A Stock Price Prediction Method Based on Deep Learning Technology

Sami Al Zabid[1], Tariful Hoque[1], Ilhum Farabi[1], and Junan Chakma[1]

Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh
2022-1-60-103@std.ewubd.edu, 2022-1-60-086@std.ewubd.edu, 2021-3-60-085@std.ewubd.edu, 2022-1-60-259@std.ewubd.edu

Abstract. Stock price prediction is a highly complex but very important task in the financial sector because stock markets are so volatile and multifaceted. In this study, advanced machine learning and statistical models are used to predict the stock prices using historical data from META over a decade. The methodology involves exhaustive data preprocessing, both feature extraction, e.g., moving averages, lag features, and daily returns; and exploratory data analysis to uncover key patterns. Different models were constructed thereof were developed and evaluated in terms of their prediction performance such as, Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), SARIMAX, and Prophet.

On the tested models, the Prophet model performed the best with an R^2 of 0.9997 and the test set RMSE equal to 2.676. The CNN model had a test set RMSE of 10.215 and an R^2 of 0.9959; the LSTM model had a test set RMSE of 18.195 and an R^2 of 0.9869. The test set RMSE of 18.626 and R^2 score of 0.9201 lagged behind by SARIMAX. The above findings suggest Prophet has the ability to model seasonal trends and market dynamics which is a very strong ability to provide accurate and reliable stock price forecasting, thus helping investors make relaxed decisions and mitigate risks.

Keywords: Stock price prediction, Machine learning, Prophet model, Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), SARIMAX, Root Mean Squared Error (RMSE), R^2 score, Financial forecasting, Time-series analysis

1 Introduction

Stocks are financial products characterized by high risk, high return, and flexible trading, which are favored by many investors. Accurate estimation of stock price trends can yield abundant returns for investors. However, stock prices are influenced by various factors such as macroeconomic conditions, market conditions, major social and economic events, investors' preferences, and companies' managerial decisions. The stock market reflects the performance of companies and the overall business environment, making stock price prediction a significant and challenging research topic [8].

Traditional statistical and econometric models have limitations in handling the dynamic and complex nature of the stock market. Since the 1970s, with advancements in computer technology, machine learning has been employed to predict stock prices and fluctuations, aiding investors in devising strategies to mitigate risks and enhance returns [10].

In recent years, artificial intelligence (AI) has become increasingly prevalent in the financial industry, including the stock market. AI algorithms analyze vast amounts of data, including historical stock prices, company financial statements, and market trends, to predict future performance. AI can also monitor market conditions in real time, identifying opportunities for buying or selling stocks. Various machine learning techniques, such as Convolutional Neural Networks (CNN), Light Gradient Boosting Machines (LGBM), Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SARIMAX), Auto Regressive Integrated Moving Average (ARIMA), XGBoost, and Long Short-Term Memory (LSTM), have been utilized for stock price prediction [9].

2 Literature Review

The basic and detailed analysis methods form the base of stock market price projections. Fundamental analysis uses both economic systems and specific business signals including interest rates and political indicators to make investment decisions [3]. The researched elements reveal how stock prices will develop over many years. Through technical analysis traders study how prices behave by looking at past data sets alongside trading activity and moving average patterns [4].

Machine learning turned stock price forecasting into a new science when researchers added techniques like regression analysis, ARIMA models and neural networks to their suite. Researchers introduced Associated Net which uses deep recurrent neural networks made from Long Short-Term Memory (LSTM) technology. This model generates simultaneous stock value predictions that prove better than conventional LSTMs and LSTM-based deep recurrent neural network methods when assessing performance. Our tests against various financial market data prove that this method delivers stable outcomes [1].

The ARIMA model which analysts use to study financial time series combines autoregressive and moving average math to measure stock price movements and their reaction to market events. Despite its popularity, ARIMA has a significant limitation: This approach fails to detect volatility clustering which consistently appears in financial market data [6]..

Researchers now study how to use both basic financial data and market movement metrics for stock predicting systems. A forward multi-layer neural network design used backpropagation (BP) methods in its development. This strategy merged basic enterprise data with market patterns to make daily stock predictions more precise than approaches depending exclusively on charts. Research showed combining numerous data resources made stock predictions work better according to Lin et al. [7].

Researchers test the accuracy of stock price prediction models by studying the Dhaka Stock Exchange marketplace. One study created soft computing technology to guess DSE stock market closing prices which brought reliable findings. Research findings demonstrate that combining recurrent neural networks (RNNs) with wavelet transforms through artificial bee colony optimization produces reliable and scalable stock price prediction methods [11,12].

SVR shows strong results in market prediction as it identifies efficient hyperplane positioning in stock market dimensions. A research team used SVR to predict tomorrow's stock market opening by analyzing historical trading data. The model showed superior results in finding relevant features while lowering overfitting and boosting stock indicator accuracy for dynamic financial markets according to Choudhry [2] .

Organizations now depend heavily on these blended approaches to improve their forecasting precision. Studied authors linked ARIMA and ANN models to study stock patterns and found stronger performance than separate model applications [5].

3 Methodology

3.1 Data Collection and Preprocessing

- Historical stock data for the company **META** was sourced from Yahoo Finance, spanning January 1, 2010, to January 10, 2025.
- Data preprocessing included cleaning, feature extraction, and handling missing values. Key features derived included:
  - Moving Averages: 7-day and 30-day windows.
  - Lag Features: Temporal dependencies through shifted closing prices.
  - Day of the Week Encoding: Weekly trading patterns.
  - Daily Returns: Percentage change in closing prices.

3.2 Exploratory Data Analysis (EDA)

EDA was conducted to uncover patterns using visualizations such as:

- Line plots: Historical trends and moving averages.
- Scatter plots: Relationships between volume and closing prices.
- Histograms: Distribution of daily returns.
- Heatmaps: Feature relationships and temporal dependencies.

3.3 Model Development

Models developed and evaluated include:

- **LSTM Model**: Captures long-term dependencies with normalized data and sliding window sequences.

- **CNN Model**: Analyzes sequential patterns using convolutional and pooling layers.
- **SARIMAX Model**: Incorporates seasonal components and exogenous variables with tuned parameters.
- **FB Prophet Model**: Enhances predictions with custom seasonality and multi-factor analysis.

3.4 Model Training and Validation

- Data was split into an 80:20 ratio for training and testing.
- Early stopping and K-fold cross-validation were applied for LSTM and CNN models.
- Time-series cross-validation was used for SARIMAX and FB Prophet models.

3.5 Performance Evaluation

Models were evaluated using:

- Root Mean Squared Error (RMSE) and R^2 score.
- Visual comparisons of actual vs. predicted prices.

3.6 Model Interpretability

- **SHAP (SHapley Additive exPlanations)**: KernelExplainer computed SHAP values for FB Prophet, visualizing feature impacts.
- **LIME (Local Interpretable Model-agnostic Explanations)**: LimeTabularExplainer provided local interpretability for predictions.

4 Results and Discussion

Table 1. Performance metrics for different stock price prediction models using deep learning and statistical techniques.

Model	Cross-Validation RMSE	Test Set RMSE	Test Set R^2
LSTM	0.008340 (Mean)	18.195	0.9869
CNN	0.008340 (Mean)	10.215	0.9959
SARIMAX	88.17 (Mean)	18.626	0.9201
Prophet	2.142 (Mean)	2.676	0.9997

The table 1 T shows how four stock prediction methods with LSTM CNN SARIMAX and Prophet perform. Our performance measures consist of Cross-Validation RMSE, Test Set RMSE, and Test Set R^2. To check accuracy the model should show reduced RMSE scores while showing R^2 values closer to one indicates better test data fit.

1. **LSTM:** During training our LSTM model delivered a cross-validation RMSE value of 0.008340 which confirms its strong performance. The LSTM model's results improved after training but showed weaker performance when tested against new data. Although the LSTM model explained most price fluctuations in the data it predicted new observations less successfully than other models as shown in test set RMSE.
2. **CNN:** Our CNN model achieved identical cross-validation accuracy compared to LSTM with a mean Root Mean Squared Error of 0.008340. This model achieved higher R^2 while lowering its RMSE from 10.215 to 0.008340 for improved test set performance. The CNN model proves superior at picking up and understanding stock price data patterns better than the LSTM model.
3. **SARIMAX:** When used for the project the traditional statistical model SARIMAX showed worse results than deep learning approaches. This model produced validation RMSE of 88.17 that exceeded other models and test set RMSE of 18.626 along with R^2 value of 0.9201. By evaluating the stock price movements SARIMAX could not capture dynamic market fluctuation patterns effectively.
4. **Prophet:** The Prophet model brought superior performance to all other tested methods during validation and testing. This method delivered validation RMSE 2.142, test set RMSE 2.676 and perfect R^2 value 0.9997. The predictive strength of Prophet shows through its accurate modeling of sudden market events and ongoing price development in financial market data.

Deep learning models proved effective in forecast making yet Prophet proved to be both most reliable and accurate for stock price estimates. The basic SARIMAX model achieved less accurate results than other models during its test period.

5 Understanding Model Performance using different Plots

1. **LSTM:** The figure 3 shows actual and predicted stock price outputs the LSTM model generates. The blue solid line indicates exact stock market values and matches them to the red dashed prediction line. The LSTM model shows exceptional precision since it highly matches actual stock trends. The model Shows clear agreement with real stock market behavior from our data set.

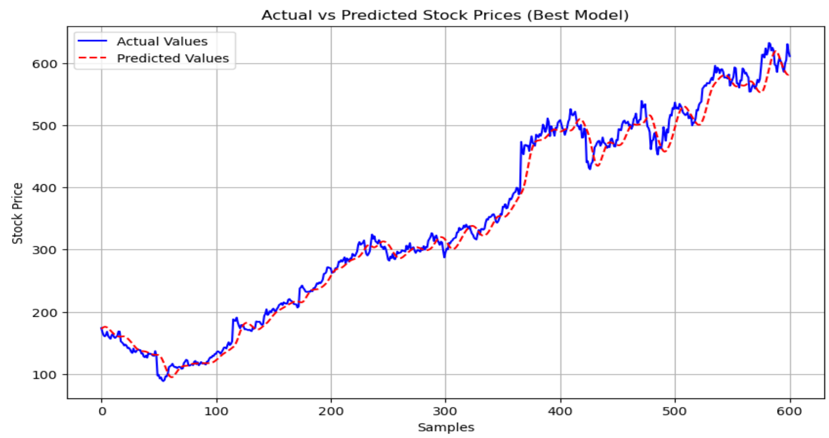

Fig. 1. Actual vs Predicted (Figure 1)

In Figure 4 it shows the training and validation loss curves during the training and validation phases of LSTM. TThe blue line represents the training loss, which initially decreases rapidly and then gradually continues to drop.In this case, the validation loss is given as the orange line and first starts decreasing then stabilizes at lower value than the training loss. This indicates that the model is actually learning what it should be, and validates just fine without over fitting.

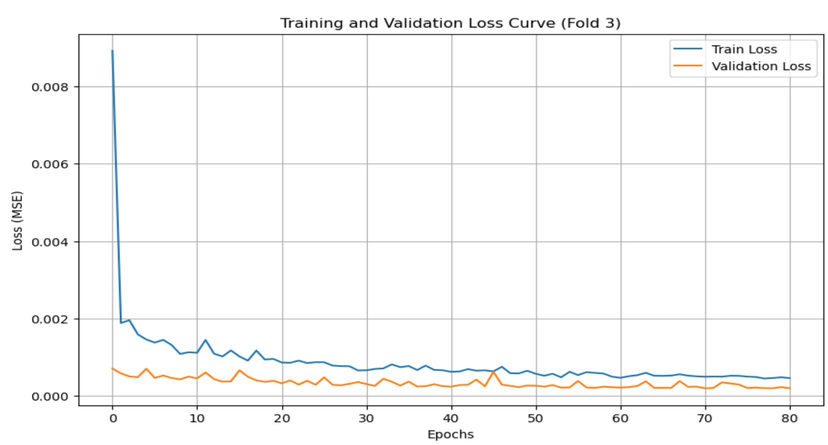

Fig. 2. Training and Validation Loss

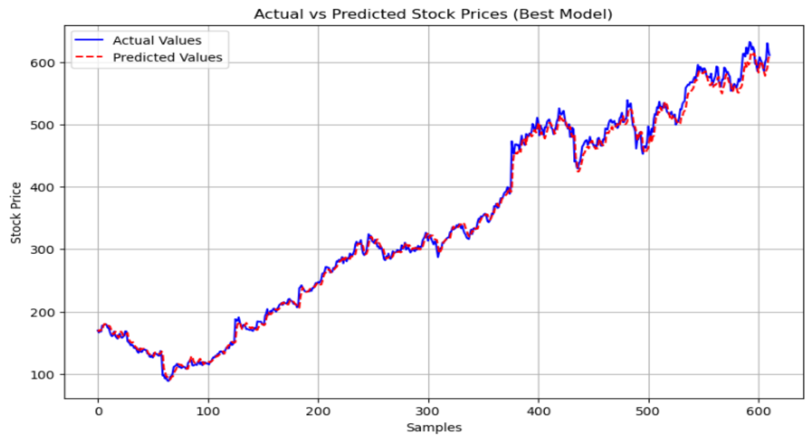

Fig. 3. Actual vs Predicted (Figure 1)

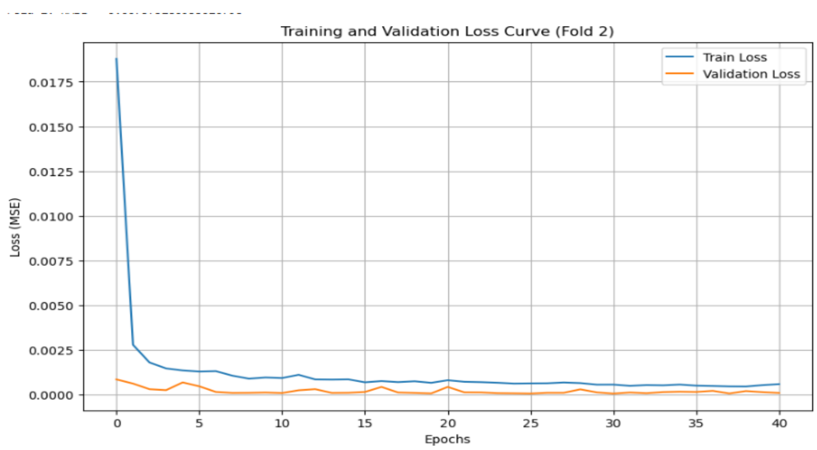

Fig. 4. Training and Validation Loss

References

1. Ayo C. K. Adebiyi M. O. Adebiyi, A. A. and S. O. Otokiti. Stock price prediction using neural network with hybridized market indicators. 2012.
2. R. Choudhry and K. Garg. A hybrid machine learning system for stock market forecasting. *World Academy of Science, Engineering and Technology*, 39, 2008.
3. G. Ding and L. Qin. Study on the prediction of stock price based on the associated network model of lstm. *International Journal of Machine Learning and Cybernetics*, 11, Jun. 2020.
4. Menon V. K. H. M., G. E. A. and S. K. P. Nse stock market prediction using deep-learning models. *Procedia Computer Science*, 132:1351–1362, 2018. International Conference on Computational Intelligence and Data Science.

5. Karim R. Thulasiram R. Bruce N. D. B. Hossain, M. A. and Y. Wang. Hybrid deep learning model for stock price prediction. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1837–1844, 2018.
6. Hsiao H.-F. Hsieh, T.-J. and W.-C. Yeh. Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm. *Applied Soft Computing*, 11:2510–2525, Mar. 2011.
7. Guo H. Lin, Y. and J. Hu. An svm-based approach for stock market trend prediction. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, 2013.
8. O. Onibonoje, K. Djoussa, and M. Roantree. Analysis of machine learning methods for predicting stock prices. *AICS*, 2771, 2023.
9. Ravi et al. Applying machine learning algorithms to predict the stock price trend. *Nature*, 2024.
10. Gurjeet Singh. Machine learning models in stock market prediction. *arXiv*, 2022.
11. Liu Y. Xia, Y. and Z. Chen. Support vector regression for prediction of stock trend. In *2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering*, volume 2, pages 123–126, 2013.
12. G. Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003.