



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: II Month of publication: February 2020

DOI: <http://doi.org/10.22214/ijraset.2020.2006>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Implementation of Extended Deep Neural Networks for Stock Market Prediction

S. Kokila¹, Dr. A. Senthilrajan²

¹M.Phil Research Scholar, ²Professor, Department of Computer Application, Alagappa University, Karaikudi.

Abstract: In Machine Learning (ML) research prediction of variations in the stock price index is considered a significant technique. Exact prediction of prices and values in the stock market is a high economic advantage. This work presents the review of feasible techniques for predicting stock values with accuracy. Primarily we have to concentrate on a dataset of the stock market and its value like prices from past year. Then these were sent to pre-processing and comes out with exact analysis. Further, the data will reviewed under random forest, support vector machine on the dataset and results will be achieved. This work examines the value of the prediction system in this world and the accuracy of the given values. This work talks on the ML model to predict the longevity of stock in this contemporary market. The exact estimation of stock will be a success for the stock market and provide pragmatic remedies to the issues that investors face.

Keywords: Stock Market Prediction, Machine Learning, Support Vector Machine, Random Forest

I. INTRODUCTION

Stock markets are highly valid financial institutions of the capitalist economy. It permits the firms to increase the invested amount from public to fund their development of business. Meanwhile, investors might also face various risk factors for its unstable nature. So, private investors as well as financial institutions like banks, tries to predict stock values earlier then make earlier marketing decision.

If the stock prices are estimated earlier then it might be a helpful thing for tradees to make investment and take decision regarding their trade. And the gain we get from stock markets depend on this kind of market analysis.

The prediction must be rigid, valid and accurate. So the research on prediction are done on the basis of regular life and suits with the reality of the realm besides it also take account office all variable and performance of stock value. Prediction are done on the basis of methods like, Fundamental Analysis, Technical Analysis, Machine Learning, Market Mimicry, and Time series aspect structuring in this digitalized era this prediction moves to technological world. The significant of implementation in ML is [3] Artificial Neural Networks, Recurrent Neural Networks. ML involves artificial intelligences are applied to promote the system to learn from previous experiences without being programmed repeatedly. ML involves traditional methods for prediction with the usage of algorithms like Backward Propagation, are called as Back propagation errors. Ensemble learning techniques are used by researchers, it consume less time and price while other network lacks to predict future heights [3]. These predictions are for predicting stock values [1].

The random forest, is also for predicting stock values; algorithm uses ensemble learning strategy for both classification and regression. It takes the average of number of data, it improves predictive accuracy as well as decreases over-fitting of the data.

II. PROBLEM DEFINITION

Predicting the values of goods in stock market is to give an idea for the consumers; related to the goods and price of it; this is what called as stock market prediction. It is mostly products with the help of quarterly financial ratio from data. So for this prediction we must get through numerous data; we can not predict it with single data. Through machine learning technique, we get through various data for the purpose of predicting the value of stock market and its trends.

III. RELATED WORK AND LITERATURE SURVEY

A. Survey of Stock Market Prediction Using Machine Learning Approach

Prediction of stock market is a significant thing of current period. Though technical analysis did not give absolutely result it is used for prediction. People invests their money on the basis of prediction, by knowing all drawbacks too; for this vast data undergoes prediction process. It is a technique of regression with its advantage and limitations; among that linear regression is highlighted. It works with the utilization of least squares approach, yet it is also with some demerits like diminishing a handicapped version of the least squares loss function. So, this approach is used for nonlinear models. [1]

B. Impact of Financial Ratios and Technical Analysis on Stock Price Prediction Using Random Forests

Machine learning and artificial intelligence techniques are also used for the predicting stock market trends. So researches are done to find the technique to predict this with fine accuracy. Though there are numerous ways to predict this stock price are available, each techniques gets differ; even for the single unique data output will get differs. The work used for this research is done on the basis of random forest algorithm is being used to predict the price of the stock using financial ratios form the previous quarter. This is just one way of looking at the problem by approaching it using a predictive model, using the random forest to foresee stock values. Various factors that influence stock's price are sentiments of the investor, public opinion, news from various outlets, and even events that cause the entire stock market to fluctuate. This stock market prediction are done with accuracy through the usage of financial ratio and analysis of investor's mind and sentiment. [2]

C. Stock Market Prediction via Multi-Source Multiple Instance Learning

Though prediction of stock market is tough task, there are several web techniques are available to make this as a simple one. Owing to interconnected data format it will become a simplified one to track the rappo between various variable and roughly sketch the investment details. This pattern of investment of different work places have some similar features and it leads to foresee stock market with an accuracy using datasets. The prediction of stocks can be done by using technical historical data, and sentiment of investors, to analyze their mental conflict for investing in specific stocks. Prediction also done by getting details of significant events through web news to get its impact on stock prices. [3]

D. Predicting Stock Price Direction Using Support Vector Machines

Financial organizations like bank and business men have tried their hand in various process and over rule market for their case or for their investors but not often some gets unpredictable profit than normal. The challenges of predicting stock values and the facts involved need double focus. It also gives benefits with dollars for these crews. [6]

IV. PROPOSED APPROACH

This research work aims to predict stock values using methods machine learning algorithms it includes Random Forest and Support Vector Machines.

This research work "Stock market price prediction" it have predicted the stock market price with the random forest and SVM algorithm. In this work, through ML technique we predict the future stock value by analyzing the past datasets. So, we utilizes datum from past year for training; and the issues were solved by dual ML libraries.

And those are numpy and scikit; while the former focuses on clearing and manipulating the data for analysis the later focuses on the purpose of analysis and predicting the data.

The datum collected from the past year was utilized for training purpose, in that 80% was used for training and the balance 20% was used for testing purpose. Basically supervised learning model involves in the process of learning through patterns and exploring the relationships between data of training set and generates them the test data.

This research also make use of python pandas library for processing data with the combination of various datasets in a data frame, which permits to develop a data for extraction.

The data frame features were mentioned and the closing price will be allotted for some day. This work done with all modules to predict the stock value under random forest model then come with the prediction of object variable, considered as a value of a given day. This work also quantified the accuracy by using the predictions for the test set and the actual values. The research work includes various field of research like, data pre-processing, random forest, and so on.

V. METHODOLOGIES

A. Classification

In this area the set of data gets analyzed and categorized under its particular column on the basis of its similar attributes. Through the datasets or values under analysis this method concludes the process from observed values. In the case of multiple input, the output also expected in various form. This work includes classifiers like, random forest classifier, SVM classifier.

- 1) *Random Forest Classifier:* It is a classifier besides a supervised algorithm. It yield results on the basis of developing decision trees. The basic approach of this classifier is to proceed with decision aggregate through random subset decision tress as well as gives a final result depends on votes of the random subset of decision trees.

- a) *Parameters*: It includes random forest classifier and mentioned $n_estimators$, it denotes total number of decision trees, and other hyper parameters addobo-score to mention generalization accuracy of the this method, $max_features$ consists of number of features for the best-split. $min_weight_fraction_leaf$ is the minimum weighted fraction of total weights of all the input samples needed for leaf node. Samples considered to be in equal weight, at the time providing samples without its exact weight.
- 2) *SVM classifier*: It is a discriminative classifier. It works on the basis of supervised learning i.e. a labeled training data. Output are in hyper plane mode to categorize new dataset. They are SVM with the aid of learning algorithm for the purpose of both classification and regression.
 - a) *Parameters*: mSVM classifier, consists of parameters like kernel parameter, gamma parameter and regularization parameter. The former can be mentioned as linear and polynomial kernels used to calculate the prediction line. Then the later focus on the prediction of new input and calculates it by dot product between the input and support vector.
 - i) C parameter is called as regularization parameter; it works to determine the accuracy of model in quantity and quality basis of increases or decreases. $c=10$ is considered as a default value. If the value gets deceased, the it ends in miscalculation.
 - ii) Gamma parameter measures the influence of a single training on the model. Lower value denotes distance from the plausible margin at same time higher value denotes nearness from the plausible margin.
- 3) *Random Forest Algorithm*: It is used for predicting stock market. It is an easiest and the most flexible methods of machine learning algorithm, it leads to exact prediction. It is commonly used in classification tasks. Owing to its high volatility in stock market, predicting became a challenging task. For prediction of stock we apply random forest classifier, with equal hyper parameters as a decision tree. The decision tool has its similarity with that tree. And the decision taken by this are on the basis of events like outcome, resource cost, and utility. It denotes an algorithm where it randomly selects different observations and features to develop numerous decision tree and make use of aggregate decision trees as outcomes. Then the data gets divided into parts on the basis of questions on a label or an attribute. The data we are using for this is a previous set of data that we collected from public sector; in this 80% is for training machine and 20% is for testing purpose. Its base purpose is to know patterns and establish the relationship between the data from training set and reproduced thing from the test data.
- 4) *Support Vector Machine Algorithm*: Support machine algorithm mainly focuses on identifying an N-dimensional space that differentiates other data points. N denotes the number of features. Various hyper planes can be chosen between two data points. This algorithm focuses to obtain plane with maximum margin, it refers to the distance between data points of classes under discussion. It has benefits like, giving reinforcement for future data points to easy classification. Decision boundaries which are there to categories data points are known as hyper planes, on the basis of data points in hyperlanes; which are attributed to various classes. Its dimension rests on number of attributes, if it is two and the hyperplane is a line, if it is three the hyperplane is two dimensional.

VI. SYSTEM ARCHITECTURE

Kaggle is an online community to analyze data and predict it with datasets of various area; regards to data miners. Data scientist oblige to develop the best models for predicting and depicting the information, which permits the user to make use of database to build their modules angelic science engineers to overcome real life issues.

The dataset used in the proposed project has been downloaded from Kaggle. This data will be presented in raw format. The data is all about the collection of stock market and its data of a few companies.

Primarily it deals with conversion of raw data into processed data with feature extraction, as it is in raw data with multiple attributes yet only few attributes are utilized for this purpose of prediction. Primarily it goes with feature extraction, where the key attributes are listed from the group of attributes in the dataset.

Measured data and builds derived values are the initiation phase for feature extraction. It initiates, both informative and non-redundant, facilitating regards learning and generalization steps. Feature extraction is a dimensionality reduction process, here the feature of raw data get reduced by deleting the unwanted datum, to make the information precisely for denoting a information which are collected.

This extraction process done on the basis of classification process, there data gets divided and classified into segments. It is a matter of classifying things on the basis similar qualities and characteristics. The training data set is made for the purpose of training modules and test data are used for prediction. And this is classified like the form where that training data should have higher proportion than the test data.

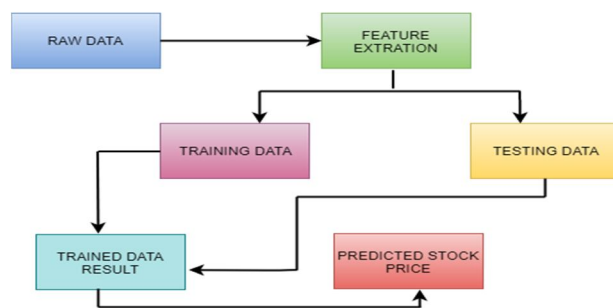


Fig.1 System Architecture

The random forest algorithm uses group of random decision trees for analyzing data. In layman terms, from the total number of decision trees in the forest, a cluster of the decision trees look for specific attributes in the data, called data splitting. Our research aims to predict the stock price by analyzing history of data.

A. Module Identification

The models and techniques involved in this research was segmented as follows:

- 1) Data Collection
- 2) Pre Processing
- 3) Training the Machine
- 4) Data Scoring

VII. EXPERIMENT RESULTS

The experiment are going done on the xlxs file, with raw dataset to find our predictions. The eleven columns otherwise the eleven attributes are there to mention increase or downfall of the stock value. The attributes, involves in this are (1) HIGH, explains the peak value of previous year stock. (2) LOW, it is different HIGH then provides the down value of past year stocks (3) OPENP it denotes the stock's value in the starting day of trade (4) CLOSEP mentions the value of stock before the day of closing of trade. Some other attributes are YCP, LTP, TRADE, VOLUME and VALUE, yet the former attributes have its valuable role in our research.

	Date	Open	High	Low	Close	Adj Close	Volume
0	2019-08-02	87.839996	89.230003	87.339996	88.610001	88.610001	3820100
1	2019-08-01	87.019997	87.610001	84.440002	86.330002	86.330002	5809000
2	2019-07-31	86.070000	86.709999	81.919998	82.190002	82.190002	5220800
3	2019-07-30	82.690002	86.269997	82.480003	85.599998	85.599998	4328600
4	2019-07-29	84.099998	85.580002	82.839996	84.629997	84.629997	3364300
5	2019-07-28	85.980003	86.949997	83.919998	84.989998	84.989998	3902000
6	2019-07-27	84.379997	86.250000	84.379997	85.849998	85.849998	4363200
7	2019-07-26	87.250000	87.699997	85.370003	85.510002	85.510002	3796500
8	2019-07-25	85.620003	88.470001	85.620003	88.040001	88.040001	3398900
9	2019-07-24	88.470001	90.279999	88.010002	88.680000	88.680000	3335900
10	2018-11-05	88.980003	89.820000	88.279999	89.470001	89.470001	2735500
11	2018-11-06	90.000000	90.250000	88.830002	89.639999	89.639999	2187400
12	2018-11-07	90.220001	92.330002	89.650002	92.160004	92.160004	3792200
13	2018-11-08	92.089996	92.849998	91.529999	92.559998	92.559998	2460300
14	2018-11-09	92.410004	93.599998	91.739998	92.349998	92.349998	4254100
15	2018-11-12	92.360001	92.690002	91.029999	91.199997	91.199997	2903200
16	2018-11-13	91.629997	93.379997	91.629997	92.599998	92.599998	3696600
17	2018-11-14	93.029999	93.470001	91.379997	91.790001	91.790001	2387000
18	2018-11-15	91.400002	92.660004	89.019997	91.980003	91.980003	3691500
19	2018-11-16	91.839996	92.599998	90.070000	92.260002	92.260002	2837900
20	2018-11-19	92.320000	92.949997	91.160004	92.089996	92.089996	2684600
21	2018-11-20	91.199997	92.790001	90.449997	91.769997	91.769997	3232500
22	2018-11-21	91.919998	93.379997	91.820000	92.860001	92.860001	3771600

Fig.2 Raw Data

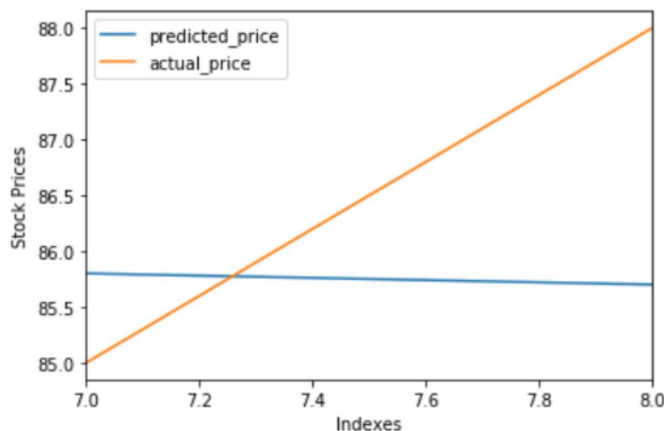


Fig.3 Random Forest & SVM Price Prediction

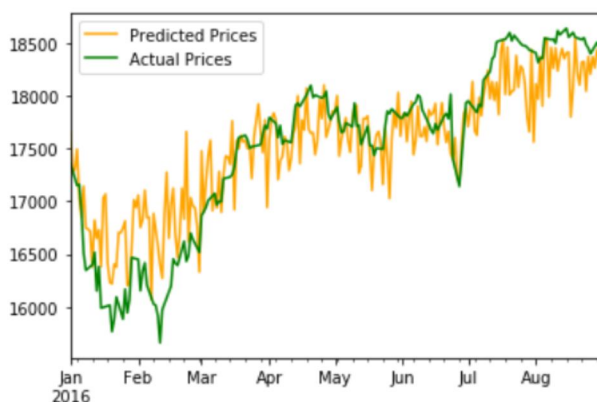


Fig.4 Proposed work accuracy Prediction

The following table shows the works of various techniques on Stock Market Prediction having Stock Market dataset.

Table. I Comparative Study Of Techniques Of Stock Market Prediction

Techniques	Accuracy
Naive Bayes	84.3%
K-nearest Neighbor (KNN)	80.7%
Decision Tree (DT)	82.1%
Proposed Work	91.9%

VIII. CONCLUSION

Through these Machine Learning algorithms perfect algorithm for the purpose of predicting the stock market value is found on the basis of research conducted with various data from collected data history. After several attempts done on sample data, the specified algorithm which is suitable for exact prediction stock values; it will be useful for both investors and brokers. This research work projects and proceeds with ML method and the value of stock market goods are foreseen finely while comparing with past modules and research works. By including various things like, parameters and financial ratios, multiple instances, etc mend way and leaf a scope for further research. As per the amount of parameters included in the research process, the prediction of stock market value comes with accuracy. This kind of algorithms are processed to analyze the contents public reviews and patterns as well as the relationships between the customer and the corporate employee. The corporation's performance structure were predicted through traditional algorithms and data mining techniques.

REFERENCES

- [1] Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017.
- [2] Loke.K.S. "Impact Of Financial Ratios And Technical Analysis On Stock Price Prediction Using Random Forests", IEEE, 2017.
- [3] Xi Zhang¹, Siyu Qu¹, Jieyun Huang¹, Binxing Fang¹, Philip Yu², "Stock Market Prediction via Multi-Source Multiple Instance Learning." IEEE 2018.
- [4] Vivek Kanade, Bhausaheb Devikar, Sayali Phadatare, Pranali Munde, Shubhangi Sonone. "Stock Market Prediction: Using Historical Data Analysis", IJARCSSE 2017.
- [5] Sachin Sampat Patil, Prof. Kailash Patidar, Asst. Prof. Megha Jain, "A Survey on Stock Market Prediction Using SVM", IJCTET 2016. https://www.cs.princeton.edu/sites/default/files/uploads/Saahil_magde.pdf
- [6] Hakob GRIGORYAN, "A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA)", DSJ 2016.
- [7] Raut Sushrut Deepak, Shinde Isha Uday, Dr. D. Malathi, "Machine Learning Approach In Stock Market Prediction", IJPAM 2017.
- [8] Pei-Yuan Zhou, Keith C.C. Chan, Member, IEEE, and Carol Xiaojuan Ou, "Corporate Communication Network and Stock Price Movements: Insights From Data Mining", IEEE 2018.
- [10] Scikit Learn Team. 2016. Stochastic Gradient Descent [ONLINE] Available at <http://scikit-learn.org/stable/modules/sgd.html> [Accessed 20 May 2017].
- [11] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features." In European conference on machine learning, pp. 137-142. Springer, Berlin, Heidelberg, 1998.
- [12] B. Pang, and L. Lee. "Opinion mining and sentiment analysis." Foundations and Trends® in Information Retrieval 2, no. 1-2, 2008
- [13] H. Tang, S. Tan, and X. Cheng. "A survey on sentiment detection of reviews." Expert Systems with Applications 36, no. 7, 2009
- [14] A. Go, R. Bhayani, and L. Huang. "Twitter sentiment classification using distant supervision." CS224N Project Report, Stanford 1, no. 12 2009.
- [15] A. Pak, and P. Paroubek. "Twitter as a corpus for sentiment analysis and opinion mining." In LREC, vol. 10, no. 2010, pp. 1320-1326. 2010.
- [16] D. Davidov, O. Tsur, and A. Rappoport. "Enhanced sentiment learning using twitter hashtags and smileys." In Proceedings of the 23rd international conference on computational linguistics: posters, pp. 241-249. Association for Computational Linguistics, 2010.
- [17] T. J. RANI, K. ANURADHA, and P. V. REDDY. "Sentiment Classification on Twitter data using Word N Gram Model."
- [18] B. M. Jadav, and V. B. Vaghela. "Sentiment analysis using support vector machine based on feature selection and semantic analysis." International Journal of Computer Applications 146, no. 13, 2016.
- [19] A. Balahur, "OPAL at SemEval-2016 Task 4: the Challenge of Porting a Sentiment Analysis System to the "Real" World." In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pp. 262-265. 2016.
- [20] G. Briones, K. Amarasinghe, and B. McInnes. "VCU-TSA at Semeval-2016 Task 4: Sentiment Analysis in Twitter." In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pp. 215-219. 2016.
- [21] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov. "SemEval-2016 task 4: Sentiment analysis in Twitter." In Proceedings of the 10th international workshop on semantic evaluation (semeval-2016), pp. 1-18. 2016.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)