NTK’s Stock Prediction

Nagalakshmi BM

Stock Prediction

Importance of Stock Price Prediction

- Investment Decision-Making:Enables informed choices for buying, holding, or selling stocks.
- Risk Management:Assists in assessing market risks and volatility to mitigate potential losses.
- Market Efficiency:Contributes to fair trading practices by ensuring prices reflect available information.
- Strategic Planning for Companies:Aids companies in financial planning and investment opportunities.
- Behavioral Insights:Provides understanding of investor behavior and market sentiment.

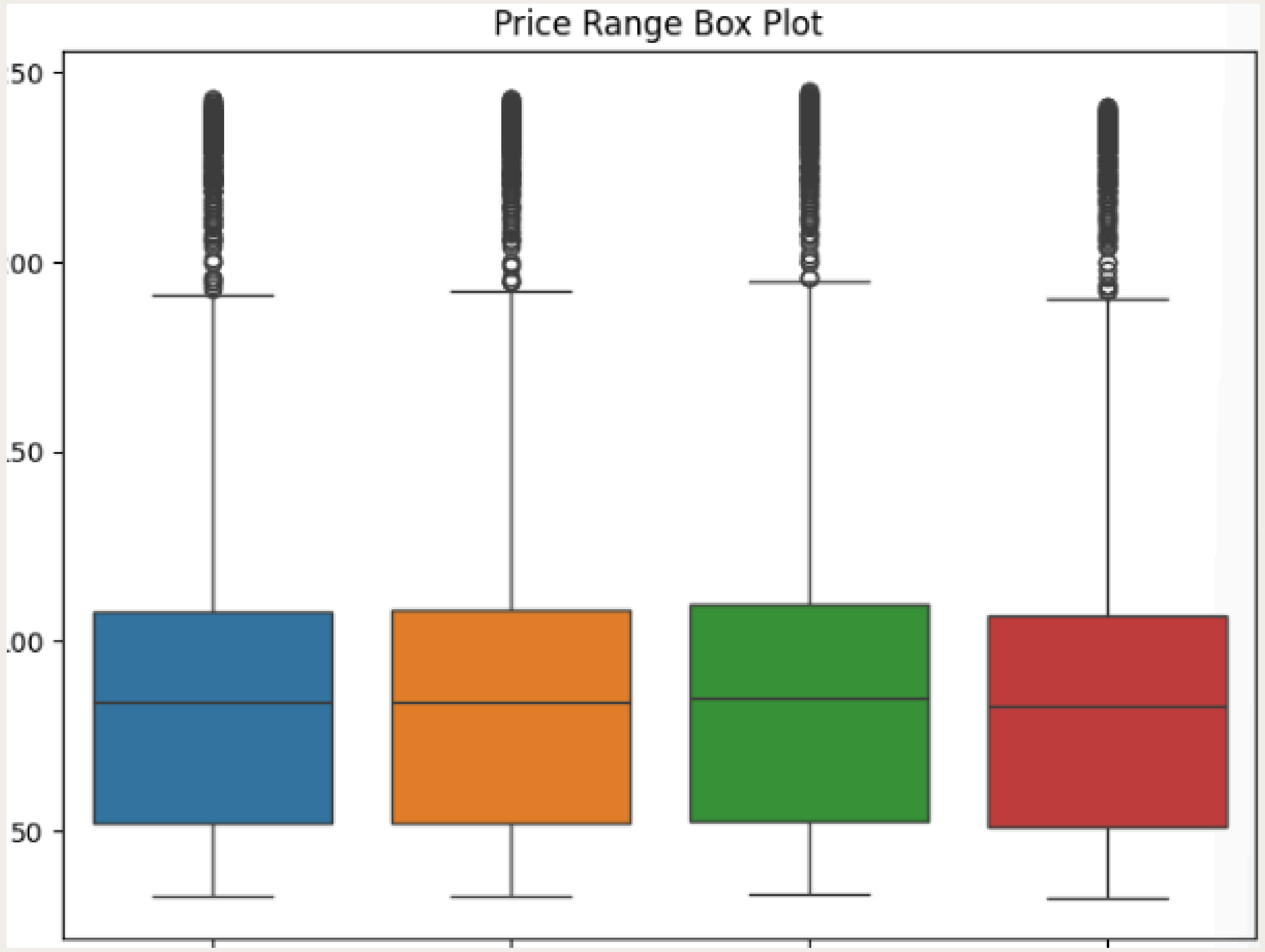
..

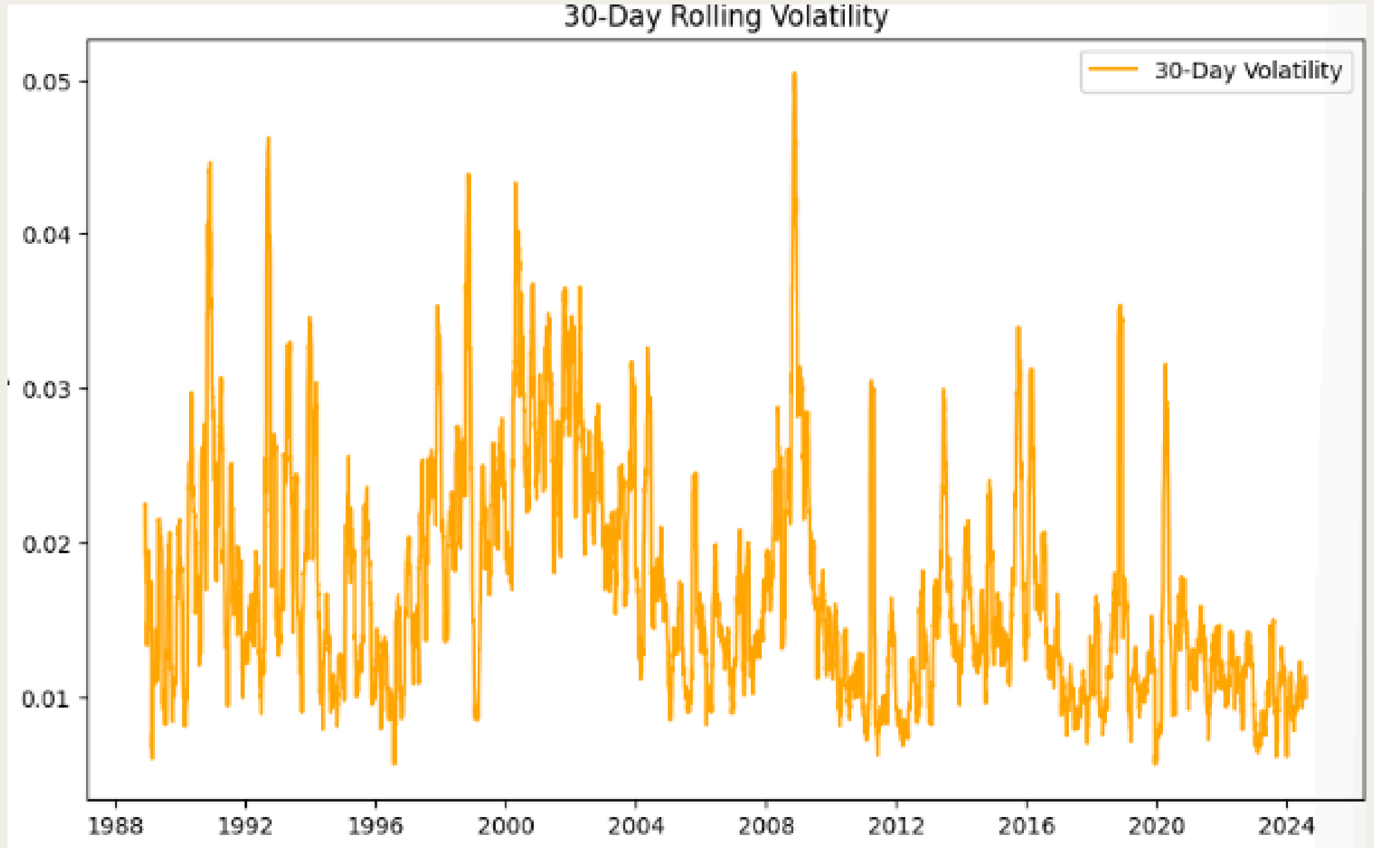
- Market Volatility:Unpredictable factors like economic indicators and political events complicate predictions.
- Data Quality and Availability:Incomplete or inaccurate data can lead to flawed models.
- Model Complexity:Balancing model complexity to avoid overfitting or underfitting is challenging.
- Feature Selection:Identifying relevant features requires domain knowledge and thorough analysis.
- Changing Market Conditions:Models must adapt to dynamic financial environments.
- Psychological Factors:Investor psychology can lead to irrational behaviors that are hard to quantify.

Challenges Stock Market

EDA and issues:

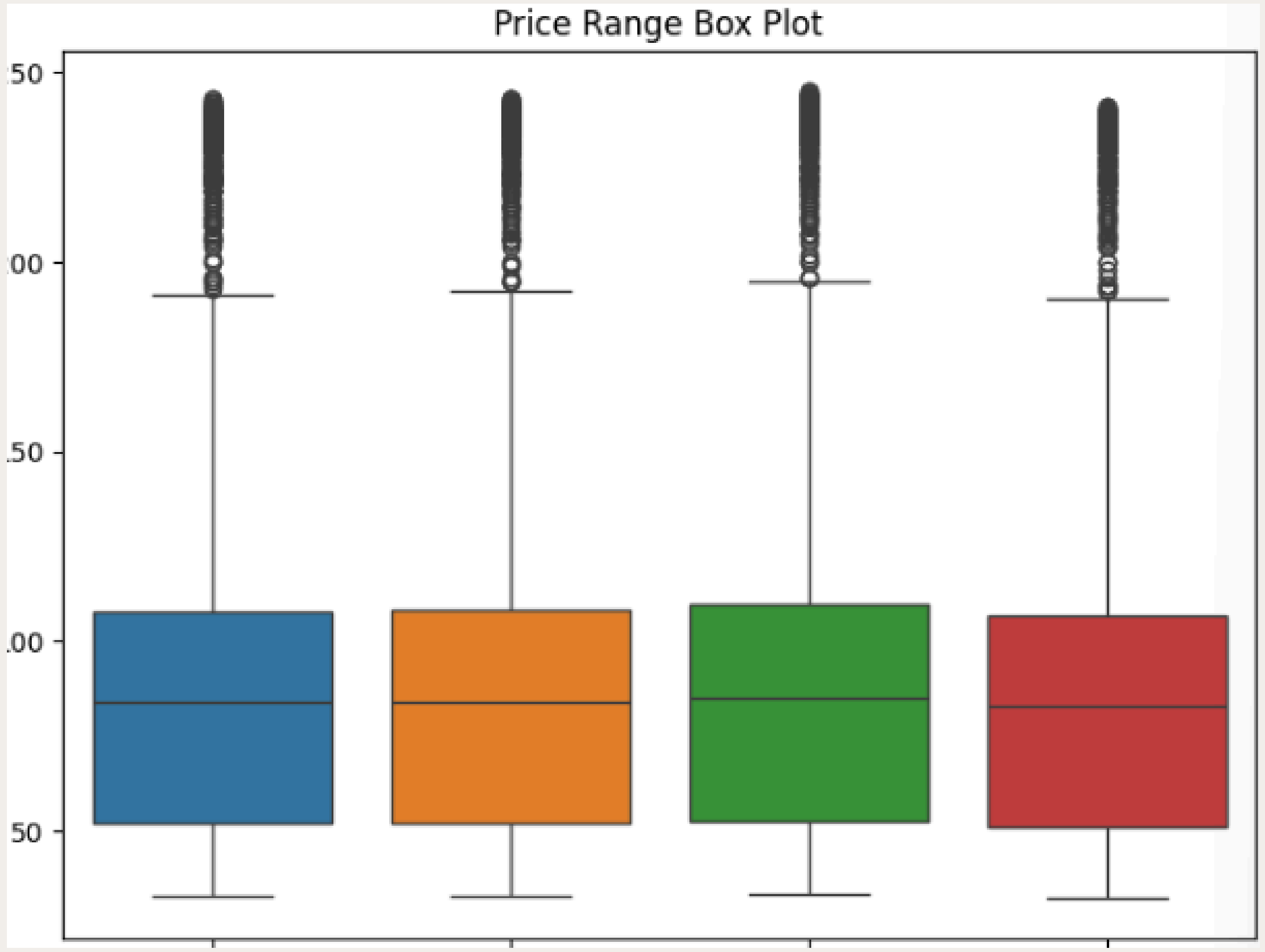
- *Data Characteristics*: The dataset consists of 9016 entries, including attributes such as closing prices, opening prices, high prices, low prices, trading volumes, and percentage change rates.
- *Volatility and Skewness:* The data exhibits **significant volatility**, characterized by a positive skewness across closing prices, opening prices, high prices, and low prices. This suggests the presence of days when stock prices spiked higher than average, a common occurrence in financial markets.
- *Outliers:* A total of **229 outliers** were identified in the dataset. These outliers can significantly impact model performance and may require further investigation or treatment.
- *Rolling Statistics:* The 30-day rolling mean and standard deviation of closing prices indicate considerable **fluctuations over time**. This reinforces the notion of volatility within the stock t.
- *Positive Skewness:* The trading volume also demonstrates positive skewness, suggesting that there are days with exceptionally **high trading activity**. Such spikes may correlate with specific market events or news that drive investor interest.

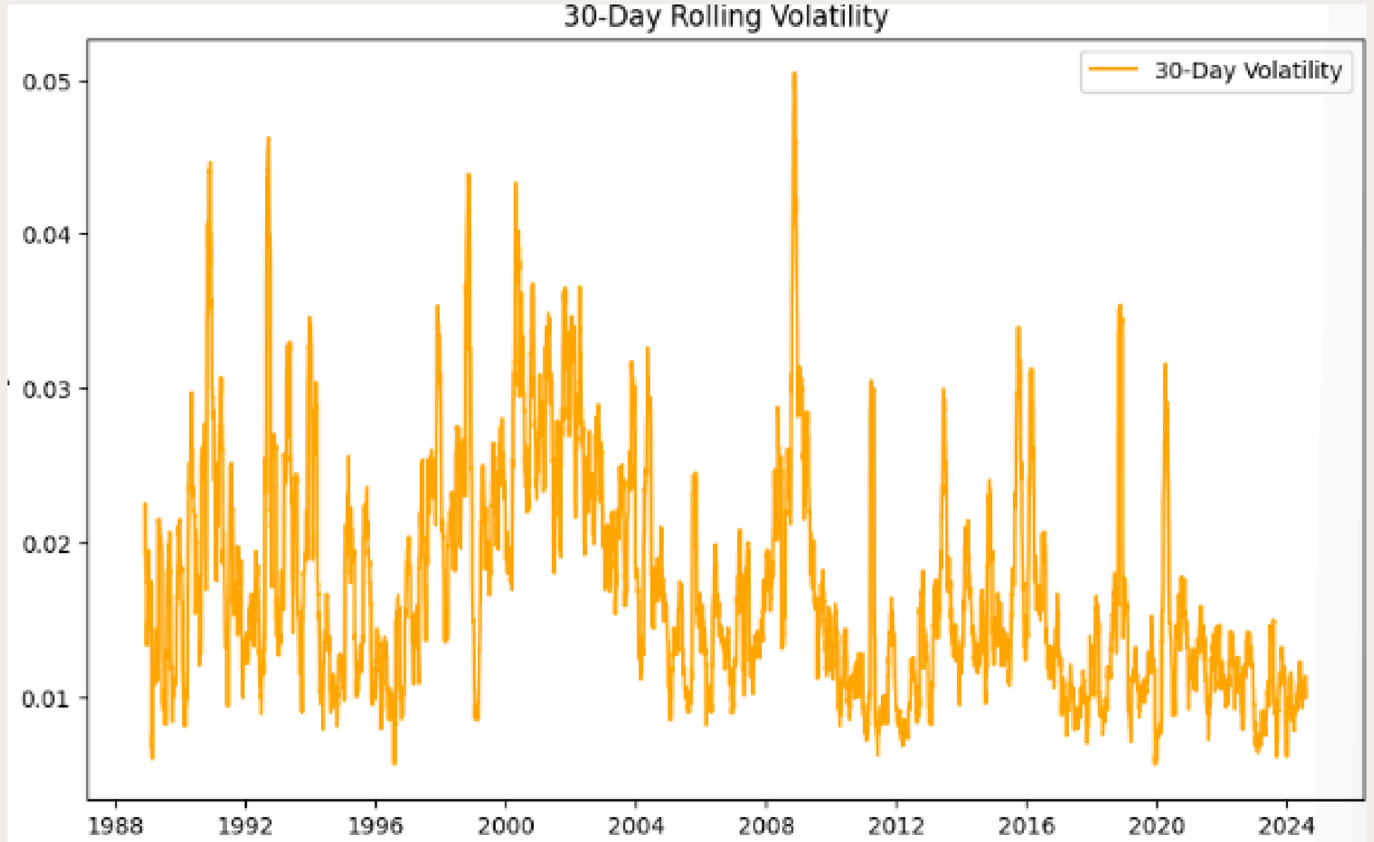


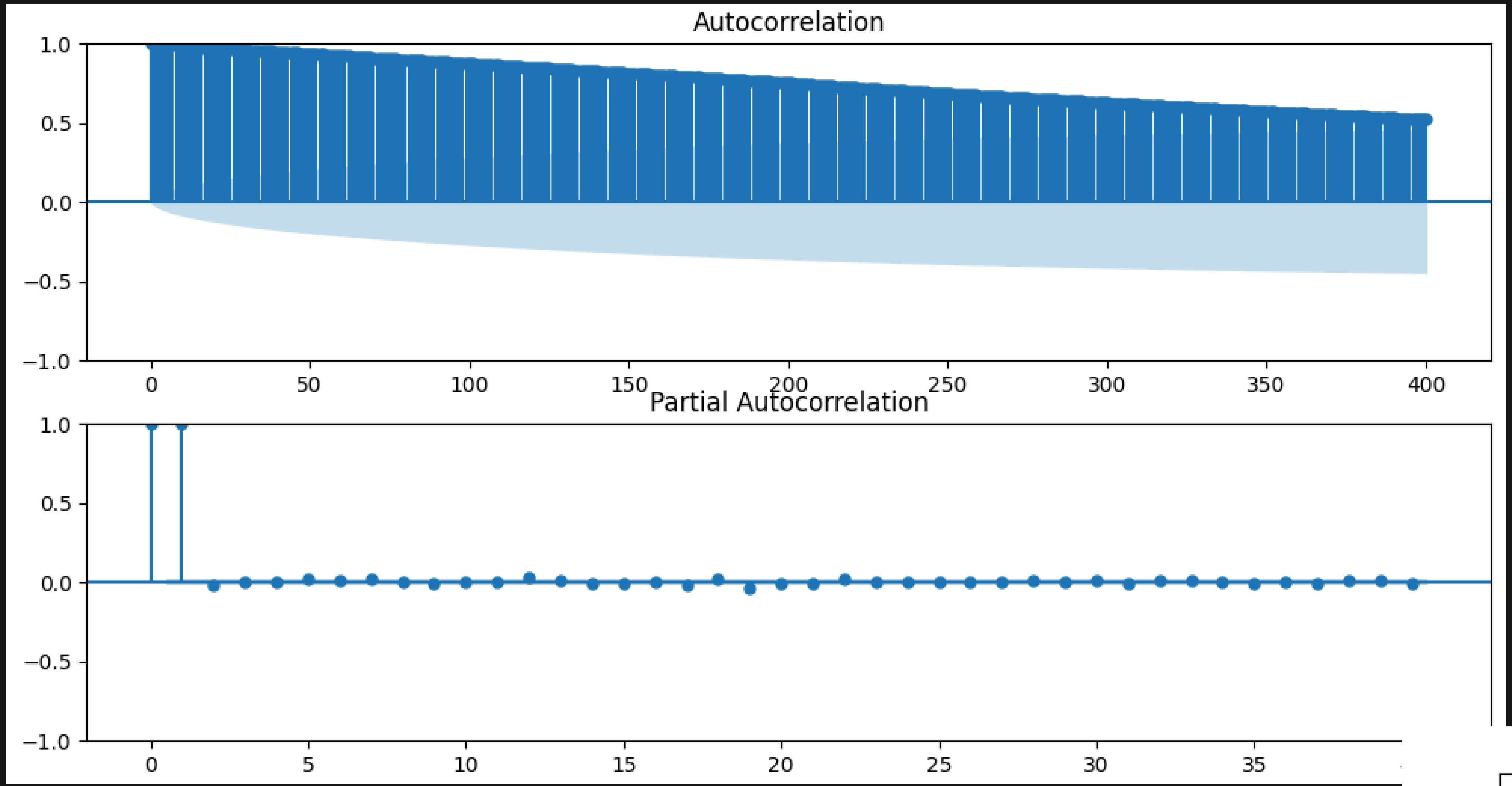


EDA and issues:

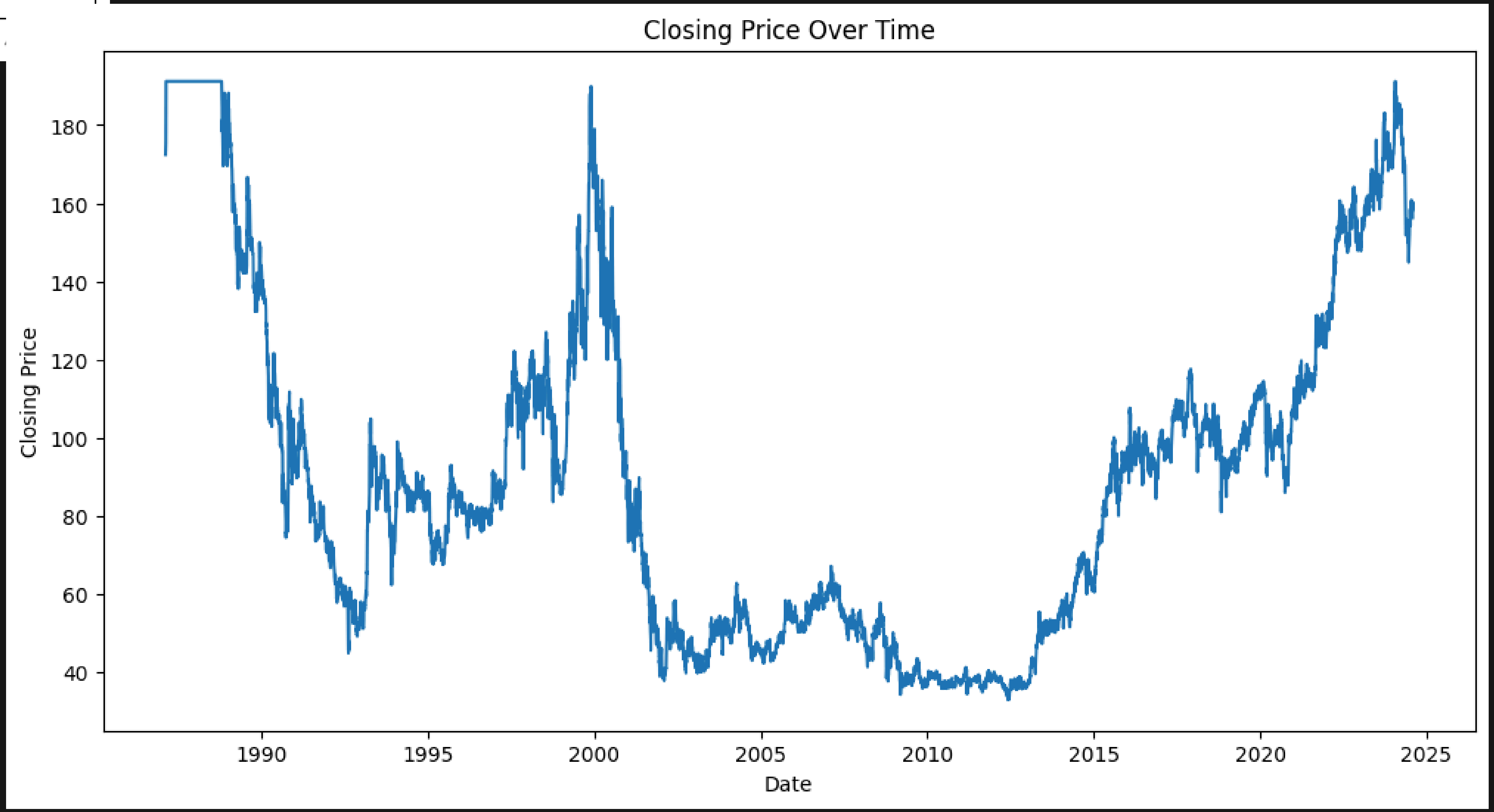
- *Data Characteristics*: The dataset consists of 9016 entries, including attributes such as closing prices, opening prices, high prices, low prices, trading volumes, and percentage change rates.
- *Volatility and Skewness:* The data exhibits **significant volatility**, characterized by a positive skewness across closing prices, opening prices, high prices, and low prices. This suggests the presence of days when stock prices spiked higher than average, a common occurrence in financial markets.
- *Outliers:* A total of **229 outliers** were identified in the dataset. These outliers can significantly impact model performance and may require further investigation or treatment.
- *Rolling Statistics:* The 30-day rolling mean and standard deviation of closing prices indicate considerable **fluctuations over time**. This reinforces the notion of volatility within the stock t.
- *Positive Skewness:* The trading volume also demonstrates positive skewness, suggesting that there are days with exceptionally **high trading activity**. Such spikes may correlate with specific market events or news that drive investor interest.







Technical Overview: ARIMA and LSTM Model Implementation



Autocorrelation Function (ACF) Analysis

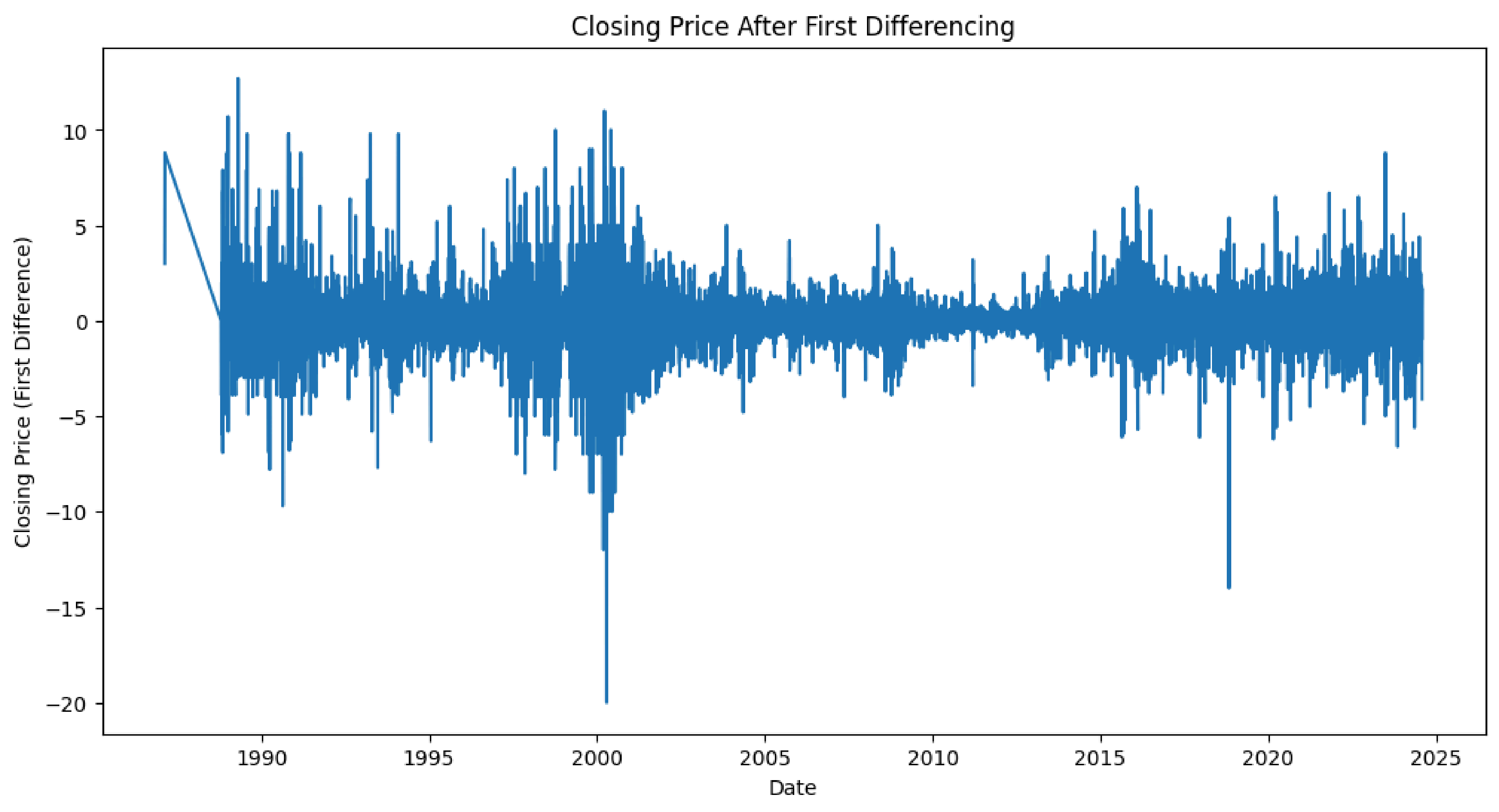
1. ACF Insights: The ACF plot indicates persistent correlation across many lags, suggesting that the closing prices are non-stationary. This behavior is typical in stock prices, which often exhibit trends over time.
    - Model Selection: The sharp cutoff at lag 2 implies that an AR(2) model could effectively capture the autoregressive behavior of the closing prices. This leads us to consider an ARIMA(2,1,0) model:2 for the autoregressive term
    - 1 for differencing to achieve stationarity
    - 0 for the moving average component

Stationarity

Augmented Dickey-Fuller (ADF) Test

- To assess the stationarity of the price series, the ADF test was performed:ADF Statistic: -2.2515
- p-value: 0.1881
  - Critical Values:1%: -3.4311
  - 5%: -2.8619
  - 10%: -2.5669

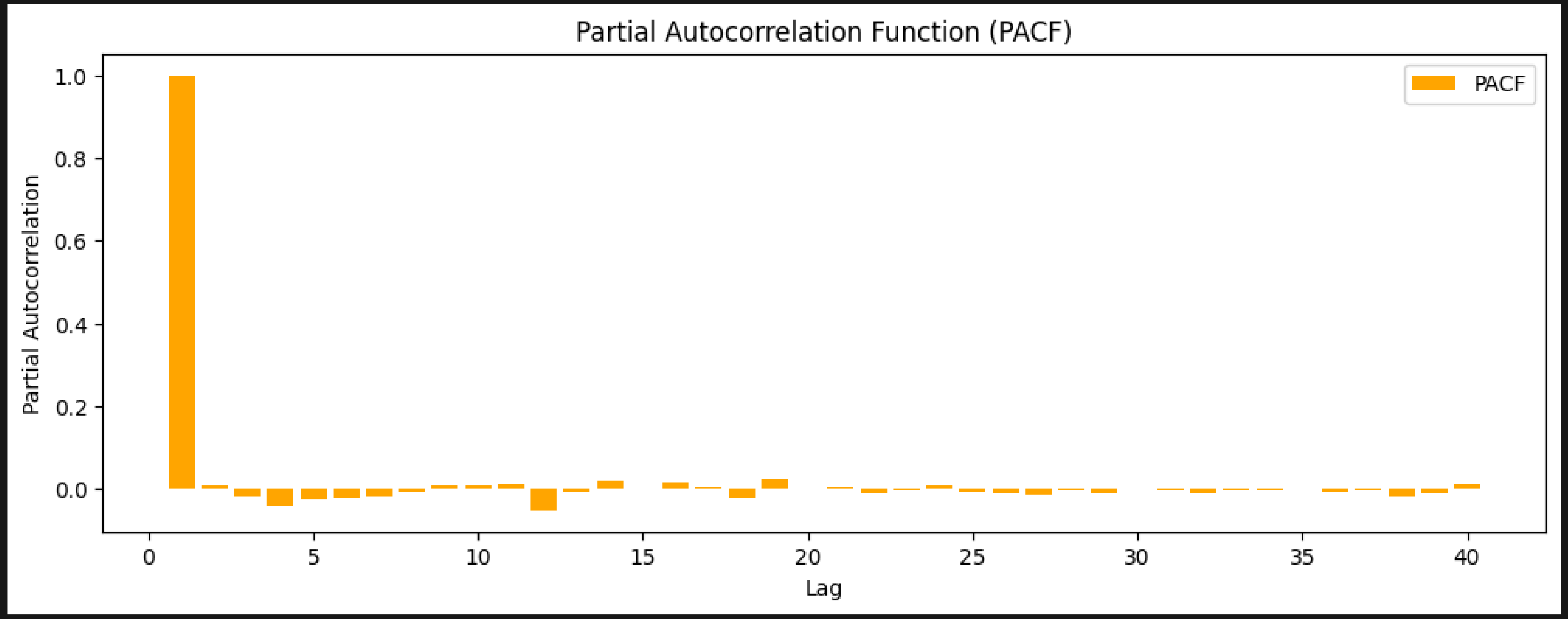
Given that the p-value (0.1881) is greater than the significance level of 0.05, we fail to reject the null hypothesis of non-stationarity. The ADF statistic does not exceed the critical values at both the 5% and 10% levels, confirming that the series is non-stationary.

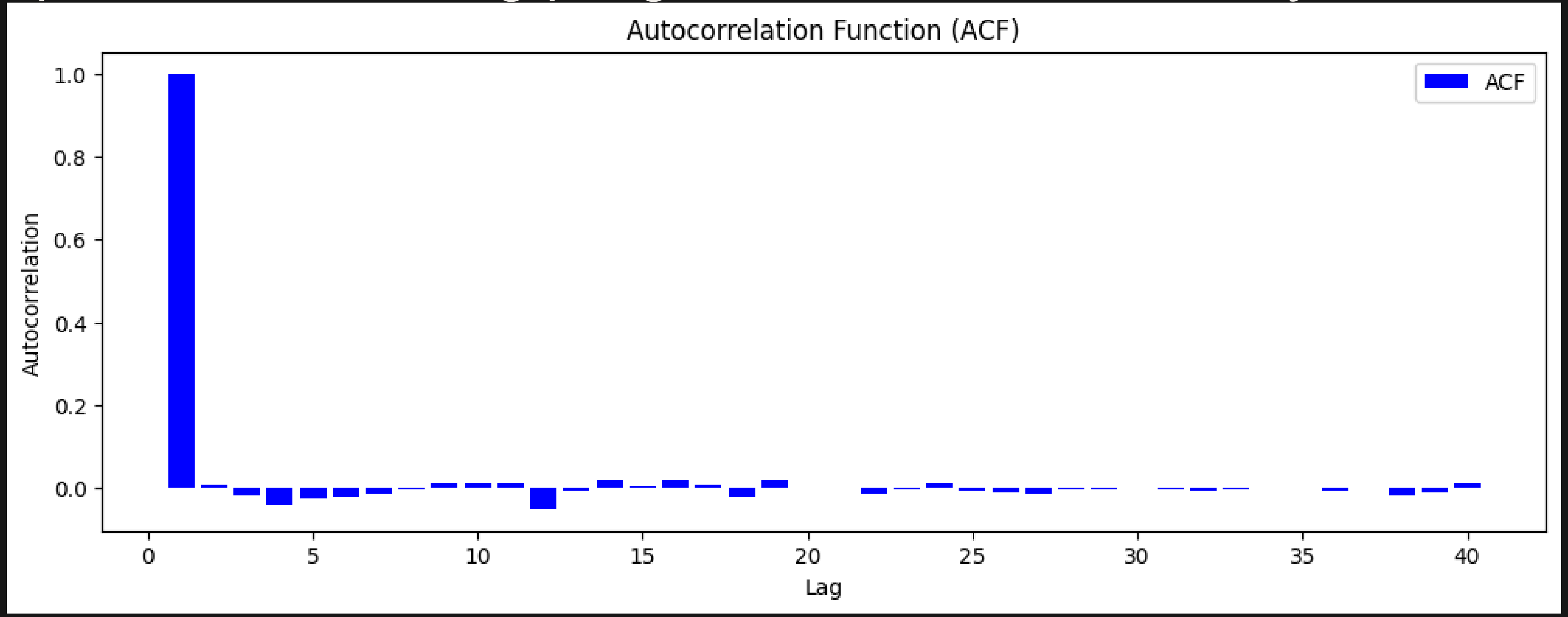


Differencing for Stationarity

To address non-stationarity, first differencing was applied

The plot of the differenced series shows a more stable pattern, indicating progress toward stationarity.





SARIMAX Model Results

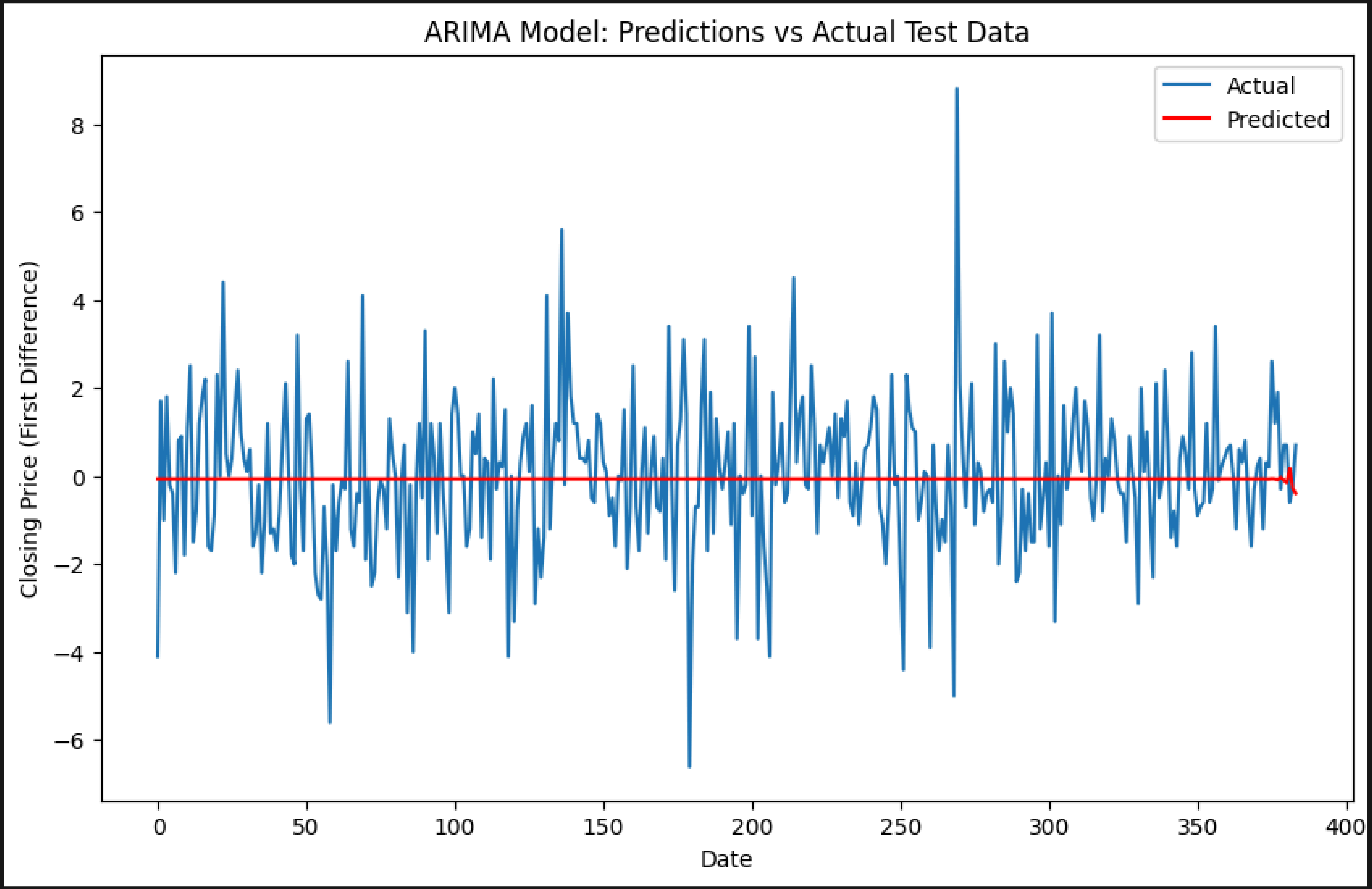
- The SARIMAX results for the ARIMA(2,1,0) model are as follows:Dependent Variable: closing_price_diff1
- No. Observations: 8400
- Log Likelihood: -17434.147
- AIC: 34874.295
- BIC: 34895.403

Coefficients:

- ar.L1: -0.6388 ($p < 0.001$)
- ar.L2: -0.3148 ($p < 0.001$)
- sigma^2: 3.7193 ($p < 0.001$)

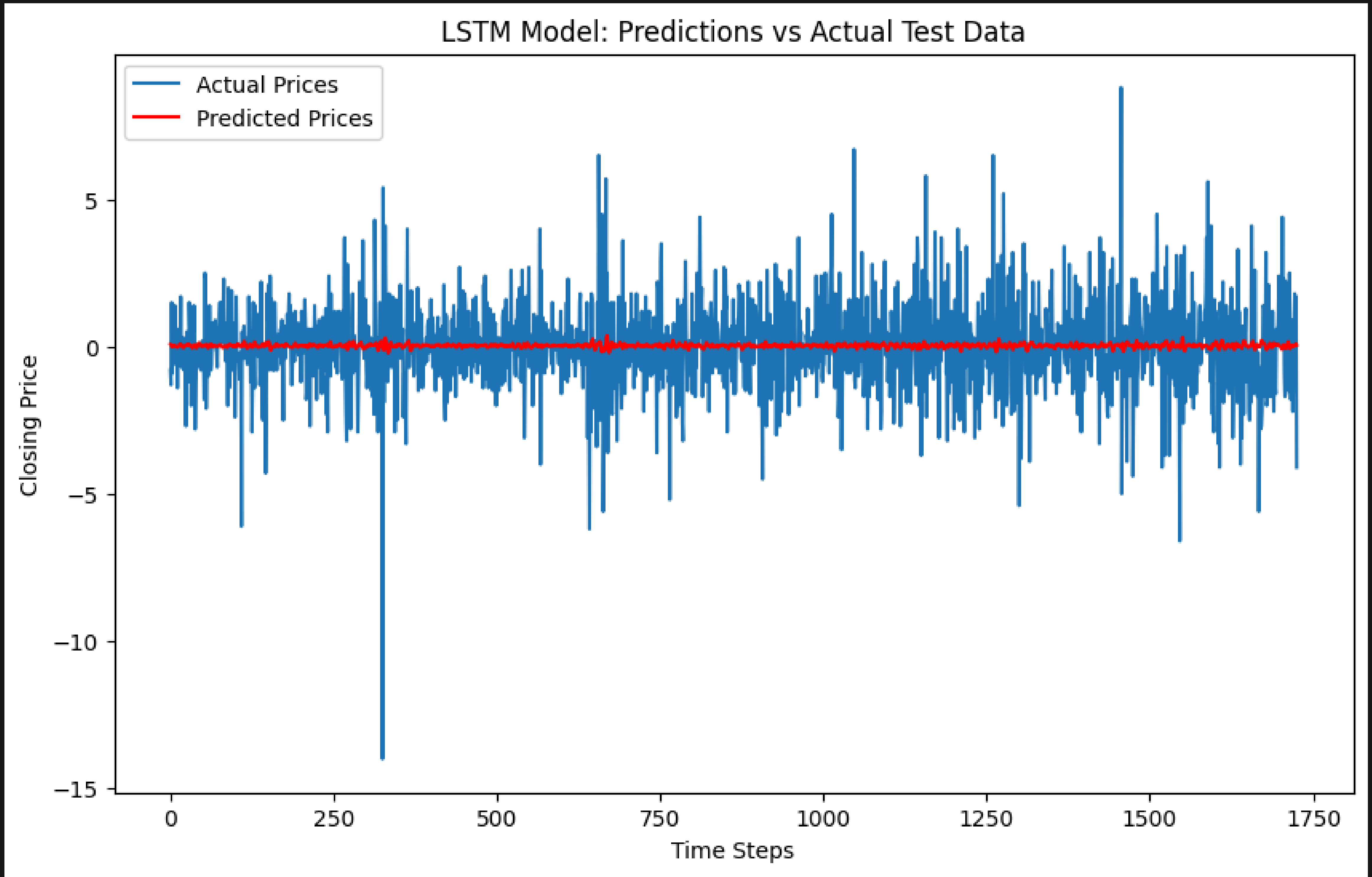
Model Diagnostics:

- Ljung-Box Test: Prob(Q): 0.00 indicates significant autocorrelation.
- Jarque-Bera Test: Prob(JB): 0.00 suggests non-normality in residuals.
- Heteroskedasticity Test: Prob(H): 0.00 indicates presence of heteroskedasticity.



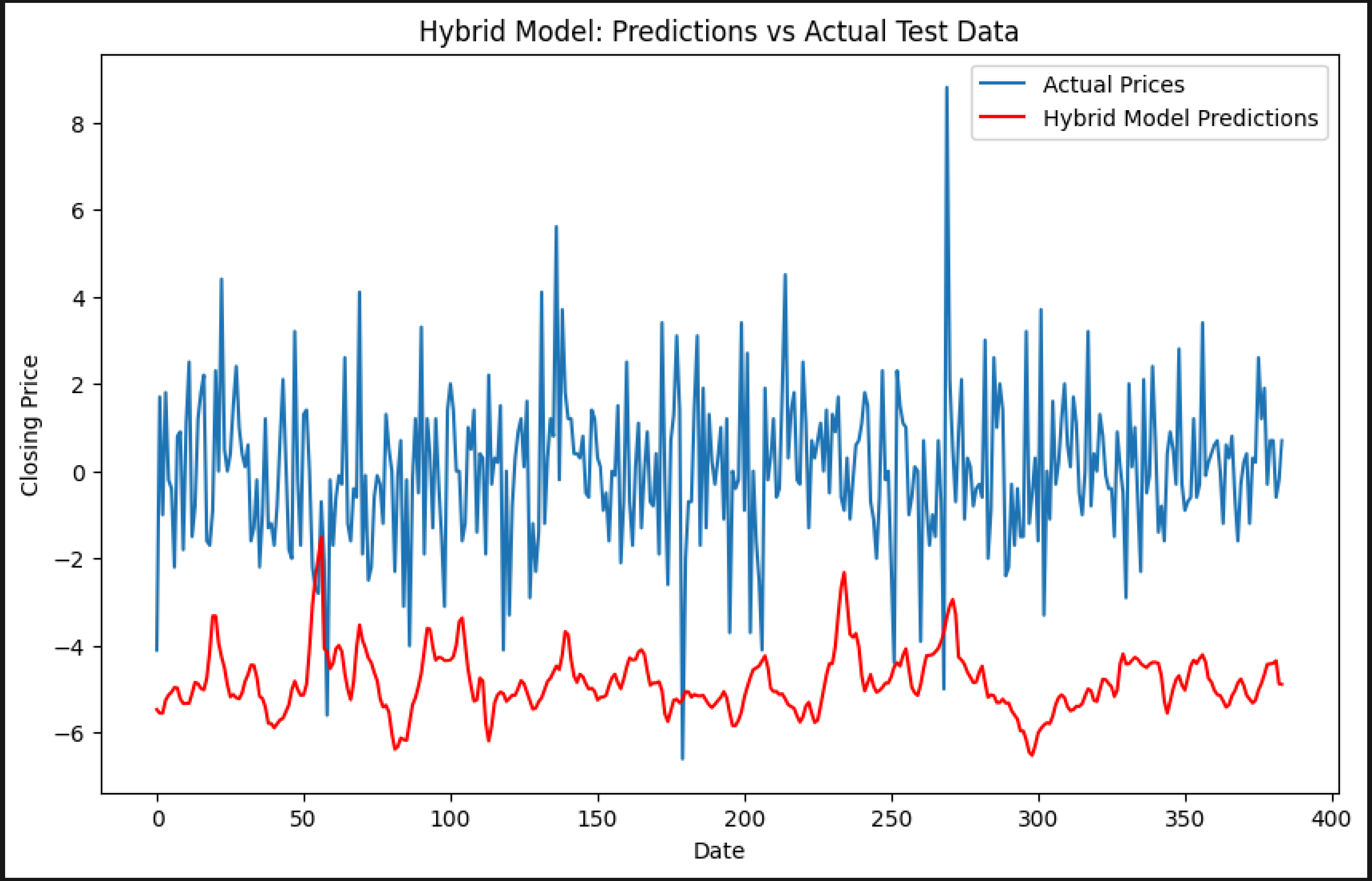
Limitations of ARIMA Model

- The ARIMA model faced challenges in accurately predicting market performance due to several factors:Volatility and Skewness: The high volatility and skewness in the data can complicate predictions.
- Market Sentiment: The model did not account for market sentiment or external events that may influence stock prices.



Implementation of LSTM Model

- To improve prediction accuracy and capture spikes more effectively, an LSTM model was implemented:LSTMs are designed to handle sequences and can learn long-term dependencies, making them suitable for time series forecasting in volatile environments.



Combining

Combining ARIMA and LSTM model prediction result

Conclusion

- In this analysis, we explored the performance of both ARIMA and LSTM models for predicting stock prices, particularly focusing on the challenges posed by market volatility. The evaluation metrics for the combined ARIMA and LSTM model yielded the following results:Mean Absolute Error (MAE): 4.8973
- Mean Squared Error (MSE): 27.0938
- Root Mean Squared Error (RMSE): 5.2052
- R-squared: -8.2559

Key Insights

1. Model Performance: While the combined ARIMA and LSTM model demonstrated improved performance compared to traditional ARIMA alone, it still struggled with accuracy due to the inherent volatility of stock prices.
2. Volatility Challenges: The financial market is characterized by rapid fluctuations and unpredictable movements, which significantly impact forecasting accuracy.
3. Future Considerations: Despite these challenges, the integration of LSTM with ARIMA has shown potential for capturing complex patterns in stock data. Further refinements and tuning may enhance predictive capabilities.

Conclusion Statement

In summary, while the evaluation metrics indicate that the model is not fully accurate due to stock market volatility, the combination of ARIMA and LSTM has provided better results than using ARIMA alone. Continued exploration of advanced modeling techniques may yield even more robust predictions in future analyses.

Thank You

@reallygreatsite reallygreatsite.com +123 456 7890