iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4251

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: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Survey on Stock Prediction Using Machine Learning Approaches.

Jay Borade[1], Ayush Khanvilkar[2], Rahul Kanvinde[3], Akshay Narkhede[4], Anuj Odedra[5]
[1, 2, 3, 4, 5]*Assistant Professor Information Technology at Fr CRCE, Mumbai University*

Abstract: Over the years stock price prediction has always attracted people investing in share market because of the direct financial benefits. Stock price prediction is a very challenging task because of the volatile nature of stock market. The stock prices are affected due to many reasons like company related news, political, social, economical conditions and natural disasters[1]. An intelligent trader predicts the stock price and buys a stock before the price rises, or sells it before its value declines. There are many models developed for predicting the future price of stock but each one has its drawback. In this paper we analyze the accuracy of Linear Regression, Support Vector Regression and Decision Tree Classifier by comparing the predicted values with the actual values over a period of time.
Keywords: Machine Learning Algorithms, stock prediction, Support Vector Regression

I. INTRODUCTION

The market hypothesis tells us that stock prices are random and unpredictable. Investment banks have used predictive models which use past market data to predict stock prices and market indexes. The services of these banks are not affordable to the common people. The common people who want to invest in stocks via their Demat account are unable to use these services. To invest money in the stock market we need to have an idea whether the prices of stocks are going to increase or decrease in the next couple of days. Several computing techniques need to be combined in order to predict the nature of the stock market.
Machine Learning algorithms can give you patterns and statistics from a dataset. These algorithms help traders to devise a strategy to buy the most profitable stocks. This paper aims at giving a comparative study of various stock prediction algorithms .

II. MACHINE LEARNING APPROACHES FOR STOCK PREDICTION

A. *Data Source*

1) *Environmental Setup:* In order to undertake the experiments and evaluate the results from the experiments, NASDAQ index was selected as our research domain[2]. The Apple stock (AAPL) was selected to analyze the results and make predictions. The historical data from the year 2004 till date was selected as the dataset. 80% of data was used for training and the remaining for testing. The original dataset contains 6 attributes: Date, Open, High, Close, Adj Close and Volume[2]. The goal was to predict the Closing price of Apple stock (AAPL) for 35 days.

```
Date,Open,High,Low,Close,Adj Close,Volume
18-08-04,2.179286,2.275,2.177857,2.267143,1.534757,91163800
19-08-04,2.250714,2.275714,2.168571,2.193571,1.484952,97230000
20-08-04,2.193571,2.213571,2.177857,2.2,1.489305,79195200
23-08-04,2.204286,2.233572,2.185714,2.22,1.502844,63665000
```

Fig 1: The above figure shows the format of historical data selected for Apple stock(AAPL).

2) *Python:* The library scikit-learn is used for model selection and matplotlib for data visualization.

B. *Overview of the Models used:*

1) *Linear Regression:* Linear regression is a method which is used to model a relationship between a dependent variable (y), and an independent variable (x). In this paper, the independent variable is date and the dependent variable will be the price of the stock. Thus, Linear regression analyzes two separate variables in order to define a single relationship.
2) *Equation:*

$$Y = a + bx$$

Where,

Y is the predicted value or dependent variable.
x is the coefficient or independent variable.
a is the y-intercept.
b is the slope of the line.[3][4]

C. Prediction Graph

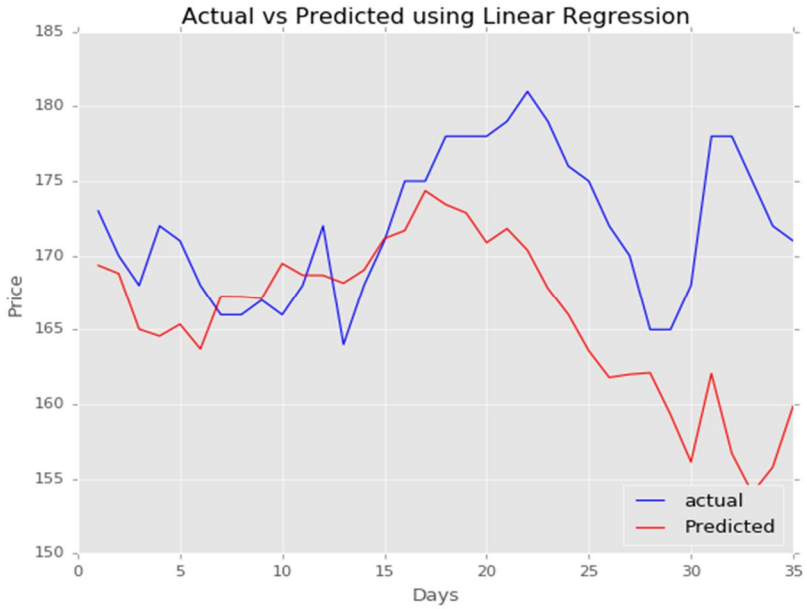


Fig 2: The graph above shows the comparison between the actual value and predicted value for AAPL stock.

The above graph is for the time span of 35 days for both actual and predicted price. The graph shows significant difference between actual and predicted values. After 15 days the predicted and actual value is in the same range which is between 170 and 175. Initially the difference is minimum but as we reach till 30 days massive difference is observed between the values. At the end the predicted value is 160 where as the actual value is 172. Thus we can conclude that Linear Regression is not giving accurate results.

1) Advantages

a) It is highly valuable in economic and business research therefore it is advantageous in stock prediction.

b) Linear Regression analysis is most applied technique of statistical analysis and modeling[5].

2) Drawbacks

a) Linear regressions are meant to describe linear relationships between variables. So, if there is a nonlinear relationship, then you will have a bad model.

D. Support Vector Regression (SVR)

Support vector machines are divided into support vector machines and support vector regression. Originally designed for the classification, but in recent years, support vector regression algorithm in the research also showed an excellent performance. Support Vector Regression is a machine learning tool which can build a regression model on the historical time series to predict the future trend of the stock price[6].

1) SVR Kernel Analysis: After the input selection process, the next step is to apply the kernel function. The output of SVR mainly depends on the kernel selection and it's parameter settings. Here we have used rbf kernel function which produce good output results.

2) Equation

Linear Kernel: $K(X, Y) = X^T Y$

Polynomial Kernel: $K(X, Y) = (\gamma . X^T Y + r)^d , \gamma > 0$

Radial Basis Function (RBF): $K(X, Y) = \exp(\| X-Y \|^2/2\sigma^2)$

Here r, γ, σ are kernel parameters.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

E. Prediction Graph

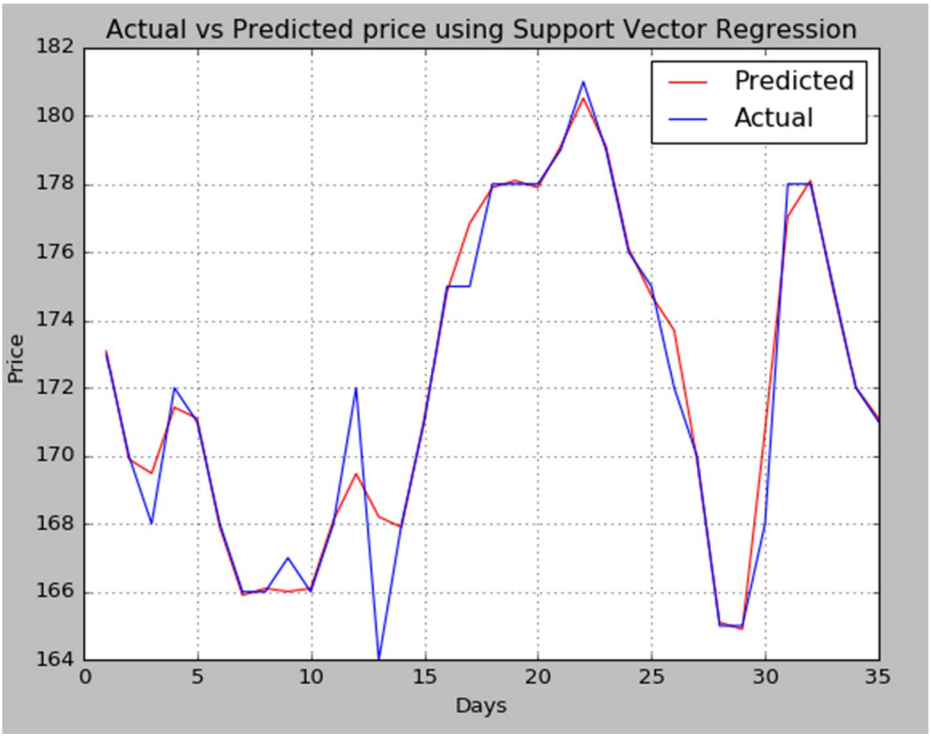


Fig 3: The graph above shows the comparison between the actual value and predicted value for AAPL stock.

The above graph is for the time span of 35 days for both actual and predicted price. For the most part, the graph of actual values and predicted values are overlapping. After 20 days the predicted and actual value is 178 and after 35 days the predicted and actual value is 171. In comparison to other algorithms support vector regression produces accurate results.

1) Advantages

a) One of the advantages of SVR is that it can be used to avoid difficulties of using linear functions in the high dimensional feature space and optimization problem is transformed into dual convex quadratic programs[7].

b) It uses the kernel trick, so you can build in expert knowledge about the problem via engineering the kernel.

2) Drawbacks

a) Serious problem with SVMs is the high algorithmic complexity and extensive memory requirements like speed and size, both in training and testing[8].

3) Decision Tree: A decision tree algorithm performs a set of recursive actions before it arrives at the end result.

ID3: Iterative Dichotomiser3 is widely known as ID3 algorithm. J. Ross Quinlan developed the ID3 algorithm. ID3 is a decision tree generation algorithm. For a given dataset containing attributes and entries ID3 algorithm generates the Decision tree using Shannon Entropy. The algorithm is as follows:

a) Consider a target attribute.

b) Calculate target attribute‟s Entropy

c) Calculate the Information Gain of all remaining attributes in the dataset.

d) Determine the attribute with highest Information Gain as the node.

e) Make the residual dataset by eliminating the selected node.

f) If all the attributes in the dataset are completed go to Step 7, else go to Step 8.

g) Generate the Decision Tree.

h) Continue from Step 2.[9]

4) Equations

$H(S) = - \sum_{i=1}^{n} p(xi) log2p(xi)$

$IG(S,A) = H(S) - \sum_{teT} p(t) * H(t)$

F. Prediction Graph

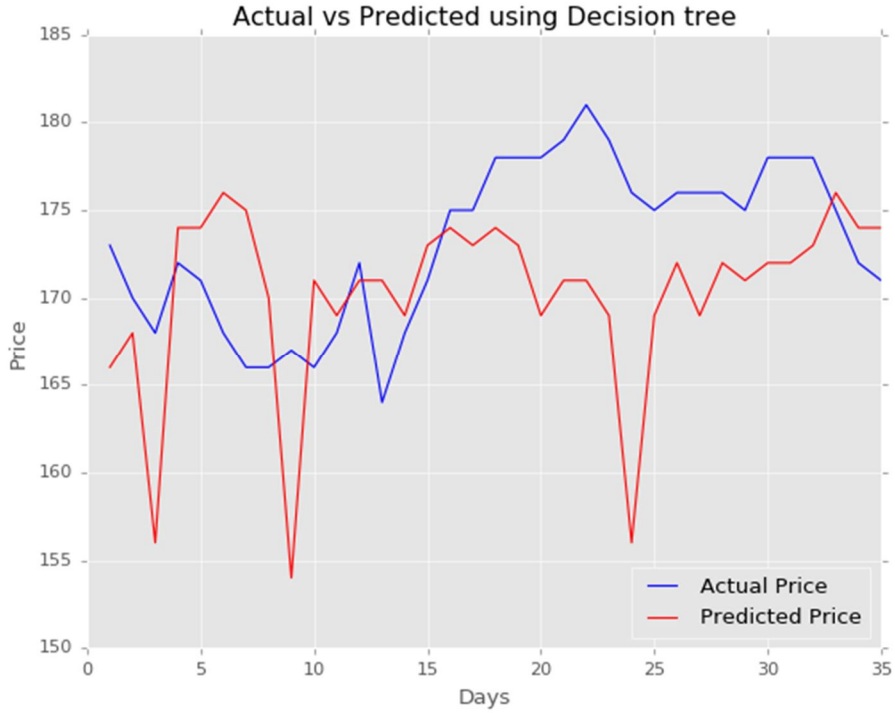


Fig 4: The graph above shows the comparison between the actual value and predicted value for AAPL stock.

The above graph is for the time span of 35 days for both actual and predicted price. After and before the point of 16 days there is minimal overlapping in the graph of actual and predicted data. Apart from the 16 days and 32 days mark there is major difference in the two graphs. Eventually after 35 days the actual data and predicted graph are very close. The actual price is 171 and the predicted value is 174. Thus we can conclude that Decision Tree has performed better than Linear Regression but in overall Support Vector Regression is the better algorithm.

1) Advantages

a) In ID3 algorithm, finding leaf nodes enables test data to be pruned, reducing number of tests.

2) Drawbacks

a) In case of ID3, tree cannot be updated when new data is classified incorrectly, instead a new tree must be generated.

b) Only one attribute at a time is tested for making decision.

III. RESULTS & CONCLUSION

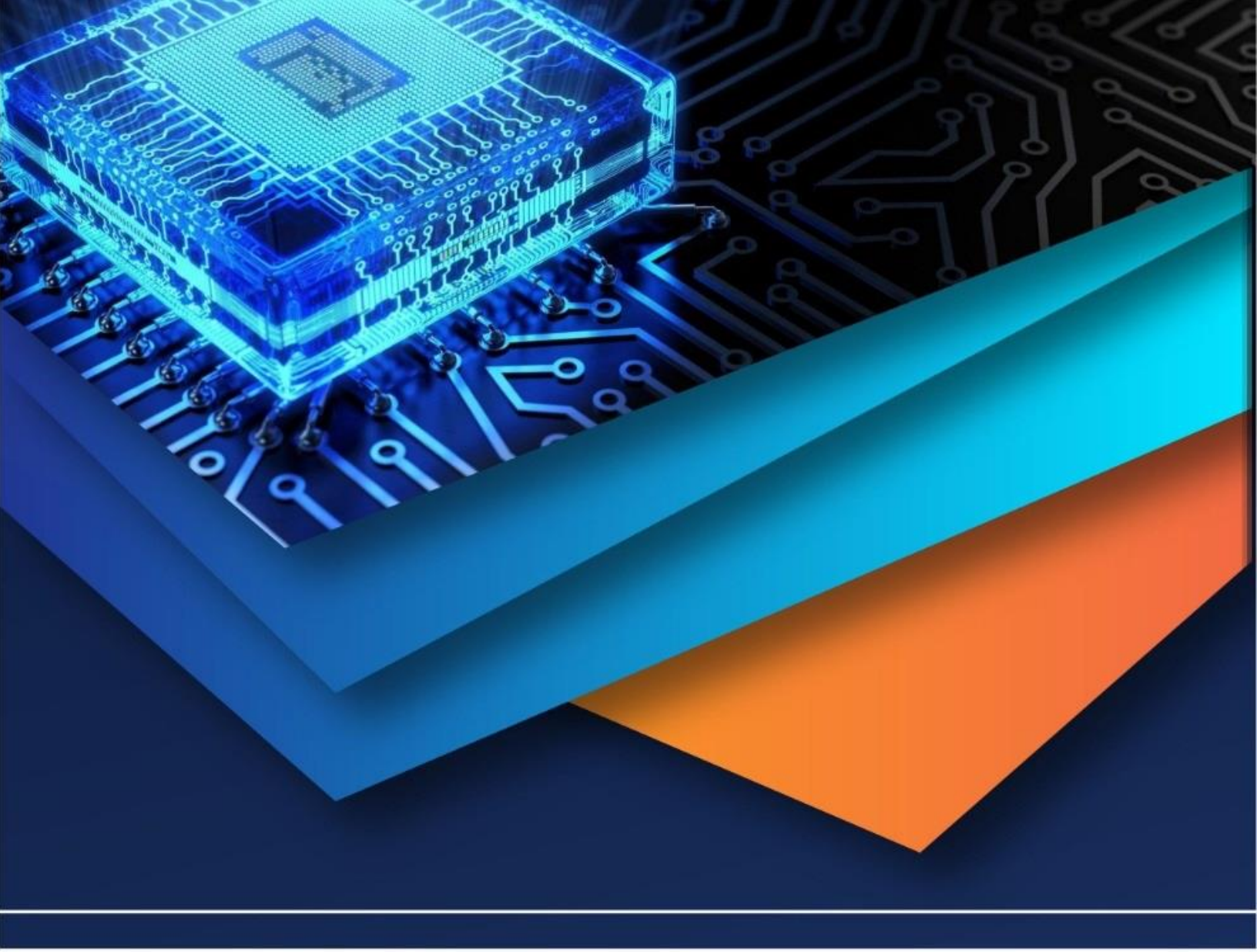
Table I: Actual Versus Predicted Values

	Linear Regression			SVR			Decision Tree		
Days	Actual	Predicted	Accuracy %	Actual	Predicted	Accuracy %	Actual	Predicted	Accuracy %
5	171.0	165.6	96.84	171.0	171.2	99.88	171.0	173.9	98.30
15	171.5	170.9	99.65	171.5	171.5	100	171.5	172.5	99.42
25	175.0	163.8	93.60	175.0	174.5	99.71	175.0	168.9	96.51

The results from Table 1 shows that SVR algorithm gives the best predicted values for the AAPL stock. In this paper we illustrated the comparison of various machine learning algorithms for stock price prediction. This paper compared the models using the actual, predicted values and the percentage accuracy. Thus, the comparative analysis conclude that SVR is the best algorithm for stock prediction because it has got maximum accuracy of around 99.86%.

REFERENCES

[1] Dr. P. K. Sahoo & Mr. Krishna charlapally, "Stock Price Prediction Using Regression Analysis", International Journal of Scientific & Engineering Research, Volume 6, Issue 3, March-2015.
[2] Phayung Meesad & Risul Islam Rasel, "Predicting Stock Market Price Using Support Vector Regression", International Conference on Informatics, Electronics & Vision (ICIEV 2013). May 2013.
[3] http://programmingforfinance.com/2018/01/predicting-stock-prices-with-linear-regression.
[4] https://www.investopedia.com/articles/trading/09/linear-regression-time-price.asp.
[5] http://www.info4mystery.com/2016/07/advantages-of-regression-analysis.html
[6] Yaqing Xia, Yulong Liu & Zhiqian Chen,"Support Vector Regression for prediction of stock trend, Information Management", Innovation Management and Industrial Engineering (ICIII), 2013 6th International Conference on 23-24 Nov. 2013.
[7] http://kernelsvm.tripod.com/
[8] http://www.svms.org/disadvantages.html
[9] I. S. L. Sarwani, N. Siva Nandhalahari, & S. Sobha Sri, "Comparative Analysis of Id3 and Naïve Bayes Algorithm on Stock Market Prediction", International Journal of Computer Applications (0975 – 8887) Volume 143 – No.2, June 2016.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY