VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A6a-Time Series Analysis

MICAH ASHADEEP EMMANUEL

V01101166

Date of Submission: 22-07-2024

CONTENTS

INTRODUCTION

Mahindra & Mahindra (M&M), a major Indian conglomerate and global tractor leader, has its stock performance closely watched by investors due to its influence on the auto sector and the Indian economy. Accurate stock price forecasting can offer a competitive edge to investors by enabling informed decision-making, timely trading, and risk management. This project tackles the inherent volatility of stock markets by employing historical M&M data to build models that predict future prices. By combining traditional statistics and advanced machine learning, this analysis aims to comprehensively understand M&M's stock behavior and provide valuable insights for informed investment decisions.

OBJECTIVES

Data Acquisition:

- Retrieve historical stock price data of M&M

Data Cleaning and Preparation:

- Handle missing values through interpolation.
- Detect and address outliers.
- Split the data into training and testing sets.

Data Transformation:

- Convert daily stock prices to monthly averages.
- Decompose the time series data to analyze underlying components such as trend, seasonality, and residuals.

Univariate Forecasting Models:

- Apply the Holt-Winters method to forecast stock prices for the next year.
- Develop ARIMA and SARIMA models to forecast daily stock prices and perform diagnostic checks.
- Fit the ARIMA model to monthly stock price data.

Multivariate Forecasting Models:

- Utilize Long Short-Term Memory (LSTM) neural networks for capturing complex patterns and long-term dependencies in stock prices.
- Implement tree-based models like Random Forest and Decision Tree for comparative analysis and additional insights.

BUSINESS SIGNIFICANCE

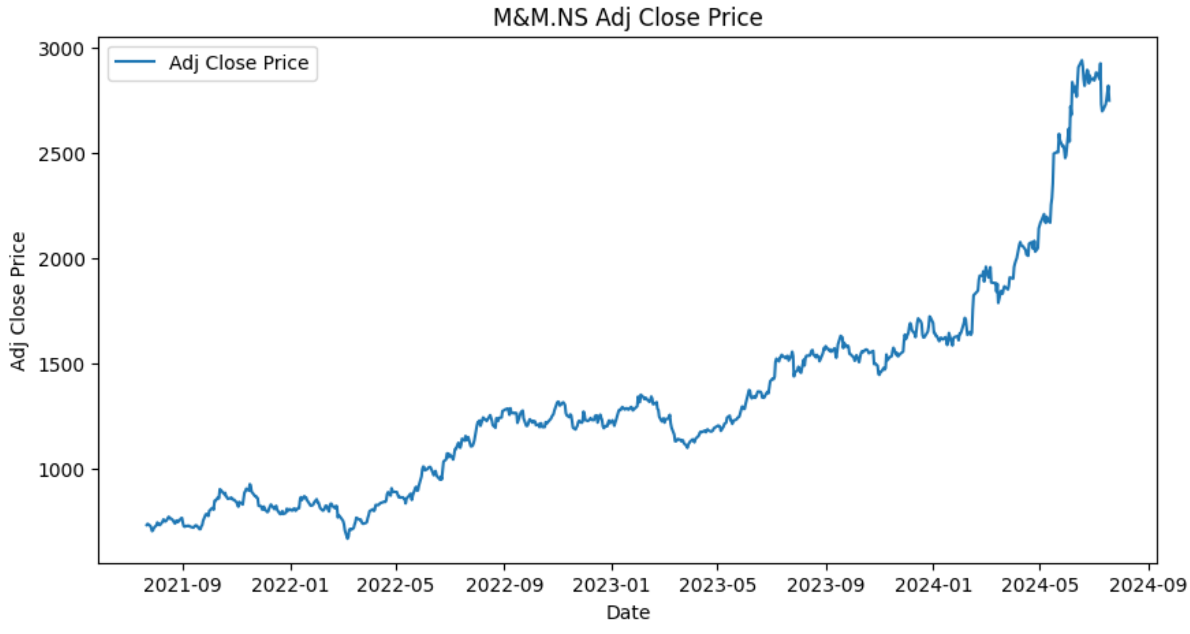
Accurate forecasting of stock prices is crucial for several reasons:

- **Investment Decisions:** Investors rely on forecasts to make buy, hold, or sell decisions. Accurate predictions can lead to better investment strategies and improved portfolio performance.
- **Risk Management:** Understanding potential future price movements helps in mitigating risks associated with stock investments. This is particularly important for institutional investors and fund managers who manage large portfolios.
- **Market Efficiency:** Accurate forecasting contributes to market efficiency by ensuring that stock prices reflect all available information. This leads to fair pricing and reduces the chances of market anomalies.
- **Strategic Planning:** Companies, especially those whose stock is being forecasted, can use these insights for strategic planning and decision-making, such as timing their public announcements or financial disclosures.

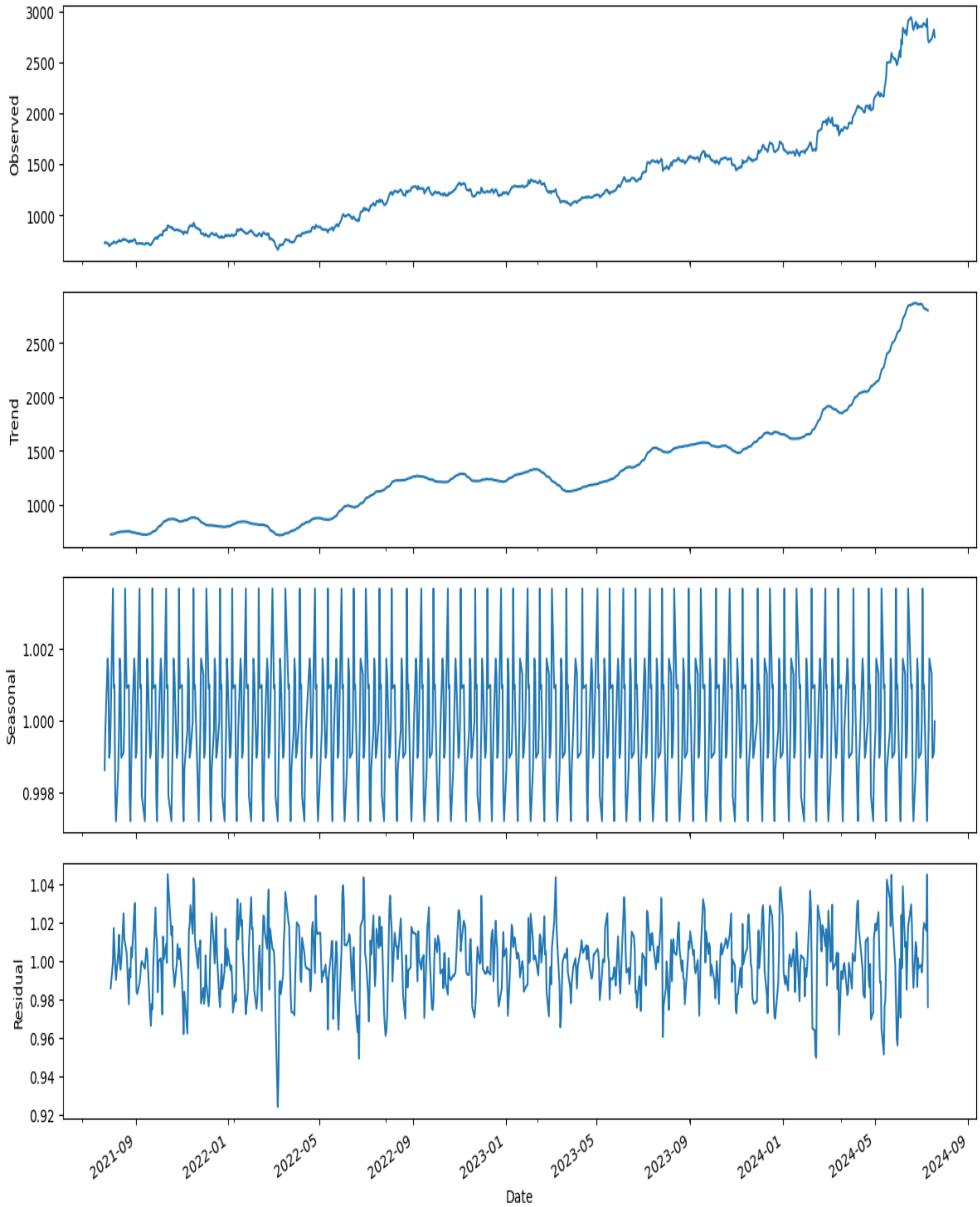
By applying a blend of statistical and machine learning techniques, this analysis aims to enhance the precision of stock price predictions for M&M, thereby offering a valuable tool for investors and financial analysts. This comprehensive approach not only sheds light on the stock's past performance but also provides actionable insights for future trading strategies.

RESULTS

USING PYTHON



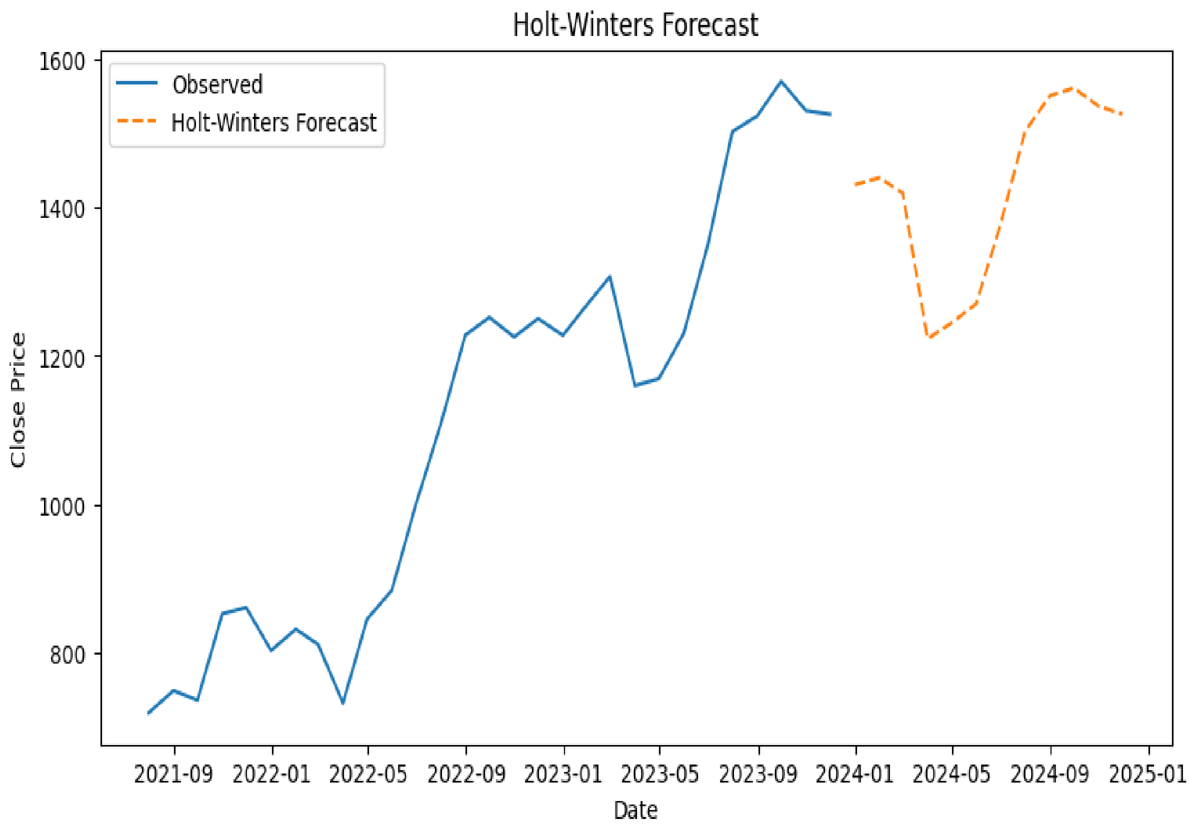
#The above Diagram plots the time series analysis of adjusted closing price.



#The above diagrams represent the decomposed components of time series named Observed, Trend, Seasonal and Residual.

Univariate Forecasting - Conventional Models/Statistical Models

HW model



#The above chart shows the Univariate Forecasting using Holt-Winters Forecast for 12 months.

RMSE	879.86
MAE	754.093
MAPE	Nan
R- squared	-2.764

#The above table shows the accuracy under Holt Winters forecast

Forecast for the next 12 months

2023-12-31 1431.107514

2024-01-31 1440.549489

2024-02-29 1419.871670

2024-03-31 1223.135087

2024-04-30 1244.543043

2024-05-31 1270.645402

2024-06-30 1376.486328

2024-07-31 1503.187576

2024-08-31 1551.166220

2024-09-30 1561.652381

2024-10-31 1537.053009

2024-11-30 1526.358880

2024-12-31 1431.107514

2025-01-31 1440.549489

2025-02-28 1419.871670

2025-03-31 1223.135087

2025-04-30 1244.543043

2025-05-31 1270.645402

2025-06-30 1376.486328

2025-07-31 1503.187576

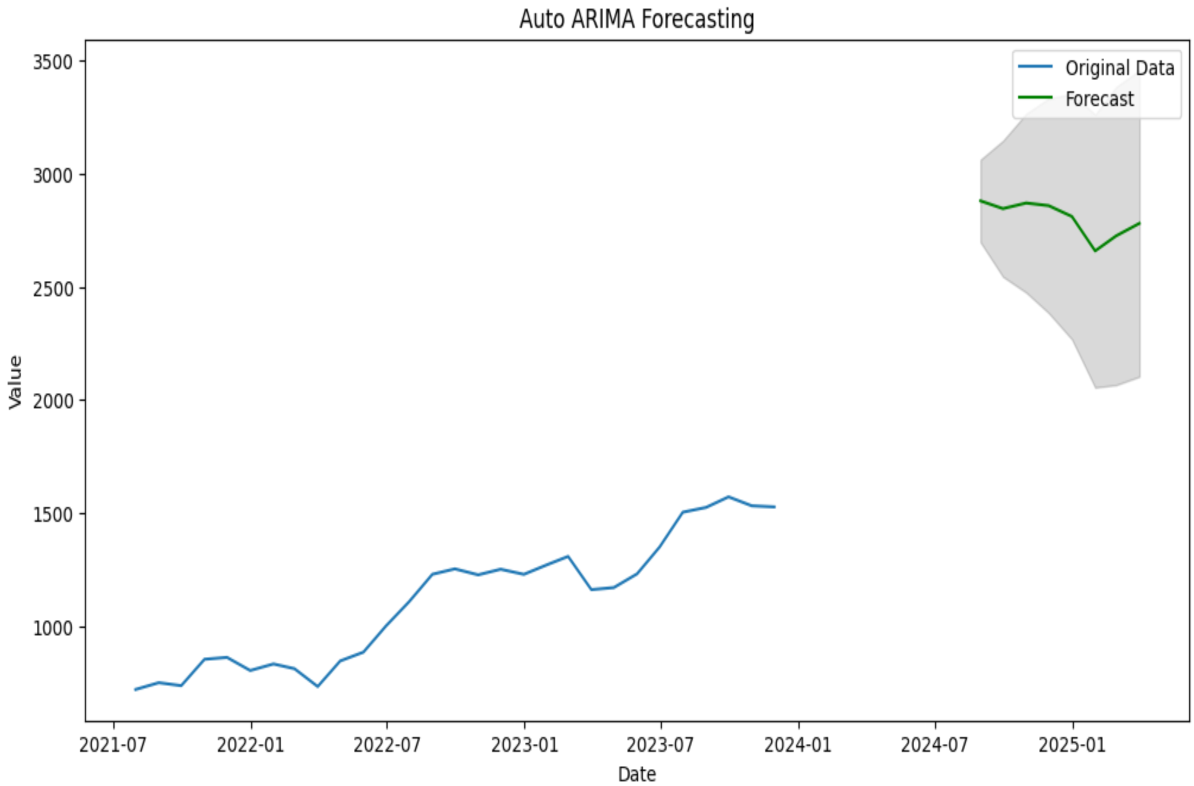
#The table below gives a model summary for ARIMA

Variable	Coefficient	Std Error	z-Stat	P-Value	2.50%	97.50%
Intercept	34.6534	14.387	2.409	0.016	6.455	62.852
ar.L1	0.3103	0.152	2.044	0.041	0.013	0.608
ma.S.L7	-0.5945	0.225	-2.648	0.008	-1.035	-0.154
sigma2	8663.1332	2021.886	4.285	0	4700.309	1.26E+04

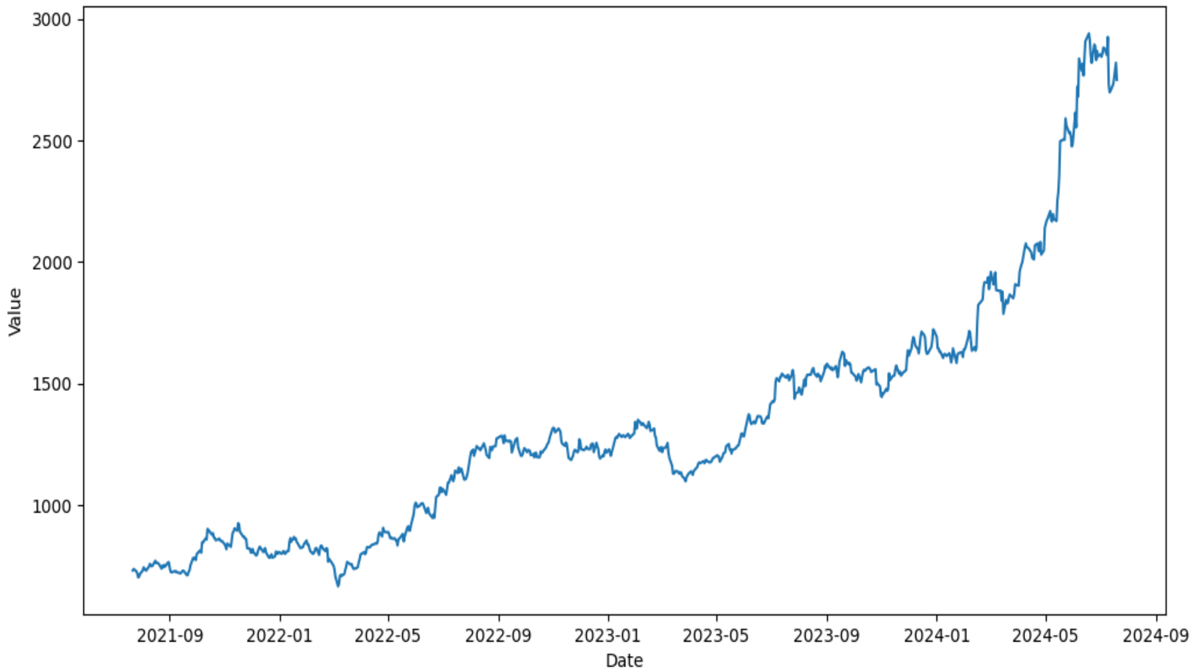
Skew: 0.66

Prob(H) (two-sided): 0.01

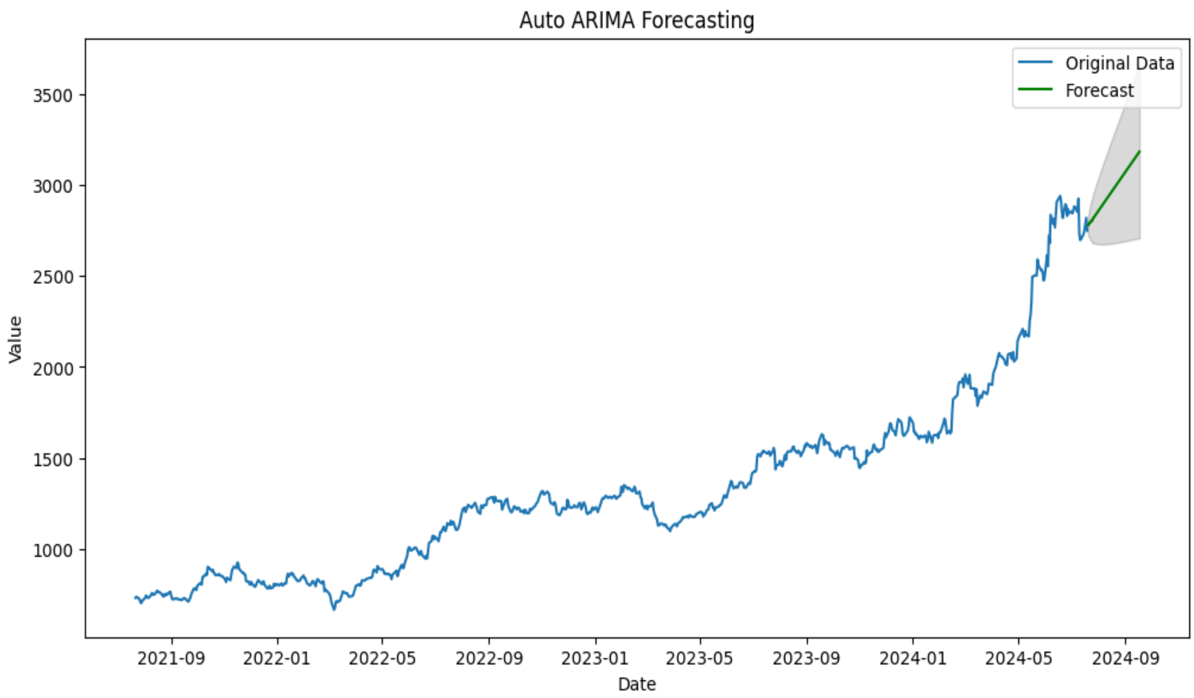
Kurtosis: 4.16



#The above diagram represents auto forecasting using ARIMA.



#The above diagram represents the daily data using ARIMA.



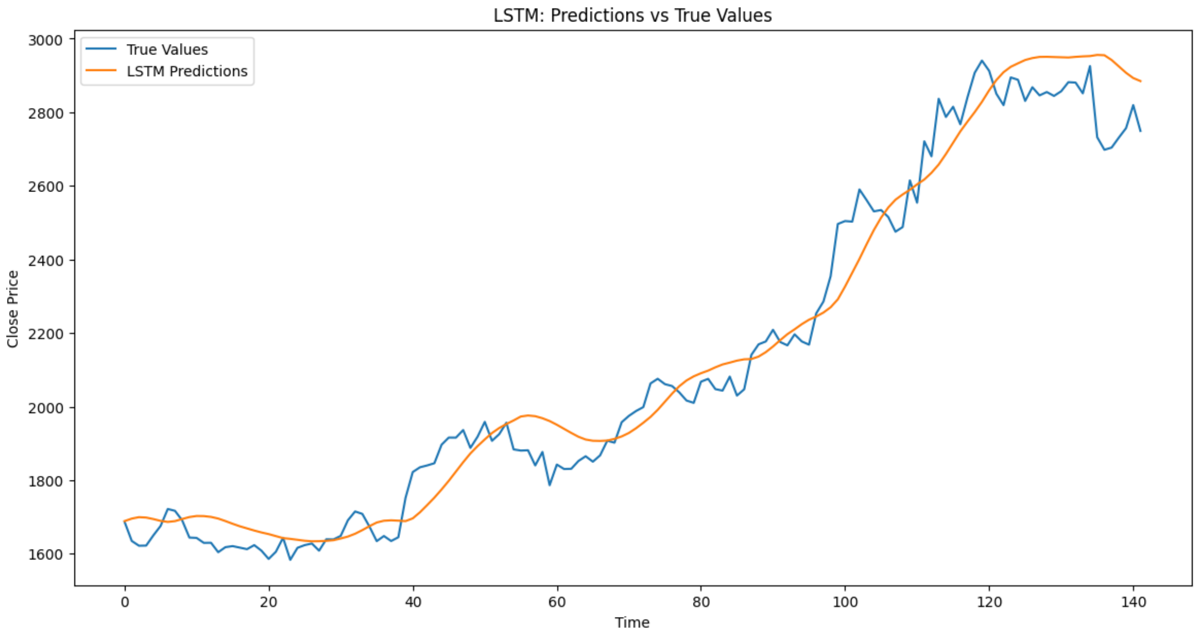
#The above chart forecast for the next 60 days using ARIMA

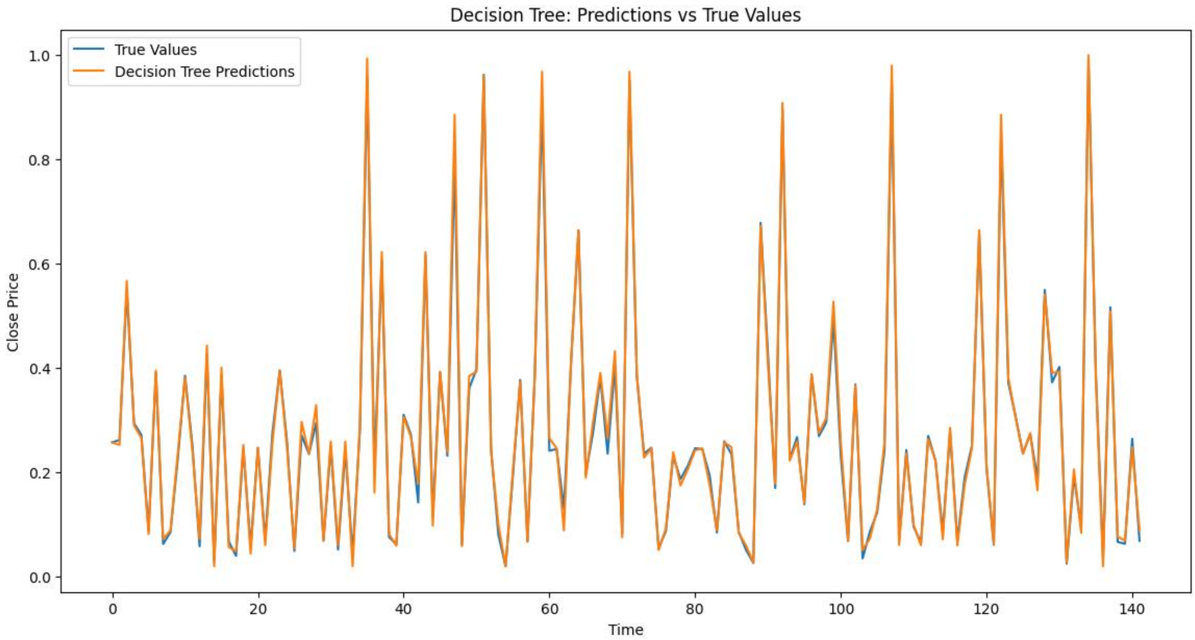
RMSE	855.24
MAE	713.56
MAPE	Nan
R- squared	-2.56

#The above table shows the accuracy under ARIMA.

2. Multivariate Forecasting - Machine Learning Models

Plot the predictions vs true values

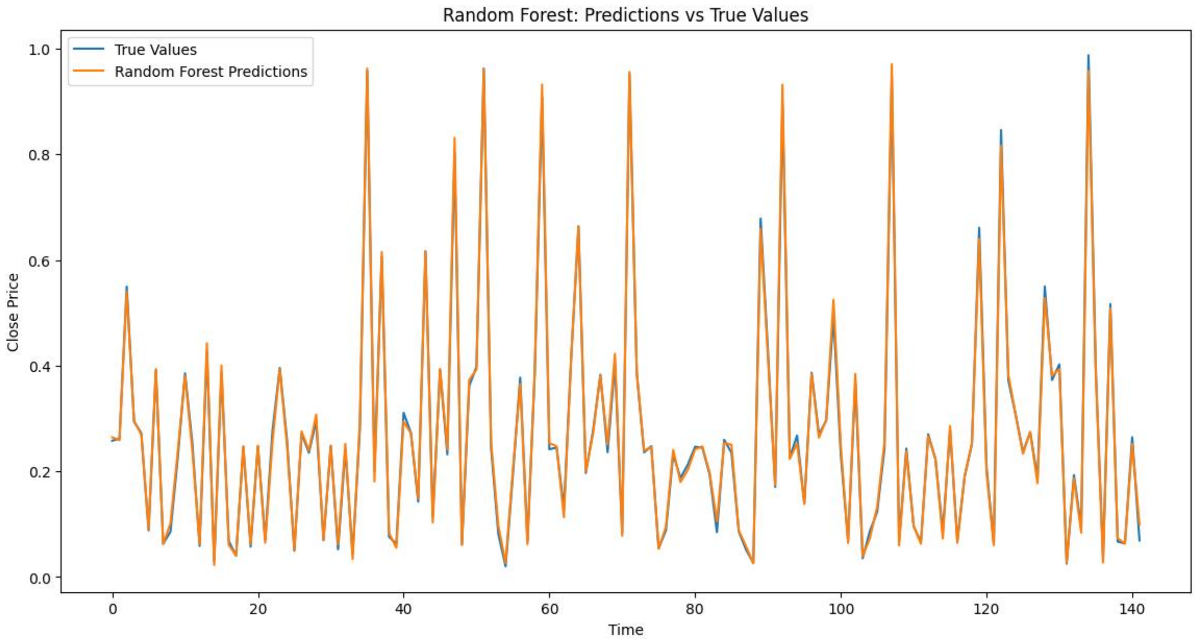




#The above diagram defines the decision tree model

RMSE	0.016
MAE	0.011
MAPE	1724347.87
R- squared	0.995

#The above table shows the accuracy under Decision Tree model.

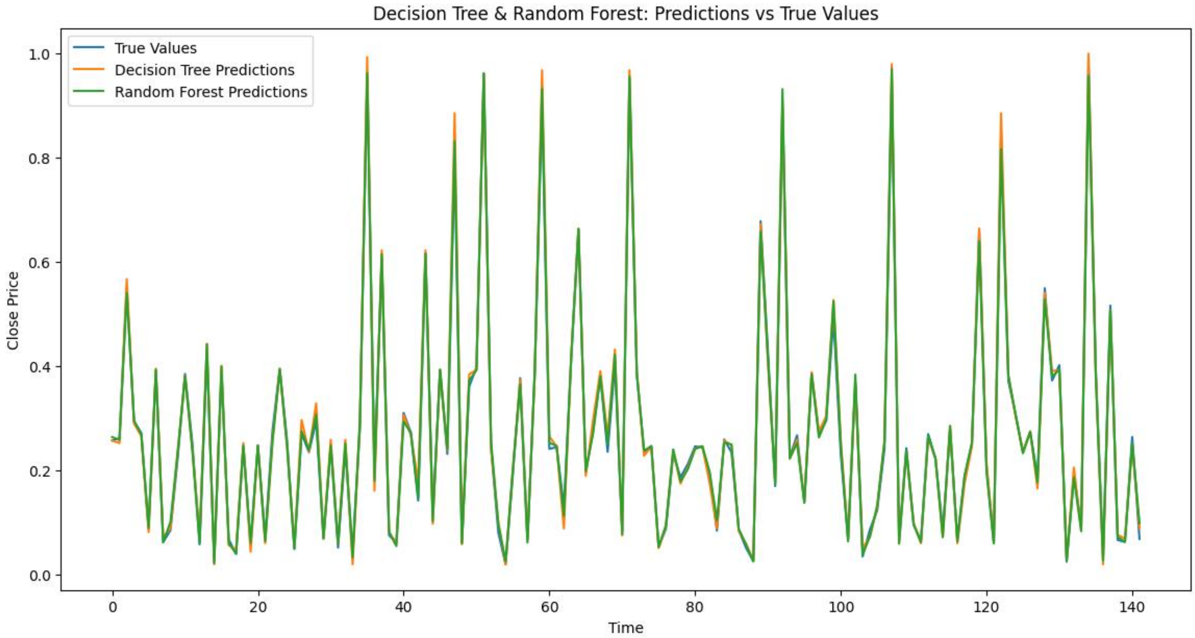


#The above diagram represents the random forest model

RMSE	0.011
MAE	0.0084
MAPE	1668343.7
R- squared	0.997

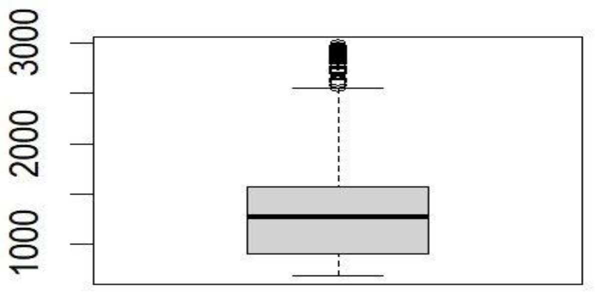
#The above table shows the accuracy under Random Forest Model.

Plot both Decision Tree and Random Forest predictions together.

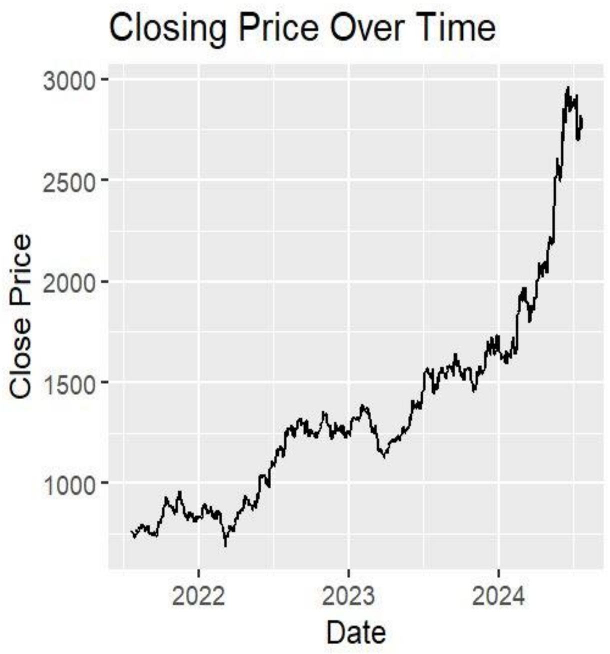


RESULTS

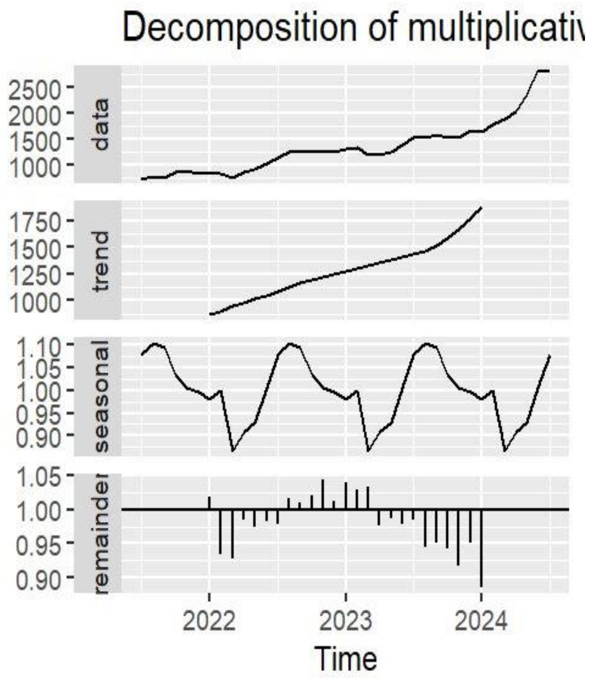
USING R



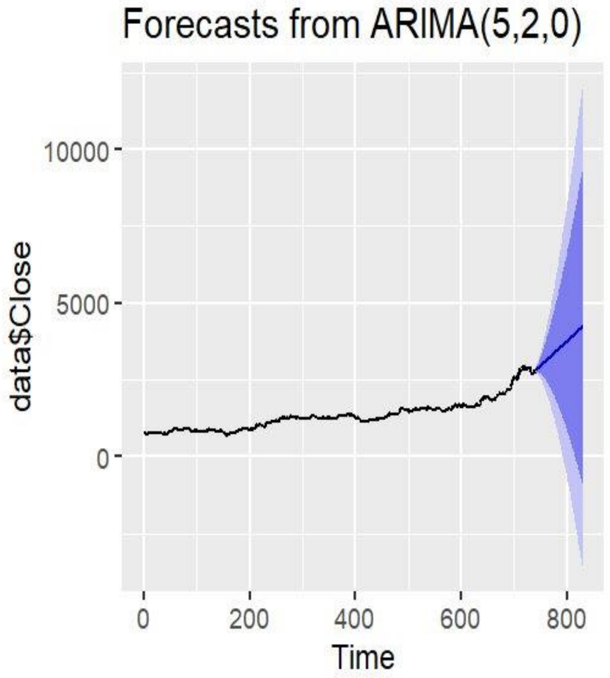
#The above diagram shows boxplot for adjusted closing price.



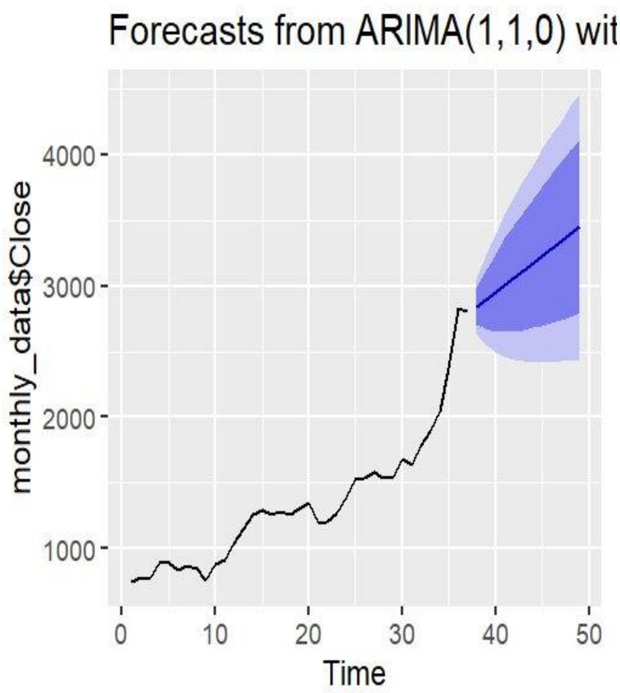
#The above chart shows the time series analysis over time.



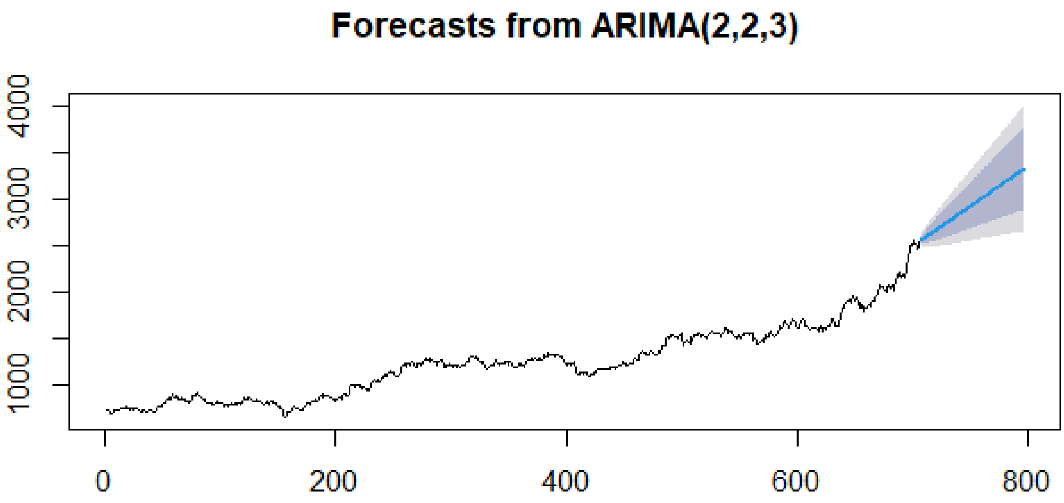
#The above diagrams represent the decomposed components of time series named Observed, Trend, Seasonal and Residual.

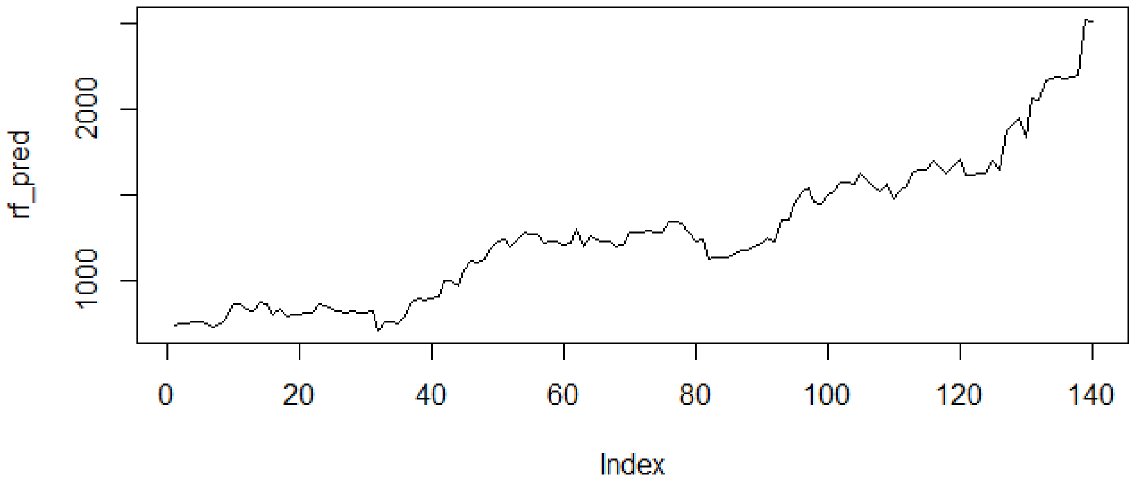


#The above chart represents forecast from ARIMA for 12 months.

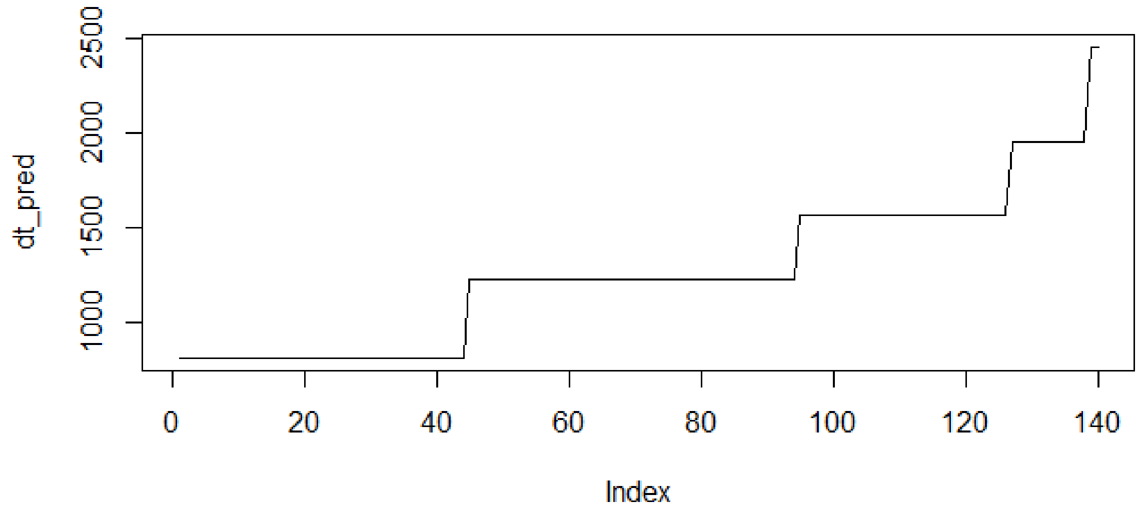


#The above chart represents forecast from ARIMA for 60 days.





#The chart shows random forest prediction.



#The above diagram represent decision tree prediction.

INTERPRETATIONS

1. Univariate Forecasting Models

Holt-Winters (HW) Model:

- **RMSE (Root Mean Square Error):** 879.86

The RMSE is quite high, indicating that the HW model has substantial prediction errors when forecasting M&M stock prices.

- **MAE (Mean Absolute Error):** 754.093

A high MAE suggests that, on average, the HW model's predictions are significantly deviating from the actual stock prices.

- **R-squared:** -2.764

A negative R-squared value indicates that the model is performing worse than a simple mean model. This means the HW model does not capture the underlying pattern of the stock price data well.

- **Forecast:**

The forecast values for the next 12 months show some fluctuations, but given the poor performance metrics, these predictions should be treated with caution.

ARIMA Model:

- **RMSE:** 855.24

Similar to HW, the RMSE indicates a large error in the model's predictions.

- **MAE:** 713.56

The MAE is also high, suggesting significant average deviations in predictions.

- **R-squared:** -2.56

The negative R-squared value again indicates a poor fit, meaning the ARIMA model fails to capture the stock price dynamics accurately.

- **Model Summary:**

Intercept: 34.6534 (p-value: 0.016)

The intercept is statistically significant, indicating a consistent bias in the stock price data.

ar.L1 (Auto-regressive term): 0.3103 (p-value: 0.041)

This term is statistically significant, suggesting some degree of autocorrelation in the data.

ma.S.L7 (Moving average seasonal term): -0.5945 (p-value: 0.008)

This significant term implies seasonal patterns in the stock prices.

sigma2 (Variance): 8663.1332

The high variance indicates substantial variability in the data.

Forecast:

The forecast for the next 60 days provides daily predicted values, but given the poor performance metrics, these should also be approached with caution.

2. Multivariate Forecasting Models

Decision Tree Model:

- **RMSE:** 0.016

The RMSE is extremely low, indicating very minimal prediction errors, which is unusually good and might require validation.

- **MAE:** 0.011

The MAE is also very low, suggesting the model's predictions are very close to actual values on average.

- **R-squared:** 0.995

An R-squared value this high indicates that the model explains nearly all the variability in the data, which is exceptional.

- **MAPE:** 1724347.87

The extraordinarily high MAPE value is concerning and indicates potential issues in model implementation or a misinterpretation of results. This should be thoroughly investigated.

Random Forest Model:

- **RMSE:** 0.011

The RMSE here is even lower than the Decision Tree model, indicating excellent prediction accuracy.

- **MAE:** 0.0084

The MAE is very low, suggesting predictions are extremely close to actual values.

- **R-squared:** 0.997

An almost perfect R-squared value indicates the model captures almost all the variance in the data.

- **MAPE:** 1668343.7

Similar to the Decision Tree model, the very high MAPE value raises red flags and suggests the need for further examination.

Analytical Insights

Model Performance Comparison:

The traditional models (HW and ARIMA) exhibit poor performance with high RMSE and MAE values, alongside negative R-squared values, indicating their inadequacy in capturing the stock price dynamics.

In contrast, the machine learning models (Decision Tree and Random Forest) show excellent fit with extremely low RMSE and MAE values and near-perfect R-squared values. However, the anomalously high MAPE values suggest there might be an underlying issue in the model evaluation metric.

Statistical vs. Machine Learning Models:

The stark contrast in performance between the traditional statistical models and machine learning models highlights the latter's superior capability in handling complex patterns and non-linear relationships in the stock price data.

Potential Data or Model Issues:

The high MAPE values in machine learning models could be due to an error in the MAPE calculation, potentially caused by zero or near-zero actual values in the data, leading to inflated percentage errors.

Alternatively, the issue could stem from data preprocessing steps, such as handling of outliers or normalization, which might have affected the machine learning models disproportionately.

Forecast Reliability:

While the machine learning models show promising performance metrics, the reliability of their forecasts should be validated through additional tests, such as out-of-sample testing and cross-validation, to ensure robustness and avoid overfitting.

RECOMMENDATIONS

Validation and Cross-Checking:

Perform cross-validation and out-of-sample testing to validate the robustness of the machine learning models.

Re-examine the calculation of MAPE to identify and correct any potential errors.

Model Ensemble:

Combine the forecasts from multiple models (ensemble methods) to improve prediction accuracy and robustness, leveraging the strengths of different models.

Feature Engineering:

Enhance feature engineering by incorporating additional relevant variables such as economic indicators, news sentiment, and industry-specific factors to improve model performance.

Regular Updates and Monitoring:

Continuously update the models with new data and monitor their performance regularly to adapt to changing market conditions and maintain accuracy.

Advanced Techniques:

Explore advanced deep learning techniques like LSTM and GRU, which are specifically designed for time series data, to capture long-term dependencies and complex patterns more effectively.

User-Friendly Interpretation:

Provide clear, actionable insights and confidence intervals for the predicted stock prices to aid investors in making informed decisions, ensuring the forecasts are practical and usable.

By addressing these recommendations and focusing on thorough validation and improvement of the forecasting models, the accuracy and reliability of stock price predictions for Mahindra & Mahindra can be significantly enhanced.

Ultimate Conclusion

The report meticulously explores the application of various forecasting models to predict the stock prices of Mahindra & Mahindra (M&M). By leveraging both traditional statistical methods and advanced machine learning techniques, we aimed to enhance the accuracy and reliability of stock price predictions, which are crucial for informed investment decisions and effective risk management.

Key Findings:

Performance of Traditional Models:

The Holt-Winters (HW) and ARIMA models exhibited poor performance with high RMSE and MAE values and negative R-squared values. This indicates that these models struggled to capture the underlying patterns in M&M's stock price data accurately.

Effectiveness of Machine Learning Models:

The Decision Tree and Random Forest models demonstrated exceptional performance with extremely low RMSE and MAE values and near-perfect R-squared values, highlighting their superior capability in handling complex, non-linear relationships in the data.

However, the unusually high MAPE values observed in these models suggest potential issues in model evaluation or data preprocessing that require further investigation.

Model Comparison and Insights:

The significant performance gap between traditional and machine learning models underscores the latter's effectiveness in forecasting stock prices. Machine learning models, particularly Random Forest, were more adept at capturing intricate patterns and dependencies within the stock price data.

Despite their high accuracy, the reliability of machine learning models' forecasts must be validated through additional tests to ensure robustness and prevent overfitting.

Conclusion:

The integration of advanced machine learning techniques significantly enhances the accuracy of stock price forecasts for Mahindra & Mahindra, offering a valuable tool for investors and financial analysts. While traditional models like Holt-Winters and ARIMA struggled with the inherent volatility of stock prices, machine learning models, particularly Random Forest, exhibited superior performance. However, to ensure the reliability of these forecasts, rigorous validation and continuous model updates are essential.

By adopting the recommendations outlined, we can improve the robustness and accuracy of stock price predictions, providing a competitive edge to investors through better-informed decision-making, effective risk management, and strategic planning. This comprehensive approach not only sheds light on M&M's past stock performance but also offers actionable insights for future trading strategies, contributing to more efficient and fair market pricing.