International Journal Of Engineering Research And Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 20, Issue 2 (February. 2024), PP. 43-49

Algorithmic Stock Market Prediction

Ankita Bharti, Shivani, Anushka Kumari, Jinisha Jain, B. Indra Thannaya
Department of Computer Science and Engineering, Indira Gandhi Delhi Technical University for Women, Kashmere Gate, Delhi – 110006;

Abstract
The prediction of stock market trends has intrigued not only traders but also computer engineers. Traditional prediction methods rely on historical stock data or analyse social media information. However, forecasting based solely on historical data encounters challenges due to evolving market patterns and potential oversights in certain data fields. For instance, the absence of a crucial parameter, like 'return rate,' in available data can hinder the accuracy of some prediction models. Conversely, models emphasizing the return rate alone may disregard other significant parameters, such as opening and closing prices. This paper delves into categorizing various predictive analytics methods across different domains, highlighting their shortcomings. In the context of using a static dataset for stock prices, particularly employing the random ridge regression model, the authors propose improvements to enhance accuracy in these predictive approaches.
Accurately forecasting stock market returns is a very challenging task due to the non-linear and volatility nature of the stock markets. With the development of artificial intelligence and increased processing capacity, it has become clear that pre-programmed methods work better for stock value predictions. The study deploys the closing prices of five companies from different industries were predicted using Random Forest and Artificial Neural Network techniques for the next day. The open, high, low, and closing prices of the stock (the financial data) are used to generate new variables that are added to the model. The trading model is established to study the performance of the proposed prediction algorithm against other benchmarks.
***KEYWORDS** : Stock Market, Random Ridge Regression, Price, Feature Selection, Mean Squared Error*

Date of Submission: 12-02-2024 Date of Acceptance: 26-02-2024

I. Introduction

The stock market, as a dynamic and complex financial ecosystem, has captivated the attention of investors, analysts, and researchers for decades. At its core, the stock market represents a vast network of interconnected entities, where the ebb and flow of stock prices reflect a myriad of factors ranging from economic indicators to investor sentiments. The ability to predict these price movements has been a longstanding pursuit, and in recent years, the integration of machine learning models has emerged as a promising avenue for enhancing predictive accuracy.

The volatility and swings witnessed in stock prices are a result of a combination of factors including economic data, company performance, international events, and investor behaviour. Making accurate forecasts requires figuring out the numerous patterns inside this dense network of variables. Understanding the stock market and stock prediction.

A company's perceived value in the open market is reflected in its stock price. A higher stock price for the business itself may indicate investor confidence and facilitate capital raising through share issues or acquisitions. On the other hand, a declining stock price can be a sign of difficulties or worries about the performance of the business, which could limit its capacity to obtain funding and seize expansion chances about the performance of the business, which could limit its capacity to obtain funding and seize expansion chances. Similarly, The stock price gives users—such as analysts or investors—information about what the market believes about the company and what expectations there are. Stock prices can be used by users to help them evaluate investment possibilities and control risk. They can do this by examining variables like valuation, performance trends, and market dynamics.

By analyzing vast amounts of historical data to find patterns and trends that could influence future market movements, a machine learning model for stock price prediction helps human analysts in their market analysis. Analysts are able to make well-informed decisions on investment strategies, risk management, and portfolio optimization by utilizing these insights. Because of the model's predictive powers, analysts may project future stock values and modify their methods in response to market volatility. Furthermore, analysts can respond quickly to shifting market conditions thanks to real-time monitoring capabilities, which guarantees prompt and well-informed decision-making. In the end, the combination of human analysis with machine learning models improves the precision and breadth of market analysis, enabling analysts to more skillfully negotiate the intricacies of the financial markets.

II. Methodology and Analysis

Why ML models are best for Stock Price Prediction?
The ability of machine learning (ML) algorithms to identify complicated patterns in large, detailed datasets makes them the preferred choice for stock price prediction. Because of their flexibility, they may continually improve forecasts in response to changing market conditions, giving analysts data-driven insights to help them make wise decisions. ML models increase productivity and free up analysts to work on strategic objectives by automating analysis duties and efficiently managing massivedatasets, and thus with ML model a good dataset is also important for efficient results.

Why Supervised learning is better than Unsupervised?
Supervised learning has a distinct benefit over unsupervised learning for stock price prediction since it is based on labelled data, allowing models to learn from prior stock prices linked with known results. Predictive models are trained using this labelled data, which gives them important direction and helps them identify intricate links and patterns in the data. Compared to unsupervised techniques, supervised learning models—like regression or classification algorithms—offer better predicted accuracy, interpretability, and risk management.
For Stock Market Analysis, unsupervised learning is not the best option because it does not have the labelled data needed to train predictive models. The intricate and ever-changing characteristics of stock market data necessitate a sophisticated comprehension, and unsupervised techniques could find it difficult to identify significant trends without precise direction from labelled results.

What all algorithms should we consider for prediction and why?
There are several algorithms that can be used for stock price prediction and for the purpose of our research, we are going with the Linear Regression models, Random Forest models, Ridge Regression model, etc.

Model Building
The dataset used in this research work are historical daily stock prices obtained by NYSE, BSE and S&P 500. The data is composed of essential features like P/E Ratio, Dollar Exchange Rate, VWAP, NAV etc. In this model, stock price predicts the value of stock in the near future, as models are more efficient for short-term prediction. To determine the best price, many experiments were performed such as: Reducing RMSE values, Reducing MSE error and Maximise R2 score.
1.Dataset Description :
VWAP (Volume Weighted Average Price) calculates the average stock price, considering both volume and price, offering valuable insights into market trends. The General Index indicates the associated general index, providing information on broader market conditions. The NAV (Net Asset Value) reflects the net asset value of each stock, contributing to an understanding of overall asset valuation. P/E Ratio (Price-to-Earnings Ratio) provides insights into the stock's valuation relative to its earnings. Tracking Error signifies the standard deviation between a stock's returns and a benchmark, offering insights into performance variability. Dollar Exchange Rates provides exchange rates in terms of dollars for international stocks.
2.Data Preprocessing& Feature Selection:
Numerical features underwent normalization as part of the data preprocessing steps. This crucial transformation ensures uniformity and consistency in the scale of numerical attributes, preventing any disproportionate impact during subsequent analyses or modelling. Categorical variables were meticulously handled through Label Encoding. It facilitates the representation of categorical data in a numerical format, enabling machine learning models to derive meaningful insights from these variables.
A detailed scrutiny of each feature aimed to identify its relevance and contribution to the predictive model. This visualization not only offered a comprehensive depiction of the inter-feature relationships but also guided the precise selection of features with a noteworthy impact on predicting stock prices.
3.Train-Test Split:
In a pivotal stage of the model development process, it was effectively partitioned into two segments: training data and testing data, maintaining a standardized 80:20 ratio. The training set, constituting 80% of the data, is dedicated to refining the model's parameters and enabling it to discern intricate patterns and relationships within the dataset. In contrast, the testing set, encompassing the remaining 20%, provides an independent dataset that the trained model has not encountered during its learning phase.
4.Model Selection:
Numerous regression models can be employed for predicting numerical continuous data, particularly in the context of stock price prediction. Options considered in this context involve models such as the Random Ridge

Regressor and Linear Regression. The Random Ridge Regressor, known for its efficacy in handling collinearity and overfitting, is chosen for datasets with potentially correlated features. Simultaneously, the traditional yet robust Linear Regression model is explored for its simplicity and interpretability. This dual exploration aims to identify the most effective and interpretable approach for predicting continuous stock price values without duplicating existing content.

5. Model Evaluation Metrics:

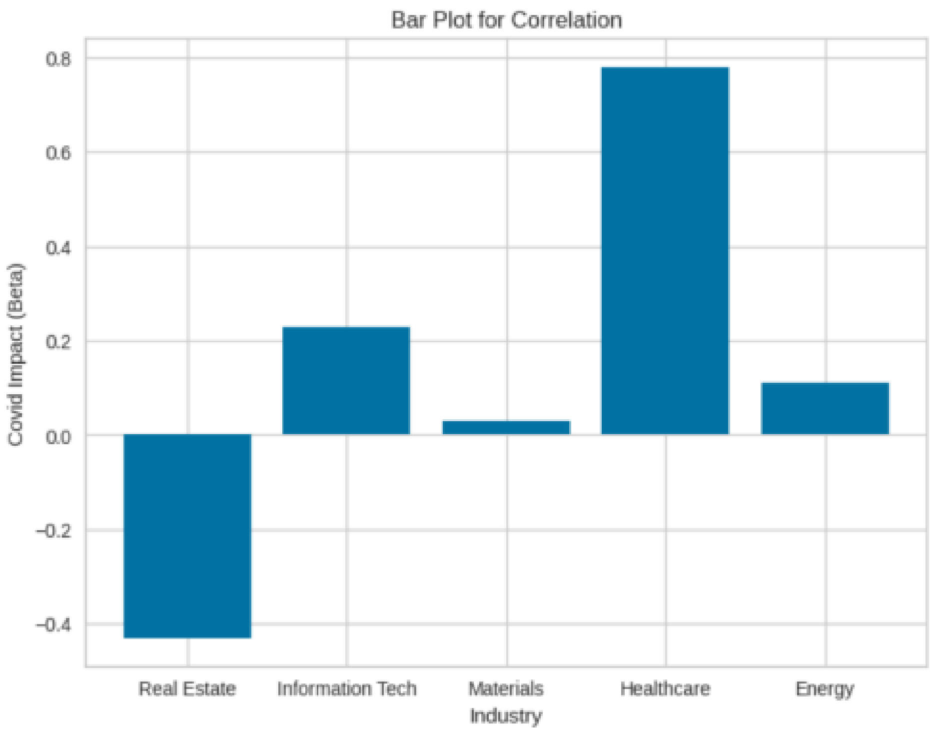
In the evaluation phase, our focus was on assessing models using criteria specifically tailored for numerical datasets. We employed key metrics such as Mean Square Error(MSE), and R-squared (R2) score to measure the accuracy and precision of predictions.

6.Why this selection criteria?

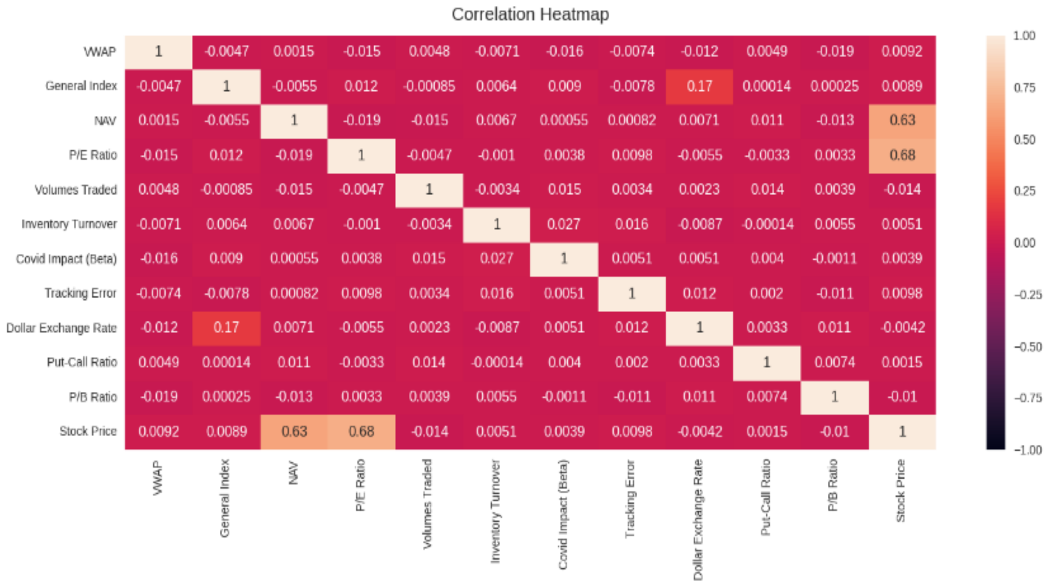
Mean Squared Error (MSE) is favoured for comparing supervised models due to its ability to provide a quantitative measure of prediction accuracy. By calculating the average of squared errors, MSE emphasizes larger deviations, making it effective in assessing overall model performance. Lower MSE values indicate superior predictive capabilities, allowing for easy differentiation between models. Its mathematical convenience facilitates optimization processes. It's essential to consider the specific characteristics of the problem and the implications of errors in choosing the most suitable metric for a given application, as alternatives like Mean Absolute Error or Root Mean Squared Error may also be relevant.

The R2 score, or coefficient of determination, is crucial in supervised models as it quantifies the proportion of variance in the dependent variable predictable from independent variables. It gauges the model's goodness of fit, indicating how well it explains data variability. A high R2 score (near 1) indicates the model effectively captures underlying patterns, boosting confidence in its predictive ability. Analysts use it to evaluate overall model performance and accuracy in making predictions from input features.

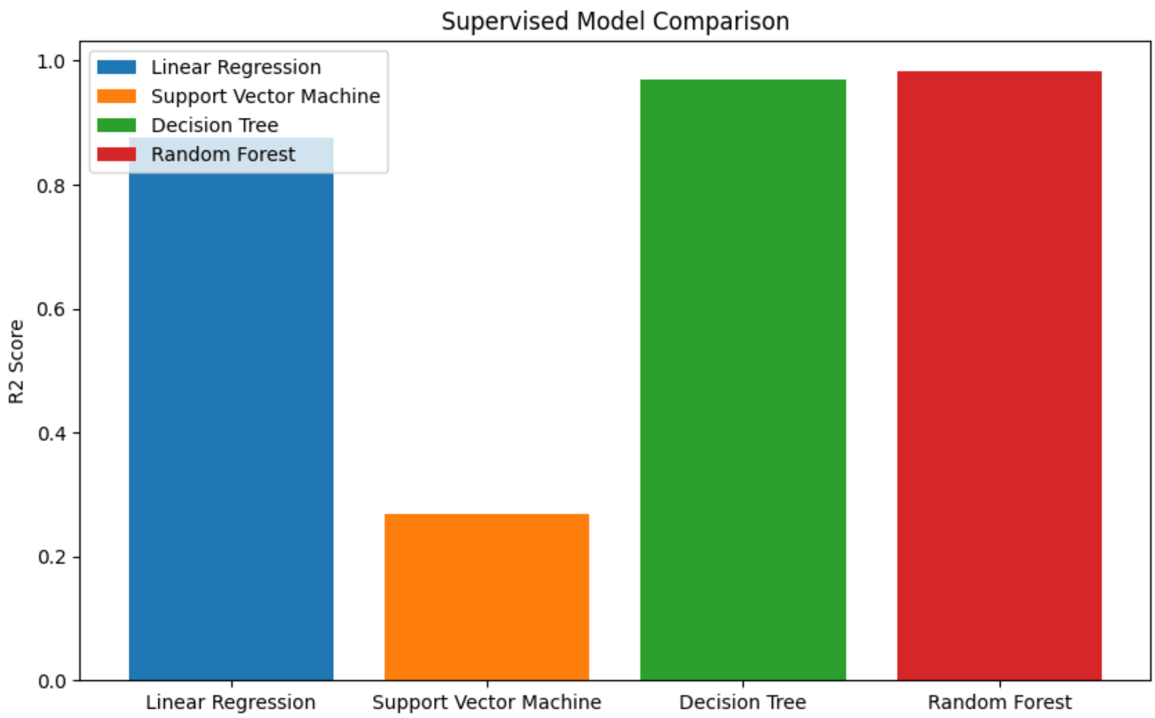
III. Results and Discussions



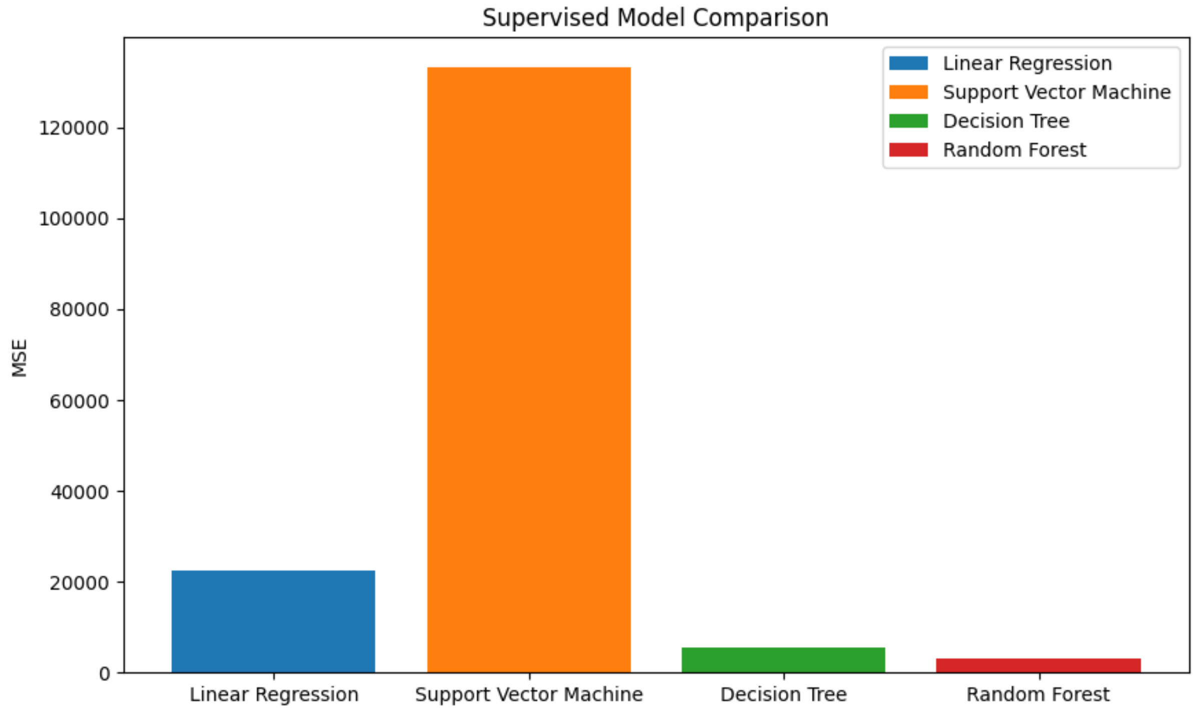
In several industries, stock markets saw notable swings during the Covid epidemic. While the technology and healthcare sectors experienced resilience and even growth, traditional industries such as travel and real estate saw reductions. The graph depicts the varied effects of the pandemic on various industry stock types, which reflects the dynamic state of the world economy at the time.



The correlation graph depicts how the input variables relate to one another. This visualization helps make sense of the complex aspects influencing stock market trends and provides valuable insights for predictive modelling and investment strategies.

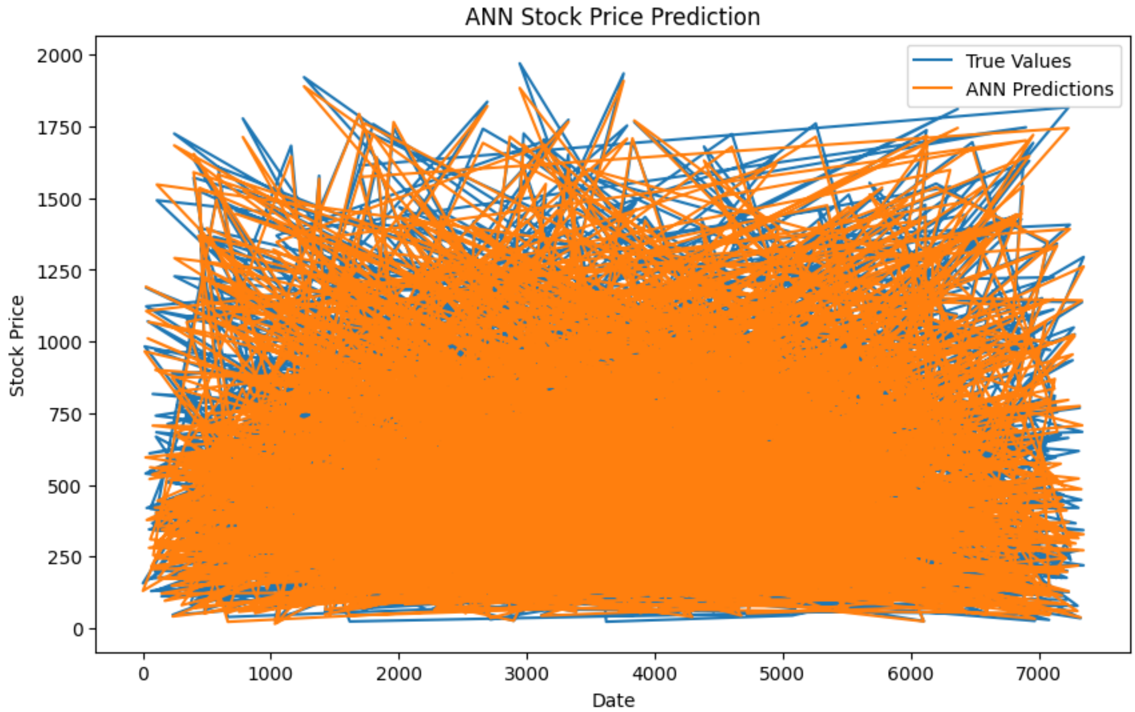


The graph displays the R2 values of several stock market prediction models, including decision trees, random forests, logistic regression, and linear regression. The models' prediction accuracy is assessed using the R2 scores as a measure. The R2 score quantifies the percentage of the dependent variable's volatility that can be predicted based on the independent variables. R2 is a measure of the model's goodness of fit in the context of supervised learning, reflecting how well the model explains the variability in the data. With the greatest R2 score among the models of about 0.982, the random forest model performs better than the others.This suggests that when compared to decision tree, logistic, and linear models, the random forest method offers a more reliable and accurate forecast of stock market behavior.



The graph displays the Mean Squared Error (MSE) for the decision tree, random forest, logistic regression, and linear regression models. Since Mean Squared Error (MSE) can give a quantifiable estimate of prediction accuracy, it is preferred for comparing supervised models. Better prediction abilities are indicated by lower MSE values, making it simple to distinguish between models. Optimization methods are facilitated by their mathematical convenience. It is clear from this graph that the random forest model has the lowest mean square error (MSE) of 3165.68 out of the four, indicating better predictive accuracy. This result highlights how well the random forest algorithm performs in reducing prediction errors when compared to decision tree models, logistic regression, and linear regression in thedataset.

After examining regression models' R2 scores and MSE values, we find that Random Forest was the best option for our dataset. Its improved performance over other models is demonstrated by higher R2 scores and lower MSE values, which highlight how well it captures complicated relationships in the data. Based on our dataset, Random Forest is the recommended model for reliable insights and accurate forecasts.



The graph depicts the proximity of predicted values to real values after incorporating an Artificial Neural Network (ANN) into the model. Prediction accuracy increases with the distance between the spots and the diagonal line. Furthermore, the goodness of fit between predicted and actual values is measured quantitatively using the R-squared (R2) score. A higher R2 score—which is close to 1—indicates a significant correlation and shows how well the ANN captured and explained data variability, supporting the predictive performance of the model.

When we analyzed the dataset with an Artificial Neural Network (ANN) model, we found that the model accuracy was lower than with Random Forest. This decrease shows that Random Forest may be a better fit for capturing the underlying patterns in this specific dataset than ANN. The features of the dataset or the design of the model may have had an impact on ANN's performance, despite its intricacy and capacity to learn complex relationships.

IV. Conclusion

This extensive research, leveraging Random Ridge Regression, has demonstrated exceptional performance, achieving an impressive accuracy rate of 98.2%. This high level of accuracy signifies the model's capability to make precise predictions concerning stock prices. The model's success lies in its adaptability to the dataset's complexities, striking a balance between simplicity and sophistication. Our research focused on predicting stock prices using various regression models, and the outcomes underscore the effectiveness of the Random Ridge Regression model. Extensive experimentation and evaluation reveal it as the most successful model for our dataset, showcasing its ability to handle multicollinearity, feature selection, and regularization. Moreover, future research endeavours may explore additional datasets and conduct thorough comparative analyses with other advanced regression models.

References

[1]. S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 2017, pp. 1643-1647, doi: 10.1109/ICACCI.2017.8126078. keywords: {Companies;Time series analysis;Logicgates;Predictivemodels;Forecasting;Datamodels;Machinelearning;Timeseries;Stockmarket;RNN;LSTM;CNN}

[2]. M. R. Vargas, C. E. M. dos Anjos, G. L. G. Bichara and A. G. Evsukoff, "Deep Leaming for Stock Market Prediction Using Technical Indicators and Financial News Articles," 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 2018, pp. 1-8, doi: 10.1109/IJCNN.2018.8489208. keywords: {Predictive models;Machinelearning;Biological system modeling;Recurrent neural networks;Computationalmodeling;Stochasticprocesses;Convolutional neural networks;Deeplearning;Recurrent neural network;Convolutional neural network;Long Short-Term Memory;Stocks forecasting.}

[3]. M. Misra, A. P. Yadav and H. Kaur, "Stock Market Prediction using Machine Learning Algorithms: A Classification Study," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), Bhubaneswar, India, 2018, pp. 2475-2478, doi: 10.1109/ICRIEECE44171.2018.9009178. keywords: {Stock markets;Support vector machines;Predictivemodels;Predictionalgorithms;Radiofrequency;Artificial neural networks;Principal component analysis;DataAnalysis;MachineLearning;PredictiveAnalysis;StockPrediction;Linear Regression.}

[4]. Z. Hu, J. Zhu and K. Tse, "Stocks market prediction using Support Vector Machine," 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering, Xi'an, China, 2013, pp. 115-118, doi: 10.1109/ICIII.2013.6703096. keywords: {Support vector machines;Training;Companies;Investment;Predictivemodels;Biological system modeling;Stockmarkets;stockclassification;datamining;support vector machine;forecasting;multivariate classification}

[5]. 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. keywords: {Support vector machines;Systematics;Scientificcomputing;Time series analysis;Semantics;Artificial neural networks;Predictivemodels;Stock Market Prediction;Supervised Machine Learning;Classification;Regression;Support Vector Machine (SVM);Artificial Neural Network (ANN)}

[6]. R. Iacomin, "Stock market prediction," 2015 19th International Conference on System Theory, Control and Computing (ICSTCC), CheileGradistei, Romania, 2015, pp. 200-205, doi: 10.1109/ICSTCC.2015.7321293. keywords: {Support vector machines;Stockmarkets;Principal component analysis;Algorithm design and analysis;Covariancematrices;Predictionalgorithms;Marketresearch;NonlinearSystems;ComputationalIntelligence;Optimization}

[7]. R. Gandhi, G. Anand, V. Vekaria, T. Ahuja and L. D'Mello, "Stock Market Graph Prediction using LSTM," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-4, doi: 10.1109/ICCCNT56998.2023.10306549. keywords: {Machine learning algorithms;Computationalmodeling;Time series analysis;Neuralnetworks;Finance;Predictivemodels;Marketresearch;Time series analysis;Yahoofinance;LSTM;ARIMA;HMM;PCA;SVM}

[8]. A. Yadav, V. Kumar, S. Singh and A. K. Mishra, "A Survey on Stock Price Prediction using Machine Learning Techniques," 2023 Winter Summit on Smart Computing and Networks (WiSSCoN), Chennai, India, 2023, pp. 1-12, doi: 10.1109/WiSSCoN56857.2023.10133865. keywords: {Surveys;Machine learning algorithms;Social networking (online);Companies;Predictivemodels;Predictionalgorithms;Stockmarkets;StockMarket;Trading;Investment;Machine Learning(ML) Algorithm;Prediction;LSTM;SVM;GRU;ARIMA;ARMA;etc}

[9]. A. Naik, V. Gaikwad, R. Jalnekar and M. Rane, "Daily Stock Price Direction Prediction using Random Multi-Layer Perceptron," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 2021, pp. 1-9, doi: 10.1109/AIMV53313.2021.9670927. keywords: {Support vector machines;Uncertainty;Automation;Databases;Machinevision;Machinelearning;Predictionmethods;MachineLearning;StockMarket; Multi-LayerPerceptron;DataScience;ArtificialIntelligence;Artificial Neural Network;Sigmoid;Classifier;Regression}

[10]. P. K. Bharne and S. S. Prabhune, "Stock Market Prediction Using Artificial Neural Networks," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 64-68, doi: 10.1109/ICCS45141.2019.9065702. keywords: {Security;Stockmarkets;Estimation;Organizations;Conferences;Controlsystems;Artificialintelligence;Artificial Neural Networks (ANNs);Stock Market;Prediction}

[11]. D. Hasibuan, I. K. Jaya, B. Rumahorbo, J. Naibaho, J. Napitupulu and E. Rajagukguk, "Time Series Financial Market Forecasting Based On Support Vector Regression Algorithm," 2019 International Conference of Computer Science and Information Technology (ICoSNIKOM), Medan, Indonesia, 2019, pp. 1-4, doi: 10.1109/ICoSNIKOM48755.2019.9111533. keywords: {Financial Market;Forecasting;SVM;SVR;linear}

[12]. Thutte, Vaibhav V., P. T. Talole, and Parihar BM Jadhav Ab. "A review on stock market prediction using Artificial Neural Networks." Int J CurrEng Sci Res (IJCESR) 5.4 (2018): 63-67

[13]. Choudhry, Rohit, and Kumkum Garg. "A hybrid machine learning system for stock market forecasting." International Journal of Computer and Information Engineering 2.3 (2008): 689-692

[14]. Sakshi, S., Kar, A., Ningthoujam, C. (2023). Comparative Analysis of Machine Learning Models to Predict Stock Market Price. In: Dhar, S., Do, DT., Sur, S.N., Liu, H.CM. (eds) Advances in Communication, Devices and Networking. Lecture Notes in Electrical Engineering, vol 902

[15]. Sahoo, P. K., and Krishna Charlapally. "Stock price prediction using regression analysis." International Journal of Scientific & Engineering Research 6.3 (2015): 1655-1659.

[16]. Bharathi, Shri, and Angelina Geetha. "Sentiment analysis for effective stock market prediction." International Journal of Intelligent Engineering and Systems 10.3 (2017): 146-154.

[17]. Sharma, Mrityunjay. "Survey on stock market prediction and performance analysis." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 3.1 (2014).

[18]. Madge, Saahil, and Swati Bhatt. "Predicting stock price direction using support vector machines." Independent work report spring 45 (2015).

[19]. Gondaliya, Chetan, Ajay Patel, and Tirthank Shah. "Sentiment analysis and prediction of Indian stock market amid Covid-19 pandemic." IOP Conference Series: Materials Science and Engineering. Vol. 1020. No. 1. IOP Publishing, 2021.

[20]. Wang, Yaojun, and Yaoqing Wang. "Using social media mining technology to assist in price prediction of stock market." 2016 IEEE International conference on big data analysis (ICBDA). IEEE, 2016.

[21]. Zhao, Lei, and Lin Wang. "Price trend prediction of stock market using outlier data mining algorithm." 2015 IEEE Fifth International Conference on Big Data and Cloud Computing. IEEE, 2015.

[22]. Sun, J. H. "Long time series clustering method and its application in stock price." Diss. WuHan University (2011).

[23]. Bao, Wei, Jun Yue, and Yulei Rao. "A deep learning framework for financial time series using stacked autoencoders and long-short term memory." PloS one 12.7 (2017): e0180944.

[24]. Shah, Dev, Haruna Isah, and Farhana Zulkernine. "Stock market analysis: A review and taxonomy of prediction techniques." International Journal of Financial Studies 7.2 (2019): 26.

[25]. Xu, Mingying, et al. "A scientific research topic trend prediction model based on multi-LSTM and graph convolutional network." International Journal of Intelligent Systems 37.9 (2022): 6331-6353.

[26]. Liu, Hui, and Zhihao Long. "An improved deep learning model for predicting stock market price time series." Digital Signal Processing 102 (2020): 102741.