A Comparative Research on Machine Learning Algorithms for Stock Market Prediction and Recommendation

[1]SUMERA JABEEN, [2]MOHUMMAD ABDUL WAHED, [3]SHAIK SAIDABI, [4]NADIGOTTU SRAVANI

[1]*Assistant.Professor, Dept of CSE, CMR Engineering College, Hyderabad*
[2]*Assistant professor, Dept of CSE, Malla Reddy Institute of Engineering and Technology, Hyderabad*
[3,4]*Assistant.Professor, Dept of IT, CMR Engineering College, Hyderabad*

Abstract-The financial industry relies heavily on the stock market. The demand for it is always increasing. Securities exchange expectation is the most common way of deciding the future worth of organization stock or other monetary instruments exchanged on a monetary trade. For certain many years Counterfeit Brain Organization (ANN), which is one shrewd information mining strategy has been utilized for Stock Value Expectation. It has been held in high regard as the most accurate factor. The various machine learning models for stock price prediction are examined in this paper. For this project, we trained the American Airlines stock data that was already available. Python is the programming language that we have utilized in this paper. The Support Vector Machine (SVM), Random Forest (RF), and ANN are the Machine Learning models utilized in this project. 70% of the data are used for training, while 30% are used for testing. Stock data for the past three years can be found in the dataset. The simulation results demonstrate that Random Forest outperforms other methods. As a result, it can be implemented in real time.

Keywords-*Random Forest, Artificial Neural Network, SVM, Stock market prediction*

I. INTRODUCTION

In the past few decades, many advances have been made in the field of data analytics. Researchers are now able to predict stock prices with higher accuracy due to analytical predictive models. These predictive techniques utilize data from previous stock price movements and look for patterns that could indicate future stock price changes in the market. The use of these machine learning techniques will allow investors to make better decisions and invest more wisely by maximizing their returns and minimizing their losses. In this blog post, you will learn about some of the popular machine learning techniques in relation to making stock price movement (direction of stock price) predictions and classify whether a stock is a buy, sell, or hold. The stock price prediction problem is a fairly complex problem and different techniques can be used appropriately to achieve good prediction accuracy.

Machine learning techniques used for predicting stock prices involve analyzing historical data to predict the likelihood of a future event occurring or forecast future performance. This is done by looking at patterns in the data which includes current and past information and finding the best fit predictive models. Machine learning model with optimal performance can be trained by tuning with different algorithms and associated parameters / hyper parameters. However,

predictions made with machine learning models are not usually as reliable as those made by humans. However, the predictive models can be used as augmented intelligence for investors to be take informed decision on investing in stocks that will provide them with higher returns and minimize losses. In recent years, the advent of machine learning and deep learning has provided a greater level of model performance accuracy for stock prices prediction. Financial analysts and investors have used predictive analytics techniques to improve their ability to predict the price of stocks in the market with relative accuracy.

One approach you could take is utilizing data from past transactions in your market or combining it with other investment strategies like technical analysis, fundamental analysis, and quantitative easing. The use of a model based on historical data provides more solid information because it has been previously tested for accuracy in the market; this eliminates incorrect or incomplete data which provides unreliable predictions about future prices movements. Machine learning techniques are also useful at analyzing how different stocks move together over time–for example, if there's an upswing due to changing sentiments then investors would want to know what caused the positive change so they can invest with more confidence.

In fundamental analysis (FA), the machine learning models can be trained using data related to companies' financial statements and macroeconomic and microeconomic factors. The models can be used to predict the stock price movement at any given point in time. One can use supervised learning models such as bagging, boosting ensemble classifiers, or deep learning neural networks for making predictions on whether to buy, sell or hold. Feature engineering is key to building a high-performance model and one can go about using feature selection (random forest)/feature extraction (PCA) techniques. The accuracy of these models depends on two factors – the input data used and the type of algorithm selected. For most cases, it is recommended that they be trained for 4-5 years with large quantities of stock price data before results can be expected.

While using machine learning models for stock price prediction, caution must be taken as the model performance accuracy can vary depending on the amount of data available for model training. For example, there are cases where models could show almost 95% accuracy in one year and then display around 60% accuracy in the next year. This is largely due to the fact that prices in the market are not always clear-cut and there can be changes in certain variables between one period and another, like volume of transactions or number of employees, which will affect price movements. When making decisions regarding using predictive models for stock price prediction, it is recommended to backtest these predictive models and evaluate the quality of the prediction output. This will increase investors' confidence in the model, giving it a more solid foundation for future uses.

II. RESEARCH SURVEY

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]

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 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, CaseBased Reasoning classifier (CBR), and Neural Networks (NN) from which Neural has given the most appropriate prediction. Sumeet et al 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.

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.

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

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].

III. COMPARATIVE ANALYSIS

Treating stock data as time-series, one can use past stock prices (and other parameters) to predict the stock prices for the next day or week. Machine learning models such as Recurrent Neural Networks (RNNs) or LSTMs are popular models applied to predicting time series data such as weather forecasting, election results, house prices, and, of course, stock prices. The idea is to weigh out the importance of recent and older data and determine which parameters affect the "current" or "next" day prices the most. The machine learning model assigns weights to each market feature and determines how much history the model should look at to predict future stock prices.

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.

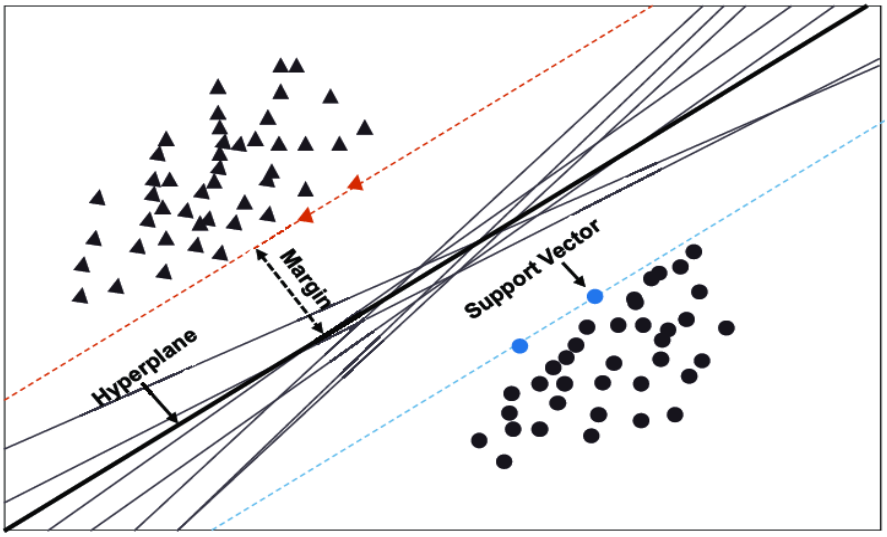

Fig.1 Support Vector Machine Model

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.

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.

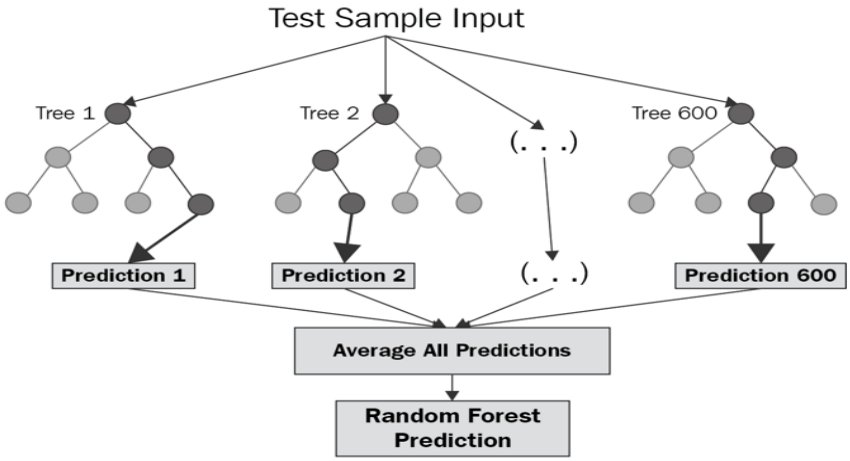

Fig.2 Random Forest model

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].

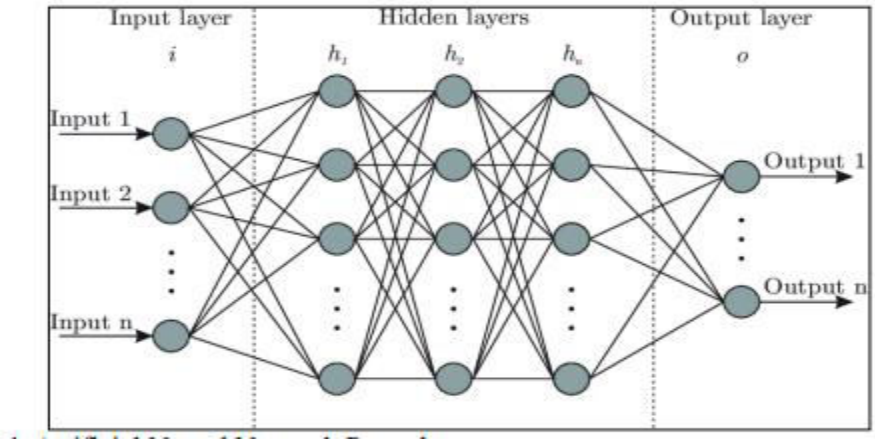

Fig3. Artificial Neural Network Methodology

Before putting the algorithms into practice, let's clarify the metric to measure the performance of our models. Stock price prediction being a fundamental regression problem, we can use RMSE (Root Mean Squared Error) or MAPE (Mean Absolute Percentage Error) to measure how close or far off our price predictions are from the real world.

Looking closely at the formula of RMSE, we can see how we will be able to consider the difference (or error) between the actual (At) and predicted (Ft) price values for all N timestamps and get an absolute measure of error.

$$RMSE = \sqrt{\frac{1}{N} * \sum_{t=1}^{N} (At - Ft)^2}$$

On the other hand, MAPE looks at the error concerning the true value – it will measure relatively how far off the predicted values are from the truth instead of considering the actual difference. This is a good measure to keep the error ranges in check if we deal with too large or small values. For instance, RMSE for values in the range of 10e6 might blow out of proportion, whereas MAPE will keep error in a fixed range.

$$MAPE = \frac{1}{N} * \sum_{t=1}^{N} \left| \frac{At - Ft}{At} \right|$$

Table 1: Comparison of SVM, RF & ANN models

	SVM		RF		ANN
ACCCURACY	95.66		97.65		98.21
F-measure	94.32		95.67		97.58
PREDICTION RATE	91.00		93.00		98.64

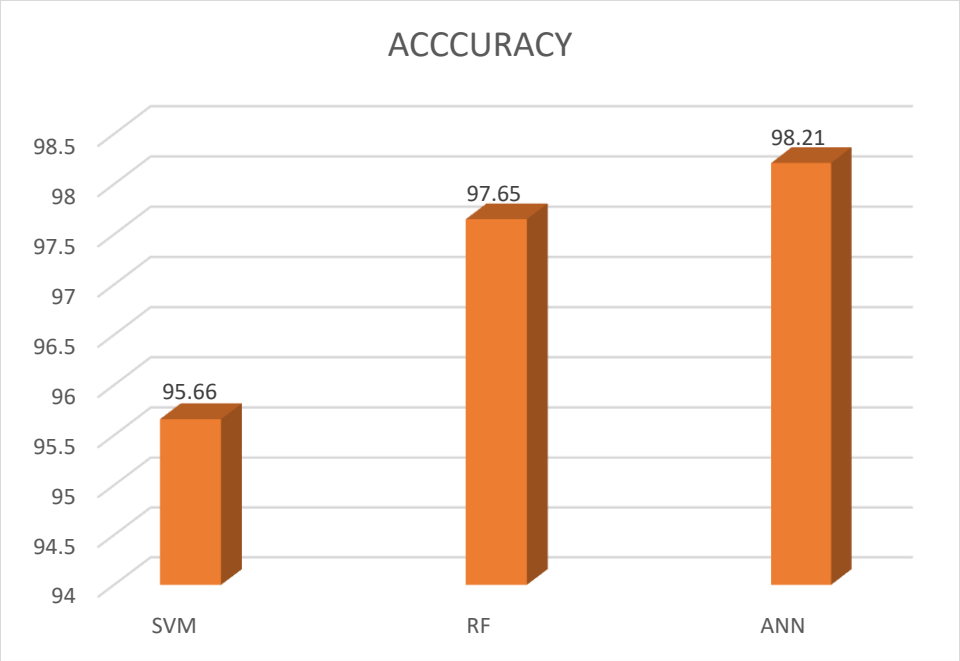

Fig.4 *Comparison of SVM, RF & ANN models by accuracy*

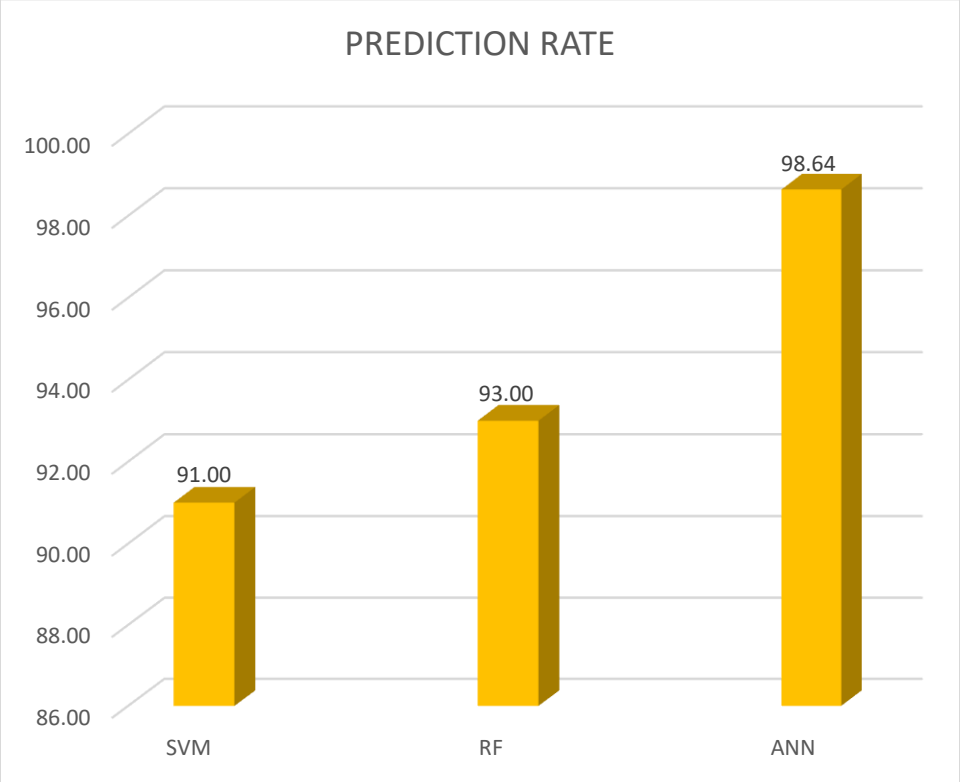

Fig.5 *Comparison of SVM, RF & ANN models by Prediction rate*

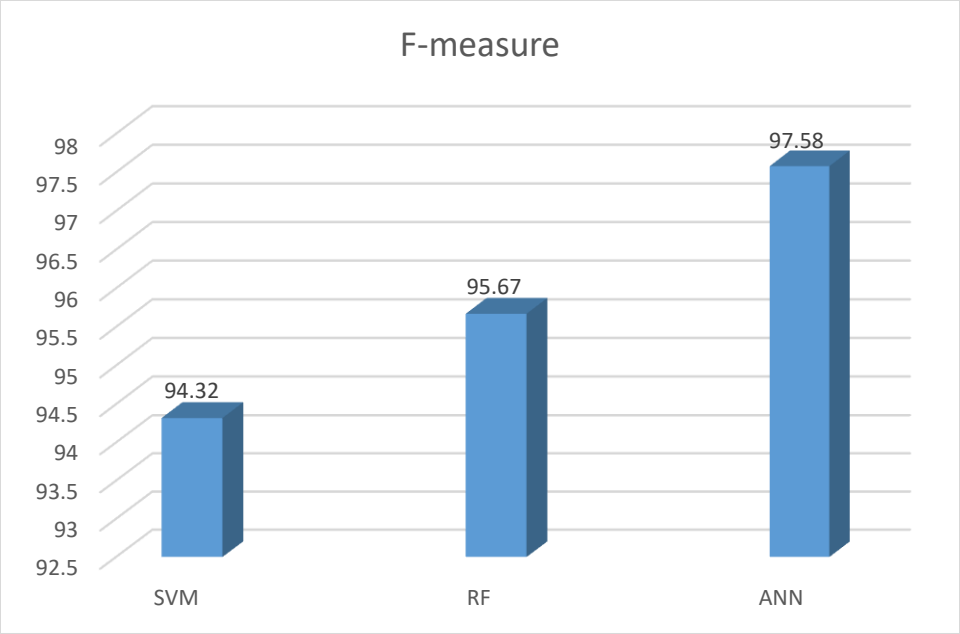

Fig.6 *Comparison of SVM, RF & ANN models by F-Measure*

IV. CONCLUSION

Machine learning techniques such as fundamental analysis, technical and sentiment analysis for stock price prediction can be very successful when applied correctly. It is important to apply these models in a way that will reduce errors and maximize results.

Predictingstockmarketreturnsisachallengingtaskduetoconsistentlychangingstockvalueswhicha redependentonmultipleparameterswhichformcomplexpatterns.Thehistoricaldatasetavailableonco pany'swebsiteconsistsofonlyfewfeatureslikehigh,low,open,close,adjacentclosevalueofstockprices ,volumeofsharestradedetc.,whicharenotsufficientenough.Toobtainhigheraccuracyinthepredictedpr cevaluenewvariableshavebeencreatedusingtheexistingvariables.ANNisusedforpredictingthenextd ayclosingpriceofthestockandforacomparativeanalysis,RFisalsoimplemented.Thecomparativeanal sisbasedonRMSE,MAPEandMBEvaluesclearlyindicatethatANNgivesbetterpredictionofstockpric esascomparedtoRF.ResultsshowthatthebestvaluesobtainedbyANNmodelgivesRMSE(0.394),MA PE(0.67)andMBE(0.011).Forfuturework,deeplearningmodelscouldbedevelopedwhichconsiderfin ancialnewsarticlesalongwithfinancialparameterssuchasaclosingprice,tradedvolume,profitandlosss tatementsetc.,forpossiblybetterresults.

REFERENCES

[1] Li,Lei,YabinWu,YihangOu,QiLi,YanquanZhou,andDaoxinChen.(2017)"Researchonmachinelearningalgorit hmsandfeatureextractionfortimeseries."IEEE28thAnnualInternationalSymposiumonPersonal,Indoor,andMo bileRadioCommunications(PIMRC):1-5.

[2] 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/WIECONECE54711.2021.9829585.

[3] Wanjawa, Barack Wamkaya, and Lawrence Muchemi. "ANN model to predict stock prices at stock exchange markets." arXiv preprint arXiv:1502.06434 (2014).

[4] Reddy, V. Kranthi Sai. "Stock market prediction using machine learning." International Research Journal of Engineering and Technology (IRJET) 5.10 (2018): 1033-1035.

[5] Ravikumar, Srinath, and Prasad Saraf. "Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms." 2020 International Conference for Emerging Technology (INCET). IEEE, 2020.

[6] 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.

[7] 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.

[8] Hiransha, M., et al. "NSE stock market prediction using deep-learning models." Procedia computer science 132 (2018): 1351-1362.

[9] 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.

[10] Kohli, Pahul Preet Singh, et al. "Stock prediction using machine learning algorithms." Applications of Artificial Intelligence Techniques in Engineering. Springer, Singapore, 2019. 405-414.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] Hegazy, Osman, Omar S. Soliman, and Mustafa Abdul Salam. "A machine learning model for stock market prediction." arXiv preprint arXiv:1402.7351 (2014).