iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VII **Month of publication:** July 2020

DOI: http://doi.org/10.22214/ijraset.2020.7016

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VII July 2020- Available at www.ijraset.com

Predicting and Visualizing Financial Time Series using Machine Learning Techniques

Sneha S.R[1], Sneha Daniel[2], Anitha E[3]
[1,2,3]*Department of Computer Science and Engineering, St. Joseph's Institute of Technology, OMR, Chennai-600119, India*

Abstract: ***The prediction of an exchange direction may function an early recommendation system for short-term investors and as an early financial distress warning system for long-term shareholders. Forecasting accuracy is the most important thing about selecting any forecasting methods. Research efforts in improving the accuracy of forecasting models are increasing since the last decade. The suitable stock selections those are suitable for investment could be a very difficult task. The key factor for every investor is to earn maximum profit on their investments. In this paper Support Vector Machine (SVM) is proposed. SVM could be a specific kind of learning algorithm characterized by the capacity control of the choice function, the employment of the kernel functions and therefore the scarcity of the answer. In this paper, there's an investigation to predict the financial movement with SVM. To gauge the forecasting ability of SVM, a comparison is finished with Decision Trees and for breaking the barrier of the issue in understanding the financial data, they are visualised using data visualisation techniques and R. Statistical data visualisation plays a crucial role in the sector of business higher cognitive process. With enormous amount of complex data from cloud and business requirement, number of graphs needed for higher cognitive process increased many folds. These methods are applied on 2 years of information retrieved from yahoo finance. The results are wont to analyse the stock prices and their prediction accuracy.[10]***

Keywords: Support Vector Machine, Decision Tree, Yahoo Finance, Data Visualisation, R

I. INTRODUCTION

Prediction is performed on statistic within financial contexts, specifically the return of a private stock from companies listed on the Yahoo Finance. the value of a private stock is analysed to predict changes that may happen so a choice is taken. To perform an adequate forecasting analysis historical prices, from companies listed within the securities market, are obtained through sites like Yahoo Finance.

This application will predict the stock prices for the subsequent trading day. After predicting the subsequent trading day, these predictions are visualised using R-GGPLOT2. This project implements using two different machine learning techniques. One using Support Vector Machine (SVM) and also the second using Decision Trees. Both the systems are trained using 75% of two years historic data so test the model to test which system yields better output using the remaining 25% of historic data. This solution uses a special algorithm and different technique to perform prediction. This uses Support Vector Machine with C type classification and Radial Basis Function (RBF) kernel.[10] C may be a regularization parameter that controls the trade-off between the achieving an occasional training error and an occasional testing error that's ability to generalize the classifier to unseen data. Decision tree could be a form of supervised machine learning. The tree may be explained by two entities, namely decision nodes and leaves. The leaves are the choices or the ultimate outcomes. the choice nodes are where the information is split. There are two main styles of decision trees: Classification trees (Yes/No types) and Regression tress (Continuous data types). This proposed approach makes use of Classification tree.

Input data is taken from the Yahoo finance by using stock's ticker symbol and so the system will take last 2 years stock data of the company using quant-mod package in R. The system is implemented using two different machine learning techniques. One using Support Vector Machines and also the second implementation using Decision Trees.

both the systems are trained using 75% of two years of historic data so test our model to check which systems yield better output using the remaining 25% of historic data. Our solution uses a special algorithm and different technique to perform the prediction. We are using Support Vector Machines with C type classification and Radial Basis Function (RBF) kernel. It uses SVM and Decision Trees which have better performance than Neural Network. Moreover, using SVM will takes away the burden of matching this price pattern with historic patterns and also SVM trains faster than a NN and includes a lower computational cost. Also, other solution uses the financial data because it's without using any indicators, whereas our solution uses many indicators like EMA, RSI, MACD, SMI, CCI, ROC, CMO, WPR and ADX to urge better results.[10]

Two algorithms, decision trees and SVM were employed in this study and only 1 dataset from Yahoo Finance was applied to coach and test the models. The system can only predict the direction(up/down) for the following trading day. within the future, SVM regression model are employed in order to predict the worth moment for the long run and also the results are going to be compared with other data processing techniques by applying different dataset from different stock market index.

Top of Form

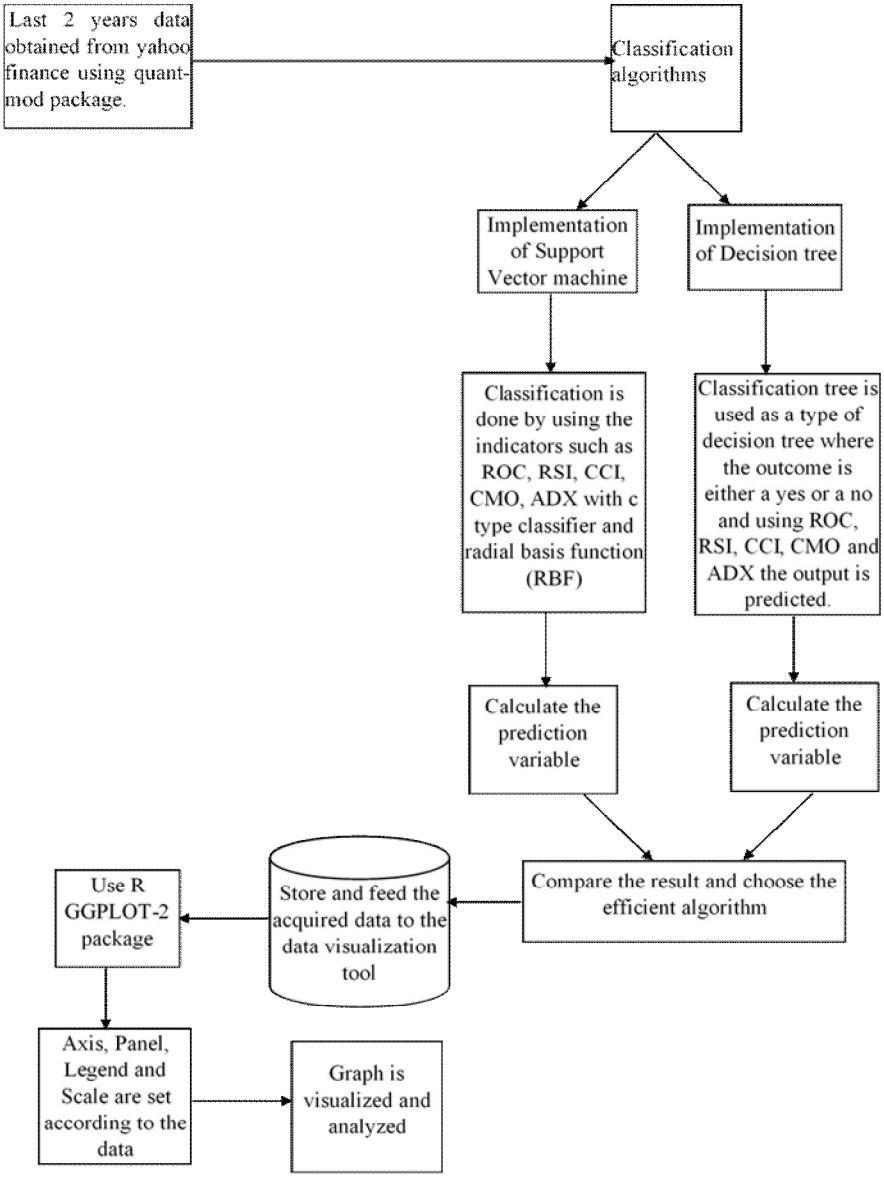


Figure 1: System Architecture for Predicting and Visualising Financial Time Series Using Machine learning techniques

II. METHODOLOGY

A. Collection of Input Data

Input data is taken from the Yahoo finance using the following steps:

1) *Step 1:* This project is considering Standard and Poor (S&P) of 500 companies. The list of S&P's can be obtained from Wikipedia.
2) *Step 2:* Use Stock's ticker symbol to get the data from Yahoo Finance.
3) *Step 3:* System will take last 2 years of the stock data of the company using quant mod package in R.
4) *Step 4:* Further the data is divided into two parts: training data and testing data where 75% of the data is used for training and 25% is used as testing data.

B. Generation of the Output

Following steps is performed to generate the output:

1) *Step 1:* Using Stock ticker's symbol get the last 2 years of data.
2) *Step 2:* Provide the data to the system.
3) *Step 3:* Train the system.
4) *Step 4:* System will predict the output.

C. Data Visualization

Define all specifications required for generating graphs in XL file.

1) *Step 1:* Prepare the data file according to XL file.
2) *Step 2:* Specify index range or specific index numbers in R-GGPLOT 2 function.
3) *Step 3:* Users can select any type of graph such as scatter, series, mean or box plots.
4) *Step 4:* The R function runs in a loop taking specifications mentioned in the XL file and data from data.

D. Work Flow

The entire system would be implemented in Python/Java and R language using open source libraries.

1) *Step 1:* System Architecture diagram roughly summarizes the entire process that takes place in the prediction of financial time series and the machine learning techniques.
2) *Step 2:* Obtaining a set of data for the last 2 or 3 years from the yahoo finance or Kaggle sites. The data is obtained either in the excel or csv format.
3) *Step 3:* We will be using 2 algorithms for the prediction process to compare and select the best algorithm.
4) *Step 4:* The Support Vector Machine (SVM) & Decision trees are the 2 algorithms that will be used here. Support-vector machines (SVMs, also support-vector networks [1]) are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis.
5) *Step 5:* There are certain indicators that must be calculated accordingly to obtain the better algorithm, such as ROC, RSI, CMO, ADX, EMA and few other.
6) *Step 6:* Calculating the prediction variables and obtaining their output, Build their decision tree. Here we also know that 25% of the data is used for testing purposes and the other 75% is used for training of the data.

E. Decision Tree Algorithm

1) *Step 1:* Get the desired data [Date, Open, High, Low, Close, Volume, Adjusted]
2) *Step 2:* Calculate all the desired indicators:

a) *RSI (Relative Strength Index):* Assuming a five-period RSI, a zero RSI value means prices has moved lower all five-period. there have been no gains to live. RSI is 100 when the typical loss equals zero. this suggests prices moved higher all five-periods. there have been no losses to live.
b) *EMA Crossover:* it's calculating the movement of EMA short period (Opening price)- EMA large period (EMA over 5 days).
c) *CCI:* An oscillator utilized in technical analysis to assist determine when an investment vehicle has been overbought and oversold. The Commodity Channel Index (CCI), first developed by Donald Lambert, quantifies the connection between the asset's price, a moving average (MA) of the asset's price, and normal deviations (D) from that average.
d) *ROC:* the value rate of change (ROC) may be a technical indicator that measures the proportion change between the foremost recent price and therefore the price "n" periods within the past.[6]
e) *CMO:* A Collateralized Mortgage obligation (CMO) could be a sort of security within which principal repayments are organised in line with their maturities and so into different classes supported risk. A Collateralized Mortgage Obligation (CMO) could be a special purpose entity that receives the mortgage repayments and owns the mortgages it receives income from (called a pool).
f) *WPR:* Syndrome R (WPR) in technical analysis could be a momentum indicator measuring overbought and oversold levels, almost like a stochastic oscillator. it had been developed by Larry Williams and compares a stock near the high- low range over a specific period time, usually 14 days.
g) *ADX:* The typical Directional Index (ADX) is an indicator utilized in technical analysis as an objective value for the strength of the trend. ADX is non-directional so it'll quantify a trend's strength no matter whether it's up or down. ADX is typically plotted

during a chart window together with two lines referred to as the DMI (Directional Momentum Indicators). ADX derived from the connection of DMI lines.

3) *Step 3:* Calculate the prediction variable (Up/Down).
4) *Step 4:* Build the classification decision tree from the information calculated from the above steps. the result was a variable like 'fit' or 'unfit'. Here the choice variable is Categorical. the simplest decision tree algorithm is that the ID3 algorithm. ID3 stands for Iterative Dichotomiser 3.[10]
5) *Step 5:* Prune the tree to get rid of any overfitting of the info.
6) *Step 6:* Give test data to trees.

F. Support Vector Machine Algorithm

This is a C- classification Support Vector Machine with RBF- kernel.

1) *Step 1:* Read the required data [Date, Open, High, Low, Close, Volume, Adjusted]
2) *Step 2:* Get the Historical data from the Yahoo finance using quant-mod package. This quant-mod package for R is designed to assist the quantitative trader in the development, testing and deployment of statistically based trading models.
3) *Step 3:* Calculate all the required indicators: RSI, EMA Crossover, CCI, ROC, CMO, WPR and ADX
4) *Step 4:* Provide the data collected from yahoo finance as 75% training and 25% as testing.
5) *Step 4:* Provide the data from the above steps to train SVM (RBF, C=1, gamma= ½)
6) *Step 5:* Provide the test data and display the results.
7) *Step 6:* Compare the output from step 5 and 6 and show the observations.

G. Visualisation Using R-GGPLOT 2

Using GGPLOT- 2, we can represent the output in a graphical form.

1) *Step 1:* Obtain the plot specifications of the data in XL format.
2) *Step 2:* Get the data file required for the plotting of the graph.
3) *Step 3:* R GGPLOT-2 is a statistical programming language package for R. Using the R function, we can plot the graph.
4) *Step 4:* We can select any type of graph such as scatter, series, mean and box plot we can represent the data obtained in a graphical format.3
5) *Step 5:* Output is obtained

III.RESULTS AND DISCUSSION

A. Decision Tree Implementation

The implementation of the algorithm is done to generate the decision tree, confusion matrix, ROC curve, area under curve, accuracy and root mean square error and the following figure 2 represents a decision tree of the company Apple whose historical stock data has been collected from Yahoo Finance. [8]

The rectangles in the tree are the nodes of the decision tree, where the topmost node is the root node of the tree. The node values are depicted by the prediction variable, either it is a up or a down. The variables used in the tree construction are CCI, EMAcross, ROC, RSI3 and WPR.

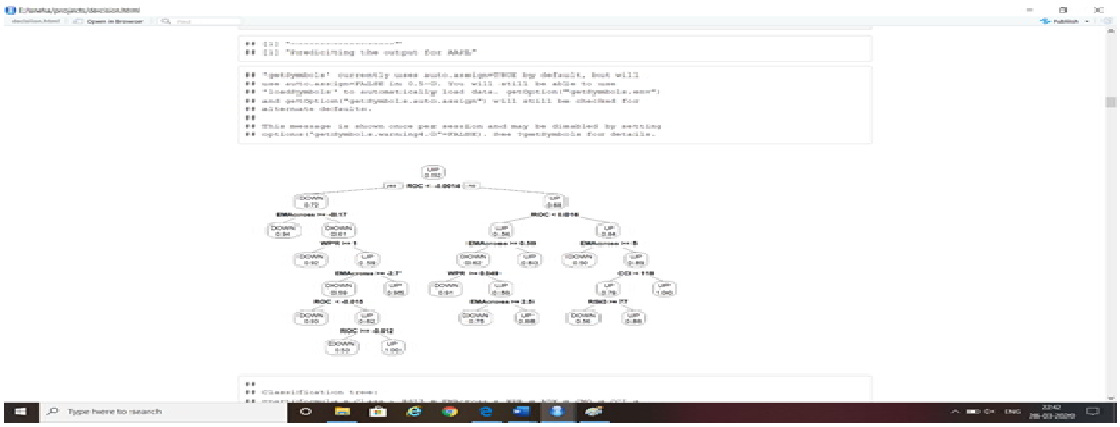


Figure 2: Decision Tree for Apple

B. Support Vector Machine Implementation

The implementation of the algorithm is done to generate the confusion matrix, ROC curve, area under curve, accuracy and root mean square error and the following figure 3 represents the Confusion Matrix, Support Vector Machines, Accuracy and Root Mean Square Error of the company Apple whose historical stock data has been collected from Yahoo Finance.

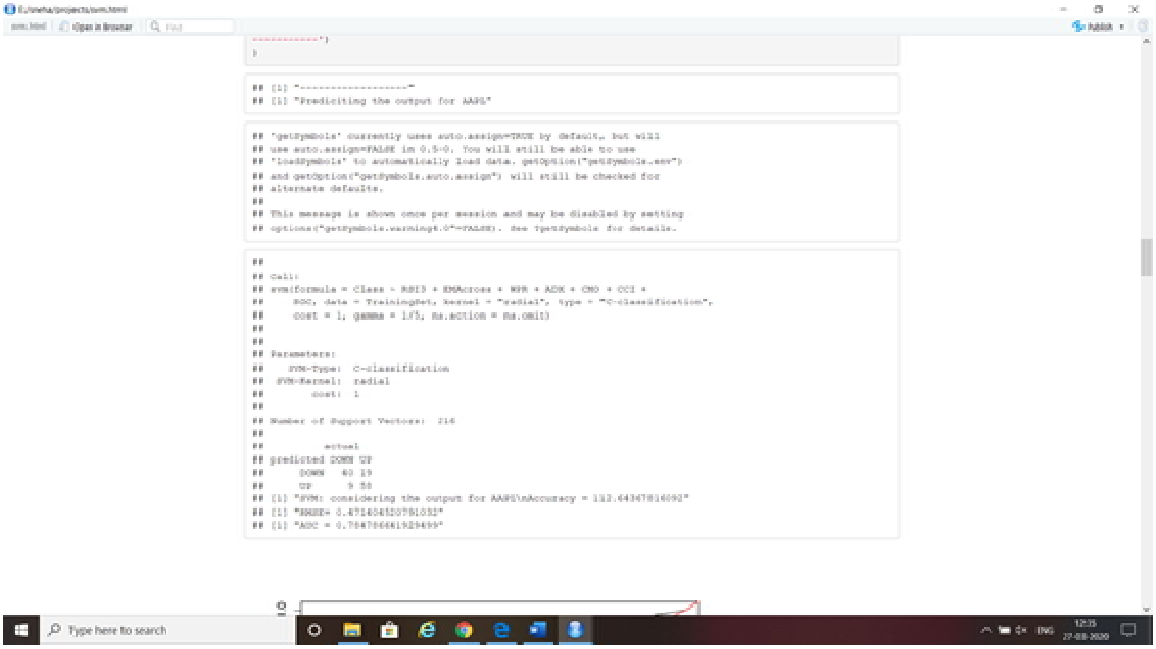


Figure 3: Generation of Support Vector Machine

The Confusion Matrix is a 2*2 matrix for 2 classes (Positive and Negative). The classes are as follows; True Positive, False Positive, False Negative and True Negative. The outcomes for Apple are shown in figure 3 with values 40 for True Positive, 19 for False Positive, 5 for false Negative and 58 for True Negative.

This also shows that the system uses 216 Support Vector Machines for the entire implementation. C-Classification is the type of SVM, and radial is the kernel of the SVM.

C. Data Visualisation

The close analysis of the output of Decision trees and SVM algorithm reveals that the SVM gives better results than decision trees.[9] Figure 4 shows the performance analysis for the 7 companies, that were tested and trained. Two years of knowledge is downloaded from Yahoo Finance, of which 75% is employed to coach the system and therefore the remaining 25% is employed for testing. The RSI could be a short-to intermediate-term indicator when used in line with its classic application, which is most typically a 14-day timeframe.[10] It generates values that fall on a scale from 0 10 100, with high and low levels (overbought/oversold) considered to be at 70 and 30, respectively. It compares the magnitude of recent gains to recent losses to measure the worth momentum of monetary assets, including stocks, commodities and financials.

Traders can use RSI to come up with trade signals, assess sentiment or as a complement to other analysis tools. Many indicator functions and their permutations were tested while training and testing the system. Of all the indicator function tested, those which gave the most effective prediction result were selected. The system performs all right for the prediction. With the chosen combination of indicator variables, we were able to get the maximum accuracy of 87.4% for Facebook stock.[1]

The prediction accuracy depends on the choice of the prediction variable. We tried multiple indicator variables and their permutation and selected the permutation which gave best result.

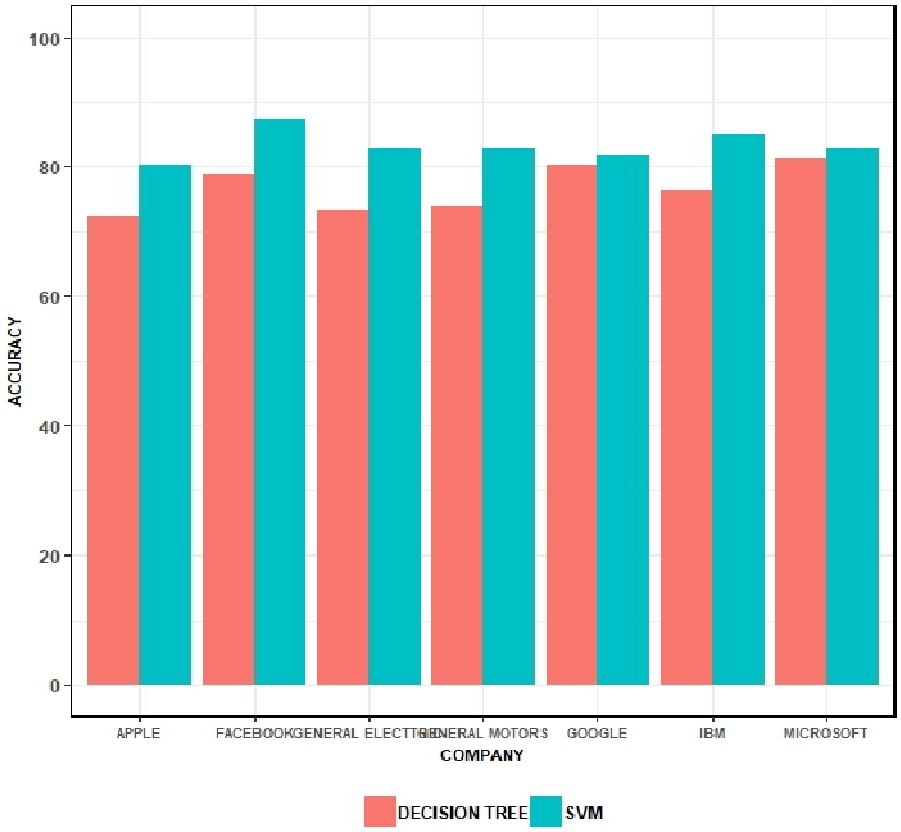


Figure 4: Visualisation of the result

IV. CONCLUSION

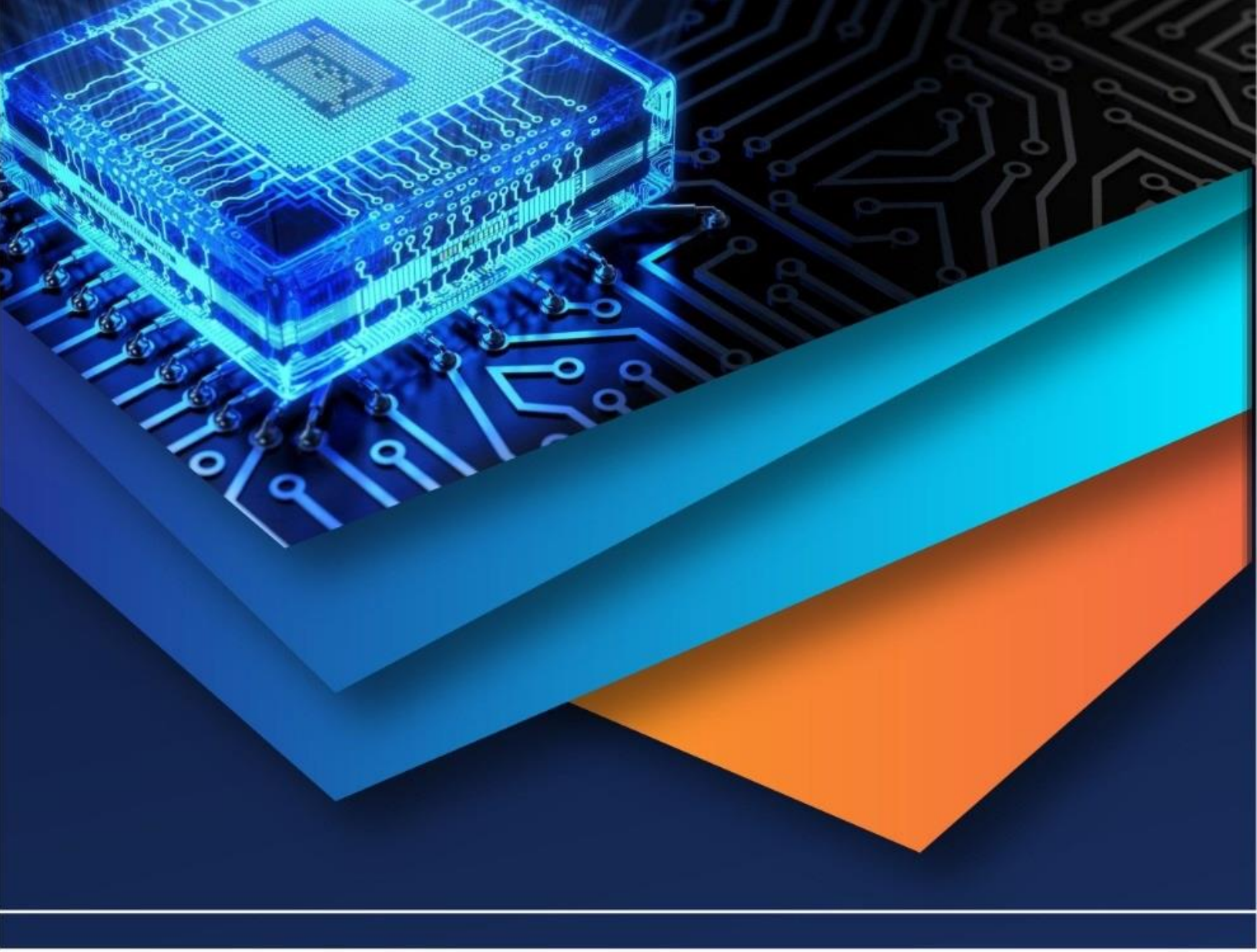
In this paper, we studied the utilization of decision trees and support vector machine to see the financial movement direction. of those both algorithms, we all know that Support Vector Machine (SVM) gave us better results. SVM could be a promising style of tool for financial forecasting. SVM is superior to other individual classification methods in forecasting daily movement direction. this can be a transparent message for financial forecasters and traders, which might cause a financial gain. However, each method has its own strength and weaknesses. during this model, the principal components identified by the SVM are used together with internal and external financial factors in SVM for forecasting. We also observed that the selection of the indicator function can dramatically improve or reduce the accuracy of the prediction system. Also, a specific Machine Learning Algorithm may be better suited to a specific variety of stock, like technology, whereas the identical algorithm might give lower accuracies while predicting energy stocks. Two algorithms, decision trees and SVM were employed in this study and only 1 dataset from Yahoo finance was applied to coach and test the models. The system can only predict the direction for next trading day. within the future, SVM regression model is employed in order to predict the value moment for the long run and also the results are going to be compared with other data processing techniques by applying different dataset from stock market index. Thus, we are able to conclude that the SVM plays an improved role in achieving an efficient and accurate outcome for the system model. The SVM obtains the information from about 7 companies and together with their indicators it's going to obtain the higher outcome of the system.[5].
The SVM allows the system to play a significant role within the algorithmic process of the Machine Learning Algorithms in their system. The output of the system should be the direction of stock for the following day. except for the higher understanding of the prediction model we are displaying the confusion matrix, root mean squared error and area under curve.

V. ACKNOWLEDGMENT

I would like to thank Dr Neelesh S Upadhaye, Associate Professor, Department of Mathematics, Indian Institute of Technology, Madras, India for providing me all the facilities during my work.

REFERENCES

[1] Dhanya & Janardhanan (2011) "Price Discovery in Indian Futures Market – An Empirical Study", Anvesha, 4 (1), 57-63.
[2] Dr. Prema Chandran (2016) "A Study on the volatility and returns of the Indian banking Sector Index with Reference to NSE Nifty" ISSN: 2321-7782 (Online) International Journal of Advance Research in Computer Science and Management Studies.
[3] M.Gocken, M.Ozcalici, A.Boru and A.T. Dosdogru (2016) Integrating metaheuristics and artificial neural networks for improved stock price prediction. Expert Syst Appl 44:320–331.
[4] Hakob Grigoryan (2017) "Stock Market trend predictions using SVM & Variable selection methods." DOI: https://doi.org/10.2991/ammsa-17.2017.45.
[5] Ji Soo Yi, Youn ah Kang, John T. Stasko and Julie A.Jacko (2007) "Toward a Deeper Understanding of the Role of Interaction in Information Visualization".
[6] Jui-Sheng Chou and Thi-Kha Nguyen, "Forward Forecast of Stock Price Using Sliding-window Metaheuristic-optimized Machine Learning Regression", IEEE Transactions on Industrial Informatics, DOI 10.1109/TII.2018.2794389.
[7] Lin Qian Yu and Feng Shao Rong (2010) "Stock Market Forecasting Research Based on Neural Network and Pattern Matching" DOI: 10.1109/ICEE.2010.490
[8] Mahmoud K. Okasha, "Using Support Vector Machines in Financial Time Series Forecasting", International Journal of Statistics and Applications 2014, 4(1): 28-39 DOI: 10.5923/j.statistics.20140401.03.
[9] Malikkarjunappa and Afsal (2008) "The Impact of Derivatives on Stock Market Volatility: A Study of the Nifty Index" Asian Academy of Management Journal of Accounting and Finance (AAMJAF), 2008, vol. 4, issue 2, 42-66.
[10] Pawan Kumbhare; Rohit Makhija; Hitesh Raichandani "Stock Market prediction", Santa Clara University.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY