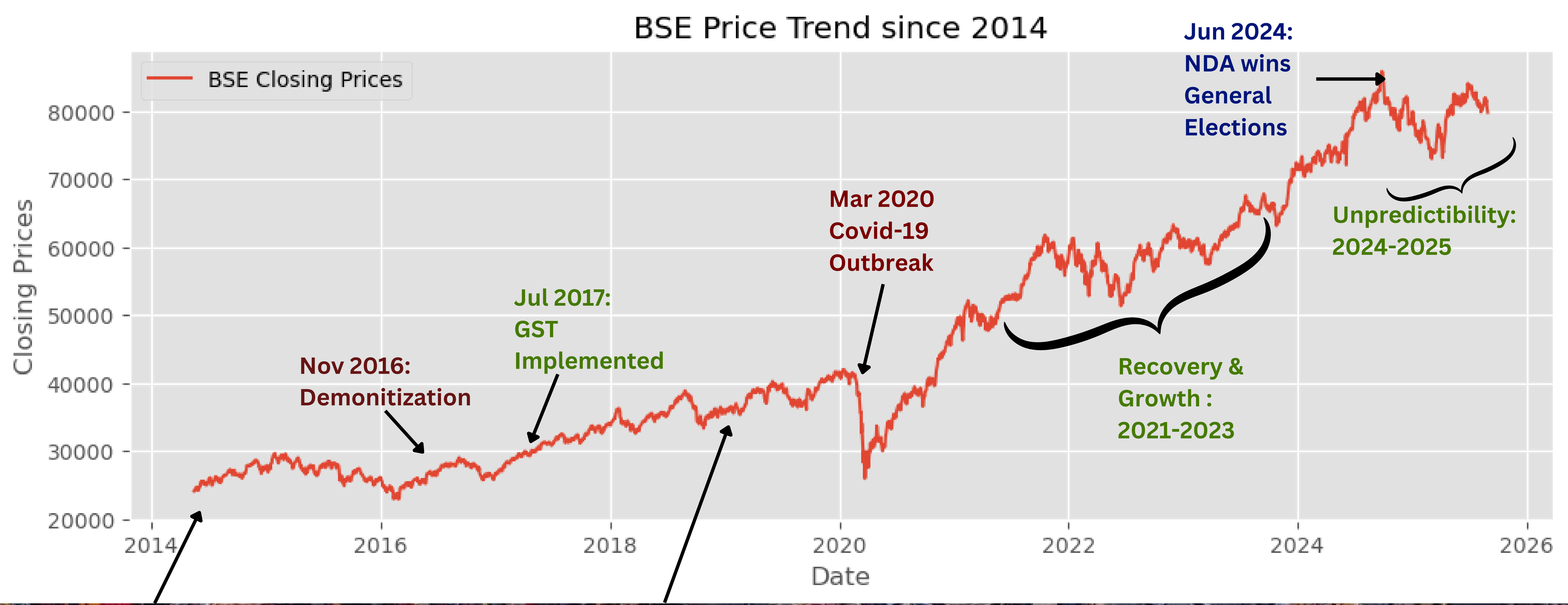
Data ,Policy And Market:
A Time Series Perspective

OBJECTIVES

- The primary objective of this project is to use time series analysis to model and forecast the behavior of the BSE SENSEX from 2014 to 2025.
- This analysis aims to understand the market's response to major political and economic events during this period.
- This provides a Time Related understanding of the relationship between policy decisions and market sentiment.
- To reveal how well a traditional time series model and Modern Deep-learning based models captures the impact of non-linear, unpredictable events on the market.

TIMELINE OF SENSEX (2014-2025)

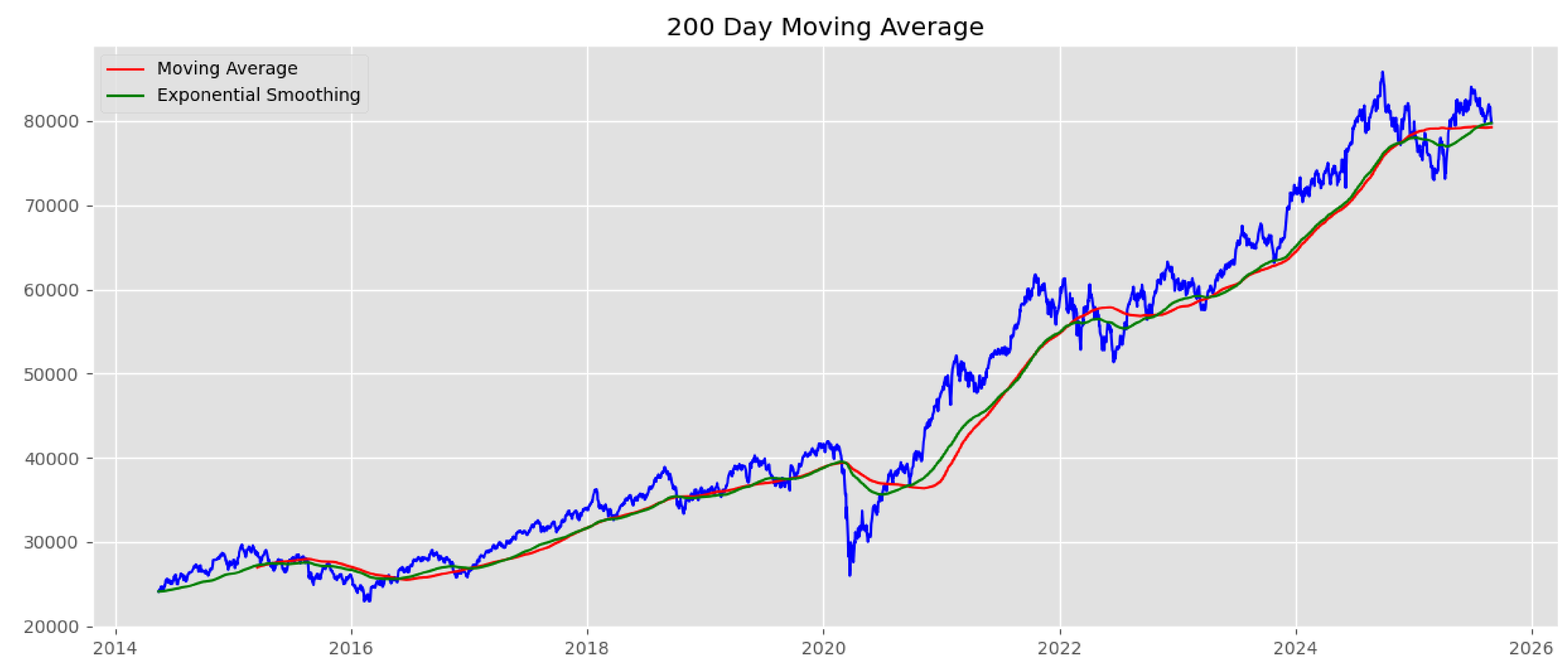


May 2014: BJP wins General Elections

June 2019: BJP wins General Elections

- *The SENSEX tracks the performance of 30 of India's largest and most actively traded companies, acting as a key indicator of the overall health of the Indian economy.

SMOOTHING TREND



- **A** **Moving Average (MA)** **is a statistical technique used in time series analysis to smooth out short-term fluctuations and highlight longer-term trends.**

- **Exponential smoothing** **uses a weighted average of past observations to predict future values. It assigns exponentially decreasing weights to older data points**

Insights

- **2014–2019**

Sensex consistently traded above the 200-DMA after 2014 elections → signaled investor confidence and a multi-year bull run.

- **Early 2020 – COVID-19 Crash**

Sharp fall in March 2020 pushed Sensex well below the 200-DMA → strongest bearish signal of the decade, reflecting panic selling.

- **Late 2020 – Recovery Rally**

By June–July 2020, Sensex crossed back above the 200-DMA, confirming recovery and start of a fresh uptrend.

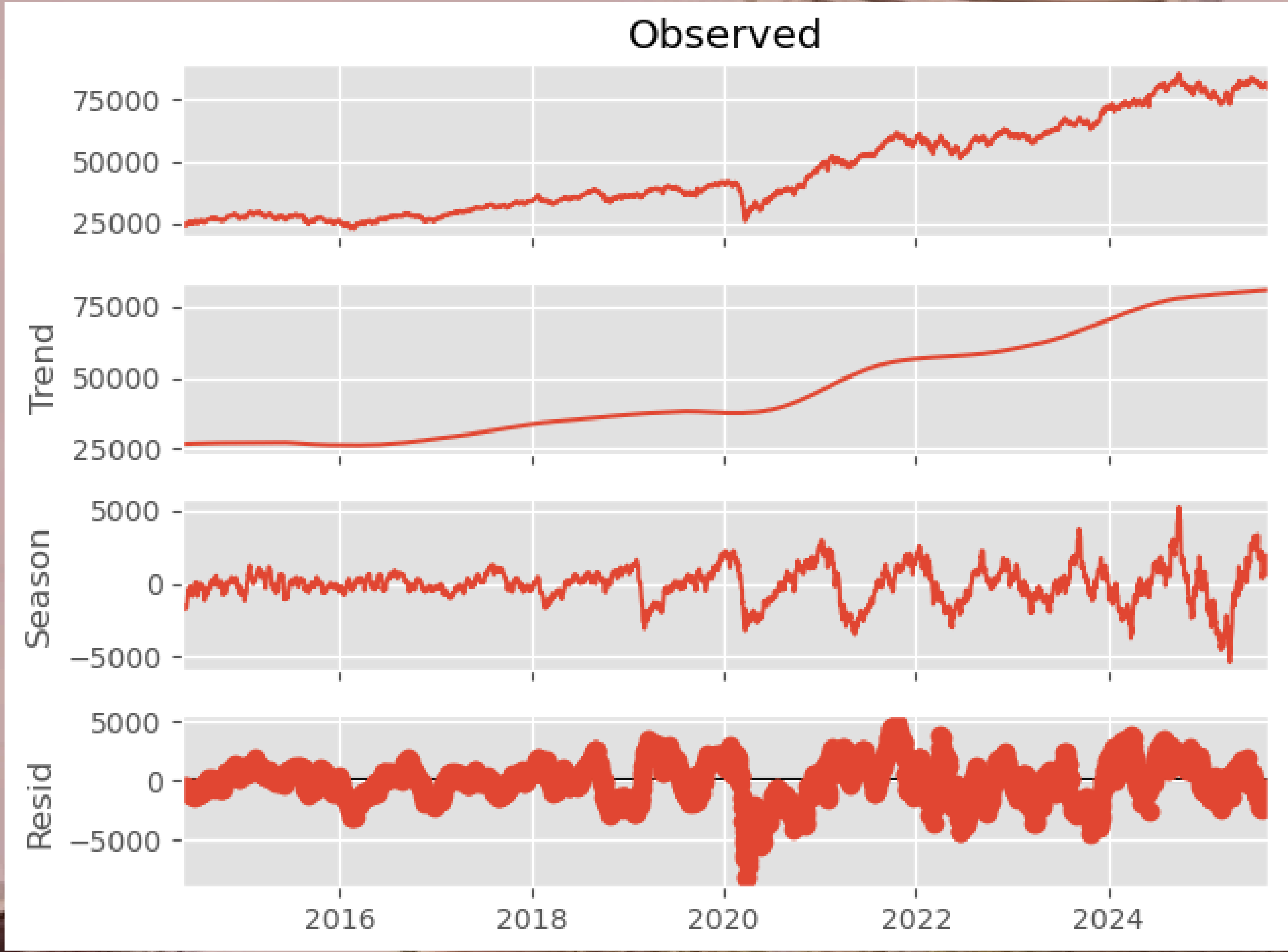
- **2021–2022 – Post-COVID Bull Run**

Sensex surged far above its 200-DMA, with only minor pullbacks → strong bullish momentum, supported by liquidity and earnings recovery.

- **2023–2025 – Resilient Consolidation**

Temporary corrections tested the 200-DMA, but Sensex repeatedly rebounded from it as support, showing long-term bullish resilience.

SEASONAL DECOMPOSITION OF TIME SERIES BY LOESS (STL)

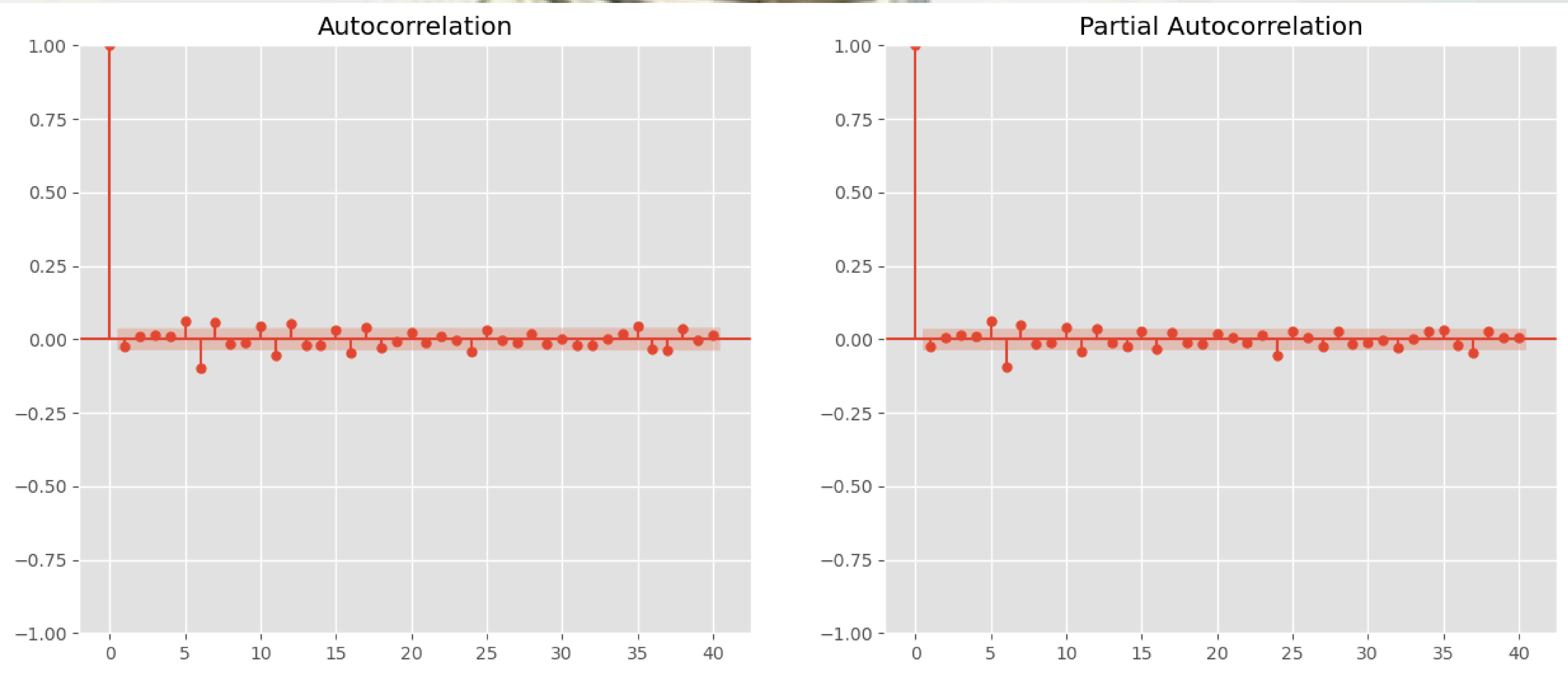


- Better to find seasonal patterns using STL decomposition
- Stl handles outliers
- only handles additive models
- Classical decompostion in multiplicative decomposition

- Sensex shows a strong long-term upward trend, especially post-2020 recovery
- Moderate seasonal patterns indicate recurring market cycles linked to fiscal and economic events.
- Short-term shocks in residuals reflect market volatility from global and domestic events.

AUTOCORRELATION & PARTIAL AUTOCORRELATION

Plotting ACF & PACF plots

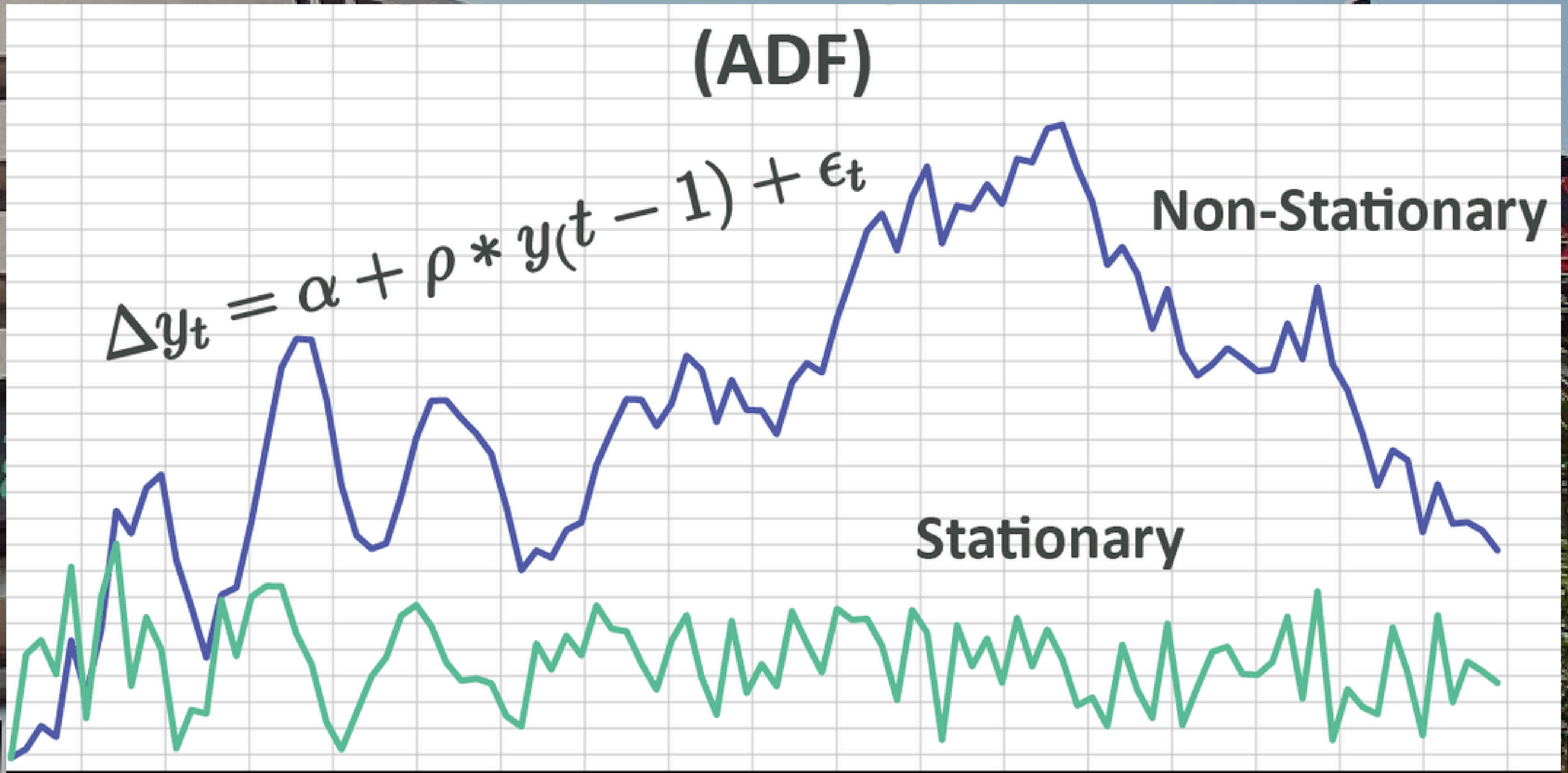


- **Both ACF and PACF values lie within the confidence bands (no significant spikes), indicating the residuals are white noise. (p=0) & (q=0)**
- **ARIMA (or time series) model has adequately captured all autocorrelation in the data, and no further AR (p) or MA (q) terms are needed.**

SEASONALITY DETECTION

A STATIONARY TIME SERIES IS ONE WHERE STATISTICAL PROPERTIES — LIKE THE MEAN AND VARIANCE — ARE CONSTANT OVER TIME.

- **KPSS test ($p=0.01 < 0.05$)** : Rejects null hypothesis of stationarity The Sensex data is non-stationary (contains trend/seasonality).(5%:0.146)
- **ADF test ($p=0.0 < 0.05$)** : Rejects null hypothesis of unit root. The data becomes stationary after differencing. (ADF Statistic: -21.8)
- **Ljung-Box test (all p-values=0.0)** : Strong autocorrelation remains. There is significant seasonality/serial correlation in the data.



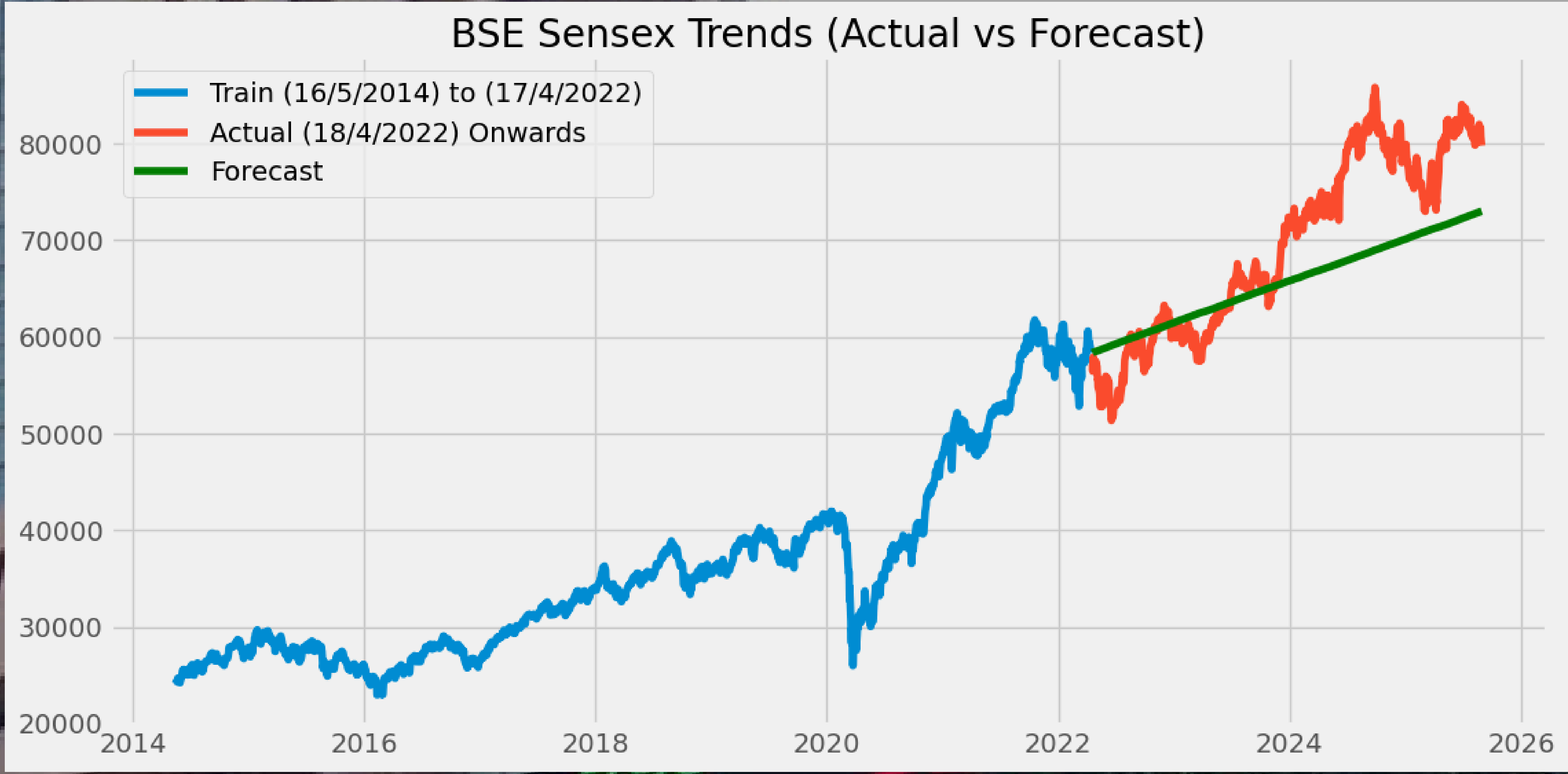
KPSS) TEST FIGURES OUT IF A TIME SERIES IS STATIONARY AROUND A MEAN OR LINEAR TREND,

AutoArima

ARIMA(0,1,0)

$$Y_t = Y_{t-1} + \epsilon_t$$

Combination with minimum AIC specifies the optimal lags



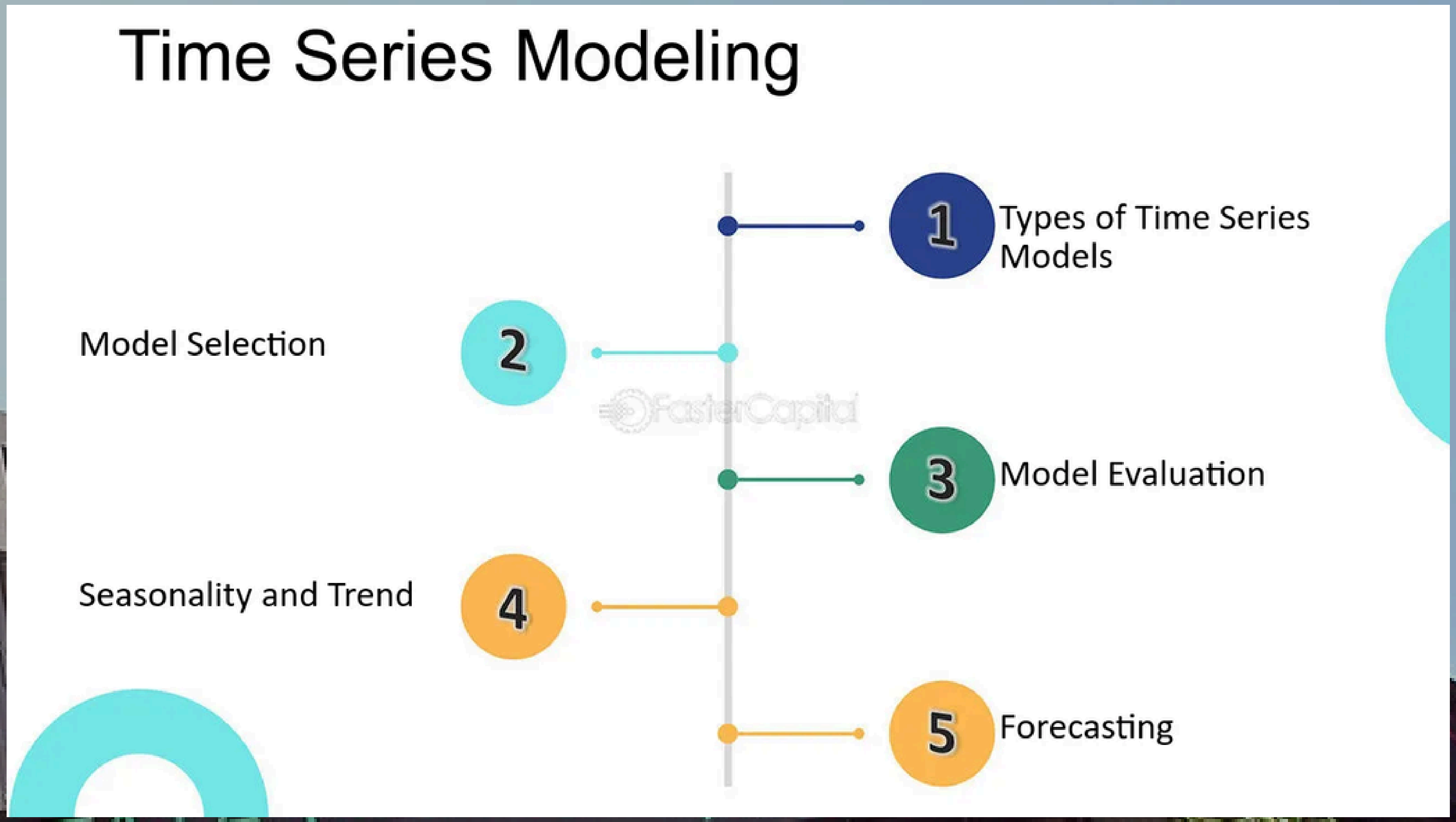
- **Best model: ARIMA(0,1,0)(0,0,0)[0] intercept**
- **AIC : 28679**

- **RMSE:6701.636070442376**

- From **2014 to 2022 (blue line),** the BSE Sensex shows steady growth with a sharp dip in **2020 (COVID-19 impact)** followed by a strong recovery.
- The **forecast (green line)** predicts a smooth, steady upward movement till 2026.
- **Actual values (red line) after April 2022** are more **volatile**, often rising above or falling below the forecast.
- The model **captures the long-term upward direction** but **underestimates short-term fluctuations and market volatility**.
- While forecasts indicate stable growth, real markets remain unpredictable due to economic, political, and global factors.

```
                              SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 2218
Model:               SARIMAX(0, 1, 0)   Log Likelihood              -16514.127
Date:                Thu, 11 Sep 2025   AIC                          33032.254
Time:                        19:42:58   BIC                          33043.662
Sample:                             0   HQIC                         33036.421
                               - 2218
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept     17.4671      9.113      1.917      0.055      -0.394      35.328
sigma2      1.728e+05   2246.676     76.910      0.000    1.68e+05    1.77e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.21   Jarque-Bera (JB):              8372.10
Prob(Q):                              0.65   Prob(JB):                         0.00
Heteroskedasticity (H):               4.79   Skew:                            -0.84
Prob(H) (two-sided):                  0.00   Kurtosis:                        12.37
===================================================================================

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

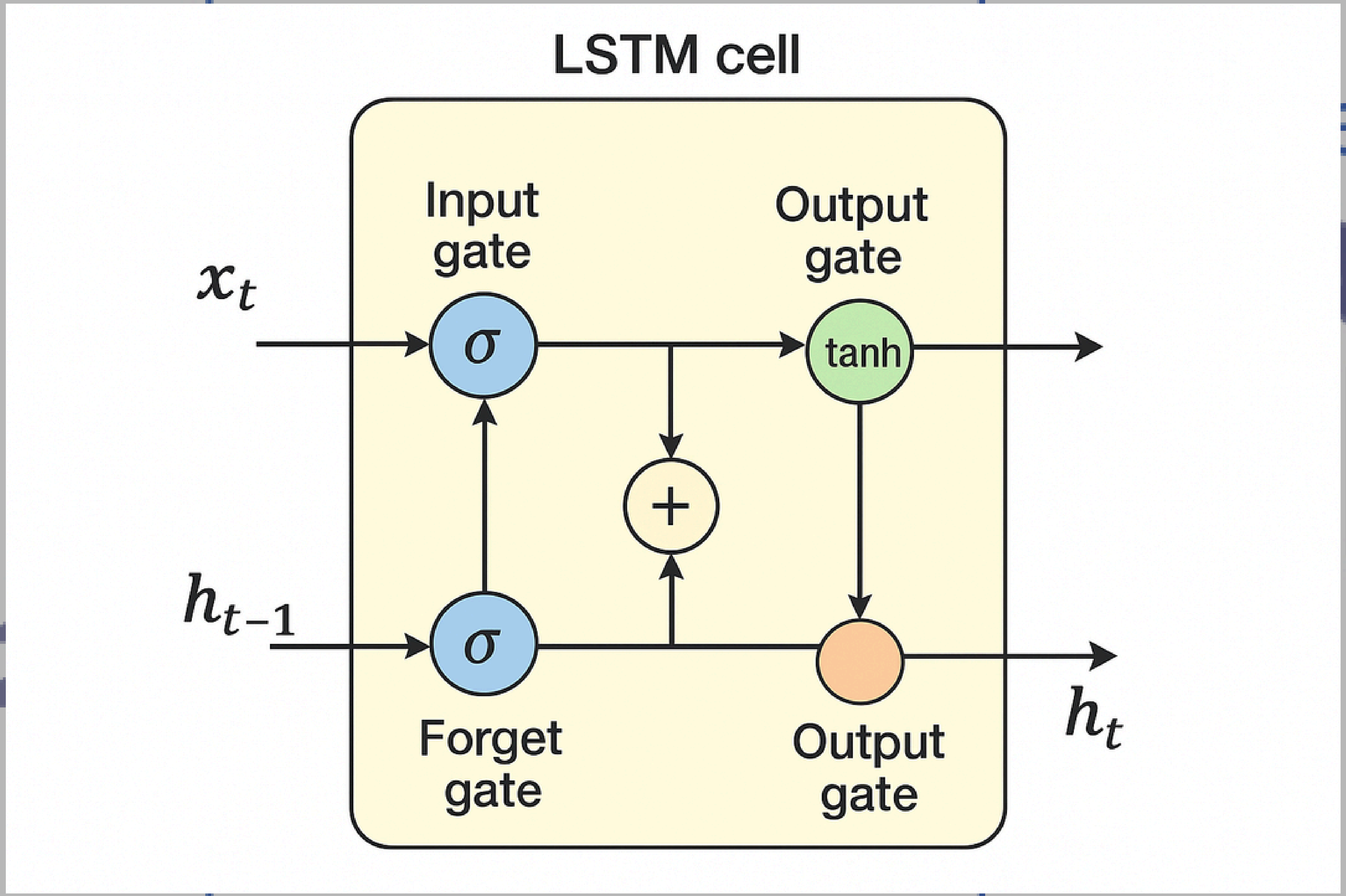


- **Selected model: SARIMAX(0,1,0) – confirms a random walk with drift process.**
- **Intercept significance: Intercept ≈ 17.47 (p=0.055) - marginally significant - reflects steady upward drift in Sensex.**
- **Error variance: Sigma² ≈ 1.73e+05, showing high market volatility in residuals.**
- **Diagnostic tests:**
- **Ljung-Box (p=0.65) - No significant autocorrelation left - Model captures patterns well.**
- **Jarque-Bera (p=0.00) - Residuals are not normally distributed (fat tails).**
- **Heteroskedasticity present (p=0.03) - Volatility clustering exists, common in financial time series.**

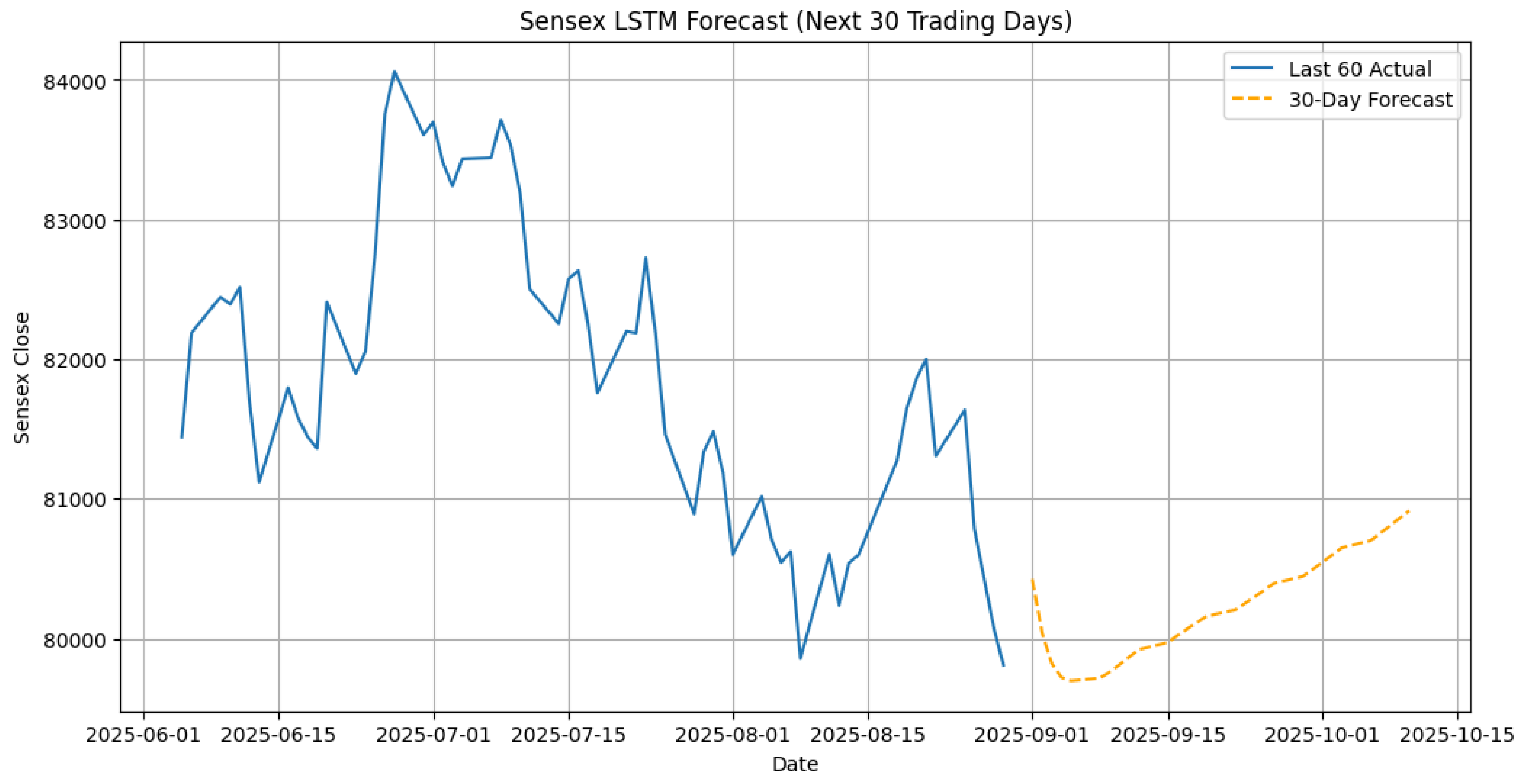
LSTM (LONG SHORT TERM MEMORY)

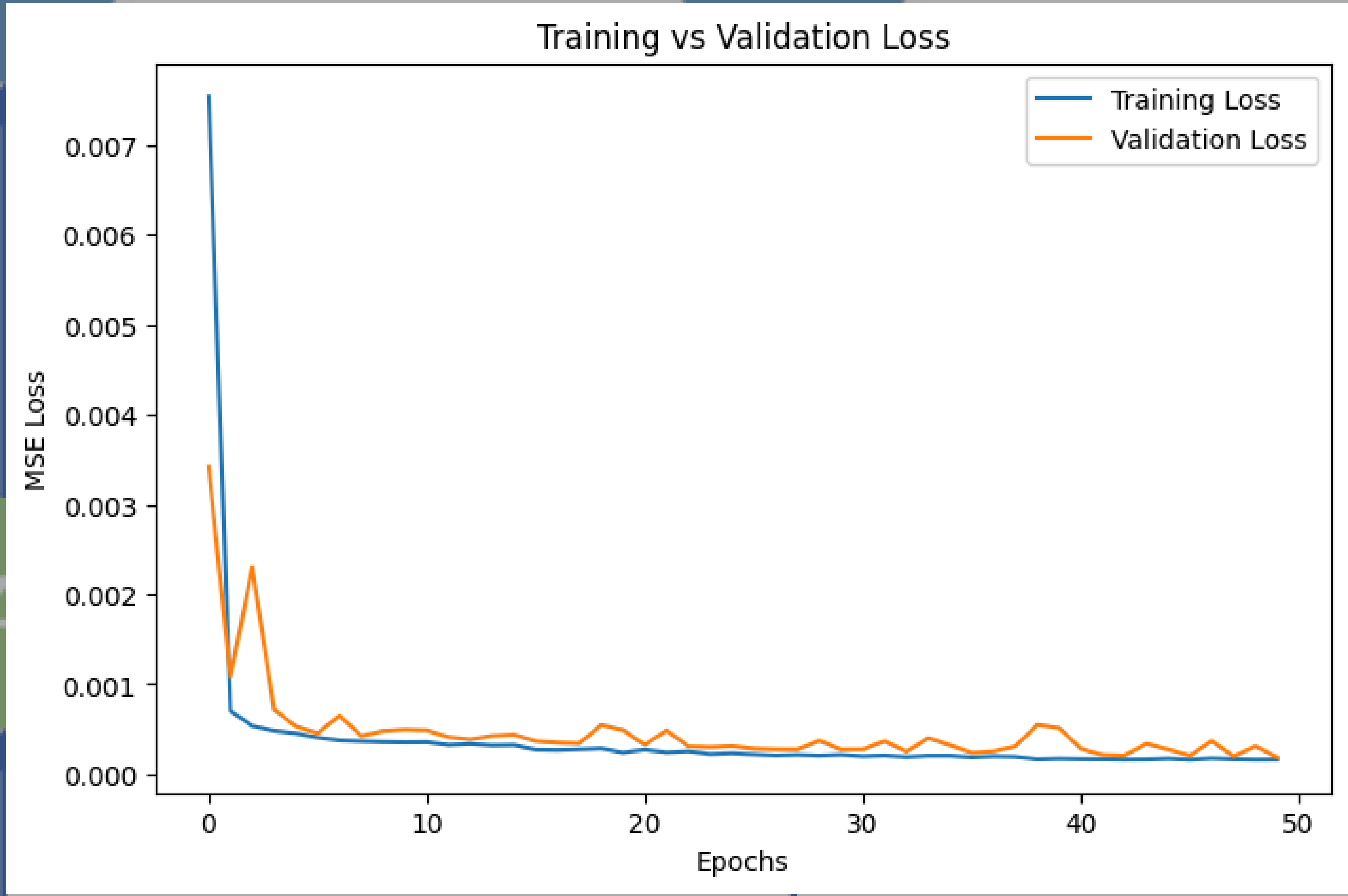
Long Short-Term Memory, is a type of deep learning neural network that excels at processing sequential data by remembering important information over long periods, overcoming the limitations of traditional recurrent neural networks (RNNs).

- LSTM layers: Single or stacked (units=50–100) with relu or tanh activations.
- Dropout: 0.2–0.3 to avoid overfitting.
- Dense layer: 1 neuron for output (next-day close).
- Optimizer: adam; Loss function: MSE.
- Model Training
- Epochs: ~50–100; Batch size: ~32.
- Validation on test split.
- Early stopping (patience=5–10) can be used to avoid overfitting.



LSTM Validation & Forecast

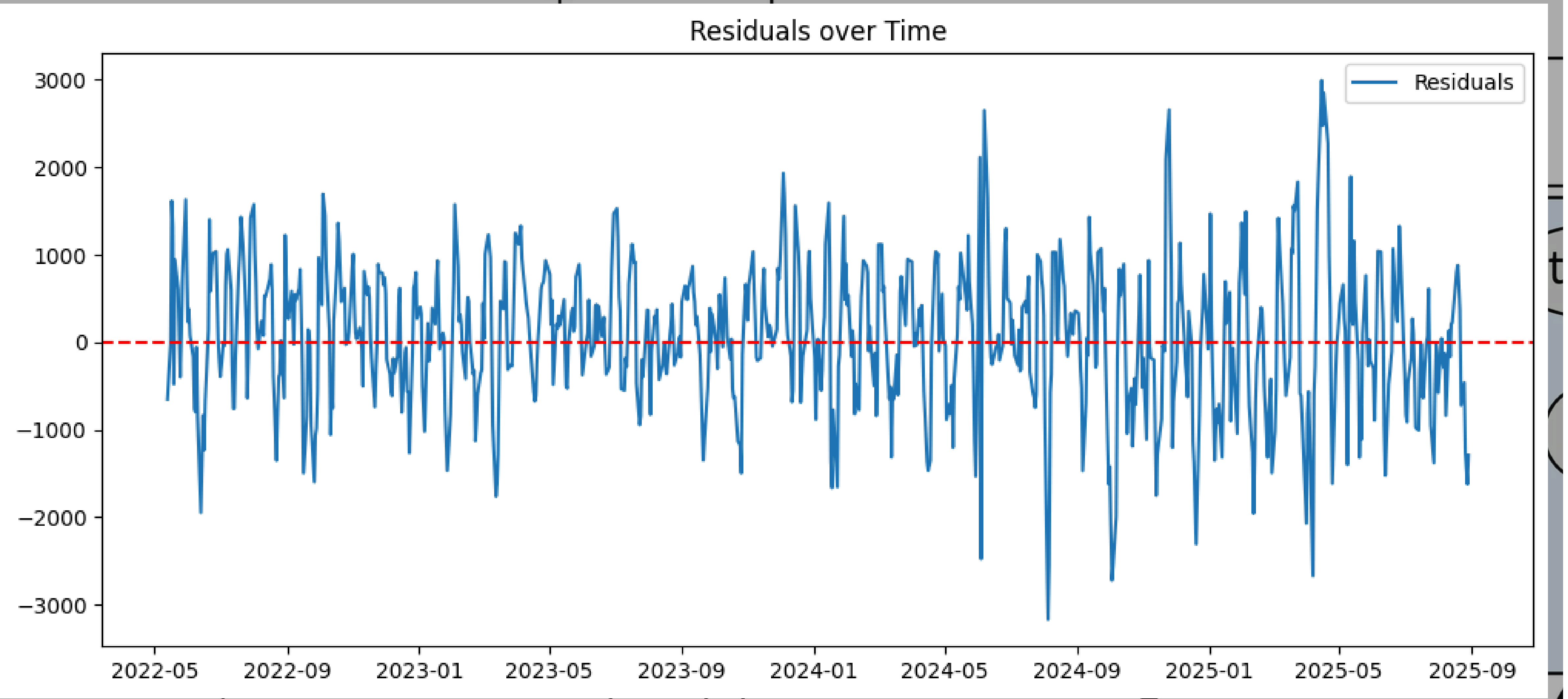


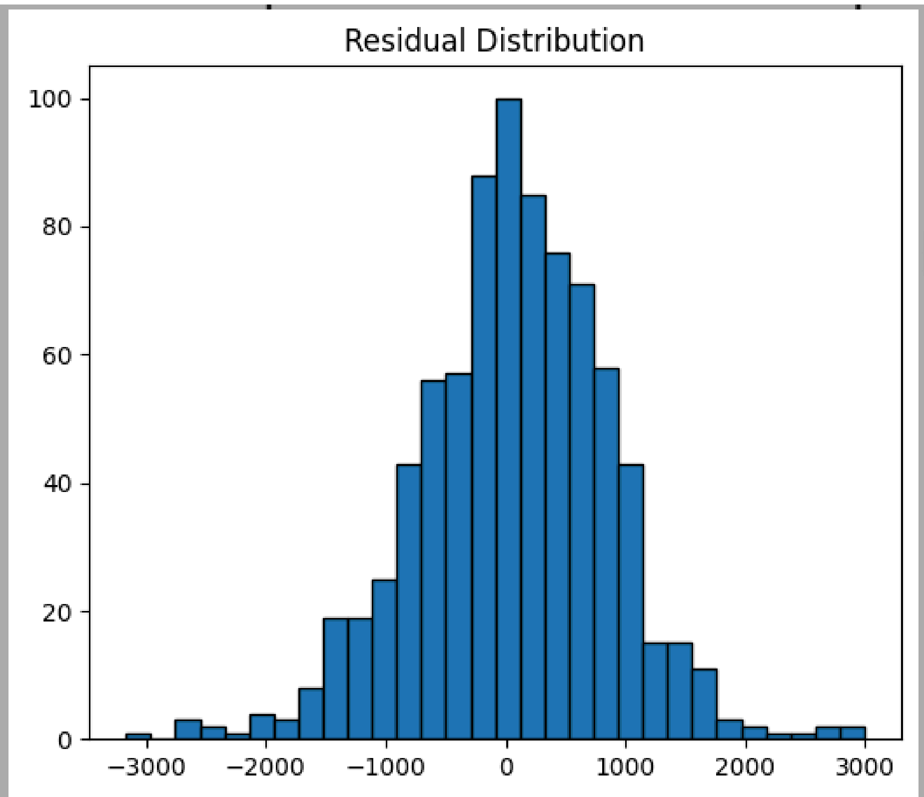


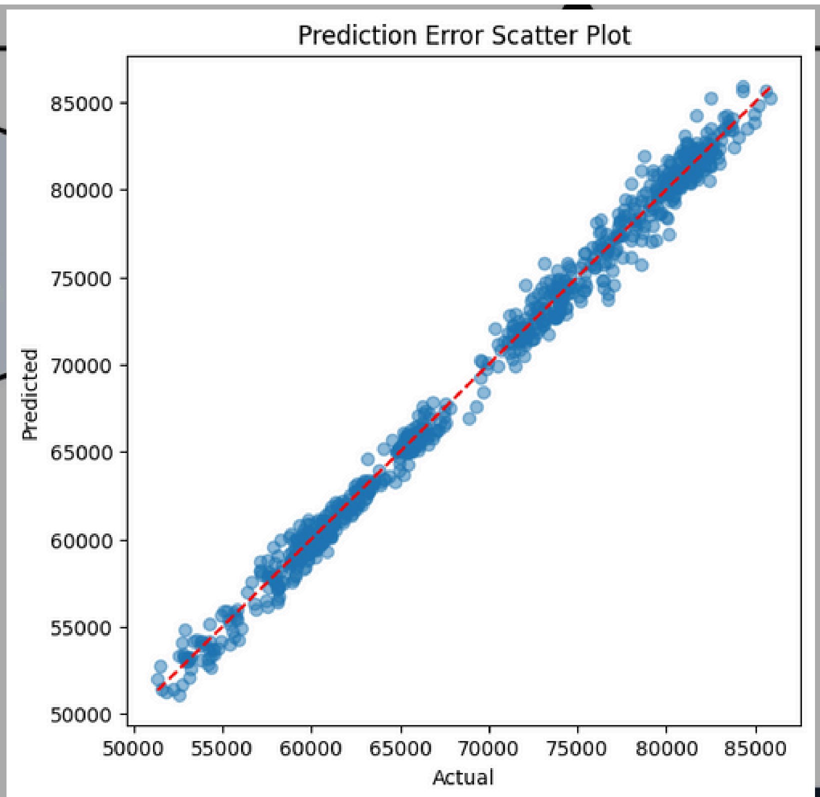
- The model captures recent downward movements before predicting recovery.
- The forecast suggests a gradual upward trend in the near term.

- Training and validation loss both converge rapidly within the first 10 epochs.
- The validation loss closely follows the training loss throughout.
- The final loss values are near zero, showing minimal error.
- No significant gap indicates the model is not overfitting.

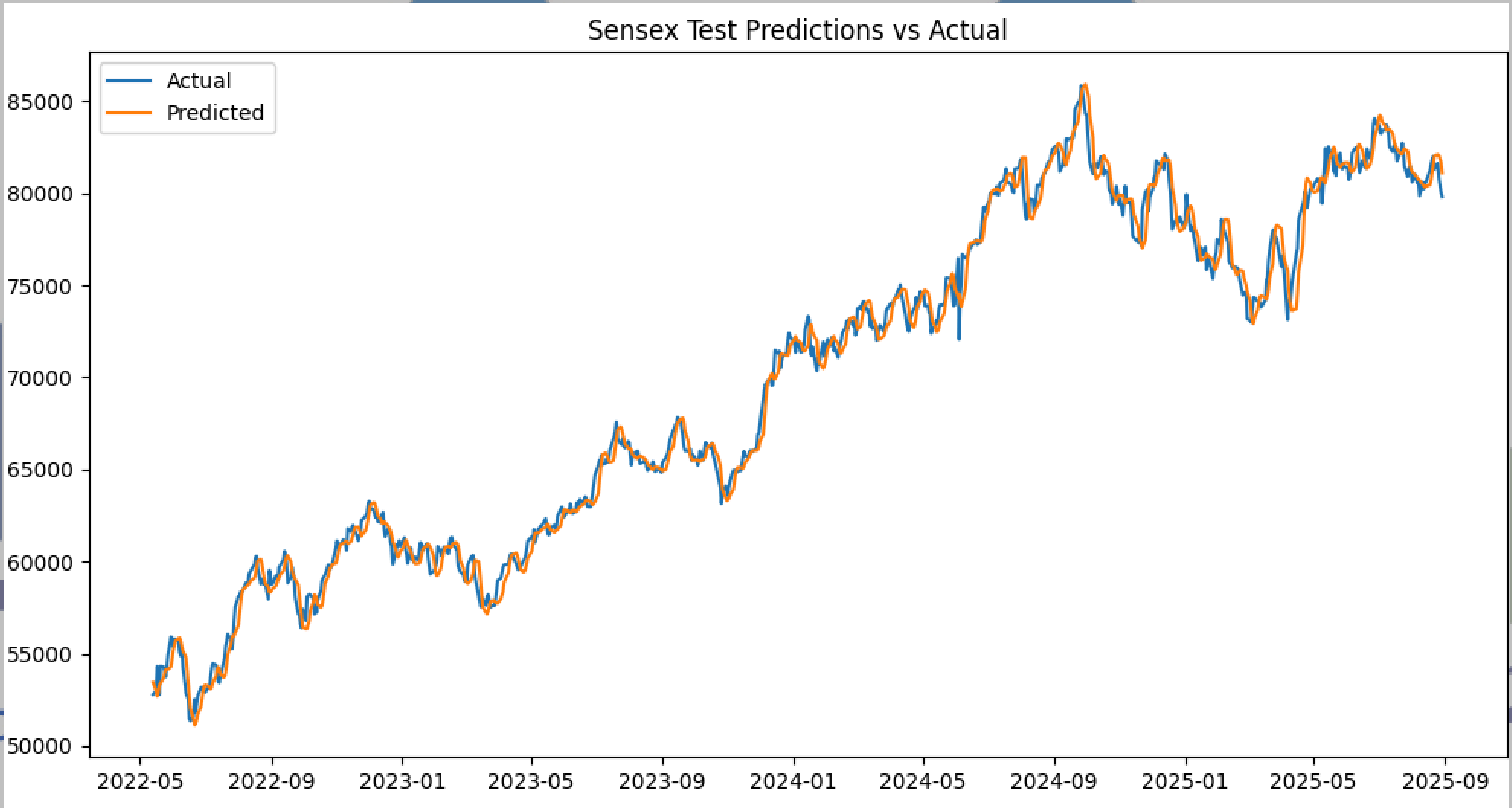
Model Evaluation







- **The spread looks reasonable, and there are no obvious extreme outliers, suggesting that the error variance is stable.**
- **A few spikes exist, but they are scattered and do not show autocorrelation.**
- **LSTM is doing an excellent job of capturing the true relationship between inputs and outputs.**
- **MAE = 620.10: On average, the model's predictions are off by about 620 units.**
- **RMSE = 805.21: Larger errors exist, but not excessively, since RMSE is close to MAE.**
- **MAPE = 0.90%: The average percentage error is less than 1%, which is extremely low.**
- **R^2 = 0.992: The model explains ~99.2% of the variance in the data, showing excellent fit.**



Key Insight: LSTM performs better than simple ARIMA-type models for stock market data, as it adapts to complex, non-linear patterns. However, it still struggles with extreme volatility.

- **Close Tracking: The LSTM forecast closely captures both upward and downward trends.**
- **Short-term Accuracy: LSTM is effective at predicting short-term fluctuations, including rises and falls, though sometimes it slightly lags behind sharp movements.**
- **Smoothing Effect: LSTM forecast captures the general direction well but underestimates sudden spikes and dips.**

REFERENCES

- https://www.nature.com/articles/s41598-024-69418-z
- https://towardsdatascience.com/exploring-the-lstm-neural-network-model-for-time-series-8b7685aa8cf/
- https://www.datacamp.com/tutorial/arima
- https://www.machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras
- https://www.sciencedirect.com/science/article/pii/S2212827121003796
- https://www.geeksforgeeks.org/machine-learning/time-series-analysis-and-forecasting/
- https://medium.com/@iamryanahmad/why-akaike-information-criterion-aic-is-a-better-tool-than-acf-and-pacf-for-identifying-lags-in-a86cf6ac63d2

Thank You