Stock Market Price Prediction using Machine Learning Techniques

Article *in* American International Journal of Sciences and Engineering Research · February 2024

DOI: 10.46545/aijser.v7i1.308

1 author:

Mahfuz Islam Khan Jabed
The Millennium University
1 PUBLICATION **0** CITATIONS

SEE PROFILE

Sciences and Engineering Research

AIJSER VOL 7 NO 1 (2024) E-ISSN 2641-0311 P-ISSN 2641-0303

Available online at www.acseusa.org
Journal homepage: https://www.acseusa.org/journal/index.php/aijser
Published by American Center of Science and Education, USA

STOCK MARKET PRICE PREDICTION USING MACHINE LEARNING TECHNIQUES

Mahfuz Islam Khan Jabed [(a)1]

[(a)] *Student, Department of CSE, The Millennium University, Dhaka, Bangladesh; E-mail: mahfuzislam1410@gmail.com*

ARTICLE INFO

Article History:

Received: 1st October 2023
Reviewed & Revised: 2nd October 2023 to 30th January 2024
Accepted: 1st February 2024
Published: 3rd February 2024

Keywords:

LSTM, Facebook Prophet, Random Forest Regressor, Machine Learning, RNN.

JEL Classification Codes :

H54, P42, G17, C88

ABSTRACT

Predicting stock market prices is a challenging task in the financial sector, where the Efficient Market Hypothesis (EMH) posits the impossibility of accurate prediction due to the inherent uncertainty and complexity of stock price behaviour. However, introducing Machine Learning algorithms has shown the feasibility of stock market price forecasting. This study employs advanced Machine Learning models that can predict stock price movements with the right level of accuracy if the correct parameter tuning and appropriate predictor models are developed. In this research work, the LSTM model, which is a type of Recurrent Neural Network (RNN), time series forecasting Facebook Prophet algorithm and Random Forest Regressor model have been implemented on 10 Dhaka Stock Market (DSEbd) listed companies and six international giants for predicting the stock and forecasting the future price. The dataset of domestic companies is extracted from the graphical representation of the DSEbd website, and the international companies' dataset is imported from Yahoo Finance. In this experiment, Facebook Prophet demonstrates a long period of forecasting with reasonable accuracy, capturing daily, weekly, and yearly seasonality, including holiday effects for market trend analysis. Remarkably, the LSTM model exhibits significant accuracy, yielding the best results with evaluation metrics, including RMSE (0.35), MAPE (0.50%), and MAE (0.30). The experimental results underscore the efficiency of LSTM for future stock forecasting, observed over 15 days of upcoming market prices. A comparison of the results shows that the LSTM model efficiently forecasts the next day's closing price.

INTRODUCTION

The stock market constantly fluctuating in the fast-paced financial sector, finding it challenging to formulate reliable stock predictions. Investors and traders feel compelled to seek out creative approaches to reduce investment risks and maximize profits because of the possibility of generating substantial profits. Machine learning (ML) is becoming a strong tool for intelligent investment management that can help investors do better at buying and selling stocks. It can also be used to make decisions and handle portfolios in new ways. This research investigation investigates three distinct models: Facebook Prophet, which is intended for time series forecasting and focuses on weekly, yearly, and seasonal trends; Long Short-Term Memory (LSTM), a type of Recurrent Neural Network that can forecast values in the future by learning patterns from sequential data; and the Random Forest Regressor model from Ensemble Learning, that develops multiple decision trees during training to improve overall accuracy and generalization. This study distinguishes between model performances using standard assessment metrics, such as Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE), using historical data from ten domestic and six worldwide companies: Amazon, Apple, Microsoft, Google, Netflix, and Tesla. The ultimate objective is to help organizations estimate stock requests and improve future strategies. The study's findings indicate the Facebook Prophet algorithm's ability to give prolonged predicting periods with reasonable accuracy. For market trend research, it also records seasonality on a daily, weekly, and yearly scale, encompassing holiday impacts. In the meanwhile, LSTM shows accurate forecasting for the majority of company stock prices with careful parameter adjustment for each dataset with a 15-day future prediction observation. The best MAPE is

[1]*Corresponding author: ORCID ID: 0009-0001-2141-2894*

https://doi.org/10.46545/aijser.v7i1.308

To cite this article: Jabed, M. I. K. (2024). STOCK MARKET PRICE PREDICTION USING MACHINE LEARNING TECHNIQUES. *American International Journal of Sciences and Engineering Research*, *7*(1), 1–6. https://doi.org/10.46545/aijser.v7i1.308

produced by LSTM, which is better to other models according to previous study findings. Over the past few decades, there has been a fundamental improvement in information technology that has changed the path that organizations take. Financial markets, one of the most fascinating innovations, have a direct impact on the economy of the country (Hiransha et al., 2018). The global stock market value has topped 68.654 trillion US dollars, according to a (World Bank, 2018) study.

LITERATURE REVIEW

For the purpose of predicting stock prices, a number of deciding algorithms, such as fuzzy systems, decision trees, neural networks, and genetic algorithms, have been developed. There is a significant relationship between public mood and pattern discovery, which makes pattern identification an important tool for stock market trend prediction. Previous research on price forecasting has looked into combining artificial neural networks (ANN) with technical indicators. In the realm of data mining, Association Rule Mining (ARM) is a well-studied method. Despite this, decision trees are acknowledged for being useful in financial decision-making. Stock market data prediction is based on a Neural Network Approach in research by Yoon and Swales (1991), including both quantitative and qualitative elements for decision-making. With a 77.5% accuracy rate, the four-layer deep neural network beats the conventional MDA model; nevertheless, because of hidden units, it has trouble interpreting the importance of input parameters. The Autoregressive Integrated Moving Average (ARIMA) model is used for time series forecasting in another research by Pai and Lin (2005). In order to overcome these difficulties, the study uses neural networks and Support Vector Machines (SVM) in recognition of ARIMA's limits in capturing nonlinear patterns. (Akhtar et al., 2022) investigates machine learning algorithms, such as Random Forest and SVM, for stock market prediction, emphasizing dataset preprocessing, introducing a predictive model for stock lifespan, and achieving an 80.3% accuracy with the Stock Prediction algorithm.

MATERIALS AND METHODS

Data Collection

The historical data of 10 companies listed on the Dhaka Stock Market has been manually extracted from the Dhaka Stock Exchange (DSE, 2018) official website. This dataset covers a 6-month period. Additionally, data for 6 international giants' companies has been collected from yahoo Finance, spanning a 10-year period. The dataset includes information such as Open, High, Low, Close, Adjacent Close, and Volume for the international companies, as well as Price and Date columns for the 10 domestic companies. Table 1 and Table 2 display the statistic of international and domestic companies' datasets, respectively, along with their training and testing periods.

Table 1. Time Period of International Companies dataset

	Dataset	Training Dataset	Testing Dataset
Time Period	01/01/2012 – 21/09/2023	01/01/2012 – 19/05/2021	20/05/2021 – 21/09/2023

Table 2. Time Period of Domestic Companies dataset

	Dataset	Training Dataset	Testing Dataset
Time Period	02/02/2023 – 01/08/2023	02/02/2023 – 01/07/2023	02/07/2023 – 01/08/2023

Additional Data Included

Three new data columns have been generated, encompassing moving averages (MA) over varying timeframes. These data columns include 5-day, 10-day, and 15-day MA for domestic companies and 10-day, 20-day, and 60-day MA for international companies.

Long Short-Term Memory (LSTM Model)

The Long Short-Term Memory (LSTM) model is a sophisticated data mining technique designed to identify fundamental trends and analyse complex patterns in unstructured data. LSTM model is implemented as a neural network composed of LSTM units, each consisting of a cell state, input gate, forget gate and output gate. These essential components collaboratively interact with the input layer, incorporating features such as historical stock prices, date/time information, and relevant market indicators. The input gate determines the information to be stored in the long-term memory cell, the forget gate decides what to retain or discard from the past, and the output gate shapes the final output based on the current input and stored information. This dynamic interaction enables LSTMs to effectively capture and retain crucial patterns and dependencies in sequential data, establishing their suitability for predicting stock market prices. The specific number of LSTM units is subject to variation based on task complexity and dataset characteristics, often determined through iterative experimentation and turning in the model development process.

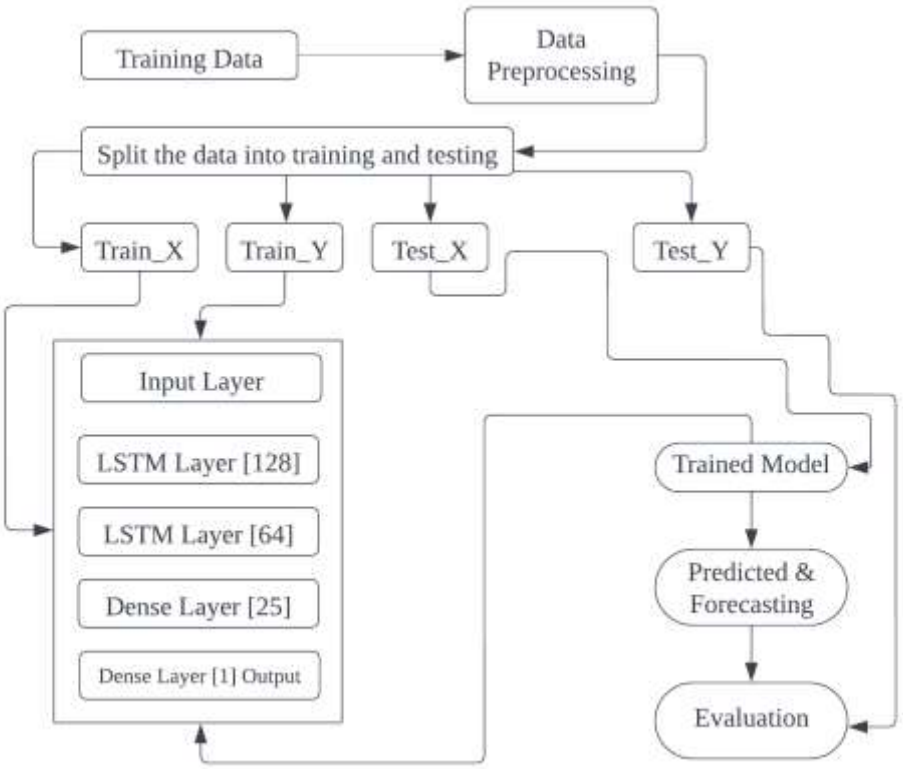


Figure 1. Proposed Recurrent Neural Network (LSTM)

Facebook Prophet

Facebook Prophet, a powerful time series predictor, identifies nonlinear trends, including seasonal and holiday effects. Ideal for datasets with strong seasonal results and limited history, Prophet stands out in handling missing data and trend shifts. It ensures quick and accurate forecasts with minimal computation time, making it equivalent to Stan models. The methodology involves data collection, pre-processing and input structuring with "ds" for data and "y" for the target variable. The model initialized with holidays, fits to the training dataset, generating and evaluating predictions. Parameter tuning balances flexibility, and the user-friendly final model is deployed for real-world predictions, accommodating various "human" seasons.

Random Forest Regressor

Random Forest Regressor methodology involves collecting historical time series data, pre-processing for cleanliness, and structuring it with relevant features and a target variable. The model is then initialized with parameters and trained to learn patterns. Feature importance analysis highlights key variables, followed by validation and hyperparameter tuning. Predictions are made on unseen data, and performance is evaluated using regression metrics. Visualization aids interpretation, and upon satisfactory results, the model is deployed for real-world predictions. The Random Forest Regressor, recognized for its robustness, proves effective in accurate time series regression.

RESULTS AND DISCUSSIONS

To evaluate the model's forecasting accuracy and overall efficiency, a comparison is conducted among three techniques (LSTM, PROPHET, RANDOM FOREST) across 10 Dhaka Stock Exchange-listed companies, namely Agrani Insurance Co. Ltd, Apex Footwear Limited, Berger Paints Bangladesh Ltd, Delta Life Insurance Company Ltd, Deshbandhu Polymer Limited, Dhaka Insurance Limited, Eastern Insurance Company Ltd, Jamuna Oil Company Limited, Premier Bank Ltd, Unilever Consumer Care Limited and six international giants, including Amazon, Apple, Google, Microsoft, Netflix and Tesla. To ascertain the accuracy of the predicted closing prices, Metrics such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) are utilized to assess errors and accuracy.
RMSE is calculated utilizing equation 1.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (1)$$

Where: n is the number of observations, y_i is the actual value of the observation and $\hat{y}_i$ is the predicted value. RMSE is a metric to measure the average accuracy of a predicted model by quantifying the differences between predicted and observed values.

MAPE is calculated utilizing equation 2.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)}{y_i} * 100 \quad (2)$$

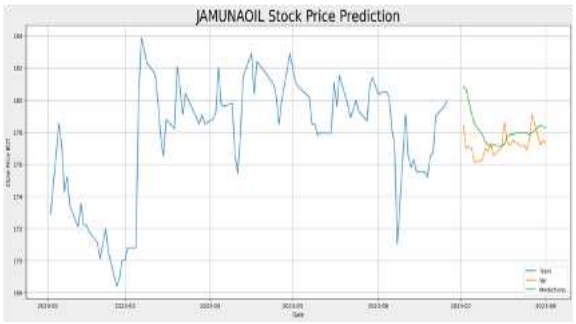
Where: n is the number of data points, y_i is the actual value and $\hat{y}_i$ is the forecasted value. MAPE (Mean Absolute Percentage Error) is used to measure the accuracy of a forecasting method by calculating the percentage difference between predicted and actual values.

MAE is calculated utilizing equation 3.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \quad (3)$$

Where: n is the number of observations, y_i is the actual value and $\hat{y}_i$ is the predicted value. MAE (Mean Absolute Error) is used to measure the average magnitude of errors between predicted and actual values, providing a straightforward assessment of prediction accuracy. The effectiveness of these models is evaluated by providing a 15-day forecast of closing prices. In the course of this research, the forecasted prices for the next 15 days are compared with the actual closing prices observed after 15 days when new data becomes available. The analysis is based on the notable accuracy achieved in predicting the stock prices, where the forecasted closing price exhibit significant similarity with the real data. This comparison serves as a validation of the models' predictive capabilities.

Eastern Insurance Company Ltd Stock

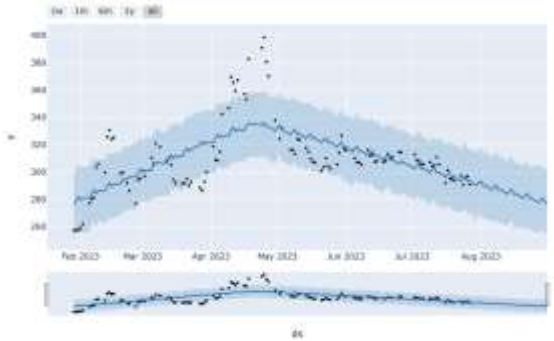


Jamuna Oil Company Ltd Stock

Netflix Stock

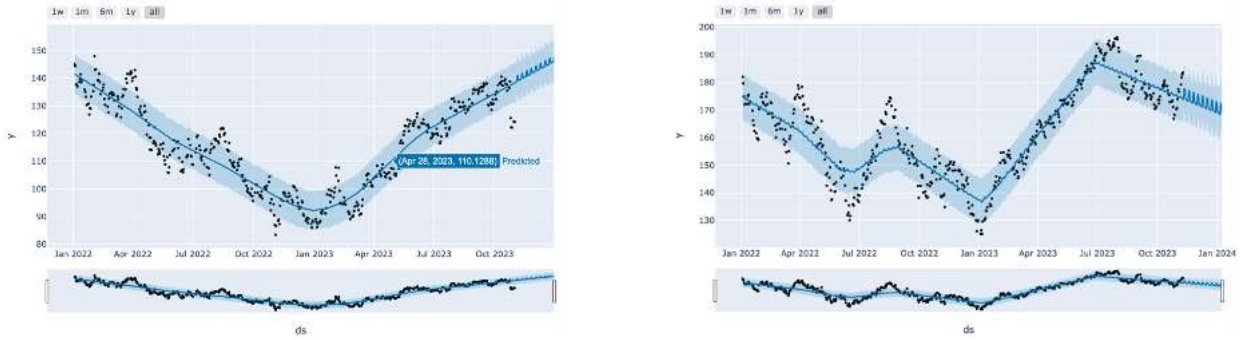
Tesla Stock

Figure 2. Original v/s Predicted Stock Closing Price using LSTM Algorithm



Apex Footwear Limited Stock

Delta Life Insurance Company Ltd Stock



Google Stock Apple Stock

Figure 3. Original v/s Predicted Stock Closing Price using Facebook Prophet Algorithm

Figure 2 displays the predicted closing prices of stocks for two domestic companies and two international companies using LSTM. In Figure 3, graphs compare the original closing prices of stocks with their predicted values using Facebook Prophet. Table 3 presents a comprehensive analysis of RMSE, MAPE and MAE values obtained from LSTM, Facebook Prophet and Random Forest Regressor model. The results emphasize the strong predictive performance of LSTM in forecasting stock prices, showcasing its effectiveness in the studied context

Table 3. Comparative Analysis of the RMSE, MAPE and MAE values derived from the utilization of LSTM, Facebook Prophet and Random Forest Regressor.

	LSTM			**Facebook Prophet**			**Random Forest Regressor**		
Company Name	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**
Agranins	1.83	1.54%	1.44	2.53	3.65%	3.85	2.21	3.33%	3.13
ApexFoot	1.36	1.28%	1.78	2.64	3.30%	4.69	3.83	3.60%	3.94
BargerPBL	2.98	1.91%	2.97	1.60	1.46%	2.05	1.59	1.05%	1.84
DeltaLife	1.82	1.68%	1.84	3.26	5.19%	3.24	1.59	1.22%	1.34
DeshBandhu	1.21	1.26%	1.55	2.73	2.63%	3.55	2.36	1.74%	2.17
Dhakains	1.71	1.63%	1.79	2.79	3.40%	2.08	1.53	1.47%	1.28
Easternins	2.11	2.74%	2.97	3.31	3.48%	5.61	2.78	1.89%	2.71
JamunaOil	1.48	0.65%	1.16	2.18	2.89%	2.58	1.20	1.09%	1.17
PremierBan	0.35	0.50%	0.30	0.63	1.18%	1.02	1.03	1.09%	1.02
UnileverCL	0.11	0.78%	1.06	1.24	2.65%	2.34	1.99	2.02%	3.63
Amazon	1.96	2.50%	1.71	1.43	1.97%	2.38	1.42	1.05%	1.66
Apple	1.96	1.46%	1.31	2.49	3.21%	5.07	1.37	1.08%	1.08
Google	1.51	2.36%	1.85	1.46	1.80%	3.32	1.20	1.74%	1.86
Microsoft	3.49	2.34%	2.88	1.51	1.94%	2.18	2.22	2.62%	1.69
Netflix	1.72	1.00%	1.63	2.35	2.85%	4.18	2.07	1.11%	2.16
Tesla	1.87	1.61%	1.70	3.51	2.55%	5.92	1.25	1.61%	2.08

The Comparative analysis on about LSTM (RNN), Facebook Prophet (Statistical Model) and Random Forest Regressor (Ensemble Learning Model) and their performance in Table 3 shows the Recurrent Neural Network (LSTM) provides better values for RMSE, MAPE and MAE.

CONCLUSIONS

This research explores the use of machine learning models, specifically LSTM, Facebook Prophet, and Random Forest Regressor, to improve stock market sentiment forecasting. The study uses historical datasets from platforms like Yahoo Finance and the company's website. LSTM is found to be a powerful tool for short-term forecasting, with impressive metrics like RMSE, MAPE, and MAE. This suggests that LSTM holds promise for traders and investors in making informed decisions in the fast-paced world of stock market investments. However, the study acknowledges its limitations due to the unpredictable nature of financial markets and the ever-evolving landscape. The study contributes to the evolving discourse on stock market prediction, providing a nuanced understanding of machine learning models' capabilities and limitations. Further exploration and refinement of predictive techniques are essential for staying ahead in the dynamic financial market landscape.

Author Contributions: Conceptualization, M.I.K.J.; Methodology, M.I.K.J.; Software, M.I.K.J.; Validation, M.I.K.J.; Formal Analysis, M.I.K.J.; Investigation, M.I.K.J.; Resources, M.I.K.J.; Data Curation, M.I.K.J.; Writing – Original Draft Preparation, M.I.K.J.; Writing – Review & Editing, M.I.K.J.; Visualization, M.I.K.J.; Supervision, M.I.K.J.; Project Administration, M.I.K.J.; Funding Acquisition, M.I.K.J. Authors have read and agreed to the published version of the manuscript.
Institutional Review Board Statement: Ethical review and approval were waived for this study, due to that the research does not deal with vulnerable groups or sensitive issues.
Funding: The authors received no direct funding for this research.
Acknowledgments: Throughout the duration of the course, I wish to extend my profoundest appreciation to my course supervisor, Dr. Md. Zahedul Hasan, for his indispensable counsel, encouragement, and mentorship. I would like to express my sincere gratitude to Lecturer Aparna Das for her invaluable

assistance in the domain of Machine Learning, which significantly advanced this research. Associate Professor K. M. Anwarul Islam a top-level scientist who secured the 2nd position on the AD Scientific Index, World Scientist Rankings 2023 for Bangladesh in the field of Business & Management / Business Administration at my university, deserves special recognition, for providing invaluable guidance, support, and assistance throughout the completion and publication of this paper. He is also a Research Fellow and Visiting Professor at INTI School of Business, INTI International University, Malaysia. He has authored or presented over 150+ scientific papers in reputed journals indexed in ABDC, ERA, ABS, Scopus-Q1/Q2, WoS-SCIE, and SSCI. His astute advice and constructive feedback were significant in ensuring the publication success of this manuscript.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

Akhtar, M. M., Zamani, A. S., Khan, S., Shatat, A. S. A., Dilshad, S., & Samdani, F. (2022). Stock market prediction based on statistical data using machine learning algorithms. *Journal of King Saud University-Science*, *34*(4), 101940.

Dhaka Stock Exchange (DSE). (2023). Dhaka Stock Market Historical Data. Retrieved from https://www.dsebd.org

Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2018). NSE stock market prediction using deep-learning models. *Procedia computer science*, 132, 1351-1362.

Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, *33*(6), 497-505.

Yoon, Y., & Swales, G. (1991, January). Predicting stock price performance: A neural network approach. *In Proceedings of the twenty-fourth annual Hawaii international conference on system sciences* (Vol. 4, pp. 156-162). IEEE.

World Bank (2018). Market capitalization of listed domestic companies (current US$). Retrieved from https://data.worldbank.org/indicator/CM.MKT.LCAP.CD

Publisher's Note: ACSE stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.