Time series analysis and forecasting in Stock market – A survey

Nakshatra Pawar
Dept of Computer science and engineering,
MITWPU, Pune, India
pawarnakshatra57@gmail.com

SECTION 1

***Abstract*— Predicting stock market movements can be problematic because time series data in the financial market are inherently non-parametric, complicated, noisy, chaotic, and unpredictable. Smart models, however, have been made possible by advancements in computers [5] and can help expert analysts and investors reduce the risk of their investments. Therefore, this survey paper, focuses on using technical analysis and deep learning algorithms to predict stock market movements. In this paper, a liner autoregressive model (AR) [2] is used.**

Keywords—time series; forecasting; analysis; autoregressive model; ARIMA, LSTM

I. INTRODUCTION

Forecasting and time series analysis are essential for foreseeing stock market moves. [1] The stock market's well-known intricacy, dynamic nature, and volatility make it challenging to make exact estimates. But advances in computing power and data analytics have made it possible to approach this issue in a sophisticated manner. By modelling and examining consecutive data sets that have been gathered through time, time series analysis seeks to discover patterns, trends, and correlations. In order to create educated hypotheses for future price movements on the stock market, previous stock price and volume data must be analyzed. [3] The assumption that stock prices display different cyclical patterns and behaviors throughout time forms the basis of this method. When all is said and done, time series research and developing procedures are invaluable resources for forecasting changes in the financial exchange. In order to navigate the complex and dynamic world of the stock market [8], analysts and investors can get insights into potential trends and make informed decisions by applying sophisticated models and prior data.
This study's goal is to examine and compile pertinent methods that are already in use, the goal is to reduce stock price forecast error rates.

II. MOTIVATION

The Utilizing historical data, spotting patterns, and putting technical indicators to use allow stock forecasters to gain insightful knowledge into potential market movements. With the use of this information, analysts and investors may make well-informed judgements, effectively manage risks, and perhaps increase their investment returns.

III. NEED

The need for time series analysis-based stock forecasting arises from the dynamic and complex character of stock market data. By utilizing real-world instances and patterns, time series analysis enables academics and financial backers to control risks, create informed forecasts, and enhance speculating methods.

IV. CONCERN

However, A few things should be taken into consideration when using time series techniques to forecast the stock market. Due to noise, consistency issues, and outliers in stock market data, predictions may be less precise. Second, non-stationarity, including seasonality and shifting patterns, makes modelling difficult. Thirdly, depending only on verifiable data recognizes that present examples will reflect the past, but shifting economic conditions may make historical links less stable. Strong model validation is required since the models are sensitive to parameter choices. Finally, forecasting is difficult because of the financial markets' innate unpredictability and uncertainty. Instead, then being precise predictors of what will happen in the future, models should be used as instruments for making decisions.

V. OBJECTIVE

- Price Change Prediction: Time series methods are used in stock market forecasting to project future price changes for equities. This entails examining verifiable information instances, patterns, and correlations in order to create precise forecasts about whether stock values will grow, decline, or remain stable.

- Finding Patterns and Trends: Another goal is to locate and recognize basic patterns and occurrences in stock market data.

- Risk The board: Time series methods for stock market forecasting are also intended to aid in risk control.

- Financial market time series forecasting techniques also try to design venture strategies. Investors may find opportunities to buy or sell stocks at a profit by accurately predicting changes in stock price. This might lead to increased dynamics regarding timing stock exchanges, progression of venture classifications, a more successful portfolio across the board, as well as more speculative earnings.

Overall, the main goal of stock market forecasting using time series with previous data, strategies are used to precisely anticipate future stock price fluctuations.

VI. ORGANISATION OF PAPER

This survey paper is organized into six sections.

Section I, the Introduction, provides an overview of the motivation, need, and concerns regarding stock market forecasting using time series, as well as the objectives of this survey paper.

Section II, Related Work, provides an overview of existing literature related to stock market prediction.

Section III, Methodology, gives the insights of methods used.

Section IV, Results, presents the outcomes of the paper, and gives insights of the most powerful and less error method for stock forecasting.

Section V, Discussion, discusses the findings of the survey, including the advantages and limitations of ARIMA method used in the paper for stock analysis and forecasting using time series

Section VI, Conclusion and Future Work, summarizes the highlights the survey's important findings and suggests future study topics in this area.

SECTION 2

I. LITERATURE SURVEY

The Time series analysis and forecasting have received a lot of interest in both academia and business because of its many applications and the need for accurate projections across a variety of areas [6] . There has been a ton of research in this field, spanning from conventional statistical techniques to cutting-edge machine learning methodologies, according to a thorough study of the literature.

The development of machine learning has completely changed time series analysis and forecasting, allowing the use of cutting-edge algorithms for modelling and prediction tasks. RNN [5], in particular LSTM networks, have become quite popular due to their capacity to recognize long-term dependencies and manage complex sequential data. The analysis of time series has also demonstrated the potential of deep learning architectures like convolutional neural networks (CNNs) [1].

- "A Comparative Study of Time Series Forecasting Techniques for Stock Market Prediction"[7]-Investigated and compared the performance of various time series forecasting techniques, such as ARIMA, LSTM, and SVM, in predicting stock market movements.

- In the research [1] "Forecasting Stock Market Returns using a Combination of LSTM and Random Forest", a crossover technique combining LSTM and Arbitrary Woods models is introduced as a method for forecasting returns on stock exchanges. The review combines LSTM's abilities to identify far-off events with irregular backwoods' advantages in managing nonlinear connections. In terms of prediction accuracy, the hybrid technique outperforms independent models when it trains the models on historical stock market data and aggregates their estimates. Integrating LSTM with Random Forest allows traders and investors to gain crucial insights about stock market returns that are more detailed and accurate.

- The ensemble learning algorithms for stock market prediction are thoroughly evaluated in the paper "Ensemble Learning for Stock Market Prediction: A Review" [5]. In comparison to individual models, the review shows that ensemble techniques like bagging, boosting, and stacking are more effective in increasing prediction accuracy and robustness. The researchers go over the theoretical underpinnings of ensemble learning, talk about relevant articles and experiments, and emphasise the advantages of ensemble approaches for gathering a variety of data and minimising prediction errors. Additionally, they address difficulties and worries associated with making clothing choices, supplying vital information to experts and pros looking to improve their securities exchange gauging models.

- In addition, [9] "Hybridizing Neural Networks and Technical Indicators for Stock Price Prediction" - Explored the fusion of neural networks with technical indicators to enhance stock price prediction accuracy, considering both short-term and long-term trends.

- The paper titled "Deep Learning Techniques for Stock Market Prediction: A Survey" [11] offers a comprehensive examination of deep learning-based stock market prediction. Convolutional neural networks (CNN), deep belief networks (DBN), and LSTM are a few of the models addressed in this study. Deep learning models have the capacity to detect complex scenarios and occurrences in securities exchange information, significantly improving prediction accuracy. Current developments, model layouts, information preparation strategies, and execution assessment techniques are covered in the article. The paper is a helpful resource for professionals and academics interested in using deep learning methods for stock market forecasting.

- The paper "Forecasting Stock Market Volatility Using GARCH Models: A Literature Review"[13] research presents a comprehensive analysis of the Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models for the prediction of stock market volatility. In this review, the various estimating approaches, assessment strategies, and GARCH variations applied in different research are addressed. It illustrates the benefits and drawbacks of using GARCH models to simulate stock market volatility dynamics. In summary, the article helps us understand how to predict volatility and manage risk in financial markets by giving us helpful information on how to utilise GARCH models to predict volatility in the stock market.

- In order to anticipate the stock market, the research "Deep Learning for Stock Market Prediction: A Comparative Study" [4] compares recurrent neural networks (RNNs) with convolutional neural networks (CNNs) deep learning models. The study investigates how effectively these models can anticipate stock price fluctuations and identify temporal linkages. The study sheds light on how well deep learning performs in stock market prediction tasks by comparing the performance of several deep learning architectures. The results contribute to our understanding of the advantages and disadvantages of deep learning models for forecasting stock exchange behaviour and help specialists and professionals choose the best strategies for stock price prediction.

II. METHODOLOGY

- Machine Learning: Time series analysis and forecasting heavily utilize machine learning techniques. These problems may be dealt with in addition to dealing with various time series, missing data, non-stationarity, and data patterns and relationships. To make accurate forecasts, ensemble approaches and the right kind of validation are necessary. In addition to providing insightful data in a variety of sectors, machine learning offers efficient tools for precise forecasting.

- Deep Learning: For time series analysis and forecasting, deep learning approaches have proved helpful. Both long-term trends and temporal correlations may be accurately captured by these models in sequential data. Even better results can be achieved by incorporating outside data such as sentiment analysis or technical indicators. However, deep learning's training necessitates a lot of data and computing power, as well as rigors model construction and regularization. In general, deep learning outperforms time series analysis in terms of its ability to recognize complicated patterns in stock market data.

- Autoregressive Model: Forecasting and time series analysis frequently employ AR models. They assume that values of the present and those of the future are logically related. How many lagged data are considered depends on the order of the models. AR models can represent short-term dependency because of their simplicity. They require stationarity in order to calculate model parameters from previous data and predict future values. They are crucial in economics, finance, and weather forecasting.

- ARIMA: The ARIMA (Autoregressive Coordinated Moving Normal) approach is a well-known one for time series analysis and estimation. Its three components are moving average, autoregression, and distinguishing. Autoregression finds the connection between the variable and its previous properties, whereas differencing corrects the data by eliminating patterns and irregularity. The moving average component accounts for the relationship between subsequent residuals. ARIMA may be used to predict future values by estimating the model's parameters from historical data. It is a useful tool for predicting and studying time series patterns in many different sectors.

- LSTM: Long Short-Term Memory (LSTM) recurrent brain organisation engineering successfully detects long-term circumstances and real-world occurrences in subsequent information. It is intended to use LSTM, which is entirely distinct from traditional RNNs, to resolve the evaporating slope problem, which prevents the learning of long-range situations. With the use of memory cells and gating mechanisms, it can selectively store and remove information at each time step. The model can record important information over extended periods of time because the memory cells maintain an internal state. LSTM is a powerful tool for modelling and predicting challenging sequential data since it has demonstrated value in a variety of applications, such as natural language processing, speech recognition, and time series analysis.

III. PROPOSED METHODOLOGY

ARIMA Model: Due to its ability to identify temporal linkages and patterns in stock market data, the ARIMA (Autoregressive Integrated Moving Average) model is particularly useful in stock prediction. It is especially helpful for evaluating patterns since it may combine autoregressive components that reflect the linkages between historical and current prices to find and evaluate underlying stock cost trends. ARIMA models may detect recurring patterns over distinct time periods by merging moving normal sections, which can also depict irregularity and repetitive patterns in financial exchange information. Additionally, ARIMA models are effective for displaying and forecasting financial exchange volatility by looking at the residuals or mistakes, which address unexplained variance in the data. This leads to the development of insights on volatility patterns and prudent risk management decisions.

Additionally, ARIMA models are suitable for short-term stock cost estimation, using verified data and slack values to generate estimates for the near future. Although ARIMA models have limitations, such as accepting direct connections and fixed information, they are frequently used in conjunction with other methods or as a component of more advanced guaging models in stock expectation experiments.

- Architecture:

Autoregression (AR), differencing (I), and moving normal (Mama) are the three main components of the engineering of an ARIMA (Autoregressive Incorporated Moving Normal) model for stock cost expectation.

1. Autoregression (AR): The autoregressive component captures the relationship between the current impression and its prior attributes. It assumes that there is a linear relationship between the past and present values of the variable. How many lagged data are used in the model depends on the order of autoregression, denoted by "p." The current error term is increased by multiplying the results of each lagged observation by the relevant coefficient.
2. Differencing (I): To make the time series information fixed, the differencing part is utilized to eliminate patterns and irregularity. Stationarity is fundamental for ARIMA since it relies upon the series' factual properties staying consistent across time. The letter "d," which denotes the order of differencing, indicates the number of differencings required to achieve stationarity.
3. Moving Average (MA): The moving typical part catches the reliance between the blunder terms or residuals over a progression of time spans. It demonstrates the data noise or short-term oscillations that the autoregressive and differentiating components are unable to explain. The moving average's order, "q," indicates the number of lag residuals included in the model.

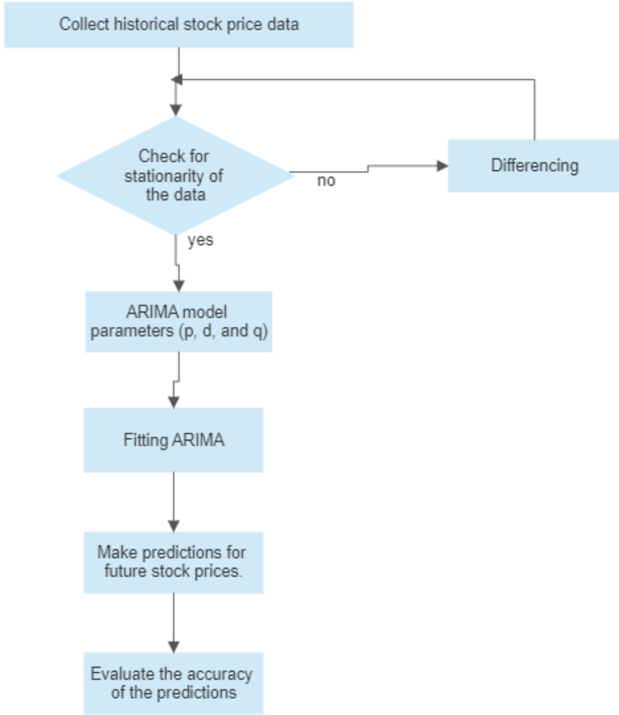

Fig 1. Architecture of ARIMA Model

IV. IMPLEMENTATION and RESULT

For implementation of ARIMA model downloaded the csv dataset and worked out the operations like preprocessing on the dataset, implementing ARIMA model to predict stock market prices.

The Fig.2 mentioned below shows the closing price of the stocks, from the dataset

Fig 2. Closing Price of stocks

From fig3, graph we can say that our data is stationary as the rolling mean and standard deviation are almost straight.

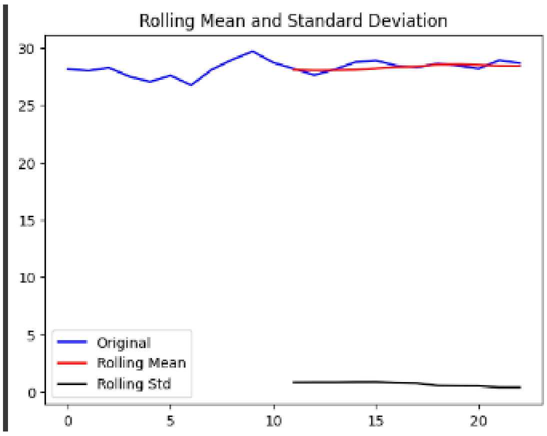

Fig 3. Stationarity checking

Results of the Dickey-Fuller Test are:

Test Statistics	-5.343665
p-value	0.000004
No. of lags used	8.000000
Number of observations used	14.000000
critical value (1%)	-4.012034
critical value (5%)	-3.104184
critical value (10%)	-2.690987

Now working ARIMA model on dataset, by dividing it into training and test data, below graph shows both the data.

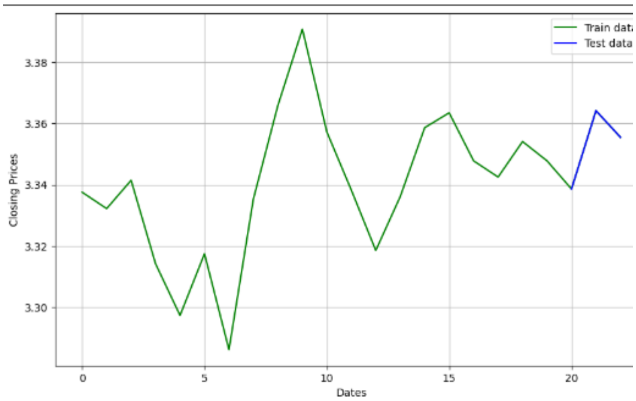

Fig4. Training and Test Data

Now to choose ARIMA model's parameters, we have used auto ARIMA. Given below in fig5 are results

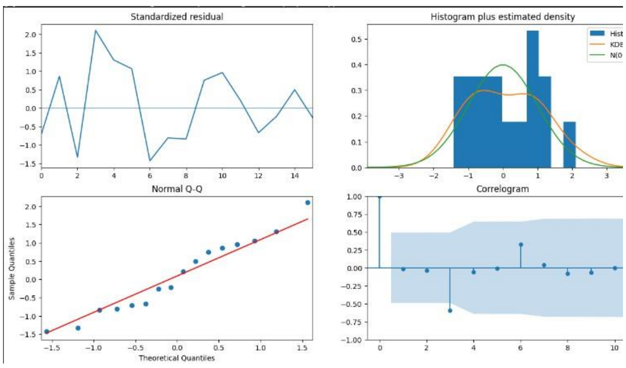

Fig5. Finding Parameters

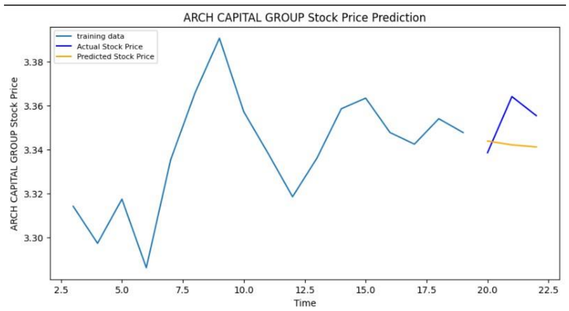

Fig6. Result predicting future prices

Hence, we have successfully deployed the model, and the future stock predictions are shown in the graph by yellow line, which is decreasing.

Now to check the accuracy of the model, We will compare the three errors, namely, MAE, RMSE, MAPE, as shown in fig7.

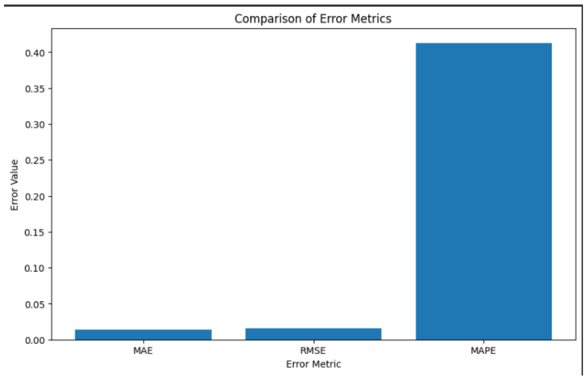

Fig7 Comparison between MAE, RMSE, MAPE

Model	Accuracy
ARIMA	97.50%
single-layers	95%
LSTM	66.80%
RNN	89%
CNN	67%

Table 1 Comparison of different models

V. LIMITATIONS

There are a number of issues with using ARIMA (Autoregressive Integrated Moving Average) models to forecast the stock market. The linearity and stationarity on which ARIMA models are built may not completely represent the dynamic and complex nature of stock market data. ARIMA models could have a hard time accurately representing stock values because of their tendency to show clustering, rapid oscillations, and non-linear patterns in their volatility. Second, it's possible that ARIMA models won't be capable of handling intricate connections and protracted correlations between various market components. Predictions over longer time horizons may be less precise due to the model's limited capacity to include seasonality and long-termtendencies.

VI. CONCLUSION

In Conclusion, we saw the working of the ARIMA model for stock market prediction and compared its accuracy with that of LSTM model, we see that ARIMA yields the best performance, i.e., it achieves the smallest mean square error and mean absolute error on the test set. In contrast, the LSTM neural network performs the worst of the three models. Additionally, ARIMA models provide insights into volatility modelling, allowing one to assess and anticipate volatility in the stock market. ARIMA models may provide insight into unpredictable designs and work with educated navigation when it comes to taking a chance with the board by analyzing the residuals or errors, which address unexplained variation in the information.

VII. FUTURE WORK

The next task is to create a model that can produce time expectations from time series data. The model must be modified in order to do this by refining its parameters and taking in fresh data. The precision and power of the model has to be increased in order to properly capture the information's transient instances and components. To do this, it may be necessary to examine a variety of models, combine cutting-edge methods like consideration components or sporadic brain organizations, and optimize hyperparameters. The objective is to create a model that can effectively and reliably predict time on various time series datasets.

VIII. REFRENCES

1. Vijh, S., Pandey, A. K., Vijh, G., & Kumar, S. (2021). *Stock Forecasting for Time Series Data using Convolutional Neural Network. 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence).* doi:10.1109/confluence51648.

2. Agarwal, U., & Sabitha, A. S. (2016). Time series forecasting of stock market index. 2016 1st India International Conference on Information Processing (IICIP). doi:10.1109/iicip.2016.7975381.

3. AUDELIANO WOLIAN LI AND GUILHERME SOUSA BASTOS, “Stock Market Forecasting Using Deep Learning and Technical Analysis: A Systematic Review”, Received September 30, 2020.

4. ZHENYU LIU, ZHENGTONG ZHU, JING GAO, AND CHENG XU, “Forecast Methods for Time Series Data: A Survey”, received June 1, 2021, accepted June 15, 2021, date of publication June 21, 2021, date of current version July 2, 2021.

5. John-Syin Ang, Kok-Why Ng, Fang-Fang Chua, “Modeling Time Series Data with Deep Learning: A Review, Analysis, Evaluation and Future Trend”, 2020 8th International Conference on Information Technology and Multimedia (ICIMU).

6. Rajkomar et al., “Scalable and accurate deep learning with electronic health records,” npj Digit. Med., 2018.

7. X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,” in Advances in Neural Information Processing Systems, 2015.

8. J. C. B. Gamboa, “Deep Learning for Time-Series Analysis,” arXiv Prepr. arXiv1701.01887, 2017.

9. Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, “Recurrent Neural Networks for Multivariate Time Series with Missing Values,” Sci. Rep., vol. 8, no. 1, pp. 1–12, 2018.

10. F. Karim, S. Majumdar, H. Darabi, and S. Chen, “LSTM Fully Convolutional Networks for Time Series Classification,” IEEE Access, vol. 6, pp. 1662–1669, 2017.

11. F. Karim, S. Majumdar, H. Darabi, and S. Harford, “Multivariate LSTM-FCNs for time series classification,” Neural Networks, vol. 116, pp. 237–245, 2019.

12. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, “Multi-Scale Convolutional Neural Networks for Time Series Classification,” J. Syst. Eng. Electron., vol. 28, no. 1, pp. 162–169, 2017.

13. S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, “DeepSense: A unified deep learning framework for time-series mobile sensing data processing,” 26th Int. World Wide Web Conf. WWW 2017, pp. 351– 360, 2017.

14. G. Lai, W.-C. Chang, Y. Yang, and H. Liu, “Modeling Long-and ShortTerm Temporal Patterns with Deep Neural Networks Multivariate Time Series, Neural Network, Autoregressive models ACM Reference Format,” arXiv, no. July, 2018.

15. S. Y. Shih, F. K. Sun, and H. yi Lee, “Temporal pattern attention for multivariate time series forecasting,” Mach. Learn., vol. 108, no. 8–9, pp. 1421–1441, 2019.

16. S. Li et al., “Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting,” no. NeurIPS, 2019.

17. S. Huang, D. Wang, X. Wu, and A. Tang, “DSANet: Dual SelfAttention Network for Multivariate Time Series Forecasting,” Proc.

28th ACM Int. Conf. Inf. Knowl. Manag., pp. 2129– 2132, 2019. pg. 12

18. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by backpropagating errors,” Nature, vol. 323, no. 6088, pp. 533–536, 1986.