Hochschule Emden/Leer
Abteilung Maschinenbau
MSR-Labor

Nifty 50 Stock price Prediction

Using ARIMA and LSTM

Sandesh Nonavinakere Sunil , Nishan Chandrashekar Poojary

26. Juni 2025

1. Problem Description
2. Application Description
3. Dataset
4. Methodology
5. Challenges
6. Solutions
7. Result
8. References

Problem Description

University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Problem Description I

- In recent years, financial forecasting has seen a paradigm shift with the increasing adoption of machine learning and deep learning techniques.
- With the expansion of algorithmic trading and accessi- ble computing resources, even individual investors now seek tools that assist in making informed trading decisions.
- Predicting next-day opening and closing prices of indices like NIFTY 50 remains challenging due to market volatility, data limitations, and model complexity.

Application Description

Application Description I

- 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.

Dataset

Dataset I

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.

Click this to view the data

Methodology

Methodology I

1. Input Data: yFinance API (Open, Close)

↓

2. Preprocessing: Clean, scale, COVID dummy

↓

3. Model Selection: ARIMA, LSTM (Keras)

↓

4. Training: ARIMA: .pkl, LSTM: .keras

↓

5. Prediction: Forecast next-day prices

↓

6. Evaluation: RMSE, MAPE, visual plots

↓

7. Deployment: Streamlit GUI, API fetch

↓

8. Logging & Docs: Log errors, Doxygen

Abbildung: NIFTY 50 Forecasting Pipeline

Challenges

University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Challenges I

Stock market forecasting involves several inherent challenges, many of which are amplified in short-term prediction:

- **Non-stationarity of data**: Financial time series data are typically non-stationary, meaning their statistical properties change over time [Tsa10]. 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.

Challenges I

- **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 [ZZQ20].
- **Interpretability**: Deep learning models often function as black boxes, offering little interpretability compared to traditional models, which can be a barrier to trust in high-stakes financial applications [Lip16].

Challenges I

- **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 [GBK18].
- **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 [Bal+15].

Solutions

Solutions I

- **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.
- **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.

Solutions I

- **Single-day prediction horizon:** Restricting predictions to the next trading day improves reliability and makes the system more practical for real-world use.
- **Automated Data Retrieval:** To overcome the limitations of static datasets and improve adaptability, the system integrates the yfinance API, enabling automated access to up-to-date NIFTY 50 stock market data for continuous model input.

Result

Result I

- The ARIMA model outperformed the LSTM model in terms of both Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), affirming its suitability for structured, stationary time series like index-level market data.
- While LSTM is often celebrated for handling complex, nonlinear dependencies, in this use case, ARIMA proved more effective due to the predictable nature of market index movements and limited feature dimensionality.
- **Click here to view the Application**

Result I

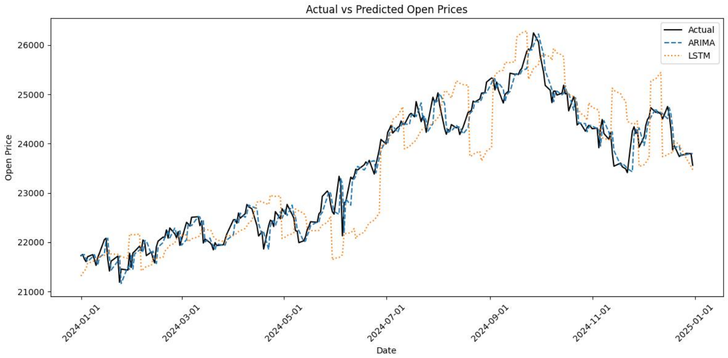

Abbildung: Actual vs Predicted Open

Result I

```
====== 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
```

Abbildung: ModelOutput

Vielen Dank

für Ihre Aufmerksamkeit

References

References I

[Bal+15] Michel Ballings u. a. “Evaluating multiple classifiers for stock price direction prediction”. In: *Expert Systems with Applications* 42.20 (2015), S. 7046–7056.

[GBK18] M. U. Gudelek, F. Bolukbasi und A. S. Kocamaz. “A Deep Learning Based Stock Trading Model with 80% Accuracy Rate”. In: *Procedia Computer Science* 140 (2018), S. 150–157.

[Lip16] Zachary C. Lipton. “The Mythos of Model Interpretability”. In: *arXiv preprint arXiv:1606.03490* (2016).

[Tsa10] Ruey S. Tsay. *Analysis of Financial Time Series*. John Wiley und Sons, 2010.

[ZZQ20] Zhe Zhang, Yuying Zheng und Dong Qi. “Deep Learning for Forecasting Stock Returns: A Long Short-Term Memory Approach”. In: *IEEE Transactions on Neural Networks and Learning Systems* 31.7 (2020), S. 2263–2274.