Research Journal of Engineering Technology and Management (ISSN: 2582-0028), Volume 02, Issue 01, March-2019
Available at www.rjetm.in/

Stock Market Analysis Using Classification Algorithm

Vaishnavi Kulkarni [1], Nikhil Prakash[2], Komal Kumari[3], and Chirag Harinkhede[4]
[1,2,3,4]Department of Computer Engineering , SPPU, MIT AOE, Pune, India

[1]vaiiishnaviii@gmail.com, [2]nikhilprakash647@gmail.com, [3]kk200215@gmail.com,[4]chirag18.h@gmail.com

Corresponding Author: Vaishnavi Kulkarni

Manuscript Received: Manuscript Accepted:

Abstract

Abstract-Stock market analysis is popular and important in financial studies. Share market is an untidy area for predicting stock prices, as there are no specific rules to estimate the prices of shares. Generally fundamental and technical analysis is used to analyze shares, but none of the above mentioned methods are proved as consistently acceptable prediction tool. One of the application of machine learning algorithms are to analyze stock market data. It is a recent trend in research. The project will demonstrate how different classification and regression algorithms can be used to determine the particular sectors of the stock in trade and the accuracy of the data compared. Machine Learning techniques are used to analyze whether the price of the stock in the coming future will be higher than its price on a given day based on historical data while providing an in-depth understanding of the used algorithms which will help the customers to compare various stocks of their interests and take their own decisions on the daily basis..

Keywords: *Machine learning, classification, Stock Market, pre- diction.*

I. INTRODUCTION (HEADING 1)

Funding or investment is the technology and art of growing money with the aid of placing money to work. The stock marketplace has been a center of attraction for the investors as well as funders for an extended period of time. It historically provided the very best returns of any financial asset which was near 10 percent over a long time. In the stock market, its miles possible to make multiple returns as well as to lose the precept and pass bankrupt. The major key to fulfillment is to buy and sale the stock at the proper time for the right cost. Numerous studies were conducted to find out the way to predict the future stock price and marketplace instructions. Several strategies from economic evaluation as well as data analysis were carried out. In current years, the classification algorithm turns out to be an extensive technique used for stock forecasting. Unluckily, even results from classification algorithm model were liked in a time period of accuracy; lots of them had been no longer positioned into practice. This trouble occurred due to the inability of category or classification algorithms to explain its reasoning; consequences have been opaque to human interpretation. Similarly, proposed systems in addition to discovered knowledge have been now not on hand to retail investors or traders.

Forecasting of the stock marketplace has been a vital subject matter in one of a kind fields of computational sciences because of its viable monetary profit. Stock marketplace is a place wherein excessive capital is invested and organizations do buy and selling of their shares. Inventory marketplace forecasting poses the task of disproving the green market hypothesis, which states that the marketplace is efficient and can't be expected. Researchers have labored hard to prove the truth that monetary markets are predictable. With the advancement and availability of era, inventory markets at the moment are greater reachable to investors. Various fashions were proposed, each in industry and academia, for stock market prediction ranging from machine learning knowledge of, to data mining, to statistical models.

In this study, time collection forecasting models are constructed to forecast the market analysis index by using the usage of a back propagation classification algorithm (a wildly used technique). Models are built in several works flows a good way to look at a suitable span of time for the inventory market information.

II. LITERATURE SURVEY

A lot of research has been done in the field of Stock Market Analysis.

Sahaj Singh Maini and Govinda.K(ICISS 2017) proposed an approach for prediction of stock market shares using machine learning methods like the Support Vector Machine and the Random Forest model .The context-driven from sources like news articles and the data set from the time line 2000 to 2016 was referred for prediction of Dow Jones Industrial Average Index. The study predicts that the Random Forest model using a 1-gram model for text analysis produced an accuracy of 84.3 percent- age and on using a 2-grammodel produced an accuracy of 86.2 percentage. The linear Support Vector Machine(LSVM) using 1-gram model and 2-gram model for text analysis produced predictions with an accuracy of 82.2 percentage and 84.6percentage,while the nonlinear Support Vector Machine produced predictions with an accuracy of 85.1 percentage for both 1-gram and 2- gram models [1].It was observed from the results that the Random Forest Model outperforms the Support Vector Machine for the given dataset.

Pankaj Kumar and Dr.Anju Bala proposed a model for stock market using machine learning algorithms such as Decision tree, Linear model, Random forest and further their results have been compared using the classification evaluation parameters such as H, AUC, ROC, TPR, FPR, etc. The study predicts that for the binary classification Random forest is the most effective model to predict as it yields the highest accuracy of 54.12 percent whereas decision tree and linear model gives the accuracy of 51.87 percent and 52.83 percent respectively. It also predicts that the study of problem-solving of binary classification data[2] based on the machine learning models gives the best 2 out of models which is experimentally used.

Ching-Hsue Cheng, You-Shyang Chen (ICMLC 2007) proposed a model for predicting RGR of companies which employs Multilayer perceptron, Bayes net, Decision Tree C4.5 [3], and Rough sets techniques..They used the RGR dataset in the Taiwan stock market. The process was based on using revenues, assets, profits, income, and other data as condition attributes to determine the potential future growth of its revenue. The historical results indicate that because of accuracy and understandable tools the rough sets outperform the listing methods.

Lamartine Almeida Teixeira and Adriano Lorena Incio de Oliveira (IEEE 2009) proposed a technique that involves a combination of some well-known tools like, Stop Gain, nearest neighbor classifier, RSI filter and Stop Loss. The results that would be generated by using a buy and hold strategy is compared with the results obtained[4] from above technique. The key performance measure in this comparison was profitability. For most of the stock data, the above method generates considerably higher profits than buy and hold.

Radu Iacomin (ICSTCC 2016) proposed a new algorithm on predicting the stock markets. PCASVM was implemented to both eliminate the false predictions and to determine what features are important. Comparing to the simple methods from SVM and evolving to GASVM and PCASVM[5], the solution to the main problem and sub-issues was more efficient and showed promising results for a real prediction using recent datasets.

Si, Deng, X., J., A., Liu, B., Mukherjee, Li, Q., Li, H. proposed a method for the stock market prediction on the basis of Twitter feeds sentiments which was experimented on SP 100 index dataset[6]. To understand the topic set daily, the continuous Dirichlet Process Mixture method was adopted. Twitter sentiment time-series and stock index were then regressed together to make a prediction.

Sneha Soni(2010) proposed a technique where a combination of three supervised machine learning algorithms are used for classification of stock market data [7] which are quadratic discriminant analysis (QDA), classification and regression tree (CART) and linear discriminant analysis (LDA). In section IV and V of IJCSE Vol.02, No. 09, 2010, experimental results shows that after comparison misclassification rate for LDA and QDA shows 74.26% and 76.57% respectively and for regression tree and classification shows 56.11%.

The work proposed by Sneha Soni (2010) and Shailendra Shrivastava is unique in comparison to other works in literature in this paper as for classification of Indian stock market data they used a combination of supervised machine learning algorithms instead other works contains unsupervised machine learning algorithms [7]. From this paper, it was concluded that classification and regression tree, supervised machine learning algorithm is best as compared to linear and quadratic discriminant analysis.

Hiral R. Patel and Satyen M. Parikh proposed the development of the prediction model on crude commodity based on its price movement due to news released by various sources. The decision strategy would be driven by analyzing stockprice fluctuations[8]. The paper decides the prediction method to be used in the model by performance comparison of following prediction techniques: Regression Modelling techniques, Classification Techniques, Statistical Techniques. The objective of the proposed work was to study the machine learning techniques to examine the effect of various kinds of government, policy-related, corporate released, global political and financial environment-related news on the different sectors for financial products and forecast the up and down trends[8]. The Paper focuses mainly on concepts like Multi Linear Regression(MLR) and Neural Network(NN). After implementation, the model achieved 82% of accuracy for MLR model and NN Model gives 70% accuracy for prediction and the same model applied in different tools with different methods the neural network with back propagation gives the best constant accurate result.

Ching-Te Wang and Yung-Yu Lin in 2015 12th International Conference on FSKD proposed a Web robot to capture data from the stock market. The system explores and analyzes the information to predict stock prices in the seesaw process. Using a group of cement, medical industries as the examples [9], this paper discusses the topics of Web robot, Genetic Algorithm and Support Vector Machine, which can provide a framework for data analysis and predict the stock market. Support Vector Machine (SVM) is an effective classification of supervision learning. SVM can map data

from high dimension into low-dimension space. When the data are classified in the procedure, SVM can collect non-overlapping data and then distinguish each other to a classification situation. The main characteristic of this system is to collect data automatically by using web robots and establish regular expressions, XPath to analyzing web pages [9]. From the experimental results, the system has shown better performance. Consequently, the method can crawl the valuable data, analyze huge information efficiently and provide the function to predict the prices in the stockmarket.

Girija V Attigeri ,Manohara Pai M M, Radhika M Pai, Aparna Nayak in 2015 in their paper considered both technical and fundamental [10]. Technical analysis is done using historical data of stock prices by applying machine learning and fundamental analysis is done using social media data by applying sentiment analysis.

III. System Architecture

The figure below shows the architecture of the proposed system.

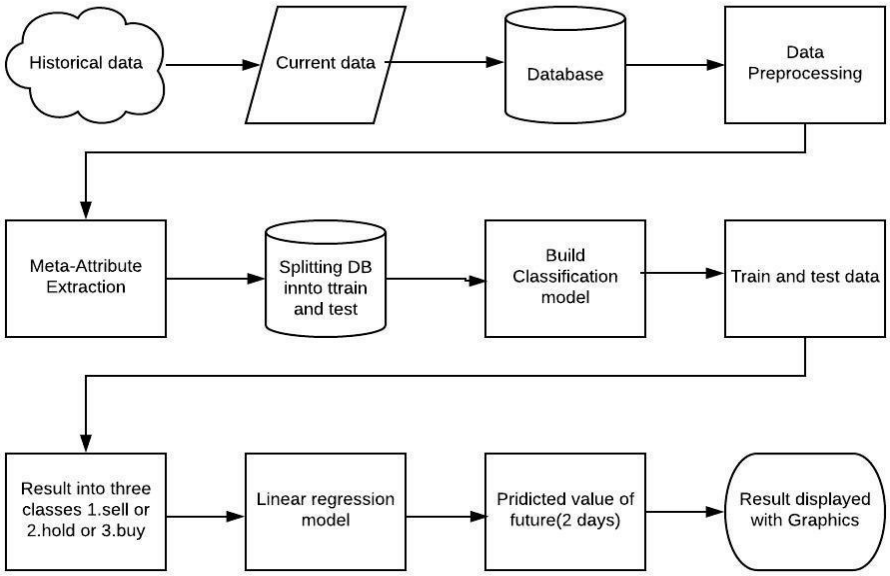

Figure 1: System Architecture

The brief description of the Architecture is as follows:

The proposed model is based on on the study of historical data, technical indicators, classification algorithm like Support Vector Machine (SVM) to be used in the classification and Regression Algorithm i.e linear regression to be used in pre- diction of daily stock prices. The architecture of the proposed model contains input vectors that represent the historical data and derived technical indicators and output represents stocks status and the next price. A web-based stock market prediction tool, which integrates all the findings in the dissertation, will be developed for the investors to provide them more valuable information for decision making.

IV. Proposed Solution

A. Mathematical Model

We in this work, study by varying the parameter n to see exactly how trends in volatility and momentum, both of the particular stock and the index, can be used to predict future changes in that stock.

Let n1 be parameter n for the index, and n2 be for the given stock, where

$$n1, n2 \in \{15, 45, 60, 80, 260\}$$

These represent one week, two weeks, one month, one quarter, and one year.

In each iteration we supply some combination of n1, n2, use these parameters to calculate our feature sets, train on the training data, and then predict on the testing set and check accuracy of results. We run 25 iterations, one for each combination of n1,n2.

In order to calculate the feature we average over the past n1 days for index and n2 days for stock, we start calculating feature vectors on the $d = (max(n1, n2) + 1)^{th}$date.

For example, if n1=6, n2=12, then d=13 and we start from the 13th date.

B. Features used in SVM

σs : Stock price volatility. This is an average over the past n days of percent change in the given stock's price per day.

$$\frac{\sum_{i=t-n+1}^{t} \frac{C_i - C_{i-1}}{C_{i-1}}}{n}$$

Stock Momentum : This is an average of the given stock's momentum over the past n days. Each day is labeled 1 if closing price that day is higher than the day before, and − 1 if the price is lower than the day before.

$$\frac{\sum_{i=t-n+1}^{t} y}{n}$$

σi : Index volatility. This is an average over the past n days of percent change in the index's price per day

$$\frac{\sum_{i=t-n+1}^{t} \frac{I_i - I_{i-1}}{I_{i-1}}}{n}$$

Index Momentum : This is an average of the index's momentum over the past n days. Each day is labeled 1 if closing price that day is higher than the day before, and − 1 if the price is lower than the day before

$$\frac{\sum_{i=t-n+1}^{t} d}{n}$$

We let Ct be the stock's closing price at time t , where t is the current day, and define It as the index's closing price that day. The stock's directional change on a given day is labeled as y∈{− 1,1}, and the index's directional change is defined as d ∈ {− 1, 1}. We use these features to predict the direction of price change between t and t + m, where m ∈{1, 4 , 15,45,60,80,260}

We are calculating the features to try and predict the price direction m days in the future, where m ∈{1, 4 , 15,45,60,80,260}.

As a result we skip the last m dates since we do not have the price m days after them. There are a total of MYears trading days between MYears and NYears, so we have a total of MYears− d − m days. The total set of feature vectors is called X

. We also have a set of output vectors y . y is calculated by finding the price direction on each of the $MYears - d - m$ days. We then split X and Y into the training and testing sets , which we call Xtrain, ytrain, Xtest, ytest.

We supply the feature vectors Xtrain as well as their corresponding output vectors ytrain to the SVM model. This is the training phase. We then supply only the testing feature vectors Xtest and have the model predict their corresponding output vectors. We then compare this output to ytrain.

C. Learning Algorithm Used

Support Vector Machines are one of the best binary classifiers. They create a decision boundary such that most points in one category fall on one side of the boundary while most points in the other category fall on the other side of the boundary. Consider an n -dimensional feature vector x = (X1,...,Xn). We can define a linear boundary (hyperplane) as

$$\beta 0+\beta 1X1+..+\beta nXn=\beta 0+n\sum i=1\beta iXi = 0$$

then elements in one category will be such that the sum is greater than 0, while elements in the other category will have the sum be less than 0. With labeled examples,

$$\beta 0+\sum ni=1\beta iXi=y,$$

where, y is the label. In our classification, $y \in \{-1,1\}$.

We can rewrite the hyperplane equation using inner products

$$y=\beta 0+\sum \alpha iyix(i)*x$$

where, $*$ represents the inner product operator.

The SVM replaces the inner product with a more general kernel function K which allows the input to be mapped to higher-dimensions. Thus in an SVM,

$$y=\beta 0+\sum \alpha iyiK(x(i),x$$

V. CONCLUSION

The paper summarizes important techniques in machine learning which are relevant to stock prediction. It presents computational results of supervised machine learning algorithms that is classification analysis using a set of rules on stock market data with the committed intention for maximizing earnings of market analyst and investors to make a selection for selling, buying or holding inventory (stock). A good way to examine the classification performance of various data is to apply classifiers on identical data and the outcomes can be compared to the idea of misclassification and accuracy of outcome. The paper proposes a system to extract knowledge from data and performing a prediction to advise the customer for investment.

ACKNOWLEDGEMENT

We would like to thank our guide Prof.Sumit Khandelwal, Department of Computer Engineering, MIT Academy of Engineering, for his guidance and support during this proposed work. We would also like to thank our Dean Mrs.R R Badre for her support and encouragement.

REFERENCES

[1] Sahaj Singh Maini and Govinda.K Stock Market Prediction using Data Mining Techniques.Proceedings of the International Conference on Intelligent Sustainable Systems(ICISS2017)IEEE Xplore Compliant
- Part Number:CFP17M19-ART, ISBN:978-1-5386-1959-9

[2] Pankaj kumar and Dr. Anju Bala Intelligent Stock Data Prediction using Predictive Data Mining Techniques at CSED Department, ThaparUniver- sity, Punjab,India

[3] CHING-HSUE CHENG, YOU-SHYANG CHEN FUNDAMENTAL ANAL- YSIS OF STOCK TRADING SYSTEMSUSING CLASSIFICATION TECH- NIQUES at Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong,

19-22 August2007

[4] Lamartine Almeida Teixeira and Adriano Lorena Incio de Oliveira. Pre- dicting Stock Trends through Technical Analysis and Nearest Neighbor Classification at Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, USA - October 2009.

[5] RaduIacomin.Stock Market Prediction at 2015 19th International Confer- ence on System Theory, Control and Computing(ICSTCC), October 14-16, Cheile Gradistei, Romania

[6] Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H., Deng, X. (2013).ExploitingTopic based Twitter Sentiment for Stock Prediction. ACL (2) ,2013 , 24-29

[7] Sneha Soni and Shailendra Shrivastava Classification of Indian Stock Mar- ket Data Using Machine Learning Algorithms at (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 09, 2010, 2942- 2946

[8] Hiral R. Patel and Satyen M. Parikh.Prediction Model for Stock Market using News based different Classification, Regression and Statistical Tech- niques at 978-1-5090-5515-9/16/ 2016IEEE

[9] Ching-TeWang and Yung-Yu Lin.The Prediction System for Data Analysis of Stock Market by Using Genetic Algorithm at 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)

[10] Girija V Attigeri ,ManoharaPai M M, Radhika M Pai, AparnaNayak Stock Market Prediction: A Big Data Approach at 978-1-4799-8641-5/15/ 2015IEEE