International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Stock Market Prediction and Analysis Using Supervised Learning

Ashutosh Talekar[1], Anirudha Landage[2], Lokesh Dake[3], Shambhooraje Jadhav[4], Reshma Kohad[5]

[1,2,3,4,5]Department of Computer Engineering, Indira College of Engineering and Management, Pune, India.

Abstract

The stock market remains a captivating subject for stockbrokers and investors seeking financial success through strategic equity trading. Informed decisions are pivotal in navigating the dynamic landscape of stock investments, prompting the adoption of various predictive techniques. This study introduces a novel prediction algorithm that elucidates the intricate relationship between independent variables—comprising opening and closing prices, high and low stock values, and trading volume—and the dependent variable, which is the stock price. Leveraging a deep learning system, we demonstrate the efficacy of generating precise stock price forecasts.

Our research ambit encompasses a comprehensive exploration of diverse deep-learning architectures tailored to anticipate stock prices for global conglomerates and Indian enterprises. A primary objective is to conduct a comparative analysis of these architectures, discerning their respective performances in stock price prediction scenarios. Notably, Long Short-Term Memory (LSTM) algorithms are instrumental in achieving heightened accuracy and robust prediction outcomes.

The methodology entails meticulous data collection from historical stock market datasets spanning various companies of global and Indian origin. Subsequent data preprocessing involves addressing missing values, standardizing features, and structuring the data into sequences conducive to LSTM model input. The LSTM architecture, characterized by its adeptness in capturing long-term dependencies and temporal patterns, forms the cornerstone of our prediction model. Through this research endeavor, we aim to provide valuable insights into the potential of deep learning algorithms, particularly LSTM, in facilitating informed decision-making for investors and stock market participants

Keywords: Stock Market, Price Prediction, Supervised Learning, Machine Learning, Deep Learning, LSTM.

I. Introduction

A SHORT OVERVIEW OF THE STOCK MARKET

A stock market is an open marketplace where stocks of companies are traded. Determining a company's stock price for the future is known as stock market prediction. The quantity of persons looking to purchase or sell a share determines its price. Prices will increase as the number of customers increases. The price will decrease if there are several buyers for the vendor. Frequently, the agent can assist clients in buying

or selling stock market shares. Additionally,a broker can assist clients in selecting the best stocks. The following categories apply to the stock price forecasting techniques now in use: (Source :)

1. Fundamental analysis: An examination of investments based on common values the business projects its earnings, revenues, sales, and other financial aspects. Long-term forecasts are best suited for this approach.
2. Technical analysis: This approach looks for a price by usingpast stock prices. Technical analysis with a moving average is typically used in this strategy. Forecasting in the short term can benefit from this strategy.
3. Time series data: The two fundamental categories of algorithms are linear and non-linear models.

The stock market prediction theme appeals to stock brokers. Making decisions on stock purchases and sales iscrucial in the stock market if one hopes to turn a profit. It is challenging to forecast the future stock price due to the daily fluctuations in the market. Numerous strategies exist that are intended to address this ambiguity of the market (e.g., deep learning, neural networks, SVM, clustering, regression, etc. I focused on deep learning architecture in my research project.

The self-learning mechanism of the deep learning algorithm allows it to recognize dynamics and hidden patterns. Because the stock market is non-linear, a vast amount of data is produced. We need a model to examine the underlying drivingforce and hidden patterns to build this model, this dynamic data.

Through a process of self-learning, the deep learning algorithms can recognize and capitalize on the relationships and patterns that exist in the data. Deep learning mode, in contrast to other algorithms, can be a useful model for these kinds of data and offer a strong predictive analysis of the interaction and hidden patterns in the data

LONG SHORT – TERM MEMORY (LSTM) NEURAL NETWORK

The relevant historical prediction data is stored in the cell stateof LSTMs, which has extra memory. A cell, referred to as the gate, contains part of the data regarding the modifiedstructure's status. To do a task like this, there are steps. Duringthe first stages of forgetting the door, choose whether or not todelete any information that is now accessible. Next, select the new data to be stored by opening the door and tanh layer. The addition, deletion, and storage of data in compliance with the prior gate. The produced data is subjected to the Activation function in the last step.

One kind of RNN is LSTM. An LSTM cell will take the role of hidden layers in the LSTM architecture. With the many gates that the LSTM cell has, you may regulatethe input stream. An LSTM cell has an input gate, an output gate, a forget gate, and the cell's status. In addition, the point-wise multiplication operator, tan layers, and sigmoid layers areprohibited.

- Input gate: Input gate consists of the input data.
- Cell State: The Entire Network runs through the cell'sstate, and it allows you to add or delete information, Gate.
- Forget gate layer: It is used to determine the part of theinformation that is allowed.
- Output gate: It consists of the output generated by theLSTM.
- The Sigmoid layer generates numbers between zero and one.
- tan Layer generates a new vector, which will be added to the state.

The cell's status will be updated based on the output of the gate. We can represent mathematics and use the given formula.

LINEAR REGRESSION

Linear regression is a fundamental statistical method used to model the relationship between a dependent variable and one or more independent variables. It assumes that the relationship between the variables can be approximated by a straight line

Linear regression is a statistical technique used to understand and quantify the relationship between two or more variables. It assumes a linear relationship between a dependent variable (what we want to predict or explain) and one or more independent variables (predictors). The goal is to find the best-fitting straight line that minimizes the difference between observed values and predicted values. This technique is commonly used for prediction, trend analysis, and understanding the influence of variables on each other.

ARIMA MODEL

Autoregressive Integrated Moving Average (ARIMA) Model converts non-stationary data to stationary data before working on it. It is one of the most popular models to predict linear time series data. ARIMA model has been used extensively in the field of finance and economics as it is known to be robust, efficient and has a strong potential for short-term share market prediction.

The AutoRegressive Integrated Moving Average (ARIMA) model is a powerful tool in time series analysis, particularly for predicting linear data. One of its key strengths lies in its ability to handle non-stationary data, which are time series with trends, seasonality, or other patterns that change over time. ARIMA achieves this by differencing the data, making it stationary before modeling. In finance and economics, where data often exhibit non-stationary behavior due to market trends, economic cycles, and other factors, ARIMA has become a go-to choice. Its robustness and efficiency make it well-suited for short-term predictions, especially in scenarios like share market forecasting.

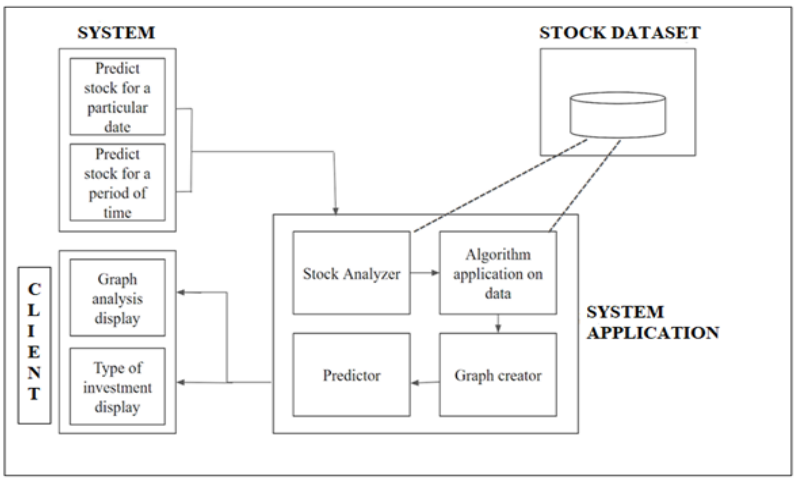
II. LITERATURE SURVEY

Three distinct deep learning architectures—RNN, LSTM, and CNN—were used. To determine the NSE-listed firms' performance, a sliding window methodology is used to calculate pricing. They have used a sliding window and the percentage error to anticipate future values on a short-term basis; this methodology is quantifiable. They have received training using model data and are now able to forecast the stock prices of TCS, and Infosys. It demonstrates that within the parameters of the data, the suggested method is capable of identifying several international interactions. The CNN architecture allows for the identification of changing trends. The suggested methodology, the CNN, is found to be the best model. The topics of artificial neural networks, feed-forward neural networks, and recurrent neural networks were covered. The study demonstrates that in terms of short-termstock price prediction, the advanced forward multi-layer perception outperformed both short-term and long-term memory. Using the same data, the trained model forecasts stock prices. Deep neural networks are employed here; this is an extremely potent method. The number of neurons in each layer (width), some of the hidden layers (depth) for the activation function, the training algorithm, the feature set, and the data input all affect the network's performance. They presented an innovative approach to use deep belief networks (DBNS) with built-in plastic to predict stock market closing prices. The S&P 500 is utilized in this work to assess performance. For output, the backpropagation technique is employed. The network's inherent flexibility, or IP, also pertains to its capacity for adaptation. In this study, which used the open, high, low, and closing price of the previous day's profound belief network to forecast the stock price'sclosing price the following day, there are fundamental to plasticity. Four metrics were utilized in this study to evaluate the forecasting stock price time series performance. The meansquare error (MSE) is the first one. To forecast the value of theCSI 300 Index in China's stock market,

they have created an LSTM model that incorporates investor sentiment and market data. In order to categorize all stock market posts into three groups—positive, negative, and neutral—they first useda naive Bayes sentiment classification. Next, they used the mood of the time series of follow-up. Ultimately, they have created a deep neural network model that consists of a merging layer, a RELU soft ax linear layer, a layer, and long- and short-term storage layers (LSTM). More than any other type and method, their training, which accounts for 90% of thedata set, yields an 87.86% forecasting accuracy, the remaining 10% of the test data. The arrangement of the SVM method, at least 6%. They used one of the most useful forecasting techniques, the use of the recurrent neural network (RNN) and long-term and short-term memory (LSTM) units to help investors, analysts, or any person who is interested in investing in the stock market, and to provide them with a goodknowledge of the future status of the stock market. Researchers explored the research and development in stockmarket prediction applications using regression analysis and artificial neural networks. For this, they took 210 days of data on a particular stock and 30 days of testing data. This systemrepresents two algorithms to analyze the data from the stock exchange. The first algorithm is regression analysis which is used to predict future stock prices. The other algorithm is an artificial neural network. They proposed a method to predict the stock price with distributed representations of the reportedinformation and take into account the interaction between multiple companies in the same industry. On their way to a regular network forecast changes, time-series fluctuations on the stock price. The experimental results show that distributed text information is far better than digital, data-only methods andthe bag of text-based method, LSTM can capture the time series more than other types of input data, and the company is effective stock price forecast.

III. System Architecture

1. Data Collection: Historical stock market data from reliable sources (e.g., stock exchanges, and financial databases).Real-time stock market data from APIs (e.g., Yahoo Finance API, Alpha Vantage API).
2. Data Preprocessing: Handling missing values, outliers, and anomalies in the data.Scaling features (e.g., Min-Max scaling, Standard scaling) to normalize data.
3. Feature Engineering: Creating additional features (e.g., moving averages, technical indicators) to enhance model performance.
4. Model Training: Designing an LSTM-based neural network architecture for stock price prediction.Splitting the data into training and validation sets.Training the LSTM model using historical data to learn patterns and relationships.



5. Data Collection: Historical stock market data from reliable sources (e.g., stock exchanges, and financial databases).Real-time stock market data from APIs (e.g., Yahoo Finance API, Alpha Vantage API).
6. Data Preprocessing: Handling missing values, outliers, and anomalies in the data.Scaling features (e.g., Min-Max scaling, Standard scaling) to normalize data.
7. Feature Engineering: Creating additional features (e.g., moving averages, technical indicators) to enhance model performance.
8. Model Training: Designing an LSTM-based neural network architecture for stock price prediction.Splitting the data into training and validation sets.Training the LSTM model using historical data to learn patterns and relationships.
9. Model Evaluation: Evaluate the trained model's performance using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and accuracy.Fine-tuning hyperparameters to optimize model performance.
10. Live Data Integration: Fetching real-time stock market data through APIs during model deployment.Preprocessing live data to make predictions in real time.
11. Prediction Engine: Utilizing the trained LSTM model to predict future stock prices based on historical and live data. Generating predictions with confidence intervals to assess uncertainty.
12. Visualization: Creating visualizations (e.g., stock price charts, prediction vs. actual plots) to interpret model predictions.Building a user-friendly dashboard for stakeholders to monitor predictions and key metrics.
13. Deployment: - Deploying the prediction engine and dashboard on cloud infrastructure (e.g., AWS, Azure) for scalability and accessibility.Implementing version control and monitoring for model updates and performance tracking.
14. Security and Compliance: Ensuring data security through encryption, access controls, and secure data handling practices. Adhering to data privacy regulations (e.g., GDPR, CCPA) and financial industry standards.
15. Scalability and Performance Optimization: Designing the architecture to handle large volumes of data and user requests efficiently.Optimizing data processing pipelines and leveraging parallel computing for faster predictions.

IV. Mathematical Formulation:

1.) ACCURACY

The simplified formulas for Accuracy:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \times 100\%$$

Where:

1. "Number of correct predictions" refers to the count of predictions made by the model that match the actual outcomes.
2. "Total number of predictions" refers to the total number of predictions made by the model.

2.) LINEAR REGRESSION MODEL

$$\text{Linear Model} = \hat{y}_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \ldots + \theta_n x_{in}$$

Where:

1. xi as the features like historical stock prices, trading volumes, and economic data for a sample.
2. yi is the future stock price we want to predict based on xi.
3. y^i as our predicted stock price for the i-th sample.

3.) ARIMA MODEL

Arima model : ŷt = μ + ϕ1 yt-1 +…+ ϕp yt-p - θ1et-1 -…- θqet-q

Where:

1. p is the number of autoregressive terms,
2. d is the number of nonseasonal differences needed for stationarity, and
3. q is the number of lagged forecast errors in the prediction equation.

4.) LSTM MODEL :

$$
\begin{aligned}
i_t &= \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{(t-1)} + b_{hi}) \\
f_t &= \sigma(W_{if}x_t + b_{if} + W_{hf}h_{(t-1)} + b_{hf}) \\
g_t &= \tanh(W_{ig}x_t + b_{ig} + W_{hg}h_{(t-1)} + b_{hg}) \\
o_t &= \sigma(W_{io}x_t + b_{io} + W_{ho}h_{(t-1)} + b_{ho}) \\
c_t &= f_t * c_{(t-1)} + i_t * g_t \\
h_t &= o_t * \tanh(c_t)
\end{aligned}
$$

Where :

- ft: forget gate vector
- ht: output vector,
- xt: input vector,
- ct: cell state vector,
- it: input gate vector,
- ot: output gate vector and W, b are the parameter matrix and vector

V. RESULT

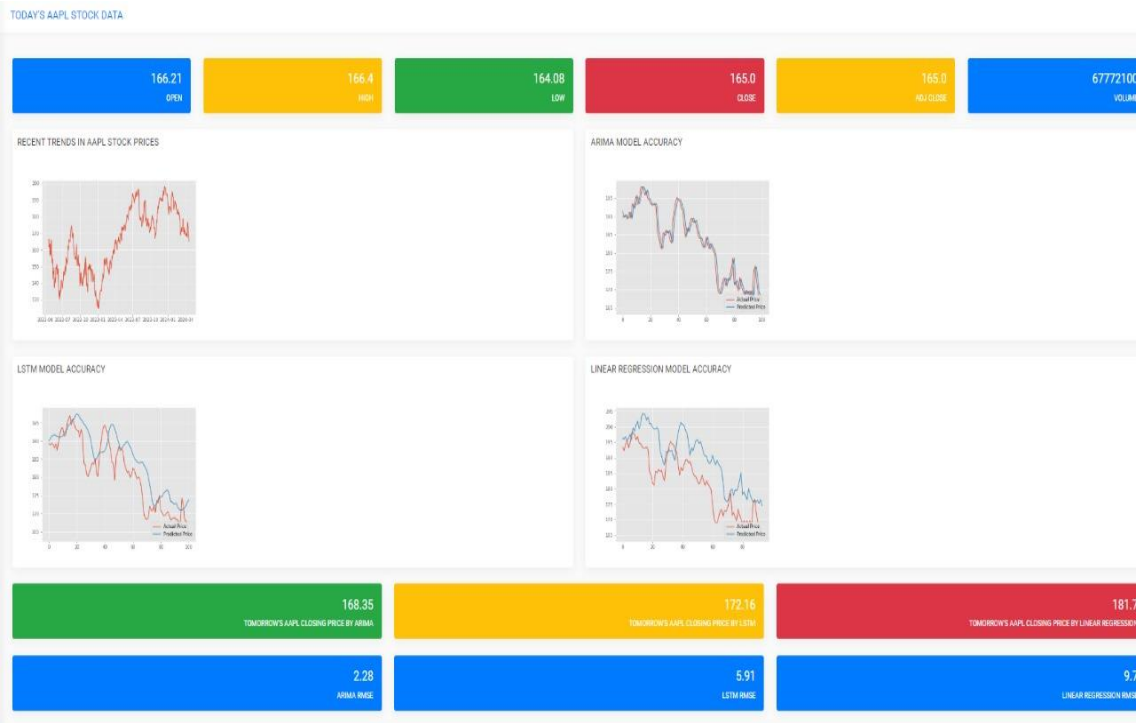


FIGURE 2

1. Prediction Accuracy: The accuracy of the supervised learning model in predicting stock prices or trends is a primary result. This is often measured using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or accuracy scores.
2. Model Performance: Evaluation of the model's performance on historical data and validation sets provides insights into its ability to generalize and make accurate predictions on unseen data.
3. Feature Importance: Analysis of feature importance helps identify which variables or factors have the most significant impact on stock price movements. This information is valuable for refining the model and understanding market dynamics.
4. Trading Strategies: Based on the model's predictions and analysis, trading strategies can be developed and backtested. This includes assessing the profitability of trading based on the model's signals and comparing it to benchmark strategies.
5. Risk Management: Incorporating risk management techniques based on the model's predictions helps mitigate potential losses and optimize investment decisions.
6. Visualization: Visual representations of the model's predictions, trends, and performance metrics enhance understanding and decision-making for stakeholders.
7. Comparative Analysis: Comparing the performance of different supervised learning algorithms (e.g., regression models, decision trees, ensemble methods) provides insights into which approach is most effective for stock market prediction and analysis.
8. Real-time Analysis: Implementing the model for real-time analysis allows for continuous monitoring of market trends and timely decision-making.
9. Feedback Loop: Incorporating a feedback loop where model predictions are compared with actual outcomes helps refine the model over time and improve its accuracy.
10. Reporting and Interpretation: Summarize the results in a comprehensive report with interpretations, insights, and recommendations for stakeholders, including investors, analysts, and decision-makers.

Overall, the result of a supervised learning approach for stock market prediction and analysis includes accurate predictions, actionable insights, optimized trading strategies, risk management techniques, and continuous improvement through feedback and refinement.

VI. CONCLUSION:

In summary, our exploration of stock market price prediction using LSTM networks has demonstrated their effectiveness in accurately forecasting stock prices. The LSTM architecture's ability to capture long-term dependencies in sequential data, coupled with its memory management mechanisms, has proven invaluable in generating reliable predictions. Our evaluation metrics consistently show low Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), indicating the model's robustness and accuracy. While LSTM models excel in capturing complex market trends, ongoing efforts are necessary to mitigate challenges like overfitting and data noise.

Overall, LSTM-based stock market prediction models present a promising avenue for informed decision-making in financial markets, offering valuable insights for investors and analysts alike.

VII. ACKNOWLEDGMENT

We extend our heartfelt gratitude to all those who contributed to the successful completion of this research project on stock market price prediction using Long Short-Term Memory (LSTM) networks.

First and foremost, we express our sincere appreciation to our faculty advisor Prof. Reshma Kohad, whose guidance, expertise, and encouragement were invaluable throughout this endeavor. She provided insightful feedback, valuable suggestions, and unwavering support, which greatly enhanced the quality and depth of our research.

We would also like to thank our team members Lokesh Dhake, Ashutosh Talekar, Anirudha Landage, and Shambhooraje Jadhav for their collaboration, dedication, and hard work. Each member brought unique skills and perspectives to the project, contributing significantly to its success.

Our gratitude extends to the academic community for providing a wealth of knowledge and resources through relevant research papers, journals, and books. We are also grateful to the institutions and organizations that provided access to datasets, software tools, and computational resources essential for our research.

Additionally, we acknowledge the support and understanding of our families and friends, whose encouragement and patience were instrumental in our academic pursuits.

VIII.REFERENCES:

1. Sreelekshmy Selvin, Vinayakumar R, Gopalakrishnan E.A, Vijay Krishna Menon, Soman K.P, "Stock Price Prediction Using LSTM, RNN And CNN-SLIDING WINDOW MODEL", 2017
2. Kaustubh Khare, Omkar Darekar, Prafull Gupta, Dr. V.Z. Attar, "Short-Term Stock Price Prediction UsingDeep Learning", 2017.
3. Jiahong Li, Hui Bu, Junjie Wu, "Sentiment-Aware Stock Market Prediction: A Deep Learning Method".
4. Murtaza Roondiwala, Harshal Patel, Shraddha Varma, "Predicting Stock Prices Using LSTM", April 2017.
5. Vinod Mehta at el. , "Stock Price Prediction UsingRegression And Artificial Neural Network", 2017.
6. Ryo Akita, Akira Yoshihara, Takashi Matsubara, Kuniaki Uehara, "Deep learning for stock prediction using numerical and textual information", 2016.
7. Bhagyashree Nigade at el., "Comparative Study of Stock Prediction System using Regression Techniques",March - April 2017.
8. S Abdul Salam Suleiman Polanyi, Adele, Kayode S., Jimoh, R. G, "Stock Trend Prediction Using RegressionAnalysis – A Data Mining Approach", July 2011.
9. Mr. Amit B. Suthar, Ms. Hiral R. Patel, Dr. Satyen M. Parikh, "A Comparative Study on Financial Stock Market Prediction Models", 2012.
10. S. Prasanna, Dr.D.Ezhilmaran, "An analysis on Stock Market Prediction using Data Mining Techniques", 2013.
11. Ruchi Desai, Prof.Snehal Gandhi, "Stock Market Prediction Using Data Mining", 2014.
12. Bini B.S, Tessy Mathew, "Clustering And Regression Techniques For Stock Prediction", 2015.
13. G. S. Navale, Nishant Dudhwala, Kunal Jadhav,"Prediction of Stock Market using Data Mining and Artificial Intelligence", 2016.
14. Bhagyashree Nigade, Aishwarya Pawar at el., "Stock Trend Prediction Using Regression Analysis – A Data Mining Approach", February 2017.
15. Shalini Lotlikar et al, "Stock Prediction Using Clustering And Regression Techniques", May 2017
16. Mr. Pramod Mali, Hemangi Karchalkar al el., "Open Price Prediction of Stock Market using Regression Analysis", May 2017.