International Journal Of Multidisciplinary Research In Science, Engineering And Technology (IJMRSET)

Volume 1, Issue 1, April 2020

Machine Learning Algorithms for Predicting the Stock Market Daily Returns

S.Rajina Begam,

Department of Computer Science and Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India

ABSTRACT: Big data analytic techniques developed with machine learning algorithm are yielding an important role in numerous application. However, some studies have focused on forecasting daily returns of the stock market, especially when using machine learning techniquesto perform the powerful analyses. This paper represents a big data analytics process to predict the daily return direction of the SPDR S&P 500 ETF using Machine Learning algorithm. Machine Learning algorithm such as Support Vector Regression (SVR) and Auto Regression and Integrated Moving Average (ARIMA) along with traditional Artificial Neural Network (ANN) are then deployed over the entire dataset, to predict the daily direction of future stock market index returns. . Moreover, a set of hypothesis testing procedures are implemented on the classification, and the simulation results show that the ANN datasets give significantly higher classification accuracy than those using the SVR and ARIMA with the artificial neuralnetwork.

KEYWORDS: Daily stock return forecasting, machine learning algorithms, Artificial neural networks (ANNs), Support vector Regression (SVR) and ARIMA (Auto Regressive Integrated Moving Average).

I. INTRODUCTION

Stock price prediction mechanisms are fundamental to the formation of investment strategies and the development of risk management models. Models based on the Neural network provide the best results.Stock market are affected by many factors. Some of the factors are 1) economic variables, like interest rates, exchange rate and commodity prices. 2) Industrial variables like growth rate of industries and consumer prices. 3) company specific data, like changes in company's policies, income statements, and dividend yields.

Many of the methodology are applied in predicting the daily returns of the stock, some of the methods includes support vector regression, ARIMA (Auto regression Integrated Moving Average) and also other time series algorithm.Stock price prediction using support vector regression on daily and up to the minute prices by Herbert Kimura, Vinicius Amorim Sobreiro, Close Bruno Miranda Henrique [1]. The support vector machine (SVM) is a training algorithm for learning classification and regression rules from data.

This method uses training observations for building a linear model by the non-linear classification thresholds, which maps variables on a greater number of dimensions. Support vector Regression provide the much accurate results when using the kernel function. The mostly used kernel function is linear, polynomial and the sigmoidal. As the sigmoidal function is used for predicting using the support vector regression.

In this study, the daily return direction of the SPDR S&P 500 ETF is predicted based on methodology of ARIMA (Auto Regressive Integrated Moving Average) and multilayer feed forward neural network algorithms. This method begins by feature selection and preprocessing the data to manage missing values and therefore the mismatched samples. The entire dataset is transferred or preprocessed then employed for prediction using ARIMA and Multilayer feed forward neural network. Finally the results of prediction by ARIMA and neural network is compared with the standard SVR model.

II.LITERATURE SURVEY

In [1] proposed a new approach using Support Vector Regression(SVR) on daily and up to the minute data to improve the prediction accuracy. The model used the historical data by employing the prices with both daily and up-to-the minute frequencies. The price of the stock was foreseen exploitation the SVR(Support Vector Regression) model by using the costs with each daily and up-to-the minute frequencies. international intelligence agency uses the 2 kernel model like linear and correlation kernels. international intelligence agency predictions exploitation linear kernel, correlation values are on the brink of one for nearly all stocks and perpetually updated models exploitation linear and radial kernels.

In [2] used Support Vector Machine(SVM) algorithm. SVM method associate degree learning rule that has the fascinating characteristics of the management of the choice operate and the utilization of the kernel methodology. Author took survey of stock price for the four company-specific and 6 political economy factors which will influence the stock trend are selected for stock statistical analysis method. Then SVM was employed in analyzing the link of those factors and predicted the stock performance. The conclusion of the project is that, SVM work on the large dataset values and doesn't offer problems of over fitting. Best and most sensible commercialism or trading model.has proposed a work based on the different techniques of preprocessing or input selection technique for SVR to predict the stock price and the stock market trends. Differing kinds of windowing operations are used as data preprocessing techniques. The different kinds of the windowing techniques were used to feed more reliable input to the regression model. The various window operations are Rectangular, Flatten window and De-Flatten window that manufacture the generic information from statistic data. The RBF kernels function was employed to provide the fast and correct results.

In [3] used Support Vector Machine(SVM) algorithm. SVM method associate degree learning rule that has the fascinating characteristics of the management of the choice operate and the utilization of the kernel methodology. Author took survey of stock price for the four company-specific and 6 political economy factors which will influence the stock trend are selected for stock statistical analysis method. Then SVM was employed in analyzing the link of those factors and predicted the stock performance. The conclusion of the project is that, SVM work on the large dataset values and doesn't offer problems of over fitting. Best and most sensible commercialism or trading model.

[4] has proposed a work based on the different techniques of preprocessing or input selection technique for SVR to predict the stock price and the stock market trends. Differing kinds of windowing operations are used as data preprocessing techniques. The different kinds of the windowing techniques were used to feed more reliable input to the regression model. The various window operations are Rectangular, Flatten window and De-Flatten window that manufacture the generic information from statistic data. The RBF kernels function was employed to provide the fast and correct results.

In [5] proposed method of predicting the future values of the stock by two machine learning called Support Vector Regression and Neural Networks. The stock exchange is foreseen by two methodology like support vector machine and neural networks. The author predict the future value by two methods and compare to seek out the correct methodology of prediction. Support vector regression uses the RBF kernel. For prediction and classification two layer neural network with linear function is employed.

[5] has proposed ARIMA Model for predicting the daily returns of the stock. Model used New York Stock Exchange (NYSE) and Nigeria Stock Exchange(NSE) data for prediction. The closing price of the data is foreseen by ARIMA and found to be efficient and robust. ARIMA was primarily used for short- term and linear statistic data prediction. The most effective ARIMA model is confirmed by uses the standards Relative small BIC (Bayesian Information Criterion).

[6] had survey the effectiveness of time series modeling ARIMA in forecasting stock prices. ARIMA model was the the general modification of Auto Regressive Moving Average (ARMA) model. The best ARIMA model is assessed as ARIMA(p, d, q), where p states the autoregressive elements, d denotes the integrated elements and q denotes moving average parts of the data set. All this three-parameters p, d, q all nonnegative integers. The accuracy was measured in proportion by uses the Mean Absolute Error (MAE). The paired T-test is conducted to visualize the distinction between the accuracy of the expected and therefore the training data.

In [7] has proposed a work of forecasting stock market indexes using principle component analysis and stochastic time effective neural networks. The financial time series was predicted by the Stochastic Time effective function Neural Network (STNN) with the preprocessed data done by Principal Component Analysis (PCA). The information was extracted from the input data and so integrates the STNN model to perform the prediction. STNN compared with the standard Back Propagation (BP) network and PCA-BPNN which offer higher results.

[8] developed a new approach of Short term stock price forecasting using kernel principal component analysis and support vector machines. The prediction capability of the stock value was improved by integrating Kernel Principal element Analysis (KPCA) and Support Vector Regression. KPCA scale back the dimension and SVR to built a decision system. KPCA would handle the non-linear method. The two form of kernel were used in SVR model such as polynomial kernel and the Gaussien kernel.

[9] proposed new approach based on the hybrid approach of opinion mining and clustering techniques. The objective was to predict the stock movement of National Stock Exchange (NSE) exploitation the opinion mining and clustering methodology. Taken the historical stock values from varied organization with the utmost capitalization and therefore the general sentence of the organization was consolidated into the stock prediction model. Collecting the opinion of shareowner were extracted and cluster the related opinion using the agglomeration(clustering) algorithms. The two output are generated, one from agglomeration based prediction and second output from sentimental analysis.

[10] proposed new approach of stock price prediction using Hybrid approach of rule based algorithm and financial news. Predicting the closing price of the BSE (Bombay Stock Exchange). Data processing techniques to

contend with the missing information values and enhance the feature of interest by including the moving average and relative strength index. Pattern matching utilized to generalize the similar section by ant colony optimization rule, that facilitate to predict the future movement then index calculated.

[11] proposed the work based on predicting Stock Market Index Movement by Optimized Artificial Neural Network Model. The author defined the prediction of next day's value exploitation by optimized artificial neural network (ANN) model. Optimize the prediction accuracy by the genetic algorithms (GA). ANN uses the Back Propagation (BP) algorithmic. The network was initially allotted with the initial weight and reason. Compare the ascertained result with the target result. The weight and bias were re-adjusted using the gradient technique of backpropagated. The BP algorithm had significant drawbacks, they're slowness in convergence associated an inability to escape local optima and dimensionality of the data was additionally higher. The GA was employed to optimize the weight and bias of the neural network. The optimization method by GA is costly, so would use the likelihood or probability model in future and additionally to propose an investment strategy (portfolio) supported the prediction outcomes.

[12] The author proposed the event of the exchange forecasting model based on neural network. The baseline neural network was designed by the generalized feed forward model. The performance of the neural network by feed forward methodology was evaluated using the capitalization stock of the varied industries with the multi-layer perceptron. The result found that the network with a lot of layer or the perceptron given the great result once compared to the baseline neural networks. The network was additionally found to be lacking in assigning the weight and the bias. So, the forecasting model would improved by introducing the enhanced technique in a sequent order.

III. METHODOLOGY

A.Data collection and selection.

The data collection involves the selection of applicable data for prediction. The data employed in the proposed method was collected from certified source. The data of the proposed system is historical data of the S&P 500 ETF that is collected from yahoo.finance.com. It is daily data for S&P 500 Index stock. The parameters used as an indicator are the price of the stock at the beginning of the day (open), the highest price of the stock during the day (high), the lowest price of the stock during the day (low) and the price of the stock at the end of the day (close). The closing price of the stock in the following day is used as output. The dataset covers the time from November 2008 up to 7th November 2019. The dataset is 2770 records. 1939 records are used in training the network while 831 records are used in testing the network.

B. Data normalization

The data quality is the most significant issue for prediction. The data might contain the missing values and mismatched samples, the prediction accuracy can dissent while not preprocessing them; so it's vital to preprocess the data. The data employed in the proposed methodology contains concerning 2770 trading days. Normalization is that the method of scaling the data values between the range of zero to one. The other significance of the normalization method is to make the data values to have the same scales, so each values is taken as equal important.

Proposed system uses the min-max normalization. Min-max normalization is the strategy which linearly transform the entire values between the range of 0 to 1.

The min-max normalization is given by equation (1)

Norm_feature1(i) = (feature1(i) - min(feature1)) / (max(feature1) - min(feature1)) → (1)

Where,

Norm_feature1 = calculated normalized value of the feature1 within range of 0 to 1.

min(feature1) = minimum value in the feature1 max(feature1) = maximum value in the feature1

C. Prediction algorithm

1. SVR Algorithm

In the existing system Support Vector Regression is used for predicting the daily direction of stock. SVR is the algorithm used for non-linear time-series prediction. The SVR contains three parameters like ε, C and the gamma. ε defines the loss function that permit the appropriate degree of error. Additionally, known as the hyper plane. The SVR uses the two hyper plane lines. The general regression model is that the function has the equation of the line. However, the utilization of SVR uses the equation of the line and to fit the training data with the value between -ε to +ε that is termed the loss function.

2. ARIMA Model

In the proposed system ARIMA model is an efficient technique of forecasting the stock exchange daily returns with the short-term. The model predicts the future price value from the previous price. The model contain threshold like p, q and d. Therefore, it form the linear combination like equation (2)

$Pt = \alpha 0 + \alpha 1\ Pt\text{-}1 + \ldots \alpha p\ Pt\text{-}p + \varepsilon t - \beta 1 \varepsilon t\text{-}1 - \ldots\ldots - \beta q\ \varepsilon t\text{-}q \rightarrow (2)$

Where,

Pt – the real price of the stock at the time 't'.

α and β are the coefficients.

ε are the random error at 't'.

p and q are the auto regression and moving average coefficients.

3. Artificial Neural network

The regression type technique the Multi-Layer Perceptron (MLP) is employed. MLP has three layer like input, hidden and output layer. The input value multiplied with weight and fed as input with the addition of bias to the primary layer known as input layer. The output of the primary layer is created by applying the transformation function. To train the data,Levenberg-Marquardt (LM) training algorithm is employed.

IV RESULT AND IMPLEMENTATION

SVR:

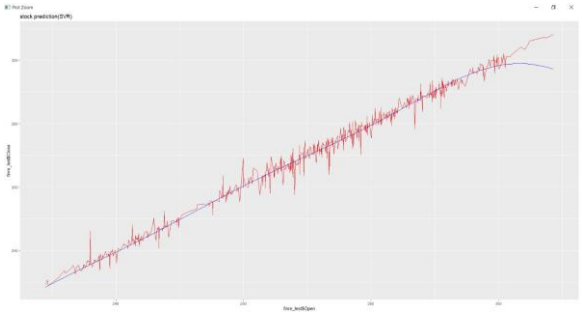

Fig 1: Output graph of actual and predicted price by SVR model.

Fig 1 shows the output graph of the actual and therefore the expected price. The red line within the graph show the actual price of the stock and the blue line outline the predicted values. The parameter are tuned by l-fold cross validation method to reduce the error and to estimate the percentage of the error rate. Finally by this method the percentage of error rate is calculated as 1.99.

ARIMA:

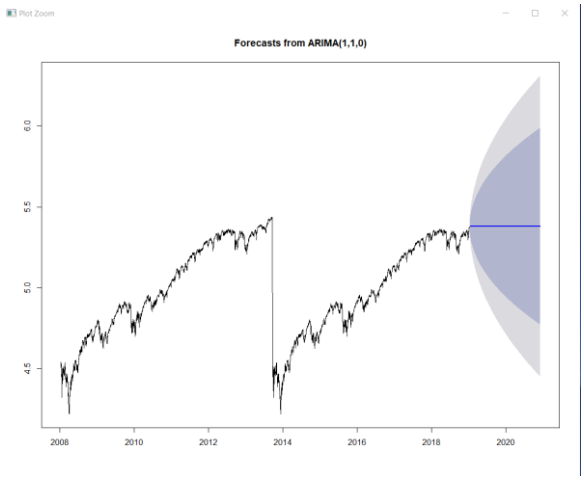

Fig 2: Output graph of forecast data by ARIMA

Primary steps to find the order of the parameter "p" and "q" to determine the best fit ARIMA model, using the auto correlation and partial auto correlation function. From the correlation function, the best fit ARIMA model of the

proposed system is (1,1,0) and by this best fit the future value is forecasted. The percentage of error rate is found to be 0.0635 using the cross- validation method.

Neural network:

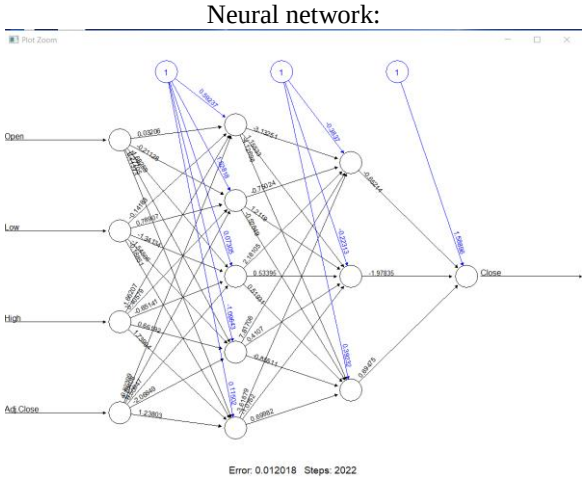

Fig 3: Structure of the proposed neural network

Fig 3 visualizes the structure of the proposed neural network. The model has two hidden layers. Number of neuron in first hidden layer is 5 and that of second hidden layer is 3. The black line shows the connection of the input values with weights. The blue line determines the bias term.

The closing price of the test dataset is predicted using the network model by predefined function called compute(). The predicted close price compared with the actual price. The cross-validation method is used to evaluate the percentage of the error rate.The percentage of error rate is found to be 0.012.

In our implementation the stock price future prediction is done using three machine learning algorithms such as SVR, ARIMA and neural network. The optimized algorithms are selected by evaluating the percentage of error rate. Cross-validation is the optimized technique to better estimate the test error of a predictive model. RMSE is the standard deviation of the prediction error. Prediction error is the measure of difference between the observed and the target output.

Algorithm	% of error rate
SVR	1.99
ARIMA	0.068
Neural network	0.012

Table 1: Percentage of error rate

From the table the neural network algorithm produces the less error rate compared to ARIMA and SVR.
Thus, for better prediction with high accuracy the neural network algorithm is the optimized one.

V.CONCLUSION AND FUTURE WORK

The proposed system using Multilayer feed forward neural network and ARIMA is found to be more accurate in predicting the stock market daily returns compared to the existing SVR algorithm. The method helped in gaining the more profit for the companies and guide the trader in investing in the market. On further improved with an algorithm like Principle Component Analysis (PCA)- Deep Neural Network (DNN) algorithm to predict the stock returns.

REFERENCES

1. Bruno Miranda Henrique, Vinicius Amorim Sobreiro and Herbert Kimura “Stock price prediction using support vector regression on daily and up to the minute prices” the journal of finance and data science, Vol 4, No. 3, September 2018.
2. SachinSampatPatil , Prof. Kailash Patidar, Asst. Prof. Megha Jain “A Survey on Stock Market Prediction Using SVM” International Journal of Current Trends in Engineering & Technology, Vol 2, No.1, February 2016.
3. PhayungMeesad and Risul Islam Rasel “Predicting stock market price using support vector Regression”, International Conference on Informatics, Electronics & Vision, At Bangladesh University of Engineering and Technology, May 2013.

4. Ray Lekhani "Stock Prediction using Support Vector Regression and Neural Networks", International Journal of Advance Research, Ideas and Innovations in Technology, Vol 3, No. 6, 2017.
5. Ayodele A. Adebiyi, Aderemi O. Adewumi, Charles K. Ayo "stock Price Prediction Using the ARIMA Model", 16th International Conference on Computer Modelling and Simulation, At Cambridge University, United Kingdom, March 2014.
6. Prapanna Mondal, Labani Shit and Saptarsi Goswami "Study of effectiveness of time series modeling (arima) in forecasting stock prices", International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.4, No.2, April 2014.
7. Jie wang, Jun wang "Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks, Science direct, Elsevier- neurocomputing, Vol 15, May 2015.
8. AnassNahil, AbdelouahidLyhyaoui "Short-term stock price forecasting using kernel principal component analysis and support vector machines: the case of Casablanca stock exchange", International conference on intelligent computing in data science, Elsevier, Vol 127, 2018.
9. Vivek Singh Rajput, Sarika Sanjay Bobde "Stock market prediction using hybrid approach" International Conference on Computing, Communication and Automation (ICCCA), 2016.
10. JageshwerShriwas, Dr. Samidha Dwivedi sharma "Stock price prediction using Hybrid approach of rule based algorithm and financial news" International journal of computer technology and applications, Vol 5, 2014.
11. Qiu M, Song Y "Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model", PLOS ONE, journal pone, May 2016.
12. Sriram Lakshminarayanan "An integrated stock market forecasting model using neural networks", International journal of business forecasting and marketing intelligence, Research gate journal, 2008.
13. Hiransha M, Gopalakrishnan E.A, Vijay Krishna menon, soman K.P "NSE Stock Market Prediction Using Deep-Learning Models" International conference on computational intelligence and data science, Elsevier journal, 2018.
14. Kang Zhang, Guoqiang Zhong, Junyu Dong, Shengke Wang, Yong Wang "Stock Market Prediction Based on Generative Adversarial Network", International conference on identification, information and knowledge in the internet of things, Elsevier journal, 2018.
15. Wen-Chyuan Chiang, David Enke, Tong Wu, Renzhong Wang "An adaptive stock index trading decision support system", Conference on Expect system with application, Elsevier journal, vol 59, October 2016.