

Forecast the Stock Market Using Time Series and Regression Models

K.Lalitha¹, R.Indhumitha², N.Indra³, M.Kaviprabha⁴

¹Associate Professor, Department of Information Technology, Panimalar Engineering College, Chennai, India
^{2,3,4,5} Department of Information Technology, Panimalar Engineering College, Chennai, India

Date of Submission: 18-06-2020

Date of Acceptance: 03-07-2020

ABSTRACT

Stock Market prediction is a complex and burdensome task but it is of great advantage for all the stock market investors. The main reason behind this difficulty in stock market prediction is that the price of the stock keeps fluctuating with respect to time on the basis of various parameter like economics, politics, natural disasters, man-made disasters and market psychology. In order to predict we try to analyze the future value of the stock of each company. Now we are in an attempt to predict the stock using Machine learning algorithms like Time Series algorithms and Regression algorithm on the bases of the historical stock data. The factor we have taken into consideration is close price of the stock market. In this project we forecast the best prediction value using ensemble techniques over all the algorithms applied.

Keywords: Stock price, prediction, machine learning algorithm.

I. INTRODUCTION

Today stock market has gained the interest of many people. Stock market has been a place where the people are allowed to buy and sell stock in companies as investments. Having a stock in a company is like sharing the ownership in the company. Stock market is a platform which allows investors to buy and sell stocks. The growth of the company also influence the shares brought by the investors. If the company gains a profit then the profit is shared with all the investors in the ratio of the shares they have brought. similarly the losses of the company is also shared with the investors. The exchange of stock is allowed between the investors. The stock price often fluctuates due to many factors like economics, politics, natural disaster, manmade disasters and market psychology. Stock prediction has now gained attention of many investors as it is hugely profitable. Assume that you are interested in buying a stock in a company but you don't know anything about stock market. In

such a situation our system act a medium which predicts the company's stock for the next upcoming month thereby giving an idea for the user on which is the great time to buy a stock in that company at a profitable rate. This application helps the investors to get a clear picture of the stock markets and the future fluctuation in it. This application also helps to analyze between the stock of various companies which helps the investors to make investment in the profitable company. Our project is to forecast the stock for the next month using the 2 years of historical data. Here we use various time series forecasting models and regression models based on the nature of the dataset the best models are chosen to forecast for the next month by ensemble.

The objective of this project is to predict the stock of nifty 50 with least mean absolute percentage error thereby forecasting with higher accuracy.

II. LITERATURE SURVEY

1. A Prediction Approach For Stock Market Volatility Based On Time Series Data

In this paper the author is using ARIMA model. ARIMA does not require a set of predictor variable. ARIMA is a method in which we analyze the future data using the previous historical dataset. In this method we use a time series data. The time series data is one in which we have date as one attribute and one additional attribute. This data must not have a null value if it has then we must change the null value into an appropriate value. But the ARIMA performs shows some poor result when its sizes are small for long-term predictions. The ARIMA is good only for shorter period of prediction using small sizes. When the data set is large and we want a short term prediction then it won't be better to use ARIMA model. It can't be used for all the type of dataset.

2. Survey of Stock Market Prediction Using Machine Learning Approach

This paper presents a new design in which the author advises to use Support Vector regression in order to predict the stock price fluctuation. Support vector regression works well with small dataset and it is not efficient for larger dataset system, thus support vector regression won't give more accurate prediction in terms of stock market forecast where we usually use a larger set of historical dataset for prediction. Another major disadvantage of support vector regression is that it gives a poor result in case of noisy data.

3. Stock Market Analysis Using Supervised Machine Learning

In this paper the author recommends the supervised learning technique to predict the stock markets. Using this method we can get to know very specifically about the definition of the classes, which means that we can train the classifier in a way which has a perfect decision boundary to differentiate different classes accurately. We can specifically determine how many classes we want to have in case of supervised learning. The major disadvantage of supervised learning is that it takes a lot of compilation time. In case of supervised learning we need to train the model with a lot of dataset which is a burdensome task to accomplish.

4. Short-Term Prediction For Opening Price Of Stock Market Based On Self-Adapting Variant Pso-Elman Neural Network

In this paper the author suggest the usage of neural network in prediction. Neural networks are versatile and may be used for each regression and classification issues. Any information which may be made numerical can be employed in the model, as neural network could be a mathematical model with approximation functions. Neural networks area unit are sensible to model with nonlinear information with sizable amount of inputs; for instance, images. It's reliable in Associate in Nursing approach of tasks involving several options. It works by rending the matter of classification into a stratified network of easier components.

Once trained, the predictions area unit pretty quick.

Neural networks is trained with any range of inputs and layers. Neural networks work best with a lot of information points. The disadvantage is Neural networks are black boxes, meaning we cannot know how much each independent variable is influencing the dependent variables. It is computationally very expensive and time consuming to coach with traditional

CPUs. Neural networks depend a lot on training data. This results in the matter of over-fitting and generalization. The mode relies more on the training data and should be tuned to the info .

III. EXISTING SYSTEM

In existing system stock market prediction works well with few set of data and gives a poor performance with other set of data as they use only one or two machine learning algorithm and each of this algorithm has its own advantages and disadvantages. Some works best with large dataset and some of the existing system works well with smaller dataset. There is no existing system which gives great performance irrespective of the dataset size.

IV. PROPOSED SYSTEM

In this project we are going to use some time series algorithm and some regression algorithm. Here we are going to dynamically choose a three best algorithm, of the 3 chosen algorithm we will apply the average ensemble method to give an 90% accurate prediction. The three best algorithm is chosen based on the MAPE (Mean Average Percentage Error) value.

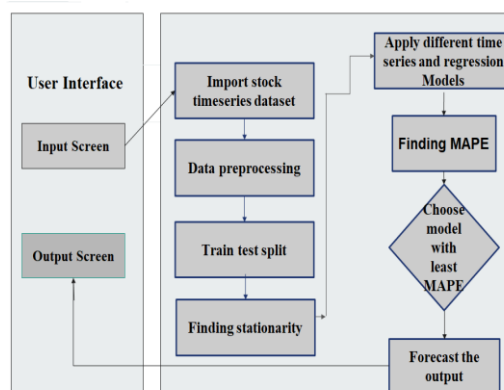


Fig. 1: High level architectural diagram

V. MODULES

1. DATA PREPROCESSING MODULE
2. ALGORITHM APPLICATION MODULE
3. ENSEMBLE MODULE

1. DATA PREPROCESSING MODULE

In order to apply time series model, the dataset must be stationary data. The stationary data is one in which the stationary property are constant very time. This property includes mean, median and variance. Before converting the data into stationary data, the data is decomposed into original, trend , seasonal and residual part. Then the data set is

split into train and test part in order to undergo prediction. Here 80% of the dataset is taken as train part and 20% is taken as test part.

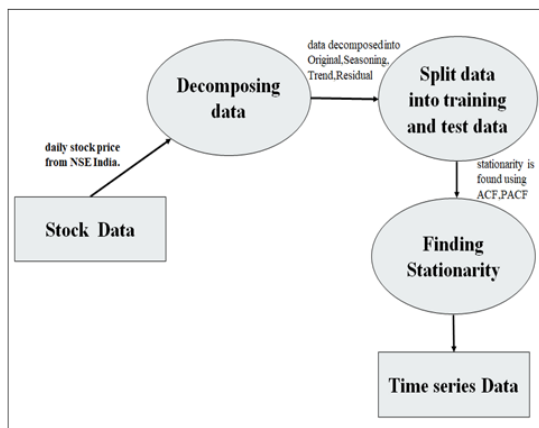


Fig. 2: Data flow diagram for data preprocessing

2. ALGORITHM APPLICATION MODULE

In this method we tried to use various time series algorithms and regression algorithm to improve the performance of our system. The time series algorithm used are moving average, ARIMA, auto regression, exponential smoothening, holts winter. The regression algorithm used are linear regression, random forest, support vector machine, XGBoost, decision tree, k-nearest neighbour. The best of the algorithm is chosen using the MAPE. The prediction is 90% accurate when the MAPE value is less than 10%.

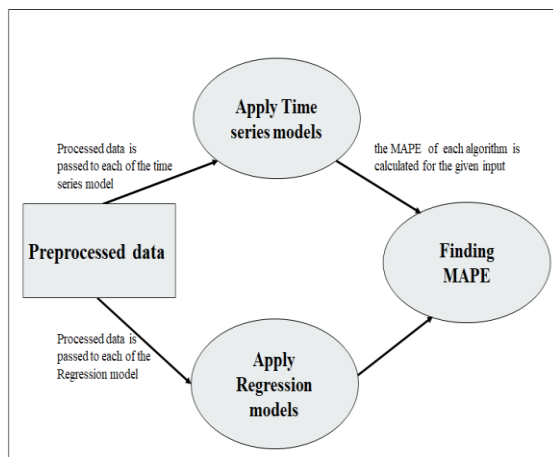


Fig. 3: Data flow diagram for algorithm application

3. ENSEMBLE MODULE

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model. The goal of any machine learning problem is to

seek out one model which will best predict our wanted outcome. Rather than making one model and hoping this model is the best accurate predictor we will make, ensemble methods take a myriad of models under consideration and average those models to produce one final model.

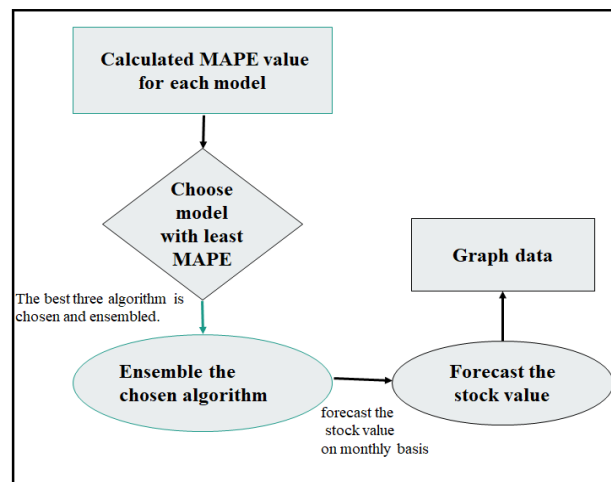


Fig. 4: Data flow diagram for ensemble

VI. FIGURES AND TABLES

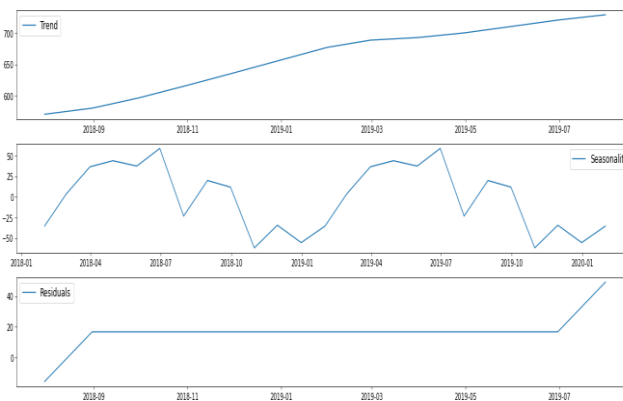


Fig.5: Trend, seasonal, residual split

```
MOVING AVERAGE= 18.995553641027414
AUTO REGRESSION= 14.8628438868928
HOLTS FORECASTING= 3.6887541242342333
EXPONENTIAL SMOOTHING= 5.959712456288877
LINEAR= 18.196748888888888
RANDOM FOREST= 4.741742048832427
SUPPORT VECTOR= 18.391198828888152
[02:32:07] WARNING: /workspace/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squarederror.
XGB= 4.715891535273229
DECISION TREE= 6.114884735144751
XNN= 4.531116182338885
```

Fig. 6: MAPE value

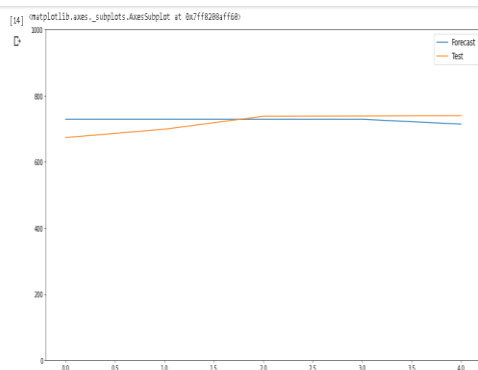


Fig . 7: test and forecast value

VII. FUTURE ENHANCEMENT

This works better for dataset having nearly a stable growth or decline. If there is a sudden drastic change in the dataset the accuracy of the prediction may be poor. Thus a new algorithm can be created to overcome this effect of unstable dataset. The prediction is only based on the close value of the stock it could be enhanced to include the other attributes lie open, high, low and so on. Only one dataset applied to train and test the models. The system can only predict the direction(up/down) for the next trading month. In the future, regression model and time series model will be used in order to predict the price moment for the future and the results will be compared with other data processing techniques by applying different dataset from different stock market index .

VIII. CONCLUSION

This application used for stock prediction given an output with more accurate value then the existing models. The MAPE value is found to be less than 5%. The technique of dynamically choosing the best prediction models has lead to the more accurate prediction. We can also improvise the prediction by adding ever some more time series and regression models in this application . We must work on data which is noisy. This application is developed in order to help the investors analyze the stock market and make their decisions wisely in buying and selling of stocks.

REFERENCES

[1]. Anderson, D. R. (2007). Model based inference within the life sciences: a primer on evidence. Springer Science & Business Media.
[2]. Ariyo, A. A., Adewumi, A. O., and Ayo, C. K. (2014). Stock price prediction using the arima model. In Computer Modelling and Simulation (UKSim), 2014 UKSim-AMSS 16th International Conference on, pages 106–112.

IEEE.
[3]. Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.
[4]. Box, G. E. and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical association, 70(349):70–79.
[5]. Gupta, A. and Dhingra, B. (2012). Stock market prediction using hidden markov models. In Engineering and Systems (SCES), 2012 Students Conference on, pages 1–4. IEEE.
[6]. Hassan, M. R. (2009). A combination of hidden markov model and fuzzy model for stock market forecasting. Neurocomputing, 72(16-18):3439–3446.
[7]. Kryzanowski, L., Galler, M., and Wright, D. W. (1993). Using artificial neural networks to pick stocks. Financial Analysts Journal, 49(4):21–27.
[8]. Labiad, B., Berrado, A., and Benabbou, L. (2016). Machine learning techniques for short term stock movements classification for moroccan stock exchange. In Intelligent Systems: Theories and Applications (SITA), 2016 11th International Conference on, pages 1–6. IEEE.
[9]. Mostafa, M. M. (2010). Forecasting stock exchange movements using neural networks: Empirical evidence from kuwait. Expert Systems with Applications, 37(9):6302–6309.
[10]. Schöneburg, E. (1990). Stock price prediction using neural networks: A project report. Neurocomputing, 2(1):17–27.
[11]. Wei, L.-Y. (2013). A hybrid model based on anfis and adaptive expectation genetic algorithm to forecast taiex. Economic Modelling, 33:893–899. [12] Wong, B. K., Bodnovich, T. A., and Selvi, Y. (1997). Neural network applications in business
[12]. Yoon, Y., Guimaraes, T., and Swales, G. (1994). Integrating artificial neural networks with rule-based expert systems. Decision Support Systems, 11(5):497–507.
[13]. Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. International journal of forecasting, 14(1):35–62.
[14]. Angadi M, Kulkarni A. Time Series Data Analysis for Stock Market Prediction using Data Mining Techniques with R International Journal of Advanced Research in Computer Science,2015



**International Journal of Advances in
Engineering and Management**

ISSN: 2395-5252



IJAEM

Volume: 02

Issue: 01

DOI: 10.35629/5252

www.ijaem.net

Email id: ijaem.paper@gmail.com