iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VIII **Month of publication:** August 2020

DOI: https://doi.org/10.22214/ijraset.2020.30756

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: 7.429
Volume 8 Issue VIII Aug 2020- Available at www.ijraset.com

Prediction Analysis of Stock Price using Machine Learning Approach

Ms. Neha[1], Mr. Roop Lal[2]
[1]*M.tech Scholer,* [2]*Assistant Professor, St.Soldier institute of engineering & management,Jalandhar (PTU)*

Abstract: ***Stock market is a real-time, complex, dynamic that poses a challenge to the individual investors, institutional investors, and financial institutions.***
Now days we can predict the market behavior precisely using machine learning based optimization techniques, database technology that helps to predicts stock market index more accurately. Though many specialized machine learning algorithms are already established, there are still scope for developing innovative models or systems which can cater the rising needs of investors.
The objective of this thesis is aligned towards the designing of some efficient machine learning based hybrid models that are more suitable and better implementations for analysis in the field of financial markets. In this research work we have tried to contribute to this objective by proposing some novel hybrid models and validated them by applying into for predicting the stock market price of different company's share.
This study also covers the scope and efficiency of different machine learning methodologies for attending the persisting problems of Stock Exchange Marketing paradigms.
The objective of the thesis is to identify forecasting techniques that helps to predict future stock returns on the basis of RMSE and R-Square parameters using supervised algorithms.
Keywords: Stock Variables, Analytics, Supervised Learning

I. INTRODUCTION

The nature of stock market in common is unpredictable which may depend on the long and short term future state. Companies need money for their establishment dumps their partial ownership in terms of share at any of the agency for sale. By this way many companies dump their share at different agency for sale.
Since large number of publics are involved and to make the trading in systematic way, govt had involvement and created an organization named as stock exchange.
Through stock exchange, a company can enlist their company's name for selling of their share and the public can purchase as per their choice and ability to investment.
The business originated during selling and purchasing of share creates a market known as share market or stock market A stock market is an open market for companies or for individuals to raise money. Stock market helps companies to purchase or sell their shares.
The cost of shares relies on the interest and supplies of shares. This procedure of purchasing and selling of shares is called trading/exchanging. The main aim is to reduce this unpredictability and the Stock Market Anticipation is used in this process.
Necessity is the cause of innovation. So a new concept of arrangement of finance was floated and termed as "share marketing". So companies invite the public to be partnership with the company by investing a token of money. Instead of choosing one or two partner of large investment, company prefers to be partner with more number of investor with small investment. For this process, company releases bond of face value of rupees 10/- or 100/- or 1000/- each.
These bonds are termed as Share. An investor can purchase any number of shares below the limit mentioned in the advertisement. For the safe guard of public money, Government has framed some rules that must be followed by the companies before release of shares.
The time series type of forecasting is a statistical concept, where the outcomes of several past data is studied, the direction of change of every outcome data is marked and then the future outcome is forecasted. The Method of Time Series Forecasting plays vital role in forecasting financial market, rain fall and many diversified sectors. The share trading is defined as purchasing and selling of shares of companies.

A. *Stock Market Trading Theory*

The two most vital hypotheses in stock market forecast are Efficient Market Hypothesis (EMH) (Kadir Can Yalçın, 2010) and random walk theory (Seyyed Ali PaytakhtiOskooe, 2011).

1) *Efficient Market Hypothesis (EMH):* In 1970, EMH invented by Eugene Fama that incorporates three structures: weak, semi-strong, strong efficiency. *Weak EMH*, future expenses can't be foreseen by basically looking at verifiable expenses. Speculators can't yield returns over the long run by utilizing speculation methods focused around historical costs or other verifiable data. Under weak efficiency specialized investigation procedures won't have the ability to dependably make wealth returns. Then again, a few types of fundamental analysis provide excess returns. The *semi-strong form of EMH* goes above and beyond by consolidating all historical and current information into the cost. The *strong form of EMH* in corporate historical, public, and private data, for example, insider data, in the share cost.

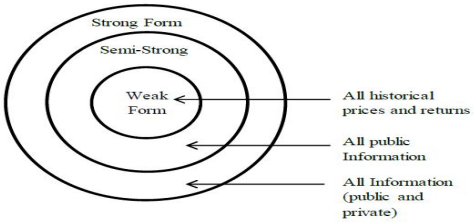


Fig 1.1 EPH

2) *Random Walk Theory (RWT):* Random walk theory is same as semi-strong Efficient Market Hypothesis where all public data accessible to everybody. RWT guarantees that using such data, future forecast is not much successful. The RWT expresses that stock value movement not rely on stock warehouse i.e. past sock. With the upcoming of efficient computing infrastructure, stock exchanging companies develops effective algorithmic trading frameworks that can abuse the hidden pricing patterns when enormous measures of information are present.

B. *Stock Market Prediction*

The predictors of stock market concentrate on creating methodologies which effectively foresee stock costs utilizing overall characterized trading methodologies. A fruitful anticipation model is the particular case that works with best exactness having least input requirements and minimum complex model.

1) *Fundamental Analysis Based Stock Price Prediction:* Fundamental analysis is a system for figuring out the future cost of a stock which a financial specialist wishes to purchase. It is the assessment of the key obliges that impact the enthusiasm of the economy, mechanical areas and associations. It tries to predict future development of the capital market utilizing signals from the Economy, organization and Company. It requires an analysis of the market from a more extensive point of view. The assumption behind fundamental analysis is that a prospering economy develops modern advancement which prompts improvement of associations. Assessment of genuine worth of a stock is made by considering the procuring capability of the company which relies on upon speculation environment and elements identifying with particular industry, competitiveness, nature of administration, operational proficiency, benefit, capital structure and profit. During fundamental analysis speculators predict future changes deals, GDP and other commercial ventures and firms performance parameters. To perform the fundamental analysis, the investors and analysts use two ways alternatively names: “Top-down” and “Bottom-up” forecasting approaches (Table 1.1).

Table 1.1 Forecasting Approaches Used In Fundamental Analysis

Forecasting approach	Sequence of forecasting
top-down	Initially, the investors do the analysis and prediction for the economy, next for the industries and at last for the organization, the anticipation of industry is based on the economy whereas the anticipation of the organization is based on the prediction of both economy and industry.
bottom-up	Just a reverse of top-down approach. First the investors initiative the analysis part and go for the companies forecast, then for industries and finally target the economy.
Combined	Here both methods merge together. First, the investors do the analysis and prediction for the economy using top-down approach and after using the bottom-up approach.

2) *Premises of Technical Analysis*

a) *Market action Discounts Everything:* The important feedback of the technical analysis is that it just consider price of the share regardless of the fundamental factors of the company. Basically the technical analysis believe that at any given time, the price of the stock reflects everything that has or could affect the organization. Also the technical analysts believe that the fundamental of the organization along with the financial components are prices into the stock.

b) *Price Moves –Trends:* Price developments are accepted to for patterns following in technical analysis. This infers after an pattern has been made, the future value advancement is increasingly inclined to be a similar way as the pattern than to be against it.

c) *Repeat History Tends Itself:* History has a tendency to repeat itself, such as price movement is the core consumption in Technical Analysis. The recursive nature of price developments is credited to market psychology; or else, market competitors have a tendency to give a reliable response to comparable market boosts over time. Technical analysis uses chart patterns to dissect market movements and comprehend patterns. Though numerous numbers of these charts has been utilized for more than 100 years, they are still accepted to be relevant because they demonstrate the patterns in price developments that frequently repeat themselves.

3) *Price Charts:* A price chart is plotted using prices versus time span. The high, low and closing costs are needed to structure the price plot the basic requirements are different costs (closing cost/high cost/low cost). Vertical bar top represents the high cost and vertical bar bottom represent the low cost. Short horizontal line that crossing the vertical bar represents the closing cost.

Fig 1.2 Price Chart

4) *Trend Lines:* Trend lines are an imperative mechanical assembly in specialized investigation for both trend ID and confirmation. A trend line is a straight line that partners at least two value focuses and afterward connects into the future to go about as a line of help or obstruction. Support is the price level at which request is sufficiently strong to keep the price from declining further. Resistance is the price level at which offering is sufficiently solid to keep the price from rising further.

a) *Uptrend Line:* An uptrend line has a positive incline and is confined by partner at least two focuses. The subsequent low should be higher than the first for the line to have a positive incline. Uptrend lines as continue as help and show that net (demand less supply) is growing even as the value climbs. An expanding cost combined with expanding request is very bullish, and shows a strong assurance from the buyers. Till the costs remain steady over the pattern line, the uptrend is seen as powerful and flawless. A break underneath the uptrend line shows that net-request has incapacitated and an adjustment in pattern could be quick as appeared in Fig 1.3.

Fig 1.3 Uptrend Line

b) Downtrend Line: A downtrend line represented using negative decline and is shaped by interfacing at least two high focuses. The subsequent high should be lower than the first for the line to have a negative decline as appeared in Fig 1.4

Fig 1.4 Downtrend Line

C. Technical Indicators

Technical indicators give unique identification of the efficient direction of the price action of the stock. Indicators serve three principle capacities: to alarm, to affirm and to foresee. There are several indicators being used today.

1) Moving Averages: Moving averages are familiar and flexible for recognizing price trends. They smooth out the changes in market prices, which in turn make it easier to estimate the upcoming trends. Most commonly using moving averages types are: simple and exponential. A simple moving average is calculated using the average cost of the stock over a predefined number of periods as shown in Fig 1.5 using example constraint.

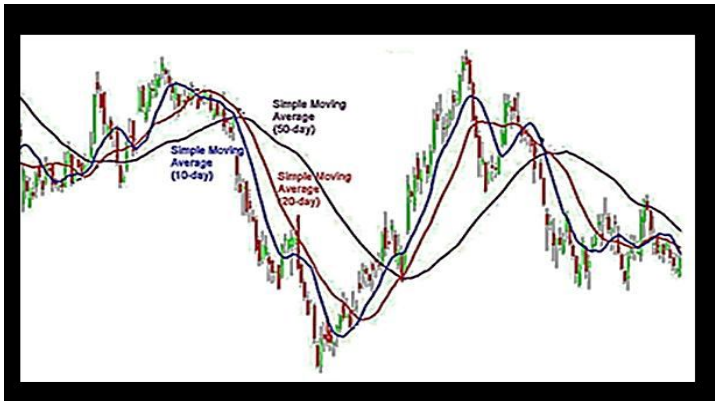


Fig 1.5 Simple Moving Averages

An exponential moving average performs as same as simple moving average, except a “multiplier” is included to give the latest price(s) more weight. Fig 1.6 shows exponential moving average chart.

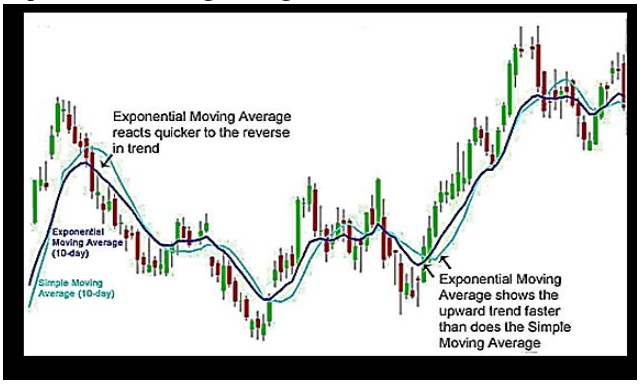


Fig 1.6 Exponential Moving Averages

D. Machine Learning Techniques

Machine learning (ML) is a data analytics technique which acquires knowledge from the previous practice and provides information directly to the systems without being programmed explicitly. Although the term Machine Learning (ML) was developed by author Samuel in 1959, its advancement and expansion has never been more instrumental than it is today. The prospect of using machine learning for decision making in numerous fields has a important function in the upliftment of the people.

Supervised learning is a concept, consist of a target that to be predicted among a set of predictors provided. Based on the outcome as well as input set of variables, also called training samples, a mapping is constructed which maps to required output. The training undergoes a set of iterative steps of inferring and goes on up to the replica acquire a required benchmark of precision. Some illustrations of supervised learning are Linear Regression, Logistic Regressions, Decision Tree, and SVM.

Unsupervised learning concepts don't have any target or outcome variable in the training datasets to draw inferences. Since the samples given to the learner are unlabeled, there is no evaluation of the accuracy of the output by the relevant algorithm. It is commonly used for clustering problems. Some of the examples that come under this category are Apriori Algorithm and K-Means.

Reinforcement learning algorithms uses previous practice to acquire the best knowledge and convert it in taking appropriate decisions in business. In this case, machine has been exposed to an atmosphere of training continuously applying trial and error to discover which action will yield the maximum reward. Markov Decision Process is one such example of reinforcement learning. The categories of problems that can be addressed with machine learning techniques are regression, classification and clustering. Regression and classification can be categorized as supervised learning whereas clustering as unsupervised learning. If the forecasting outcome is a continuous one then it considered as regression one. In classification type the forecasting output is represented as Boolean type. Clustering refers to arrangement alike attributes in same group.

The authors Banerjee, et al. (2013) explained basically data analytics, which is also called analysis of data, may be categorized into four types:

Descriptive analytics juggles raw historical data to draw conclusion about past and the output is represented by using only right or wrong without stating the reason. It basically handles quarries of type "what happened"

Diagnostic analytics tries to locate interdependencies and classify the patterns. It handles the question of type "why something happened", historical data are measured against other data.

Predictive analytics applies the conclusion drawn from above two categories for identifying about the trends and behavior of data and predicts the future.

Prescriptive analytics is based on the concept of optimization that helps to understand the uncertainties to make better decisions and achieve the best outcomes.

II. REVIEW OF LITERATURE

Zuoquan Zhang et al, (2012) proposed a study of stock market perdition using big data. Author examined how big data received from huge datasets in structured, semi-structured, unstructured format and how these data format used in stock market domains using machine learning algorithms.

Grinberg examined that how the data is labeled from OSN pages. Author proposed some web services models in which training and deployment can be easily implemented.

Lai, L. and Liu, J. (2014) predict the stock market using SVM. Author also used three more algorithms: Least Square, SVR and GARCH using wavelet kernel to sort out financial time series problems.

Lin et al, (2013) implemented the SVM for stock market forecasting. Author used the variables such as feature subset, stock indicator, and Taiwan stock market datasets and conclude model performs better than traditional stock market forecast system.

Kim et al, (2006) projected new hybrid model of Artificial Neural Network and Genetic Algorithm for attribute discretization. Attribute discretization concepts to convert continuous data into discrete data using certain thresholds. Rightly discretized data makes the learning simple and developed the quality of the learned outcomes.

Khashei and Bijari, (2011) presented a hybrid ANN using auto-regressive ARIMA-ANN model that divide a time series data into linear and nonlinear form to capture different forms of relationship of the data.

Khalid Alkhatib et al (2013) implemented non-linear regression using KNN for targeting Jordanian stock trade for forecasting stock price.

Dase & Pawar (2010) discussed the literature review for stock market predictions using Artificial Neural Network and analyze an Artificial Neural network may be suitable for proposed activity.

Chen, M.Y. (2013) performed a hybridization of NLICA and SVR with PSO for forecasting the value of share. In the planned model, the NLICA was applied to find out the relevant and de-noised features out of observed primal data, SVR was applied for building forecast model, and PSO was employed to decide optimal parameters of SVR.
Guresen et al (2011) implemented the ANN model, integrated Fama and French model for Chinese stock market prediction.
Amit Ganatret & Kosta (2010) evaluate stock market forecast using neural network. Authors used R language to constraint the stock variables such as inflation, global indices, closing price, interest rate, currency rate and crude price.
Abhishek Gupta et al (2014) proposed a model that used combined features of classification and clustering techniques. For clustering author used K-means algorithm and for classification used horizontal partition decision tree. This combined method produced more accurate and efficient result.
Das et al (2016) implemented another hybrid model by integrating SVM and TLBO for forecasting the price of future in stock market.
S Baadel et al. (2019) discussed phishing as a classification problem. Author also discussed recent trending machine learning techniques (dynamic self-structuring neural network, associative classifications, dynamic rule-induction, etc.) that used in various anti-phishing models.
A Alsaeedi et al. (2019) proposed a Evaluation Framework for Twitter Sentiment Analysis (EFTSA) whose evaluation based on multiple dataset that help to compare the Twitter sentiment approaches between/against others.

III. RESEARCH FINDINGS

For the prediction of financial market by implementation of different statistical and machine learning based optimization models, an existing research gap has been identified specifically when the learning model is trained with very large number of data set samples. Even if SVR is recognized as a proficient method for simplification in such scenarios, more appropriate hybrid models are required for the better prediction accuracy.

IV. TECHNIQUES/ METHODOLOGIES- STOCK PRICE PREDICTION

This section will give you the detailed analysis of each process involved in the project. Each sub section is mapped to one of the stages in the project.

A. Data Pre-Processing

The data pre-processing stage involves

1) Data Parameterization
2) Data transformation using normalization
3) Data Cleaning
4) Data Integration

B. Feature Selection and Feature Generation

In Feature Selection, select features according to the k highest scores, using linear model under single regressor, and sequentially go for many regressors. Select K Best Algorithm, with f regression as the scorer for evaluation.

1) *Work Division:* The work division is currently done based on a per algorithmic implementation and testing. This will be further decided as the project progresses.
2) *Datasets:* Nasdaq, Quant Quote minute resolution data files, Google Finance, Yahoo Finance
3) *Libraries:* Numpy, Scipy, Pandas, Matplotlib, Scikit-Learn, Seaborn.

V. RESULT ANALYSIS

Result analysis can be evaluated on the basis of two performance metrics such as Root Mean Square Error (RMSE) and r^2 score value.

A. Root Mean Squared Error (RMSE)

Also called Root Mean Squared Deviation (RMSD) is the frequently used measurement of difference between predicted value and actual value i.e. is also called residuals.
RMSE used to identify how much errors arise between two dataset

$$RMSErrors = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

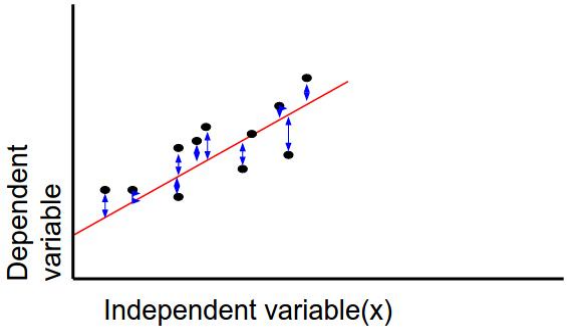


Fig: RMSE

B. R-Squared Value (r^2 value)

R2 value lies between 0 and 1 range, and for more accurate value its value should becomes higher. R-squared value also known as the coefficient of determination, or the coefficient of multiple determinations. R-Squared value is the proportion of the variance in the dependent variable that is predictable from the independent variable(s).

Table 4.1 : Classifier evaluation

Algorithm	RMSE Value	R-squared Value
Gradient Boosting Regressor	1.2745	0.96145
KNeigbhours Regressor	0.000390	-117.0117
Adaboost Regressor	2.9883	0.909611
Bagging Regressor	1.32996	0.959771
Random Regressor	1.43254	0.956669

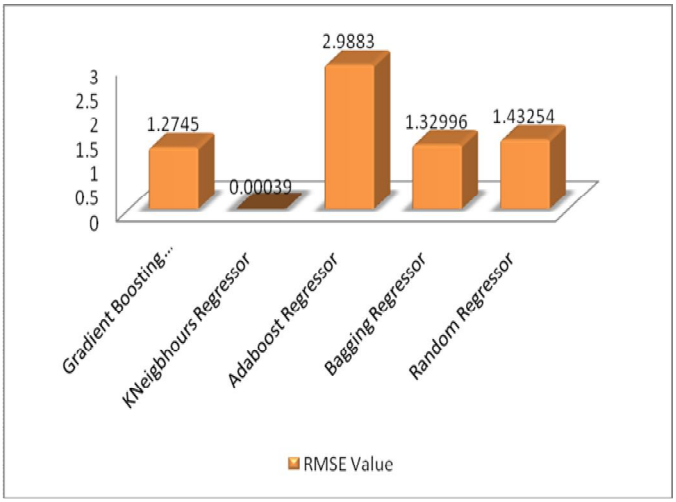


Fig. 4.1. Comparison Graphs RMSE Value - Different Models

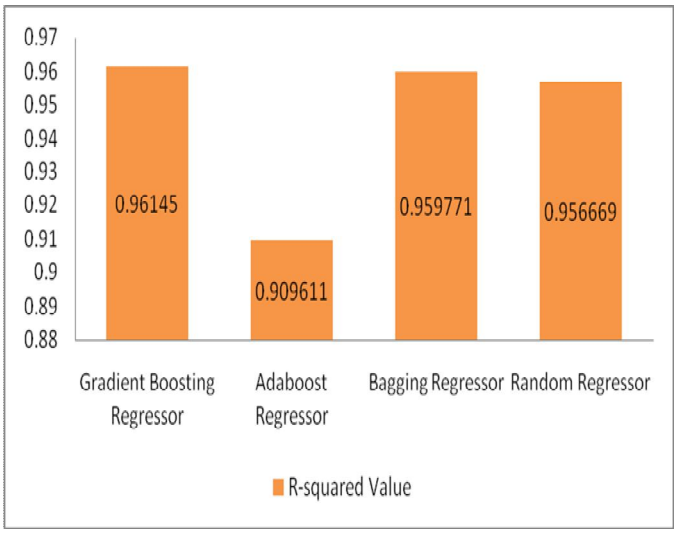
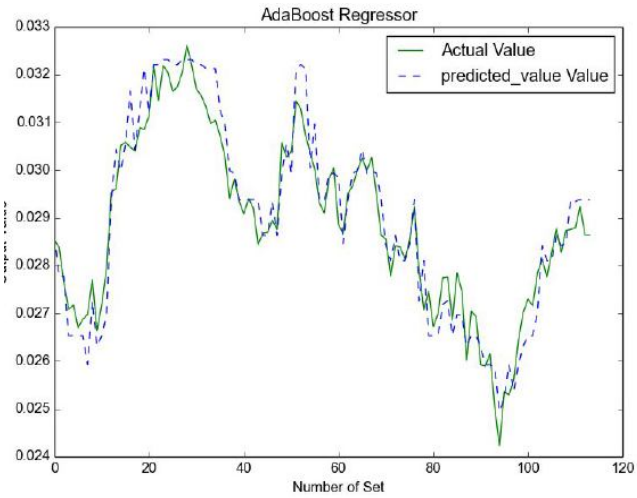
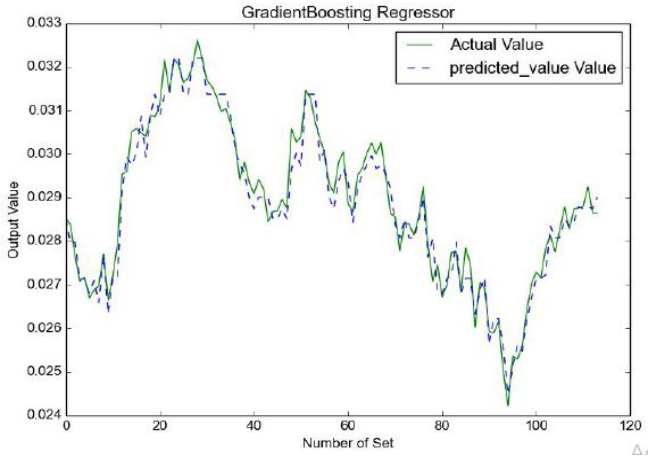


Fig. 4.2 Comparison Graphs R-squared Value - Different Models

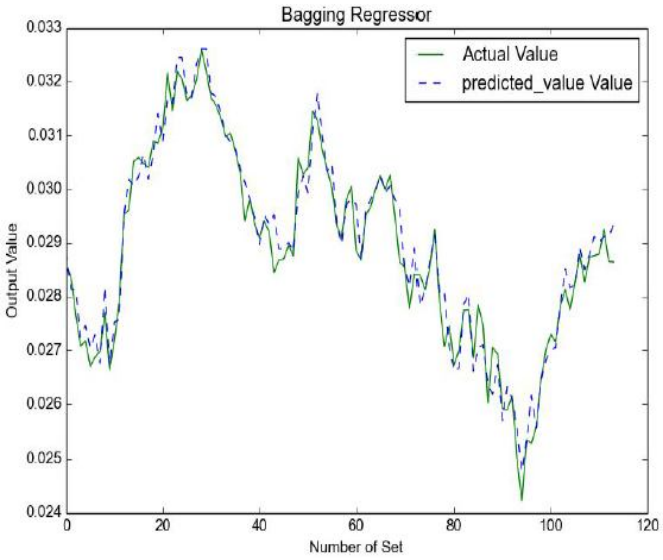
Results: Adaboost Regressor



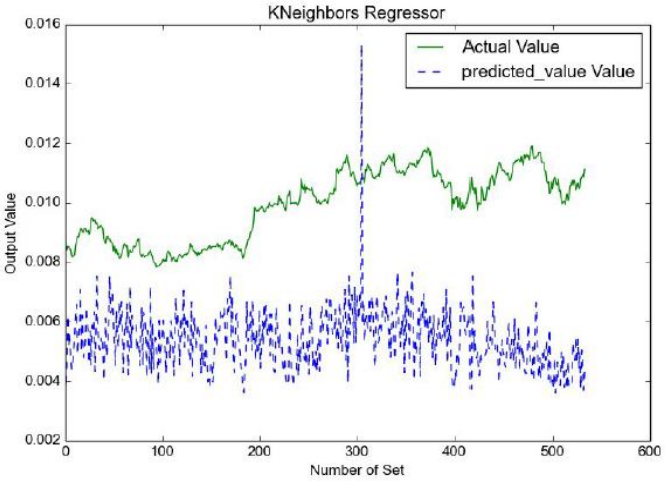
Results: Gradient Boosting Regressor



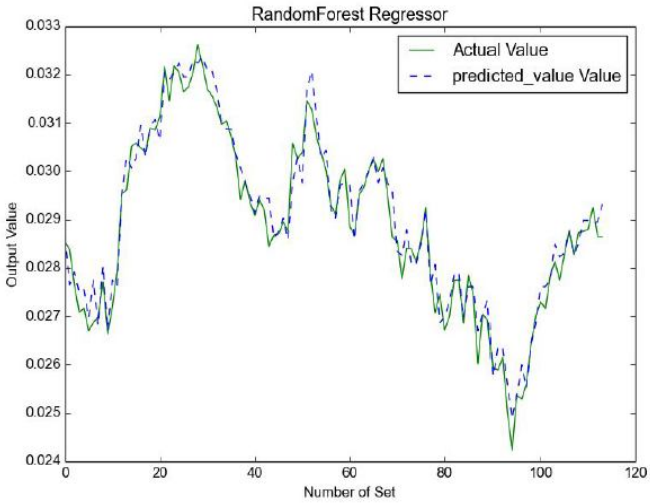
Results: Bagging Regressor



Results: K Neighbours Regressor



Results: Random Forest Regressor

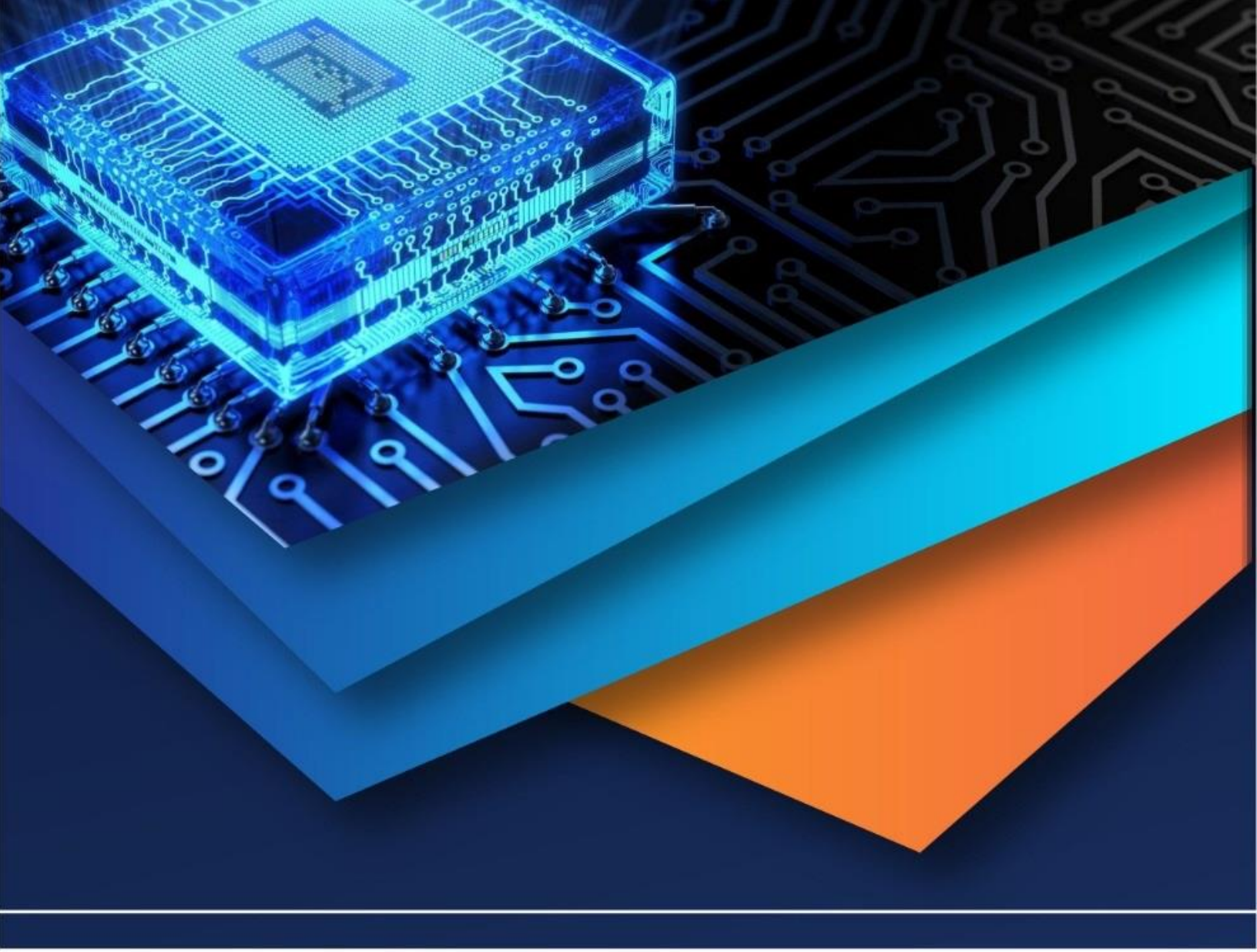


VI. CONCLUSION

On the basis of result found for stock price prediction it is found that Gradient Boosting Regressor algorithm performs the best in consistent manner. Bagging Regressor also performs better for stock price prediction so it also play vital role for researcher for future prediction sing some more hybrid models. Other data mining procedures and optimization methods, for example, ant colony optimization, particle swarm optimization can be applied to anticipate the future qualities and contrasting the outcomes and the strategies utilized in this paper. Hence, in this research work data mining and developmental methodologies have been applied to anticipate the future stock worth.

REFERENCES

[1] Zuoquan Zhang, Menggang Li & Rongquan Bai 2012, 'An integrated model for stock price prediction based on SVM and ARMA', Journal of Theoretical and Applied Information Technology, vol. 44, no.1, pp.51-54
[2] Grinberg, M. (2018). Flask web development: Developing web applications with python.O 'Reilly Media, Inc.
[3] Lucas Lai & James Liu 2014,'Support Vector Machine and Least Square Support Vector Machine Stock Forecasting Models', Computer Science and Information Technology, vol. 2, no.1, pp. 30-39
[4] Lin, Yuling, Haixiang Guo & Jinglu Hu 2013, An SVM-based approach for stock market trend prediction', Neural Networks , The International Joint Conference on, IEEE, pp.1-7
[5] Hyun-jung Kim & Kyung-shik Shin 2006,'A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets', Applied Soft Computing, vol.7, no.2, pp.569-576
[6] Khashei, M. and Bijari, M. (2010). An artificial neural network (p, d, q) model for time series forecasting, Expert Systems with Applications, Elsevier, Vol. 37: 479–489.
[7] Khalid Alkhatib, Hassan Najadat, Ismail Hmeidi & Mohammed K Ali Shatnawi 2013,'Stock Price Prediction Using K-Nearest Neighbour Algorithm', International Journal of Business, Humanities and Technology, vol.3, no.3, pp.32-44
[8] Dase,RK & Pawar,DD 2010, 'Application of Artificial Neural Network for stock market predictions: A review of literature', International Journal of Machine Intelligence, vol. 2, no. 2, pp-14-17
[9] Chen, M.Y. (2013). A hybrid ANFIS model for business failure prediction using PSO and subtractive clustering, Informatics and Computer Science Intelligent Systems Applications, Vol. 220, Elsevier: 180-195.
[10] Guresen,E , Kayakutlu,G & Daim,TU (2011) ,'Using artificial neural network models in stock market index prediction', Expert Systems with Applications, vol.38, no.8, pp.10389–10397
[11] Amit Ganatr & Kosta,YP (2010), 'Spiking Back Propagation Multilayer Neural Network Design for Predicting Unpredictable Stock Market Prices with Time Series Analysis', International Journal of Computer Theory and Engineering, vol.2, no.6, pp.963-971
[12] Abhishek Gupta, Samidha & Sharma,D (2014), 'Clustering-Classification Based Prediction of Stock Market Future Prediction', International Journal of Computer Science and Information Technologies, vol. 5 , no.3, pp.2806-2809
[13] Das, S. P., Achary, N.S. and Padhy, S. (2016). Novel hybrid SVM-TLBO forecasting model incorporating dimensionality reduction techniques, Appl Intell, Springer, Vol. 45:1148–1165.
[14] S Baadel, J Lu (2019),Data Analytics: intelligent anti-phishing techniques based on Machine Learning,Journal of Information & Knowledge Management.
[15] A Alsaeedi(2019),EFTSA: Evaluation Framework for Twitter Sentiment Analysis,JSW.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY