

USEFUL DECISION MAKING FOR COMMON STOCK SELECTION BY USING REGRESSION TECHNIQUES

Rama Devi Burri¹, V. B. V. N. Prasad², SaiManvitha Enadula,

¹Department of Information Technology, Institute of Aeronautical Engineering, Dundigal-Telangana-500043.

²Department of Engineering Mathematics, Koneru Lakshmaiah Education Foundation (KLEF), Vaddeswaram, Green fields, Guntur, Andhra Pradesh, India -522302.

³Indian Institute of Information Technology design and manufacturing, Kurnool, A.P., India

E-mail: vbvnprasad@kluniversity.in

E-mail: ramaburri5@gmail.com

DOI : 10.48047/IJFANS/11/ISS4/121

Abstract: Decision making in Prediction of stock market performance is the most complicated thing. There are so many issues to influence the prediction of the stock market; they include the Physical, Physiological and Rational behavior. All these factors combine to make difficult to predict the share price. By means of features like most up-to-date announcements regarding to the organization, their periodical returns. Machine Learning (ML) techniques encompass prospective to come across patterns and insights. The prediction procedure considered as irrespective of these external factors and only considers the internal factors and variables. Price predictions are typically evaluated on the basis of statistical criteria; accurate predictions of stock market help the investors more reliable and motivated towards the business of buying and selling shares. Mathematical methodologies uses interpreter variables to predict the conclusion of a variable by using statistical approaches like regression techniques, we can able to expect the consequences of the stock price. The successful prediction will direct to real life solutions for stock investors.

Keywords: Stock selection, Decision making, Prediction model, Machine Learning, Regression techniques.

Introduction:-

A stock or share market is the combination of both investors and stock sellers takes into concern. Stock Market is one of the most complicated and stylish way to do business, in order to gain the maximum profits by reducing the losses in visualizing the predictions in Stock market. We have 14 different sectors in stock market like IT, Banking, Financial and so on. The value of the stock depends on its face value, overall market value and company capital. In the earlier day's the prediction of share values is more composite as it includes complicated calculations such as finding pivot values total turnovers, which consumes much time. Due to the tremendous development in technology, now it becomes much uncomplicated to calculate and predict the values and consume very less time. By the usage of machine learning algorithms which are optimal in predicting future share values more accurately and helps stock investors in choosing the best one to buy or sell at a particular instance of time to make them more profitable. The stock values are highly impacted by the external factors such as political conditions and decisions taken by the governments etc. However this prediction process is irrespective of those external factors and only considers the internal factors and variables. Accurate predictions of stock market help investors more reliable and motivated towards the business of buying and selling shares.

Methodology: "Machine learning is one of the applications of Artificial Intelligence(AI)" by which we can provide the ability to the system to learn automatically and improve from experience, by using machine learning we are not required to program the system explicitly. Machine learning algorithms are classified into 4 types

they are “Supervised, unsupervised, semi supervised and reinforcement”. Regression is a statistical approach in finding the relationship between variables. We have different types of regression techniques.

Regression: Regression is a collection of mathematical techniques that are designed to approximate the correlation between a independent variable and dependent variable. Regression is generally used in the prediction process of continuous variables or values. We have various types of regression techniques in order to predict stock value; they are linear regression, Multiple Linear Regression, Decision trees and Random forest regressions.

Linear Regression (LR): The Linear regression (LR) is fundamental “Machine learning algorithm”, able to apply on the data. “The linear regression model precedes an equation that determines association between independent and dependent variables”.

“The equation for linear regression can be written as:

$$Y = \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_n X_n$$

Here, x_1, x_2, \dots, x_n represent the independent variables while the coefficients $\theta_1, \theta_2, \dots, \theta_n$ represent the weights”.

Multiple Linear Regression (MLR): MLR mathematical methodology for numerous variables to predict the outcome of a variables. “This is one of the significant regression algorithms, to model the linear relationship connecting a single dependent continuous variable and more than one independent variable”, the extension of linear regression. Main objective behind (MLR) formulation is to design positive correlation among the ‘autonomous and non autonomous variables’. This is based on several assumptions which are described as “The independent variables are not highly correlated and the residuals should be normally distributed with mean ‘0’ and variance σ ”.

“The formula for multiple linear regression is $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon$

Where for $i=n$ observations.

Here y_i is independent variable,

X_i is explanatory variable,

β_0 is a constant term and y- intercept.

β_p is the slope coefficient for each explanatory variable.

ε is Residual is the model’s error term”.

Decision Tree Regression: Decision tree Structure is like to flow chart. It is “supervised non parametric learning technique and used for both **classification** and **regression** tasks”. These are created through an algorithmic advance. This is one of the mainly used methods for “supervised learning”. It decomposes a dataset into subsets and incrementally develops a corresponding decision tree. A decision node comprises two or more branches, each express the values for testing the attributes. Leaf node reflects a mathematical endpoint decision. The highest decision node called root node and predicts the data. Decision trees are capable of handling both statistical and unambiguous data. By exchanging Knowledge Gain with Standard Deviation Reduction can be used to create a decision tree regression.

- “Standard deviation for two attributes: $S(T,X) = \sum P(c)S(c)$ where $c \in X$
- Standard deviation reduction: $SDR(T,X) = S(T) - S(T,X)$ ”.

Random Forest Regression: Random forest Regression has the capability to perform both classification and regression by using multiple decision trees. It is combination of various machine learning algorithms so is called ensemble, random forest regression is ensemble model. “Random forest is Supervised Learning algorithm which uses ensemble learning method for classification and regression”. The numeral features can divide at every node is partial to few proportion of the overall. This ensures that the ensemble model does not rely too heavily on any individual feature, and makes fair use of all potentially predictive features. The quantity of features while to split at each module is biased to some proportion of the sum, and prepares fair use for all potential predictive features.

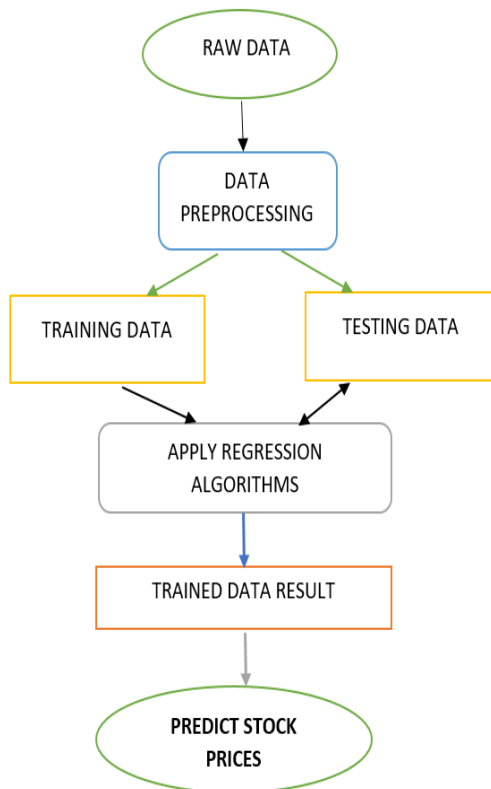
Formula behind the implementation of random forest algorithm implementation is

$$f G(x) = f_0(x) + f_1(x) + f_2(x) + \dots$$

Where G is the sum of sample models f_i , Random forest is the combination of decision trees .and the collection of trees is termed as a forest. The more number of trees, more robust and accurate algorithm we will get.

SYSTEM ARCHITECTURE:

Kaggle is the most part reputed online data analysis communities and is a renowned source of the predictive modeling. It consists of the datasets from dissimilar fields which are the contributions of various data miners. The datasets for this are taken from the kaggle web source. The data set is the group of share values of a specific company.



The raw data is collected from Kaggle Data set, after that pre-processing data set can be split in to training data and testing data. On Both these training and testing data sets regression algorithms are applied. After that trained data results resolve, by the consequences one can has the capability of predicting the Stock price.

Random forest regression algorithm is uses the combining several random decision trees which analyses the data. The main aim of our predictive model is to get accurate stock values with the given data.

Correlation of Open Price to Turnover: Here the attributes are Open price, closing price, Day’s low and high, Last traded value, Date, Total Traded Quantity and Turnover. We have data set which contains 4943 records having the combination of the above mentioned eight attributes. The following figure describes about the raw data. Data Table with Date, “Open Price, Days High price, Low price, Last traded price, Closing Price, Total Traded Quantity, and Turnover”.

1	Date	Open	High	Low	Last	Close	Total Trad	Turnover
2	7/1/2018	223.9	223.9	218.15	218.15	219.15	1042	2.29
3	7/2/2018	53.3	54.05	53.3	53.9	53.9	5200	2.8
4	7/3/2018	52	52.85	51.65	52.75	52.75	8100	4.25
5	7/4/2018	53.5	54.25	53.5	54.2	54.2	12000	6.45

Fig1.Close Price to turnover

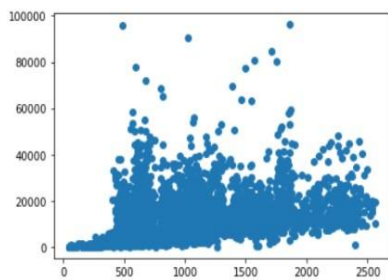
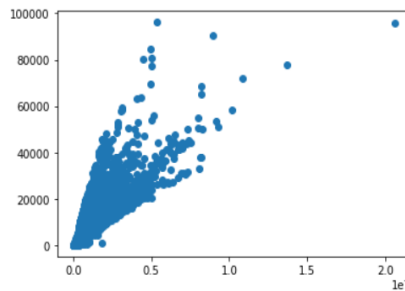
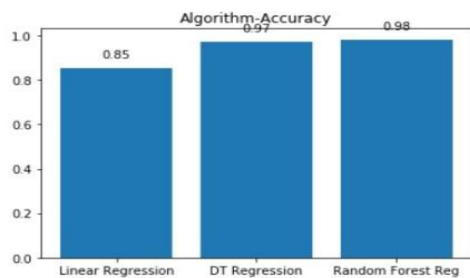


Fig 2. Correlation of Total Trade Quantity to Turnover



Total Traded Quantity is considered as the base attribute with the help of remaining attributes the algorithm is qualified with 80% data set and the trained algorithm is tested with the left over data set. The qualified algorithm is functional dynamic data for the prediction of the Turnover as it predicts the result.

This is done by means of three algorithms they are” Multiple linear regression, Decision tree and Random forest”s, they predict outcome with the accuracy values of 0.854, 0.971 and 0.98. The Random Forest Regression Gives the more accurate value, when compared to the remaining, so Random Forest Regression is the best way algorithm to Decision on the stock value.



Conclusion:

In this paper, we make use of regression algorithms for Decision making for the share value of an Individual company which is to be selected; Regression algorithms are used in carrying out the results of continuous variables. By calculating the accurateness of the various algorithms, Random forest algorithm is the best suitable algorithm for predicting the stock values. The algorithm will be a great asset to all types of brokers, Investors and shareholders as well as being trained with enormous gathering of past data and sample data being tested.

References:

- [1] D. Shubhrata, D. Kaveri, T. Pranit, and S. Bhavana, "Stock Market Prediction and Analysis Using Naïve Bayes," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 4, no.11, pp. 121–124, 2016.
- [2] 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.
- [3] K. Hiba Sadia, "Stock Market Prediction Using Machine Learning Algorithms," *International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-4, April 2019*
- [4] Rama Devi Burri, Ram Burri, Ramesh ReddyBojja, Srinivasarao Buraga "Insurance claim Analysis using Machine learning Algorithms," *International journal of innovative technology and Exploring Engineering (IJITEE), Volume-8, Issue-6S4, April-2019, ISSN: 22278-3075.*
- [5]. Ms.sonalpatil, Mr.Mayur Agrawal, Ms.Vijaya R. Baviskar "Efficient Processing of Decision Tree using ID3 & improved C4.5 Algorithm", *International Journal of Computer Science and Information Technologies, Vol. 6 (2) , 2015, 1956-1961.*
- [6] Dr.B.Rama Devi, Yashoda muralikrishna, Raviteja Reddy, ChittiBabu "Share market Prediction using Machine Learning Algorithms" *Test Engineering and Management Volume-83 , Issue-March-April 2020, April 2020, ISSN: 0193-4120.*
- [7] Rama Devi Burri, Y.Venkata Raghava Rao, V.B.V.N. Prasad "Machine learning methods for software Defect Prediction a Revisit" *International journal of innovative technology and Exploring Engineering (IJITEE), Volume-8, Issue-8, June-2019, ISSN: 2278-3075.*
- [8] Rashid, T., & Abbas, S. (2017). Low-power CMOS amplifier for biomedical signal processing: A review. *IEEE Access*, 5, 2697-2717.
- [9] Khan, M., & Rahman, S. (2019). A low-power CMOS OTA for ECG signal amplification. In *Proceedings of the IEEE International Conference on Microelectronics (ICM)* (pp. 141-144).
- [10] Lim, S., & Kim, H. (2021). A low-power OTA for wearable biosignal monitoring devices. In *Proceedings of the IEEE International Conference on Electronics, Circuits and Systems (ICECS)* (pp. 1-4).
- [11] Patel, A., & Desai, U. (2018). Low-power OTA design for portable EEG amplifiers. In *Proceedings of the IEEE International Conference on Innovations in Electronics, Signal Processing and Communication (IESC)* (pp. 1-5).

