****

**CHAPTER 1**

**1 Introduction**

**1.1 Data Mining**

Data Mining is a computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems.The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.It is the analysis step of the "knowledge discovery in databases" process, or KDD.the goal is the extraction of patterns and knowledge from large amounts of data, not the extraction (mining) of data itself.

Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

* Massive data collection
* Powerful multiprocessor computers
* Data mining algorithms

**Types of Data Mining:**

1. Spatial Data Mining
2. Temporal Data Mining
3. Web Mining
4. Text Mining
5. Visual Data Mining

From these mining types we focus on Temporal Data Mining.

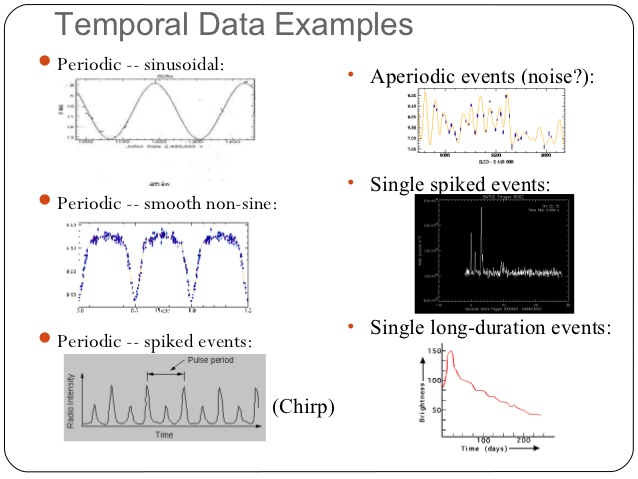
**1.2 Temporal Data Mining**

Temporal Data Mining is a single step in the process of Knowledge Discovery in Temporal Databases that enumerates structures (temporal patterns or models) over the temporal data, and any algorithm that enumerates temporal patterns from, or fits models to, temporal data is a Temporal Data Mining Algorithm.It refers to the extraction of implicit, non-trivial, and potentially useful abstract information from large collections of temporal data.

Temporal data are sequences of a primary data type, most commonly numerical or categorical values and sometimes multivariate or composite information. Examples of temporal data are regular time series (e.g., stock ticks, EEG), event sequences (e.g., sensor readings, packet traces, medical records, web log data), and temporal databases (e.g., relations with time stamped tuples, databases with versioning). The common factor of all these sequence types is the total ordering of their elements. They differ on the type of primary information, the regularity of the elements in the sequence, and on whether there is explicit temporal information associated to each element (e.g., timestamps).

**Types of Temporal Data Mining:**

* Classification
* Clustering
* Association
* Prediction

****

*Figure 1: Temporal Data Examples*

**1.3 Time Series Prediction**

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values.Time series data have a natural temporal ordering.A [stochastic](https://en.wikipedia.org/wiki/Stochastic" \o "Stochastic) model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart.

**1.3.1 Support Vector Machine**

Support vector machine (SVM) is a machine learning approach for pattern recognition. The main objective of the SVM is to find a hyper plane,which divides data with the widest margin.The SVM is formed as a convex optimization problem.In [machine learning](https://en.wikipedia.org/wiki/Machine_learning" \o "Machine learning), support vector machines are [supervised learning](https://en.wikipedia.org/wiki/Supervised_learning" \o "Supervised learning) models with associated learning [algorithms](https://en.wikipedia.org/wiki/Algorithm" \o "Algorithm) that analyze data used for [classification](https://en.wikipedia.org/wiki/Statistical_classification" \o "Statistical classification) and [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis" \o "Regression analysis). Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-[probabilistic](https://en.wikipedia.org/wiki/Probabilistic_classification" \o "Probabilistic classification) [binary](https://en.wikipedia.org/wiki/Binary_classifier" \o "Binary classifier) [linear classifier](https://en.wikipedia.org/wiki/Linear_classifier" \o "Linear classifier).

**1.3.2 Linear Regression**

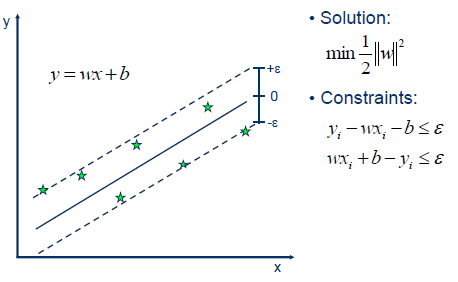
Linear regression is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression.

In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models.

If the goal is prediction, or forecasting, or error reduction, linear regression can be used to fit a predictive model to an observed data set of y and X values. After developing such a model, if an additional value of X is then given without its accompanying value of y, the fitted model can be used to make a prediction of the value of y.

**1.3.3 Support vector regression**

Support vector regression is a discriminative regression technique much like any other discriminative regression technique. You give it a set of input vectors and associated responses, and it fits a model to try and predict the response given a new input vector. Consider we have stock prices for N days. Then, for each day we could construct a feature vector, which, in a simple case, could be be the previous day's price and the current day's price. The response for each feature vector would be the next day's price. Thus, given yesterday's price and today's price the objective would be to predict the next days price.With a time series, an import step is determining what your "feature vector" x will be; each x is called a "feature" and can be calculated from present or past data, and each y, the response, will be the future change over some time period of whatever you're trying to predict.



*Figure 2; Support Vector Regression model*

**1.3.4 Moving Average method**

Moving Average method assumes that the best predictor of what will happen tomorrow is the average of everything that has happened up until now.In financial applications a simple moving average (SMA) is the unweighted [mean](https://en.wikipedia.org/wiki/Arithmetic_mean" \o "Arithmetic mean) of the previous *n* data. However, in science and engineering the mean is normally taken from an equal number of data on either side of a central value. This ensures that variations in the mean are aligned with the variations in the data rather than being shifted in time.Equation for predicting the value of Y at time t+1 based on data up to time t is:



Yt =data in time t

m=total number of time intervals.

The period selected depends on the type of movement of interest, such as short, intermediate, or long-term. In financial terms moving-average levels can be interpreted as [support](https://en.wikipedia.org/wiki/Support_(technical_analysis)" \o "Support (technical analysis)) in a falling market, or [resistance](https://en.wikipedia.org/wiki/Resistance_(technical_analysis)" \o "Resistance (technical analysis)) in a rising market.

**­­**

**CHAPTER 2**

**2.Literature Survey**

Chen and Lu[1] used geographic information system (GIS) is introduced in this study to manage and visualize the spatial distribution of demand data and forecasting results. A flexible model is implemented in GIS, through which training data are prepared with user-desired sizes for the spatial grid and discretized temporal steps.They have applied artificial neural network and sinusoidal Regression for the forecasting of prehospital emergency medical demand. The results from these approaches, as a reference, could be used for the preallocation of ambulances.

Bhaskaran and Nair[2], paper demonstrates the skill level of a wavelet neural network in improving numerical ocean wave predictions of significant wave height (Hs) and peak wave period having practical applications in operational centers.The time series of error between numerical and corresponding measured values was first constructed, and using a wavelet neural network, the errors were predicted for future time steps. The predicted errors when incorporated into the model values provided the updated prediction of Hs and Tp. The study signifies that numerical estimations could be significantly improved using this procedure. The results provide quite satisfactory predictions with a lead time varying from 3 to 24 h.

Kasabovand and Fellow[3],presents spiking neural networks (SNNs)for remote sensing spatiotemporal analysis of image time series,which make use of the highly parallel and low-power-consuming neuromorphic hardware platforms possible.It presents the development and testing of a methodological framework which utilizes the spatial accumulation of time series ofModerate Resolution Imaging Spectroradiometer 250-m resolution data and historical crop yield data to train an SNN to make timely prediction of crop yield.Our method was able to predict the yield around six weeks before harvest with a very high accuracy.

Suganthan and Srikanth[4],presents accurate wind speed forecasting is essential for power dispatch planning, unit commitment decision, maintenance scheduling, and regulation. However, wind is intermittent and wind speed is difficult to predict. The EMD is used to decompose the wind speed time series into several intrinsic mode functions (IMFs) and a residue.Subsequently, a vector combining one historical data from each IMF and the residue is generated to train the SVR. The proposed EMD-SVR model is evaluated with a wind speed data set. The proposed EMD-SVR model outperforms several recently reported methods with respect to accuracy or computational complexity.

Amin Hedayati Moghaddam[5],used artificial neural network (ANN) in forecasting the daily NASDAQ stock exchange rate was investigated. Several feed forward ANNs that were trained by the back propagation algorithm have been assessed. The methodology used in this study considered the short-term historical stock prices as well as the day of week as inputs. Daily stock exchange rates of NASDAQ from January 28, 2015 to18 June, 2015 are used to develop a robust model. First 70 days (January 28 to March 7) are selected as training dataset and the last 29 days are used for testing the model prediction ability. Networks for NAS-DAQ index prediction for two type of input dataset (four prior days and nine prior days) were developed and validated.

**CHAPTER 3**

**3.Problem Definition and Background**

**3.1 Existing Approach**

Support vector machines (SVMs) are promising methods for the prediction of financial time series because they use a risk function consisting of the empirical error and a regularized term which is derived from the structural risk minimization principle. This study applies SVM to predicting the stock price index.The experimental results show that SVM provides a promising alternative to stock market prediction.

**3.2 Problem Statement**

Since SVM cannot predict the accurate time series data.So, our goal of this research is to provide information to decision makers for predicting time series data using Support vector Regression.

**CHAPTER 4**

**4.Requirements**

**4.1. Hardware Support**

The project has the following minimum hardware requirements:

Number of Processors : 1

Processor : Intel® Core(TM)

Architecture : i4-4010U

CPU Speed : 1.70 GHz

RAM : 4 GB

**4.2. Software Requirements**

The list of software required for the project is:

(i) Windows Operating System

(ii) Java8

(iii)Weka 3.8

(iv)R Studio

**4.2.1. Windows Operating System**

The project was decided to be developed using Java and R language. JVM is available for Linux, Windows, MAC, Solaris operating systems. Out of those, we decided to choose Windows 8 operating system,, as it is more stable and reliable with long term support.

**4.2.2. Java8**

Java is platform independent meaning project compiled in one machine can be run in any other machine that has a version of JVM available for it regardless of the underlying architecture. It is an object oriented GPL licensed programming language.

Java 8 is the only version of Java that is currently supported. Main reason for choosing Java 8 is because of ease of prototyping in it.

**4.2.3 Weka 3.8**

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.Weka is a workbench that contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to these functions. This original version was primarily designed as a tool for analyzing data from agricultural domains,but the more recent fully Java-based version (Weka 3.8), for which development started in 1997, is now used in many different application areas, in particular for educational purposes and research.

**Advantages of Weka include:**

* Free availability under the GNU General Public License.
* Portability, since it is fully implemented in the Java programming language and thus runs on almost any modern computing platform.
* A comprehensive collection of data preprocessing and modeling techniques.
* Ease of use due to its graphical user interfaces.

**Weka packages:**

|  |  |  |
| --- | --- | --- |
| **Package Name** | **Purpose** | **Usage** |
| Libsvm | Classification, Regression | A wrapper class for the libsvm tools |
| WekaExcel | Converter | WEKA MS Excel loader/saver |
| timeSeriesFilters | Filters, Time Series | Time Series Filters |
| timeseriesForecasting | Time series | Time series forecasting environment. |

**4.2.4 R Studio**

R is a programming language and software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. Polls, surveys of data miners, and studies of scholarly literature databases show that R's popularity has increased substantially in recent years.

**Package:**e1071

**Package description:** Functions for latent class analysis, short time Fourier transform, fuzzy clustering, support vector machines, shortest path computation, bagged clustering, naive Bayes classifier

**4.3. Dataset**

The dataset for our experiment were crawled from UCI repository.

*Table 1:Stock market dataset*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date | open | high | low | close |
| 01-Jan-14 | 10312.81 | 10338.24 | 10254.12 | 10269.29 |
| 02-Jan-14 | 10272.36 | 10365.12 | 9936.96 | 9977.82 |
| 03-Jan-14 | 9900.59 | 9936.77 | 9760.24 | 9807.16 |
| 06-Jan-14 | 9812.43 | 9912.25 | 9752.26 | 9820.78 |
| 07-Jan-14 | 9855.68 | 9970.51 | 9803.99 | 9869.15 |
| 08-Jan-14 | 9910.21 | 9924.71 | 9715.71 | 9735.82 |
| 09-Jan-14 | 9750.72 | 9795.46 | 9525.58 | 9537.91 |
| 10-Jan-14 | 9548.05 | 9640.07 | 9392.79 | 9415.33 |
| 13-Jan-14 | 9443.98 | 9595.85 | 9443.98 | 9561.01 |
| 14-Jan-14 | 9568.6 | 9686.44 | 9469.67 | 9501.12 |
| 15-Jan-14 | 9546.92 | 0 | 9542.86 | 9689.03 |
| 16-Jan-14 | 9748.37 | 9798.14 | 9655.4 | 9713.12 |
| 17-Jan-14 | 9702.9 | 9758.83 | 9656.47 | 9688.56 |
| 20-Jan-14 | 9695.81 | 9765.87 | 9632.31 | 9700.22 |
| 21-Jan-14 | 9745.99 | 9820.37 | 9743.63 | 9761.91 |
| 22-Jan-14 | 9759.99 | 9803.19 | 9676.27 | 9710.82 |
| 23-Jan-14 | 9808.28 | 9969.64 | 9808.28 | 0 |
| 24-Jan-14 | 9836.14 | 9857.61 | 9609.81 | 9624.23 |

**CHAPTER 5**

**5.Proposed Approach**

**5.1 Overview of Proposed Design**

Normalization is accomplished through applying some formal rules either by a process of synthesis or decomposition. a non-linear function is leaned by linear learning machine mapping into high dimensional kernel induced feature space.Regression is that presenting the solution by means of small subset of training points.The below figure 3 gives the overview of proposed design.

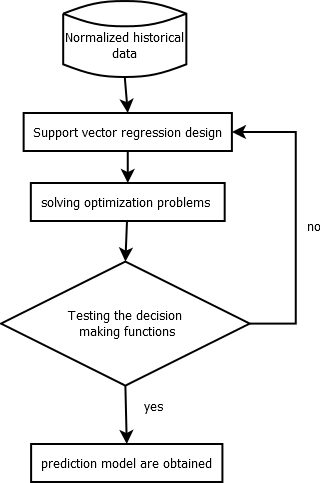


Figure 3: Proposed Design

**5.2 Normalizing historical data**

Database normalization, or simply normalization, is the process of organizing the [columns](https://en.wikipedia.org/wiki/Column_(database)" \o "Column (database)) (attributes) and [tables](https://en.wikipedia.org/wiki/Table_(database)" \o "Table (database)) (relations) of a [relational database](https://en.wikipedia.org/wiki/Relational_database" \o "Relational database) to reduce [data redundancy](https://en.wikipedia.org/wiki/Data_redundancy" \o "Data redundancy)and improve data integrity.

Normalization involves arranging attributes in tables based on [dependencies](https://en.wikipedia.org/wiki/Dependency_theory_(database_theory)" \o "Dependency theory (database theory)) between attributes, ensuring that the dependencies are properly enforced by database integrity constraints. Normalization is accomplished through applying some formal rules either by a process of synthesis or decomposition. Synthesis creates a normalized database design based on a known set of dependencies. Decomposition takes an existing (insufficiently normalized) database design and improves it based on the known set of dependencies.

When a fully normalized database structure is extended to allow it to accommodate new types of data, the pre-existing aspects of the database structure can remain largely or entirely unchanged. As a result, applications interacting with the database are minimally affected.

**5.3 Support Vector Regression Design**

Support Vector Machine can be applied not only to classification problems but also to the case of regression. Still it contains all the main features that characterize maximum margin algorithm: a non-linear function is leaned by linear learning machine mapping into high dimensional kernel induced feature space. The capacity of the system is controlled by parameters that do not depend on the dimensionality of feature space.

In the same way as with classification approach there is motivation to seek and optimize the generalization bounds given for regression. They relied on defining the loss function that ignores errors, which are situated within the certain distance of the true value. This type of function is often called – epsilon intensive – loss function.

**5.4 Solving Optimization Problems**

One of the most important ideas in Support Vector Classification and Regression cases is that presenting the solution by means of small subset of training points gives enormous computational advantages. Using the epsilon intensive loss function we ensure existence of the global minimum and at the same time optimization of reliable generalization bound.

In SVM regression, the inputIMG_256is first mapped onto a *m*-dimensional feature space using some fixed (nonlinear) mapping, and then a linear model is constructed in this feature space.Using mathematical notation, the linear model (in the feature space) *f(x,w)* is given by



Where gj(x) ,j=1,2,…m denotes a set of nonlinear transformations

*b* is the “bias” term

SVM regression performs linear regression in the high-dimension feature space using ξ-insensitive loss and, at the same time, tries to reduce model complexity by minimizing *w2*. This can be described by introducing (non-negative) slack variables ξi , i=1…n,to measure the deviation of training samples outside *w*-insensitive zone.

Thus SVM regression is formulated as minimization of the following functional:



**5.5Algorithm Description**

**5.5.1 Algorithm 1 -Support Vector Regression**

In our SVR algorithm we have given a set of input vectors and associated responses, and it fits a model to try and predict the response given a new input vector. Consider we have stock prices for N days. Then, for each day we could construct a feature vector, which, in a simple case, could be be the previous day's price and the current day's price. The response for each feature vector would be the next day's price. Thus, given yesterday's price and today's price the objective would be to predict the next days price.With a time series, an import step is determining what your "feature vector" x will be; each x is called a "feature" and can be calculated from present or past data, and each y, the response, will be the future change over some time period of whatever you're trying to predict.

Below is the code to make predictions with Support Vector Regression:

* reg.model <- lm(open~Date,data=t1)
* t1$forecast.open.predicted <- predict(reg.model,data=t1)
* points(t1$Date, t1$forecast.open.predicted,

col = "green",

pch=18)

The function will automatically choose SVM if it detects that the data is categorical.

This time the predictions is closer to the real values.

* error <- data$open - predictedY
* svrPredictionRMSE <- rmse(error)

As expected the RMSE is better than the tuned Support vector regression values.

**5.5.2 Algorithm 2 – Simple Moving Average**

In financial applications a simple moving average (SMA) is the unweighted [mean](https://en.wikipedia.org/wiki/Arithmetic_mean" \o "Arithmetic mean) of the previous *n* data. However, in science and engineering the mean is normally taken from an equal number of data on either side of a central value. This ensures that variations in the mean are aligned with the variations in the data rather than being shifted in time. An example of a simple equally weighted running mean for a n-day sample of closing price is the mean of the previous *n* days' closing prices. If those prices are *pm,pm-1,…..pm-(n-1)*





The period selected depends on the type of movement of interest, such as short, intermediate, or long-term. In financial terms moving-average levels can be interpreted as [support](https://en.wikipedia.org/wiki/Support_(technical_analysis)" \o "Support (technical analysis)) in a falling market, or [resistance](https://en.wikipedia.org/wiki/Resistance_(technical_analysis)" \o "Resistance (technical analysis)) in a rising market.

If the data used are not centered around the mean, a simple moving average lags behind the latest datum point by half the sample width. An SMA can also be disproportionately influenced by old datum points dropping out or new data coming in. One characteristic of the SMA is that if the data have a periodic fluctuation, then applying an SMA of that period will eliminate that variation

**CHAPTER 6**

**6. Implement­ation**

**6.1 Weka**

**6.1.1 Preprocessing**

The dataset used here is the stock market data which available in arff format.

* Loading the data. We can load the dataset into weka by clicking on open button in preprocessing interface and selecting the appropriate file.
* Once the data is loaded, weka will recognize the attributes and during the scan of the data weka will compute some basic strategies on each attribute. The left panel shows the list of recognized attributes while the top panel indicates the names of the base relation or table and the current working relation .
* Clicking on an attribute in the left panel will show the basic statistics on the attributes for the categorical attributes the frequency of each attribute value is shown, while for continuous attributes we can obtain min, max, mean, standard deviation and deviation etc.,
* The visualization in the right button panel in the form of cross-tabulation across two attributes. Note:we can select another attribute using the dropdown list.
* Selecting or filtering attributes Removing an attribute-When we need to remove an attribute,we can do this by using the attribute filters in weka.In the filter model panel,click on choose button,This will show a popup window with a list of available filters. Scroll down the list and select the “weka.filters.unsupervised.attribute.remove” filters.
* Next click the textbox immediately to the right of the choose button.In the resulting dialog box enter the index of the attribute to be filtered out.
* Make sure that invert selection option is set to false.The click OK now in the filter box.you will see “Remove-R-7”.
* Click the apply button to apply filter to this data.This will remove the attribute and create new working relation.
* Save the new working relation as an arff file by clicking save button on the top panel.(stockmarket.arff)

*Table 2:Preprocessed dataset*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date | open | high | low | close |
| 01-Jan-14 | 10312.81 | 10338.24 | 10254.12 | 10269.29 |
| 02-Jan-14 | 10272.36 | 10365.12 | 9936.96 | 9977.82 |
| 03-Jan-14 | 9900.59 | 9936.77 | 9760.24 | 9807.16 |
| 06-Jan-14 | 9812.43 | 9912.25 | 9752.26 | 9820.78 |
| 07-Jan-14 | 9855.68 | 9970.51 | 9803.99 | 9869.15 |
| 08-Jan-14 | 9910.21 | 9924.71 | 9715.71 | 9735.82 |
| 09-Jan-14 | 9750.72 | 9795.46 | 9525.58 | 9537.91 |
| 10-Jan-14 | 9548.05 | 9640.07 | 9392.79 | 9415.33 |
| 13-Jan-14 | 9443.98 | 9595.85 | 9443.98 | 9561.01 |
| 14-Jan-14 | 9568.6 | 9686.44 | 9469.67 | 9501.12 |
| 15-Jan-14 | 9546.92 | 9703.9 | 9542.86 | 9689.03 |
| 16-Jan-14 | 9748.37 | 9798.14 | 9655.4 | 9713.12 |
| 17-Jan-14 | 9702.9 | 9758.83 | 9656.47 | 9688.56 |
| 20-Jan-14 | 9695.81 | 9765.87 | 9632.31 | 9700.22 |
| 21-Jan-14 | 9745.99 | 9820.37 | 9743.63 | 9761.91 |
| 22-Jan-14 | 9759.99 | 9803.19 | 9676.27 | 9710.82 |
| 23-Jan-14 | 9808.28 | 9969.64 | 9808.28 | 9889.53 |
| 24-Jan-14 | 9836.14 | 9857.61 | 9609.81 | 9624.23 |

**6.1.2. SMOreg**

SMOreg is a function used for forecasting the result in regression.It implements the support vector machine for regression.

**Class column**

Choose the column that contains the target variable.

**Preliminary Attribute Check**

The Preliminary Attribute Check tests the underlying classifier against the DataTable specification at the inport of the node. Columns that are compatible with the classifier are marked with a green 'ok'. Columns which are potentially not compatible are assigned a red error message.

**Classifier Options**

C: The complexity constant C. (default 1)

N:Whether to 0=normalize/1=standardize/2=neither. (default 0=normalize)

I:Optimizer class used for solving quadratic optimization problem (default weka.classifiers.functions.supportVector.RegSMOImproved)

K:The Kernel to use.

(default: weka.classifiers.functions.supportVector.PolyKernel)

T: The tolerance parameter for checking the stopping criterion.

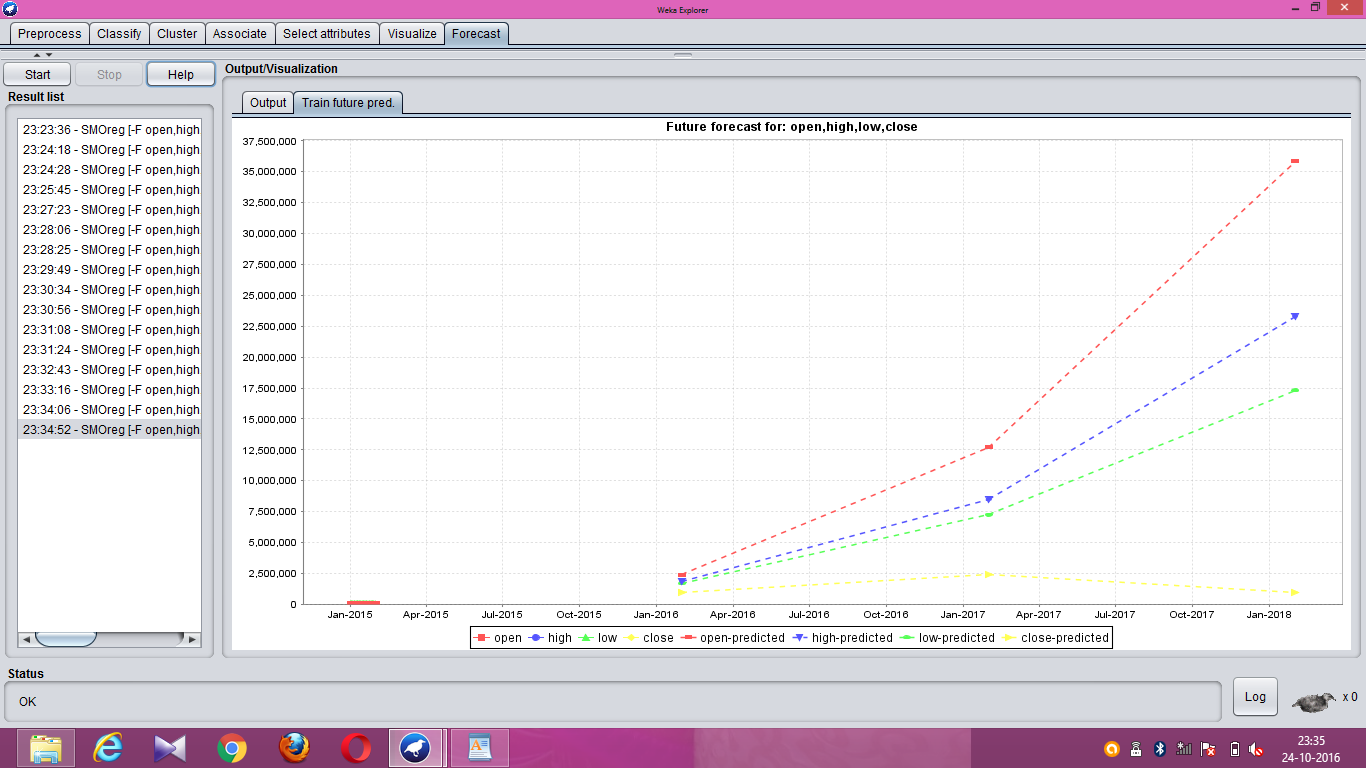
(default 0.001)

P: The epsilon for round-off error. (default 1.0e-12)

L: The epsilon parameter in epsilon-insensitive loss function.

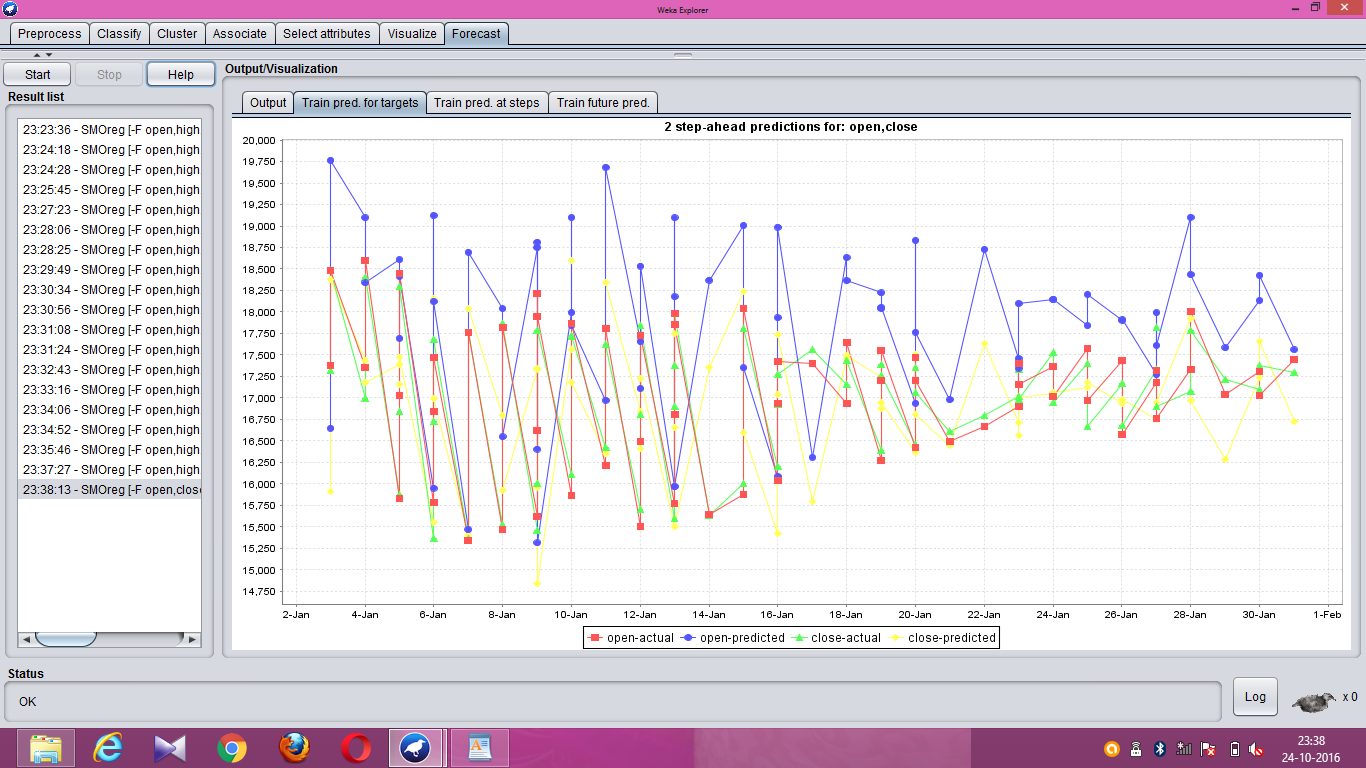
(default 1.0e-3).

The future forecast for all the attributes in stock market data is forecasted using weka and result is show below:

**

*Figure 4 :Future forecast using stock market data*

The future forecast for open and close attributes in stock market data is forecasted using weka and result is show below:

**

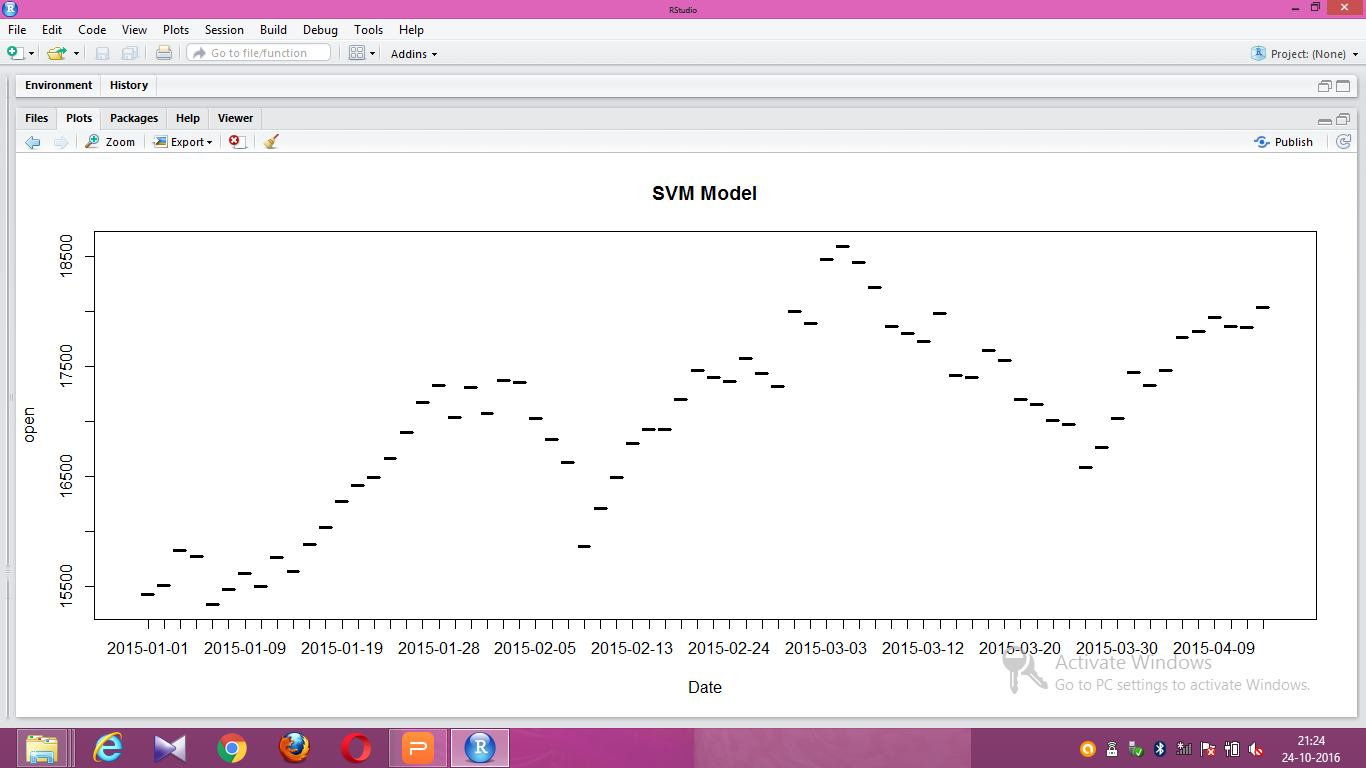
*Figure 5 Prediction of open and close attribute in stock market data*

**6.2 Support vector machine**

**Proposed procedure for SVM:**

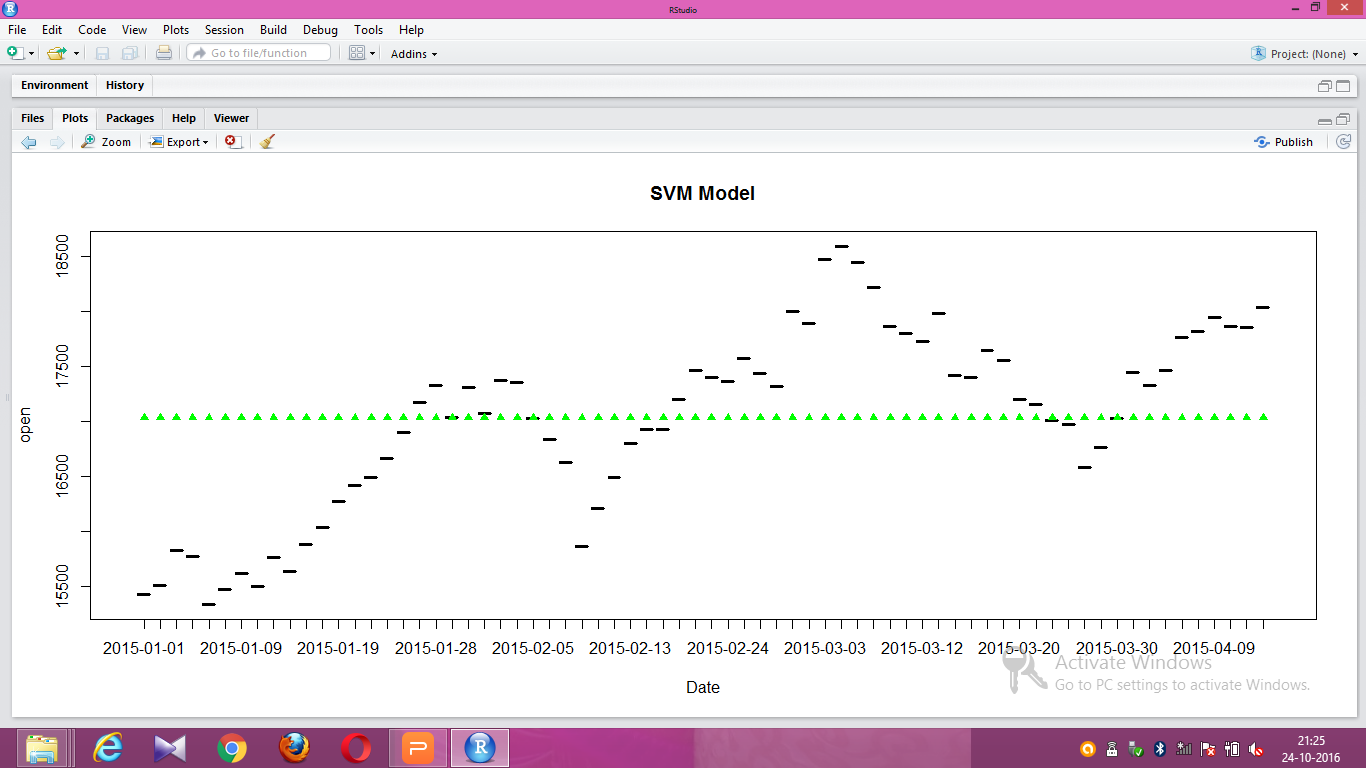
* Transform data to the format of an SVM package
* Conduct simple scaling on the data
* Consider the RBF kernel K(x, y) = e −γkx−yk
* Use cross-validation to find the best parameter C and γ
* Use the best parameter C and γ to train the whole training set
* Test

The future forecast for all the attributes in stock market data is forecasted using svm and result is show below:



*Figure 6:SVM model experimented in R*

The future forecast for open attribute in stock market data is forecasted using svm and result is show below:



*Figure 7:Prediction of open attribute in stock market data using SVM*

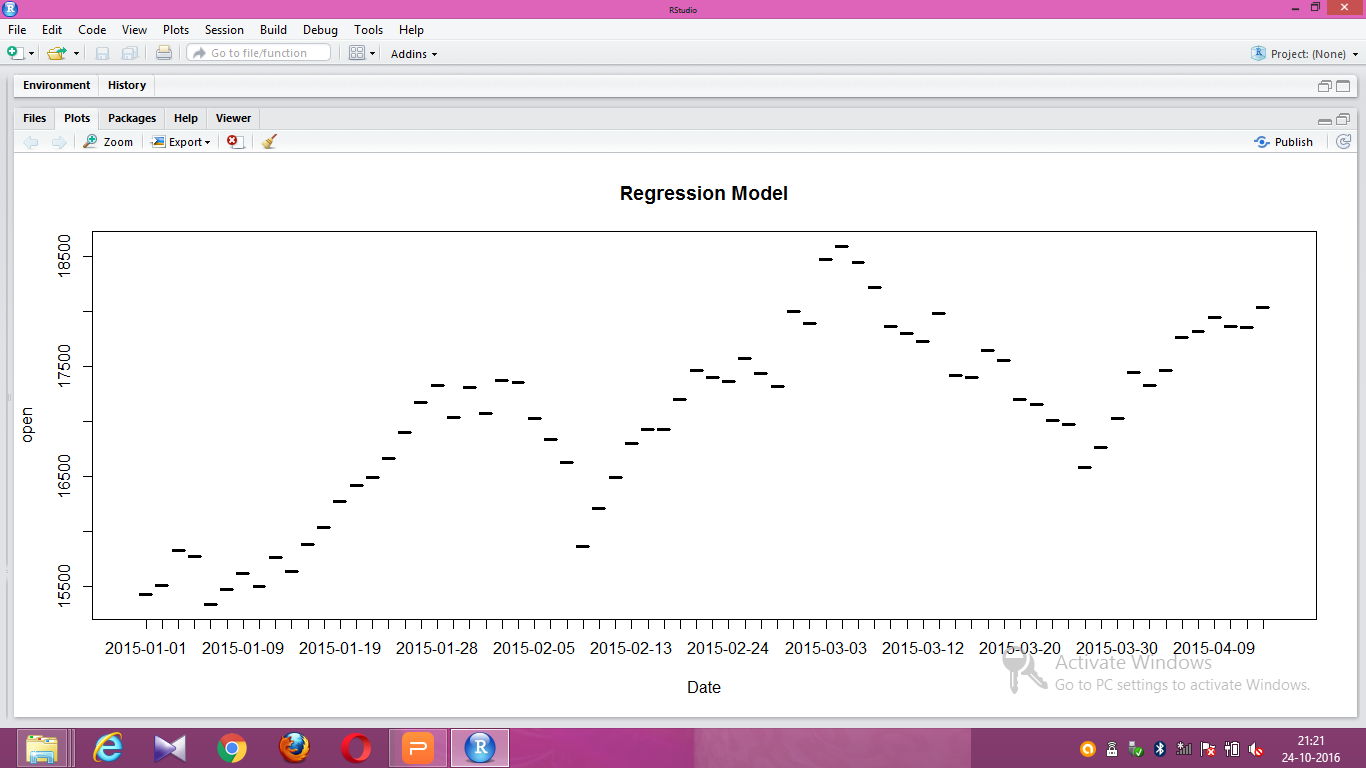
* Since,the selection of the kernel function parameters - for Gaussian kernels the width parameter [sigma] - and the value of [epsilon] in the [epsilon]-insensitive loss function.
* However, from a practical point of view perhaps the most serious problem with SVMs is the high algorithmic complexity and extensive memory requirements of the required quadratic programming in large-scale tasks.

**CHAPTER 7**

**7 Experiments and Results**

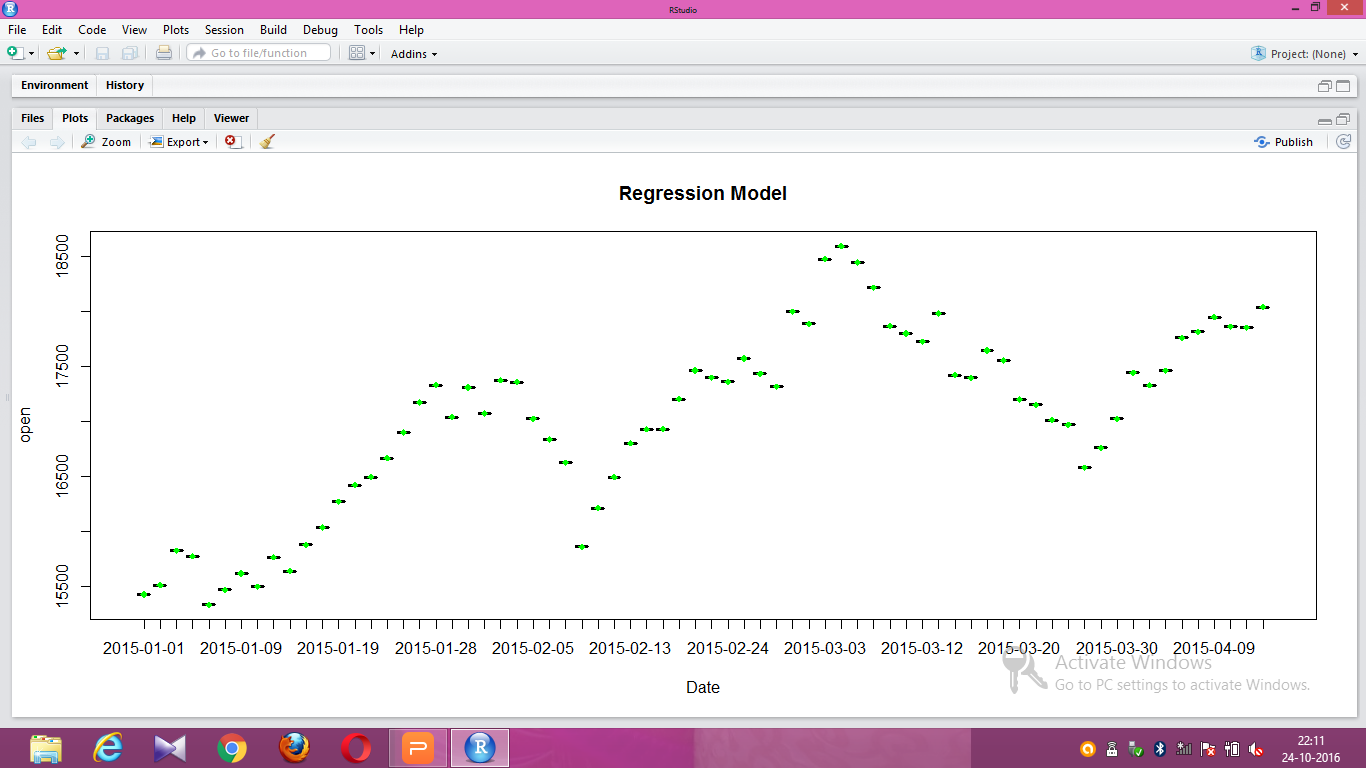
**7.1 Experiments**

**7.1.1 Regression model**

Support Vector Regression attempts to minimize the generalization error bound so as to achieve generalized performance. The idea of SVR is based on the computation of a linear regression function.The future forecast for all the attributes in stock market data is forecasted using svr and result is show below:

*Figure 8: Regression model of stock market data experimented in R*

The future forecast for open attribute in stock market data is forecasted using svm and result is show below:

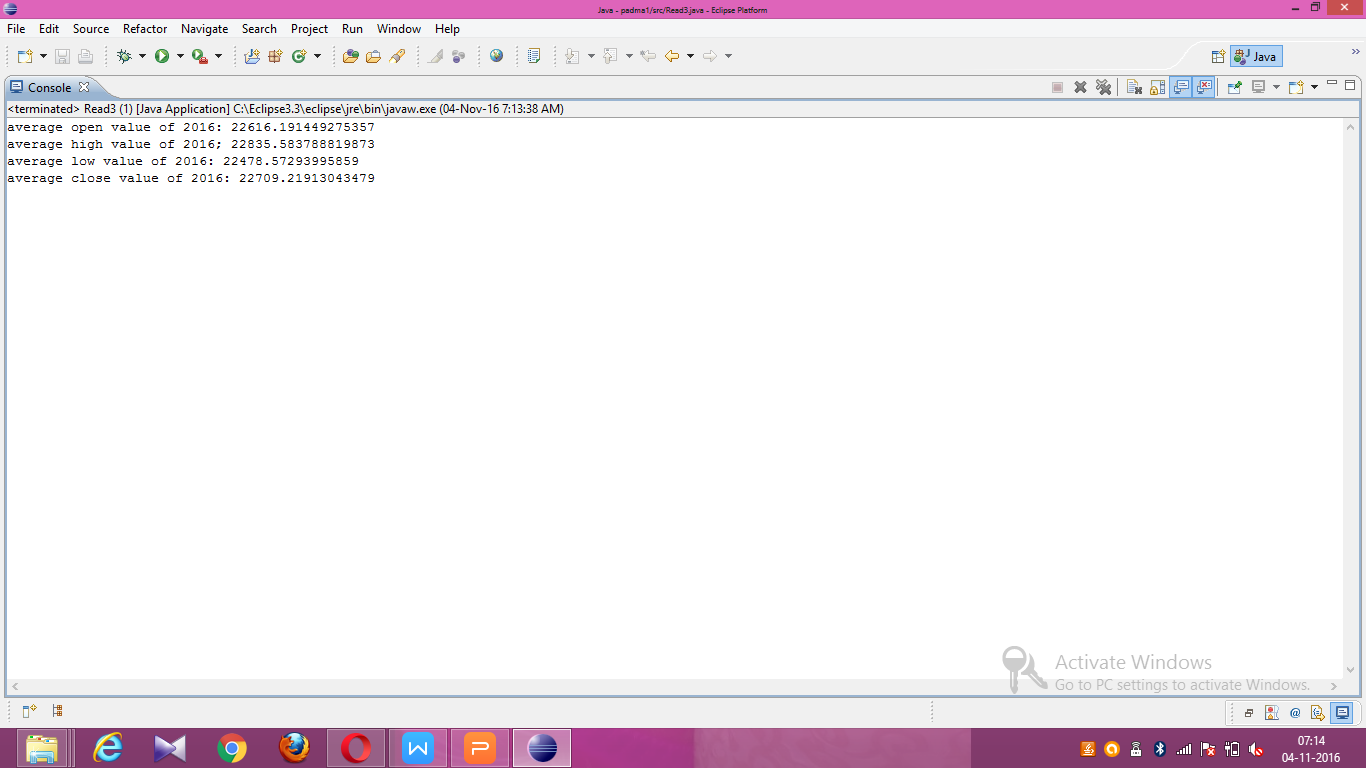


*Figure 9: Prediction of open attribute in stock market data using SVR*

**7.1.2 Moving Average Method**

Moving average is a calculation to analyze data points by creating series of averages of different subsets of the full data set.

The future forecast for all the attributes in stock market data is forecasted using moving average method and result is show below:

**

*Figure 10: Moving Average predicted value*

**7.2 Results Analysis**

The predicted value obtained from the Support vector regression and Moving Average method gives appropriate values of the present years and the future values is also predicted for the stock market.Our method was able to predict the shares around two years with a very high accuracy. Our methodology provided an average accuracy of 87%.





*Figure 11: Comparison of RMSE in MA and SVR*

**CHAPTER 8**

**8. Conclusion**

In this paper we predicted the future values in the stock market using Support vector regression and Moving Average method.The predicted value obtained from the algorithm will be nearly equal to the future values in the stock market.

**CHAPTER 9**

**9. REFERENCES**

1. *Albert Y. Chen,Tsung-Yu Lu,”Demand Forecast Using Data Analytics for the Preallocation of Ambulances”,.IEEE Journal of Biomedical and Health Infomatics, vol. 20, no. 4, July 2016.*
2. *Aditya N. Deshmukh, M. C. Deo, Prasad K. Bhaskaran, T. M. Balakrishnan Nair, and K. G. Sandhya,”Neural-Network-Based Data Assimilation to Improve Numerical Ocean Wave Forecast,IEEE Journal of Oceanic Engineering”, vol. 41, no. 4, October 2016.*
3. *Pritam Bose, Nikola K. Kasabov, Lorenzo Bruzzone, Fellow,”Spiking Neural Networks for Crop Yield Estimation Based on Spatio temporal Analysis of Image Time Series”,IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 11, November 2016.*
4. *Ye Ren, Ponnuthurai Nagaratnam Suganthan, and Narasimalu Srikanth,”A Novel Empirical Mode Decomposition With Support Vector Regression for Wind Speed Forecasting",IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 8, August 2016.*
5. *Amin Hedayati Moghaddama, Moein Hedayati Moghaddamb, Morteza Esfandyaric.”Stock market index prediction using artificial neural network”,Journal of Economics, Finance and Administrative Science 21 (2016) 89–93.*