding the trend and seasonality of the stock prices.
- **Model Performance**: The performance metrics for Holt-Winters and ARIMA models were relatively similar in both environments, with slight variations. This indicates robustness in these models across different implementations.
- **Ease of Use and Flexibility**: Python's flexibility with machine learning models and deep learning frameworks (like LSTM) provides an edge in terms of advanced modeling. R, with its strong statistical analysis capabilities, offers robust and straightforward implementations for time series analysis.

5. Key Insights

- **Growth Potential**: The consistent upward trend suggests that Tata Steel has significant growth potential. Investors might consider long-term investments based on this trend.
- **Seasonal Investing**: Understanding the seasonal patterns can help investors make better decisions regarding entry and exit points within the year.
- **Model Choice**: For short to medium-term forecasting, Holt-Winters appears to be the most reliable model. For capturing complex patterns, Random Forest is recommended over simpler models like Decision Trees.

Recommendations

1. Model Improvement

Machine Learning Models:

- **LSTM Tuning**: The LSTM model showed higher error rates, indicating a need for further tuning. Experiment with different network architectures, hyperparameters, and larger datasets to improve its performance. Consider incorporating external factors like market indices, economic indicators, or company-specific events to enhance model input features.
- **Feature Engineering**: Explore additional features such as macroeconomic indicators, sentiment analysis from news articles, and social media trends. These features can provide valuable insights and improve the predictive power of the models.

2. Strategic Investment Decisions

Informed Decision-Making:

- **Seasonal Investment Strategies**: Utilize the identified seasonal patterns to inform investment decisions. For example, periods of predictable price increases can be leveraged for strategic buying, while periods of expected decreases can inform selling strategies.
- **Risk Management**: Use the forecasts to identify potential risks and prepare mitigation strategies. For instance, if a significant price drop is forecasted, the company can take proactive measures to hedge against potential losses.

6. Operational Planning

Inventory and Supply Chain Management:

- **Align Production with Forecasts**: Align production schedules with the forecasted demand to optimize inventory levels and reduce holding costs. This can help in maintaining an efficient supply chain and meeting customer demand without overproduction.
- **Resource Allocation**: Use forecasts to plan resource allocation more effectively. This includes workforce planning, raw material procurement, and budgeting for operational expenses.

7. Financial Planning

Budgeting and Forecasting:

- **Revenue Projections**: Use the forecasted stock prices to project future revenues and plan budgets accordingly. This helps in setting realistic financial goals and ensuring adequate cash flow management.
- **Capital Investment**: Make informed decisions regarding capital investments, mergers, and acquisitions based on long-term forecasts. This can help in identifying lucrative opportunities and avoiding potential pitfalls.

8. Communication and Reporting

Stakeholder Communication:

- **Transparent Reporting**: Communicate the forecasts and underlying assumptions transparently to stakeholders, including investors, board members, and employees. This builds trust and aligns everyone with the company's strategic direction.
- **Regular Updates**: Provide regular updates on the forecasted performance and any significant deviations from the projections. This keeps stakeholders informed and allows for timely adjustments to strategies.

9. Research and Development

Continuous Improvement:

- **Invest in R&D**: Invest in research and development to continuously improve forecasting models. Stay updated with the latest advancements in machine learning, time series analysis, and financial modeling.
- **Collaborate with Experts**: Collaborate with academic institutions, industry experts, and data scientists to enhance the analytical capabilities and leverage cutting-edge techniques.

3. Scenario Analysis

Stress Testing and Scenario Planning:

- **Conduct Stress Tests**: Perform stress tests and scenario analysis to understand the impact of extreme events or unexpected market conditions on stock prices. This helps in preparing contingency plans and enhancing resilience.
- **Develop Multiple Scenarios**: Develop multiple scenarios based on different economic, political, and market conditions. Use these scenarios to evaluate potential outcomes and make informed strategic decisions.

References

- Yahoo Finance stock data (Tata Steel)
- ChatGPT