Stock Market Price Prediction Using Machine Learning Models

Madhur Deshpande[1], Mayank Singh[2], Om Nair[3], Piyush Malhotra[4], Pranjay Tomar[5], Nirmal Mungale[6]

[1,2,3,4,5] Dept. of InformationTechnology G.H. Raisoni College ofEngineering Nagpur, India

[6]Assistant Professor Dept. of InformationTechnology G.H. Raisoni College ofEngineering Nagpur, India

Abstract- Crucial investigation of the stock showcase centres on understanding a company's genuine esteem by analysing different variables that impact its stock cost. By closely watching the stock's closing costs over time, they can distinguish designs and patterns that may flag whether a stock is underestimated or exaggerated. This approach makes a difference financial specialists pick up bits of knowledge into the company's generally wellbeing and potential for future development, empowering them to make educated choices approximately buying or offering offers based on the stock's execution in connection to its natural value.

Fundamental investigation is a key approach utilized in the stock showcase to survey a company's genuine esteem by analyzing different components that influence its stock cost. One way to do this is by following a stock's closing costs overtime to recognize designs or patterns, which can offer assistance decide if a stock is estimated as well tall or as well moo. This strategy gives financial specialists knowledge into the company's money related wellbeing and future development potential, permitting them to make more astute choices around buying or offering offers. The point is to get itwhether the stock's current cost reflects its genuine worth, directing venture choices based on this natural value.

To make strides the exactness of anticipating stock costs, progressed models such as Long Short-Term Memory (LSTM) systems and Autoregressive Coordinates Moving Normal (ARIMA) models are commonly utilized. LSTM, a sort of profound learning show, is especially solid in identifying designs over time, making it important for determining long-term patterns in the stock advertise. ARIMA, on the other hand, is a more conventional approach that employments verifiable information to make forecasts. In our investigation, we compared the execution of LSTM andARIMA by applying them to a huge dataset of stock costs and assessing their adequacy utilizing measurements like Cruel Outright Blunder (MAE) and Root Cruel Square Mistake (RMSE). The comes about appear that whereas ARIMA is dependable for short-term expectations, LSTM is way bettersuited for capturing long-term patterns and taking care of showcase instability. As monetary markets gotten to be morecomplex, having exact instruments to figure stock costs is significant for speculators and policymakers. By combining LSTM and ARIMA models with principal investigation, we can make a more vigorous framework for understanding andanticipating stock showcase developments, making a difference financial specialists make better-informed choices.

Keywords – stock, stock market, price, LSTM, ARIMA, stock price prediction.

INTRODUCTION

Foreseeing stock advertise costs has continuously been a key challenge for speculators, analysts, and budgetary examiners, as the capacity to expect cost developments can lead to critical money related rewards. The stock advertise is famously erratic, impacted by a wide extend of components, from financial arrangements to worldwide occasions and indeed advertise estimation. Asa result, finding solid ways to figure these cost changes has gotten to be significant. Customarily, models like the Autoregressive Coordinates Moving Normal (ARIMA) have been broadly utilized for short-term stock expectations. ARIMA is especially successful at recognizing straight patterns from chronicled information, making it a reliable instrument for time arrangement investigation. Be that as it may, as the stockadvertise gets to be progressively complex and eccentric, the restrictions of these conventional models—especially in capturing long-term patterns and nonlinear behaviours—are getting to be more apparent.

In later a long time, profound learning models like Long Short- Term Memory (LSTM) systems have opened up unused conceivable outcomes in time arrangement estimating. LSTM, which is outlined to work with successive information, has the one-of-a-kind capacity to hold data over longer periods and distinguish covered up designs in the information. Not at all likeARIMA, which centres exclusively on past information, LSTM learns from both short-term changes and long-term patterns,making it well-suited for capturing the complex, regularly sporadic developments of stock costs. This gives LSTM an advantage when it comes to understanding the nonlinear and unstable nature of money related markets. As these machine learning methods advance, LSTM has ended up a promising elective for long-term determining, advertising the potential for more exact expectations in an erratic showcase environment.

This inquiry about investigates the adequacy of both LSTM and ARIMA models in foreseeing stock costs. By dissecting a comprehensive dataset of chronicled stock costs, we degree how well each demonstrate performs utilizing measurements like Mean Average Error (MAE) and Root Mean Square Error (RMSE), which permit us to understand the precision of their forecasts. The objective is to get it how ARIMA, which is solid in short-term slant expectation, compares to LSTM, which exceeds expectations in recognizing long-term designs and instability.

Ultimately, this ponder points to extend our understanding of stock advertise expectation by highlighting how conventional factual strategies like ARIMA and cutting-edge profound learning methods like LSTM can be connected to real-world money related information. With markets getting to be more complex and harder to anticipate, having the right apparatuses to estimate costdevelopments precisely is more critical than ever. By comparing these two approaches, we trust to give important bits of knowledge for financial specialists and policymakers alike, empowering them to make more educated choices. Furthermore,this inquire about might serve as a establishment for future workon half breed models that combine the qualities of both ARIMA and LSTM, pushing the boundaries of exactness in stock advertise forecasts indeed advance. In the model development, we focused on four phase which includes,

1. Architecture
2. Literature Review
3. Methodology
4. Experimental Result

Architecture

In project flowchart, we initially input raw dataset of Apple Stocks and data is pre-processed through feature engineering, such as calculating simple moving averages (SMA) and scaling the data for optimal neural network input. The model is trained using historical stock price data, and its predictive performance is evaluated using metrics like Root Mean Square Error (RMSE). The research also incorporates ARIMA for comparison, highlighting LSTM's advantage in capturing long-term dependencies and non-linear relationships within stock market data. The architecture of the LSTM model includes multiple layers, starting with two LSTMlayers—one with 50 units returning sequences and another without—along with dropout layers to prevent overfitting, followed by Dense layers for final predictions. Through visualizations and performance metrics, the study demonstrates that while ARIMA is a strong baseline for short-term forecasts, LSTM offers more accuracy in predicting future trends due to its ability to learn from sequential data. The paper concludes by emphasizing the strengths of deep learning for stock prediction and suggests potential enhancements, such as incorporating additional financial indicators or further tuning model hyperparameters. The flowchart as follows:

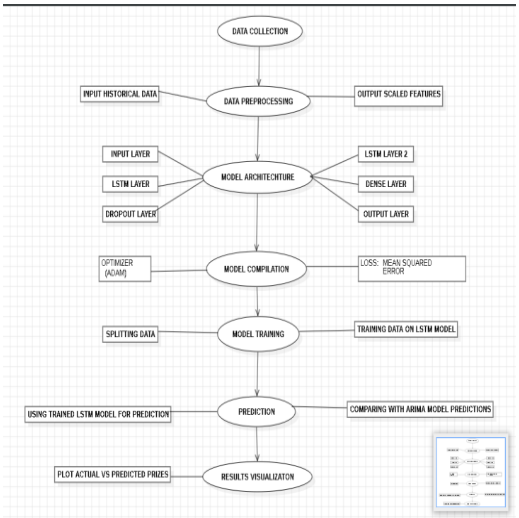

Fig1. Working of Model

LITERATURE REVIEW

Researc hPaper	Authors	Algorithm Approach	Findings
[1]	Dariusz Kobiela, Dawid Krefta, Weronika Krol, Paweł	• ARIMA • LSTM	LSTM model is more effective for long-term

	Weichbroth (2022)		
[2]	Yuxin Wang (2024)	• LSTM	LSTM gives high accuracy
[3]	Rushil Yavasani and Frank Wang (2023)	• LSTM • GRU • ARIMA	Difference in GRU and LSTM is minimal
[4]	Ishtiaq Ahammad, William Ankan Sarkar, Famme Akter Meem, Jannatul Ferdus, Md. Kawsar Ahmed, Md. R. Rahman, Rabeya Sultana, Md. Shihabul Islam (2024)	• LSTM • ARIMA	LSTM model is more effective for long term
[5]	Cristiane Orquisa Fantin, Eli Hadad (2022)	• LSTM	LSTM gives high accuracy
[6]	Berlilana1, Arif Mu’amar (2024)	• LSTM • ARIMA	LSTM is better than ARIMA for longer time period

TABLE 1: SUMMARY OF PREVIOUS WORK

The reviewed research papers explore various machine learning algorithms for stock market price prediction on applestock dataset. In the study by [1], We studied ARIMA vs LSTM on NASDAQ stock exchange data which described in detail about statistical data as well as deep learning one which are ARIMA and LSTM respectively, they used monthly and weekly average prices of companies chosen from the list of NASDAQ stock exchange, they showed the differencebetween optimum result between all three models depending on the parameters, input data, percentage error (MAPE). [4] They got the result that after analysis of the models they concluded that the ARIMA model is better in comparison to LSTM for just one feature that is historic price values and predicting more than one time period using various layers thatare adam optimizer , tanh activation function and Twice LSTMarchitecture the difference was made on the basis of comparingMAPE error .When the data was selected for a time period of 30 days ARIMA was about 3.4 times better than LSTM but when the data was selected for 90 days then then the accuracy of ARIMA decreased to 2.1 and eventually 1.8for 9 months.

[5] ARIMA's limits become apparent when dealing with the non-linear and extremely volatile character of financial markets, even if it has proved fundamental in modelling linear trends in historical price data. Because of their distinctive architecture, which makes them ideal for capturing long-term dependencies and non-linear interactions, long-term memory networks (LSTM networks) have become highly effective instruments for financial forecasting, especially in situations with high volatility and intricate data patterns. [6] Comparative research indicates that although ARIMA models are good at predicting short-term outcomes by spotting and extending linear trends, they are not very good at handling the non-lineardynamics that are typical of assets such as stocks and cryptocurrencies. On the other hand, LSTM networks have proven to be more successful in capturing these non-linearities, which makes them more useful for long-term predictions. In order to take advantage of the advantages of both methodologies, the combination of LSTM and ARIMA in hybrid models has been investigated. This has increased prediction accuracy by addressing both linear and non-linear components of financial data. [3] This study emphasizes the importance of reliable stock price forecasting in volatile markets, particularly focusing on the A-stock market. It highlights Long Short-Term Memory (LSTM) networks as a superior model for predicting stock prices compared to traditional methods like ARIMA and GRU, especially in time- series analysis. The research outlines the LSTM model architecture, data preprocessing techniques, and acknowledgeslimitations during extreme market volatility, suggesting future improvements through enhanced features and model integration. Ultimately, the study underscores LSTM's effectiveness in financial forecasting, particularly for indices like the Shanghai Composite 50. [3] This study compares the effectiveness of LSTM and ARIMA models in predicting stock market prices across various industries. LSTM, a neural network model, consistently

outperforms ARIMA, particularly in sectors with complex, non-linear data, while ARIMA remains effective for short-term forecasts in simpler markets. Using historical data, the research shows that LSTM achieves lower Mean Squared Error (MSE) and is better suited for volatile markets. The findings highlight LSTM's superior predictive power in financial forecasting, especially in challenging conditions.

METHODOLOGY

1. Long Short-Term Memory (LSTM) Model:

LSTM demonstrate improvement: A profound learning LSTM demonstrate is built with two stack ed LSTM layers. These two layers captures long term conditions. At that point we actualize Dropout regularization (20%) is connected to anticipate the overfitting. The show is prepared utilizing Adam optimizer and MSE (Mean Squared Error) as loss function. At that point we execute Hyperparameter tuning to optimize further model performance.

1. Input Layer: This is where the model takes in the stock data, such as prices (Open, Close) and other highlights like trading volume. The model uses this data to learn patterns.

2. LSTM Layer 1: The to begin with LSTM layer has 50units (or "neurons"). It learns from the sequence of stock prices over time. The setting 'return_sequences=True' implies that it will send its learned data to the following layer, helping the model keep in mind patterns from pastdays.

3. Dropout Layer 1: This layer arbitrarily turns off 20% of the neurons amid training to avoid overfitting (when the model gets to be as well specialized to the training data and doesn't work well on new data).

4. LSTM Layer 2: The second LSTM layer too has 50 units. Thislayer proceeds learning from the output of the first LSTM layer. Here, (return_sequences=False) implies it doesn't require to pass information to more LSTM layers and will presently centre on making predictions.

5. Dropout Layer 2: Another dropout layer, too turning off 20% ofneurons arbitrarily to further reduce the chances of overfitting.

6. Dense Layer: This layer with 25 neurons helps combine all thelearned data from the LSTM layers to prepare for the finalprediction.

7. Output Layer: The last part of the model predicts the stock price for the following day based on everything the model has learnedfrom the earlier layers.

2. ARIMA Model Development: The ARIMA (auto regressive moving average) model is executed as traditional statistical benchmark. ARIMA is utilized for univariant time time-series forecasting, focusing on linier patterns and model's key parameters are tuned through trial and error, grid search or minimizing AIC. The ARIMA model is prepared on same information from financial historical dataset which is utilized forLSTM to have fair comparison. As ARIMA considers linier relationship, it is not suitable for capturing non-linier dynamics often presents in stock market data. The model's forecast is madeon set of tests for comparing both of the above models.

3. Model Evaluation:

The execution of both the LSTM and ARIMA models is assessed utilizing a extend of error metrices: RMSE (Root Mean Square Error): Measures the square root of the average squared differences between anticipated values and real stocks prices.

MAE (Mean Absolute Error): Computes the average absolute difference between forecast and genuine values, giving experiences into the forecast precision.

1) Using LSTM: -

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import LSTM
```

```
# Define and compile the LSTM model
model=Sequential()
model.add(LSTM(50,return_sequences=True,input_shape=(100,1)))
model.add(LSTM(50,return_sequences=True))
model.add(LSTM(50))
model.add(Dense(1))
model.compile(loss='mean_squared_error',optimizer='adam')
```

```
model.summary()
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 100, 50)           10400

 lstm_1 (LSTM)               (None, 100, 50)           20200

 lstm_2 (LSTM)               (None, 50)                20200

 dense (Dense)               (None, 1)                 51

=================================================================
Total params: 50,851
Trainable params: 50,851
Non-trainable params: 0
_________________________________________________________________
```

Fig.5 Model summary after implementing LSTM

2) Table of comparison of LSTM and ARIMA

	LSTM	ARIMA (0,1,1)
Training RMSE	0.0238	0.0202
Test RMSE	0.0286	0.1986
Training MAE	0.0183	0.0146
Test MAE	0.0247	0.1702

Fig.6 Comparing Training values of LSTM and ARIMA

3) Final Chart for representing predicted values:

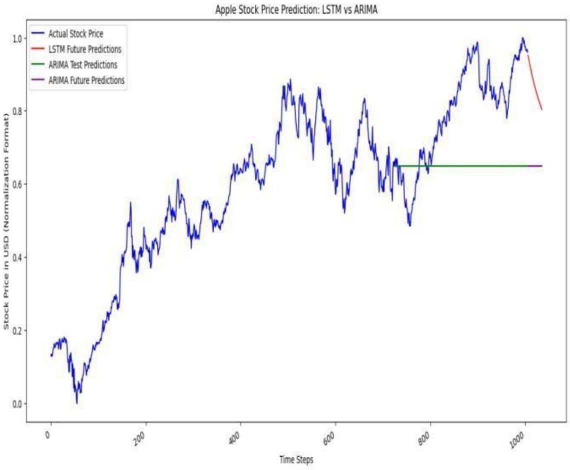

Fig.7 Chart after implementing LSTM and ARIMA model in the chart

FUTURE SCOPE

Future research can include various factors of the stock market other than only using the closing index for the price prediction model. There can usage of multiple different machine learning models to increase the accuracy of the model over a long run. As the market is highly volatile and complex and can change on different types of news, sentiment analysis can be applied on it topredict the charts and trend lines with more accuracy.

CONCLUSION

In concluding our exploration of stock market price prediction with Long Short-Term Memory (LSTM) networks and Autoregressive Integrated Moving Average (ARIMA) models, it's clear that both methods offer valuable insights, each in its ownway. LSTMs are particularly impressive when it comes to handling the complexities and unpredictability of stock prices. Their unique ability to remember important information over time makes them well-suited for adapting to the often-volatile nature of the market, resulting in forecasts that can be quite accurate. On the other hand, ARIMA remains a trusted companion in the realm of financial forecasting. Itsstraightforward approach is great for shortterm predictions, allowing analysts to easily grasp underlying trends and seasonal patterns. While it might struggle a bit in rapidly changing conditions, its simplicity and interpretability make it a go-to toolfor many professionals in the field.

This study highlights an important takeaway: instead of choosingone model over the other, we can actually benefit from combining their strengths. Imagine a hybrid approach that merges the powerful predictive capabilities of LSTM with the clarity and ease of understanding that ARIMA provides. That could take our forecasting abilities to the next level. Looking ahead, there's so much potential for growth. By incorporating additional factors, such as economic indicators or market sentiment, we can make these models even more robust. As technology continues to evolve, blending advanced machine learning techniques with traditional models will be essential for investors and analysts trying to make sense of the ever-changing stock market. Ultimately, striking the right balance between innovation and tried-and-true strategies will empower us to navigate the financial landscape with confidence and clarity.

REFERENCES

[1] Dariusz Kobiela, Dawid Krefta, Weronika Krol, Paweł Weichbroth (2022) - ARIMA vs LSTM on NASDAQ stock exchange data.

[2] Yuxin Wang (2024) - Research on Stock Price Prediction based on LSTM Modeling- A Stock Market asa Case Study.

[3] Rushil Yavasani and Frank Wang (2023) Comparative Analysis of LSTM, GRU, and ARIMA Models for Stock Market Price Prediction.

[4] Ishtiaq Ahammad, William Ankan Sarkar, Famme Akter Meem, Jannatul Ferdus, Md. Kawsar Ahmed, Md. R. Rahman, Rabeya Sultana, Md. Shihabul Islam (2024)-Advancing Stock Market Predictions with Time Series Analysis Including LSTM and ARIMA.

[5] Cristiane Orquisa Fantin, Eli Hadad (2022) - Stock Price Forecasting with Artificial Neural Networks Long ShortTerm Memory: A Bibliometric Analysis and Systematic Literature Review

[6] Berlilana1, Arif Mu'amar (2024) - Time Series Analysis of Bitcoin Prices Using ARIMA and LSTM forTrend Prediction.

[7] Kavin, B. Prabhu; Karki, Sagar; Hemalatha, S.;

Singh, Deepmala; Vijayalakshmi, R.; Thangamani, M.; Haleem, Sulaima Lebbe Abdul; Jose, Deepa; Tirth, Vineet; Kshirsagar, Pravin R.; Adigo, Amsalu Gosu - Machine Learning-Based Secure Data Acquisition for Fake Accounts Detection in Future Mobile Communication Networks

[8] Kshirsagar, Pravin R.; Manoharan, Hariprasath; Shitharth, S.; Alshareef, Abdulrhman M.; Albishry, Nabeel; Balachandran, Praveen Kumar - Deep Learning Approaches for Prognosis of Automated Skin Disease

[9] Kshirsagar, Pravin R.; Kumar, Neeraj; Almulihi, Ahmed H.; Alassery, Fawaz; Khan, Asif Irshad; Islam, Saiful; Rothe, Jyoti P.; Jagannadham, D. B. V.; Dekeba, Kenenisa - Artificial Intelligence-Based Robotic Technique for Reusable Waste Materials

[10] Manoharan, Hariprasath; Haleem, Sulaima Lebbe Abdul; Shitharth, S.; Kshirsagar, Pravin R.; Tirth, Vineet; Thangamani, M.; Chandan, Radha Raman - A machine learning algorithm for classification of mental tasks

[11] Murugamani, C.; Shitharth, S.; Hemalatha, S.; Kshirsagar, Pravin R.; Riyazuddin, K.; Naveed, Quadri Noorulhasan; Islam, Saiful; Mazher Ali, Syed Parween; Batu, Areda - Machine Learning Technique for Precision Agriculture Applications in 5G-Based Internet of Things.

[12] Mishra, Prerna; Kumar, Santosh; Chaube, Mithilesh Kumar; Shrawankar, Urmila - ChartVi: Charts summarizer for visually impaired

[13] Kshirsagar, Pravin R.; Jagannadham, D. B. V.; Alqahtani, Hamed; Noorulhasan Naveed, Quadri; Islam, Saiful; Thangamani, M.; Dejene, Minilu – Human Intelligence Analysis through Perception of AI in Teaching and Learning

[14] Dutta, Prateek; Sarkar, Abhiroop; Ambekar, Yash; Pek, HuiTing; Juwono, F. H.; Sakarkar, Gopal - A Computational Approach for Predicting the Termination of COVID-19

[15] Chandankhede, Pankaj H.; Titarmare, Abhijit S.; Chauhvan, Sarang - Voice Recognition Based Security System Using Convolutional Neural Network

[16] Ravekar, Kratika; Jaju, Santosh B.; Narsingpurwala, Abizar Zoeb - Comparative analysis of material for material discharge plate of micro-surfacing machine.

[17] Potnuru, Srikar; Sahoo, Susanta K.; Sahoo, Santosh K.; Kotta, Anand Babu; Gara, Dheeraj K.- Analytical, numerical and experimental analysis of double collar collet chuck holder by combined extrusion and forging processes.

[18] Singh, Sushil Kumar; Singh, Deepak; Kumar, Abhishek; Jain, Anuj - An Analysis of Mechanical and Viscoelastic Behaviour of Nano-SiO2 Dispersed Epoxy Composites.

[19] Choudhary S., Dorle S. - A quality of service-aware high- security architecture design for software-defined network powered vehicular ad-hoc networks using machine learning-based blockchain routing

[20] Kalbande V.P., Fating G., Mohan M., Rambhad K., Sinha A.K. - Experimental and theoretical study for suitability of hybrid nano enhanced phase change material for thermal energy storageapplications

[21] Bramhane L., Salankar S., Gaikwad M., Panchore M. - Impact of Work Function Engineering in Charge Plasma Based Bipolar Devices

[22] Manoharan H., Abdul Haleem S.L., Shitharth S., KshirsagarP.R., Tirth V., Thangamani M., Chandan R.R. - A machine learning algorithm for classification of mental tasks

[23] Shrawankar U., Dhule C.A. - Resource-Efficient Pareto-Optimal Green Scheduler Architecture

[24] Mate N., Akre D., Patil G., Sakarkar G., Basuki T.A. -Emotion Classification of Songs Using Deep Learning

[25] Patankar P., Dorle S., Wyawahare N., Thakre L.P. -Comparative Study on Design Of AI-Based Communication Protocol For VANET