

Regression Analysis for Stock Market Prediction using Weka Tool without Sentiment Analysis

Sudip Padhye* and Karuna Gull**

*K.L.E. Institute of Technology, Hubli, Karnataka
sudip.padhye@gmail.com

**K.L.E. Institute of Technology, Hubli, Karnataka
karunagull74@gmail.com

Abstract: The paper deals with prediction of stock value for upcoming week, for various companies by knowing the previous weeks stock data. Stock data for different companies are collected from NSE India, which are given as input to the various regression methods using weka tool, the output of which is tabulated. By applying the numerical methods on this tabulated data we get aggregation. Thus aggregated data which is the predicted value for the successive week, is then compared with the actual data picked from NSE after completion of that successive week. Finally statistical analysis is depicted for comparison with the help of graphs.

Keywords: Regression Analysis, Weka Tool, Stock Prediction, Numerical Formula, Technical Analysis, Sentiment Analysis, Pre-processing, Weka Forecasting.

Introduction

Introduction to stock market

The Indian stock market is controlled by two regulatory bodies – NSE (National Stock Exchange) and BSE (Bombay Stock Exchange). This paper retrieves stock data only from NSE. The National Stock Exchange (NSE) is India's leading stock exchange covering various cities and towns across the country [1]. NSE was set up by leading institutions to provide a modern, fully automated screen-based trading system with national reach. NSE offers trading in Equities, Derivatives and Debt (Corporate bonds). This paper deals with only NSE Equities. The Equity stocks can be classified on the basis of stock price (Small-cap, Mid-cap & Large-cap), based on Sector of their operation (Banking, IT, Pharmaceutical, Automobile etc.) or based on stock's fundamentals (past record, turnovers etc.). Irrespective of classification, this paper put forth an approach to predict the stock prices.

Introduction to machine learning

The Machine Learning paradigms are of three types – Supervised Learning, Unsupervised Learning and Reinforcement Learning. The paper uses Supervised Learning approach as the basis for prediction. In supervised learning, initially training is given for later testing. Supervised are classified as Classification and Regression problems. The classification problem determines discrete set of values and thus is used to classify the given data. On the other hand, regression problem are used for prediction. They use continuous set of values for the same. Hence, several regression methods are used in this paper, for predicting the stock values.

Regression analysis [2] is used when anyone wants to predict a continuous dependent variable from a number of independent variables. Regression models are often constructed based on certain conditions that must be verified for the model to fit the data well, and to be able to predict accurately.

Application of machine learning to stock prediction

Fundamental and Technical methods can be used to do the prediction in stock market for different companies.

Fundamental method [3] depends on the past records and news updates of a company which is under prediction. Fall and rise of the stock of a particular company mainly depends on the good and bad news. So consideration of these strong and weak records spread by the news bulletins drastically affect the company's stock market. So more precisely we can say sentiment analysis on data sources collected from business news channels and newspaper headlines etc. with past record analysis of the company under prediction yields more accurate results.

Technical method depends on the weekly data [7]. Thus the stock value can be predicted using Weka forecast package. More specifically doing comparison of different algorithms over considerable amount of data (33 weeks data in our case) to predict future position of stock market.

Even though inclusion of fundamental with technical methods are giving the most accurate results, but here we are applying only the Technical method to predict the stock value for a set of various companies irrespective of any type (like large cap, small cap, medium cap, sectors etc.) to start with.

Literature review

As per the related work done by us, it tells why we have chosen the regression methods for analyzing and predicting the stock data collected. The summary of the literature is given below.

“Generally, higher highs and higher lows indicate an uptrend, whereas lower highs and lower lows mean a downtrend. So, look at the trend. Look at news related to the stock”, says Shrikant Chouhan, senior vice president, technical research, Kotak Securities [13]. Thus, he tries to convince that Stock Predictions can be done using both Technical and Sentiment Analysis for better accuracy. The stock not only depends upon its movement trends but also on the news affecting that particular stock or the news regarding a sector(s) related to that stock.

Analysts and market experts take the help of various parameters to confirm if a stock is a trade pick. These include 200-day Moving Average, Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Fibonacci Retracement and Candle-stick price chart (Support and Resistance). Thus, only few of the existing methods use Machine Learning – Regression Analysis as a technique for prediction. Moreover, unlike Regression Analysis, the existing techniques don’t provide 90%+ accuracy.

According to Rupinder Kaur and Ms. Vidhu Kiran [4], the prediction was done using ANN (Artificial Neural Networks) for forecasting features like current stock price, opening price, closing price, volume, change and change percent. The System is trained using NARX Neural Networks where the errors are back propagated to the system from output layer. At the end, the predicted values are compared with the actual values using Mean Squared Error (MSE) technique by plotting the graph.

Farhad Soleimani Gharehchopogh, Tahmineh Haddadi Bonab and Seyyed Reza Khaze [5], carried out the work by undertaking a Case Study that uses Linear Regression to predict stock market. They have discussed how the trading volume is one of the important parameter affecting the dynamics of the stock market. Using this approach, they observed 61.35% accuracy of results. Hence, linear regression approach for prediction does not always give the correct results i.e. with greater accuracy of greater than 90%.

K.K. Sureshkumar and Dr.N.M. Elango [6], forecasted Indian Stock Market Price and obtained their performance analysis by describing the usage of different neural classifiers of Weka tool. The actual stock price of Infosys technology over 1000 days had been compared with the predicted value of the Gaussian, Isotonic regression, Least mean square, Linear regression, Multilayer Perceptron functions, Pace regression, Simple linear regression and SMO regression values. Later, they applied the error evaluation rate indicators such as mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE) and root relative squared error (RRSE). They concluded that isotonic regression function offers the ability to predict stock price more accurately.

Vishal Parikh and Parth Shah [8] described Stock Prediction using Automated Trading System that shows how Stock value prediction can be done using Naïve-Bayes classifier and Random Forest classifier and then the robustness is evaluated using precision, recall and accuracy.

Raj Kumar and Anil Balara [10], made Time series forecasting of Nifty Stock Market using Weka in which SMO regression offered ability to predict more accurately than other functions such as Gaussian processes, linear regression and multilayer perception.

Methodology

Concept of Regression

In Regression problems, we take input variables and try to map the output onto a continuous expected result function [11]. Regression with one variable is known as “Uni-variate Regression” whereas regression with multiple variables is called as “Multi-variate Regression”. “Uni-variate regression is used when you want to predict a single output value from a single input value. On the other hand, Multi-variate regression is used to predict from multiple input values [12].

Regression models involve the following variables [15]:

- The unknown parameters, denoted as β , which may represent a scalar or a vector.
- The independent variables, X .
- The dependent variable, Y .

In various fields of application, different terminologies are used in place of dependent and independent variables.

A regression model relates Y to a function of X and β .

$$Y \approx f(X, \beta) \quad \text{----- (1)}$$

The approximation is usually formalized as

$$E(Y | X) = f(X, \beta) \quad \text{----- (2)}$$

Various regression techniques used in this paper are,

Regression By Discretization

Regression By Discretization[16] is a regression technique that requires the class attribute of the data to be discretized or with equal width. The predicted value is the expected value for each discretized interval based on the predicted probabilities for each interval.

RBF Regressor

RBF (Radial Basis Function) Regressor[17] implements RBF Networks by training in a supervised manner by minimizing squared error with the BFGS (Broyden–Fletcher–Goldfarb–Shanno algorithm) method. All attributes are normalized into the [0,1] scale, including the class. The ridge parameter is used to penalize the size of the weights in the output layer, which implements a simple linear combination. It is also possible to use conjugate gradient descent rather than BFGS updates, which is faster for cases with many parameters, and to use normalized basis functions instead of un-normalized ones. Nominal attributes are processed using the unsupervised Nominal to Binary filter and missing values are replaced globally using Replace Missing Values.

Linear Regression

The Linear Regression[11] is a popular technique, wherein the hypothesis (decision-boundary) obtained is in the form of line. The hypothesis function has the general form:

$$h_{\theta}(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n = \theta^T x \quad \text{----- (3)}$$

where, θ is parameter vector x is the feature vector.

The cost function, which is the difference between the predicted value and the actual value, is given by,

$$J(\theta_0, \theta_1 \dots \theta_n) = \left(\frac{1}{2m}\right) \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \quad \text{----- (4)}$$

where, n is no. of features & m is no. of training sets.

Finally, the cost function can be minimized by finding minimum value of θ , which can be determined either using Gradient Descent or Normal Equation. If we compare Gradient Descent and Normal Equation then Gradient Descent requires correct choice of α and it also requires many iterations. Conversely, Normal equation doesn't contain α and also does not iterate. On the other hand, Gradient Descent can be used on any number of features (input variables) in training data, unlike the normal equation, which can be used to calculate θ values for limited number of features (upto 10,000 features) due to complexity in computation of $(X^T X)^{-1}$. Also, the Gradient Descent finds the local minima whereas Normal Equation helps in finding Global minima.

Multilayer Perceptron

Multilayer perceptron [18] is a type of Neural Network. It is a Classifier that uses back-propagation to classify instances. This network can be built by hand, created by an algorithm or both. The network can also be monitored and modified during training time. The nodes in this network are all sigmoid/logistic.

MLP Regressor

MLP Regressor[19] trains a multilayer perceptron with one hidden layer. All attributes, including the target attribute, are standardized. There are several parameters - the ridge parameter is used to determine the penalty on the size of the weights. The number of hidden units can also be specified using ridge parameter. Finally, it is possible to use conjugate gradient descent rather than BFGS updates, which may be faster for cases with many parameters. To improve speed, an approximate version of the logistic function is used as the default activation function for the hidden layer, but other activation functions can also be specified. In the output layer, the sigmoid function is used for classification. If the approximate sigmoid is specified for the hidden layers, it is also used for the output layer. For regression, the identity function is used activation function in the output layer. Also, if delta values in the back-propagation step are within the user-specified tolerance, the gradient is not updated for that particular instance, which saves some additional time. For large datasets, parallel calculation of loss function and gradient is possible when multiple CPU cores are present. Thus in this case, the data is split into batches and processed in separate thread.

SMOReg

The SVM (Support Vector Machine) is used for implementing SMOReg[20]. The SMOReg implements sequential minimal optimization algorithm for training a support vector regression using polynomial or RBF kernels. This implementation globally replaces all missing values and transforms nominal attributes into binary ones. It also normalizes all attributes by default. Thus, the coefficients in the output are based on the normalized/standardized data, not the original data.

Gaussian Processes

Gaussian processes[21] implements regression without hyper-parameter-tuning. This method applies normalization or standardization to the attributes. Any missing values are replaced by the global mean/mode. Also, Nominal attributes are converted to binary ones.

M5P

M5P[22] is a reconstruction of Quinlan's M5 algorithm for inducing trees of regression models. M5P combines a conventional decision tree with the possibility of linear regression functions at the nodes. M5P can deal effectively with enumerated attributes and missing values. First, a decision-tree induction algorithm is used to build a tree, for which a splitting criterion is used that minimizes the intra-subset variation in the class values down each branch, instead of maximizing the information gain at each inner node. The splitting procedure in M5P stops if the class values of all instances that reach a node vary very slightly, or only a few instances remain. Second, the tree is pruned back from each leaf. When pruning, an inner node is turned into a leaf with a regression plane. Third, to avoid sharp discontinuities between the sub-trees, a smoothing procedure is applied that combines the leaf model prediction with each node along the path back to the root, smoothing it at each of these nodes by combining it with the value predicted by the linear model for that node. Smoothing substantially increases prediction accuracy.

Packages required

Weka is a data mining tool. The Weka workbench is a collection of state-of-the-art machine learning algorithms and data pre-processing tools. It is designed so that users can quickly try out existing machine learning methods on new datasets in very flexible ways. The required packages can be installed using "Package Manager" in "Tools" menu of "Weka GUI Chooser". The various packages required on Weka 3.7.2 or above versions, to make regression on data are:

1. Forecast package (for finding results for building and evaluating Time series forecasting models)
2. Arff converter (to convert given .xlsx file to .arff file format)
3. Linear regression
4. Gaussian regression
5. RegressionByDiscretization
6. RBFRegressor
7. MultilayerPerceptron
8. MLPRegressor
9. SMOReg
10. M5P

Note: If any of the above mentioned package is not installed, then they can be installed using Package Manager available from Weka version 3.7.2 onwards.

Steps for weka prediction

1. Open Weka in "Explorer" Mode and choose the pre-processed file.
2. Later, click on "Forecast" menu and, in "Basic configuration", choose the outputs to predict (Avg. High and Avg. Low) in "Target Selection". Then, select "Number of time units to forecast" as 1 (in order to make prediction of upcoming 1 week), "Time stamp" as Week and "Periodicity" as Weekly.
3. In "Advanced configuration", choose different regression methods and click on "Start" to start the prediction by keeping the rest of the settings as default.
4. On predicting, the values for the next week will be presented, along with the Time stamp marked with *.
5. Weka also provides Visualization of the data along with the predicted values in the form of Line graph.

Flow Diagram

Collect the required data from the NSE for different companies like KPIT, Bajaj Auto, L&T etc. Pre-process the data as per our requirement like converting the daily stock data to weekly stock data. Then create the files containing pre-processed data in the form needed by regression methods. The output of different regression methods yield the predicted stock values for upcoming week. Now use the numerical method to normalize the data using the formula defined by the author in the equation (5).

Estimated value =

$$\frac{7(\text{RegressionByDiscretization}) + 6(\text{RBFRegressor}) + 5(\text{Linear Regression} + \text{MultiLayerPerceptron} + \text{MLPRegressor} + \text{SMOReg}) + 4(\text{GaussianProcesses} + \text{M5P})}{41} + \delta \quad \text{----(5)}$$

Where, δ is the stochastic error factor

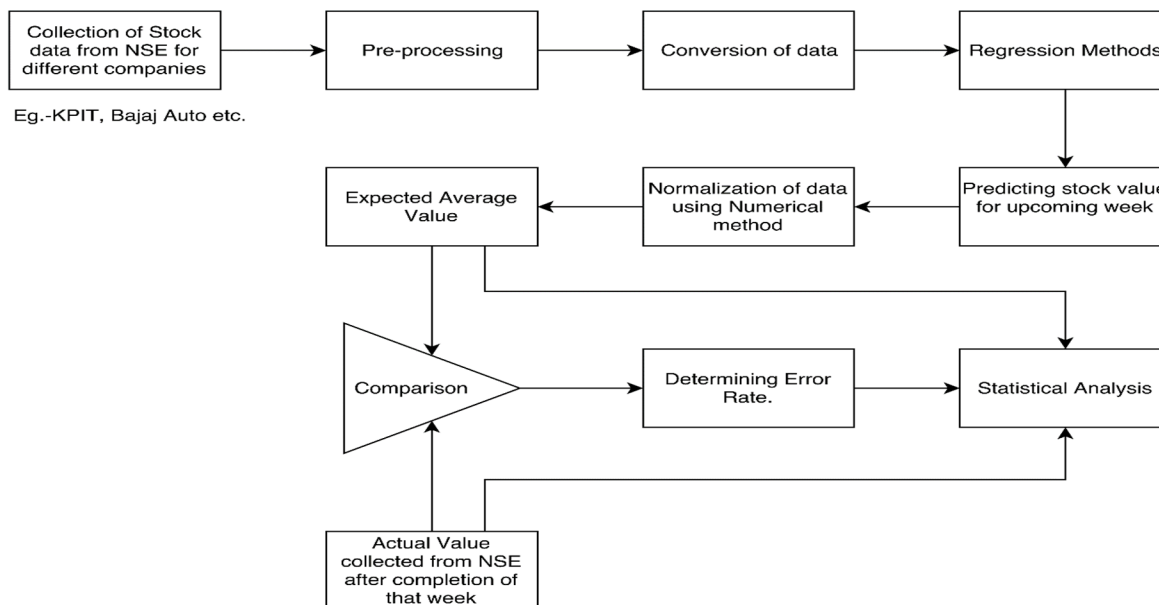


Figure 1. Technical Analysis Flow-Chart

The result of which is an expected average values. These average values are now compared with actual values picked from the NSE after completion of that week. Comparison results lead to a deviation in percentage as an error rate. Finally the graph is drawn for the average values vs. actual values and vs. the error rates.

Analysis of data

Analysis is done in following steps,

Collection of data

Raw data collected from NSE India for different companies is depicted in Table 1.

Table 1. Raw data collected from NSE for different companies

Comp any		KPIT		L&T		Bajaj Auto		Ashok Leyland		Axis Bank		LIC Hsg. Fin.	
Week	Date	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
Week 1	29/06/15	97.25	89.50	1792.7	1720	2574.05	2480.45	71.25	68.5	567.3	546	454	431.2
	30/06/15	95.80	92.95	1807.9	1768.15	2584	2531.1	72.7	71.05	566	553	456	448
	01/07/15	96.50	93.10	1818.15	1783	2555	2516	74.9	72.9	580.5	556.85	459.8	450.8
	02/07/15	95.70	93.10	1817.6	1792.15	2586.2	2528	76.4	73.65	586.65	576.1	461	450.05
	03/07/15	94.40	93.05	1823.8	1792.3	2609	2575	73.95	72.4	588.5	581	454.75	443.4
:	:	:	:	:	:	:	:	:	:	:	:	:	:
Week 33	15/02/16	123	115.5	1154.7	1067.95	2372.3	2305	90	84.7	420	395.65	432.8	414.2
	16/02/16	124	119.1	1165	1101.1	2368.7	2328.55	91	86.9	422.5	405	431.1	408.85
	17/02/16	121.3	115.2	1133.1	1082.65	2369	2263.8	90.5	85.1	406.9	394.75	414	402.05
	18/02/16	123.85	118.25	1174.65	1133.05	2417.65	2364.1	91.1	86.75	412.1	391.6	417	405.1
	19/02/16	126	117.95	1168.9	1131.4	2467	2406.15	91.5	88.85	396.9	390	421.2	406.75

Pre-processing

Raw data is pre-processed and shown in table 2.

Table 2. Pre-processing of the Raw data collected from NSE for different companies

Comp any	KPIT		L&T		Bajaj Auto		Ashok Leyland		Axis Bank		LIC Hsg. Fin.	
Week	Avg-High	Avg-Low	Avg-High	Avg-Low	Avg-High	Avg-Low	Avg-High	Avg-Low	Avg-High	Avg-Low	Avg-High	Avg-Low
1	95.67	93.13	1836.9	1800.68	2581.05	2533.66	73.84	71.7	577.79	562.59	457.11	444.69
2	105.68	99.78	1879.25	1855.75	2539.51	2487.59	73.15	71.18	584.39	573.39	451.39	439.65
3	114.79	108.32	1866.02	1828.04	2585.19	2505.16	74.86	73.11	596.78	583.21	468.3	456.61
4	111.44	105.93	1798.79	1755.25	2531.65	2483.64	81.15	78.76	598.04	581.71	491.31	473.06
5	130.42	122.98	1808.82	1765.94	2573.27	2511.98	84.63	82.18	573.33	559.26	497.2	483.05
	:	:	:	:	:	:	:	:	:	:	:	:
29	159.85	149.47	1175.4	1131.87	2376.54	2327.07	92.96	91.05	410.67	391.72	497.47	480.66
30	146.26	136.92	1128.4	1085.22	2308.04	2246.28	92.09	87.91	402.57	385.94	476.02	460.21
31	146.5375	141.4125	1121.85	1092.9	2337.875	2279.825	85.99	82.67	420.76	408.66	483.26	475.57
32	142.1	135.4	1144	1112.05	2382.03	2308.65	88.44	86.82	399.28	383.97	460.49	443.76
33	131.07	123.91	1128.48	1084.82	2376.49	2313.59	88.45	83.77	402.91	386.95	433.98	415.49

Output of Weka tool for various regression methods

The output of the various regression methods for highs and lows of various companies are shown in table 3 and table 4.

Table 3. Output (**High values**) of Weka tool for various regression methods

Stock	Regression ByDiscretization	RBF Regressor	Linear Regression	MultiLayer Perceptron	MLP Regressor	SMOReg	Gaussian Processes	M5P
KPIT	133.49	130.9485	133.5284	123.0326	131.6938	127.4935	118.1059	131.0725
LT	1158.63	1115.452	1493.4073	1118.9145	1123.462	1083.2275	1078.6941	1108.3348
BAJAJ AUTO	2369.9853	2319.142	2478.3914	2406.3816	2397.206	2375.7498	2378.303	2412.5801
ASHOK LEYLAND	88.7886	92.8258	88.6225	87.8416	87.8343	89.1433	88.418	91.3648
AXIS BANK	403.8575	407.5133	402.1328	413.0255	406.7175	398.3615	387.7654	395.3627
LIC HSG. FIN.	454.5125	420.1207	359.9999	433.3918	446.707	456.1832	453.7118	432.2268

Apply the numerical method on final data collected from Weka tool

No single method gave correct predictions always, thus there was a necessity to take average of all methods. The averages allocates equal priorities to all methods, thus we decided to go for weighted averages. We carried out 41 experiments, out of which results were provided 7 times by Regression By Discretization, 6 times by RBF Regressor, 5 times by Linear Regression, 5 times by Multilayer Perceptron, 5 times by MLP Regressor, 5 times by SMOReg, 4 times by Gaussian Processes and 4 times by M5P. Hence, we assigned weights to each method using their success frequency and derived the Numerical Method formula (Equation 6) shown below.

Table 4. Output (**Low values**) of Weka tool for various regression methods

Stock	RegressionBy Discretization	RBF Regressor	Linear Regression	MultiLayer Perceptron	MLP Regressor	SMO Reg	Gaussian Processes	M5P
KPIT	135.88	124.473	127.9339	115.8912	119.0986	115.1329	110.9317	129.8308
LT	1121.7567	1073.2648	1460.0694	1064.9707	1077.3879	1027.3228	1044.4092	1062.6313
BAJAJ AUTO	2267.1926	2254.574	2434.0913	2309.141	2308.8375	2313.1226	2312.4154	2332.3748
ASHOK LEYLAND	86.7543	77.4038	87.5913	78.27	82.4829	87.7574	84.9615	87.4764
AXIS BANK	387.145	390.3497	376.9289	386.3938	390.5927	380.2483	371.8556	387.9576
LIC HSG. FIN.	451.391	401.9892	333.3012	410.1261	432.5221	436.4709	438.6235	421.5869

Estimated value =

$$\frac{7(\text{RegressionByDiscretization}) + 6(\text{RBFRegressor}) + 5(\text{LinearRegression} + \text{MultiLayerPerceptron} + \text{MLPRegressor} + \text{SMOReg}) + 4(\text{GaussianProcesses} + \text{M5P})}{41} + \delta \quad \text{----(6)}$$

Here, δ represents small variation in data which can be minimized by considering large dataset and by using Sentiment Analysis.

Comparison of aggregate value with actual NSE value in the table form

Find the aggregation of the output for all regression methods for lows and highs and also find the deviation of result from their actual values by error as shown in table 5 and 6.

Table 5. Average high and actual values and error rate for various companies

Stock	Estimated Avg. High	Actual Avg. High	% Error
KPIT	129.16	123.63	4.47
LT	1162.10	1159.27	0.24
BAJAJ AUTO	2389.19	2398.93	-0.41
ASHOK LEYLAND	89.39	90.82	-1.58
AXIS BANK	402.58	411.68	-2.21
LIC HSG. FIN.	432.38	423.22	2.16

Table 6. Average low and actual values and error rate for various companies

Stock	Estimated Avg. Low	Actual Avg. Low	% Error
KPIT	123.20	117.2	5.12
LT	1118.75	1103.23	1.41
BAJAJ AUTO	2312.27	2333.52	-0.91
ASHOK LEYLAND	83.95	86.46	-2.90
AXIS BANK	384.44	395.4	-2.77
LIC HSG. FIN.	416.45	407.39	2.22

Experimental Results

The statistical analysis of the experimental results Vs Actual values and Error rates were depicted using various figures Fig 2 to Fig 9.

Graphs for various regression method

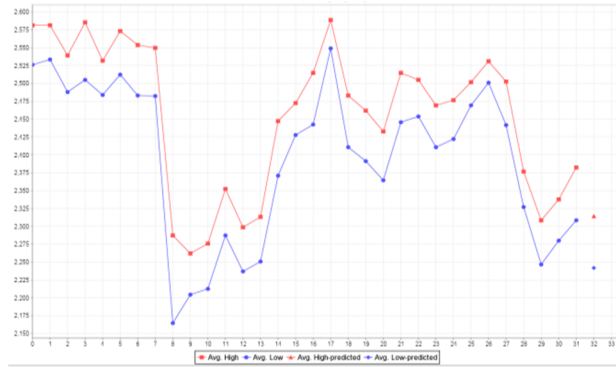


Figure 2. Bajaj auto predicted data – RBFRegressor

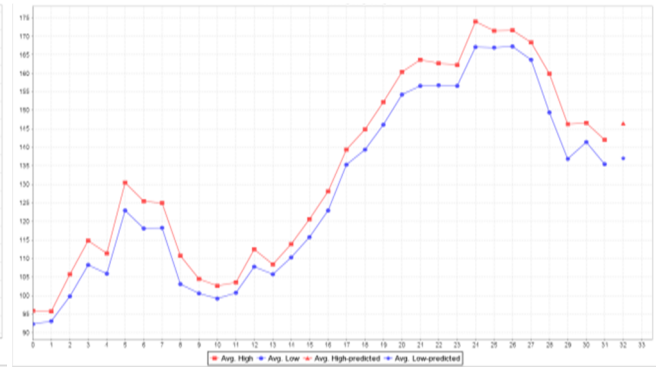


Figure 3. KPIT predicted data – RegressionByDiscretization

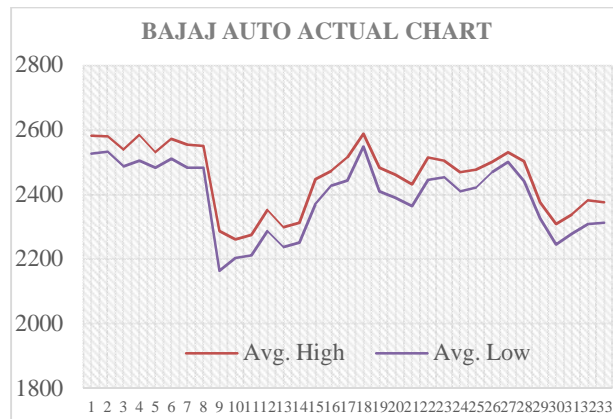


Figure 4. Bajaj Auto Actual Data

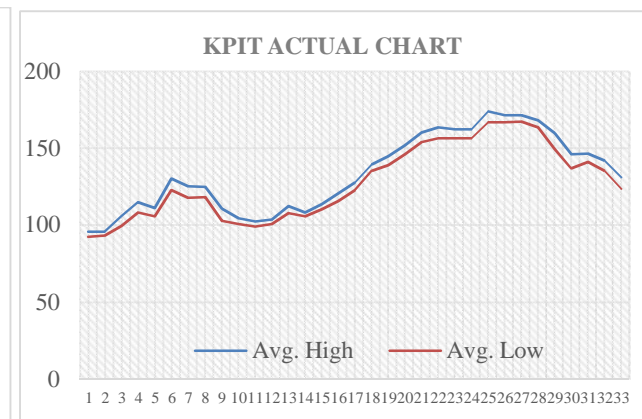


Figure 5. KPIT Actual Data

Bar chart for comparison of estimated value and actual value

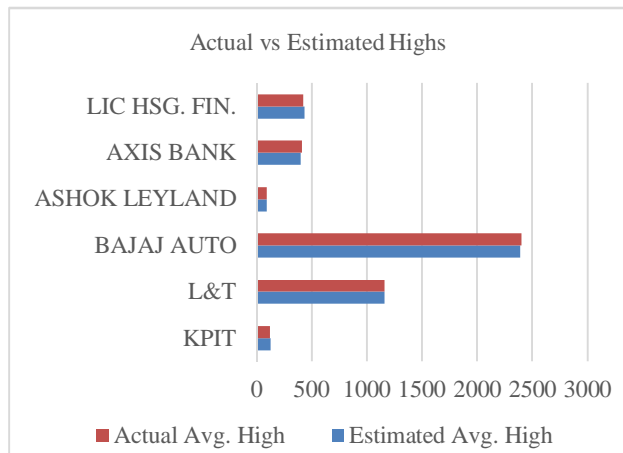


Figure 6. Actual vs. Estimated Highs

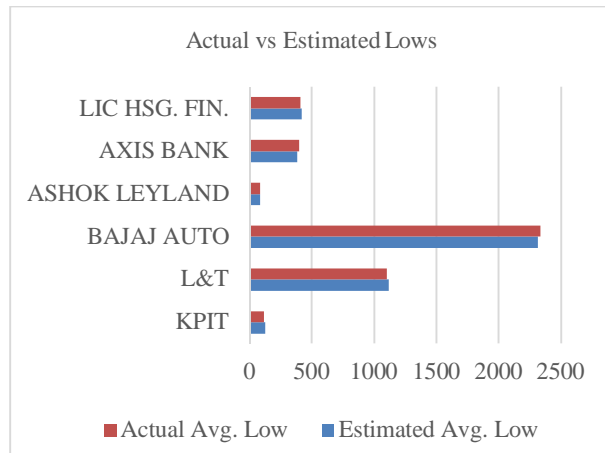


Figure 7. Actual vs. Estimated Lows

Column chart for %Error

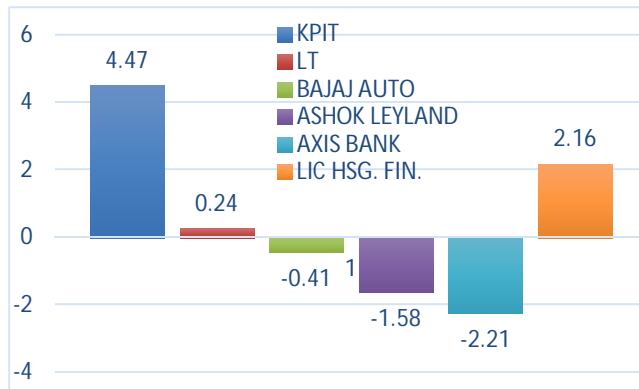


Figure 8. Error% for Highs

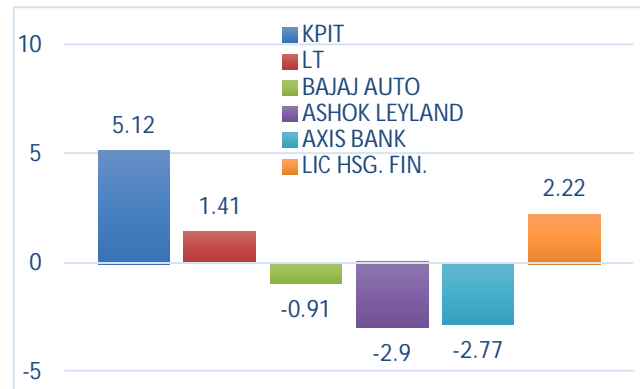


Figure 9. Error% for Lows

Performance Evaluation

Table 7. Performance Evaluation by computing error

Functions	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)	Relative Absolute Error (RAE)	Root Relative Squared Error (RRSE)
RegressionByDiscretization	20.1616	28.9586	2.68%	2.64%
RBF Regressor	21.8251	34.9426	2.89%	3.27%
Linear Regression	52.2119	106.5071	6.93%	9.69%
MultiLayer Perceptron	16.3219	22.9255	2.17%	2.09%
MLP Regressor	14.3147	19.4126	1.90%	1.78%
SMOReg	21.2613	31.9352	2.82%	2.91%
Gaussian Processes	23.5843	21.7965	3.13%	3.01%
M5P	15.1037	32.557	2.01%	1.98%
Numerical estimation	10.1396	13.1145	1.35%	1.19%

At the end, the performance of the predictor is calculated by comparing the estimated value with the actual value. There are various methods for evaluation such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE) and Root Relative Squared Error (RRSE). Formulas for calculating these errors are as given below

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x'_i| \quad \dots\dots (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - x'_i)^2}{n}} \quad \dots\dots (8)$$

$$RAE = \frac{\sum_{i=1}^n |x_i - x'_i|}{\sum_{i=1}^n x_i} \quad \dots\dots (9)$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^n (x_i - x'_i)^2}{\sum_{i=1}^n x_i^2}} \quad \dots\dots (10)$$

where, x_i is actual stock value and x'_i is the predicted stock value.

The Relative errors are represented w.r.t. the standard value (actual stock value) whereas the rest are computed w.r.t. total no. of values (n). Out of these four methods, RMSE and RAE are very common among several literatures. The MAE and RMSE are represented in decimal format while RAE and RRSE are represented in percentage. Lesser the error, more is the accuracy of it. Thus, our Numerical Estimation Formula shows least error for all four methods as compared to other individual regression functions.

Conclusion and future work

To conclude we say that only few of the existing methods use Machine Learning – Regression Analysis as a technique for prediction. But however the expected results from them is not as accurate as we got i.e. those techniques don't provide 90%+ accuracy. So taking these as main drawbacks of existing ones, we started by considering two methods fundamental and

technical to predict the stock values for upcoming week. But we end up with a completion of only the Technical method for prediction in depth.

Technical method depends on the weekly data and the Weka forecast package. More specifically the comparison is done on the data collected for different companies using various regression methods to predict the future position of the stock market. This paper began by stating that irrespective of the type or the sector of the stock, predictions with less errors can be made. In our consideration of 6 stocks, KPIT belongs to IT Sector, Ashok Leyland and Bajaj Auto belongs to automotive sector, Axis Bank is a banking stock whereas LIC Housing Finance is a Finance Stock & L&T is a diversified stock company. Thus, in spite of this variation in nature of stocks, our Numerical method formula proved successful with minimal error. The comparison result of the algorithms had given more accurate results i.e. by showing less deviation from actual values collected after completion of the week under prediction. Statistical analysis drawn show the predicted results and accuracy of the methods defined.

As a future work if Fundamental and Technical methods are combined and used for prediction of stock market position of any company they give the most accurate results. Because Fundamental method depends on the past records and news updates of a company which is under prediction. Fall and rise of the stock of a particular company mainly depends on the good and bad news. So consideration of news data really affect the company's stock market. Thus addition of sentiment analysis [9] on data collected from different news media along with past history analysis of the company under prediction will give way to most accurate results.

References

- [1] Oscar Alsing, OktayBahceci, "Stock Market Prediction Using Social Media Analysis", Thesis, Computer Science, stockholm, sweden 2015.
- [2] Kim-Georg Aase, "Text Mining of News Articles for Stock Price Predictions", Thesis, Master of Science in Computer Science, Norwegian University of Science and Technology, 2011.
- [3] Ruchi Desai, Prof.Snehal Gandhi, "Stock Market Prediction Using Data Mining", International Journal of Engineering Development and Research (IJEDR) - 2014, Volume 2, Issue 2, ISSN: 2321-9939, pp 2780-2784.
- [4] Rupinder kaur1, Ms.VidhuKiran, "Efficient Artificial Neural Network based Practical Approach of Stock Market Forecasting", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 8, August 2015, DOI 10.17148/IJARCC.2015.4855,ISSN (Online) 2278-1021, ISSN (Print) 2319 5940. Pp-263-267.
- [5] FarhadSoleimanianGharehchopogh, TahminehHaddadiBonab and Seyyed Reza Khaze, "A Linear Regression Approach To Prediction Of Stock Market Trading Volume: A Case Study", International Journal Of Managing Value And Supply Chains (IJMVSC) Vol.4, No. 3, September 2013, DOI: 10.5121/ijmvsc.2013.4303, pp-25-31.
- [6] K.K.Sureshkumar ,Dr.N.M.Elango , "An Efficient Approach to Forecast Indian Stock Market Price and their Performance Analysis", International Journal of Computer Applications (0975 – 8887) Volume 34– No.5, November 2011, pp-44-49.
- [7] PierpaoloDondio "Stock Market Prediction Without sentiment Analysis: Using a Web-Traffic Based Classifier and User-Level Analysis", 46th Hawaii International Conference on System Sciences, 2013, DOI 10.1109/HICSS.2013.498 ,1530-1605/12 \$26.00 © 2012 IEEE. pp- 3135- 3144.
- [8] Vishal Parikh, Parth Shah, "Stock Prediction and Automated Trading System", IJCSC, Vol- 6 • Issue-1 Sep - Mar 2015 pp.104-111 available at www.csjournalss.com.
- [9] Yoosin Kim, SeungRyulJeong, Imran Ghani, "Text Opinion Mining to Analyze News for Stock Market Prediction", Int. J. Advance. Soft Comput. Appl., Vol. 6, No. 1, March 2014, ISSN 2074-8523; Copyright © SCRG Publication, 2014
- [10] Raj Kumar, Anil Balara, "Time Series Forecasting Of Nifty Stock Market Using Weka", JRPS International Journal for Research Publication & Seminar Vol 05 Issue 02 March -July 2014, pp-1-6.
- [11] <https://www.coursera.org/learn/machine-learning/home/welcome>
- [12] https://onlinecourses.nptel.ac.in/noc16_cs03/preview
- [13] <http://www.businesstoday.in/moneytoday/stocks/technical-analysis-tools/story/21155.html>
- [14] <http://www.nseindia.com>
- [15] https://en.wikipedia.org/wiki/Regression_analysis
- [16] <http://weka.sourceforge.net/packageMetadata/regressionByDiscretization/index.html>
- [17] <http://wiki.pentaho.com/display/DATAMINING/RBFRegressor>
- [18] <http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/MultilayerPerceptron.html>
- [19] <http://weka.sourceforge.net/doc.packages/multiLayerPerceptrons/weka/classifiers/functions/MLPRegressor.html>
- [20] <http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/SMOreg.html>
- [21] <http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/GaussianProcesses.html>
- [22] <http://www.opentox.org/dev/documentation/components/m5p>