Stock Market Prediction Using Machine Learning Techniques

Jagruti Hota [1], Sujata Chakravarty [2], Bijay K. Paikaray [3] and Harshvardhan Bhoyar [4]

[12]*Dept. of CSE, Centurion University of Technology and Management, Odisha, India.*
[3]*School of Information & Communication Technology, Medhavi Skills University, Sikkim, India*
[4]*Faculty. of Management Studies, Sri Sri University, Odisha, India*

Abstract

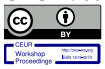
The stock market is a very important activity in the finance business. Its demand is consistently growing. Stock market prediction is the process of determining the future value of company stock or other financial instruments traded on a financial exchange. For some decades Artificial Neural Network (ANN), which is one intelligent data mining technique has been used for Stock Price Prediction. It has been trusted as the most accurate consideration. This paper surveys different machine learning models for stock price prediction. We have trained the available stock data of American Airlines for this project. The programming language that we have used in this paper is Python. The Machine Learning (ML) models used in this project are Decision Tree (DT), Support Vector Regression (SVR), Random Forest (RF), and ANN. The data here is split into 70% for training and 30% for testing. The dataset contains stock data for the last 5 years. From the simulation results, it is shown that Random Forest performs better as compared to others. Thus, it can be used in the real-time implementation.

Keywords

Machine Learning, Stock Price, Prediction, American Airlines, Support Vector Machine (SVR), Artificial Neural Network (ANN), Random Forest (RF), Decision Tree (DT).

1. Introduction

The Stock Market is the accumulation of stockbrokers, traders, and investors who sell buy or share trades. There are so many companies that provide their stock list on market, these make their stocks attractive to investors [1]. Because ever since the 16s investors are trying different techniques to get knowledge about different companies to improve their investment returns [2]. It plays a very important role in increasing a developing country's economic status like India [3]. The demand for Stock Market is growing significantly. We all know that it has been in focus for many years because of the outstanding profits [4]. Lots of wealth are traded daily through the stock market and so it is seen as one of the most profitable financial outlets [5]. Now, the stock market is one of the factors which shows a country's economy [6]. Many people invest a handsome amount of money in the share market but sometimes they tend to incur very huge losses because they depend upon the stockbrokers, who advise investors based on fundamental, technical, and time series [7]. Investors have been trying to find an intelligent idea to overcome such problems. This is where Stock Price Prediction comes into action because predicting stock prices is very necessary [2].

ACI'22: Workshop on Advances in Computation Intelligence, its Concepts & Applications at ISIC 2022, May 17-19, Savannah, United States
EMAIL: 90301120100@cutm.ac.in (A. 1); chakravartys69@gmail.com (A.2); bijaypaikaray87@gmail.com (A. 3); harshvardhan.b@srisriuniversity.edu.in (A. 4)
ORCID: 0000-0001-5843-0335 (A. 3)


Stock Price Prediction's main idea is to accurately predict the future financial outcome [5]. In the past few years, Machine Learning algorithms are seen to give promising results in various industries, so many traders are applying these techniques to their respective fields [8].

ML can be applied as a game-changer [9]. In this paper, some experimentation is done by taking different ML algorithms to predict the opening price of American Airlines stocks. The Machine learning (ML) algorithms that we have used are Random Forest (RF), Decision Tree (DT), Support Vector Regressor (SVR), and Artificial Neural Network [6]. Prediction of Stocks is based on the opening price of the day for this paper.

The remaining paper has been laid out in the following order. In section -2 literature survey has been reviewed followed by section -3 where various approaches or different machine learning algorithms used have been discussed. In section-4 the problems that occurred or that needed to be improved previously have been addressed. Section-5 represents all the information about the dataset. In section-6 the results and future works have been discussed and in section-7 the paper has been concluded.

2. Literature Survey

Since the introduction of the Stock Market so many predictors are constantly trying to predict stock values using different Machine Learning algorithms such as Support Vector Regressor (SVR), Linear Regression (LR), Support Vector Machine (SVM), Neural Networks Genetic Algorithms, and many more [5] on stocks of various companies.

There is a diversity in many papers based on different parameters. Many different ML algorithms are used by different authors based on different parameters. Some authors believe that Neural Networks have given better performance as compared to other approaches [5]. Like, in paper [12] Hiransha M and GopalKrishnan E. A has trained four models Multi-Layer Perceptron (MLP), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) and it was observed that CNN has performed better than the other three networks. On the other hand, many authors believe that Support Vector Regression which is known to solve regression and prediction problems gives better performance as seen in paper [13] by Haiqin Yang, Laiwan Chan, and Irwin King. In paper [5] Paul d. Yoo has trained 3 models Support Vector Machine, Case-Based Reasoning classifier (CBR), and Neural Networks (NN) from which Neural has given the most appropriate prediction. Sumeet et al [18] has done an approach where they have combined two distinct fields for stock exchange analysis. It merges price prediction based on real time data as well as historical data with news analysis. In this paper LSTM(Long Short-Term Memory) is used for prediction. The datasets are collected from large sets of business news in which relevant and live data information is present. Then the results of both analyses are combined to form a response which helps visualize recommendation for future increases.

So, in many papers, it has been seen that neural networks give the expected prediction value.

3. Approaches

In this project, prediction is carried out by using these ML algorithms. These are Decision Tree, Support Vector Regression, Random Forest, and Artificial Neural Network.

3.1. Decision Tree Methodology

It is a supervised ML, which is used for both regressions as well as classification. That is how it is also called CART Classification and Regression Trees. In this algorithm, two nodes are present namely Decision Node which is for making the decisions and can be divided into multiple branches and Leaf Node which gives the output of decisions and this node can't be further divided into many nodes. The following is the formula for Leaf Node:

$$Information\ Gain\ =\ Class\ Entropy - Entropy\ Attribute \quad (1)$$

Branches-Here decision rules are set by which nodes can be divided further.

For Prediction, it starts from the root node, compares values of the real attribute with the root attribute, and based on that comparison it follows the branch and jumps to the next node. This process continues until it reaches the leaf node of the tree.

Entropy-It is a metric that helps in measuring error in a given attribute. The formula to find entropy is: -

$$\text{Entropy(s)= -P(yes)log2 P(yes)- P(no) log2 P(no)} \quad (2)$$

Here, (S) implies the Total number of samples. P (yes) refers to the Probability of S and P (no) means the Probability of no.

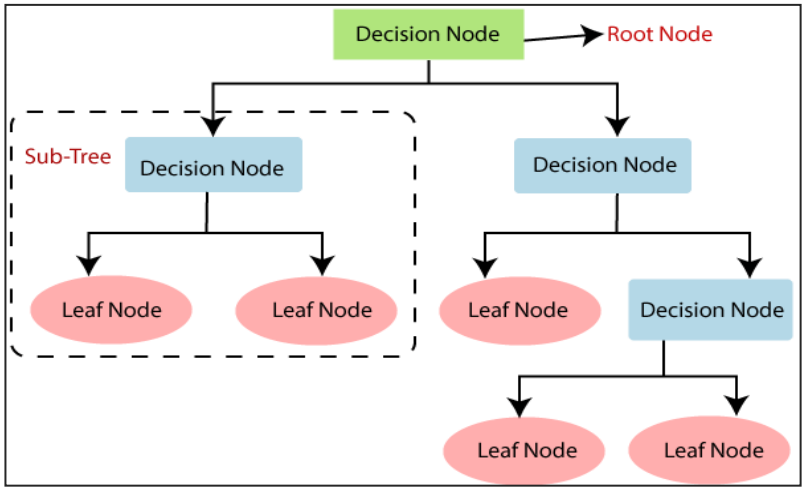


Figure 1: Decision Tree Classifier Process

3.2 Support Vector Regression Methodology

It is a Supervised Machine learning algorithm used for regression analysis. It finds the function that helps us approximate mapping based on the training sample from an input domain to real numbers. The Terminologies contained in this are Hyperplane -this is the line that is used to predict the continuous output. Kernel helps to find hyperplanes in higher dimensional space without increasing the computational cost of it and the decision boundary is a simplification line that differentiates positive examples and negative examples.

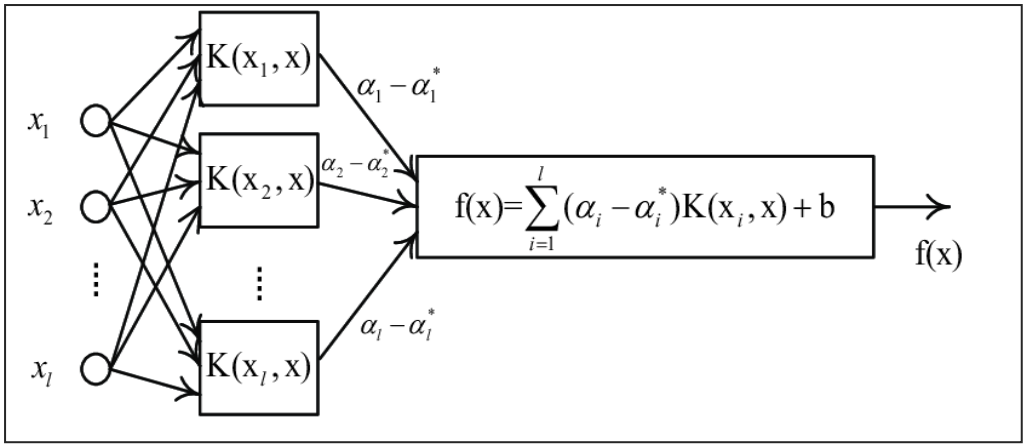


Figure 2. Support Vector Regression

3.3 Random Forest Methodology

Random forest is a supervised Machine Learning algorithm that is used for Regression analysis. This overcame the problem of overfitting as seen in the decision Tree [12]. It is an ensemble learning method. The steps for prediction are first a random k data point is picked from the training set then accordingly the decision tree is built. Then choose the number of trees we want to build and again follow the previous steps. From every new data point, make N tree Trees predict the value of Y for data points and assign new data points across all of y predicted Y values.

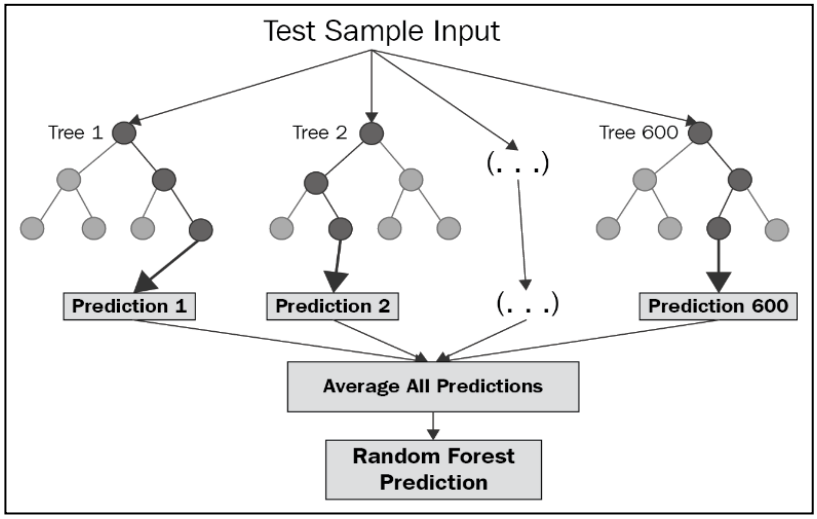


Figure 3. Random Forest Procedure

3.4 Artificial Neural Network Methodology

An artificial Neural network is an interconnection of nodes that is like the biological neuron in our body but not similar. For the last few decades, ANN has been used for Stock Price Prediction [12]. It contains three layers, first is the Input Layer – this layer takes different inputs variable from the user then, the hidden layer-This layer is present between

the input layer which identifies all hidden features and patterns and the last layer is the Output layer- This layer provides the final output. ANN takes different inputs and multiplies them with the specified weights for each with an activation function for the activation of neurons.

The formula of the transfer function is:

$$\sum_{i=1}^{n} Wi * Xi + b \qquad (3)$$

Here, b is the threshold value. Xi is input and value and Wi is the weight.

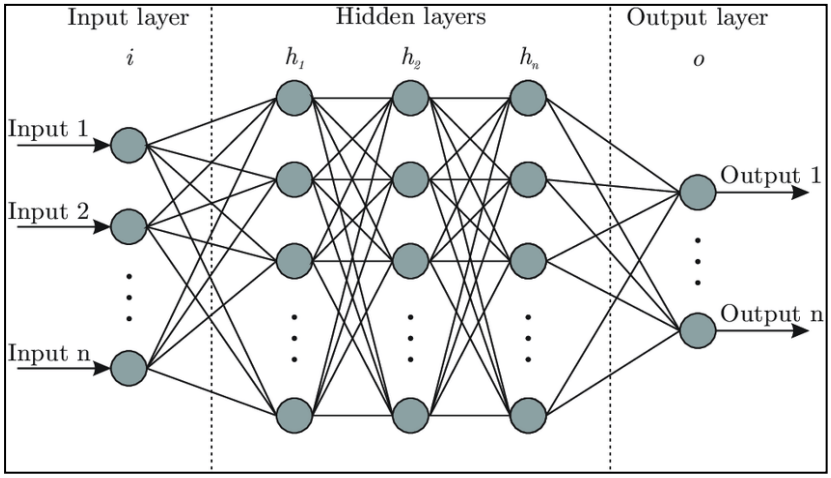


Figure 4. Artificial Neural Network Procedure

4. Problem Statement

Now, stockbrokers who execute trading mainly depend on their experience, price trends, or fundamental analysis i.e. - buy or hold to select stocks. These methods may lead to great losses to investors if they make any wrong decisions because these are personalized and short-sighted due to their limited capacity. Lack of prominent results may lead to reluctance to participate in trading by investors. So, to overcome these drawbacks it is important to have a tool that can guide us on proper trading methods and consequences. Technical and fundamental analysis are the basis of future stock market Prediction. Here, Machine Learning methods come into action. These methods can help us analyze stock prices over time and create ideas about them and then help us in prediction and can be used to model a tool.

5. Stock Market Prediction Architecture

Stock market data of American airlines from 2-08-2013 to 2-07-2018 has been used as a dataset in this project. This dataset has 1258 rows and 7 columns. Each row represents the information for a single day. For columns, the following are the feature description.

5.1 Data Preprocessing

It includes searching for essential missing or null values and replacing them with mean values Searched for categorical value and if there is any unnecessary data then those values are dropped.

5.2 Data Splitting

The processed data has been divided into 70% training data and 30% testing data using the train_test_split method. Here 881 data is taken as training data and the rest 377 is kept

for testing. The training data values are taken from the date 2013-02-08 to 2016-08-09 and the testing data are from 2016-08-10 to 2018-02-06.

Table 1:
Dataset Feature Description Table

Sl. No	Feature	Description
1.	Date	It shows the date in the format: yy-mm-dd.
2.	Open	It shows the price of the stock at market opening.
3.	High	It shows the highest price reached on that day.
4.	Low	It shows the lowest price reached on that day.
5.	Close	It shows the lowest price reached on that day.
6.	Volume	It shows the number of shares traded on that day.
7.	Name	This is the name of the stock's ticker.

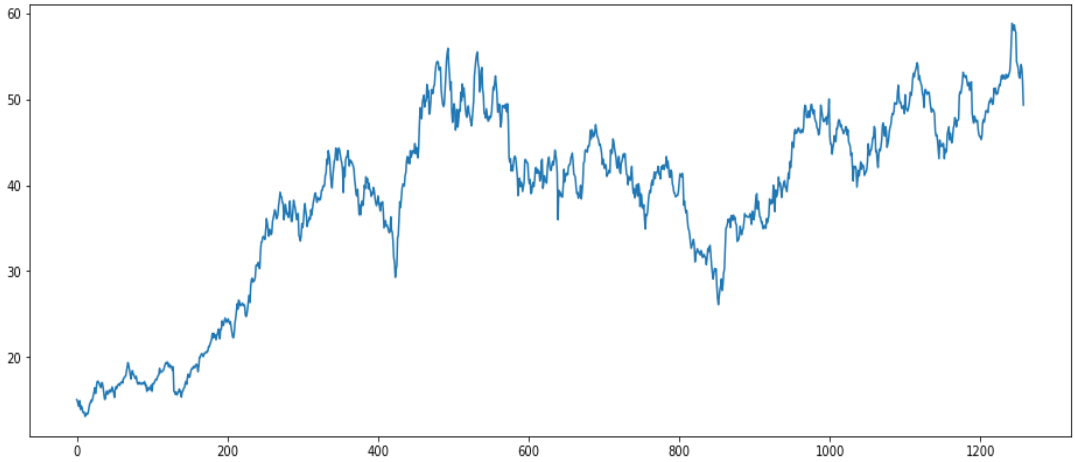


Figure 5: Opening Price Graph

5.3 Data Scaling

Standardization and Normalization are done on the data using Minmax Scaler and Standard Scaler to limit the ranges of variables to make them comparable on common grounds using ML methods.

5.4 Feature Selection

The selection of features is a very important task to predict future values. If we consider the worst features then the prediction can go wrong. In this paper, the attribute or feature used for feature extraction is the opening price or the 'open' column of American Airlines stocks. A data structure has been created with 7 timesteps and 1 output.

5.5 Prediction

We have adapted Machine Learning Approaches to find the prediction. In this case, training the model is very necessary. Random Forest, Decision Tree, and Support Vector Regression models have been used to do the prediction work.

5.6 Error Calculation

There are 4 types of error calculations present for evaluation.

In this paper, we have used the MAPE method to find the error. Performance evaluation is done using MAPE values of all the models. Following are the formulae to find the MAPE

(Mean Absolute Percentage Error), MAE (Mean Absolute Error), rRMSE (Root Mean Squared Error), and MSE (Mean Squared Error) value

$$MAPE = \frac{1}{n\sum_{i=1}^{n}\left(\frac{|A_i - P_i|}{|A_i|}\right)} \times 100 \quad (4)$$

(

$$minus - endMAE = \frac{1}{n\sum_{i=1}^{n}\left(\frac{|A_i - P_i|}{|A_i|}\right)} \quad (7) \quad (5)$$

$$rRMSE = sqrt\left(\frac{1}{n\sum_{i=1}^{n}\left(\frac{A_i - P_i}{A_i}\right)^2}\right) \quad (6)$$

$$MSE = 1/n\sum_{i=1}^{n}\left(\frac{A_i - P_i}{A_i}\right)^2) \quad (7)$$

Here, n is the sample size, Ai is the predicted value and Pi is the Predicted value.

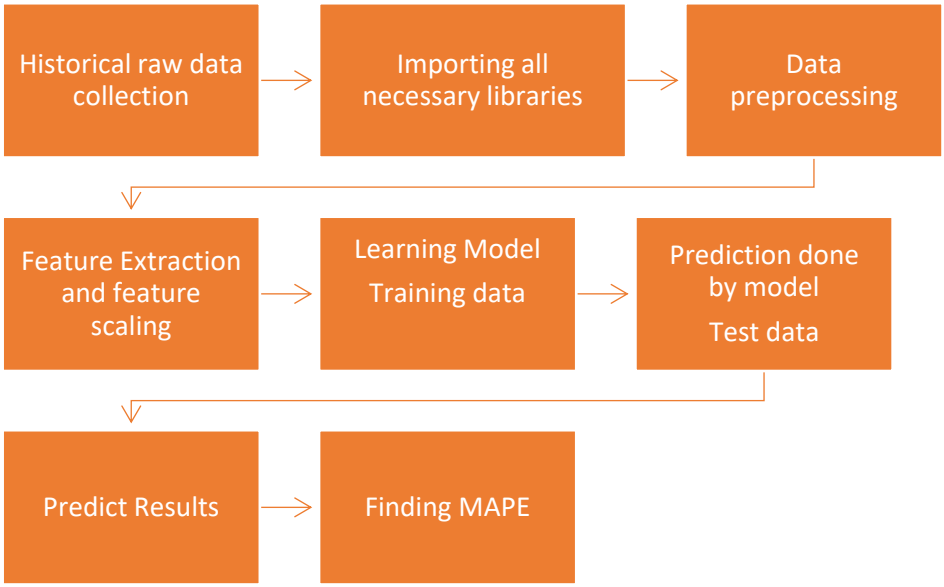


Figure 6: Architecture of Methodology

As shown in the figure above all the historical data were collected first and followed by the importation of all necessary libraries such as NumPy, Pandas, matplotlib, Seaborn, mean_squared_error, etc. In the next step, various data processing methods have been performed such as drop, isnull, etc. Then feature extraction and feature scaling techniques have been implemented using Min Max Scaler and sc.fit_transform. In the next step we have trained the data and learned the model required. In the next step various machine learning model which we have learned have been applied such as Decision Tree, Support Vector, Artificial Neural Networks, and Random Forest. Then we have got the prediction results. Out of all the 4 algorithms, Random Forest has the lowest MAPE value i.e.- 0.36

6. Results and Discussion

The main objective of this project is to examine several different prediction techniques to predict future stock prices based on past returns. And here it is visible that Random Forest

is the best algorithm for this research giving a MAPE value of 0.36. This algorithm shall be used to predict opening prices shortly. The following is the table to show the MAPE values using Machine Learning Algorithms.

Table 2

MAPE Chart

S.No	Model	MAPE
01	Decision Tree	1.60
02	Support Vector Regression	3.56
03	Random Forest	0.36
04	Artificial Neural Network	0.37

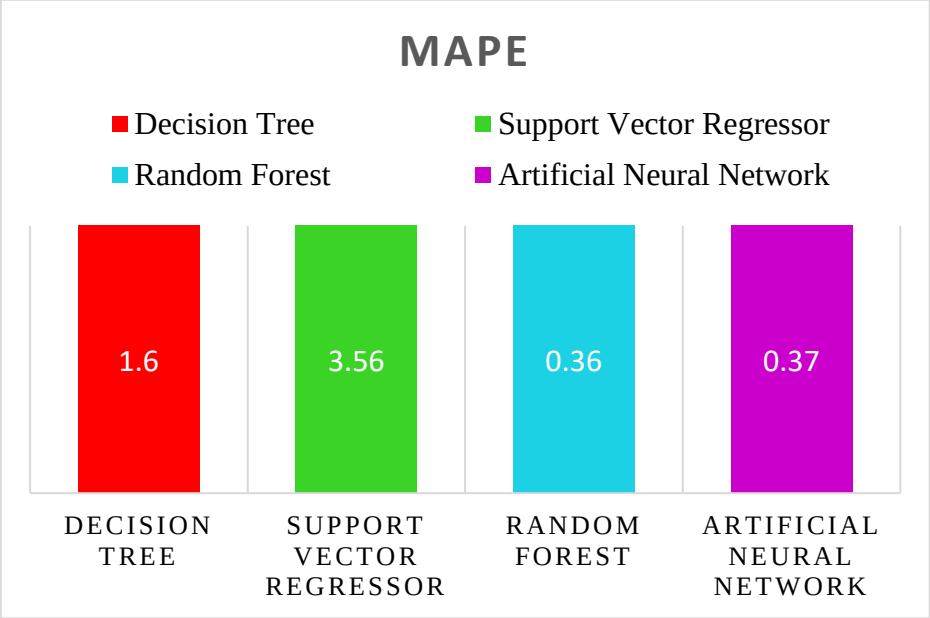


Figure 7 MAPE Comparison

7. Conclusion

The project was majorly aimed at creating an efficient tool that will help stockbrokers and investors properly invest in the stock market. Five years American Airlines stocks have been preprocessed and four machine learning algorithms have been used – Random Forest, Support Vector Regressor, Decision Tree, and Artificial Neural Network on this project. Based on calculations, estimations, and observations, we conclude that Random Forest has the lowest Mean Absolute Percentage Error (MAPE) value of 0.36 followed by Artificial Neural Networks with the value of 0.37, then Decision Tree showing MAPE value of 1.6 and the highest in SVR showing a value of 3.5. Artificial Neural Network has been used in this project, giving a MAPE value of 0.37 which is the second least MAPE value provided. So, in the future, it is intended to work on advanced ANN evolutionary techniques like Genetic Algorithm to decrease the MAPE values for better implementations.

8. References

1. Bhattacharjee, Indronil, and Pryonti Bhattacharja. "Stock Price Prediction: A Comparative Study between Traditional Statistical Approach and Machine Learning Approach." 2019 4th International Conference on Electrical Information and Communication Technology (EICT). IEEE, 2019.
2. Mehta, Yash, Atharva Malhar, and Radha Shankarmani. "Stock Price Prediction using Machine Learning and Sentiment Analysis." 2021 2nd International Conference for Emerging Technology (INCET). IEEE, 2021.
3. Sharma, Ashish, Dinesh Bhuriya, and Upendra Singh. "Survey of stock market prediction using machine learning approach." 2017 international conference of electronics, communication and aerospace technology (ICECA). Vol. 2. IEEE, 2017.
4. Hegazy, Osman, Omar S. Soliman, and Mustafa Abdul Salam. "A machine learning model for stock market prediction." arXiv preprint arXiv:1402.7351 (2014).
5. Yoo, Paul D., Maria H. Kim, and Tony Jan. "Machine learning techniques and use of event information for stock market prediction: A survey and evaluation." International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06). Vol. 2. IEEE, 2005.
6. S. Chakravarty, B. K. Paikaray, R. Mishra and S. Dash, "Hyperspectral Image Classification using Spectral Angle Mapper," 2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), 2021, pp. 87-90, doi: 10.1109/WIECON-ECE54711.2021.9829585.
7. Wanjawa, Barack Wamkaya, and Lawrence Muchemi. "ANN model to predict stock prices at stock exchange markets." arXiv preprint arXiv:1502.06434 (2014).
8. Reddy, V. Kranthi Sai. "Stock market prediction using machine learning." International Research Journal of Engineering and Technology (IRJET) 5.10 (2018): 1033-1035.
9. Ravikumar, Srinath, and Prasad Saraf. "Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms." 2020 International Conference for Emerging Technology (INCET). IEEE, 2020.
10. Pathak, Ashish, and Nisha P. Shetty. "Indian stock market prediction using machine learning and sentiment analysis." Computational Intelligence in Data Mining. Springer, Singapore, 2019. 595-603.
11. Deepak, Raut Sushrut, Shinde Isha Uday, and D. Malathi. "Machine learning approach in stock market prediction." International Journal of Pure and Applied Mathematics 115.8 (2017): 71-77.
12. Hiransha, M., et al. "NSE stock market prediction using deep-learning models." Procedia computer science 132 (2018): 1351-1362.
13. Yang, Haiqin, Laiwan Chan, and Irwin King. "Support vector machine regression for volatile stock market prediction." International Conference on Intelligent Data Engineering and Automated Learning. Springer, Berlin, Heidelberg, 2002.
14. Kohli, Pahul Preet Singh, et al. "Stock prediction using machine learning algorithms." Applications of Artificial Intelligence Techniques in Engineering. Springer, Singapore, 2019. 405-414.
15. Moedjahedy, Jimmy H., et al. "Stock Price Forecasting on Telecommunication Sector Companies in Indonesia Stock Exchange Using Machine Learning Algorithms." 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS). IEEE, 2020.
16. Mohanty, Sachi Nandan, et al., eds. Recommender System with Machine Learning and Artificial Intelligence: Practical Tools and Applications in Medical, Agricultural, and Other Industries. John Wiley & Sons, 2020.
17. Jain, Sarika, et al. "Human Disease Diagnosis Using Machine Learning." Intelligent Data Communication Technologies and Internet of Things. Springer, Singapore, 2021. 689-696.
18. Sarode, Sumeet, et al. "Stock price prediction using machine learning techniques." 2019 International Conference on Intelligent Sustainable Systems (ICISS). IEEE, 2019.