NIFTY 50 STOCK PRICE PREDICTOR

University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Sandesh Nonavinakere Sunil, Nishan Poojary

Business Intelligence and Data Analytics

Problem Description

Accurately forecasting next-day stock index values—such as the NIFTY 50—is inherently difficult due to the volatile and non-linear nature of financial markets. Existing forecasting tools largely emphasize long-term trend analysis and are often designed for expert users with technical knowledge.

This project addresses the short term prediction gap by building an easy-to-use platform that predicts the next day's opening and closing prices of the NIFTY 50 index. By integrating both ARIMA (a classical time series model) and LSTM (a deep learning-based model), the system enables side-by-side performance comparisons and model interpretability. The Streamlit-based interface allows non-expert users—like analysts, students, and retail traders—to gain quick insights from real-time data fetched via the yFinance API.

Challenges

- **Non-stationarity of data**: Financial time series data are typically non-stationary, meaning their statistical properties change over time. This poses difficulties for models like ARIMA, which assume a level of stationarity.
- **Market noise and volatility**: The stock market is influenced by a multitude of unpredictable factors—news events, macroeconomic indicators, and investor sentiment—all of which introduce noise that is hard to model.
- **Overfitting in deep learning models**: While LSTM networks can learn complex temporal patterns, they are susceptible to overfitting, especially when trained on relatively small datasets or without sufficient regularization (Zhang, 2020).
- **Static datasets limit adaptability**: Using static historical datasets (e.g., those from financial data repositories or research archives) can hinder dynamic forecasting, as newer data must be manually incorporated, delaying analysis (Gudelek, 2017).
- **Missing auxiliary features**: Common academic datasets often exclude essential fundamental indicators like P/E and P/B ratios, limiting the inclusion of broader economic context in model learning and reducing predictive potential (Ballings, 2015).

Solution

- **ARIMA model:** Used for its transparency and effectiveness in modeling linear time series data. Preprocessing steps such as differencing are applied to handle non-stationarity.
- **LSTM model:** Leveraged for its ability to capture nonlinear dependencies and temporal patterns. Techniques like dropout regularization and careful sequence windowing are used to mitigate overfitting.
- **Single-day prediction horizon:** Restricting predictions to the next trading day improves reliability and makes the system more practical for real-world use.
- **Dropout :** helps prevent overfitting in LSTM by randomly disabling 20% of the neurons during training, which forces the network to learn more robust and general patterns instead of relying too heavily on specific neurons. This improves the model's ability to generalize to unseen data.

System Setup

Hardware: Laptop/Desktop , Mouse and Keyboard (optional)

Software and Technology:

- **Development and Modeling :** Python , Pandas , NumPy , Statsmodels , PMDARIMA , Scikit-learn , SciPy , Seaborn and Matplotlib
- **Model Evaluation and Testing :** Pytest
- **Data Source :** YFinance
- **Deployment and UI :** Streamlit
- **Operating Systems :** Windows OS or Mac OS

Dataset

The dataset used in this study was programmatically retrieved via the Python yfinance API from Yahoo Finance's historical data for the NIFTY 50 Index. It spans from January 1, 2008, covering 4,000+ trading sessions across various market conditions. This automated source ensures real-time adaptability, reproducibility, and relevance for short-term financial forecasting.

Results

This plot compares the actual NIFTY 50 opening prices with predictions from both ARIMA and LSTM models. The ARIMA model closely tracks actual prices, showing strong short-term alignment, while the LSTM model exhibits more volatility and larger deviations. This visually supports the conclusion that ARIMA performs better for short-term stock index forecasting in our project.

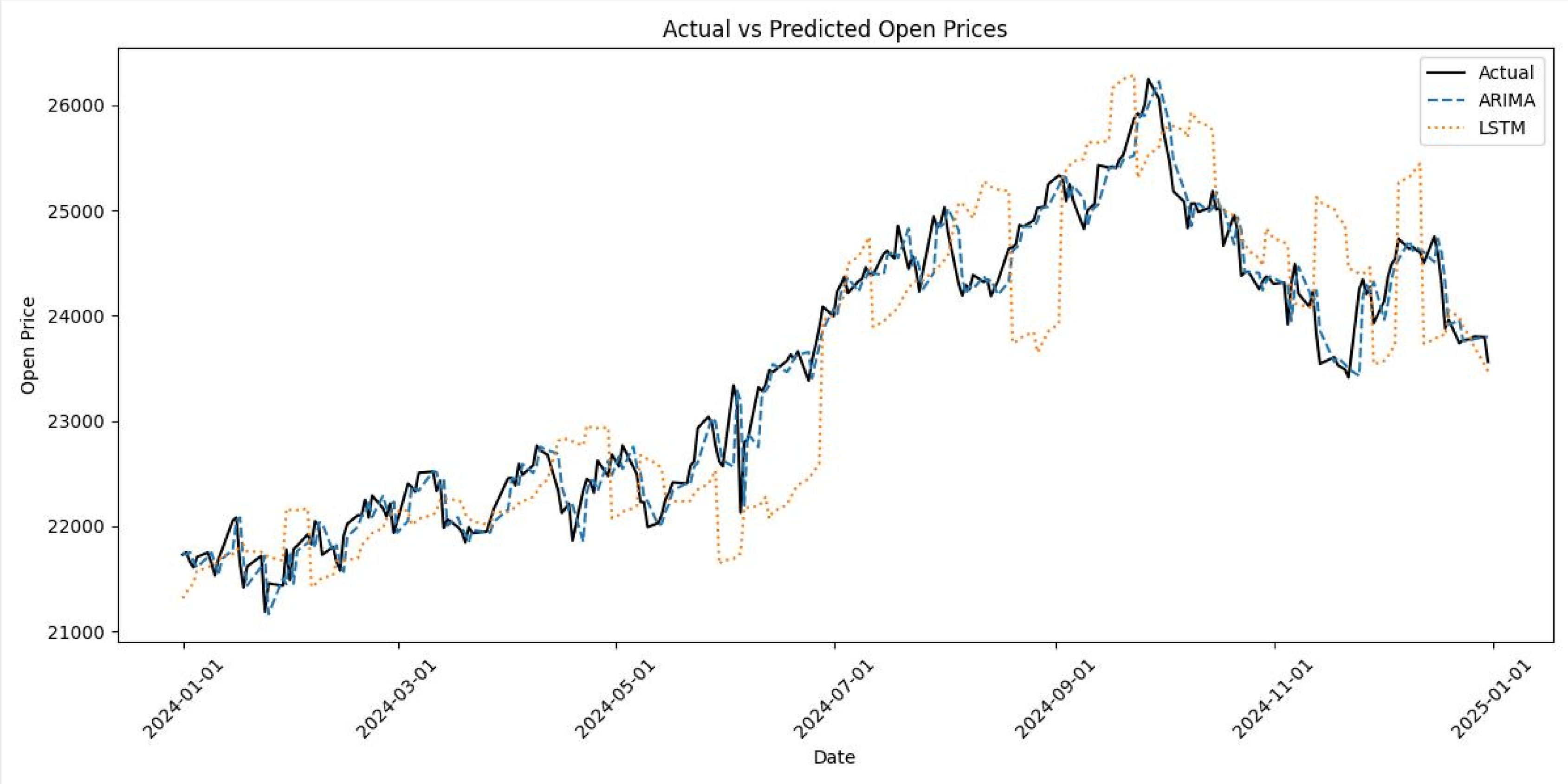

Figure 1: Actual vs Predicted Open

The comparative performance analysis between ARIMA and LSTM models for predicting NIFTY 50 index prices shows that ARIMA outperforms LSTM for both opening and closing price forecasts.

```
====== Tomorrow's Predicted Prices ======
Open - ARIMA: 21737.6504, LSTM: 21312.6850
Close - ARIMA: 21713.8015, LSTM: 21410.5345

====== Open Forecast Results ======
ARIMA -> RMSE: 215.73, MAPE: 0.01
LSTM -> RMSE: 623.52, MAPE: 0.02

====== Close Forecast Results ======
ARIMA -> RMSE: 208.23, MAPE: 0.01
LSTM -> RMSE: 608.60, MAPE: 0.02
```

Figure 3: Model Output

These results indicate that ARIMA not only provides more accurate forecasts with significantly lower RMSE and MAPE values, but it also performs better in a short-term context, making it a more reliable model for next-day stock price prediction in this project. Despite being simpler, ARIMA delivers higher precision, especially in volatile financial data scenarios.

Future Work

- **Improved Data Handling:** Introduce advanced imputation and anomaly detection methods for missing or irregular data to maintain input quality.
- **Add Financial Indicators:** Incorporate P/E, P/B, EPS, and sentiment data to capture broader market influences and improve forecast accuracy.
- **Outlier and Disruption Management:** Develop strategies to detect and manage outliers and market shocks, testing different treatments for impact on results.

References

- Zhang, Z., Zheng, Y. and Qi, D., 2020. Deep learning for forecasting stock returns: A long short-term memory approach. IEEE Transactions on Neural Networks and Learning Systems, 31(7), pp.2263–2274.
- Gudelek, M.U., Bolukbasi, F. and Kocamaz, A.S., 2018. A deep learning based stock trading model with 80% accuracy rate. Procedia Computer Science, 140, pp.150–157.
- Ballings, M., Van den Poel, D., Hespeels, N. and Gryp, R., 2015. Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications, 42(20), pp.7046–7056.