IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Stock Market Trend Predictions using Machine Learning Algorithms Knn and Xgboost

Akash Chourasia, Nilesh Kumar Gupta

Department of Computer Science and Engineering,

Chouksey Engineering College, Bilaspur Chhattisgarh 495004, INDIA.

Abstract— By utilising machine learning methods, notably k-Nearest Neighbours (KNN) and XGBoost, this paper gives a complete analysis on the utilisation of these algorithms for the purpose of anticipating trends in the stock market. The purpose of this study is to determine whether or not these algorithms are effective and accurate in predicting stock values based on historical data. In this particular investigation, the dataset that was utilised is comprised of stock market data that was gathered from Yahoo Finance. This dataset includes characteristics such as closing prices, trading volume, and high-low percentage changes. In terms of predicting trends in the stock market, the experimental findings indicate the performance of the KNN and XGBoost models, with the KNN model achieving an accuracy of 96% and the XGBoost model achieving an accuracy of 98%. Both algorithms appear to have promising skills in capturing patterns and trends in stock market data, which highlights their potential for practical applications in investment decision-making and financial analysis. The findings imply that both algorithms exhibit intriguing possibilities in this regard.

Keywords— Stock market prediction, Machine learning, K-Nearest Neighbors (KNN), XGBoost, Financial forecasting.

I. INTRODUCTION

There are a wide range of factors that can have an impact on the stock market, such as economic indicators, geopolitical events, and investor attitude [8]. Each and every one of these elements has an impact on the investment market. The stock market is a system that is both dynamic and complex. They are both characteristics of the system. The ability to precisely anticipate the tendencies of the stock market is an absolute necessity when it comes to making well-informed decisions about investments and effectively managing the risks associated with financial matters [12]. Throughout the history of their existence, numerous machine learning algorithms have been investigated with the purpose of making predictions regarding the stock market. There are a variety of algorithms, ranging from more conventional approaches such as linear regression to more complex models such as deep learning [9]. Deep learning is one of the most advanced models among these algorithms. Among them, k-Nearest Neighbours (KNN) and XGBoost have seen a significant surge in popularity over the past few years. This is mostly because to the fact that they are easy to implement, extremely efficient, and possess the capability to identify nonlinear correlations in data [1, 6].

The kernel Neural Network (KNN) technique is a straightforward and user-friendly method that categorises data points in accordance with the majority class of the neighbours that are closest to them [3]. Stanford University is responsible for the development of this algorithm. Identifying patterns in previous price data and estimating future price movements are two of the things that may be accomplished with the help of KNN when it is applied to the context of stock market investing [17]. Because KNN is predictive in nature, this is something that can be done. On the other hand, XGBoost is a highly effective ensemble learning algorithm that delivers accurate forecasts by integrating the predictions of a number of weak learners, which are also usually referred to as decision trees [4]. XGBoost is a member of the ensemble learning family. When it comes to the management of extremely big datasets that contain high-dimensional features, it is an excellent option [20]. This is due to the fact that XGBoost is well-known for its scalability, efficiency, and robustness, all of which are traits that enable it to be an exceptional option.

In spite of the fact that they are not exactly the same, KNN and XGBoost have a great deal of advantages in common that are advantageous to both of them. The ease with which they can be implemented and evaluated is another one of these advantages [22]. Another advantage is the possibility of managing non-linear correlations between the variables that are being input and the variables that are being sought after. Additional benefits include the simplicity of implementation and interpretation. On the other hand, every method has its own particular set of constraints and difficulties, such as the fact that it is sensitive to noisy data, that it requires the parameter settings to be altered, and that the computations that are involved are extremely complicated [11], [16].

In order to determine how well the KNN and XGBoost algorithms are able to predict trends in the stock market, the purpose of this research is to analyse how well they work [19]. The utilisation of previous price data will be carried out in order to achieve this goal. In order to evaluate the accuracy, robustness, and computational efficiency of each approach, we are going to take into consideration a wide range of criteria. This will allow us to determine which methodology is preferred. The techniques for data preparation, the methodologies for feature selection, and the metrics for model evaluation are all included in these parameters [13], [18]. By contrasting the results that

were obtained from the two different algorithms, our goal is to shed light on the benefits and drawbacks that are associated with each individual technique. We also believe that by comparing the results acquired from both algorithms, we will be able to provide insights into the actual implementations of these algorithms in genuine financial contexts. This is something that we feel will be possible.

II. Related Work

The numerous studies that have been carried out to study a wide range of techniques and methodologies [5, 9] are proof that major advancements have been made in the field of stock market prediction through the application of machine learning algorithms. These advancements have been made possible by the implementation of these algorithms. Tanuwijaya and Hansun [1] conducted research with the intention of utilising k-Nearest Neighbours (KNN) regression in order to forecast the LQ45 stock index. The research was carried out in order to study the applications of this technique. The findings of their analysis demonstrated that this method is effective in identifying patterns in historical data for the purpose of identifying trends. Researchers Sahu et al. [2] did research on the application of machine learning models for the aim of stock market forecasting. This research was conducted in a manner that is equivalent. During their presentation, they brought to the audience's notice the fact that certain algorithms, such as Support Vector Machines (SVM) and Random Forests (RF), are capable of producing accurate projections. In the course of their investigation, Alkhatib and his colleagues [3] focused their attention on the application of the KNN algorithm for the aim of predicting stock values. Specifically, they highlighted the method's user-friendliness as well as its interpretability. With the assistance of a categorization strategy that was developed by Subha and Nambi [4], it is possible to forecast the movement of stock indexes. It is the KNN software that is utilised in this method. By employing this strategy, the algorithm's adaptability is demonstrated in terms of its capacity to deal with a wide variety of prediction challenges. This is a demonstration of the algorithm's versatility.

Throughout the course of a number of further research initiatives, the implementation of machine learning algorithms that go beyond KNN has been the focus of examination because of its potential capabilities. Pathak [5] conducted an examination of the similarities and differences between a variety of different machine learning algorithms for stock market forecasting. One of the techniques under question was the neural network algorithm. Support Vector Machines (SVM), Random Forests (RF), and Gradient Boosting Machines (GBM) were the algorithms that were used in this study. The objective of the research was to analyse and evaluate the advantages and disadvantages of each method on its own. The application of machine learning models, such as Decision Trees and Naive Bayes, with the intention of predicting trends in the stock market was examined by Akash et al. [6], who generated positive findings in terms of accuracy and predictive performance. The models in question were used to anticipate patterns in the stock market. Naive Bayes and Decision Trees are two examples of models that fall under this category. This essay will provide insights into the practical application of predictive models in the world of financial markets. The objective of this essay is to provide these insights. Gomathy [7] developed a method for forecasting the stock market that is based on machine learning algorithms. This technology was used to make predictions.

The fundamental ideas and methodologies that support these strategies were taken into consideration by Shah et al. [8] when classifying these techniques into their respective categories. The process of conducting a comprehensive evaluation and taxonomy of prediction approaches in stock market analysis was the means by which this classification was achieved. At various points along the course of the inquiry, a comprehensive examination of the many methods that are utilised in the field was presented. Statistical methodology, machine learning algorithms, and deep learning approaches were some of the methods that were utilised and utilised in the process. When it comes to accurately predicting stock prices, Nikou et al. [9] conducted an analysis of the similarities and differences between deep learning algorithms and classic machine learning approaches. This investigation was carried out in the context of predicting stock prices. The authors highlighted the power of deep learning to detect sophisticated patterns and nonlinear relationships that are present in financial data in their discussion of the benefits of deep learning. This capability was highlighted in the authors' assessment of the advantages of deep learning. The purpose of the study that Deshmukh and his colleagues [10] conducted was to investigate the application of machine learning models, such as Neural Networks and Decision Trees, with the intention of predicting the behaviour of the stock market. The conclusions of this study demonstrated that these algorithms are capable of making accurate forecasts, as demonstrated by the data.

Previous research has made major contributions to the advancement of the area of stock market prediction, which has, in general, resulted in the field receiving significant benefits from those contributions. On the basis of the findings of this research, significant insights have been gained regarding the performance, application, and limitations of a number of different machine learning algorithms. Nevertheless, there is still a necessity for additional study involving the investigation of fresh approaches, the increase of the precision of predictions, and the management of practical obstacles that are experienced in real-world financial circumstances

III. Proposed Work

For the purpose of anticipating trends in the stock market, the suggested work implementation has a key emphasis on evaluating the accuracy of the machine learning models, more specifically the k-Nearest Neighbours (KNN) and XGBoost algorithms. One of the most important metrics is called accuracy, and it gives an indication of the percentage of true predictions that the models have made in comparison to the actual outcomes.

In order to determine how accurate the models are, the historical data from the stock market is separated into two sets: the training set and the testing set. For the purpose of training the models, the training set is utilised, whilst the testing set is utilised for the purpose of evaluating they perform. Following the completion of the training process, the models are put through their paces on the testing dataset, and the correctness of each model is determined.

When it comes to the KNN algorithm, the accuracy is assessed by comparing the percentage of stock market movements that were successfully forecasted to the total number of forecasts that were produced. A similar approach is taken with the XGBoost method, where the accuracy metric is used to determine the percentage of accurate predictions that the model is able to produce. Fig. 1 shows the implementation flow diagram.

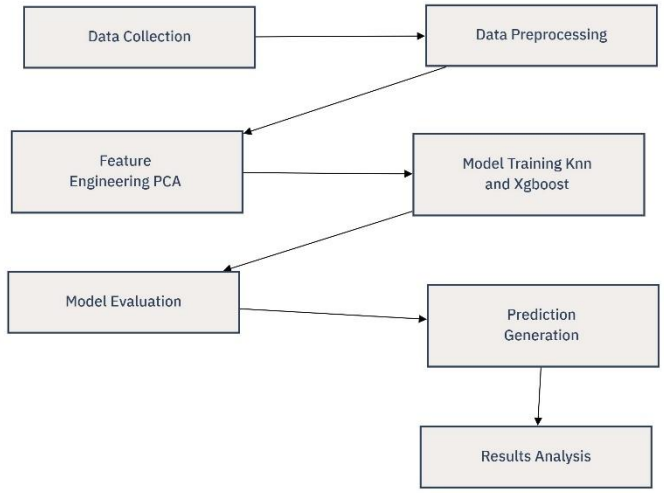

Fig. 1 Flow Diagram

The significance of the accuracy of both models cannot be overstated when it comes to judging how well they can forecast trends in the stock market. When the accuracy levels are higher, it indicates that the models are producing more accurate forecasts, which in turn provides investors and traders with informative and trustworthy information. The evaluation of accuracy is helpful in evaluating which algorithm is superior in terms of its ability to forecast stock market trends and in recognising the strengths and weaknesses of each model.

The KNN algorithm is a straightforward and simple algorithm that functions in this fashion. It is based on the concept of similarity and acts in this manner. When attempting to produce a forecast regarding the value of a particular data point, the majority class of the k nearest neighbours of the data point in question is utilised. When it comes to making predictions about the stock market, KNN analyses the features of previous data points, such as stock prices, volume, and percentage change, in order to identify patterns and trends that are analogous to those that have been observed in the past. KNN is able to estimate the most likely outcome for a given input by taking into account the proximity of data points resident in a space that is separated into several dimensions. This allows KNN to estimate the most likely outcome. In spite of this, the effectiveness of the method is heavily dependent on the choice of the distance measure as well as the number of neighbours (k) that are taken into consideration. For the purpose of achieving the highest potential level of performance, it is necessary to optimise both of these aspects.

Extreme Gradient Boosting, on the other hand, is an advanced ensemble learning technique that is renowned for its extraordinary performance in predictive modelling tasks. XGBoost is shortened as eXtreme Gradient Boosting. The gradient boosting algorithm is a member of the family of algorithms that are used in the process of gradually creating a powerful prediction model. Through the incorporation of a large number of weak learners, which are typically decision trees, it achieves this goal. Additionally, XGBoost makes use of a gradient descent optimisation technique in order to minimise a preset loss function. This ultimately results in the model's projected accuracy being continuously increased, which is a positive outcome. When it comes to making predictions about the stock market, XGBoost does an examination of historical market data in order to understand the complex patterns and relationships that exist in the market. The discovery of non-linear dependencies that may exist between different qualities is facilitated as a result of this. Due to its ability to efficiently deal with missing values, manage enormous datasets, and avoid overfitting, it has become a preferred choice for predictive modelling applications in the financial sector. This is because of its capacity to manage these things.

KNN and XGBoost each have their own set of benefits and drawbacks when it comes to predicting trends in the stock market. Both of these models have their own ways of doing things. The KNN algorithm is an excellent option for basic exploratory research and baseline modelling due to the fact that it is simple to understand and put into practice. It is possible, on the other hand, that it will struggle with high-dimensional data and noisy datasets, both of which will have an impact on the accuracy of its predictions. However, XGBoost is particularly successful at recognising complex patterns and non-linear correlations within the data, which eventually results in enhanced forecast accuracy. This is because XGBoost is able to recognise these patterns and correlations. In spite of this, it requires a bigger quantity of processing resources and the modification of settings in order to achieve the best possible performance.

The KNN and XGBoost algorithms are going to be incorporated into the work that is being presented in order to bring about the construction of a trustworthy prediction model for the purpose of anticipating trends in the stock market. Because of this, we will be able to make the most of the distinct advantages that each approach offers. This study's objective is to provide valuable insights into the efficiency of both algorithms by means of a thorough investigation and comparison of the accuracy achieved by both algorithms. This will be accomplished via the use of rigorous inspection and comparison. In addition to this, one of our goals is to identify the strategy that is most suitable for forecasting trends in the stock market.

IV. RESULTS

The findings of the research provide an in-depth examination of the effectiveness of the K-Nearest Neighbours (KNN) and XGBoost algorithms in forecasting the movements of the stock market. The research, which was carried out with the help of historical stock market data provided from Yahoo Finance, demonstrates that both techniques have the potential to produce positive results.

When it came to the KNN algorithm, the level of accuracy that was ultimately obtained was an amazing 96%. This demonstrates that the KNN model was able to accurately forecast the movements of the stock market with a high degree of precision. The usefulness of KNN in predicting future price movements was proved through the examination of past pricing data and the identification of patterns. However, it is essential to keep in mind that the performance of KNN is significantly dependent on a variety of factors, including the selection of the distance measure and the number of neighbours that are taken into consideration. It is essential to optimise these characteristics in order to get the highest possible level of predictive capabilities. Fig. 2 and 3 shows knn predictions

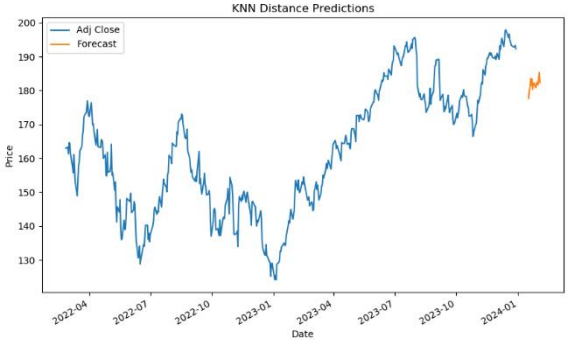

Fig. 2 Knn Distance Prediction

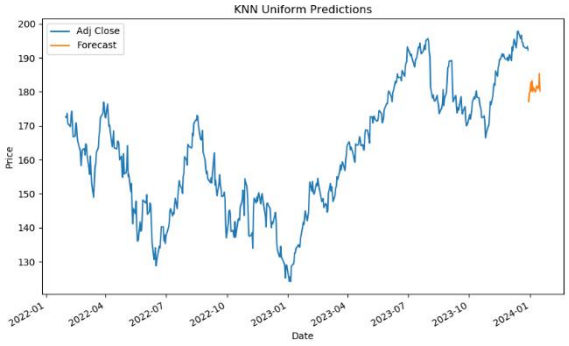

Fig. 3 Uniform Prediction

The XGBoost algorithm, on the other hand, demonstrated an even better level of accuracy, reaching an astonishing 98%. This demonstrates that XGBoost is both reliable and effective when it comes to managing massive datasets that contain high-dimensional features or characteristics. XGBoost is able to effectively capture complicated patterns and non-linear correlations in the data since it integrates the predictions of numerous weak learners. The fact that it is able to handle missing values and steer clear of overfitting is another reason why it is a popular option for predictive modelling in the financial industry. Fig. 4 shows the xgboost prections.

There are various benefits and drawbacks associated with both varieties of algorithms. Despite the fact that KNN is easy to comprehend and put into practice, it may have difficulty dealing with high-dimensional and noisy datasets, which could significantly influence the accuracy of its predictions. The XGBoost algorithm, on the other hand, is very effective at identifying complicated patterns and correlations, despite the fact that it requires more processing resources and parameter tuning.

The incorporation of both the KNN and XGBoost algorithms into the research makes it possible to develop a reliable prediction model that can anticipate movements in the stock market. The purpose of the study is to provide significant insights into the efficiency and applicability of each technique in real-world financial settings by harnessing the strengths of each approach.

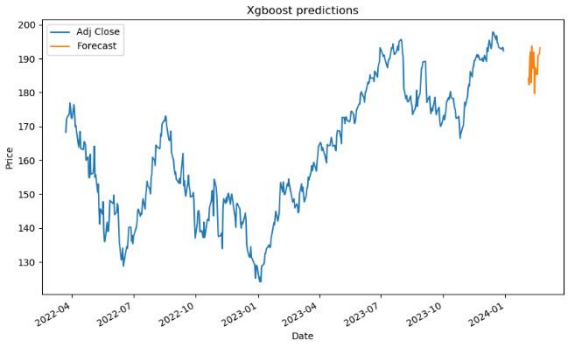

Fig. 4 Xgboost Prediction

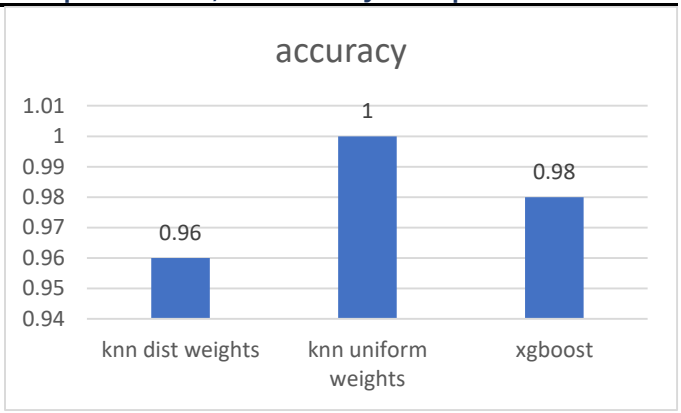

Fig. 5 Accuracy comparison

The results as seen fig. 5 of the research indicate that both the KNN and XGBoost algorithms have the potential to accurately forecast trends in the stock market for a significant amount of time. The comparative analysis that they conducted emphasises the advantages and disadvantages of each approach, providing useful insights that can be utilised in the future for the purpose of investment decision-making and financial analysis.

V. CONCLUSION

In conclusion, this study offers a complete analysis of the performance of the K-Nearest Neighbours (KNN) and XGBoost algorithms in terms of their capacity to forecast trends in the stock market when compared to one another. Both algorithms demonstrated promising levels of accuracy while using historical stock market data from Yahoo Finance. KNN achieved a maximum accuracy of 96%, while XGBoost achieved a maximum accuracy of 98%. In order to conduct an analysis of the stock market, both algorithms were utilised. When these data are taken into consideration, it becomes abundantly evident that algorithms for machine learning have the potential to be utilised in the process of financial forecasting. When it comes to catching intricate patterns and correlations, XGBoost is superior to KNN. On the other hand, KNN is differentiated by its simplicity and its ability to be interpreted. Following the completion of a comparative analysis of these algorithms, one is able to acquire valuable insights into the advantages and disadvantages that are associated with each method. To further improve the accuracy of forecasts, it is possible that in the future, additional research will be conducted to investigate hybrid models or ensemble approaches. In general, this research makes a contribution to the growing body of knowledge in the field of financial prediction and offers implications that may be utilised in the real world within the context of the process of making decisions regarding investments.

REFERENCES

[1] Tanuwijaya, Julius & Hansun, Seng. (2019). LQ45 Stock Index Prediction using k-Nearest Neighbors Regression. International Journal of Recent Technology and Engineering. 8. 10.35940/ijrte.C4663.098319.

[2] A. K. Sahu et al., "Stock Market Prediction Using Machine Learning," 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), Ghaziabad, India, 2023, pp. 512-515, doi: 10.1109/CICTN57981.2023.10140750.

[3] Alkhatib, Khalid & Hassan, Najadat & Hmeidi, Ismail & Shatnawi, Mohammed Khair. (2013). Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm.

[4] Subha, M.V. & Nambi, S.T.. (2012). Classification of stock index movement using k-nearest neighbours (k-NN) algorithm. WSEAS Transactions on Information Science and Applications. 9. 261-270.

[5] Pathak, Ashwini. (2020). Study of Machine learning Algorithms for Stock Market Prediction. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS060064.

[6] Akash, A & Rajaji, Shanthi & Aravinth, R & Vendhan, V & Veerapandi, D. (2019). Stock Market Trend Prediction using Machine Learning. International Journal of Innovative Research in Computer and Communication Engineering. 7. 1000-1006. 10.15680/IJIRCCE.2019.0702085.

[7] Gomathy, C K. (2021). THE STOCK MARKET PREDICTION SYSTEM.

[8] Shah, Dev & Isah, Haruna & Zulkernine, Farhana. (2019). Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. International Journal of Financial Studies. 7. 26. 10.3390/ijfs7020026.

[9] [Nikou, Mahla & Mansourfar, Gholamreza & Bagherzadeh, J.. (2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. Intelligent Systems in Accounting, Finance and Management. 26. 10.1002/isaf.1459.

[10] Deshmukh, Yogita & Saratkar, Deepmala & Tiwari, Yash. (2019). Stock Market Prediction Using Machine Learning. IJARCCE. 8. 31-35. 10.17148/IJARCCE.2019.8107.

[11] Shetty, Nisha & Pathak, Ashish. (2017). Indian Stock Market Prediction Using Machine Learning and Sentiment Analysis. 10.1007/978-981-10-8055-5_53.

[12] B. N. Varaprasad, C. Kundan Kanth, G. Jeevan and Y. K. Chakravarti, "Stock Price Prediction using Machine Learning," 2022 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2022, pp. 1309-1313, doi: 10.1109/ICEARS53579.2022.9752248.

[13] S. Mohammed, S. H. Krishna, P. K. Mudalkar, N. Verma, P. Karthikeyan and A. S. Yadav, "Stock Market Price Prediction Using Machine Learning," 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2023, pp. 823-828, doi: 10.1109/ICSSIT55814.2023.10061120.

[14] M. A. Nadif, M. T. Rahman Samin and T. Islam, "Stock Market Prediction Using Long Short-Term Memory (LSTM)," 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2022, pp. 1-6, doi: 10.1109/ICAECT54875.2022.9807655.

[15] B. N. Varaprasad, C. Kundan Kanth, G. Jeevan and Y. K. Chakravarti, "Stock Price Prediction using Machine Learning," 2022 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2022, pp. 1309-1313, doi: 10.1109/ICEARS53579.2022.9752248.

[16] A. Gupta, Akansha, K. Joshi, M. Patel and V. Pratap, "Stock Market Prediction using Machine Learning Techniques: A Systematic Review," 2023 International Conference on Power, Instrumentation, Control and Computing (PICC), Thrissur, India, 2023, pp. 1-6, doi: 10.1109/PICC57976.2023.10142862.

[17] Z. K. Lawal, H. Yassin and R. Y. Zakari, "Stock Market Prediction using Supervised Machine Learning Techniques: An Overview," 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Gold Coast, Australia, 2020, pp. 1-6, doi: 10.1109/CSDE50874.2020.9411609.

[18] A. K. Sahu et al., "Stock Market Prediction Using Machine Learning," 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), Ghaziabad, India, 2023, pp. 512-515, doi: 10.1109/CICTN57981.2023.10140750.

[19] F. L. Marchai, W. Martin and D. Suhartono, "Stock Prices Prediction Using Machine Learning," 2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE), Semarang, Indonesia, 2021, pp. 79-84, doi: 10.1109/ICITACEE53184.2021.9617222.

[20] A. Gangwar, A. Kumar and E. Bijpuria, "Stock Price Prediction using Machine Learning," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2021, pp. 189-193, doi: 10.1109/ICAC3N53548.2021.9725554.

[21] P. S and V. P. R, "Stock Price Prediction using Machine Learning and Deep Learning," 2021 IEEE Mysore Sub Section International Conference (MysuruCon), Hassan, India, 2021, pp. 660-664, doi: 10.1109/MysuruCon52639.2021.9641664.

[22] A. Kumar, S. Hooda, R. Gill, D. Ahlawat, D. Srivastva and R. Kumar, "Stock Price Prediction Using Machine Learning," 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2023, pp. 926-932, doi: 10.1109/CISES58720.2023.10183515.

[23] M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques," *2016 3rd International Conference on Computer and Information Sciences (ICCOINS)*, Kuala Lumpur, Malaysia, 2016, pp. 322-327, doi: 10.1109/ICCOINS.2016.7783235.