



Stock Price Prediction Using Machine Learning

Shaliny Paramesvaran¹, Dr. G. Paavai Anand²

M.Tech, 1st Year CSE, SRMIST, Vadapalani, Chennai, India¹

Asst. Professor, CSE, SRMIST, Vadapalani, India²

Abstract: Stock price prediction is one of the most challenging tasks in financial analysis due to the market's highly volatile and nonlinear nature. In this study, we propose a machine learning-based approach for predicting stock prices using historical data. The model utilizes regression-based and deep learning algorithms to capture temporal patterns and market trends. Specifically, we employ Linear Regression, Random Forest, and a Long Short-Term Memory (LSTM) network to model the time-series behavior of stock prices. The LSTM model is particularly suited for this task as its architecture allows it to effectively learn and remember long-term dependencies inherent in sequential financial data.

The project aims to assist investors and analysts in making informed decisions by forecasting future stock prices with reasonable accuracy. The performance of all models is rigorously evaluated using standard metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 -score, on a held-out test set. The results demonstrate that the Long Short-Term Memory (LSTM) network significantly outperforms traditional regression models in capturing sequential dependencies in stock market data, achieving the highest R^2 score of 0.93 and the lowest RMSE of 0.018. This robust performance underscores the suitability of deep learning for complex financial forecasting.

Keywords: Stock prediction, Sequential Financial Data, Temporal patterns, Forecasting, Deep learning, LSTM.

I. INTRODUCTION

Stock market prediction has gained significant attention from researchers and investors alike due to its potential in financial planning and investment strategies. The ability to forecast future stock movements offers a considerable advantage in optimizing portfolio returns and managing risk. However, stock prices are influenced by a myriad of complex factors, making this task notoriously challenging. These factors include macroeconomic indicators (like interest rates and GDP growth), company-specific performance (like earnings and debt), investor sentiment, and overall market volatility.

Traditional statistical models, such as Autoregressive Integrated Moving Average (ARIMA) or simple linear regression, often fail to capture these complex, nonlinear relationships and the time-dependent structure inherent in financial time series data. These models assume stationarity and linearity, which rarely hold true in dynamic stock markets.

To overcome these limitations, the financial sector has increasingly turned to machine learning (ML) and deep learning (DL) algorithms. These advanced techniques can model intricate dependencies efficiently. This project specifically focuses on applying sophisticated machine learning techniques to historical stock market data to forecast future prices. We utilize both traditional regression models (Linear Regression and Random Forest) as baselines, alongside the state-of-the-art deep learning architecture, the Long Short-Term Memory (LSTM) network. The project aims to analyze the comparative performance of these models, demonstrating the superior ability of LSTMs to handle the sequential and temporal nature of stock price movements. This analysis provides valuable insights into the efficacy of deep learning for practical financial forecasting applications.

II. LITERATURE SURVEY

The existing body of research provides a strong foundation for the application of advanced computational methods in stock price prediction, moving beyond traditional statistical models.

Early success in applying soft computing methods was demonstrated by using fuzzy logic systems to predict stock trends with improved accuracy [1]. A crucial comparison of machine learning algorithms, including Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Random Forest, was conducted, concluding that the use of ensemble learning techniques improves predictive stability [2].

A pivotal contribution came from Fischer & Krauss who demonstrated the effectiveness of Long Short-Term Memory



(LSTM) networks in capturing temporal dependencies in financial time series data, setting a key benchmark for deep learning approaches [3]. Building on this, a study utilized CNN and LSTM hybrid architectures to analyze stock price movements based on technical indicators [4]. Further, research found that a hybrid CNN-LSTM model exhibited superior accuracy and performance compared to standalone RNN, CNN, MLP, and LSTM models [5], highlighting the synergistic benefits of multi-architecture integration.

More recently, research has introduced sophisticated architectures for greater robustness. A novel hybrid model, LSTM-mTrans-MLP (LSTM-modified Transformer-Multi-Layer Perceptron), was proposed, integrating LSTM for long-term sequence capture and a Transformer with a self-attention mechanism [6]. For modeling inter-stock dependencies, an innovative LSTM-Graph Neural Network (GNN) hybrid model was introduced, achieving a significant reduction in Mean Squared Error (MSE) over a standalone LSTM [7]. A comprehensive comparison of LSTM and Gated Recurrent Unit (GRU) models augmented with technical indicators consistently found that LSTM generally achieves a slightly lower error [8]. To address this trade-off, a Regularized GRU-LSTM hybrid approach was introduced to leverage the strengths of both [9].

Beyond purely historical prices, the importance of incorporating external context has been reinforced. One study addressed stock trend prediction by integrating news sentiment analysis with technical indicators [10]. Similarly, a methodology combining a genetic algorithm with a Random Forest classifier demonstrated the crucial role of feature optimization [11]. An investigation using LSTM and Random Forest models augmented by Sentiment Analysis from news headlines and macro-economic parameters showed a marked improvement in accuracy [12]. The application of Deep Reinforcement Learning (DRL) was advanced by a novel method that incorporates community-aware sentiments and a knowledge graph [13]. The growing role of large language models in capturing market sentiment was highlighted by the use of FinBERT and GPT-4 alongside Logistic Regression for stock index prediction [14]. Finally, a comparative study examined the effectiveness of Decision Trees, Random Forests, and SVM using a dataset of insider trading transactions, focusing on recursive feature elimination (RFE) [15].

The application of Reinforcement Learning (RL) for optimizing trading strategy was initially proposed by Lee (2001) [16]. Most recently, an innovative DRL framework was presented that integrates NLP and historical data into a Deep Q-Network (DQN) to make optimal sequential decisions, focusing on real-world profit maximization [17].

These studies consistently highlight that machine learning models, particularly deep neural networks, outperform traditional statistical methods such as ARIMA or linear regression when tasked with predicting dynamic, nonlinear stock market data.

These studies highlight that machine learning models, particularly deep neural networks, outperform traditional methods such as ARIMA or linear regression when predicting dynamic stock data.

III. METHODOLOGY

The proposed methodology involves several stages:

1. Data Collection: Historical stock price data is collected from Yahoo Finance or Kaggle.
[https://finance.yahoo.com/quote/AAPL/history\\$](https://finance.yahoo.com/quote/AAPL/history$)
2. Data Preprocessing: Cleaning, normalization, and feature extraction (Open, High, Low, Close, Volume, and moving averages).
3. Feature Engineering: Creating additional features such as daily returns, moving averages, and relative strength index (RSI).
4. Model Building: Training machine learning models such as Linear Regression, Random Forest, and LSTM.
5. Evaluation: Using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 -score to evaluate performance.
6. Prediction: Forecasting future stock prices and visualizing results.

IV. DATASET DESCRIPTION

The dataset consists of daily historical stock price data for selected companies (e.g., Apple Inc., Google, or NIFTY 50).

Attributes include:

- Date: Trading date
- Open: Price at the beginning of the trading day



- High: Highest price during the day
- Low: Lowest price during the day
- Close: Price at the end of the trading day
- Volume: Number of shares traded

Dataset Source: Yahoo Finance / Kaggle

V. DATA COLLECTION AND PREPROCESSING

Data is collected using Python's yfinance library. Missing values are handled using forward fill, and outliers are detected through z-score analysis.

Normalization is performed using Min-Max scaling to bring all features into the range [0,1]. The dataset is split into training (80%) and testing (20%) sets.

A sequence of past 60 days is used to predict the next day's stock price when using LSTM.

Implementation

The implementation is done in Python using Jupyter Notebook. Libraries used:

- pandas, numpy for data processing
- matplotlib, seaborn for visualization
- scikit-learn for machine learning models
- tensorflow / keras for deep learning models

Model Steps:

1. Build baseline models (Linear Regression, Random Forest).
 2. Construct LSTM model with layers: 50 LSTM units → Dropout (0.2) → Dense(1).
 3. Compile model using Adam optimizer and Mean Squared Error loss.
 4. Train model with batch size 32 for 100 epochs.
- Evaluate models using RMSE and visualize predictions vs. actual values.

VI. RESULT AND DISCUSSION

The LSTM model achieved a lower RMSE compared to traditional models:

Table 1. Metrics Comparison of Model Performance

Model	RSME	RME	R ² Score
Linear Regression	0.034	0.027	0.85
Random Forest	0.028	0.022	0.89
LSTM	0.018	0.014	0.93

The results indicate that the LSTM model effectively captures temporal dependencies, leading to better predictive performance. The predicted price trends closely align with actual market movements. Visual analysis of predicted vs. actual data confirms the robustness of the deep learning approach.

VII. CONCLUSION

This project successfully demonstrates that machine learning and deep learning models can be effectively applied to the highly challenging task of stock price prediction, yielding forecasts with good accuracy. The primary objective of applying these advanced algorithms to historical financial time series data was achieved, providing a clear comparison between traditional and deep learning approaches.

The empirical results conclusively show that the Long Short-Term Memory (LSTM) network provides superior performance compared to the traditional Linear Regression and Random Forest models. Specifically, the LSTM's lower Root Mean Squared Error (RMSE) of 0.018 and higher R² score of 0.93 validate its effectiveness in capturing the sequential dependencies and complex, nonlinear temporal patterns inherent in stock market data. The internal



gating mechanism of the LSTM is crucial for remembering relevant information over long sequences, a feature that simple regression models lack.

However, the conclusion must be tempered by the recognition that perfect prediction remains impossible. External factors, such as sudden political events, global economic shocks, and shifts in market sentiment, are not fully captured by the historical price data alone and still introduce significant uncertainty in all predictions.

Therefore, future work should focus on integrating multi-modal data sources to enhance the prediction model's robustness. This includes incorporating sentiment analysis of financial news, social media data, and macro-economic indicators (as suggested in the literature) to provide the model with a richer, contextual understanding of market drivers, which could further minimize predictive error and align forecasts more closely with real-world market volatility.

REFERENCES

1. Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert Systems with Applications*, 36(2), 5932–5941.
2. Abraham, M. S., Jose, M. S., Kunjupillai, M. L., & Jacob, L. (2025). Artificial Intelligence Models for Predicting Stock Returns Using Fundamental, Technical, and Entropy-Based Strategies: A Semantic-Augmented Hybrid Approach. *PMC - PubMed Central*.
3. Altuner, A. B., & Kilimci, Z. H. (2022). A novel deep reinforcement learning based stock price prediction using knowledge graph and community aware sentiments. *Turkish Journal of Electrical Engineering and Computer Sciences*, 30(4), 1506–1524.
4. Barua, R., et al. (2025). A Comparative Analysis of LSTM and GRU Models Enhanced with Technical Indicators for Stock Forecasting. *ResearchGate*.
5. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(1), 58–67.
6. Guo, Q., et al. (2023). LSTM–Transformer-Based Robust Hybrid Deep Learning Model for Financial Time Series Forecasting. *MDPI*.
7. Kabbani, M., & Usta, O. (2025). A Comparative Study of Machine Learning Algorithms for Stock Price Prediction Using Insider Trading Data. *arXiv*.
8. Kaur, G., et al. (2024). Innovative Sentiment Analysis and Prediction of Stock Price Using FinBERT, GPT-4 and Logistic Regression: A Data-Driven Approach. *MDPI*.
9. Lee, T. (2001). Stock price prediction using reinforcement learning. In *2001 IEEE International Conference on Systems, Man and Cybernetics*. IEEE.
10. Lu, W., Li, J., & Cheng, X. (2020). Forecasting of Share Prices Based on Hybrid Model of CNN and LSTM: A Multi-Factor Approach. *SciTePress*.
11. Luo, X., et al. (2025). A Comparative Study of Machine Learning Algorithms for Stock Price Prediction Using Insider Trading Data. *arXiv*.
12. Nelson, D. M. Q., Pereira, A. C. M., & de Oliveira, R. A. (2017). Stock market's price movement prediction with LSTM neural networks. *IEEE International Joint Conference on Neural Networks (IJCNN)*.
13. Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, 42(21), 7626–7634.
14. Patel, R., et al. (2023). Stock Market Prediction Using Deep Reinforcement Learning. *MDPI*.
15. Rane, A., et al. (2021). Stock Price Prediction using Sentiment Analysis and Deep Learning for Indian Markets. *ResearchGate*.
16. Sarma, N., et al. (2025). Advanced Stock Market Prediction Using Hybrid GRU-LSTM Techniques. *IJARIIIE*.
17. Sonani, M. S., et al. (2025). Stock Price Prediction Using a Hybrid LSTM-GNN Model: Integrating Time-Series and Graph-Based Analysis. *arXiv*.
18. Yahoo Finance Dataset. (n.d.). Retrieved from <https://finance.yahoo.com>