# Bangladeshi Stock Price Prediction and Analysis with Potent Machine Learning Approaches

**Submitted By**

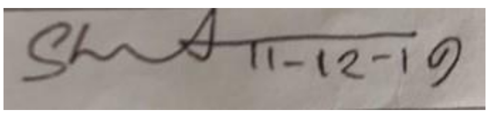

Student Name: Sajib Das
ID: 161-35-1546
Department of Software Engineering
Daffodil International University


**Submitted Date: 01/12/2019**


**Daffodil International University**
**Dhaka, Bangladesh**

## DECLARATION

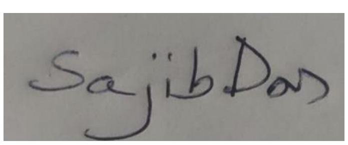
I hereby declare that, this project has been done by me under the supervision of **Supervisor MD. Shohel Arman, Lecturer, Department of SWE** at Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree.




**Supervised by:**

**MD.Shohel Arman**

**Supervisor Name**
**Designation**
Department of Software Engineering
Daffodil International University




**Submitted by:**

**Sajib Das**

**Student Name**
**ID:161-35-1546**
Department of Software Engineering
Daffodil International University

**ACKNOWLEDGEMENT**

I am thankful to my creator for allowing me to complete this research and learn so much. I am grateful to my supervisor of research, Md. Shohel Arman, for providing careful guidance starting from selecting the research scope to successfully finalizing the research work. I would also like to thank Md. Sanzidul Islam, Lecturer, SWE for his valuable comments which was always insightful. Finally, I want to express my gratitude to Professor Dr. Touhid Bhuiyan, Head of the Software Engineering faculty, for inspiring us in all means. I am also thankful to all the lecturers, Department of Software Engineering who sincerely guided me at my difficulties. I am grateful to my parents for their unconditional support and encouragement. I am thankful to my friends who supported me throughout this venture.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| Abbreviation | Explanation |
| --- | --- |
| SVR | Support Vector Regression |
| LR | Linear Regression |
| RBF | Radial Basis Function |

# ABSTRACT

Stock price forecasting, is one of the most significant financial complexities, since data are not reliable and noisy, impacting many factors. This article offers a learning machine model for the stock price prediction using Support vector machine-Regression (SVR) with two different kernels which are Radial Basis Function (RBF) and linear kernel. This study shows the Prediction and accuracy comparison between Support Vector Regression (SVR) and Linear Regression (LR) and also the accuracy comparison for different kernels of Support Vector Regression (SVR). The model has used sum squared error (SSE) to determine the accuracy of each algorithm; which has shown significant improvement than the other studies. This analysis is conducted on the price data of about five years of Grameenphone listed on Dhaka Stock Exchange (DSE). The highest accuracy was found with Linear Regression model in every case with the highest accuracy of about 97.07 percent followed by SVR (Linear) model and SVR (radial basis function) model with the highest accuracy rate of about 97.06 and 96.82 percent. In some cases, the accuracy of SVR (radial basis function) was higher than SVR (linear). But it was the Linear Regression which had the highest accuracy of all in every case.

# CHAPTER 1

# INTRODUCTION

## 1.1 Background

The stock market refers to the selection and exchange of stocks in which common shares of public companies are bought, exchanged and issued. The shares of the company are all shares in which the company's ownership is split. In proportion to the number of shares in total, a single share of the stock represents fractional ownership. The prices of stocks shift with market forces every day. This means that share prices are changing due to demand and supply. If more people would like to purchase a stock (demand) than sell it (supply), the price will increase. On the other hand, if more people wanted to sell a stock than purchase it, there would be more supply than demand, and the price would fall [12]. In other words, the more a stock has been transacted the more is valuable.

For years, stock price forecasts have focused because they can generate substantial profit. Investors have tried to predict the trends using various methods and bet in the markets. Technical analysis like RSI (Relative Strength Index), MFI (Money Flow Index), MACD (Moving Average Convergence/Divergence) etc. [1] and fundamental analysis like Investor sentiment analysis [2], EPS (Earnings per Share), Net asset value etc. are used in analyzing the trends of stock prices. Different machine learning algorithms have also used in forecasting stock prices. Tough it is a tricky task to as predict stock prices as the prices follow a random pattern. It is difficult to apply simple time series or regression techniques due to the volatile nature of the stock market. Financial institutions and traders have developed different proprietary strategies to try to beat the market for themselves or their clients, but seldom has anybody achieved consistently higher-than-average investment returns. Nonetheless, the stock forecasting challenge is so compelling as a gain of just a few percentage points will boost these institutions ' income by millions of dollars. Most predictive models have historically based on linear time series models such as ARIMA [17]. The variance underlying stock movements and other assets, however, makes linear techniques suboptimal, and non-linear models such as ARCH tend to have lower predictive errors [18].

## 1.2 Motivation of the Research

Very few researches have been done for predicting stock prices on Bangladeshi stock market using machine learning techniques. The other studies carried out on different stock markets for predicting stock price using machine learning techniques were failed to give promising accuracy. We have collected the stock price data of GrameenPhone of about last five years (1/1/2014-21/11/2019) form Stock Bangladesh website (stockbangladesh.com). The dataset contains six columns named Date, Open, High, Low, Close and volume. Where the Date stands for a date, the Open indicate the opening price of a stock on the particular date, the High and Low stands for the highest and lowest trading price of on that day. The Close indicates the closing price of the day and the volume indicates the numbers of share transacted on the day. We have used the closing prices of the previous days to predict out stock price for the future days. We have evaluated the performance of SVR (Support vector machine- Regression) with Linear & Radial Basis Function (RBF) kernels and Linear Regression with the previous price data. We tried to predict the stock price with a better accuracy than the previous studies which are mentioned in Literature Review section. Lack of studies on Bangladeshi stock price prediction was one of the key factor for us to carry out his research.

## 1.3 Problem Statement

We have seen potent machine learning algorithms like neural networks to be used in predicting stock prices. The models could not give a promising accuracy. We have used SVR (Support vector machine- Regression) with Linear & Radial Basis Function (RBF) kernels and Linear Regression to predict the stock price to get a higher accuracy and compare which of these models gives the better accuracy in predicting stock prices for Bangladeshi stock market.

## 1.4 Research Questions

The research question was:

- RQ1: Which algorithm will perform the best for predicting stock price?
- RQ2: Would it give a higher accuracy than the previous models used in various researches?

## 1.5 Research Objective

The objectives of this research are:

- To find out the best performing algorithm
- To investigate whether the accuracy of our model is higher than the previous studies.

## 1.6 Research Scope

We have found research gap in different papers that we have gone through. The accuracy was quite low. And researches predicting stock price with Bangladeshi stock market data was very low. So, it is a great scope to work with stock data of Bangladesh. Using potent machine learning algorithms like LR and SVR may increase the accuracy. So, we have Used close prices to predict stock prices by using it as the independent variable which will determine the prices of the next days. We have used Dhaka stock exchange data to execute our model.

## 1.7 Thesis Organization

In the following, in second chapter we discussed about the background works where other researches done on the same subject including the research gap. In chapter three, we discussed our proposed research methodology. In chapter four, we discussed the analysis of the results. Finally, in chapter five, we discussed the conclusions and recommendations containing findings, limitations and future work.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1 Previous literature

Various algorithms for machine learning are used to predict stock price trends. Some of them are ANN(Artificial Neural Networks) [3,4,5,6] , GA(Genetic Algorithm) [5], LS-SVM(Least Square Support Vector Machine) [1,7], Trend Estimation with Linear Regression (TELR) [8],SVM(support vector machines) [4,6,9] with different kernels, KNN( K Nearest Neighbors) [5] structural support vector machines (SSVMs)[10] . Some statistical analyses are also used like Autoregressive Integrated Moving Average (ARIMA) [11]. But none of them were able to give quite promising prediction due to the non-linearity of the data.

## 2.2 Algorithm performance of different models on Stock Price Data

Studies have tried to predict stock prices using Artificial Neural Network. In a study in 2017 [4] on Korean stock market ANN was used to predict the stock price with the highest accuracy of 81.34 percent for 20 days and 83.01 percent for 30 days moving average. In another study [6] ANN was used and got 86.69% average accuracy based on experiment carried out on three different stocks. In 2018 this study [3] ANN was used and the best SSE score was 0.6271104815 for Apple, 0.0121281374 for Pepsi, 0.0335425612 for IBM, 0.016770174 for McDonald and 0.0211154625 for LG. In this study [4] author used ANN and achieved 96.10 percent accuracy for 5-fold average prediction (Mj`). Mj is a model where $1 \leq j \leq J$, for each classifier. And Mj` is a variant model as a substitution of Mj. 92.81 percent accuracy was achieved from model Mj.

Least square support vector machine (LS-SVM) is another popular machine learning used to predict stock prices. In this study [7] author a LS-SVM model along with Particle Swarm

Optimization (PSO) and the accuracy rate of the system was around 90.5 – 93%. Another research [1] used LS-SVM with PSO optimization and LS_SVM to predict stock prices. The Mean Squared Error (MSE) was 0.5317 for Adobe, 0.6314 for oracle 0.7725 for HP and 0.7905 for American Express with PSO-LS-SVM where MSE was 0.5703 for Adobe, 0.8829 for Oracle 1.2537 for HP and 1.0663 for American Express.

Other potent Machine Learning Algorithms like SVM was used in this study [6] and found average accuracy of 89.33 percent after applying the algorithm for three company stocks. For BSE-Sensex the accuracy was 90.10% using polynomial kernel where c=1 and degree =1 and for RBF kernel the accuracy was 88.08 where c=1 gamma=4. For Infosys the accuracy was 89.59% with polynomial kernel where c=1 and degree=1 and 87.80% for RBF kernel where c=5 and gamma=1.5. Another research [9] that used SVM with RBF kernel has the mean accuracy of 53.3% to 56.8% for 10 days. And the accuracy rate can be lower than 50% for different dataset. A modified linear regression algorithm Trend Estimation with Linear Regression was used in this study [8] and Mean Absolute Percentage Error (MAPE) was 5.41% for bank data and 5.42% for overall stock data. K Nearest Neighbors or KNN is also been used to predict stock prices. A study that used KNN [5] has an accuracy of 83.52 percent with KNN. Structural support vector machines (SSVMs) is used in a research and the accuracy with training samples was higher than 78% and the accuracy with testing samples was about 50%.

## 2.3 Research gap:

None of the algorithm we have seen to used in those studies that we have discussed had an accuracy higher than 90 percent except SVM optimized with PSO [7]. In that research a highest accuracy of 90.5 to 93 percent was found. One study was found which has used Bangladeshi stock price data to predict stock price. Author has used a modified linear regression algorithm named TELR. No study was carried out on Bangladeshi Stock market data with a simple linear regression algorithm and SVR with different kernels. So we have implemented the model which will use simple linear regression and SVR with RBF and linear kernel and then compare the accuracy among those.

# CHAPTER 3

# RESEARCH METHODOLOGY

## 3.1 Framing the Question

We must ask as many important questions as possible in any data science study. We need to look at the available dataset and formulate important hypotheses and questions. In our study, looking at our dataset, we realized that closing price was an important element in the analysis of stock price data. So, we decided to use the closing price of the previous dates to use as a factor to predict the next closing prices.

## 3.2 The Proposed Model

To define our Proposed model we used the following diagram:

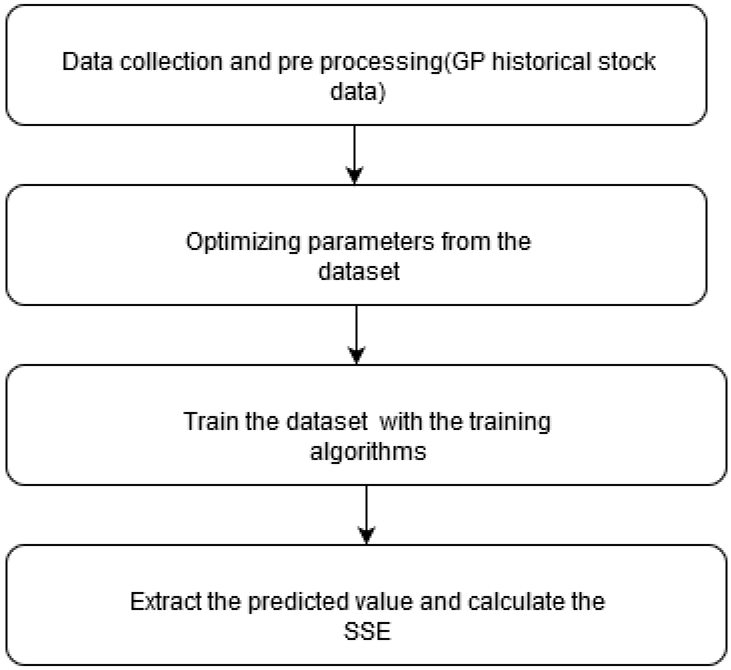



Fig1: The proposed model

As our diagram (Fig11) shows our proposed model has five stages. First, we collected the data and pre-process it. Then we optimized the parameters that we are going to use for our training algorithm. Then we tested our dataset with Linear Regression and support vector regression with 2 different kernels using the parameters that we optimized. Then we extracted our output and tested the accuracy of our algorithms with SSE. Finally, we visualized the comparison between the actual values and the predicted values.

## 3.3 Data collection and pre-processing

We collected the data form stockbangladesh.com. The website is an open source website and contains the previous price data of all the companies listed in Dhaka Stock Exchange. We have collected the price data of Grameenphone from 1st January 2014 to 21st November 2019. We got the dataset in the recent price to the oldest price form. To Process the data to our desired order we had to flip the data to get the oldest price to recent price order. We checked for missing values; there was none so we used the dataset as it was.

| Date | Open | High | Low | Close | Volume |
|---|---|---|---|---|---|
| 1/1/2014 | 79.5 | 81.6 | 79 | 79.5 | 210400 |
| 2/1/2014 | 82.5 | 85.6 | 80.5 | 83.8 | 751200 |
| 6/1/2014 | 85.4 | 85.4 | 82.1 | 83.9 | 194800 |
| 7/1/2014 | 79.5 | 85.8 | 75.5 | 85.2 | 377400 |
| 8/1/2014 | 85.9 | 89 | 84.9 | 88 | 641200 |
| 9/1/2014 | 88.3 | 90.4 | 85 | 86.2 | 378800 |
| 12/1/2014 | 87.9 | 89.3 | 86.5 | 87.4 | 368800 |
| 13/01/2014 | 87.5 | 89.3 | 83.2 | 83.4 | 493400 |
| 15/01/2014 | 83.4 | 84.8 | 81 | 82 | 540200 |

Fig2: sample data of GP

Fig2 shows a sample of the data that we collected after pre-processing.

## 3.4 Optimizing parameters from the dataset

We have used the Close column of our dataset as the input parameter to predict the stock prices. The close prices for different dates were used as the dependent variable to predict stock prices. We have split the data into 80 percent test data and 20 percent train data and stored them into different variables to use them as parameters.

## 3.5 Training Algorithms

### 3.5.1 Linear regression

We have used linear regression algorithm to train our data. Linear regression is a statistical method for modeling a relationship between two variables that corresponds to the observed data on a linear equation. One variable is regarded as an explanatory variable and the other as a dependent variable. The linear regression method can be used to predict under the assumption that the correlation between the variables will continue in the future. A linear regression equation (1) is as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \ldots \ldots \ldots \ldots \ldots (1)$$

Here, $Y_i$ is the outcome (predicted output) of the dependent variable for the $i^{th}$ test unit , $X_i$ is the independent variable which is used for prediction for $i^{th}$ test unit. $\beta_0 + \beta_1 X_i$ is the linear relation between $Y_i$ and $X_i$ $\beta_0$ is the intercept or the mean of Y when X=0 and $\beta_1$ is the slope or the change in mean when X increases by 1. The $\varepsilon_i$ represents the error term. In Our model we have used the data close column as our independent variable. So in our model $Close_i = X_i$.

The $\beta_0$ and $\beta_1$ parameters are unknown. They are estimated by using least square method. The least square method (2) is as follows:

$$b = \frac{n(\sum XY)-(\sum X).(\sum Y)}{n(\sum X^2)-(\sum X)^2} \ldots \ldots \ldots. \ldots. \ldots (2)$$

Considering the least square the best fitted line is taken and uses that as function for the prediction.

### 3.5.2 Support Vector Regression

Support vector Regression (SVR) has the same principal that Support vector machine (SVM) uses except for a few minor differences. At first, as output is a real number, the information at hand which has endless possibilities becomes very hard to predict. In the case of regression, the SVM is approximated by a range of tolerance (epsilon), which would already have requested the problem. The main idea, however, is always the same: to mitigate error, to individualize the hyperplane which maximizes the margin, taking into account which part of the error is tolerated.

The fundamental idea behind SVM is for training information from the input field to be transformed into a higher dimension of function ϕ and then a separating hyperplane with maximum margin in the function space is constructed as shown in Fig3.

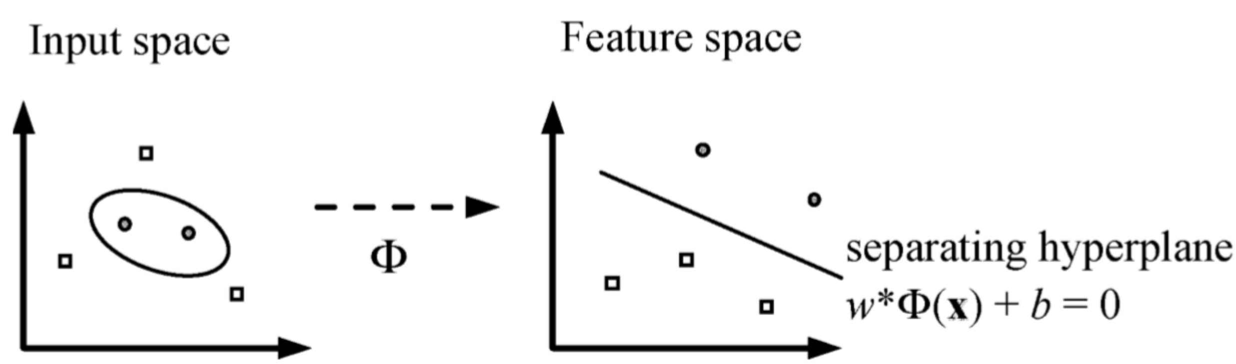



Fig3: Using a kernel function SVM mapped the data into a higher dimensional space and separated them using a hyperplane.

When it comes to SVR we can consider a set of training data $\{(x_1, y_1)\ldots(x_l, y_l)\}$ where each $x_i \subset R^n$ which represents the sample input space and has an adequate target value $y_i \subset R$ for $i = 1,\ldots,l$ , where l is the size of training data[13]. The standard SVR estimation function(1) is as follows:

$$f(x) = \big(w \cdot \Phi(x)\big) + b \qquad \ldots . \ldots \ldots \ldots (3)$$

Where $w \subset R_n$, $b \subset R$ and $\phi$ indicates a nonlinear conversion from $R_n$ to a high-dimensional space. Our objective is to find the value of w and b so that we can determine values of x by reducing the risk of regression.

$$R_{\text{reg}}(f) = C \sum_{i=0}^{\ell} \Gamma\big(f(x_i) - y_i\big) + \frac{1}{2}\|w\|^2$$

……………(4)

Here Γ () is a cost function C is a constant and the data points vector w may be formulated as:

$$w = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*)\, \Phi(x_i).$$

………………(5)

The most widely used cost function is the $\mathscr{E}$-insensitive loss function [14]. The function is in this following form:

$$\Gamma\left(f(x) - y\right) = \begin{cases} |f(x) - y| - \varepsilon \\ 0, \end{cases}$$

……………(6)

The regression risk in (4) and the -insensitive loss function (6) can be minimized by solving the quadratic optimization problem.

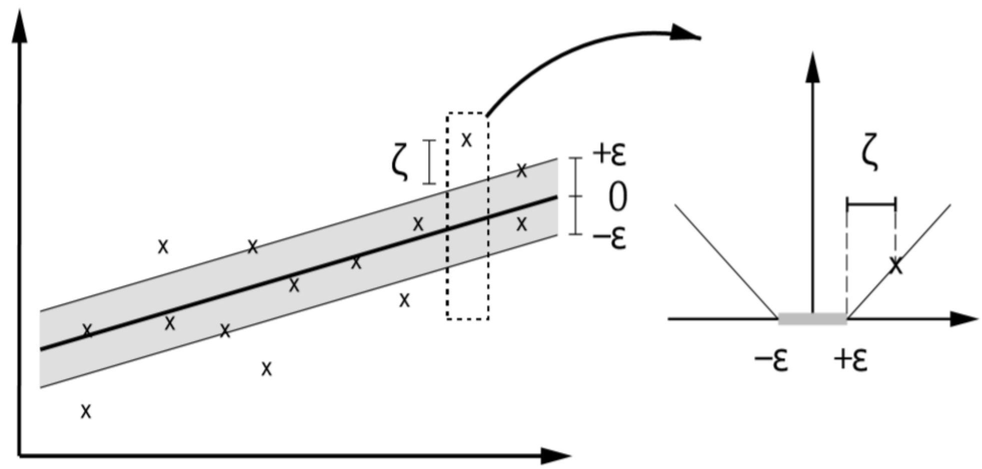



Fig4: The configuration of the soft margin loss applies to a linear SVR system

ζ is a slack variable used to calculate errors outside $\mathscr{E}$ tube. In Fig3 SVR is fitting a tube with radius $\mathscr{E}$ to the data and positive slack ζ is measuring the points that are outside the tube.

The standard formula can be rewritten by substituting (5) to (3):

$$\begin{aligned} f(x) &= \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) \big(\Phi(x_i) \cdot \Phi(x)\big) + b \\ &= \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*)\, k(x_i, x) + b. \end{aligned}$$

… …. …. …. (7)

In (6) the function $k(x_i, x)$ can substitute the dot product known as the kernel function. The kernel function is the idea is to map the non-linear data set into a higher dimensional space where a hyperplane can be found that separates the samples. There are different types of kernels in SVM. The kernels that we used for our model are Radial basis function kernel of RBF and linear kernel.

### 3.5.2.1 Radial basis function kernel

The Radial basis function kernel, or RBF kernel also known as Gaussian kernel, is a kernel that is in the form of a radial basis function. The RBF kernel on two samples x and x′, interpreted in some input space as feature vectors, is defined as follows:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right) \quad \text{… … … … (8)}$$

It can also be interpreted as:

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma\|\mathbf{x} - \mathbf{x}'\|^2) \quad \text{… … … … (9)}$$

Radial basis function takes parameters like gamma and c. The parameter gamma describes the degree to which the effect of a single example of learning exceeds. The C parameter works against maximizing the margin of the decision function from accurate identification of training instances. We have set the values of our parameters by gamma = 0.1 and c = le3.

### 3.5.2.2 Linear kernel

The Linear kernel function is the simplest kernel function. The function is given by the internal product (x, y) and an optional constant C. Algorithms using linear kernel functions are often the same as their non-kernel counterparts. It can be interpreted as:

$$k(x, y) = x^T y + c \quad \dots\dots\dots\dots (10)$$

## 3.6 Extracting Predicted value and calculating SSE

After training our dataset with the algorithms we extract the forecasted values that our algorithms predicted. We calculated the accuracy by calculating the SSE where the value is between 0 to1. 1 is the best possible value that we can get meaning an accuracy of hundred percent. The formula for calculating SSE is:

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \quad \dots\dots\dots\dots (11)$$

Where $Y_i$ is the actual value and $\hat{Y}_i$ is the predicted value

# CHAPTER 4

# RESULTS AND DISCUSSIONS

## 4.1 Data Analysis Technique

Data analysis was done using the sklearn learning framework of python with Jupiter notebook. Two regression algorithm LR and SVR were used for this research. SVR was trained using two different kernels which are RBF and linear kernel.

## 4.2. Sum Squared Error

Sum Squared error was used to determine the accuracy of the models. In SSE the lowest accuracy is zero percent or 0 in SSE value and the highest accuracy is 100 percent or 1 in SSE value. The equation used to calculate SSE is in research methodology shown in equation 11. It is also known as RSS or residual sum of squares. Residual means remaining or unexplained.

## 4.3 Data Visualization

After training and test our dataset with algorithms we have extracted the predicted value for different days. We have predicted the closing price for a 10 days, 20 days and 60 days period. The accuracy for predicting the price for less number of days was much higher than the accuracy for a higher number of days. The accuracy and predicted price changes after each iteration. We have discussed about the accuracy and the predicted price that we got for a random iteration. We have plotted the comparison between the actual value and predicted value using graphical representations. We have used matplotlib; a python library to visualize the graph for the values.

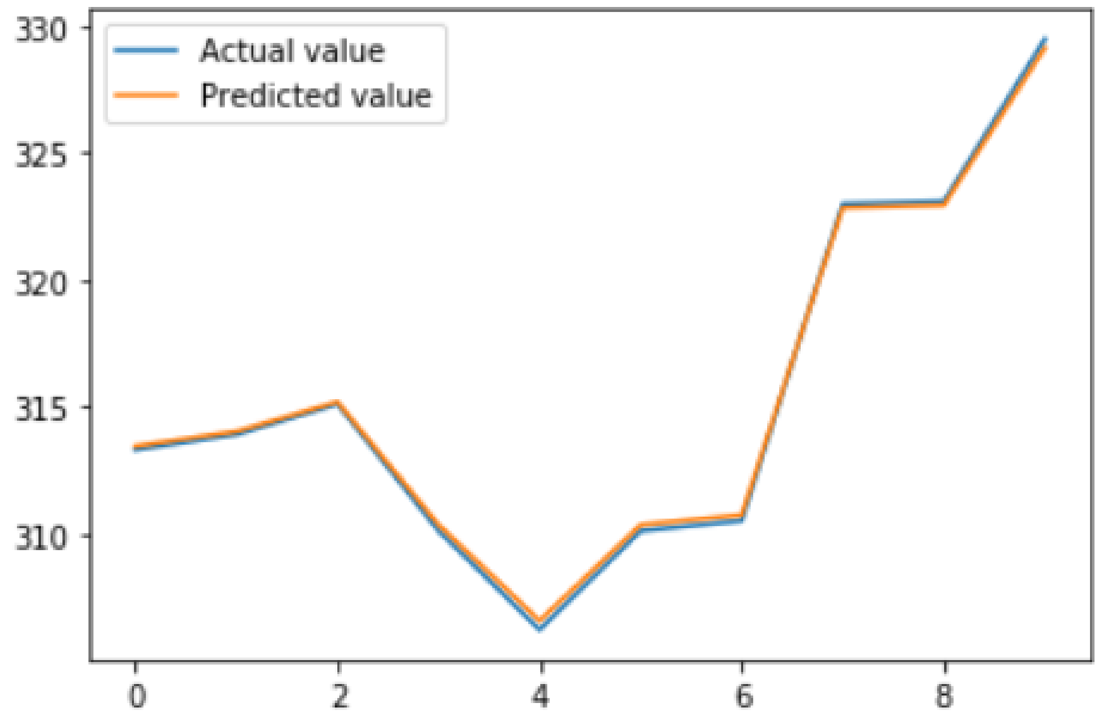



Fig5: Comparison for 10 days (LR)

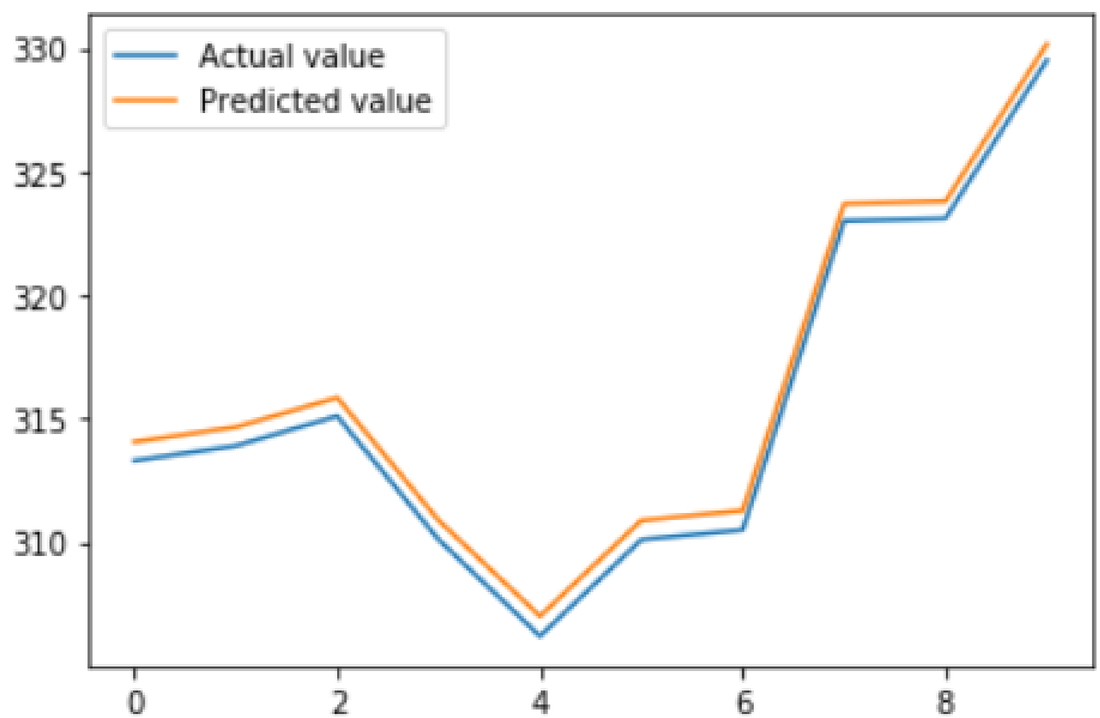



Fig6: Comparison for 10 days SVR (linear)

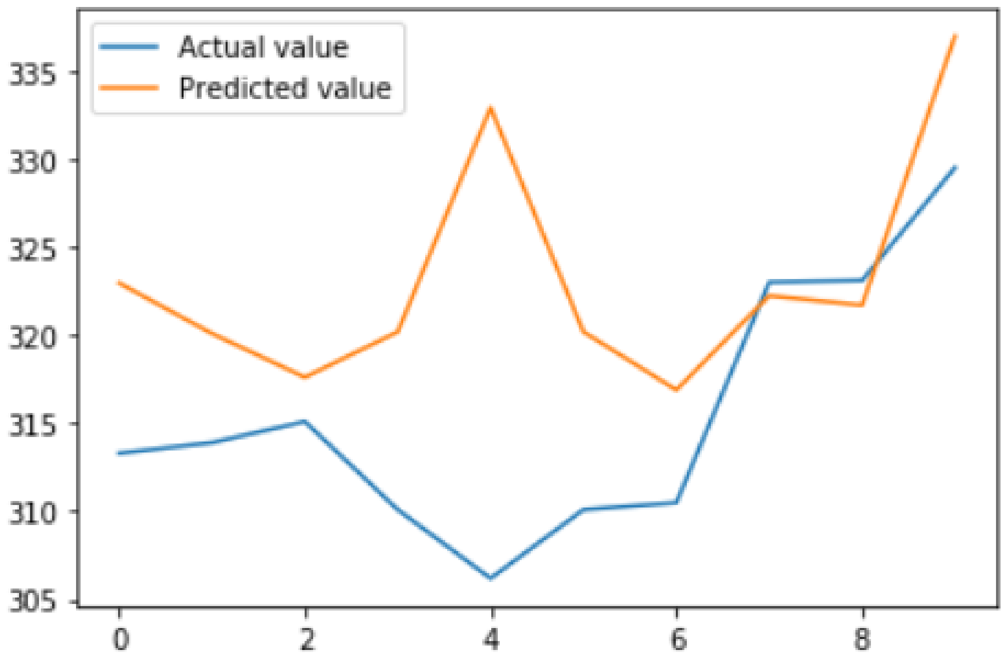



Fig7: Comparison for 10 days SVR (RBF)

After running our program we found an accuracy of 97.07 percent with linear Regression 97.06 with SVR (RBF) and 96.82 with SVR(linear) for 10 days. Fig5 shows the comparison the comparison between actual value and forecasted value for 10 days with LR. It has a 94.07 percent accuracy which is pretty high for a continuous valued prediction. The graph shows that the predicted value is almost as same as the actual value which is pretty impressive. In Fig6 we can see the graphical comparison between the actual and predicted value for SVR (linear). The accuracy is 97.06 percent which is pretty similar to the LR prediction and the predicted values are very close to the actual values. In Fig7 the graph shows the comparison between actual and predicted values with SVR (RBF). The patterns are quite close for actual and predicted values but still not as accurate as LR and SVR (linear). The accuracy is 96.86 percent which is still pretty high.

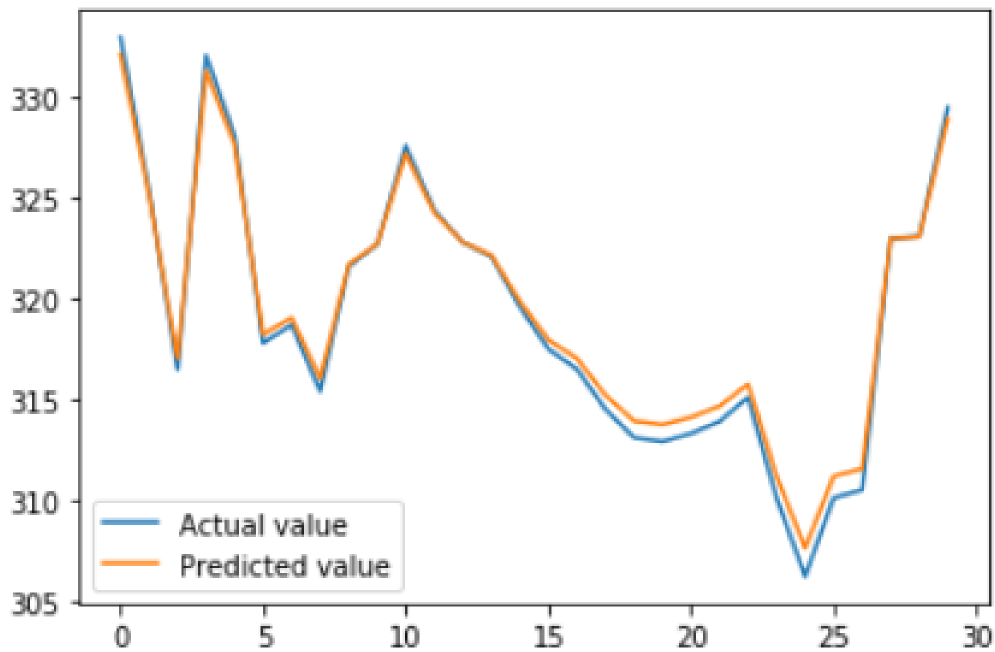



Fig8: Comparison for 30 days (LR)

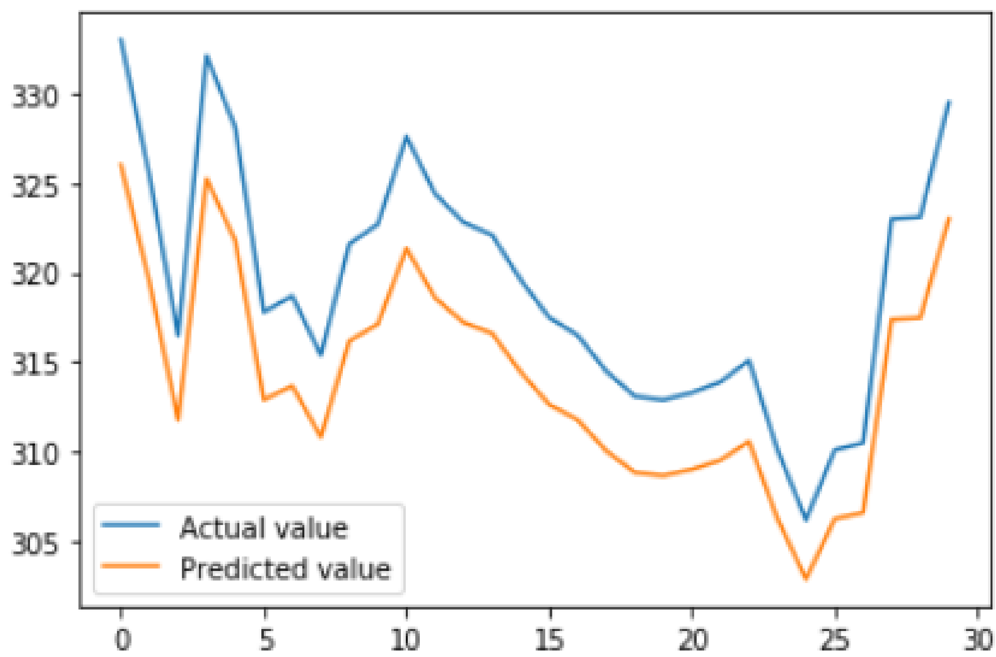



Fig9: Comparison for 30 days SVR (linear)

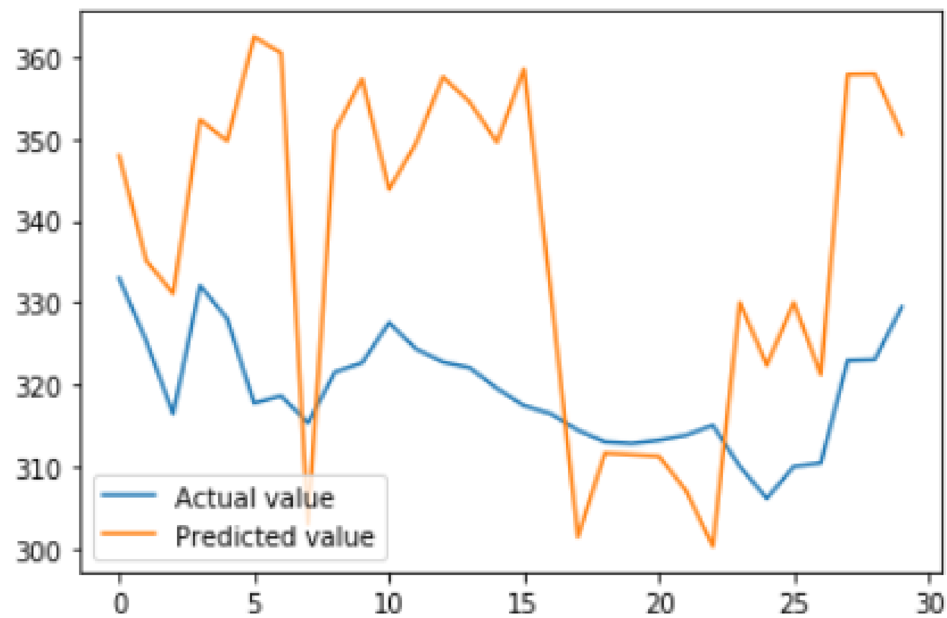



Fig10: comparison for 30 days SVR (RBF)

We have predicted the closing price for 30 days period too. For 30 days The Linear regression performed the best with the highest accuracy of 91.22 percent while the SVR (linear) has an accuracy of about 90.70 percent and SVR (RBF) has an accuracy of about 87.50 percent. The accuracy for all of the algorithms are still quite good. Fig8 shows that the actual and predict values are very close. Fig9 shows the predicted prices followed the same trend as the actual prices. In Fig10 with SVR (RBF) the prices weren't quite accurate but still pretty good for a continuous valued prediction.

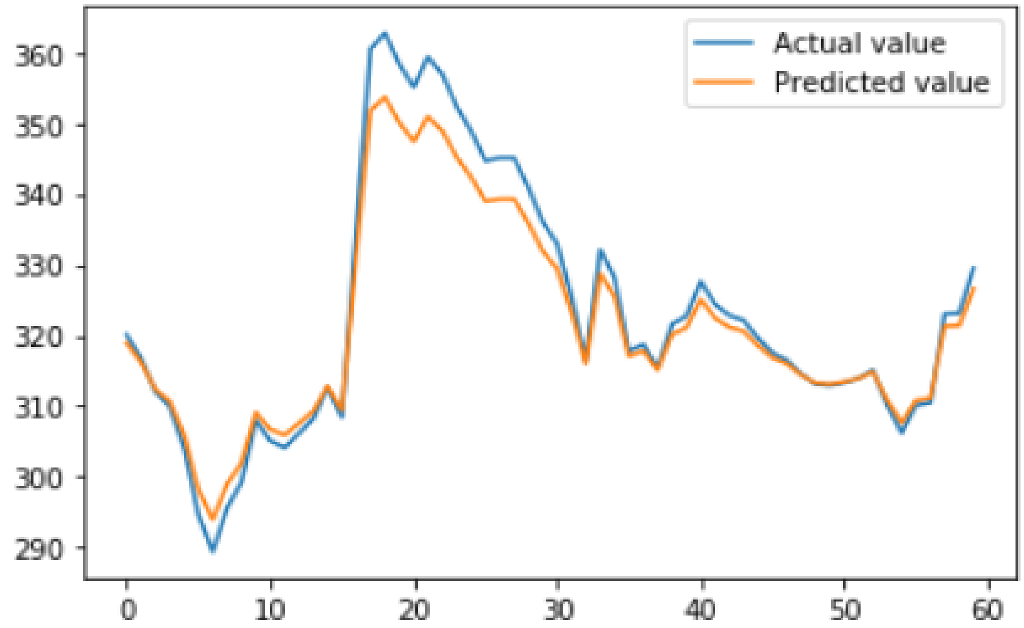



Fig11: Comparison for 60 days (LR)

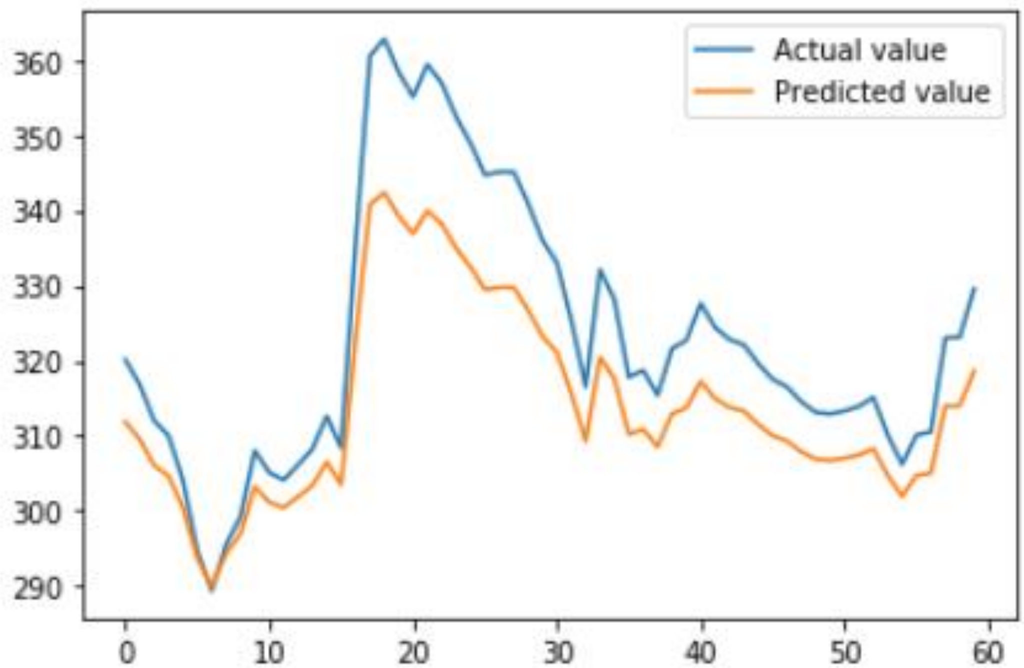



Fig12: Comparison for 60 days SVR (linear)

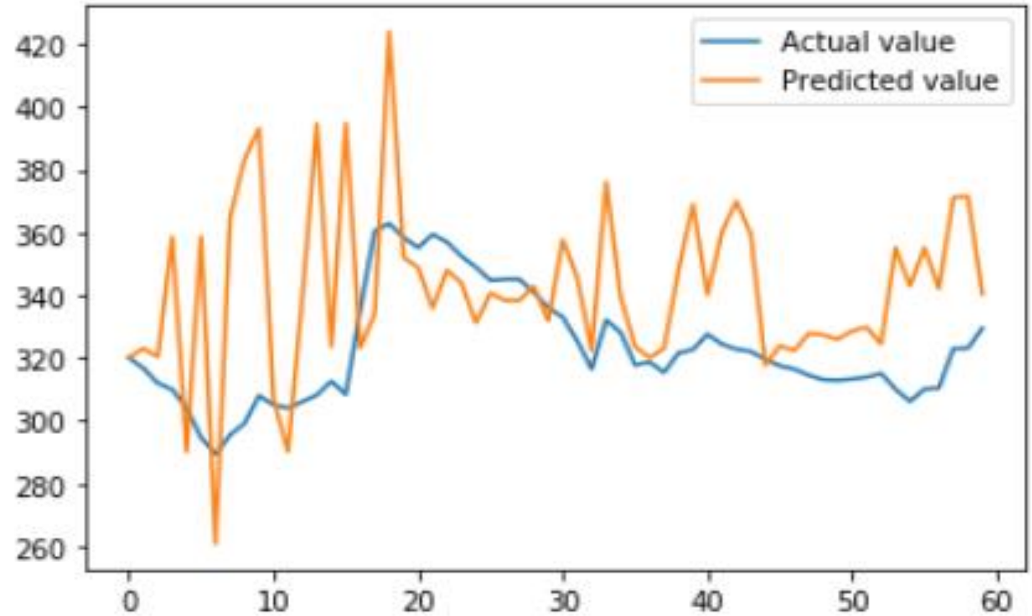



Fig13: Comparison for 60 days SVR (RBF)

For 60 days prediction like other prediction LR has the best performance with an accuracy of 79.82 percent while SVR (RBF) has a better accuracy than SVR (linear) with 78.50 percent and SVR (linear) have an accuracy of 77.53 percent. The comparison in graphical representation is Fig11, Fig12, Fig13. In Fig11 we can see that the predicted price trend followed the actual price trend. The accuracy of linear is much lower than the 10 days or 30 days prediction accuracy but it's still pretty close. For SVR (linear) the trend is similar, but the values are a little far from the actual price point. For SVR (RBF) the price points are pretty close in some points but still has a lower accuracy than LR.

## 4.4 Evaluation Metrics Table (In-Sample)

| Days\| algorithm | LR | SVR(linear) | SVR(RBF) |
|---|---|---|---|
| 10 | 97.07% | 97.06% | 96.82% |
| 30 | 91.22% | 90.70% | 87.50% |
| 60 | 79.82% | 77.53% | 78.50% |

The table shows the accuracy comparison between LR, SVR (linear) and SVR(RBF) for 10, 30 and 60 days.

## 4.5 Conclusion from In-Sample Evaluation

As per our table we can see that Liner Regression algorithm has the best performance among all. SVR (linear) and SVR (RBF) have pretty much similar performance with varying performance for different days.

# CHAPTER 5

# CONCLUSION AND FUTURE WORK

## 5.1 Findings and Contributions

Linear Regression model has performed the best to predict stock price. It has the highest accuracy for 10 days than any other studies that we have gone through. For fewer days it has a tremendous performance. SVR (linear) and SVR (RBF) has a quite impressive performance too. These too was also found to have a very high accuracy for fewer days. For 10 days each one of them had an accuracy above 95 percent.

## 5.2 Limitations

Though our models have an outstanding performance for fewer days, for a higher number of days the accuracy drops. The further the range increases the lower the accuracy goes. It is normal to have a lower accuracy for a higher number of days. LR and SVR(linear) still has an accuracy above 90 percent for ten days with the exception of SVR(RBF) with an accuracy of about 87 percent. The SVR(RBF) performed better than SVR (linear) for sixty days. So there’s still some weakness that can be improved.

## 5.3 Recommendations for Future Works

With different parameters the performance of SVR may be improved. Using technical and fundamental indicators like RSI, MACD, investor sentiment percentage, company background etc. as parameters might improve the prediction performance as these has an effect on stock price movement. Using these parameters in potent machine learning algorithms might increase the accuracy of price prediction.

# REFERENCES


1. Hegazy, O., Soliman, O. S., & Salam, M. A. (2014). A machine learning model for stock market prediction. arXiv preprint arXiv:1402.7351.
2. Guo, K., Sun, Y., & Qian, X. (2017). Can investor sentiment be used to predict the stock price? Dynamic analysis based on China stock market. Physica A: Statistical Mechanics and its Applications, 469, 390-396.
3. Ebadati, O. M. E., & Mortazavi, M. T. (2018). An efficient hybrid machine learning method for time series stock market forecasting. Neural Network World, 28(1), 41-55.
4. Pyo, S., Lee, J., Cha, M., & Jang, H. (2017). Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the Korean stock markets. PloS one, 12(11), e0188107.
5. Qian, B., & Rasheed, K. (2007). Stock market prediction with multiple classifiers. Applied Intelligence, 26(1), 25-33.
6. Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Systems with Applications, 42(1), 259-268.
7. Akash, A & Rajaji, Shanthi & Aravinth, R & Vendhan, V & Veerapandi, D. (2019). Stock Market Trend Prediction using Machine Learning. International Journal of Innovative Research in Computer and Communication Engineering. 7. 1000-1006. 10.15680/IJIRCCE.2019.0702085.
8. Efat, M. I. A., Bashar, R., Uddin, K. I., & Bhuiyan,(2018) T. Trend Estimation of Stock Market: An Intelligent Decision System. International Conference on Cyber Security and Computer Science
9. Madge, S., & Bhatt, S. (2015). Predicting stock price direction using support vector machines. Independent work report spring.
10. Leung, C. K. S., MacKinnon, R. K., & Wang, Y. (2014, July). A machine learning approach for stock price prediction. In Proceedings of the 18th International Database Engineering & Applications Symposium (pp. 274-277). ACM.
11. Kamalakannan, J., Sengupta, I., & Chaudhury, S. (2018, February). Stock Market Prediction using Time Series Analysis. In 2018 IADS International Conference on Computing, Communications & Data Engineering (CCODE) (pp. 7-8).
12. David R. Harper (2019) Forces That Move Stock Prices. https://www.investopedia.com/articles/basics/04/100804.asp Accessed 20 Nov 2019
13. Müller, K. R., Smola, A. J., Rätsch, G., Schölkopf, B., Kohlmorgen, J., & Vapnik, V. (1997, October). Predicting time series with support vector machines. In International Conference on Artificial Neural Networks (pp. 999-1004). Springer, Berlin, Heidelberg.
14. Müller, K. R., Smola, A., Rätsch, G., Schölkopf, B., Kohlmorgen, J., & Vapnik, V. (1999). Using support vector machines for time series prediction. Advances in kernel methods—support vector learning, 243-254.
15. Edwards, R. D., Magee, J., & Bassetti, W. C. (2018). Technical analysis of stock trends. CRC press.

16. Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. Neural networks, 17(1), 113-126.
17. Bontempi, G., Ben Taieb, S., & Le borgne, Y-A. (2013). Machine learning strategies for time series forecasting. In M-A. Aufaure, & E. Zimányi (Eds.), Business Intelligence*:* Second European Summer School, eBISS 2012, Brussels, Belgium, July 15-21, 2012, Tutorial Lectures (pp. 62-77).
18. Zhang, J., Shan, R., & Su, W. (2009). Applying time series analysis builds stock price forecast model. Modern Applied Science, 3(5), 152-157.
19. Verma, S. A., Thampi, G. T., & Rao, M. (2017). Inter-comparison of Artificial Neural Network Algorithms for Time Series Forecasting: Predicting Indian Financial Markets. International Journal of Computer Applications, 162(2).
20. Nayak, S. C., Misra, B. B., & Behera, H. S. (2012, February). Index prediction with neuro-genetic hybrid network: A comparative analysis of performance. In 2012 International Conference on Computing, Communication and Applications (pp. 1-6). IEEE.
21. Cai, C., Ma, Q., & Lv, S. (2012). Research on support vector regression in the stock market forecasting. In Advances in Electronic Commerce, Web Application and Communication (pp. 607-612). Springer, Berlin, Heidelberg.
22. Wu, J. Y., & Lu, C. J. (2012, June). Computational intelligence approaches for stock price forecasting. In 2012 International Symposium on Computer, Consumer and Control (pp. 52-55). IEEE.
23. Huang, C., Huang, L. L., & Han, T. T. (2012, May). Financial time series forecasting based on wavelet kernel support vector machine. In 2012 8th International Conference on Natural Computation (pp. 79-83). IEEE.