Stock Price Prediction

A Data-Driven Approach

Presented by Sakshi Sharma

01

Background

Importance of Stock Price Prediction:

- Financial markets significantly influence economic stability.
- Accurate predictions assist investors in making informed decisions, maximizing profits, and minimizing losses.

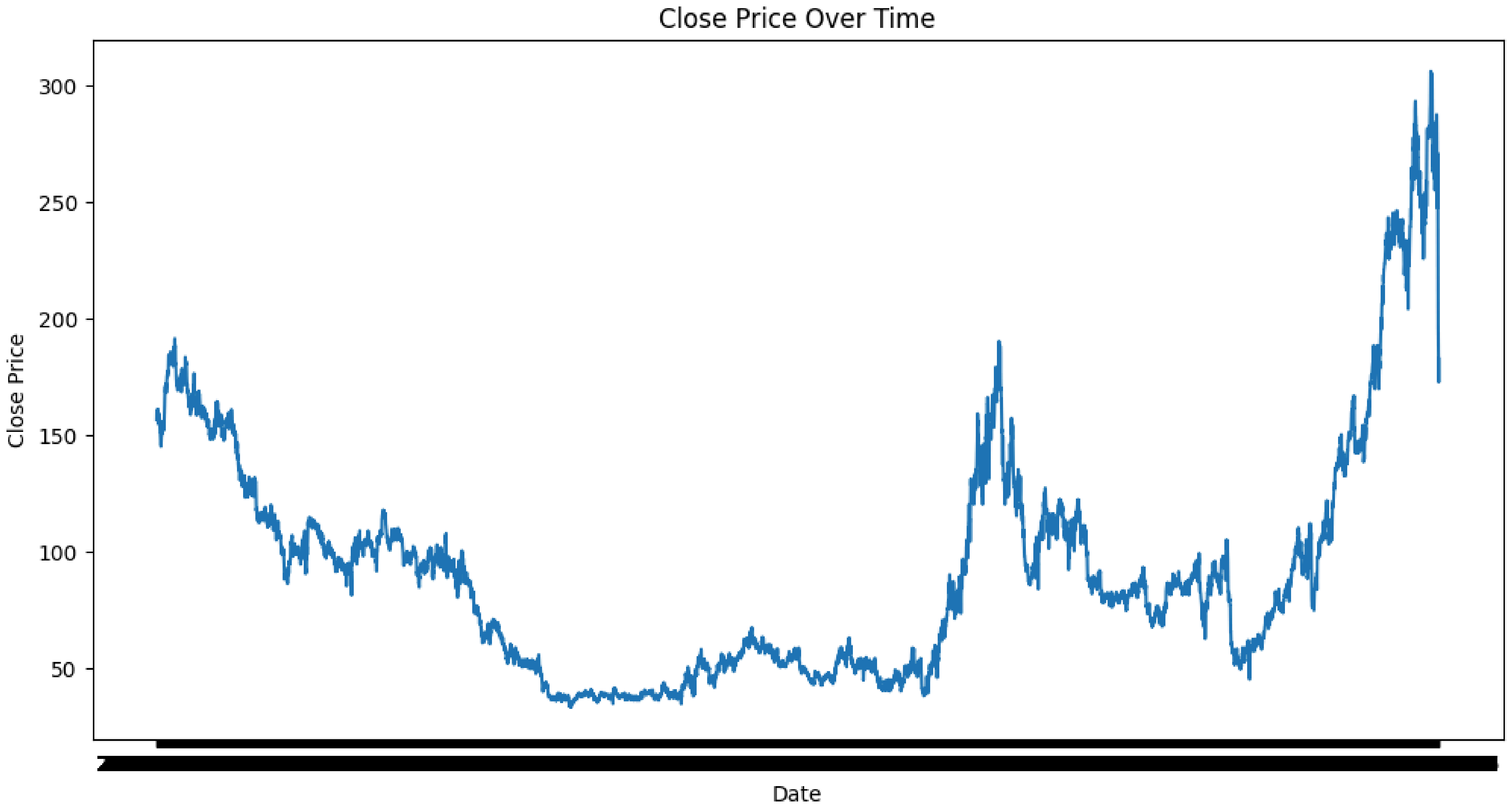
Challenges:

- Stock markets are characterized by volatility and unpredictability.
- External factors such as news events, economic indicators, and geopolitical developments can heavily influence prices.

02

Data Analysis Results

Exploratory Data Analysis (EDA):



03

Basic Statistics:

	Close Price	Open Price	High Price	Low Price	Volume \
count	9202.000000	9202.000000	9202.000000	9202.000000	9.202000e+03
mean	92.180961	92.256183	93.176451	91.330146	1.726677e+08
std	50.452228	50.598215	51.049837	50.087405	1.251280e+08
min	33.000000	33.000000	33.200000	32.200000	9.340000e+06
25%	52.000000	52.100000	52.800000	51.500000	8.073000e+07
50%	85.100000	85.100000	86.050000	84.200000	1.540150e+08
75%	110.800000	110.800000	111.900000	109.275000	2.305225e+08
max	305.900000	309.800000	311.800000	303.900000	1.280000e+09

	Change %
count	9202.000000
mean	0.017502
std	1.876667
min	-14.740000
25%	-0.940000
50%	0.000000
75%	0.900000
max	16.250000

Identified Issues:

- Missing values in the dataset addressed using forward fill methods to ensure data continuity.
- Lag features created to capture past price movements, enhancing model training.

Model Explanation

04

LSTM Overview:

Long Short-Term Memory (LSTM) Networks:

- A type of Recurrent Neural Network (RNN) designed to learn from sequences of data.

Strengths:

- Temporal Dependencies: LSTMs are particularly effective at capturing long-term dependencies in time series data due to their unique architecture that includes memory cells, input gates, output gates, and forget gates.
- Handling Non-linearity: They can model complex non-linear relationships, which are common in financial data.
- Suitability: Ideal for stock price prediction as they can learn patterns from historical prices and incorporate those patterns into future predictions.

ARIMA Overview:

Autoregressive Integrated Moving Average (ARIMA):

- A classical statistical approach for modeling time series data.

Components:

- Autoregressive (AR) part: Relies on the relationship between an observation and a number of lagged observations (previous values).
- Integrated (I) part: Differencing of raw observations to allow for the time series to become stationary.
- Moving Average (MA) part: A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Strengths:

- Simplicity: Easier to implement and interpret, making it a common choice for time series forecasting.
- Statistical Foundation: Provides a solid statistical framework, which is beneficial for understanding the underlying patterns in the data.

05

Feature Engineering Methods

For LSTM:

Scaling of Close Price:

- Use of StandardScaler to normalize input features, improving model convergence and ensuring that the model is not biased by the scale of the data.

Creation of Lag Features:

- For example, creating a feature called Prev_Close to provide the model with previous prices, allowing it to make informed predictions based on historical data.

For ARIMA:

Differencing:

- Applying differencing to stabilize the mean of the time series, making it stationary and suitable for modeling.

Lagged Variables:

- Utilizing past observations in the autoregressive part to predict future values.

Justification

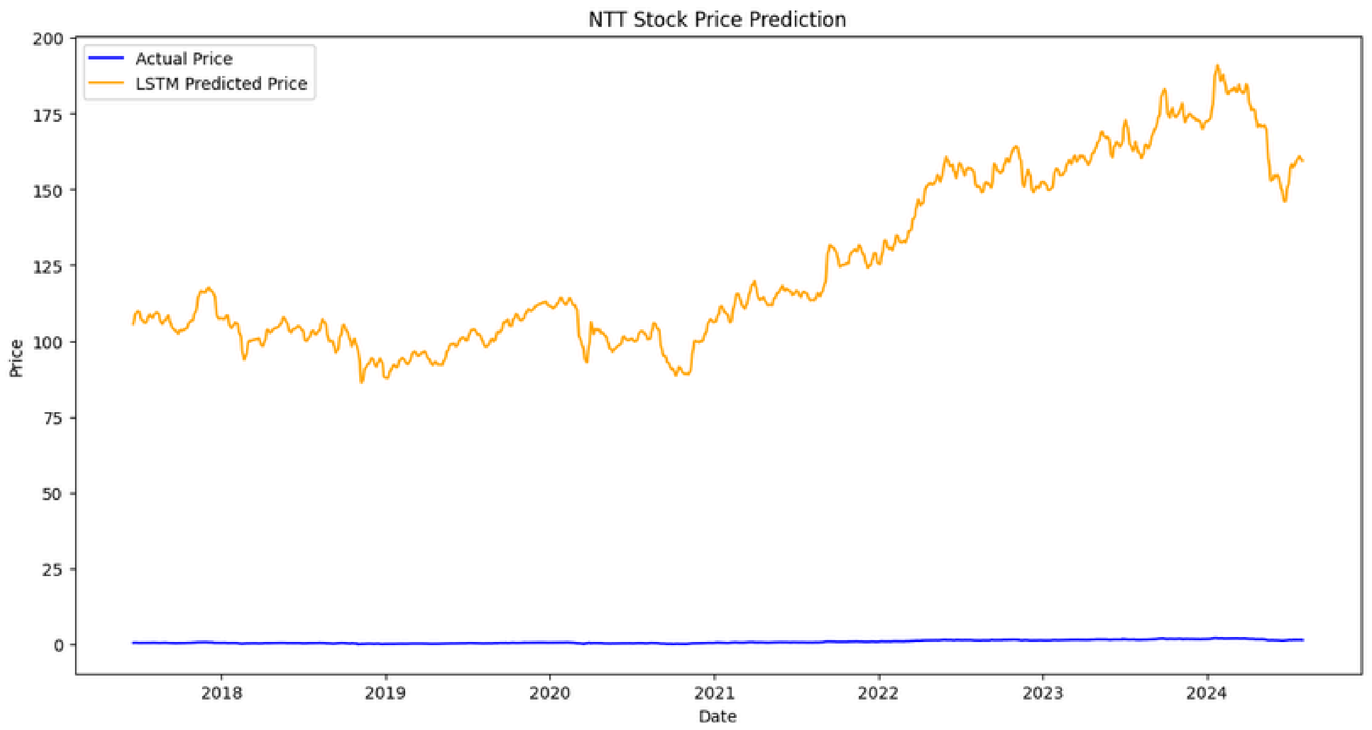
LSTM:
The ability of LSTMs to retain information over long sequences is crucial in stock data where past values significantly affect future prices. This makes them particularly suitable for capturing trends and patterns in volatile markets.

ARIMA:
While ARIMA is effective for many stationary time series, it may struggle with non-stationary data common in stock prices unless properly differenced. Additionally, it may not capture complex patterns and non-linearities as effectively as LSTMs.

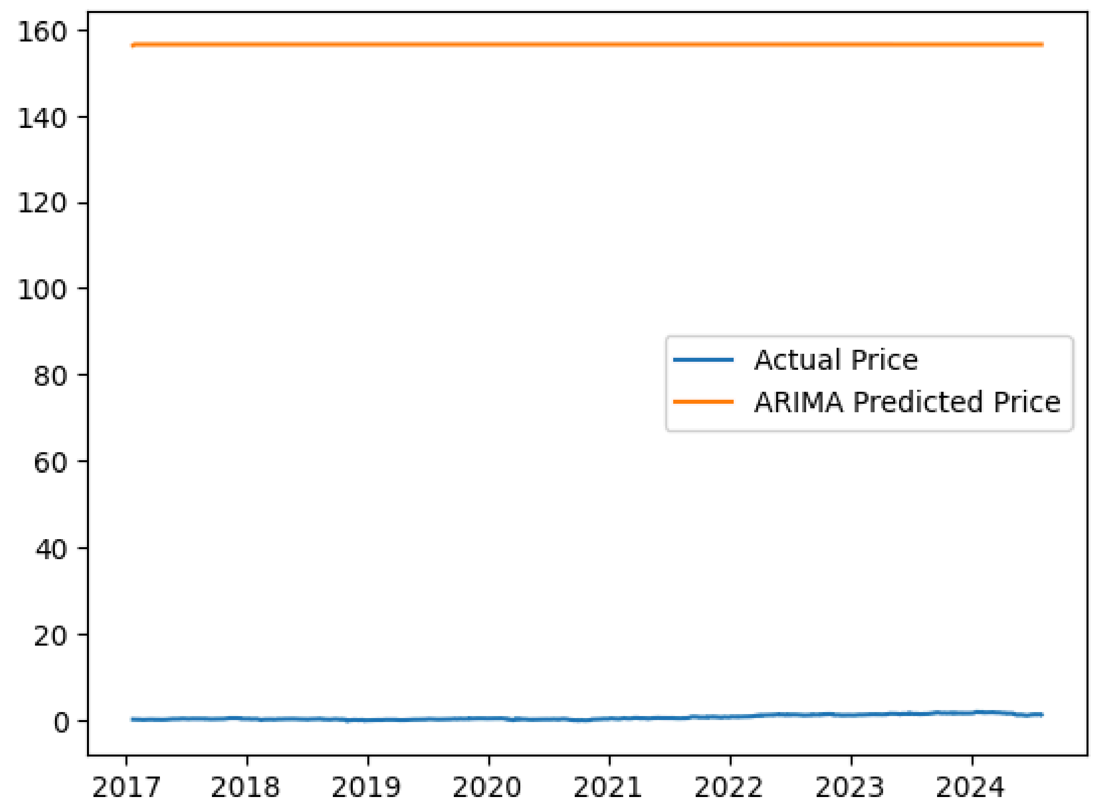
06

Comparison

LSTM



ARIMA



On comparing both the models I found LSTM is better than ARIMA

Evaluation Metrics

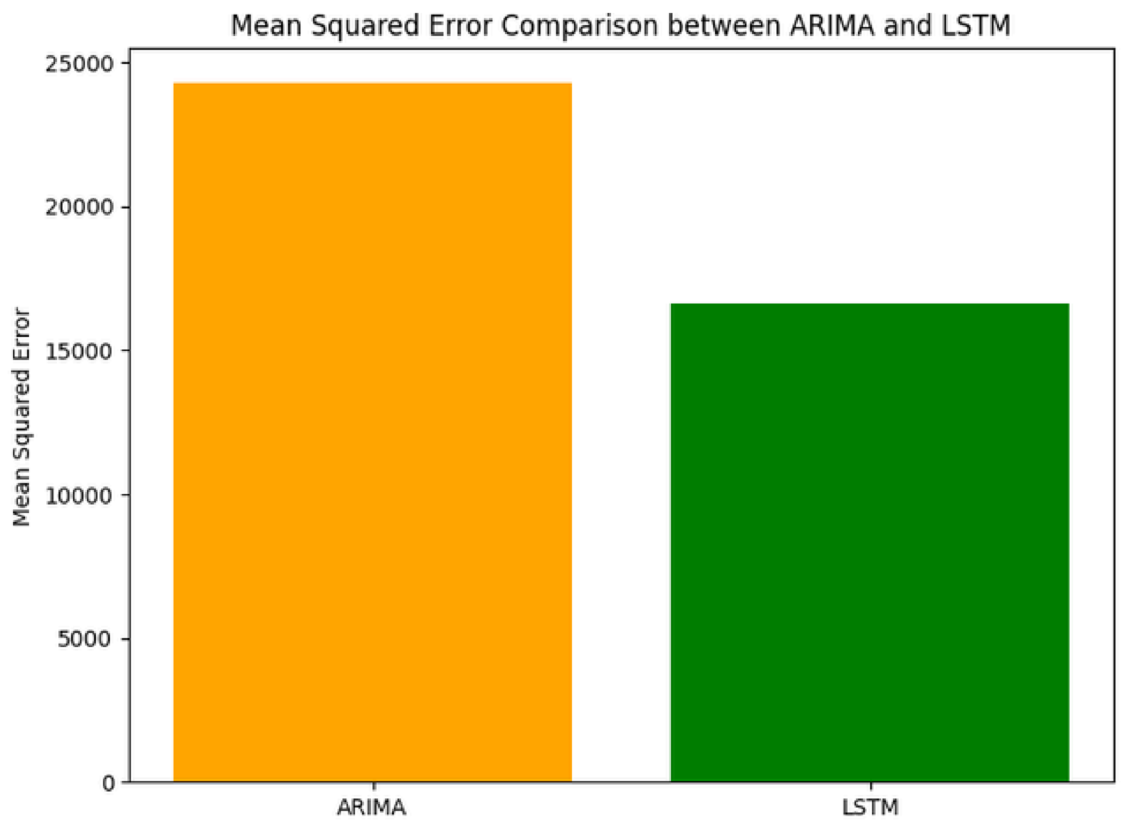
07

Selected Metrics:

Mean Squared Error (MSE): Used to quantify the difference between predicted and actual stock prices.

Comparison of the performance of the LSTM model against traditional models like ARIMA.

Results:




```
ARIMA Mean Squared Error: 24265.764565456593
LSTM Mean Squared Error: 16639.63850479618
```

Validation

08

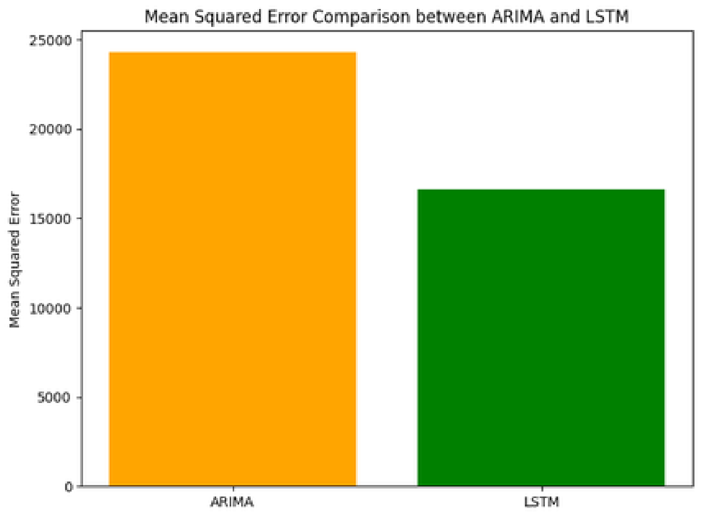
Hypotheses for Improvement:

Hypothesis 1: Increasing the number of LSTM units will enhance model performance.

Results

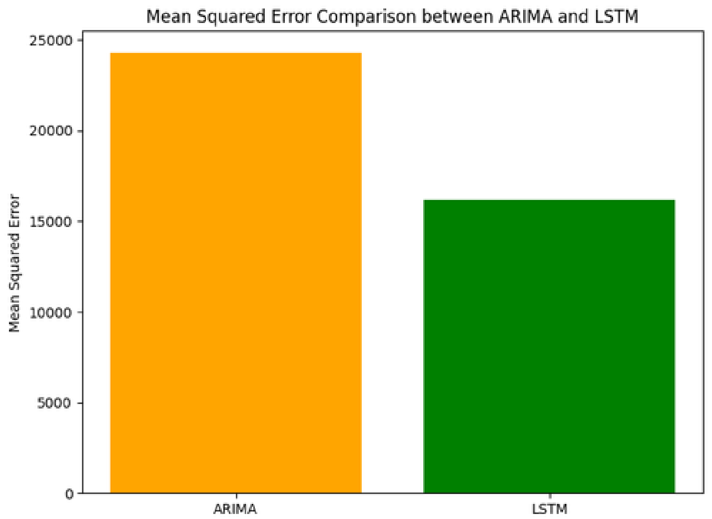
MSE Before

```
ARIMA Mean Squared Error: 24265.764565456593
LSTM Mean Squared Error: 16639.63850479618
```



MSE After

```
ARIMA Mean Squared Error: 24265.764565456593
LSTM Mean Squared Error: 16196.350248275214
```



- The MSE decreased significantly after increasing the number of LSTM units, indicating improved model performance.

Changes made:

- Adjustment: Increased the number of LSTM units in layer 1 from 50 to 100 in the model architecture.
- Rationale: This adjustment was made to enhance the model's capacity to learn complex patterns and dependencies in the stock price data, which is often necessary for accurately predicting future prices.

Analysis of Impact:

- Greater Capacity: More LSTM units enhance the model's learning ability, capturing intricate patterns in stock price data.
- Learning Complex Patterns: With 100 units, the model effectively utilizes historical information, improving predictions.
- Reduced Underfitting: Increasing units from 50 to 100 helped the model fit training data better, enhancing generalization on unseen data.

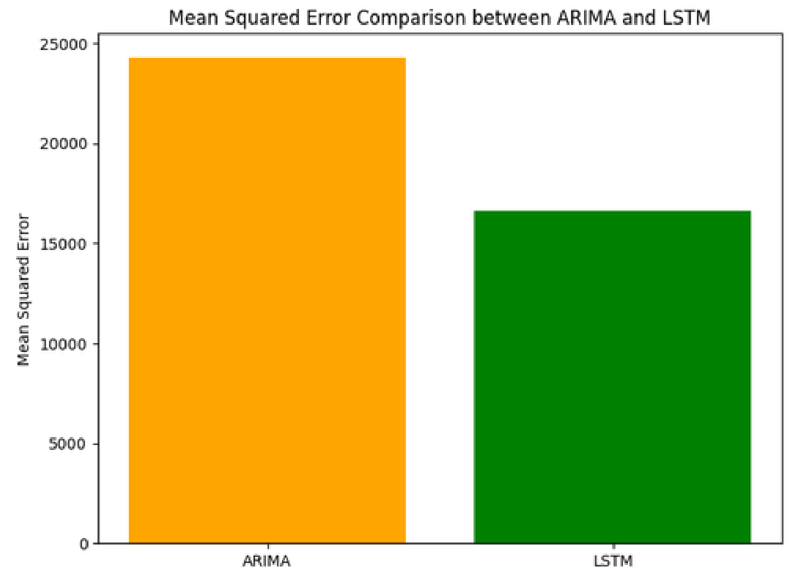
09

Hypothesis 2: Modifying batch size or the number of training epochs will lead to better convergence and accuracy.

Result:

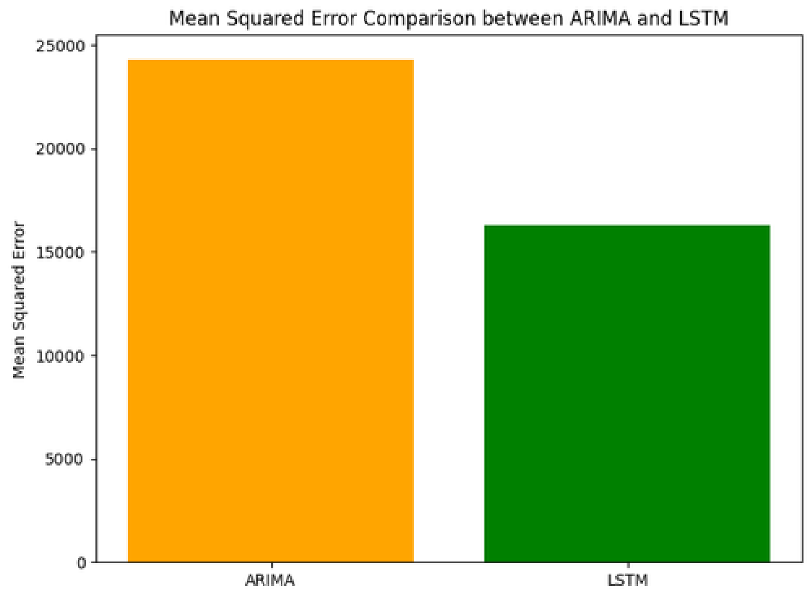
MSE Before

```
ARIMA Mean Squared Error: 24265.764565456593
LSTM Mean Squared Error: 16639.63850479618
```



MSE After

```
ARIMA Mean Squared Error: 24265.764565456593
LSTM Mean Squared Error: 16283.0972063657
```



Changes made

- Batch Size: Changed from 64 to 84.
- Epochs: Increased from 10 to 20.

Impact on Convergence and Accuracy:

- Increasing the batch size often leads to more stable gradients, enhancing model convergence.
- Increasing epochs allows the model more opportunities to learn from the data, improving accuracy over time.

Insights for Future Modeling:

- A larger batch size can help balance speed and accuracy.
- More epochs can lead to better fitting, but require monitoring for overfitting, indicating the need for regularization techniques in future experiments.

Summary

Conclusion:

- The LSTM model demonstrated superior performance over ARIMA for stock price predictions, validated by lower MSE scores.
- Emphasized the importance of hyperparameter tuning to enhance LSTM performance.

Future Outlook:

- Explore advanced models like Gated Recurrent Units (GRU) or Transformers for potentially better results.
- Consider integrating additional features, such as sentiment analysis from financial news and trading volume data, to improve prediction accuracy.

Thank You