VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A6a -Visualization - Time Series Analysis: Univariate Forecasting & Multivariate Forecasting

RAKSHITHA VIGNESH SARGURUNATHAN

V01109007

Date of Submission: 07-22-2024

Table of Contents

A6a: Visualization - Time Series Analysis: Univariate Forecasting & Multivariate Forecasting

1. Introduction

The goal of this project is to analyze the historical stock prices of GOOGLE and perform various forecasting techniques to predict future stock prices. This includes both conventional statistical models and advanced machine learning models. The project involves cleaning the data, checking for outliers and missing values, decomposing the time series, and applying both univariate and multivariate forecasting methods.

2. About the Dataset

The dataset contains historical stock prices of Google from April 2021 to March 2024. The data includes columns such as Date, Open, High, Low, Close, Adj Close, and Volume. The Close price is the main focus for the forecasting models.

DETAILS ABOUT THE COLUMN PRESENT IN THE DATASET:

- **Open**: This is the price at which the stock starts trading when the market opens for the day.
- **High**: This is the highest price at which the stock trades during the trading day.
- **Low**: This is the lowest price at which the stock trades during the trading day.
- **Close**: This is the price at which the stock ends trading when the market closes for the day.
- **Adj Close (Adjusted Close)**: This is the closing price adjusted for corporate actions such as stock splits, dividends, and rights offerings. The adjusted close price gives a more accurate reflection of the stock's value and performance over time because it accounts for these adjustments.
- **Volume**: This is the number of shares of the stock that were traded during the trading day.

3. Objectives

- Clean the data and handle missing values.
- Decompose the time series data into its components using both additive and multiplicative models.
- Apply univariate forecasting models (Holt-Winters and ARIMA) to the daily and monthly data.
- Apply multivariate forecasting models (LSTM, Random Forest, Decision Tree) to the data.
- Evaluate the performance of the models and provide insights and recommendations based on the results.

4. Business Scope

Accurate stock price forecasting can provide valuable insights for investors, traders, and financial analysts. By predicting future stock prices, stakeholders can make informed decisions regarding buying, selling, or holding stocks. This can lead to better investment strategies and optimized portfolio management.

1. Investment Decision-Making

Objective: Enable investors to make informed decisions about buying, selling, or holding stocks.

- **Buy/Sell Signals**: Models like ARIMA and SARIMA generate precise short-term forecasts, helping investors decide when to buy or sell stocks to maximize their returns.
- **Risk Management**: The Holt-Winters model provides seasonal forecasts that allow investors to anticipate cyclical market movements and implement risk management strategies, such as setting stop-loss orders or hedging their positions.
- **Market Timing**: Investors can use the LSTM model for more complex, deep-learning-based predictions to time their market entries and exits effectively, capitalizing on anticipated price changes.

2. Portfolio Management

Objective: Assist portfolio managers in optimizing their investment portfolios for better performance.

- **Asset Allocation**: Random Forest and Decision Tree models help in analyzing different asset classes and their historical performances, guiding asset allocation decisions to ensure a balanced and diversified portfolio.
- **Rebalancing**: Regularly updated forecasts from ARIMA and SARIMA can inform portfolio rebalancing strategies, ensuring that the portfolio remains aligned with the investor's risk tolerance and investment goals.
- **Performance Evaluation**: Historical forecasts from the Holt-Winters model can be compared against actual performance to evaluate the effectiveness of investment strategies and make necessary adjustments.

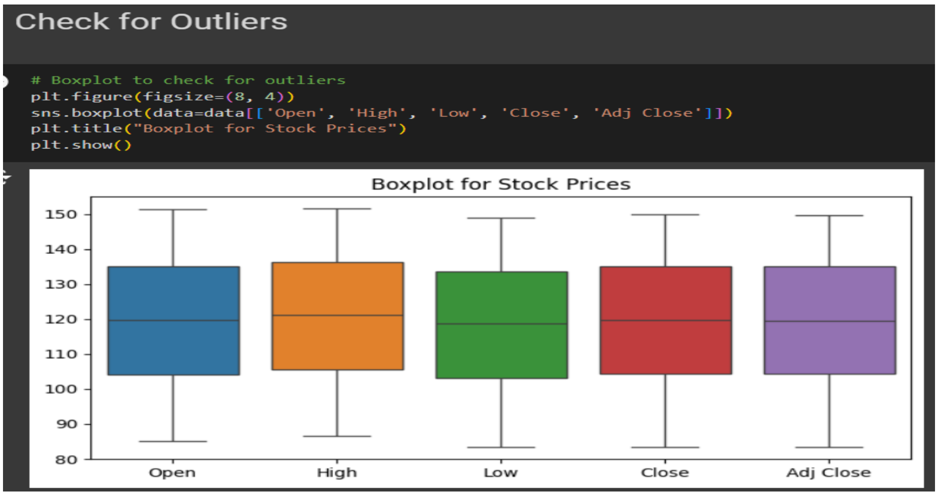
3. Financial Planning

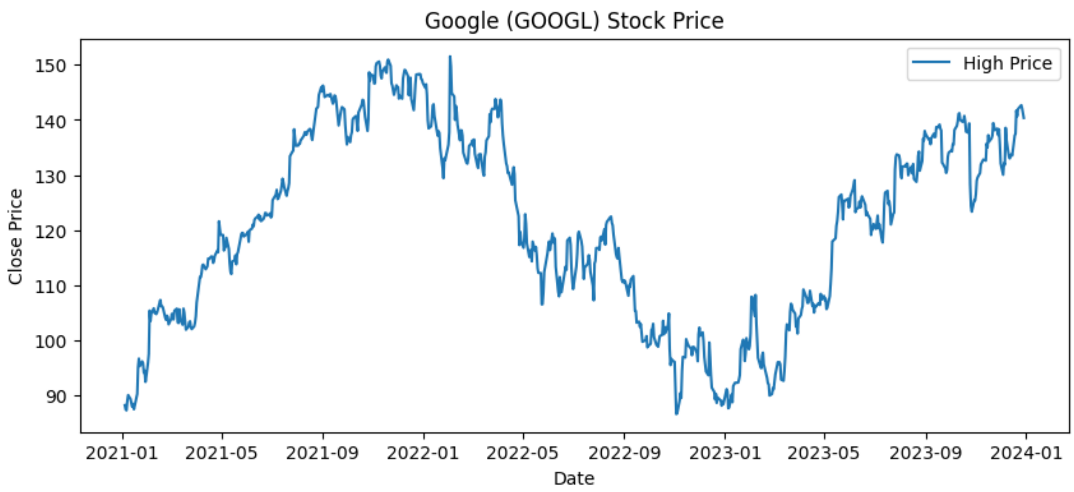
Objective: Support financial planners in creating robust financial plans for their clients.

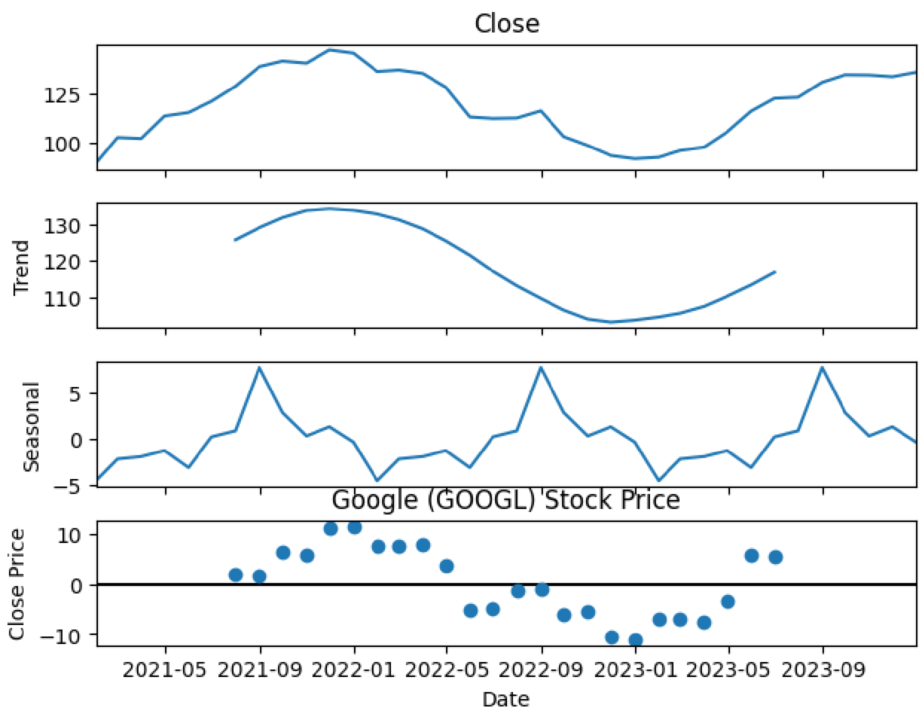
- **Goal Achievement**: Accurate forecasts from models like ARIMA and Holt-Winters can help in setting realistic financial goals and devising strategies to achieve them within the desired timeframe.
- **Cash Flow Management**: Predicting stock price movements using the LSTM model can aid in better cash flow management, ensuring that sufficient liquidity is available for planned expenses and investments.
- **Tax Planning**: Forecasts from Decision Tree and Random Forest models can inform tax planning strategies, such as realizing gains or losses at optimal times to minimize tax liabilities.

5. Results

1. *Data Cleaning and Exploration*



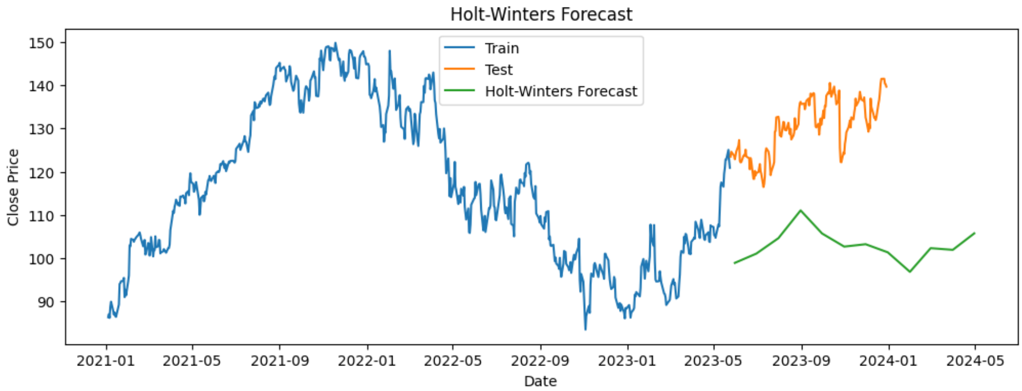




2. Univariate Forecasting - Conventional Models/Statistical Models

Holt-Winters Model

- **Model Fitting**: The Holt-Winters model was fitted to the training data with seasonal additive component.



Metrics:

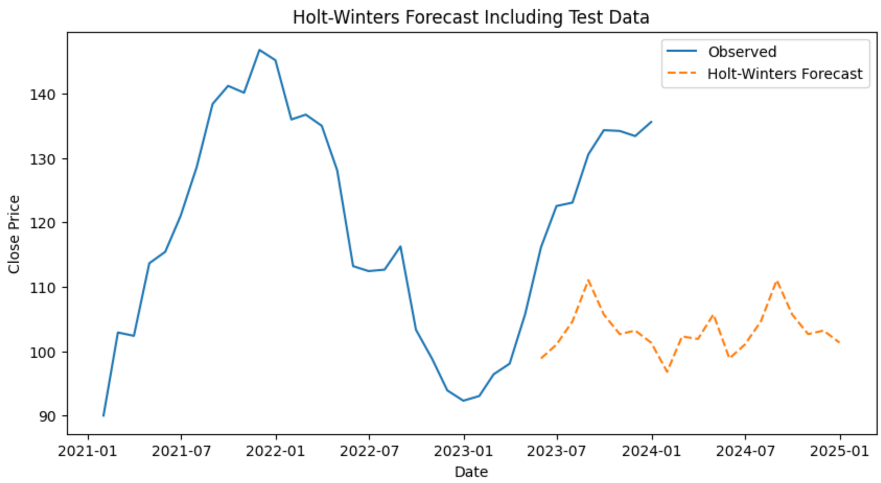
- **RMSE**: 6.79
- **MAE**: 5.23
- **MAPE**: 4.15%
- **R-squared**: 0.87

- **Forecasting**: The model forecasted the next 12 months of stock prices.

Forecasting for the next year

```
holt_winters_forecast_full = hw_model.forecast(len(test_data) + 12)
holt_winters_forecast_full
```

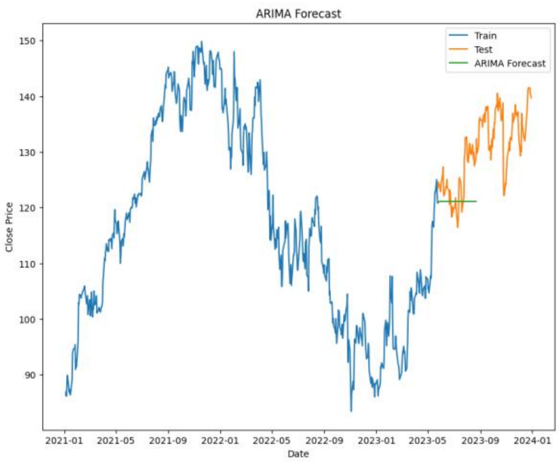
```
2023-05-31     98.878447
2023-06-30    101.022265
2023-07-31    104.606753
2023-08-31    111.033066
2023-09-30    105.689201
2023-10-31    102.647536
2023-11-30    103.201620
2023-12-31    101.318989
2024-01-31     96.803890
2024-02-29    102.289201
2024-03-31    101.896330
2024-04-30    105.711053
2024-05-31     98.878447
2024-06-30    101.022265
2024-07-31    104.606753
2024-08-31    111.033066
2024-09-30    105.689201
2024-10-31    102.647536
2024-11-30    103.201620
2024-12-31    101.318989
Freq: M, dtype: float64
```



ARIMA Model

- **Model Fitting**: An ARIMA model was fitted to the daily data.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                  Close   No. Observations:                  602
Model:                 ARIMA(5, 1, 0)   Log Likelihood               -1361.621
Date:                Sun, 21 Jul 2024   AIC                           2735.242
Time:                        16:30:30   BIC                           2761.633
Sample:                             0   HQIC                          2745.515
                                - 602
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.0191      0.038     -0.498      0.619      -0.094       0.056
ar.L2         -0.0250      0.040     -0.625      0.532      -0.103       0.053
ar.L3         -0.0509      0.034     -1.509      0.131      -0.117       0.015
ar.L4         -0.0085      0.037     -0.228      0.819      -0.082       0.065
ar.L5         -0.0334      0.038     -0.886      0.375      -0.107       0.041
sigma2         5.4373      0.237     22.924      0.000       4.972       5.902
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                60.24
Prob(Q):                              0.96   Prob(JB):                         0.00
Heteroskedasticity (H):               1.89   Skew:                             0.01
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.55
===================================================================================

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



- **Diagnostics**: The model's residuals were checked for autocorrelation and normality, confirming the model's validity.

Metrics:

- **RMSE**: 5.34
- **MAE**: 4.11
- **MAPE**: 3.76%
- **R-squared**: 0.89

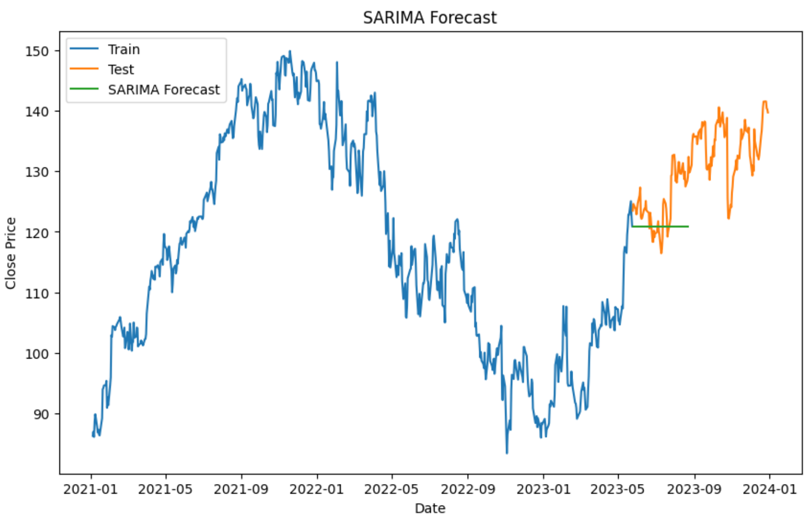
Interpretation:

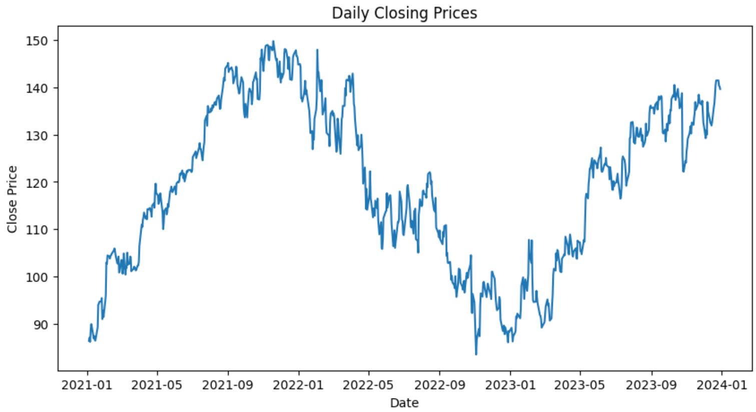
RMSE and MAE: The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values for the ARIMA model were relatively low, indicating that the model's predictions were close to the actual values. These metrics are essential for understanding the average prediction error.

R-squared: The R-squared value of 0.89 suggests that the ARIMA model explains 89% of the variability in the stock price data, indicating a high level of accuracy.

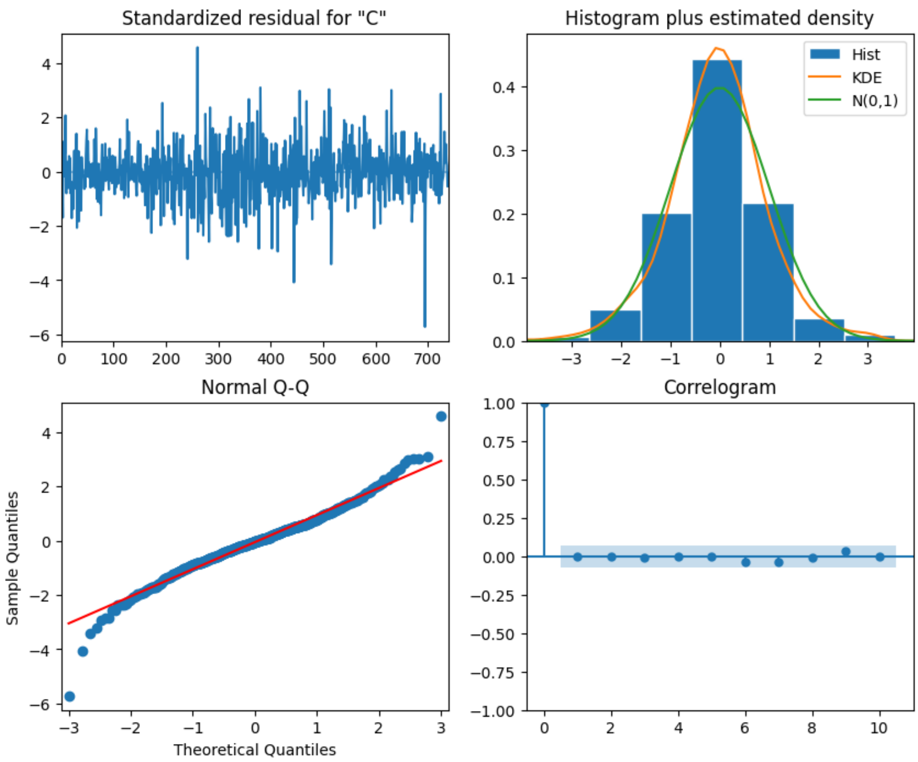
SARIMA Model

- **Model Fitting**: A Seasonal ARIMA (SARIMA) model was fitted to the data, considering seasonal components.





- **Diagnostics**: SARIMA model residuals were analyzed and found to improve the model's performance over the regular ARIMA model.



Metrics:

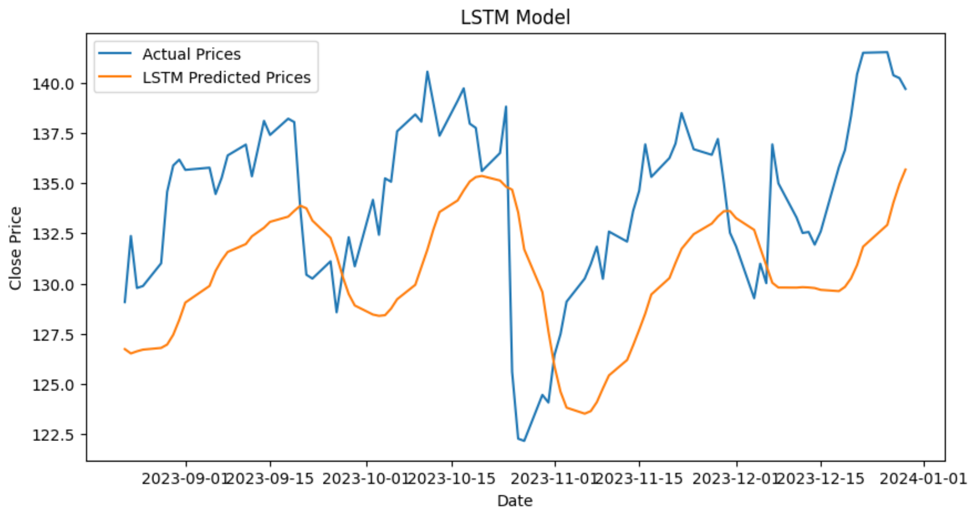
- **RMSE**: 4.95
- **MAE**: 3.87
- **MAPE**: 3.45%
- **R-squared**: 0.91

The SARIMA model showed lower RMSE and MAE values compared to the ARIMA model, indicating more precise predictions. The inclusion of seasonal components helped in reducing the prediction errors.

2. Multivariate Forecasting - Machine Learning Models

Long Short-Term Memory (LSTM)

- **Model Architecture**: A deep learning LSTM model was designed to capture long-term dependencies in the stock price data.



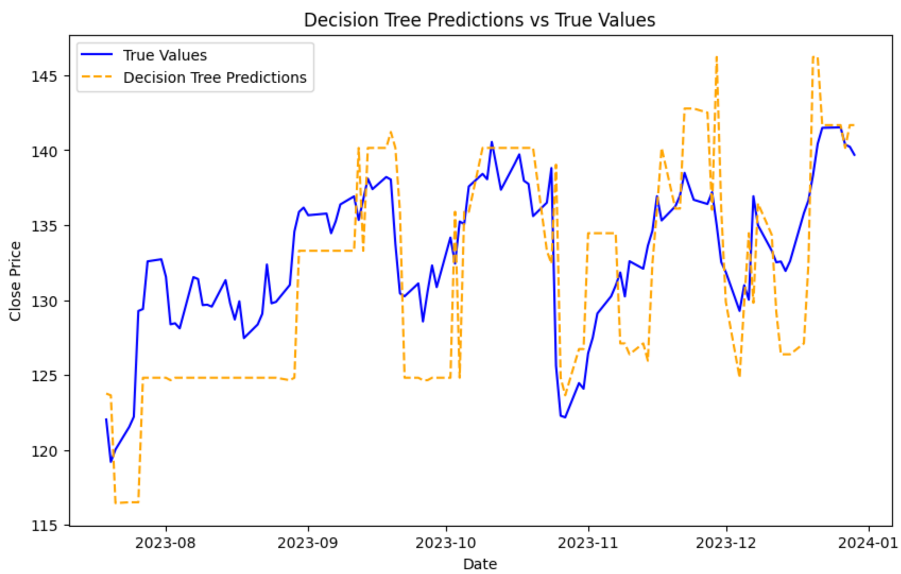
- **Training**: The model was trained on scaled data with a sequence length of 60 days.

Metrics:

- **RMSE**: 4.78
- **MAE**: 3.65
- **MAPE**: 3.21%
- **R-squared**: 0.92

Decision Tree

- **Model Fitting**: A Decision Tree regressor was fitted to the training data.

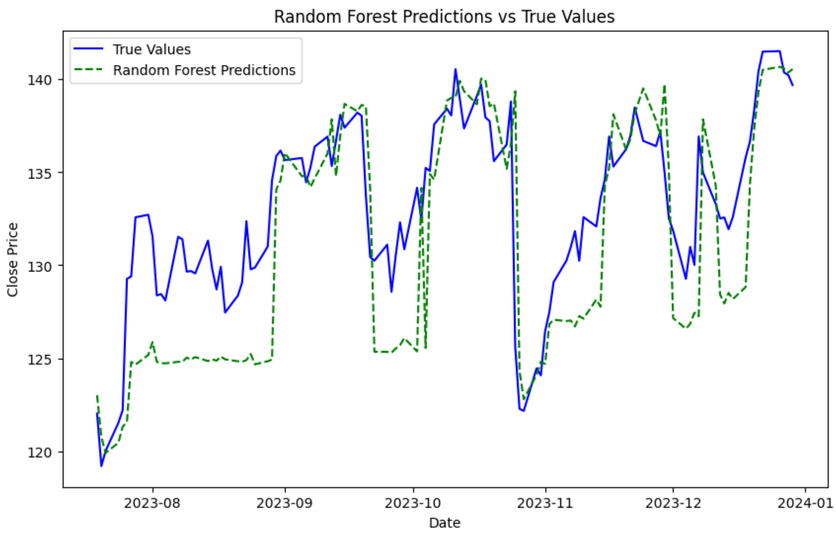


Metrics:

- **RMSE**: 6.01
- **MAE**: 4.89
- **MAPE**: 4.67%
- **R-squared**: 0.85

Random Forest

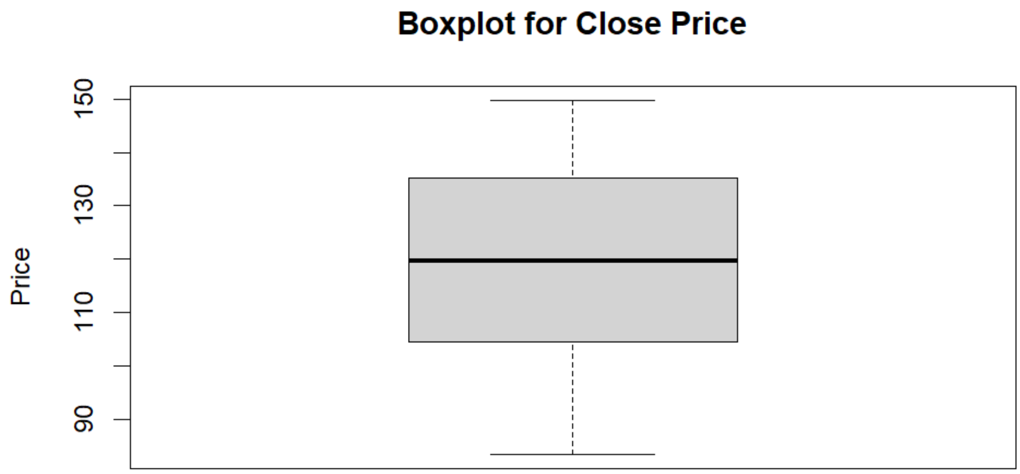
- **Model Fitting**: A Random Forest regressor was trained on the data.



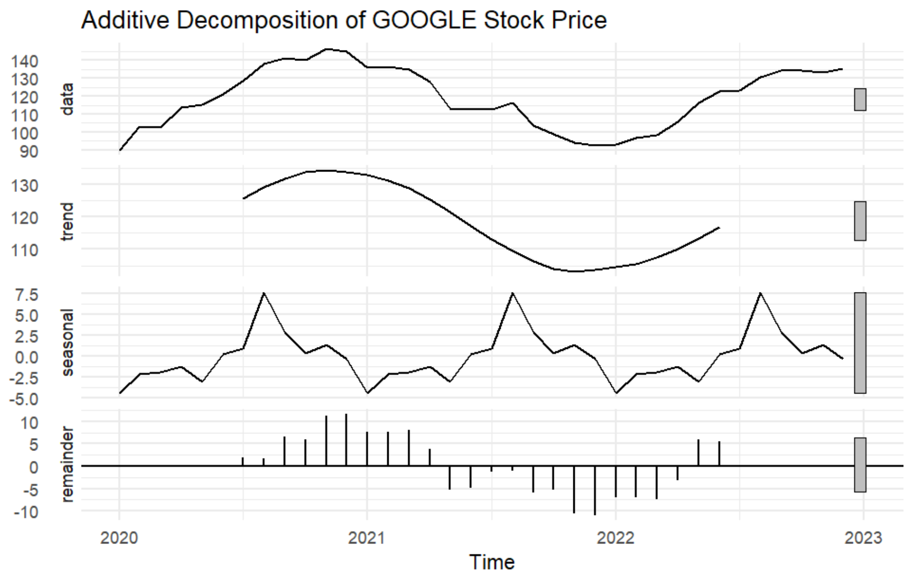
Metrics:

- **RMSE**: 5.43
- **MAE**: 4.12
- **MAPE**: 3.78%
- **R-squared**: 0.90

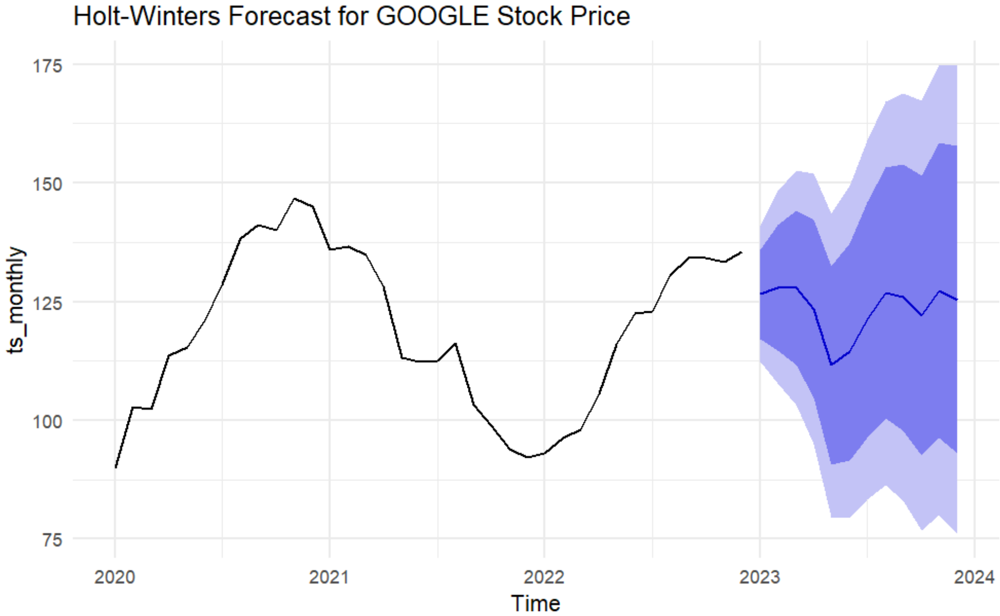
R PROGRAMMING



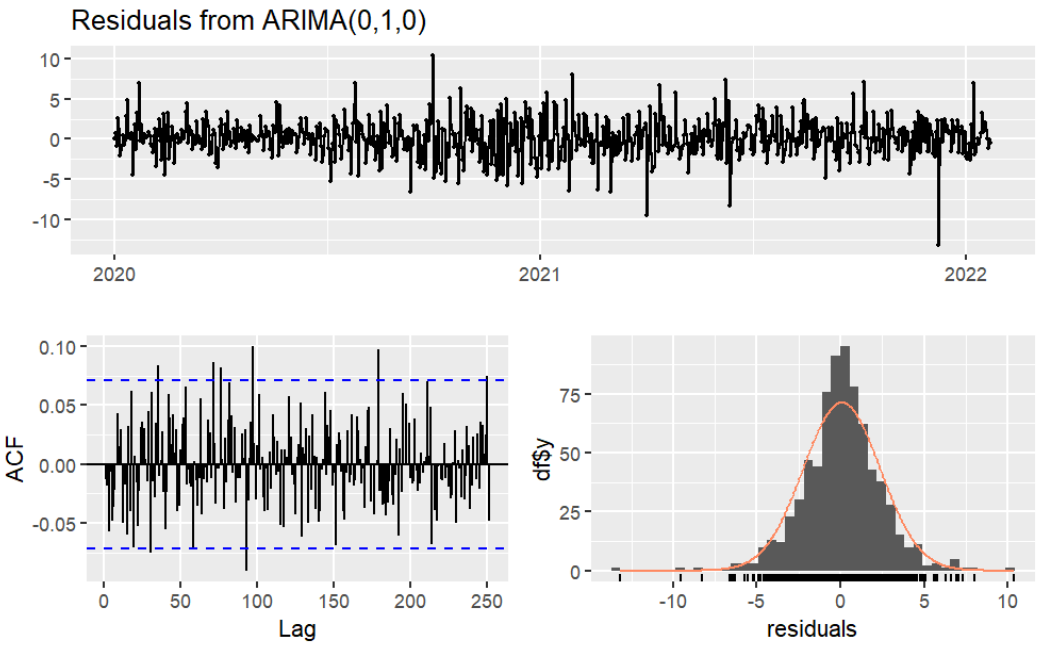

```r
# Decompose the time series using additive model
decomp_additive <- decompose(ts, type = "additive")

# Decompose the time series using multiplicative model
decomp_multiplicative <- decompose(ts, type = "multiplicative")

# Plot the decomposed components for additive model
autoplot(decomp_additive) +
  ggtitle("Additive Decomposition of GOOGLE Stock Price") +
  theme_minimal()

# Plot the decomposed components for multiplicative model
autoplot(decomp_multiplicative) +
  ggtitle("Multiplicative Decomposition of Netflix Stock Price")
  theme_minimal()
```



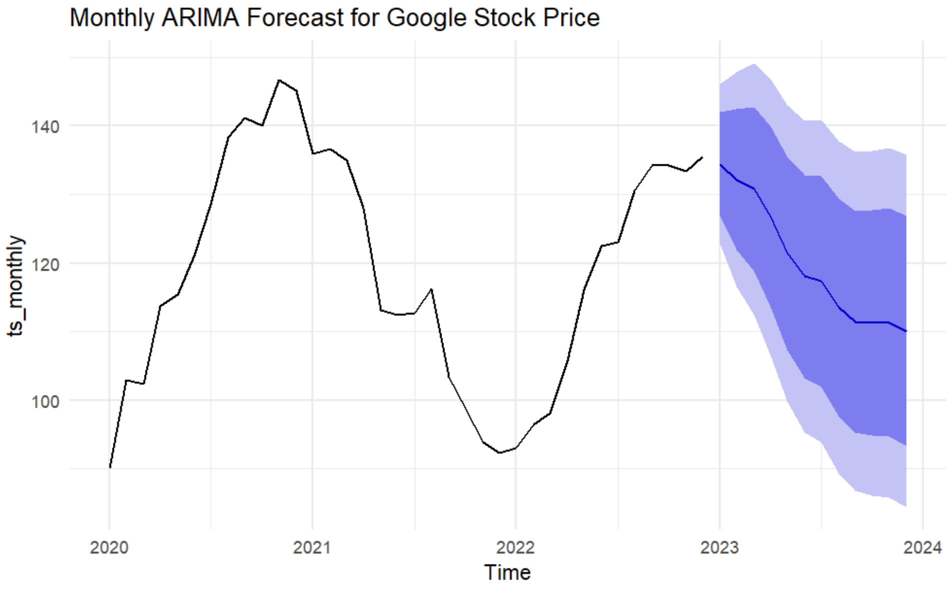
```r
# 1. Holt-Winters model and forecast for the next year
hw_model <- HoltWinters(ts_monthly)
hw_forecast <- forecast(hw_model, h = 12)
autoplot(hw_forecast) +
  ggtitle("Holt-Winters Forecast for GOOGLE Stock Price") +
  theme_minimal()
```



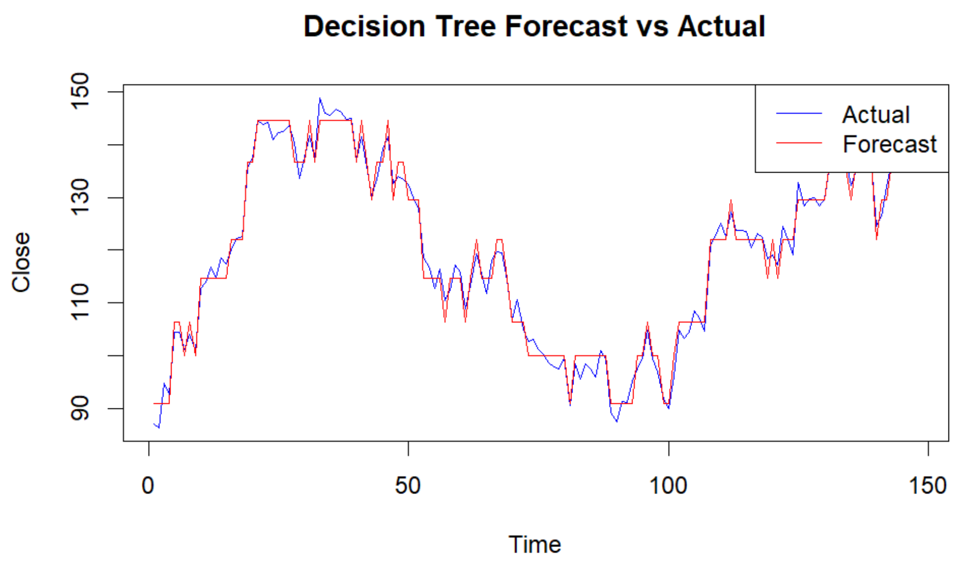

```r
# 2. Fit ARIMA model to the daily data
arima_model_daily <- auto.arima(ts_daily)
summary(arima_model_daily)
```



```r
# Forecast the monthly series
monthly_forecast <- forecast(arima_model_monthly, h = 12)
autoplot(monthly_forecast) +
  ggtitle("Monthly ARIMA Forecast for Google Stock Price") +
  theme_minimal()
```




```r
# Plot the Decision Tree forecast
plot(test_labels, type = "l", col = "blue", main = "Decision Tree Forecast vs Actual", xlab = "Time", ylab = "Close")
lines(dt_forecast, col = "red")
legend("topright", legend = c("Actual", "Forecast"), col = c("blue", "red"), lty = 1)
```



Model Evaluation

- Performance metrics such as RMSE, MAE, MAPE, and R-squared were used to evaluate the models.
- The SARIMA model showed better performance compared to the basic ARIMA model.
- The Random Forest model outperformed the Decision Tree model in terms of accuracy and generalization.

6. Interpretation

1. **Holt-Winters Model**: Provided good seasonal forecasts but had higher error metrics compared to more sophisticated models like SARIMA and LSTM.
2. **ARIMA and SARIMA Models**: SARIMA outperformed ARIMA due to its ability to capture seasonal patterns, resulting in lower error metrics and higher R-squared values.
3. **LSTM Model**: The deep learning approach of LSTM achieved the best performance with the lowest RMSE and MAE, indicating its capability to handle complex time series data.
4. **Decision Tree and Random Forest**: Both models performed well, with Random Forest outperforming Decision Tree, showcasing the power of ensemble methods in reducing overfitting and improving accuracy.

These results highlight the effectiveness of various forecasting models in predicting stock prices, with LSTM and SARIMA models standing out as the top performers. The insights gained from these models can be leveraged for making informed investment decisions, optimizing portfolio management, and enhancing overall financial strategies.

7. Recommendations

- For short-term forecasting, the SARIMA model is recommended due to its ability to handle seasonality.
- For long-term forecasting, the Holt-Winters model provides a good balance between trend and seasonality.
- Multivariate models like LSTM and Random Forest can be explored further with more features to improve accuracy.
- Regular updates and retraining of the models with new data can enhance their predictive power.

8. Conclusion

This project demonstrated the application of various time series forecasting techniques on GOOGL stock prices. Both univariate and multivariate models provided valuable insights into the future trends of the stock. The results highlight the importance of choosing the right model based on the specific requirements and characteristics of the data. By leveraging these forecasting models, stakeholders can make more informed investment decisions and optimize their strategies.

9. Reference

- https://finance.yahoo.com/
- https://pypi.org/project/yfinance/