iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** I **Month of publication:** January 2023

DOI: https://doi.org/10.22214/ijraset.2023.48931

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue I Jan 2023- Available at www.ijraset.com

Stock Market Prediction Techniques: A Literature Review

Om Nath[1], Swapan Shakhari[2]
[1, 2]*Computer Science, RCC Institute of Information Technology*

Abstract*: This literature review summarizes the existing research on the use of machine learning for stock market prediction. The review covers studies from various sources such as journals, conference proceedings, and theses. The methods used for stock market prediction using machine learning include decision trees, support vector machines, artificial neural networks, and time-series analysis. The review also highlights the advantages and limitations of these methods, as well as their applications in the stock market. The findings of the review indicate that machine learning has the potential to provide valuable insights into the stock market, but there is still room for improvement in terms of accuracy and robustness. The review concludes by suggesting future directions for research in this field.
Keywords: Stock Price Prediction, Stock Market Prediction, LSTM, Regression Model, CNN, Machine Learning.

I. INTRODUCTION

Stock market prediction has long been an important topic in finance and economics, as accurate predictions can help investors, traders, and decision makers make informed decisions and potentially achieve higher returns. With the advent of technology and the availability of large amounts of data, machine learning has become a popular approach for stock market prediction. Machine learning algorithms can learn patterns in historical stock market data and use these patterns to make predictions about future stock prices.

In recent years, there has been a growing body of research on the use of machine learning techniques for stock market prediction. This literature review aims to provide a comprehensive overview of the current state of research in this area. The review will summarize the different machine learning algorithms used for stock market prediction, discuss the factors that can impact prediction accuracy, and provide a critique of the existing studies. By synthesizing the existing literature, this review will provide insights into the strengths and limitations of machine learning for stock market prediction and suggest areas for future research.

So now, In Digital era, after introduction of Machine learning. The machine learning plays vital role in the different fields of prediction and Computer vision. Some of the popular area where machine learning plays vital role are stock market prediction[1-2], sentiment prediction[3], house rent price prediction[4], heart disease prediction[5], flight delay prediction[6], fish classification[7], image processing[7], etc. By using machine learning it's been easy for the device to train and test the huge data. In stock market, with the help of machine learning it been easy to analyse and filter the huge historical data of companies at once, and it to selects the right data set, process it, and give the maximum possible accurate data. Some of the most common and popular methods for predictions are decision trees[8], support vector machines[9], artificial neural networks[10], time-series analysis[11], LSTM[12] etc.

In Section II of this paper, some of the popular and commonly used techniques have been discussed based on the different journals. In Section III, the challenges faced by the researchers. Section IV, conclusion of the literature review.

II. STOCK MARKET PREDICTION TECHNIQUES

Stock price prediction has been a hot topic of research for many years and various approaches have been used to make predictions. Machine learning is one of the popular approaches used for stock price prediction. There have been several studies done on the use of decision trees, support vector machines, artificial neural networks, time-series analysis and LSTM.

A. Decision Trees

Decision trees are a commonly used machine learning technique for stock market prediction. The process involves creating a tree-like model that makes predictions based on a series of decision rules, each branching into more specific rules based on certain conditions. The rules are determined by identifying patterns and correlations in historical stock market data, such as prices, volume, and news events.

©IJRASET: All Rights are Reserved

The prediction is made by starting at the root node and following the decision rules along the branches of the tree until a terminal leaf node is reached, which provides the final prediction. However, the accuracy of decision tree predictions in stock market forecasting is limited by the quality of the training data and the potential for overfitting.

There are several types of decision trees used in machine learning, including:

1) *Classification and Regression Trees (CART):* This is the most basic type of decision tree and can be used for both binary and multi-class classification, as well as regression problems. It splits the data based on the feature that results in the best reduction in impurity, such as entropy or Gini impurity.
2) *ID3 (Iterative Dichotomiser 3):* This is an early decision tree algorithm that was developed for binary classification problems. It splits the data based on the feature that maximizes the information gain, which measures the reduction in entropy.
3) *C4.5:* This is an improvement on ID3 that can handle both continuous and categorical features, as well as missing values. It uses a more sophisticated approach to select the best feature for splitting, considering both information gain and gain ratio.
4) *CHAID (Chi-squared Automatic Interaction Detection):* This is a decision tree algorithm designed for categorical features, and is particularly useful in market research and customer segmentation problems. It uses a chi-squared test to determine the significance of each feature in predicting the target.
5) *Random Forest:* This is an ensemble method that combines multiple decision trees to make predictions. It addresses the overfitting problem by randomly selecting a subset of features for each tree and aggregating the predictions of individual trees to make a final prediction.
6) *Gradient Boosting Decision Trees (GBDT):* This is another ensemble method that combines multiple decision trees, but in a sequential manner. It uses the errors from previous trees to weight the data and focuses the next tree on the most difficult examples.

Kamble [8], used Random Forest because the Random Forest algorithm is a powerful machine learning technique that uses an ensemble of decision trees to make predictions. It works by first creating multiple subsets of the data using a random sampling method, known as bootstrapping. For each subset, a decision tree is constructed using a criterion such as information gain or entropy to determine the best splits.

During the tree building process, each attribute is given a vote for each split, and the attribute with the highest number of votes is selected as the splitting node. This process is repeated for each tree, until each tree reaches its maximum depth[8][13].

Finally, the predictions from each tree are combined to make a final prediction. This can be done through voting or averaging, depending on the problem type. The unique aspect of the Random Forest algorithm is its ability to provide a more robust prediction by combining the predictions of multiple decision trees, each of which only considers a subset of the data and features[8].

Hindrayani et. al.[13], used four different techniques i.e., Multiple Linear Regression, Support Vector Regression, Decision Tree Regression, and K-Nearest Regression and gave comparison data of it.

TABLE I
COMPANY TABLE

Number	Stock Code	Company Name
1	TLKM	Telekomunikasi Indonesia
2	EXCL	XL Axiata
3	FREN	Smartfren
4	ISAT	Indosat

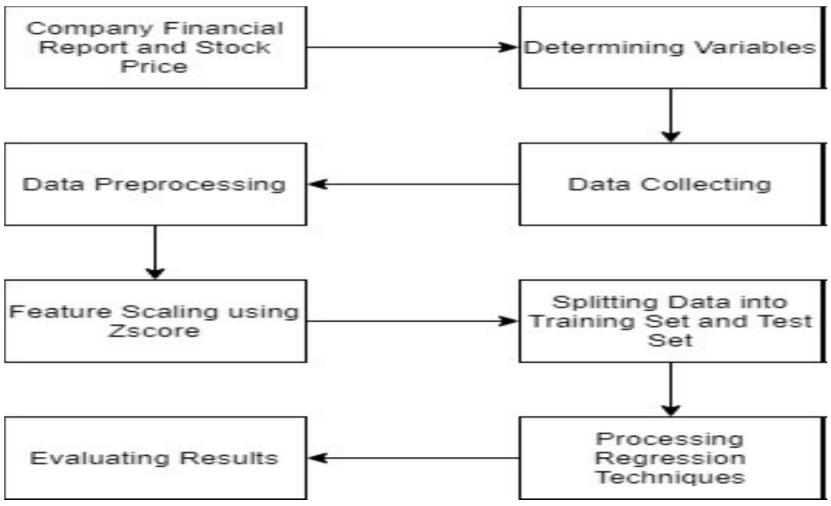


Fig 1: Research Methodology

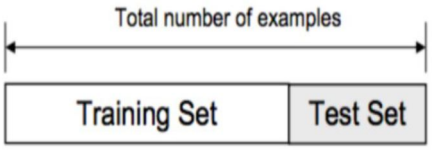


Fig 2: Train Test Split

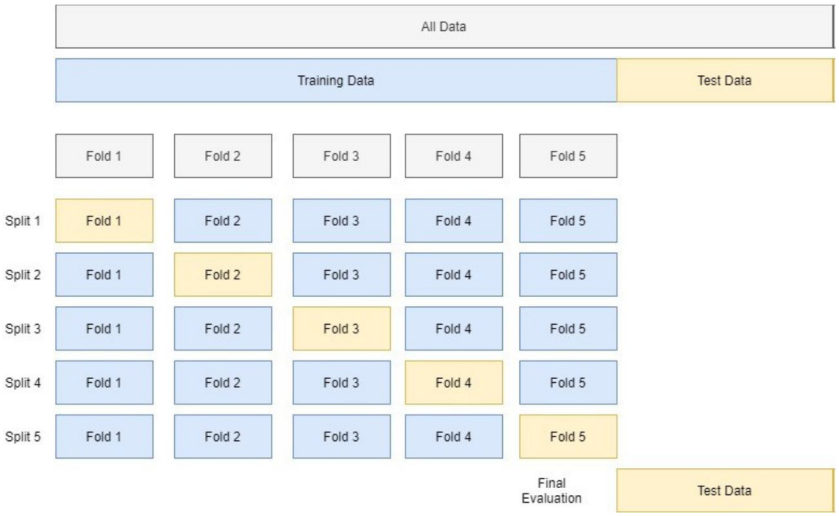


Fig 3: Cross Validation K-Fold

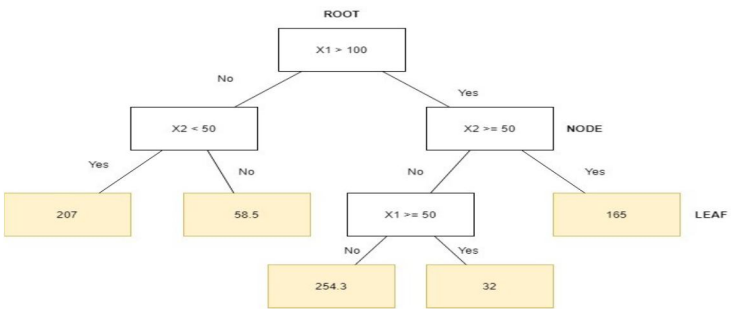


Fig 4: Decision Tree regression examples

TABLE II
CORRELATION COEFFICIENT RESULT

	TLKM	EXCL	FREN	ISAT
Multiple Linear Regression	0.757	0.291	0.692	0.822
SVR	0.555	0.513	0.763	0.84
Decision Tree Regression	0.932	0.86	0.902	0.977
K-Nearest Regression	0.919	0.853	0.899	0.973

TABLE IIII
MAPE RESULTS WITH COMMON TRAIN TEST SPLIT DATA

	TLKM	EXCL	FREN	ISAT
Multiple Linear Regression	11.474%	16.548%	73.078%	54.979%
SVR	9.265%	15.819%	50.726%	52.212%
Decision Tree Regression	2.991%	7.745%	12.859%	8.314%
K-Nearest Regression	3.208%	8.034%	12.983%	8.79%

TABLE IIIV
MAPE RESULTS WITH K-FOLD CROSS VALIDATION

	TLKM	EXCL	FREN	ISAT
Multiple Linear Regression	11.417%	16.578%	72.151%	54.795%
SVR	9.278%	15.869%	52.314%%	47.585%
Decision Tree Regression	2.966%	7.774%	12.909%%	8.28%
K-Nearest Regression	3.313%	8.179%	16.131%	9.783%

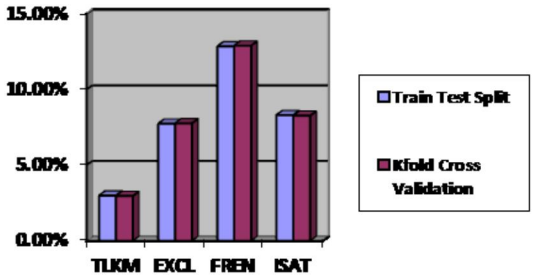


Fig. 2: MAPE Results of Decision Tree Regression

B. Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are a type of machine learning algorithm used for classification and regression tasks. In the context of stock market prediction, SVMs can be used to classify whether the stock price will increase or decrease.

The basic idea behind SVMs is to find the hyperplane (a line or a higher dimensional plane) in a high-dimensional space that best separates the data points into different classes. The SVM algorithm then optimizes the hyperplane to maximize the margin between the two classes, which helps to improve the accuracy of predictions.

SVMs can handle non-linearly separable data by transforming the input data into a higher dimensional space where a linear separation is possible. The algorithm then uses a kernel function to perform this transformation, which enables SVMs to perform well on a wide range of data sets.

There are two main types of Support Vector Machines (SVMs):

1) Linear SVM: It classifies data into two classes using a straight line or hyperplane.

2) Non-linear SVM: It classifies data into two classes using a non-linear boundary. It is often achieved by transforming the data into a higher dimension where the boundary can be linear.

Gururaj et. al., used two different techniques, Linear Regression and Support Vector Machines (SVMs). These are two popular machine learning techniques. Here is a brief comparison:[14]

Linear Regression: It is a supervised learning technique that models a linear relationship between input variables (predictors) and a continuous target variable (response). The goal is to find the line of best fit that minimizes the sum of squared errors between the predicted values and actual values.

Support Vector Machines (SVMs): It is a supervised learning technique that separates two classes in a high-dimensional space using a hyperplane. The goal is to find the hyperplane that maximizes the margin, i.e. the distance between the hyperplane and the closest data points from either class, which are known as support vectors.

Both techniques can be used for classification and regression tasks, but SVMs are more appropriate for problems where the classes are highly imbalanced, non-linearly separable, or have many outliers[14].

Kofi Nti et. al.[15], in their paper they introduced a new "homogeneous" ensemble classifier, GASVM, based on a Genetic Algorithm (GA) for optimizing SVM parameters and feature selection to predict 10-day stock price movement on Ghana Stock Exchange (GSE). The GASVM outperformed state-of-the-art models (DT, RF, NN) in accuracy metrics such as RMSE, MAE, AUC, Accuracy, and Recall. The GA was introduced as a feature selection mechanism to optimize SVM factors, resulting in improved accuracy compared to conventional SVM, Random Forest, Decision Trees, and Neural Networks.

The current study combined feature selection and parameter optimization to achieve a large search space in the Ghana stock exchange. The proposed GASVM technique achieved 93.7% accuracy and eliminates the need for manual optimization of SVM. The study only used a genetic algorithm for optimization, future research could consider other techniques and the effects of customer sentiment and financial news on stock price movement[15].

C. Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) emulate the human brain's architecture and functions to create a form of machine learning. In the context of stock market prediction, ANNs can be used to predict stock prices by analyzing a range of factors, such as historical prices, economic indicators, and news events.

An ANN is composed of layers of interconnected artificial neurons, functioning as nodes to process and pass on information. Each neuron receives inputs, performs a computation, and generates an output that is passed to the next layer of neurons. The ANN is trained using a large dataset of historical stock market data, where the inputs are the relevant factors and the outputs are the corresponding stock prices.

The ANN adjusts the weights of the connections between neurons in order to minimize the difference between the predicted and actual stock prices. This process is repeated multiple times until the ANN reaches a satisfactory level of accuracy.

There are different types of ANNs, including feedforward, recurrent, and convolutional neural networks, each with its own strengths and weaknesses. ANNs have the ability to model complex relationships and patterns in the data, making them well-suited for stock market prediction tasks.

Gurjar et. al.[16], in their paper uses Artificial Neural Networks to predict stock prices based on historical data and features such as stochastic indicator, moving averages, RSI. ANN model trained with training and testing sets to evaluate accuracy. Predicted stock prices aid in smart investment decisions and market trend analysis. Personalized user profiles maintain privacy and allow favorite stock selection. Admin can add more stocks beyond the top 50 NSE stocks. Predicted prices given for next 1, 3, and 5 days with graphical display of results for easy understanding.

D. Time-Series Analysis

Time-series analysis employs statistical techniques to examine and predict data accumulated over a period. In the context of stock market prediction, time-series analysis can be used to predict future stock prices based on historical stock data.

Time-series analysis techniques include trend analysis, seasonal analysis, and cycle analysis. These techniques aim to identify patterns and relationships in the data, such as trends, seasonality, and cycles, and to use this information to make predictions about future values.

Some common time-series analysis techniques for stock market prediction include:

1) *Moving Average:* A moving average is a simple method that smooths out fluctuations in the data by calculating the average of a set of past values.

2) *ARIMA (Auto Regressive Integrated Moving Average):* ARIMA is a statistical model that combines moving average and regression techniques to model and forecast time-series data[17].
3) *SARIMA (Seasonal ARIMA):* SARIMA is a variant of ARIMA that incorporates the effects of seasonality in the data.
4) *GARCH (Generalized Autoregressive Conditional Heteroscedasticity):* GARCH is a statistical model that is used to model and forecast time-series data with changing volatility.

It is important to note that time-series analysis is based on the assumption that patterns and relationships in the past will continue into the future, which may not always be the case. Additionally, stock market prediction is a complex and uncertain task, and no single model or technique can provide a guarantee of accurate predictions.

Stock market prediction using ARIMA involves using the ARIMA model to forecast future values of stock prices based on their historical data[17]. The process typically involves the following steps:

a) *Data Collection:* Gather historical data for the stock prices of interest.
b) *Data Pre-Processing:* Clean and pre-process the data to remove any missing values or outliers.
c) *Time Series Decomposition:* Decompose the time series data into its trend, seasonality, and residual components to better understand the underlying patterns.
d) *Stationarity Check:* Check if the time series is stationary or non-stationary. If it is non-stationary, take differences or perform other transformations to make it stationary.
e) *Model Selection:* Select the appropriate ARIMA model by determining the order of differencing (d), the autoregression order (p), and the moving average order (q) based on the ACF and PACF plots of the residuals.
f) *Model Fitting:* Fit the ARIMA model to the time series data.
g) *Model Evaluation:* Evaluate the performance of the ARIMA model using metrics such as mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE).
h) *Forecasting:* Use the ARIMA model to make predictions for future stock prices.

E. Long Short-Term Memory (LSTM)

LSTM (Long Short-Term Memory) is a type of Recurrent Neural Network (RNN) that is well-suited for processing sequential data, such as stock prices over time. LSTMs are designed to handle the problem of vanishing gradients in traditional RNNs, which can make it difficult to train the network and maintain long-term dependencies in the data.

In the context of stock market prediction, LSTMs can be used to predict future stock prices based on historical data, such as prices, volume, and news events. The network is trained using large amounts of data, adjusting the weights and biases of the neurons to minimize the prediction error. The LSTM network is able to capture long-term dependencies in the data by using a memory cell and gates that control the flow of information through the network.

The prediction is made by processing the input data step-by-step, updating the hidden state of the network at each time step to capture the dependencies between the time steps. The final prediction is then made based on the hidden state at the last time step.

Kambli et. al.[18], used LSTM and RNN methods were used on a set of companies from NSA data. The model was trained with historical and real-time data, with a focus on collecting accurate data. Intraday trading data was captured every minute for real-time training. BPTT (Backpropagation) was used to train RNN in an unrolled manner to avoid far backpropagation and simplify training.

LSTM performs stock price prediction using its memory cell state and gates. The cell state stores relevant past information and the gates (forget, input, and tanh) determine which information is eliminated or stored.

Sarode et. al.[19], model has 2 layers: input (cells=sequence) & compact LSTM, with output layer having similar of cells. It uses 4 types of LSTM learning features.

1) Historical trade details
2) Technical Analysis from historical trade details
3) Market indices movement
4) Economic fundamentals.

Parmar et. al.[20], found LSTM to be more efficient and accurate than a regression-based model in their study. The results from LSTM were promising and precise, leading to a conclusion that stock price prediction can be made more accurately and efficiently using LSTM. It's widely used for analysis with recent and current data as LSTM can remember long-term information. LSTM units and blocks overcome the vanishing gradient problem faced by general RNN by replacing traditional neurons.

III. FUTURE WORK

Future work in the field of stock market prediction using machine learning can include:

1) Improving the accuracy of predictions through the use of more advanced machine learning algorithms and techniques.
2) Integrating alternative data sources, such as news articles, social media data, and financial reports, to improve the accuracy of predictions.
3) Exploring the use of reinforcement learning techniques for stock market prediction, where the algorithm can learn from its own predictions and adjust its strategy over time.
4) Combining different machine learning models to form an ensemble model, which can improve prediction accuracy by combining the strengths of different algorithms.
5) Developing new evaluation metrics for stock market prediction models, taking into account both accuracy and risk.
6) Incorporating the effects of market sentiment and investor behaviour into predictions, as these factors can have a significant impact on stock prices.
7) Applying the methods to other financial markets, such as the forex market, to validate their effectiveness in other domains.

IV. CONCLUSIONS

The literature on stock market prediction using machine learning suggests that it is a challenging task and that even the most advanced models can only produce predictions with limited accuracy. The stock market is influenced by a variety of unpredictable factors such as news events, economic shifts, and market sentiment, making it difficult to produce reliable predictions.

Multiple machine learning algorithms have been applied to stock market prediction, including artificial neural networks, decision trees, support vector machines, and others. Among these, recurrent neural networks, specifically LSTM (Long Short-Term Memory) networks, have shown promising results in predicting stock market trends.

However, it is important to note that even LSTM models are limited in their ability to predict stock market trends with high accuracy and that other forms of analysis should be used in conjunction with machine learning predictions. Additionally, it is important to carefully pre-process and normalize the data before using it to train machine learning models to improve their predictive performance.

In conclusion, while machine learning can be a useful tool in stock market prediction, it should not be relied upon as the sole source of information and should be used in combination with other forms of analysis to make informed investment decisions.

REFERENCES

[1] I. Parmar et al., "Stock Market Prediction Using Machine Learning," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 574-576, DOI: 10.1109/ICSCCC.2018.8703332.
[2] B. Jeevan, E. Naresh, B. P. V. kumar and P. Kambli, "Share Price Prediction using Machine Learning Technique," 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C), Bangalore, India, 2018, pp. 1-4, DOI: 10.1109/CIMCA.2018.8739647.
[3] Sourav Das, Dipankar Das, Anup Kumar Kolya, "Sentiment classification with GST tweet data on LSTM based on polarity-popularity model", Indian Academy of Sciences, 2020, DOI:10.1007/s12046-020-01372-8.
[4] Lirong Hu, Shenjing He, Zixuan Han, He Xiao, Shiliang Su, Min Weng, Zhongliang Cai, “Monitoring housing rental prices based on social media:An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies, Land Use Policy”, Volume 82, 2019, Pages 657-673, ISSN 0264-8377, DOI:10.1016/j.landusepol.2018.12.030.
[5] Shah, D., Patel, S. & Bharti, S.K., “Heart Disease Prediction using Machine Learning Techniques”,. SN COMPUT. SCI. 1, 345 (2020). DOI:10.1007/s42979-020-00365-y
[6] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou and D. Zhao, "Flight Delay Prediction Based on Aviation Big Data and Machine Learning," in IEEE Transactions on Vehicular Technology, vol. 69, no. 1, pp. 140-150, Jan. 2020, DOI: 10.1109/TVT.2019.2954094.
[7] Om Nath, Arup Kadia, Sayantan Ghanta, Bipasha Bhattacharjee, Shreyashree Das, "Fish Recognition and Classification Based on Feature Vector Analysis", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 6 Issue III, March 2018, ISSN: 2321-9653, pp:2306-2309, DOI:10.22214/ijraset.2018.3532.
[8] R. A. Kamble, "Short and long term stock trend prediction using decision tree," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2017, pp. 1371-1375, DOI: 10.1109/ICCONS.2017.8250694.
[9] Strader, Troy J.; Rozycki, John J.; ROOT, THOMAS H.; and Huang, Yu-Hsiang (John) (2020) "Machine Learning Stock Market Prediction Studies: Review and Research Directions," Journal of International Technology and Information Management: Vol. 28: Iss. 4, Article 3. DOI:10.58729/1941-6679.1435
[10] P. Werawithayaset and S. Tritilanunt, "Stock Closing Price Prediction Using Machine Learning," 2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE), Bangkok, Thailand, 2019, pp. 1-8, DOI: 10.1109/ICTKE47035.2019.8966836.
[11] Y. Hua, Z. Zhao, R. Li, X. Chen, Z. Liu and H. Zhang, "Deep Learning with Long Short-Term Memory for Time Series Prediction," in IEEE Communications Magazine, vol. 57, no. 6, pp. 114-119, June 2019, DOI: 10.1109/MCOM.2019.1800155.
[12] Hum Nath Bhandari, Binod Rimal, Nawa Raj Pokhrel, Ramchandra Rimal, Keshab R. Dahal, Rajendra K.C. Khatri, “Predicting stock market index using LSTM, Machine Learning with Applications”, Volume 9, 2022, 100320, ISSN 2666-8270, DOI:10.1016/j.mlwa.2022.100320.

©IJRASET: All Rights are Reserved

[13] K. M. Hindrayani, T. M. Fahrudin, R. Prismahardi Aji and E. M. Safitri, "Indonesian Stock Price Prediction including Covid19 Era Using Decision Tree Regression," 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 2020, pp. 344-347, DOI: 10.1109/ISRITI51436.2020.9315484.

[14] Vaishnavi Gururaj, Shriya V, and Dr. Ashwini K, "Stock Market Prediction using Linear Regression and Support Vector Machines", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 8 (2019) pp. 1931-1934.

[15] Nti, Isaac Kofi, Adekoya, Adebayo Felix and Weyori, Benjamin Asubam. "Efficient Stock-Market Prediction Using Ensemble Support Vector Machine" Open Computer Science, vol. 10, no. 1, 2020, pp. 153-163. DOI:10.1515/comp-2020-0199

[16] Mruga Gurjar, Parth Naik, Gururaj Mujumdar, Prof. Tejaswita Vaidya, Mruga Gurjar, Parth Naik, Gururaj Mujumdar3, Prof. Tejaswita Vaidya, "Stock market prediction using ANN", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 03 | Mar-2018.

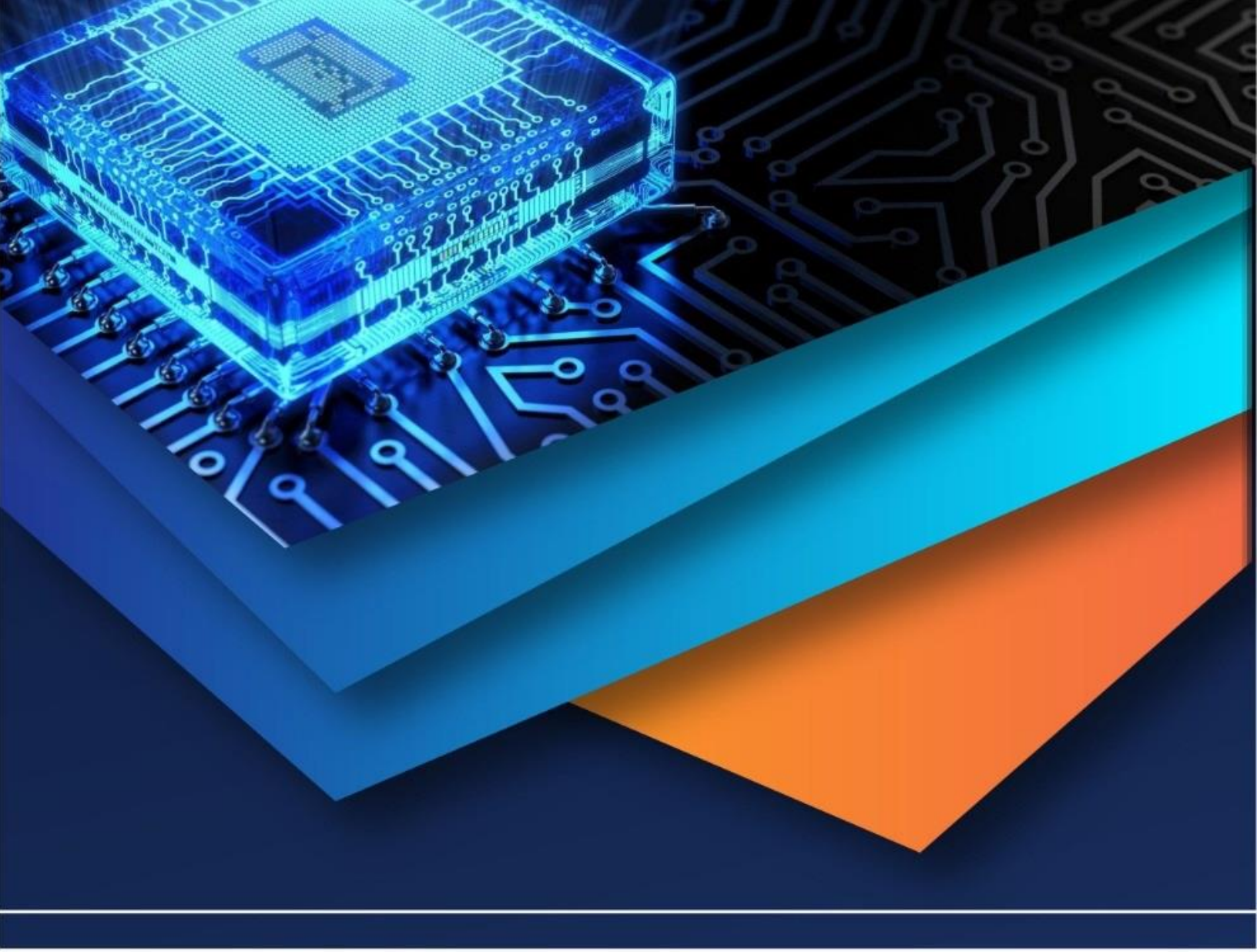
[17] Yash Mehta, Atharva Malhar, Dr. Radha Shankarmani, "Stock Price Prediction using Machine Learning and Sentiment Analysis", IEEE, 2021, DOI: 10.1109/INCET51464.2021.9456376.

[18] Jeevan B, Naresh E, Vijaya kumar B P, Prashanth Kambli, "Share Price Prediction using Machine Learning Technique", IEEE, 2018, DOI: 10.1109/CIMCA.2018.8739647.

[19] Sumeet Sarode, Harsha G. Tolani, Prateek Kak, Lifna C S, "Stock Price Predicition using Machine Learning Techniques", IEEE, pp.177-181, 2019, DOI: 10.1109/ISS1.2019.8907958.

[20] Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta, Himanshu Dhiman, Lokesh Chouhan, “Stock Market Prediction Using Machine Learning", IEEE, pp.574-576, 2018, DOI: 10.1109/ICSCCC.2018.8703332

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)