**Time Series Analysis of NIFTY 50 Index**

Prajwal B M

CH24M556

December 29, 2024

**Abstract**

This paper analyses how Time Series Analysis techniques can be applied to capture movement of an exchange traded index in a stock market. Specifically, Seasonal Auto Regressive Integrated Moving Average (SARIMA) and Long Short-Term Model (LSTM) classes of models is applied to capture the movement of Nifty 50 index, which is one of the most actively traded exchange.

**Introduction**

Time series analysis is a powerful statistical tool used to analyze time-ordered data points. In the context of financial markets, it helps in understanding and forecasting the behavior of stock indices. The Nifty 50 index, representing the weighted average of 50 of the largest Indian companies listed on the National Stock Exchange, serves as a benchmark for the Indian equity market. This report explores the application of SARIMA and LSTM models to predict the future movements of the Nifty 50 index, providing insights into their effectiveness and accuracy.

Forecasting stock price movements is a complex task due to the inherent non-linear and volatile nature of financial markets. Traditional models like ARIMA have been widely used for time series forecasting, but they often fall short when dealing with seasonal patterns and non-linear data. To address these limitations, we employ the Seasonal ARIMA (SARIMA) model, which incorporates seasonal differencing to handle seasonality in the data. Additionally, we explore the use of Long Short-Term Memory (LSTM) networks, a type of Recurrent Neural Network (RNN) designed to capture long-term dependencies in sequential data.

In this project, we utilized 8 years of historical data (2016-2023) to train our models and tested their performance against the data from 2024. The SARIMA model parameters were determined using the Akaike Information Criterion (AIC), and the model was evaluated based on its ability to forecast the Nifty 50 index's average price movement for the next 12 months. Despite the SARIMA model's limitations in capturing the complex, non-linear patterns of stock market data, it provided a baseline for comparison.

To improve prediction accuracy, we implemented an LSTM model, which involved normalizing the stock prices and creating rolling window sequences of length 50. The LSTM model was trained on the training data for 20 epochs with a batch size of 32, resulting in more accurate predictions of the Nifty 50 index. This approach highlights the potential of advanced machine learning techniques in financial forecasting, while also acknowledging the uncertainties associated with macro and micro-economic factors that can influence stock prices.

**ARIMA model**

Auto Regressive Integrated Moving Average (ARIMA) is a class of models that explains a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values. Any ‘non-seasonal’ time series that exhibits patterns and is not a random white noise can be modelled with ARIMA models.

An ARIMA model is characterized by 3 terms: p, d, q

where,  p is the order of the AR term

 d is the number of differencing required to make the time series stationary

 q is the order of the MA term



Predicted 𝑌𝑡 = Constant + Linear combination Lags of Y (up to p lags) + Linear combination of lagged forecast errors (up to q lags)

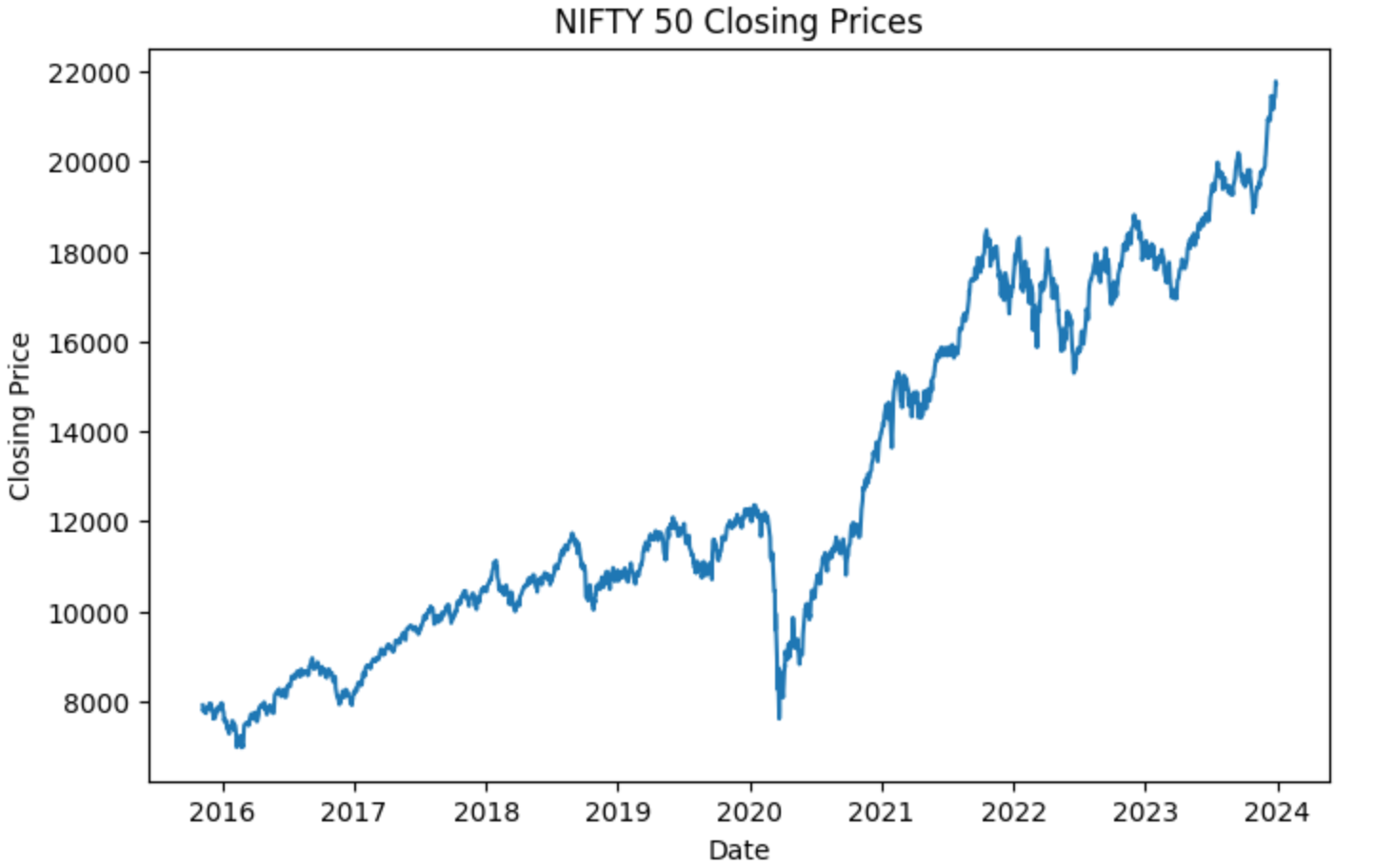
**Seasonal ARIMA (SARIMA)**

Many previous studies have focussed on using ARIMA models for forecasting stock markets. However, a shortcoming of plain ARIMA model is that it does not support seasonality. If the time series has defined seasonality, then, SARIMA which uses seasonal differencing is a more appropriate model in which values from previous season are subtracted instead of subtracting consecutive terms.

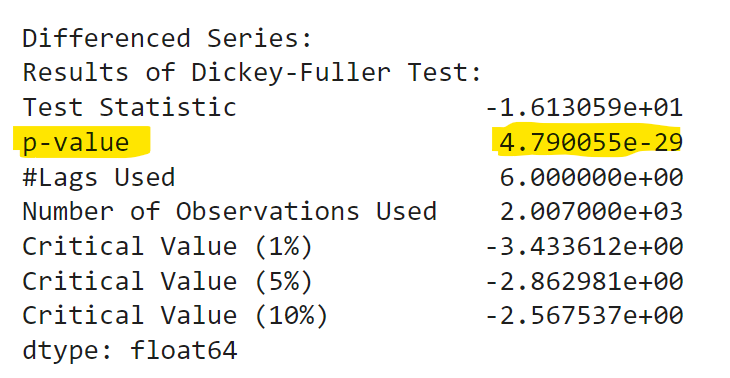
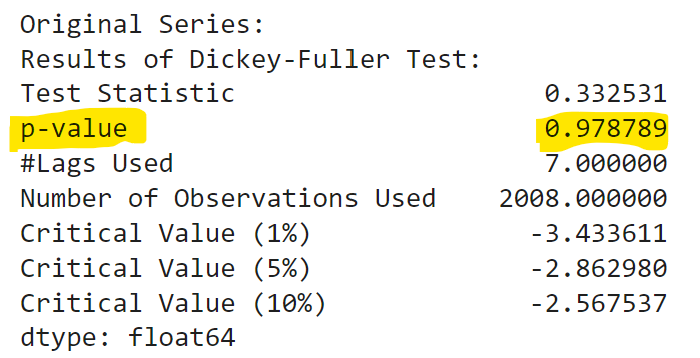
SARIMA can be represented by SARIMA(p,d,q) x (P,D,Q,m), where,

* (p,d,q) are non-seasonal terms of the model
* (P,D,Q,m) are seasonal autoregressive (SAR), order of seasonal differencing (D) and seasonal moving average (SMA) and frequency of the time series terms respectively.

**Methods & Results**

Plot of Closing price vs Date for Training Data.**Dicky-Fuller Test for Stationarity:**

By observation the time series is not stationary, so check for stationarity by Dicky-Fuller test. In order to make the series stationary take the difference. i.e. ΔYt = Yt − Yt−1



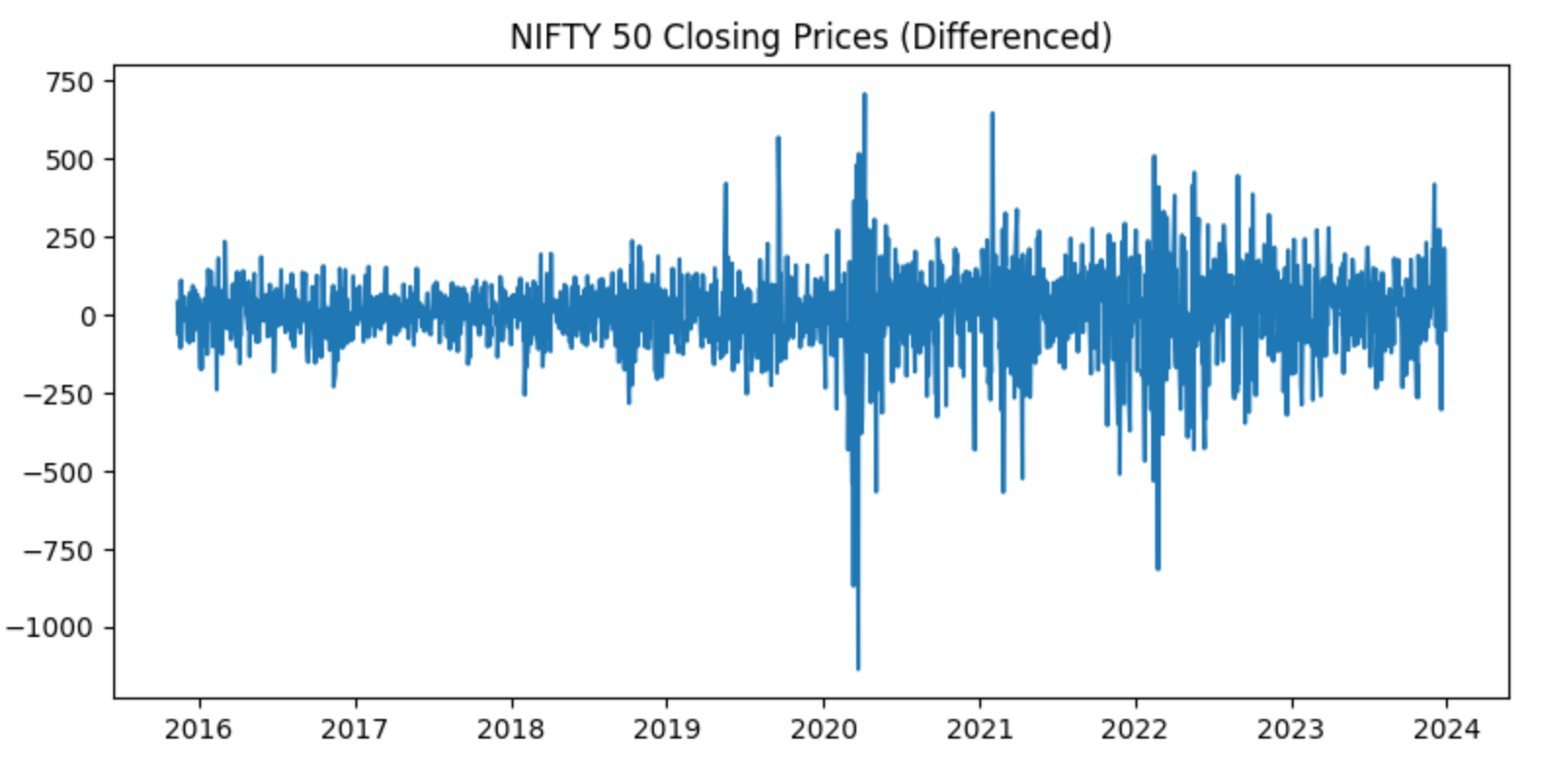
p-Value = 0.978 > 0.05

p-Value = 4.79e-29 < 0.05

Original Series: Yt

Differenced Series: ΔYt = Yt − Yt−1

**Differenced Stationary Time series data (ΔYt):**

****

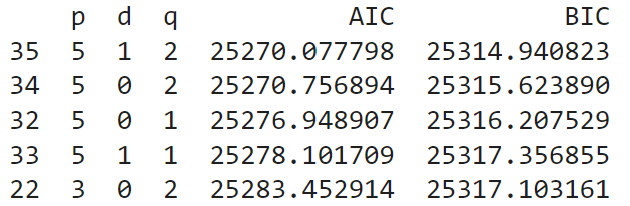
**ACF and PACF Plots for Stationary signals:**

**ACF (autocorrelation)** is the linear correlation of a signal with itself at two different points in time, ACF is just such correlation as a function of the lag h between two points of time.

**PACF (partial autocorrelation function)** is essentially the autocorrelation of a signal with itself at different points in time, with linear dependency with that signal at shorter lags removed, as a function of lag between points of time. Informally, the partial correlation between x(t) and x(t+h) is the autocorrelation between x(t) and x(t+h) without the contribution of x(t+1),x(t+2),....,x(t+h−1).



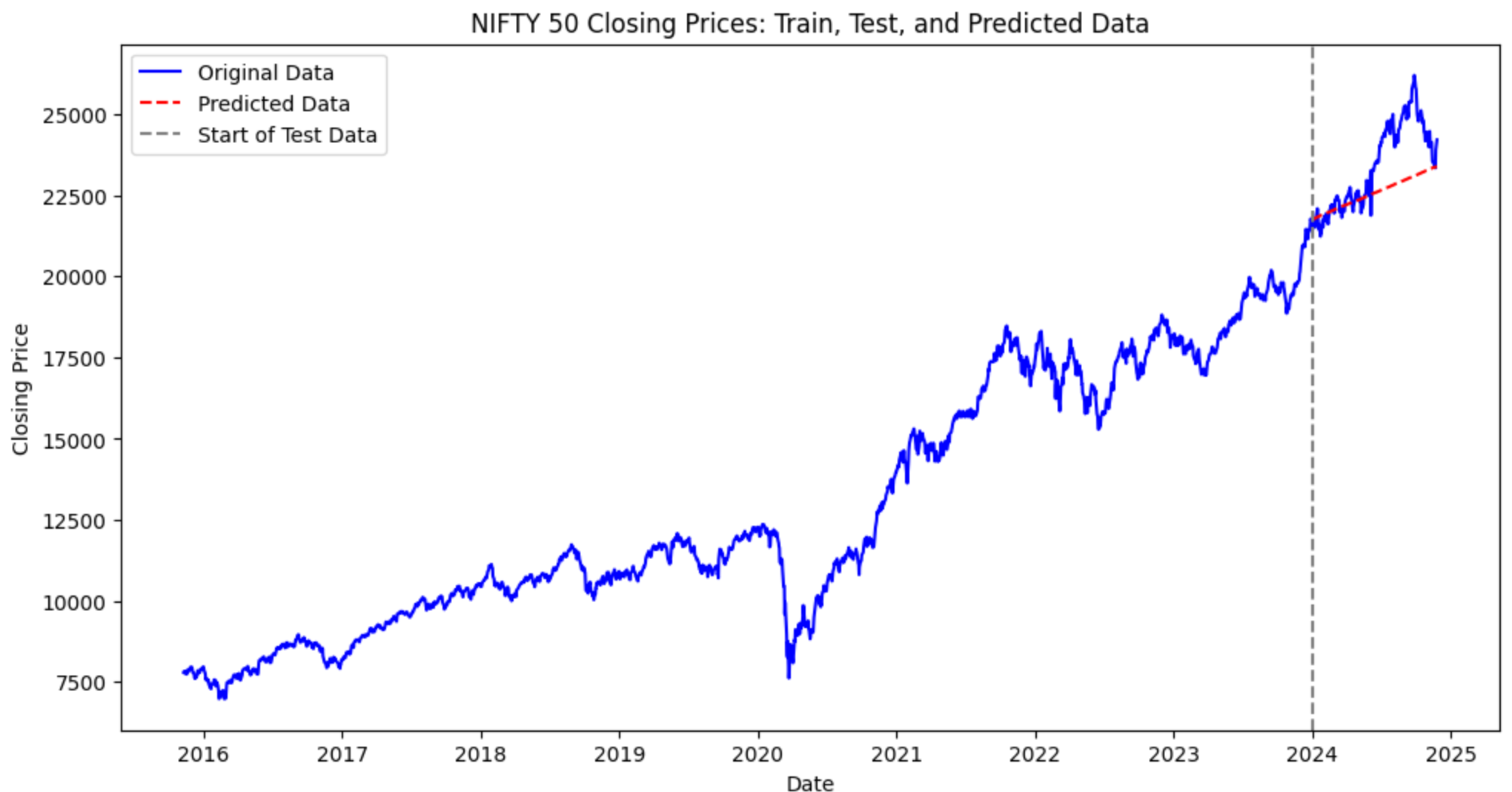
From Above plots we cannot really make out which order model to use. So we make use of **AIC(Akaike information criterion)** & **BIC (Bayesian information criterion)** to estimate the model order.



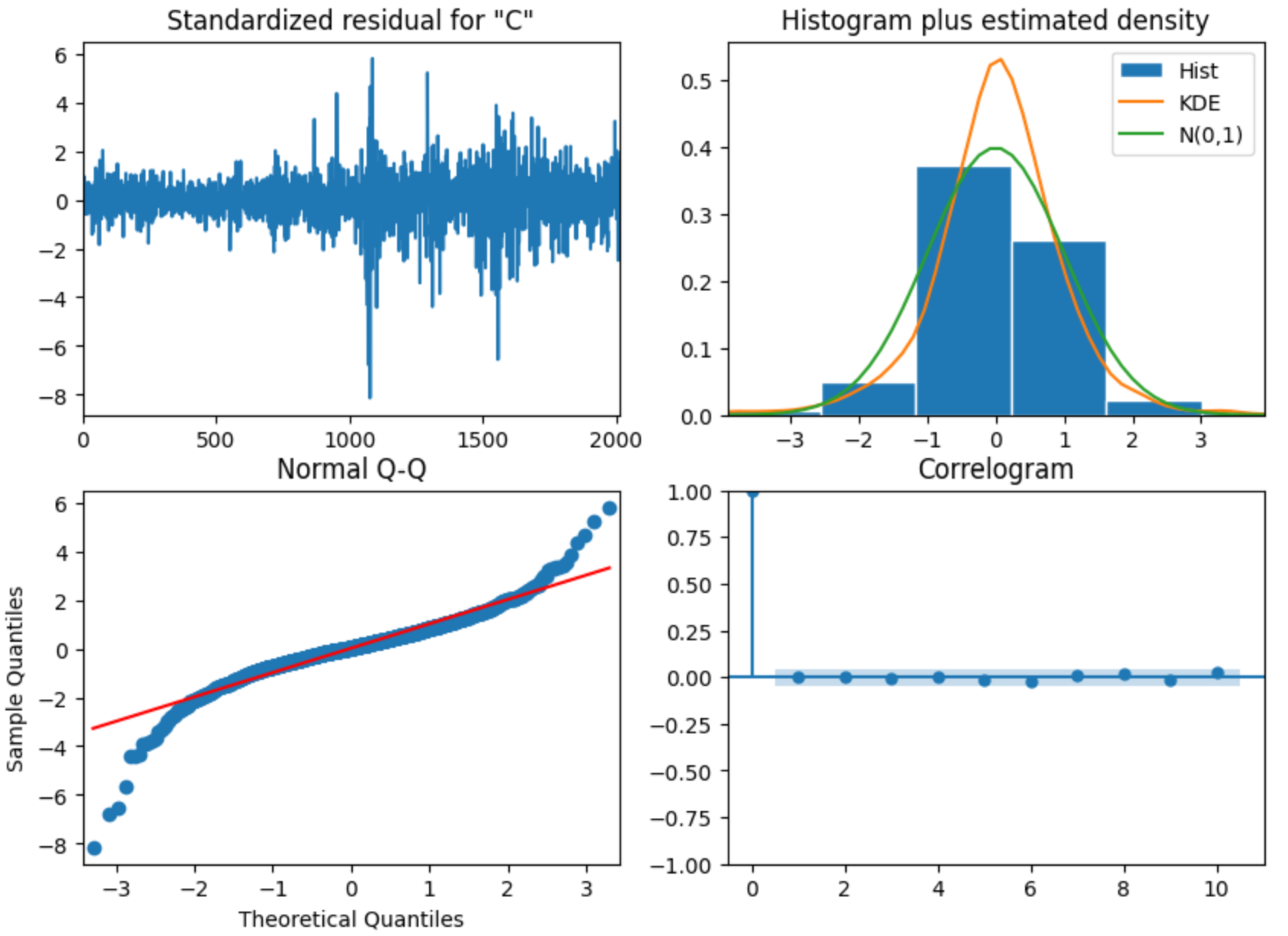
From above table we have lowest AIC value for **(p,d,q) = (5,1,2).** So, this will be our model order.

**1. SARIMA Model:**

**Fit & forecast for SARIMA Model of order (5,1,2):**

****

**SARIMA (5,1,2) Model Evaluation:**

****

**Interpretation of above plots:**

* If the model fits well the residuals will be white Gaussian noise
* Histogram & KDE should be normally distributed
* Normal Q-Q points should be almost on red line
* Correlogram of residuals should be zero for all legs

Mean Absolute error is 88.85 And prediction is not satisfactory. Stock Market Index data is a complex, non-linear data, which requires complex algorithm to comprehend.

But still SARIMA model can be used to predict the near future trends and opportunities.

**2. Long Short-Term Memory Model:**

Long Short-Term Memory (LSTM) networks are a type of Recurrent Neural Network (RNN). They are designed to handle sequences of data, making them particularly useful for tasks like time series forecasting and natural language processing.

**Steps for LSTM Model:**

1. **Data Normalization**:

MinMaxScaler scales the 'Close' prices between 0 and 1 to ensure the data is within a consistent range for the LSTM model.

1. **Create Sequences for LSTM:**

A rolling window of sequence\_length (50) is used to create sequences of past data points. Each sequence is used as an input to predict the next data point.

1. **Build the LSTM Model:**

A sequential model with two LSTM layers (each with 50 units) and a Dense layer is defined. The first LSTM layer returns sequences (return\_sequences=True) to feed into the second LSTM layer.

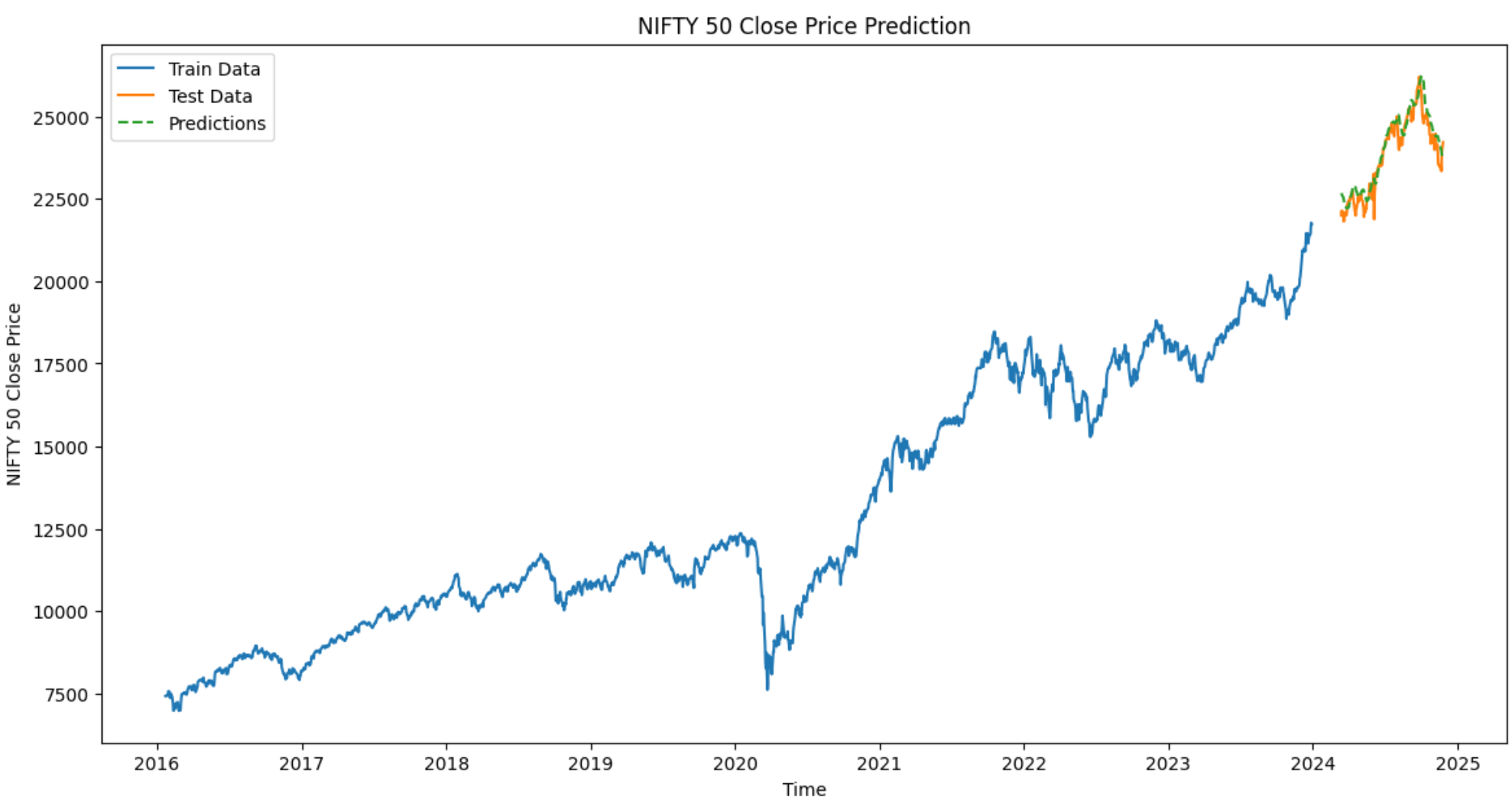
1. **Compile the Model:**

The model is compiled using the Adam optimizer and Mean Squared Error (MSE) as the loss function.

1. **Training the Model:**

The model is trained on the training data (X\_train, y\_train) for 20 epochs with a batch size of 32.

**Fit & Forecast for LSTM model:**



**LSTM Model Evaluation:**

Mean Squared Error: 158957.95

Mean Absolute Error: 299.12

Root Mean Squared Error: 39.69

Predictions form LSTM Model traces the test data with more accuracy. So, this helps us to predict Nifty 50 Index correctly.

**Conclusion:**

In this project, we have tried to forecast the stock price movement of the Nifty 50 index which is the largest and most liquid Indian securities. Seasonal ARIMA (SARIMA) model parameters were discovered using AIC criteria. To apply the time series techniques, we used 8 years data (2016-2023 for this project. And tested our model against the data of 2024. The stationarity in this time series has been removed using decomposing and differencing. After that, Seasonal ARIMA (SARIMA) parameters - p, d, q, P, D, Q, m was determined based on AIC criteria and using that model average index price movement for next 12 months forecasted and plotted.

We found the predictions from SARIMA unsatisfactory, so, we tried to use LSTM model. Before building the model, we normalised the stock price, created rolling window sequence of length 50. We used 2 layers sequential layers, wherein first LSTM layer feeds into the second LSTM layer. We trained on the training data (X\_train, y\_train) for 20 epochs with a batch size of 32.

Finally, we were able to predict Nifty 50 Index with maximum accuracy.

There can also be some uncertainties associated in forecasting stock and index price movements e.g. due to macro as well as micro-economic factors which can influence price movement.

However, these are outside the scope of the current Project.