International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 406-413 www.ijaem.net ISSN: 2395-5252

Exploratory Evaluation of Machine Learning Techniques Used for Stock Market Prediction throughout the Various Sequences of Time Periods

Chikile Srikar[1], Bhupinder Singh[2]
[1]Department of Computer Science, Lovely Professional University, Punjab.
[2]Department of Computer Science, Lovely Professional University, Punjab.

Date of Submission: 15-12-2024 Date of Acceptance: 25-12-2024

ABSTRACT— The stock market is a place where one can create fortune without physical activity. Stock market is one kind of financial market-place in which buyers and sellers trade the company's shares in the open market. Nowadays, everybody is interested in the stock market. Many people are losing money simply because of proper ignorance over the stock market. In this research, Simulations are going to check the machine learning algorithms to analyze the feature stock price and market behavior. These computer algorithms have been employed to forecast share values in the stock market to prevent losses and increase wealth. Artificial intelligence will examine information testimony for strategic insight. A simple difference between system learning and historical recursive text analysis is, while the popular algorithms simply scan a text for some specific keywords, machine learning understands what those words actually mean.

Keywords— Stock market, Machine learning, Logistic Regression, Random Forest, KNN algorithm, ARIMA.

I. INTRODUCTION

The stock market is a financial market where buyers and sellers trade shares of publicly held companies. Shares represent ownership in the company and give investors a claim on part of the company's assets and earning. Some of the variables that affect the price of a stock are macroeconomic conditions of use, corporate performance, and shareholder attitude. Equities are issues that one can buy into, hoping that over some period, the value may go up and capital gain will be realized. Stocks bear risks, too, and their values can fall, perhaps causing an investor to go through losses. Stock markets serve as a channel for the supply of capital, and therefore sources of funds for firms wishing to expand, while reflecting common economic conditions. It is well recognized that several components have recent "remembering" pertaining to the knowledge-based data applied up until this time. Sequencing pair data is a definite advantage that RNNs use frequently. Recurrent Neural Networks are well suited for stock market prediction, as they can grasp the temporal dependencies and patterns in the sequential data, such as trading volumes and historical stock prices. This allows the network to learn from historical trends and encapsulate sequential nature of the stock market data for prediction. Figure 1 describes the architecture of RNNs. In forecasting and prediction scenarios, RNNs may compute two inputs by closely looking at their internal state cell memory. Investors use the stock market as a channel to gain ownership in companies with the hope of earning returns through capital gains and dividends. People can earn money without any physical efforts.The evaluation includes the significance of feature selection and the robustness of the prediction models, and comparative analysis of predictive accuracy. The implemented models are Logistic Regression, Random Forest, KNN algorithm, and Arima. These diverse performance of the implemented models and their respective strengths and limitations are identified. And comparison of performance and accuracy of each model. The practical implications of this study extend to decision support for investors, trading professional, and financial analysts for stock market predictions.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 406-413 www.ijaem.net ISSN: 2395-5252

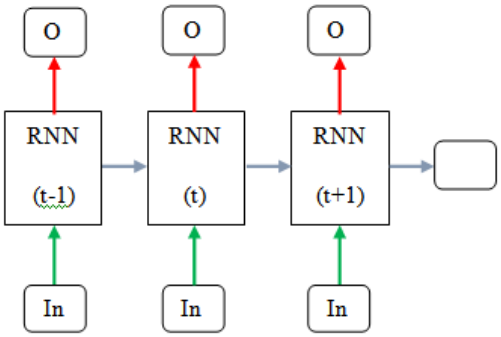

Fig. 1: Architecture of RNNs

Fig1 displays the structure of Recurrent Neural Network, it is type of neural network. In the figure the main component of Recurrent Neural Network is RNN cells. These cells are main units for Recurrent neural networks that helps to capture the sequential dependences in data. At each step time step the RNN cell process input and maintains an internal state to retain the previous memory inputs and the architecture indicates different time steps with (t-1), (t), (t+1) each step represents a point in the sequence being analyzed. The RNN cells process data sequentially through these time steps, allowing the network to learn patterns and relationships overtime. The connections between the RNN cells signify the flow information from one step to next step. This recurrent connectivity enables the network to retain memory and learn temporal dependencies sequential data like stock market prediction.

II. RELEATED WORK

In forecasting stock markets, some researchers prefer neural networks with back propagation methods. Ten years ago, statisticians were thought to be masters of making predictions concerning economies using the time series analysis. The main characteristic of a time series is that there is some sort of time dependence between the different observations. Recently, new development in artificial intelligence and with the impact of higher computation capacity the adoption rate in economic forecasting has been rising rapidly. This can be improved further by the use of various techniques like machine learning and evolutionary computation. The reason for the results of the above study was that markets are a random walk and the change in price at every step is an independent variable drawn from the same distribution.

III. PROBLEM STATEMENT

To predict the share prices in the stock market in the future, study have analyzed some research papers and given an overview and how much accuracy of the algorithms provided. Here study observed several algorithms from thus observed from some research papers where the algorithms are implemented to predict data that is needed. Everyone needs money to lead life here, stock is a place where study can generate money without any physical effort and simulation can earn money more than by doing business and jobs. Here the problem occurs due to lack of knowledge on stock market, most of the people loss money in stock market to reduce the loss of money, researchers did a survey on the stock market prediction through machine learning algorithms to reduce the loss of money.

Deep learning models with numerous hidden layers outperformed models with a single neuron, according to research by Qingfu Liu et al. [1]. DLNNs assessed pricing graphical layout charts using manual statistical methods. In his suggested piece of writing, Md. Mobin Akhtar [2] went into further detail about how a soothsaying device is used in the real world and the problems that can arise from using it irresponsibly. While the composition includes a framework education form that can be used to predict the life cycle of services in a high-demand market for corporate items, the Stock Expectation calculation is accurate 80.3% of the time. In order to predict the stock showcase list, Chaojie Wang et al. [3] took into account Transformer, the most recent deep learning algorithm. Transformer was initially developed to address the common dialect preparation problem; as of Encoder design, it has been linked to scheduling estimation. Dhanasekaren Kiran et al. [4] used deep learning and machine learning techniques to predict stock costs to within a few percentage points of accuracy. This research looks into the historical development of stock expectations as well as the most recent and efficient technique for supplying, predicting, and reducing the losses of speculators.

Yechan Han et al. [5] expounded on the importance of data labelling in the development of trading systems. NPMM labelling was developed as a solution to the problems of noise generation and information loss in up down labelling, which is widely used in trend prediction studies. NPMM labelling, on the other hand, uses all points of time to identify a particular trend by only labelling at the minimum and maximum points of time.

Dongdong et al.'s [8] strategy of pre-processing noisy starting data and concentrating on stock value in erratic markets was used. By examining the patterns, the author Park K et al. [9] developed a novel method to simplify noisy

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 406-413 www.ijaem.net ISSN: 2395-5252

financial series and discovered increases in accuracy of between 4 and 7 percent.
Sneh Kalra et al. [10] developed the process for trend prediction utilising news articles and historical data, and they estimated an accuracy range of 65.30 to 91.2%.

IV. METHODOLOGY

On the performance of machine learning algorithms mostly used programming language is "PYTHON". Python is a high level and object-oriented language it is easy to perform and it is a readable language. Python is an incredibly versatile analytical scripting language that is compatible with all common operational framework and is available as open-source software. Using this programming as a platform the machine learning models performed and evaluated. Four models are taken are into consideration those are Logistic Regression, Random Forest, K Nearest Neighbour Algorithm, Arima. Logistic Regression is chosen for its simplicity and effectiveness in binary classification tasks, Random Forest Regressor is selected for its robotic performance in handling complex relationships in data and reduces the overfitting, KNN algorithm is selected for its versatile method for pattern recognition and its intuitive approach of classifying instances based on similarity metrics, and Arima is included for capturing seasonal trends and autocorrections to predict stock market fluctuations over sequential time periods .These models are chosen for their applicability for stock market prediction and their ability to address different aspects of the dynamic and complex nature of financial markets.

V. MPELICATONS OF ALGORITHAMIC APPORACH

A. LOGISTIC REGRESSION

The method of choice for "supervised machine learning" is a logistic regression model. It is easier to use and more effective for binary and linear classification issues. A regression exhibit with a straight-forward variables that are dependent is called a logistic regression. The aforementioned model makes use of information from tests to determine the coefficients and utilises mathematical ability to determine the relationship between variables. In this model circumstance, the limit would be able to forecast the feature results using these coefficients. The logistic regression is the final requirement [6]. It uses a "S"-shaped curve to represent the non-linear relationship between x and y and the odds that y=1 will occur on the occasion that y happens. To find an algorithm that fits the information indications, it applies the logistic function. The information is illustrated by the function S-shaped curve. When y is a binary number, it is simple to use logistic regression since it is designed for binary situations with values of 0 or 1.

B. RANDOM FOREST

One type of machine learning model is the random forest. An algorithm for supervised machine learning is called Random Forest. It generates a single outcome by combining the output of several decision trees. It is a collective learning methodology that yields the class, which is the strategy [11] for the classes or mean figure of each individual tree. It works by building up innumerable trees during retraining time and applying backslide techniques and varied endeavours. Decision trees' tendency to overfit to the training dataset is compensated for by random decision forests [12].

$$Y = \frac{1}{N} * \sum_{i=1}^{N} f_i(X)$$

Where,
Y= Predicted value
N= No. decision trees in the forest
$f_i(X)$ is the prediction of the decision tree for input (X)

C. KNN Algorithm

KNN, or K Nearest Neighbour, is a machine learning algorithm. It's an algorithm for supervised machine learning. Complications with regression and classification can both be resolved with this technique. "K" stands for the number of closest neighbours to a newly discovered unexplained variable, which has been anticipated or classed.

$$\mathrm{Y} = 1/K \sum_{i=1}^{K} Y_i$$

Where,
Y = predicted value
K = Number of nearest neighbors
Y_i = Target value of the nearest neighbor to the input data point

- RMSD, or root mean square The overall RMSD is gusted into a single regarded parameter. Deviation is a fineness statistic that calculates the discrepancies between the estimated respects, Y, and honest to goodness regards, X. The method for handling "Explained Sum of Squares" is as follows.
- The average-estimated-error rate, or AEE, is the total of all RMS errors for each stock document characteristic divided by the total number of documents. The lift chart represents the shift in a data-mining visualisation when seen in comparison to a random estimator, and the shift

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 406-413 www.ijaem.net ISSN: 2395-5252

is expressed in terms of lift-score. The lifting-score for an assortment of informative file segments and distinct models may then be separated, allowing it to be selected [22] to indicate which instance is special and which level of the enlightening [23] list would benefit from employing the desires exhibit.

D. Arima

ARIMA signifies "Autoregressive Integrated Moving Average" which refers to an artificially intelligent learning algorithm. In economists and research, the ARIMA model is used for quantifying events that occur over time. This approach is applied to interpret historical data [13] or forecast data in a sequence for the future. It is defined by the three order parameters (p, d, and q). The AR(p) Autoregression regression model makes advantage of the dependent relationship between a current observation and observations from an earlier time frame.

An auto regressive (AR(p)) element is a regression equation for a time series that incorporates past values[21]. I(d) Arrangement: This method uses distinguishing measurements to take away an observation from an observation made at a previous time step in order to make the time sequence stationary.

When differentiation, a series's current values are deducted from its historical values d times [14]. The relationship between a measurement and the residual error of a moving average model applied to lag-time independent data is leveraged by the MA(q) Moving Average model.Three basic approaches are combined by the ARIMA methodology [20]: Autocorrelation: First, as indicated by the "p" esteem in the representation, auto-replace estimates of a specific time arrangement information are relapsed alone slacked esteems.

The term "differencing" (I stands for "integration") refers to the process of separating apart patterns and converting non-stationary time arrangements into ones that are stationary. The "d" esteem in the model [15] specifies this. Moving Average: The model's moving average characteristic is denoted by the letter "q," where "q" stands for the number of erroneous rate-related lag considerations.

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q} + e_t$$

Where,

Y_t = value of the time series at time (t)

$c = Constant$

$\phi_1, \phi_2, \ldots, \phi_p$ = Autoregressive coefficients representing the relationship between the current value and its past values

$\theta_1, \theta_2, \ldots, \theta_q$ = moving average coefficients representing the relationship between the current error and past errors

e_t = error or residual time

$Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$ = lagged values of the time series

$e_{t-1}, e_{t-2}, \ldots, e_{t-q}$ = lagged errors.

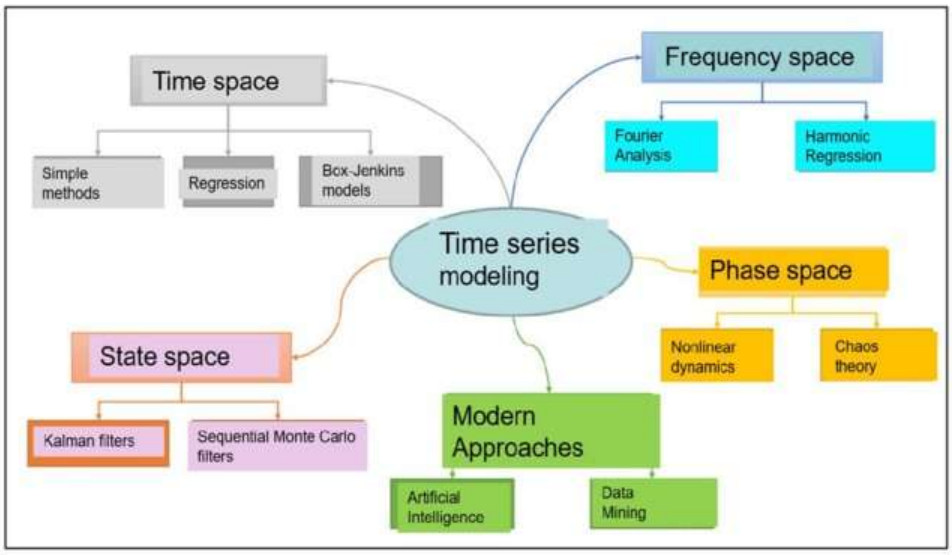

Fig. 2: Categorization of the principal categories of fundamental forecasting techniques and models

VI. MPELICATONS OF ALGORITHAMIC APPORACH

As in the papers they obtained the historical data and extensible time series data involving TATA Consultancy service stock information includes an average day's assessment for the months of February 2015 to February 2019.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 406-413 www.ijaem.net ISSN: 2395-5252

The qualities that make up the majority of the dataset are Data, Open, High, Low, Adj.close, and Volume [19]. A process diagram for evaluating the predictions made by various datasets for stock data is shown in Figure 3 .Figure 2 represents the Categorization of the principal categories of fundamental forecasting techniques and models.

Pre Processing

- Putting together a class label called Status with a numerical set of information based on the opening and closing prices of the stock; on the basis of the information below, if the stock's open price on a given day is less than its close, the number contained in the class label will be "1," which indicates a profit, and if the open price exceeds than the close, the appreciate will be "0," which indicates a loss on the stock market [16].

  "status" = 1, if shift > 0.55 0, otherwise.
- Preprocessing of the data is very important since the dataset consists of categorical attributes rather than numerical data, which might be problematic for machine learning algorithms. It involves the conversion of categorical data into a format analyzable, such as one-hot encoding or label encoding, and scaling for uniformity across values to facilitate better algorithm performance. Besides this, preprocessing also gives an indication of the data anomaly and mitigates many biases in data. In this way, processing the data enhances the outcome of machine learning strategies implemented in this work.
- 1048 combinations of observations are obtained, and data partitioning is done. The dataset is then divided into two halves. The first few sections are used to determine the model and component specialization. In the subsequent subsection, many degenerating models are exhibited and evaluated outside of the exam.
- The aim of feature selection is to select those attributes that are most relevant for a particular problem and that may be considered as contributing much in building an effective predictive model. By selecting the best features, the model can focus on the qualities that contribute most toward correctly identifying the target class. This will enhance the performance of the model by reducing noise and irrelevant data, hence improving generalization and prediction accuracy. Besides, feature selection makes the model leaner and more comprehensible. On the whole, it improves the effectiveness of the predictive modeling process.
- Following the predictive model is fitted, the cross-validation is carried out to verify the prediction error. The effectiveness of the model can be evaluated using a variety of calculations, as indicated in equations (1–5) following.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$R^2\ Score = 1 - \frac{RSS}{TSS} \quad (4)$$

$$F1\ Score = 2 * \frac{(Prec * Rec)}{(Prec + Rec)} \quad (5)$$

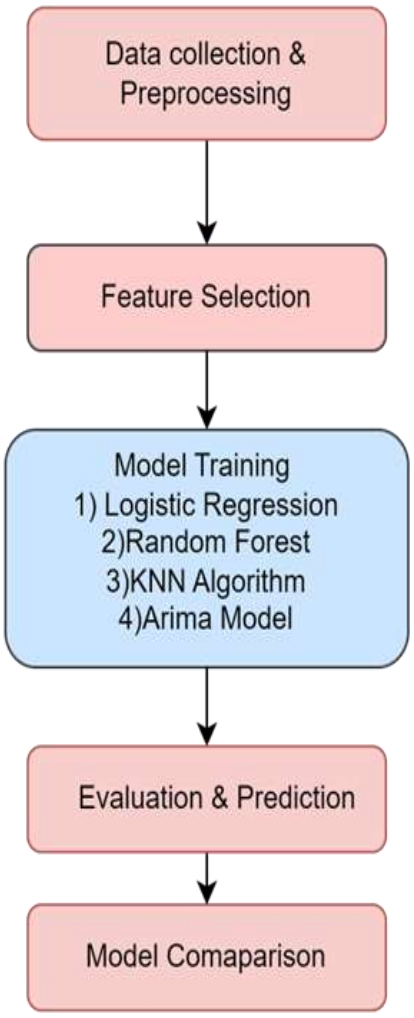

Fig 3. Flow chart of Evaluation of prediction of different datasets for stock data.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 406-413 www.ijaem.net ISSN: 2395-5252

From the above flowchart the process begins with the collection of data that is historical stock data from Yahoo finance and clean the data by handling missing values, normalizing features, and performing data transformations it is the input data for the subsequent stages of analysis and prediction. And second step is Feature Selection, choosing the relevant features from historical stock data, such as High, Low, Open, Close prices, trading volume, and any other significant indicators, third step is training the models of supervised learning. And performing each model separately and elevating the performance of each algorithm. And final step comparing each model performance and finding the model to fit for stock market prediction.

VII. MPELICATONS OF ALGORITHAMIC APPORACH

The training's data greatest weakness has been identified by studies, and testing has made it less close. The final result indicates that, within the data set, the greatest loss from height is thirty-eight percent if this technique is followed. Table 1 represents the Comparison of Accuracy Techniques implemented in related to stock prediction. Based on the Sharpe percentage of replication and the majority of drawdowns, the method's overall success is quite consistent with moderate losses [17]. The findings of this work practical implications for investors, trading professionals, and financial analysts by providing insights into the efficiency of machine learning models for prediction. And accuracy and robustness of the models was rigorously evaluated against the actual stock market movements. And understanding the requirements and limitations of each model is crucial for improving prediction accuracy and avoiding slowdowns.

Sr.no	Applied method	Reliability	Restrictions
1.	Random Forest (Ensemble Learning)	53.48%	Insufficient trees hinder accurate predictions, leading to model slowdown.
2.	Logistic Regression (non-Ensemble learning	66%	Choosing the Proper Attributes to Incorporate
3.	ARIMA	45%	Yields less precise outcomes
4.	K-Nearest Neighbour	64%	The information set needs to be normalised, and classification variables are difficult to manage.

Table 1. Comparison of Accuracy Techniques Implemented for Stock Prediction of TATA Consultancy Services Ltd

After collecting the dataset about TATA Consultancy service Ltd, then applying as discussed machine learning models. On the above comparison of above machine learning algorithms like, Random Forest has accuracy rate is 53.48%, Logistic Regression accuracy is 66%, ARIMA accuracy is 45%, K-Nearest Neighbour accuracy is 64%. On comparison of four machine learning models Logistic Regression giving more accuracy then compared remaining models that is 66% accuracy [18].

VIII. CONCLUSION

Research found that, after compressing four models, Logistic Regression provides a more accurate means of predicting and assessing the trajectory of the market's movement than any of the other approaches now in use. But, in stock market any model cannot predict with more accuracy because price of a stock is based various parameters by only collecting the previous data of a stock, research cannot predict the price of a stock but somehow, study can use these kinds of models for less risk. Furthermore, because the global indices of the entire world are completely dependent upon one another, This Study is unable to produce superior outcomes in scenarios like to a battle between two superpowers. Corporate insider trading presents a significant prediction difficulty because our research is entirely data-driven. The recurrent cost of this system lowers the profit margin due to the costly nature of broker connectivity and information from

a third-party source. The expense of administering and testing new solutions on cloud infrastructure increases since hosting machine learning systems requires a lot of resources.

REFERENCES

[1]. Qingfu Liu, Zhenyi Tao, Yiuman Tse, Chuanjie Wang,"Stock market prediction with deep learning: The case of China",Finance Research Letters,2022,Volume 46, Part A,102209, https://doi.org/10.1016/j.frl.2021.102209.

[2]. Md. Mobin Akhtar, Abu Sarwar Zamani, Shakir Khan, Abdallah Saleh Ali Shatat, Sara Dilshad, Faizan Samdani,"Stock market prediction based on statistical data using machine learning algorithms",Journal of King Saud UniversityScience,2022,Volume34,Issue4, 101940, https://doi.org/10.1016/j.jksus.2022.101940 .

[3]. Chaojie Wang, Yuanyuan Chen, Shuqi Zhang, Qiuhui Zhang,"Stock market index prediction using deep Transformer model",Expert Systems with Applications,2022,Volume208,118128

[4]. Dhanasekaren Kiran, Aluri Teja Sri, Karthikeyan Neeraj, Baskaran Hari Saravanan and Selvanambi Ramani," A Study on the Impact of Sentiment Analysis on Stock Market Prediction", Recent Advances in Computer Science andCommunications,2023;16(1):e250522202256.

[5]. Yechan Han, Jaeyun Kim, David Enke," A machine learning trading system for the stock market based on N-period Min-Max labeling using XGBoost", Expert Systems with Applications,Volume 211,2023,118581,ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.118581 .

[6]. A. A, Adebiyi, A. K, Charles, A. O, Marion, and O. O, Sunday, "Stock Price Prediction using Neural Network with Hybridized Market Indicators," J. Emerg. Trends Comput. Inf. Sci., vol. 3, no. 1, pp. 1–9, 2012.

[7]. S.Kayode, "Stock Trend Prediction Using Regression Analysis – A Data Mining Approach," ARPN J. Syst. Softw., vol. 1, no. 4, pp. 154–157, 2011.

[8]. Dongdong Lv ,Shuhan Yuan,Meizi Li and Yang Xiang,(2019).An Empirical Study of Machine Learning Algorithms for Stock DailyTrading Strategy.Mathematical Problems in Engineering ,2019.

[9]. Park K, Shin H. (2013). Stock price prediction based on a complex interrelation network of economic factors.Engineering Applications of ArtificialIntelligence,26,1550–1561 https://doi.org/10.1016/j.engappai.2013.01.009

[10]. Sneh Kalra, Jay Shankar Prasad. (2019, February). Efficacy of News Sentiment for Stock Market Prediction. International Conference on Machine Learning, Big Data, Cloud and Parallel Computing.

[11]. H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend," ICSSBE 2012 - Proceedings, 2012 Int. Conf. Stat. Sci. Bus. Eng. "Empowering Decis. Mak. with Stat. Sci., pp. 99–103, 2012.

[12]. B.Sivakumar and K. Srilatha, "A novel method to segment blood vessels and optic disc in the fundus retinal images," Res. J. Pharm. Biol. Chem. Sci., vol. 7, no. 3, pp. 365–373, 2016.

[13]. M. P. Mali, H. Karchalkar, A. Jain, A. Singh, and V. Kumar, "Open Price Prediction of Stock Market using Regression Analysis," Ijarcce, vol. 6, no. 5, pp. 418–421, 2017.

[14]. L. Khaidem, S. Saha, and S. R. Dey, "Predicting the direction of stock market prices using random forest," vol. 0, no. 0, pp. 1– 20, 2016.

[15]. T. Manojlović and I. Štajduhar, "Predicting stock market trends using random forests: A sample of the Zagreb stock exchange," 38th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2015, no. May, pp. 1189–1193, 2015.

[16]. S. B. Imandoust and M. Bolandraftar, "Forecasting the direction of stock market index movement using three data mining techniques: The case of Tehran Stock Exchange," Int. J. Eng. Res. Appl., vol. 4, no. 6, pp. 106–117, 2014.

[17]. K. Alkhatib, H. Najadat, I. Hmeidi, and M. K. A. Shatnawi, "Stock price prediction using K-nearest neighbor algorithm," Int. J. Business, Humanit. Technol., vol. 3, no. 3, pp. 32–44, 2013.

[18]. S. H. Liu and F. Zhou, "On stock prediction based on KNN-ANN algorithm," Proc. 2010 IEEE 5th Int. Conf. Bio-Inspired Comput. Theor. Appl. BIC-TA 2010, pp. 310–312, 2010.

[19]. Bhupinder Singh and Dr. Santosh Kumar Henge, "Evaluation of Neural Fuzzy Inference System and ML Algorithms for Prediction of Nifty Large Cap Companies Based Stock Values", International Conference on Intelligent and Fuzzy Systems, Springer 2021, Cham, pp 147–154.

[20]. BhupinderSingh,Dr.SantoshKumarHenge.(2020,July).Neural Fuzzy Inference Hybrid System with SVM for Identification of False Singling in Stock Market Prediction for Profit Estimation. Intelligent and Fuzzy Techniques: Smart and Innovative Solutions.