Prediction of Stock Market using Machine Learning Algorithms

Neha Bhardwaj[1], MD Akil Ansari[1]

[1]School of Computer Science and Engineering, VIT University, Vellore, Tamil Nadu, India.

---***---

Abstract - *In modern financial market, the most crucial problem is to find essential approach to outline and visualizing the predictions in stock-markets to be made by individuals in order to attain maximum profit by investments. The stock market is a transformative, non-straight dynamical and complex system. Long term investment is one of the major investment decisions. Though, evaluating shares and calculating elementary values for companies for long term investment is difficult. In this paper we are going to present comparison of machine learning aided algorithms to evaluate the stock prices in the future to analyze market behaviour. Our method is able to correctly analyze supervised algorithms and compare which algorithm performs the best to predict the future stock market prices in the market.*

Keywords: Stock market, machine learning, Supervised learning algorithms, Random forest, Logistic regression, K-NN, ARIMA.

1. INTRODUCTION

Exchanging the stocks on money markets is one of the significant speculation exercises. Already, scientists developed different stock examination system that could empower them to envision the bearings of stock esteem development. Predicting and foreseeing of significant worth future cost, in perspective of the present cash related information and news, is of colossal use to the financial pros. Financial masters need to know whether some stock will get higher or lower over particular time-period. To obtain the accurate output, the approach used is to implemented is machine learning along with supervised learning algorithms. Results are tested using different types of supervised learning algorithms with a different set of a features.

The prologue to algorithms in exchanging unquestionably changed the share trading system. Calculations made it simple to respond quick to specific occasions on the share trading system. Machine learning algorithms additionally empowered examiners to make models at anticipating costs of stocks significantly simpler. Implementation of machine learning caused that new models can be created in light of the past information. In this paper we will analyse the method for predicting stock market direction using several machine learning algorithms.

The paper is composed in mentioned accompanying ways: In Section 2, we will be reviewing the literature survey of several papers done in past few years for this topic . In Section 3, we will describe problem statement, methodology and algorithms we are using for prediction . Section 4 describes the results of our methodology on datasets of five different company . In the last section we will conclude the paper with its results.

2. LITERATURE SURVEY:

1) M.Suresh Babu et al., 2012, this paper investigates the significant clustering calculations: K-Means, Hierarchical grouping calculation and turn around K-Means and look at the execution of these noteworthy clustering calculations on part of effectively class savvy group building capacity of calculation. The proposed strategy comprises of three stages. To start with, they change over each money related report into an element vector and utilize the various levelled agglomerative grouping strategy to isolate the changed over element vectors into bunches. They consider both subjective and quantitative highlights in monetary reports. Second, they join the upsides of two grouping techniques to propose a compelling clustering strategy. Third, picking a fitting number of parts in HAC can limit the bunches produced and in this way enhance the nature of the grouping created by the K-means clustering.

2) Mahajan Shubhrata D et al., 2016, this paper is to anticipate future stock value utilizing forecast idea. In that Parse Records at that point figure anticipated esteem and send to client. Also, consequently perform activities like buy and deal shares utilizing Automation idea. For that utilization Naïve Bayes Algorithm. There is Real time Access by Download log shapes hurray back site and Store in dataset. The investigations uncover a high capability of Naïve Bayes Algorithm in foreseeing the arrival on interest in the offer market.

3) Xiao Ding et al., 2015, recommended that a deep learning technique for occasion driven securities exchange expectation. Initially, occasions are removed from news message, and spoke to as thick vectors, prepared utilizing a

novel neural tensor system. Second, a deep convolutional neural system is utilized to demonstrate both here and now and long haul impacts of occasions on stock value developments. They exhibited that deep learning is helpful for occasion driven stock value development forecast by proposing a novel neural tensor system for learning occasion embeddings, and utilizing a deep convolutional neural system to demonstrate the joined impact of long haul occasions and here and now occasions on stock value developments.

4) Adebiyi Ayodele et al., 2012, proposed investigation work to enhance the exactness of every day stock value forecast of securities exchange records utilizing artificial neural networks. The examination utilized three-layer, multilayer perceptron models (a feedforward neural system demonstrate) prepared with backpropagation calculation. This paper displays that hybridized approach can possibly upgrade the nature of basic leadership of financial specialists in money markets by offering more precise stock forecast contrasted with existing specialized investigation based approach.

5) Ayodele A. Adebiyi et al., 2012 , presents comprehensive |technique of building stock price predictive model using the ARIMA model. Published stock data obtained from New You are able to Stock Exchange (NYSE) and Nigeria Stock Exchange (NSE) {are being used} with stock price predictive model developed. Outcomes or Benefits obtained revealed that the ARIMA model has a strong potential for immediate prediction and can remain competitive favorably with existing {processes tactics for stock price prediction.

6) Peihao Li, Chaoqun Jing, et al., 2015 , through this paper they proposed a modelling process and present the estimate SSE (Shanghai Stock Exchange) Composite Index to see the model's estimation performance, which testifies to be feasible and effective. The forecast condition is essentially a direct condition that alludes to past estimations of unique time series and past estimations of the blunders. As should be obvious in the last expectation, the pattern can be accurately anticipated by the model implying that both the model and free factors are accurately chosen.

7) C.Narendra Babu and B.Eswara Reddy et al., 2014, proposed a direct half and half model utilizing ARIMA and GRACH is created which protects the information pattern and renders great forecast exactness. Appropriately, the given TSD is deteriorated into two distinctive arrangement utilizing a basic moving normal (MA) filter. One of them is displayed utilizing ARIMA and the other is demonstrated utilizing GARCH relevantly. The forecasts got from both the models are then melded to acquire the final demonstrate expectations. Indian Stock market information is considered keeping in mind the end goal to assess the exactness of the proposed display. The execution of this model is contrasted and conventional models, which uncovers that for multi-venture ahead expectation, the proposed display beats the others as far as both forecast precision and safeguarding information drift.

8) Abdulsalam Sulaiman Olaniyi et al., 2011, proposed in this paper, the serial development of stock costs over some stretch of time extricated from the everyday official rundown of Stock Exchange, are utilized as a part of building a database and estimations of factors were separated from the database to anticipate the future estimations of different factors using time arrangement information that utilized moving normal technique. They exhibited regression investigation as an information mining strategy and created instrument for abusing particularly time arrangement information in money related organization. An expectation framework has been constructed that utilizations information mining method to create intermittently estimates about securities exchange costs and ready to utilize regression investigation as an information mining system to portray the patterns of stock exchange costs and foresee the future securities exchange costs.

9) Jatinder N.D. Gupta et al., 2000, gives a review to the tasks research reader of the fundamental neural network methods, and in addition their verifiable and ebb and flow use in business. Neural networks and information mining are not enchantment answers for issues, in spite of the message indicated by merchants of programming items. They explored neural network strategies in business from the viewpoint of the tasks researcher. The three fundamental neural network ways to deal with taking care of business issues have been presented: multi-layered feedforward neural networks, Hopfield neural networks, and self-sorting out neural networks. Every one of these methods discovers regular similarity with more customary factual and tasks research strategies, and these analogies have been talked about.

10) Dinesh Bhuria et al., 2017, surveyed about stock market prediction using regression techniques and proposed productive regression way to deal with foresee the stock market cost from stock market information based. In future the consequences of multiple regression approach could be enhanced utilizing more number of factors. This examination ponder is to help the stock merchants and speculators for putting cash in the stock market. The expectation

plays an imperative part in stock market business which is exceptionally confused and testing process due to dynamic nature of the stock market.

11) Md Jan Nordin et al., 2012 , looks at the hypothesis and routine with regards to regression methods for expectation of stock value slant by utilizing a changed informational collection in ordinal information format. In this examination, all information in numerical esteems are changed into ordinal or counted qualities to frame the dataset. Regression based classifiers from WEKA are then utilized as prescient examination to test the ordinal information. The results were thought about and assessed.

12) Bhagyashree Nigade et al., 2017 , surveyed that the stock market does not have an effective programming where the best possible proposals of accessible stocks and the correct speculation investigation are displayed in an effective way. The paper clarifies the advancement and execution of a stock value prediction application utilizing machine learning algorithm and protest situated approach of programming framework improvement. The algorithm was utilized as a part of preparing an arrangement of market information gathered for the time of one thousand, two hundred and three days. And a prediction framework has been manufactured that utilizations information mining procedure to deliver intermittently gauges about stock market costs. The utilization of back engendering in neural system empowers us to decrease blunders and enhance exactness of the framework.

13) Hemangi Karchalkar et al., 2017, explained a stock value prediction technique in this undertaking and for this reason regression algorithm and question arranged approach of programming advancement is used. The winning strategies demonstrate a pattern on future development of stocks and not the conceivable cost for any stock later on. It is in this manner desirable over have an instrument that does point a bearing towards value development, as well as demonstrates the doubtlessly cost of the stock itself

14) Luckyson Khaidem et al., 2016, proposed a novel method to limit the danger of interest in stock advertise by foreseeing the profits of a stock utilizing a class of intense machine learning calculations known as ensemble learning. They have utilized four administered learning calculations, i.e "Logistic Regression, Gaussian Discriminant Analysis, Quadratic Discriminant Analysis, and SVM".

15) T. Manojlović et al., 2015, utilized the 'Random-forest' calculation to construct the model used to anticipate 5-days-ahead and 10-daysahead bearings of the CROBEX record andchose stocks. Their outcomes demonstrate that random forests can be effectively used for building predictive models for anticipating the course of securities exchange patterns.

16) Mohammad Bolandraftar et al., 2014, attempted to create three models and looked at their exhibitions in anticipating the heading of development in every day Tehran Stock Exchange (TSE) file. The models depend on three order procedures, Decision Tree, Random Forest and Naïve Bayesian Classifier. What's more, reasoned that principal investigation assumes less vital part than specialized examination during the time spent basic leadership of brokers and partners.

17) Khalid Alkhatib et al., 2013, applied "K-nearest Neighbour" algorithm and indirect relapse approach so as to anticipate stock costs for an example of six noteworthy organizations recorded on the Jordanian stock exchange to help financial specialists, administration, chiefs, and clients in making right and educated ventures choices. As indicated by the outcomes, the k-NN algorithm is vigorous with little mistake proportion; subsequently the outcomes were sound and furthermore sensible.

18) San- hing Liul et al., 2010, build up a fundamental anticipating model in view of KNN and BP Neural Network. They attempted it in anticipating the stock cost of China stock and test comes about demonstrate that the normal blunders happening in KNN-ANN algorithm are littler than those in KNN algorithm, showing that the anticipating model in light of KNN-ANN algorithm can do better in the stock forecast.

19) Tian Ye et al., 2017 , showed a stock estimating model in light of wavelet examination and ARIMA-SVR. The stock cost is disintegrated into remade part and blunder part by wavelet deterioration and wavelet remaking. At that point, the ARIMA show and the SVR display are utilized to gauge the recreated part and the blunder part separately, and the last forecast outcomes are joined to get the last expectation comes about. The day by day shutting cost of Shanghai Pudong Development Bank from January 5, 2015 to January 29, 2016, was chosen as the trial information, with a sum of 250, of which the initial 220 were the preparation set and the last 30 were the test set to do recreation try. The

test comes about demonstrate that contrasted and the single determining model, the proposed display is a powerful technique for anticipating stock value, which extraordinarily enhances the precision of estimating.

20) Robert P. Schumaker et al., 2008, discovered that two speculations have significantly affected statistical surveying: Efficient Market Hypothesis (EMH) and Random Walk Theory. In EMH, the cost of a security is an impression of finish showcase data. At whatever point a change in money related viewpoint happens, the market will in a flash alter the security cost to mirror the new data (Fama, 1964). EMH contained three unique levels of data sharing: the powerless frame, the semi-solid and the solid shape. Inside powerless EMH, just recorded information is implanted inside the present cost. The semi- solid frame goes more remote by joining chronicled and current open data into its costs. The solid shape incorporates verifiable and current open data and additionally private data. From these three structures, it was trusted that business sectors acted proficiently and that immediate value amendments would forestall expectation models.

3. PROBLEM STATEMENT:

In order to predict the stock prices in future markets, we have analyzed papers and has given an overview on how these algorithms give precise and accurate future predictions. In this paper, we used several algorithms from which we observed that not all the algorithms implemented can predict data we need. There has been a basic requirement for computerized and automized ways to deal with powerful and proficient usage of huge measure of money related information to help organizations and people in vital arranging and decision making on investments.

4. METHODOLOGY:

We are using R statistical programming language to study and perform this experiment. -R is an extremely flexible statistics programming language that is Open Source and unreservedly accessible for all standard working frameworks. R has as of late encountered a ""explosive growth in use and in user contributed software". The "user-contributed software" is one of the most exceptional and gainful parts of R, as a huge number of clients have contributed code for actualizing probably the most avant-garde measurable strategies, notwithstanding R executing basically all standard measurable investigations. As a result of R's Open Source structure and a group of clients devoted to making R of the most noteworthy quality, the PC code on which the techniques are based is straight forwardly scrutinized and improved.

By using R and implementing following Machine learning algorithms on the datasets we are predicting the stock price movement:

LOGISTICS-REGRESSION

In statistical studies, "Logistic-regression", is a regression exhibit where the dependent variable (DV) is straight out. Logistic-regression uses the trademark logarithm ability to find the association between variables and utilizes test data to find the coefficients. The limit would then have the capacity to suspect the future results using these coefficients in the logistic condition.

$$P = \frac{e^{a+bX}}{1+e^{a+bX}}$$

This last condition is the logistic curve for Logistic-regression. It models the non-straight relationship amongst x and y with a 'S'- like curve for the probabilities that y =1 - that occasion the y occurs. Logistic-Regression uses the logistic function to locate a model that fits with the information points. The function shows S shaped curve to demonstrate the information. The curve is restricted in the vicinity of 0 and 1, so it is easy to apply when y is binary.

To predict stock values example of normal cost of the Ith month, such stock cost is considered to be down trended if the Logistic-Regression esteem is I, p is close to 0 (or is proportional to 0). Sometimes, such stock cost is considered to have uptrend, if the Logistic Regression esteem I, p is almost 1(or is equal to 1). What's more, the more far from 0 I and, p must have the lesser probability the stock costs have downtrend, and the different way around. Normal logarithm tasks are performed on the condition obtained through the probability capacity of joint thickness work with n tests.
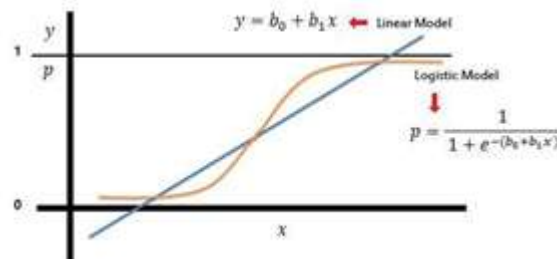
Fig-1: Logistic Regression Graph

RANDOM-FOREST

Random decision forest is an ensemble learning methodology towards classification calculations, backslide techniques and diverse endeavors, that work by building up countless trees during training time and yielding the class which is the strategy for the classes (classification) or mean figure (backslide) of each individual tree. Random decision forests redress for decision trees' inclination for overfitting to the training dataset.

The training calculation for "random-forests" applies the general procedure and tactics of bootstrap collecting, or stowing, to tree learning sets. Given, a training dataset $X = x1, ..., xn$ with reactions $Y = y1, ..., yn$, packing more than once (B times) chooses a random example with substitution of the training set and fits trees to these examples:

For, b = 1, ...B:

1.   Example, with substitution, n preparing cases from X, Y; calling these Xb, Yb..

2.   To provide training to a classification or regression-tree fb on Xb, Yb.

In the wake of preparing, forecasts for inconspicuous examples x' can be made by averaging the prospect from all the particularized regression trees on x':

"f= $\sum$ fb (x')/B"

and then again by taking larger part of vote by virtue of grouping trees. The amount of tests/tree B, is a unbound parameter. Ordinarily, two or three hundred to a couple of thousands of trees are used, dependent upon the nature and size of the preparing dataset. An ideal number of trees B can be discovered utilizing cross-approval, or by watching the "out-of-bag error"(OODB).

The plot in the above stated algorithm depicts the least number of trees which can stabilize the forest so generated. As the number of trees in the graph so obtained increases and stabilizes the accuracy and the precision of the graph increases, depicting and verifying that on the further increase in the tree number is not affecting the precision and accuracy in any of the so authorized context. The VarImpPlot graphical representation the random-forest methodology indicates and signifies the importance and the strategy to prioritize the features and attributes associated in a particular dataset. This is plotted against variable and its importance. This helps in analyzing which variable is of high importance in the data analyses and will give higher precision and accuracy for the so given dataset to be analysed.

K-NN ALGORITHM

In design acknowledgment, the "K-nearest Neighbors" count (K-NN) is the non-parametric procedure used for course of action and regression. In some couple of cases, the data includes the k nearest getting-ready cases in the part space. The yields depends upon whether K-NN can be used for regression or characterization

"R*=< RkNN =< (2-MR*/M-1) *R"

where $R*$ is "Bayes-error rate" (which is the insignificant error rate conceivable), and M is the number of classes and RkNN is the K-NN error rate in the problem. For M=2, "Bayesian-error rate" $R*$ approaches to zero, this confine reduces to "not more than double the Bayesian-error rate". A blueprint of conditions that was associated in this article for anticipating following day cost. The figuring consolidates error-evaluation, signify total of squared error, normal error, add up to closing cost when orchestrated using expected regards, k-characteristics and planning "RMS error".

a) Root-Mean-Square (RMSD) Deviation is precision metric which figures the differences between the assessed regards, Y, and honest to goodness regards, X. The total of RMSD is gisted into a singular regard measure.

b)   "Explained Sum of Squares" is handled as taken after.

c) Average-Estimated-Error Rate:

AEE is total whole of RMS errors for every variable in stock document isolated by the total number of the documents. The lift chart symbolizes the change that a data-mining- representation gives when perceived against a sporadic estimation, and the

change is imparted with respect to lift-score. By separating the lifting-score, for a combination of part of educational file and different models, it would then have the capacity to be picked which indicate is exceptional and which level of the cases inside the enlightening list would get from using the desires show.

ARIMA

'ARIMA' is extracted as Autoregressive-Integrated-Moving-Average. ARIMA is additionally referred to as "Box-Jenkins approach". Box and Jenkins asserted that non-stationary information can be made to stationary information by differencing the arrangement, Yt. The model for Yt can be composed as,

"Yt =ɸ1Yt−1 +ɸ2Yt−2…ɸpYt−p +ϵt + θ1ϵt−1+ θ2ϵt−2 +…θqϵt−q"

The ARIMA methodology joins three fundamental methods: Auto-Regression(AR): Firstly, auto-relapse the estimations of a given time arrangement information are relapsed alone slacked esteems, which is shown by the "p" esteem in the representation. Differencing, (I is for Integration) - This includes differencing the time arrangement information to dispose of the pattern and change over a non-stationary time arrangement to a standing one. This is specified by the "d" esteem in the model. Moving-Average(MA): The moving average feature of the model is termed by the "q" where q, is number of lagged-principle of the error-rate.

5. DATA AGGREGATION:

We obtain the historical data and extensible time series data from the Yahoo finance. The datasets collected are of five different – Apple Inc., Tata Consultancy service, Infosys, Verizon and Johnsons & Johnsons. The whole dataset covers daily activity summary from the period of February, 2015 to February, 2019. The dataset mainly consist of attributes- Date, Open, High, Low , Close, Adj.close, Volume.

Pre-processing:

i) Creating a Class-label by the name Status which has binary set of values depending on the Opening and closing price of the stock, if the value of Open price of the stock on a particular day is lesser than the Close then the value in the class-label will be "1" that denotes the profit and if the value of the Open is greater than the Close then the value of class-label will be "0" that denotes the Loss in the market on the following date.

"Status" =1, if change >0.5

0   , otherwise

ii) Data Preparation is done because the feature values in the dataset are categorical or factor, not numerical data so, this creates problem. The machine learning algorithms used in this project, requires numerical data. Normalization of values is also done for some algorithm to avoid biases from anomalies in dataset and adjustment of values to a common scale.

iii) Data Splitting is done as a total of 1048 pairs of observations are taken. Then dataset is separated into two segments. The initial segment (545 sets of perceptions) is utilized to decide the specialization of the parameters and models. The second section (503 sets of perceptions) is held for out-of-test assessment and examination of exhibitions among different determining models.

iv) Features election is done for selecting the right attributes to be used to build the predicting model that has the most significance on the class label. This step helped to select the best fit model that can predict the result more precisely.

v) Cross-validation is done after fitting the model to check the error rate in prediction

Table I: Table of Comparision

Sr. No.	Algorithm Used	Dataset Used	% of Accuracy	Advantages	Limitations
1.	Random Forest (Ensemble Learning)	Apple Inc (APPL)	50.7%	Pretty Flexible and easy to train.	Require more no of trees to predict accurately that makes model slow.
		Johnsons & Johnsons (J&J)	80.12%		
		TATA Consultancy service Ltd	53.48%		
		Verizon (VZ)	77.3%		

		Infosys Ltd (INFY)	76%		
2.	Logistic Regression (non-Ensemble Learning)	Apple Inc (APPL)	76%	Provides very accurate results.	Selecting right Features to Fit.
		TATA Consultancy service Ltd	66%		
		Johnsons & Johnsons (J&J)	75%		
		Verizon (VZ)	71%		
		Infosys Ltd (INFY)	72%		
3.	ARIMA (Supervised Algorithm)	Apple Inc (APPL)	52%	Visual plots are effective	Provides less accurate results
		TATA Consultancy service Ltd	45%		
		Johnsons & Johnsons (J&J)	54%		
		Verizon (VZ)	55%		
		Infosys Ltd (INFY)	31%		
4.	K-Nearest Neighbour	Apple Inc (APPL)	74%	Very Simple and Flexible.	Issue of normalization of dataset and cant handles categorical variable well.
		TATA Consultancy service Ltd	64%		
		Johnsons & Johnsons (J&J)	73%		
		Verizon (VZ)	53%		
		Infosys Ltd (INFY)	52%		

From the above table a few things turn out to be promptly recognizable. First is, that an assortment of systems have been utilized, Secondly, on which datasets those algorithms are connected with their exactness and error rate. Advantages and limitations of algorithms are also mentioned which are faced while performing the experiment. After analyzing the above table, Logistic Regression Algorithm performs the best with the prediction with an average accuracy of 74%, second is Random Forest Algorithm with the average accuracy of 70%, after this K-NN algorithm performs good which has average accuracy rate of 64%, At last ARIMA model performs with 47% in forecasting the stock market directions.
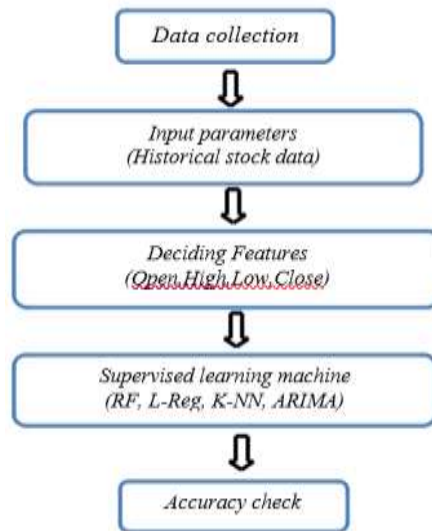
Fig-2: Flow Chart of Algorithm

6. RESULTS

After successfully implementing algorithms on the datasets, the results were interesting as shown in the comparison table. By selecting the right attributes with the minimum of standard error gives the fitted prediction model ability to predict the movement of the stock somehow correctly. To visualize the results firstly, the plotting of open vs close attribute is visualized using ggplot() plotted directly based on the apple dataset. It describes that the Open attribute is directly proportional to the close attribute.
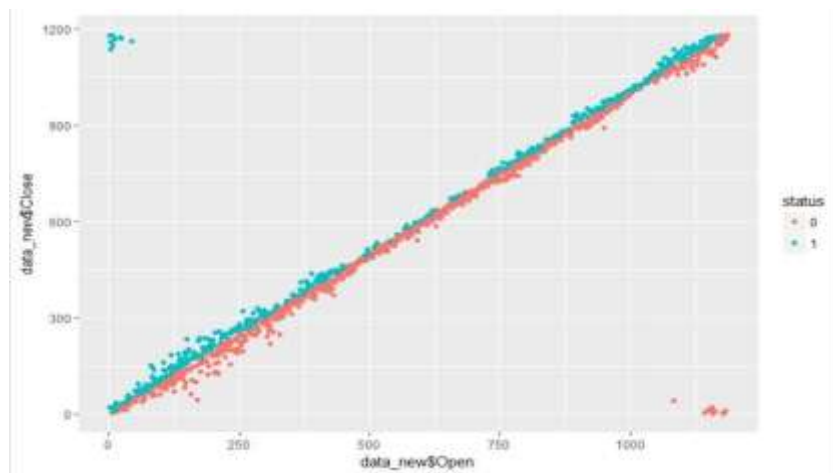
Fig-3: Plot for Open versus close attribute

Regression model with the least standard error and AIC value was used after checking many sample models which were not a good fit for model. The attributes that are used in fitting the regression on model was having a fine level of significance and likelihood with the class label. For each algorithm, we have different plots generated for a particular dataset AAPL (Apple inc.), for Logistic Regression four plots are generated, first plot shows the plot between Residuals and Fitted, second is plot between Leverage and Standard Pearson Residual, third plot shows Normal Q-Q and last is Scale-Location.
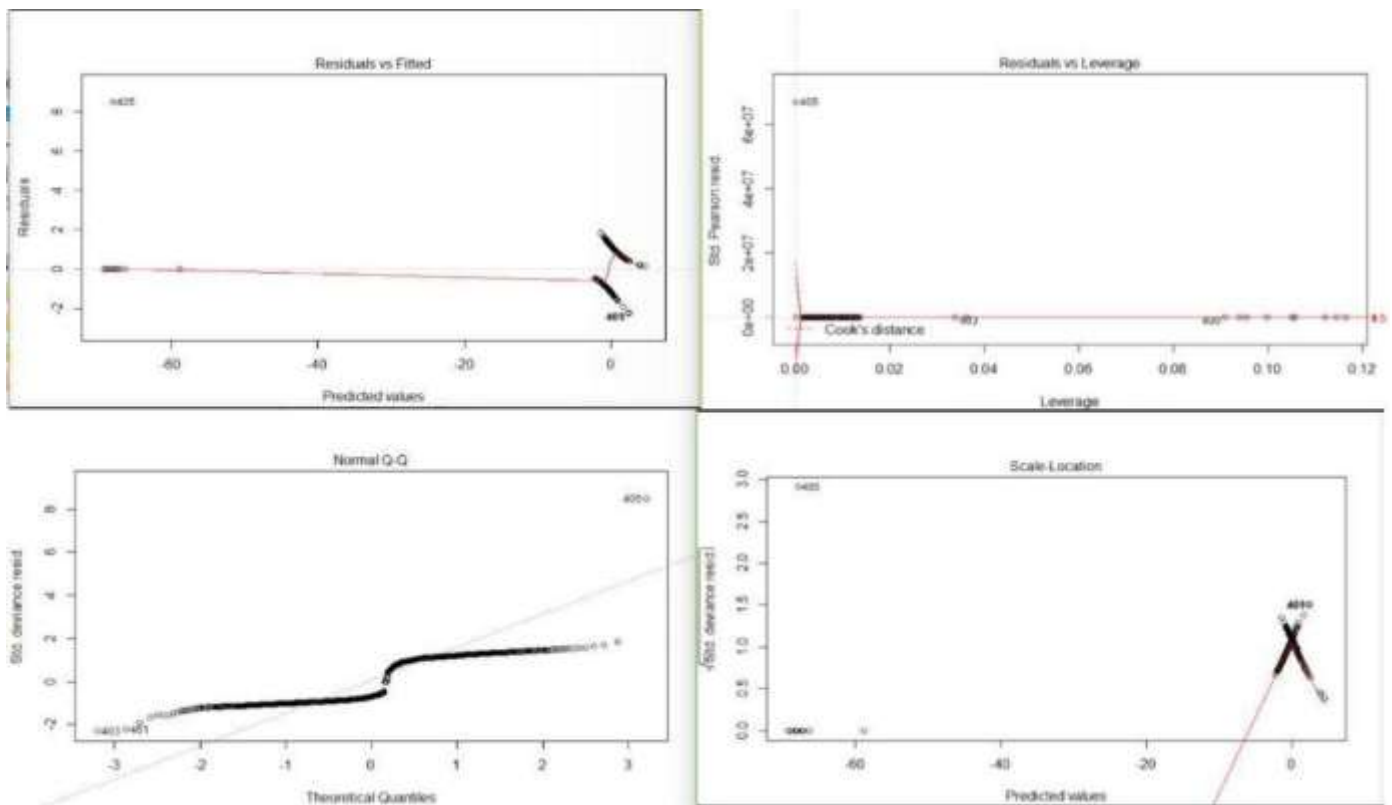
Fig-4: Logistic Regression plot for residuals, Normal Q-Q, Leverage and Standard Deviations

When implemented with ARIMA Algorithm, 2 plots are generated ACF and PACF plot. The persistence of high values in acf plot probably represent a long term positive trend. ACF plot shows the Autocorrelation function and PACF shows the Partial autocorrelation function. If the sign of the decided return measures up to the sign of the certifiable returns we will allot it in positive precision score. The accuracy level of the model comes approximately, 40% that looks like a respectable number. We can try running the model for other conceivable blends of (p,q,d) or rather use the auto. The arima-algorithm picks the best perfect parameters to run the model. The ARIMA show and associated itself for deciding stock esteem returning using R programming lingo. We furthermore cross-checked the foreseen results with the genuine returning.

Random Forest algorithm generates a plot that shows the minimum number of trees that affects the prediction and the stabilization of the number of trees after a certain number of trees. For assessment of utilized predictive-modelling representation, F measure and grouping accuracy were utilized. Grouping accuracy is computed by separating the number of effectively arranged occasions with the generally number of cases. Proposed model was cross validated multiple times for reducing error and enhancing the accuracy.

Random Forest algorithm generates a plot that shows the minimum number of trees that affects the prediction and the stabilization of the number of trees after a certain number of trees. For assessment of utilized predictive-modelling representation, F measure and grouping accuracy were utilized. Grouping accuracy is computed by separating the number of effectively arranged occasions with the generally number of cases. Proposed model was cross validated multiple times for reducing error and enhancing the accuracy.
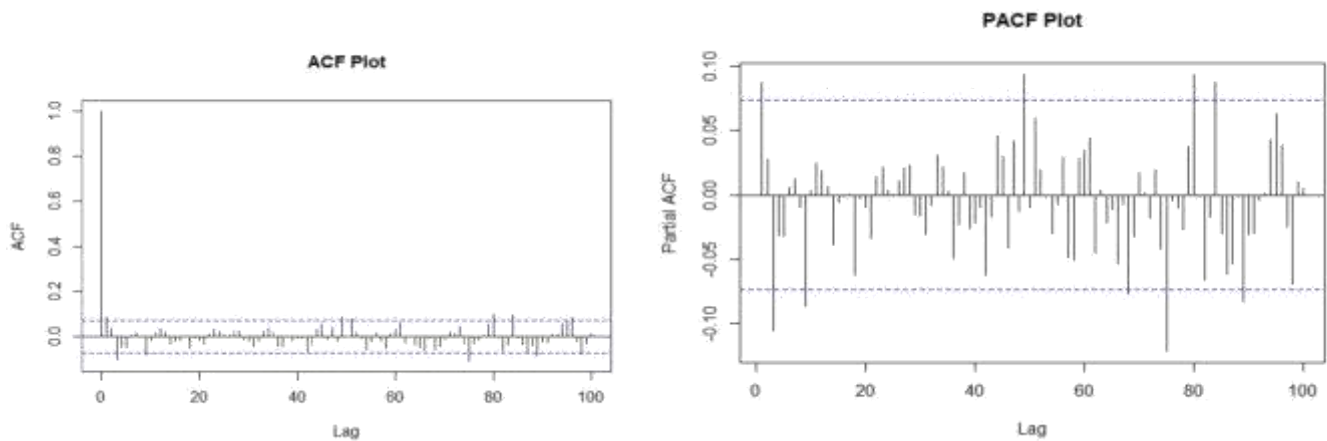
Fig-5: Plot for ACF vs LAG and Partial ACF vs LAG

K-NN algorithm gives the predictions very weakly, but better than Arima. In KNN a graph is generated about which plots have the nearest neighbor shows the in-depth for Open and Close attributes.
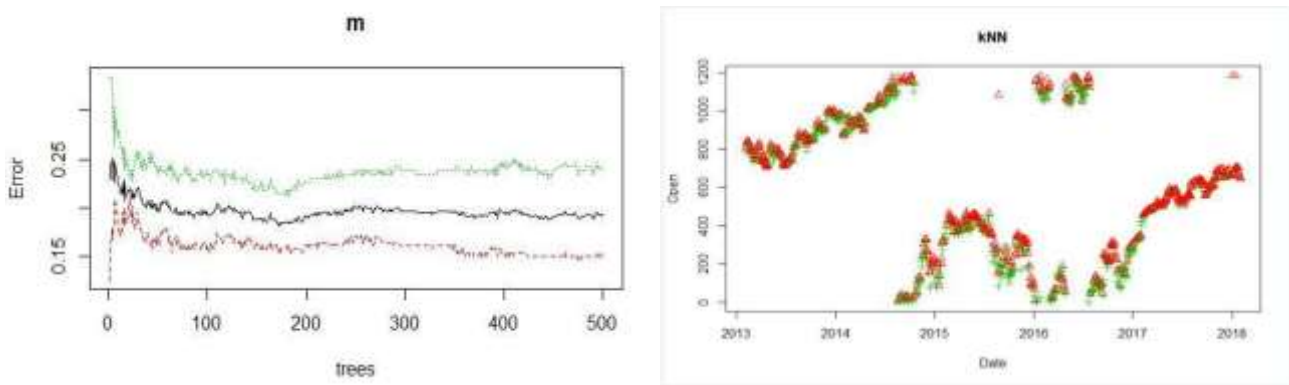

Fig-6: Random forest plot for model (trees vs error) and KNN- Plot for nearest neighbors

Combining all the accuracies, we plot a bar-graph showing the mean accuracies of all the machine-learning aided algorithms:
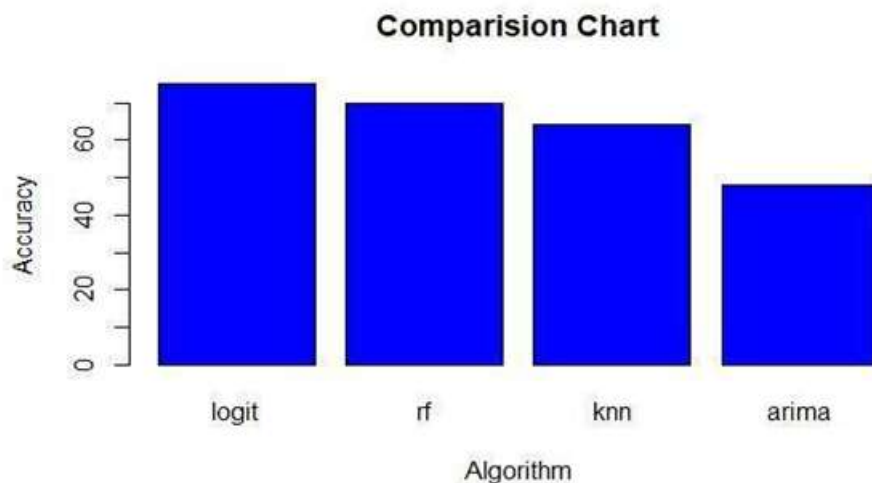

Fig-7: Comparison chart for the algorithm used

The bar plot shows that the Logistic Regression with 74% accuracy performs the best when compared to other techniques which are implemented. Then comes Random Forest algorithm with 70.1% accuracy, then we have KNN 64% with and the last comes ARIMA in terms of Accuracy.

7. CONCLUSIONS

In the above paper, we inspected advancements in economic market predictions. By looking at different predictive models, we discovered that Logistic-regression is offering the capacity to predict and analyzing market movement direction more precisely than the other existing methods. Different models, for example, Random Forest and ARIMA have additionally turned out to be well known in stock market prediction. Random Forest demonstrated its fruitful application in classification work, ARIMA on time series prediction and financial related applications. K-NN model is also applied in the experiment which also show some good results in predicting stock market directions.

This shows that it can be conceivable to utilize Logistic-Regression as an effective approach to successfully predict the future pattern of stock cost. Highlight Index Variables are creative and significant to the approach so proposed. They are just not simple to both comprehend for the private speculator and acquire from every day stock exchanging data, yet in addition improve our approach perform than or equivalent to different strategies, e.g. Random- Forest and ARIMA Prediction Model under similar states of Featured-Index Variables.

This method of choosing enhancing gathering of regression-coefficient enhances predicting accuracy to 74% which is quite high. The approach so proposed is very low in unpredictability and straight forward or figure it out. Because of utilization current month to month exchanging data, noteworthy time-viability and soli deliberateness empower clients foresee stock value pattern of straightaway month simply through considering current month to month monetary information as opposed to requiring a long haul system of examining furthermore, gathering budgetary information. In any case, one deformity still exists in our approach looking at that as some component file factors come up short on account of little esteem (surmised to zero) of this parameter. This is critical to assist examinations of Logistic-Regression representation for stocks in value incline. Also, there exists the circumstance that element file factors flop in some case, we will think about how to choose compelling factors to enhance prediction accuracy of stock value incline.

REFERENCES:

[1] C. N. Babu and B. E. Reddy, "Selected Indian stock predictions using a hybrid ARIMA-GARCH model," 2014 Int. Conf. Adv. Electron. Comput. Commun. ICAECC 2014, 2015.

[2] M. 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.

[3] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep Learning for Event-Driven Stock Prediction," no. Ijcai, pp. 2327–2333, 2015.

[4] K. A. Smith and J. N. D. Gupta, "Neural networks in business: Techniques and applications for the operations researcher," Comput. Oper. Res., vol. 27, no. 11–12, pp.1023–1044, 2000.

[5] M. S. Babu, N. Geethanjali, and P. B. Satyanarayana, "Clustering Approach to Stock Market Prediction," vol. 1291, pp. 1281–1291, 2012.

[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] A. Sharma, D. Bhuriya, and U. Singh, "Survey of Stock Market Prediction Using Machine Learning Approach," pp. 506–509, 2017.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] T. Ye, "Stock forecasting method based on wavelet analysis and ARIMA-SVR model," Inf. Manag. (ICIM), 2017 3rd Int., pp. 102–106, 2017.

[18] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock Price Prediction Using the ARIMA Model," 2014 UKSim-AMSS 16th Int. Conf. Comput. Model. Simul., no. March, pp. 106–112, 2014.

[19] P. Li, C. Jing, T. Liang, M. Liu, Z. Chen, and L. Guo, "Autoregressive moving average modeling in the financial sector," ICITACEE 2015 - 2nd Int. Conf. Inf. Technol. Comput. Electr. Eng. Green Technol. Strength. Inf. Technol. Electr. Comput. Eng. Implementation, Proc., no. 4, pp. 68–71, 2016.

[20] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news," ACM Trans. Inf. Syst., vol. 27, no. 2, pp. 1–19, 2009.