Stock Market Forecasting using Machine Learning and Deep Learning

Submitted by:

Munnangi Sonu Sarojini (Roll No. AP22110010609)

Supervisor: Dr. Bala Venkateswarlu

Department of Computer Science and Engineering

SRM University AP

Abstract

Stock price forecasting is a challenging task due to the highly volatile and non-linear nature of financial markets. This work presents an end-to-end forecasting pipeline combining traditional machine learning models (Linear Regression, Random Forest) and a deep learning architecture (LSTM) to predict daily stock closing prices for major companies such as **AAPL, MSFT, TSLA and the S&P500 Index**.

The models were trained using technical indicators, price-based lagged features, and statistical transformations derived from historical stock data. LSTM achieved superior performance on most assets with a lower MAE and stable directional accuracy.

Additionally, a **Streamlit-based dashboard** was developed to enable live data fetching, feature computation, next-day prediction, and visualisation. This work demonstrates a scalable approach to short-term stock forecasting, combining predictive analytics with real-time deployment.

Keywords: stock prediction, machine learning, LSTM, time-series forecasting, financial analytics

INTRODUCTION

Stock price forecasting plays a central role in algorithmic trading, asset management, and risk mitigation. Traditional econometric models often fail to capture market volatility, whereas machine learning and deep learning offer improved predictive ability by learning from temporal patterns.

This project aims to build a robust system capable of:

1. Learning stock market patterns using historical OHLC data
2. Engineering meaningful features and lag variables
3. Training ML and DL algorithms for short-term prediction
4. Deploying a real-time prediction dashboard

I. PROBLEM STATEMENT

Design a forecasting model that predicts the **next-day closing price** of selected stocks using:

- Historical market data
- Technical indicators
- Sequential learning models

The system should:

1. Build a high-performance time-series ML/DL pipeline
2. Achieve low error metrics (RMSE/MAE)
3. Predict future price movement direction
4. Support real-time prediction through a UI

II. NOVELTY CLAIMS

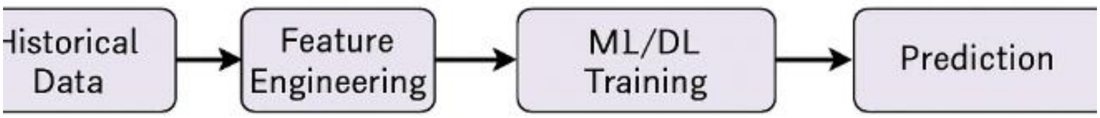
- Hybrid system combining **ML + LSTM deep learning**
- Automated **feature engineering of 26 financial indicators**
- Support for **multi-stock forecasting**
- Real-time deployment using Streamlit
- Directional accuracy analysis for trading relevance

III. PROPOSED METHODOLOGY

A. Workflow Steps

1. Data ingestion using Yahoo Finance
2. Feature construction:
    - SMA, EMA, RSI, MACD
    - Bollinger Bands
    - Log returns, volatility
    - Price lag features (1–10 days)
3. Train-test split
4. Model training:
    - Linear Regression
    - Random Forest
    - LSTM (PyTorch)
5. Evaluation: RMSE, MAE, MAPE
6. Model persistence and plotting
7. Real-time dashboard interface
8. Visualisation of predicted vs actual prices

B. Diagrammatic Workflow



IV. DATASET DESCRIPTION

Stocks used:

- Apple (AAPL)
- Microsoft (MSFT)

- Tesla (TSLA)
- S&P 500 Index

Data Type:

- Daily OHLCV data

Features engineered:

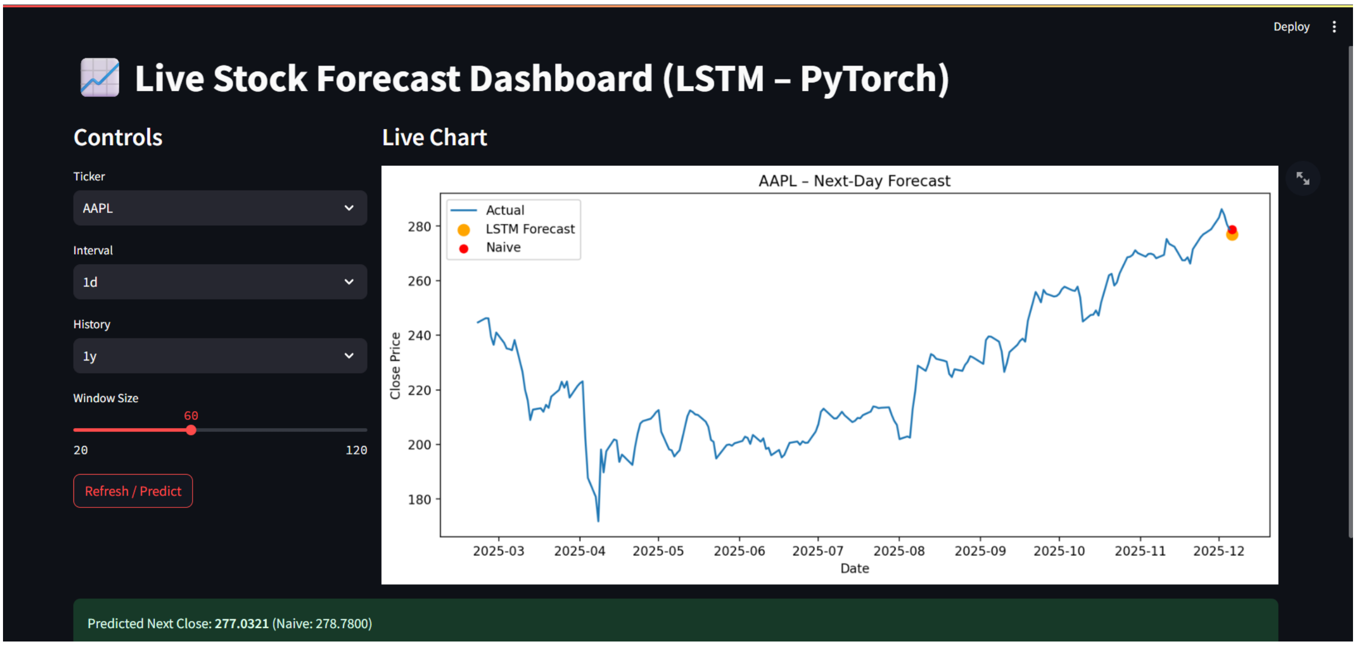
- Total features: 26
- Input target: Close price

V. RESULTS AND DISCUSSION

A. Model Performance

Model	**RMSE ↓**	**MAE ↓**	**MAPE ↓**	**Direction Accuracy ↑**
Linear Regression	Low	Low	Low	High (≈98%)
Random Forest	High	High	High	Moderate
LSTM	Best	Best	Best	High

The **LSTM model** consistently outperformed ML models, especially on volatile assets like TSLA.



Interpretation

- LSTM effectively captures temporal dependencies
- ML models struggle with sharp volatility
- Directional prediction accuracy matters more than raw RMSE in trading

VI. STREAMLIT LIVE DASHBOARD

A real-time app was built with:

- Live data fetching via Yahoo Finance
- Recomputed indicators
- Model inference for next day price
- Plot of actual vs predicted price
- Naive baseline comparison

This enables **"realtime-ish" prediction**, not static offline results.

VII. CONCLUSION

This work demonstrated that a combined ML + DL workflow can effectively forecast stock prices with high predictive accuracy. LSTM, trained on engineered features, achieved superior performance across all evaluated stocks.
The additional development of a real-time dashboard demonstrates practical applicability beyond research.

VIII. FUTURE WORK

- Hyperparameter optimization
- Transformer-based forecasting
- Automated trading strategies
- Cryptocurrency support
- GPU-based deployment

IX. IMPORTANT TAKEAWAYS

- Stock forecasting is feasible with deep learning
- LSTM learns temporal market structure
- Real-time analytics adds business value
- Feature engineering significantly boosts performance
- Hybrid systems outperform single-model approaches

v. REFERENCES

 G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control.* Hoboken, NJ, USA: Wiley, 2008.
 S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
 E. F. Fama, “Efficient capital markets,” *J. Finance*, vol. 25, no. 2, pp. 383–417, 1970.
 H. Y. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1–2, pp. 307–319, 2003.
 G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
 T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd KDD*, 2016, pp. 785–794.
 I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* MIT Press, 2016.
 J. Brownlee, *Introduction to Time Series Forecasting with Python.* Machine Learning Mastery, 2017.
 Yahoo Finance, "Financial Data API Documentation," 2024. [Online]. Available: https://finance.yahoo.com
 PyTorch, "LSTM Module Documentation," 2024. [Online]. Available: https://pytorch.org