Kernel Linkage Support Vector Regression for Stock Market Index Prediction and Analysis

G. Kavitha[a], S. Bhuvaneswari[b]

[a] Assistant Professor (SG), Department of Mathematics, Hindustan Institute of Technology & Science, Tamilnadu, India
[b] Assistant Professor, Department of Mathematics, DB Jain College, Chennai, Tamilnadu, India
Email: [a]kavithateam@gmail.com, [b]prof.karuna@gmail.com

Article History: Received:11 January 2021; Revised : 12 February 2021 Accepted: 21 March 2021; Published online: 23 May 2021

Abstract:

The study proposes a novel method for stock market index prediction and analysis based on Kernel Linkage Support Vector Regression (KLSVR). The method pre-processes the data to transform it into information with which the decision system can work. A regression model was built for the NASDAQ dataset using Support Vector Machines (SVM) on the training data and testing the model for goodness of fit. The Regression types selected for running the R code was SVM eps-Regression, SVM nu-Regression, Bound Constraint SVM eps-Regression. The training and testing were done with 70-30 combination. The experimental result validates the model proposed using minimum error analysis performance measures.

Keywords: Kernel Linkage Support Vector Regression, R, Error Analysis, Decision System, Stock Prediction

1. Introduction

Information extraction from data to predict trends and certain behavior patterns for regression problems is a challenging task. The main objective of the paper is to predict the stock market non-linear data by using a kernel linkage support vector regression in the R environment, forming a decision system using SVM eps-Regression, SVM nu-Regression, Bound Constraint SVM eps-Regression. There are numerous techniques available in the literature [1-6] for prediction and analysis for support vector regressions. The paper tries to find the goodness of fit for the output vector minimizing the error in the performance measures using the kernel linkage support vector regression in R environment.

2. Materials and Methods

The proposed study deals with the concept of support vector machines exclusively for regression purpose for analyzing the stock trend. Here the response variable is a quantitative variable. The objective of the research work is to do a numerical prediction for NASDAQ stocks. The attributes that are taken for stock prediction are Open, High, Low, Close value, Adjacent close and Volume. Regression is done using the support vector machine. Support vectors showcase the relationship between the attributes 'X' and the response variables 'Y'. The model which is built can be utilized for future analysis and prediction. The regression models used in our experimental study includes, SVM eps-Regression, SVM nu-Regression, Bound Constraint SVM eps-Regression. SVM eps-

Regression, SVM nu-Regression are executed using e1071 package in R and Bound Constraint SVM eps-Regression are executed using kernlab packages in R. SVM and KSVM functions are used for classification. We have built a regression model for the NASDAQ dataset using support vector machines on the training data and testing the model for goodness of fit. Following are the steps for SVM eps-Regression, SVM nu-Regression, Bound Constraint SVM eps-Regression in R environment.

Step 1: Load e1071 package in R with the SVM function

Step 2: Input the dataset with the required attributes and inspect the preliminaries for the content of the dataset

Step 3: The split up for the proposed experimental study is taken as 70 percent for training and 30 percent for testing.

Step 4: The support vector machine is estimated using kernels with parameter values being set.

Step 5: The NASDAQ stock index dataset is taken from January 2015 to June 2020 for prediction and analysis.

Here we would like to model the close value of the NASDAQ dataset as the response variable. We are trying to check whether there is any relationship between the attributes and the response variable. Here, we are trying to use support vector machine for this regression concept. Check the dataset for abnormalities. Support vector machine is used to carry out general regression and classification with type epsilon and nu. Here we have tried to use the close value as such. Then after transformation the transformed close value was used as response variable for the experimental setup. Results are tabulated and analyzed.

3. Experimental Result and Discussion of SVM eps-Regression, SVM nu-Regression, Bound Constraint SVM eps-Regression

The proposed methodology and experimentation have been simulated in R environment. The dimension of the stock taken for the experimental study is 1381 rows and 7 columns. The data in the rows are taken from Jan 2015 to June 2020. The attributes for the columns are chosen as open value, close value, high value, low value, adjacent close value and volume. The results are tabulated. The summary of the experimental study is given below:

The first 6 rows generated from the data population using R is given in Table 1.

Table 1: First 6 rows generated from the data population using R environment

S. No.	Open	High	Low	Close value	Adjacent Close	Volume
1.	4760.24	4777.01	4698.11	4726.81	4726.81	1435150000
2.	4700.34	4702.77	4641.46	4652.57	4652.57	1794470000
3.	4666.85	4667.33	4567.59	4592.74	4592.74	2167320000
4.	4626.84	4652.72	4613.90	4650.47	4650.47	1957950000
5.	4689.54	4741.38	4688.02	4736.19	4736.19	2105450000
6.	4744.47	4744.71	4681.24	4704.07	4704.07	1715830000

Table 2 gives the Minimum, 1st Quartile, Median, Mean, 3rd Quartile, Maximum for the attributes of NASDAQ considered in the experiment.

Table 2: The Minimum, 1st Quartile, Median, Mean, 3rd Quartile, Maximum for the attributes of NASDAQ

	Open	**High**	**Low**	**Close value**	**Adjacent Close**	**Volume**
Minimum	4219	4293	4210	4267	4267	1.494e+08
1st Quartile	5082	5106	5061	5089	5089	1.797e+09
Median	6460	6473	6428	6456	6456	1.998e+09
Mean	6537	6575	6496	6539	6539	2.162e+09
3rd Quartile	7740	7804	7699	7752	7752	2.266e+09
Maximum	10131	10222	10112	10131	10131	7.279e+09

Support vector machine with type eps-regression: 70 percent of the data were taken for training. Details of the Parameters are given below in Table 3.

Table 3: Support vector machine with type eps-regression

S.No.	SVM-Kernel	Cost	Gamma	Epsilon	Number of Support Vectors	Total Mean Squared Errors	Squared Correlation Coefficient	Cross validation fold	Correlation Coefficient
1.	Linear	1	0.16667	0.1	5	4251.914	0.9996484	967	0.9997406
2.	Polynomial of degree 3	1	0.16667	0.1	826	483504.3	0.7844045	967	0.802167
3.	Radial	1	0.16667	0.1	31	6774.571	0.9967882	967	0.9957115
4.	Sigmoid	1	0.16667	0.1	966	1479183588	0.1207931	967	0.1643545

Support vector machine with type nu-regression: 70 percent of the data were taken for training. Details of the Parameters are given below in Table 4.

Table 4: Support vector machine with type nu-regression

S.No.	SVM-Kernel	Cost	Gamma	nu	Number of Support Vectors	Total Mean Squared Errors	Squared Correlation Coefficient	Cross validation fold	Correlation Coefficient
1.	Linear	1	0.16667	0.5	65	0.01733343	1	967	1
2.	Polynomial of degree 3	1	0.16667	0.5	490	443125.9	0.791784	967	0.8006916
3.	Radial	1	0.16667	0.5	547	1298.88	0.9993769	967	0.9988068
4.	Sigmoid	1	0.16667	0.5	485	894357688	0.0996714	967	0.1450302

Support vector machine with type eps-bsvr regression: 70 percent of the data were taken for training. Details of the Parameters are given below in Table 5.

Table 5: Support vector machine with type eps-bsvr regression

S.No	SVM-Kernel	Hyper parameters	Objective function value	Training error	Cross validation error	Number of Support Vectors	Correlation Coefficient
1.	Linear	-	-0.1621	0.00092	1933.197	4	0.9997284
2.	Polynomial	degree= 1, scale=1, offset=1	-0.1622	0.001333	2469.6	4	0.9997237
3.	Gaussian Radial Basis	sigma=1.2164	-19.9282	0.003436	25772.64	80	0.9850214
4.	Hyperbolic tangent	scale=1, offset=1	-25359.01	6517.318011	10760921456	964	0.00617843
5.	Laplace	sigma=1.2570	-16.5516	0.004399	25451.49	129	0.9880375
6.	Bessel	sigma=1, order=1, degree=1	-9.4018	0.002034	6935.601	39	0.9968213
7.	ANOVA RBF	sigma=1, degree=1	-0.8776	0.001956	4435.781	18	0.9979209

After data transformation, the first 6 rows generated from the data population using R is given in Table 6.

Table 6: The first 6 rows generated from the data population after transformation using R

S. No.	Open	High	Low	Close value	Adjacent Close	Volume
1.	-59.90039	-74.239746	-56.649902	-74.24023	-74.24023	359320000
2.	-33.48975	-35.439942	-73.870117	-59.82959	-59.82959	372850000
3.	-40.01025	-14.609863	46.310058	57.72998	57.72998	-209370000
4.	62.70020	88.659668	74.120118	85.71973	85.71973	147500000
5.	54.93018	3.330078	-6.779786	-32.12012	-32.12012	-389620000
6.	-30.40039	-28.899902	-30.590332	-39.35986	-39.35986	146130000

Table 7 gives the Minimum, 1st Quartile, Median, Mean, 3rd Quartile, Maximum for the attributes after transformation of NASDAQ considered in the experiment.

Table 7: The Minimum, 1st Quartile, Median, Mean, 3rd Quartile, Maximum for the attributes after transformation of NASDAQ

	Open	High	Low	Close value	Adjacent Close	Volume
Minimum	-737.670	-469.030	-656.280	-970.290	-970.290	-3.641e+09
1st Quartile	-27.592	-20.948	-26.155	-24.383	-24.383	-1.409e+08
Median	8.065	6.925	9.055	6.205	6.205	-1.385e+06
Mean	3.793	3.785	3.660	3.645	3.645	4.235e+06
3rd Quartile	41.160	33.420	37.833	42.920	42.920	1.522e+08
Maximum	522.880	433.430	538.440	673.080	673.080	4.910e+09

Support vector machine with type eps-regression after transformation: 70 percent of the data were taken for training. Details of the Parameters are given below in Table 8.

Table 8: Support vector machine with type eps-regression after transformation

S. No.	SVM-Kernel	Cost	Gamma	Epsilon	Number of Support Vectors	Total Mean Squared Errors	Squared Correlation Coefficient	Cross validation fold	Correlation Coefficient
1.	Linear	1	0.16667	0.1	7	20.16064	0.9995733	966	0.9995156
2.	Polynomial of degree 3	1	0.16667	0.1	635	20943.69	0.2517048	966	0.4835053
3.	Radial	1	0.16667	0.1	128	3282.907	0.6181392	966	0.6317106
4.	Sigmoid	1	0.16667	0.1	959	9809026	0.4325787	966	0.3537289

Support vector machine with type nu-regression after transformation: 70 percent of the data were taken for training. Details of the Parameters are given below in Table 9.

Table 9: Support vector machine with type nu-regression after transformation

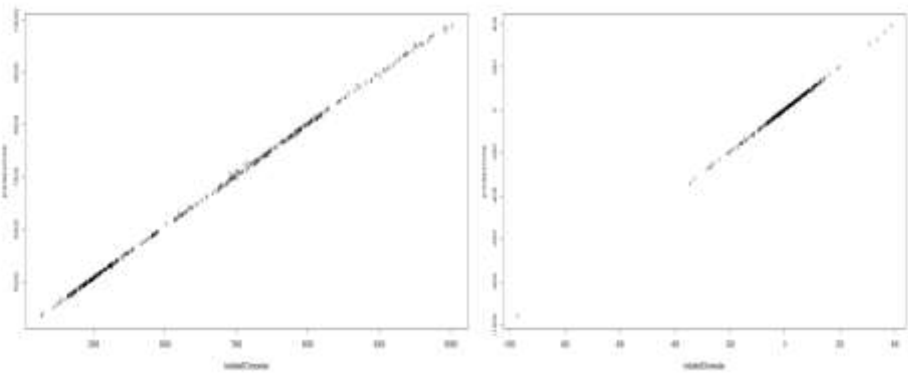
S.No.	SVM-Kernel	Cost	Gamma	nu	Number of Support Vectors	Total Mean Squared Errors	Squared Correlation Coefficient	Cross validation fold	Correlation Coefficient
1.	Linear	1	0.16667	0.5	20	4.326957e-05	1	966	1
2.	Polynomial of degree 3	1	0.16667	0.5	509	17190.02	0.292198	966	0.5283113
3.	Radial	1	0.16667	0.5	618	3203.349	0.6300255	966	0.6437896
4.	Sigmoid	1	0.16667	0.5	486	5208324	0.3423065	966	0.2702722

Support vector machine with type eps-bsvr regression after transformation: 70 percent of the data were taken for training. Details of the Parameters are given below in Table 10.

Table 10: Support vector machine with type eps-bsvr regression after transformation

S.No.	SVM-Kernel	Hyper parameters	Objective function value	Training error	Cross validation error	Number of Support Vectors	Correlation Coefficient
1.	Linear	-	-0.1621	0.00092	1933.197	4	1.038138e-05
2.	Polynomial	degree= 1, scale=1, offset=1	-0.1622	0.001333	2469.6	4	1.038138e-05
3.	Gaussian Radial Basis	sigma=1.2164	-19.9282	0.003436	25772.64	80	2.896446e-05
4.	Hyperbolic tangent	scale=1, offset=1	-25359.01	6517.318011	10760921456	964	0.0002884587
5.	Laplace	sigma=1.2570	-16.5516	0.004399	25451.49	129	2.520923e-05
6.	Bessel	sigma=1, order=1, degree=1	-9.4018	0.002034	6935.601	39	2.29781e-05
7.	ANOVA RBF	sigma=1, degree=1	-0.8776	0.001956	4435.781	18	7.475226e-06

Figure 1 gives the Support vector machine with type eps-regression and kernel Linear, Polynomial, Radial, Sigmoidal before and after data transformation



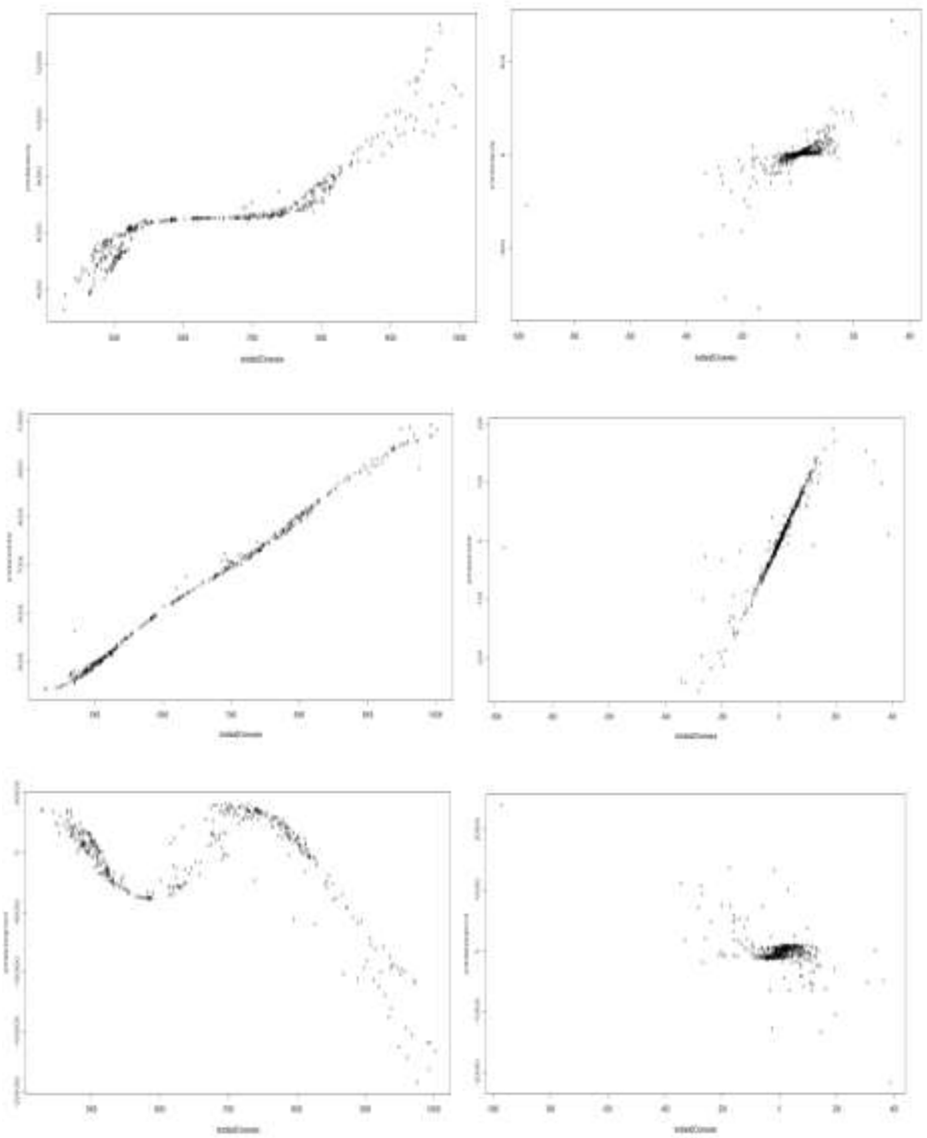


Figure 1: Support vector machine with type eps-regression and kernel Linear, Polynomial, Radial, Sigmoidal before and after data transformation

Figure 2 gives the Support vector machine with type nu-regression and kernel Linear, Polynomial, Radial, Sigmoidal before and after data transformation

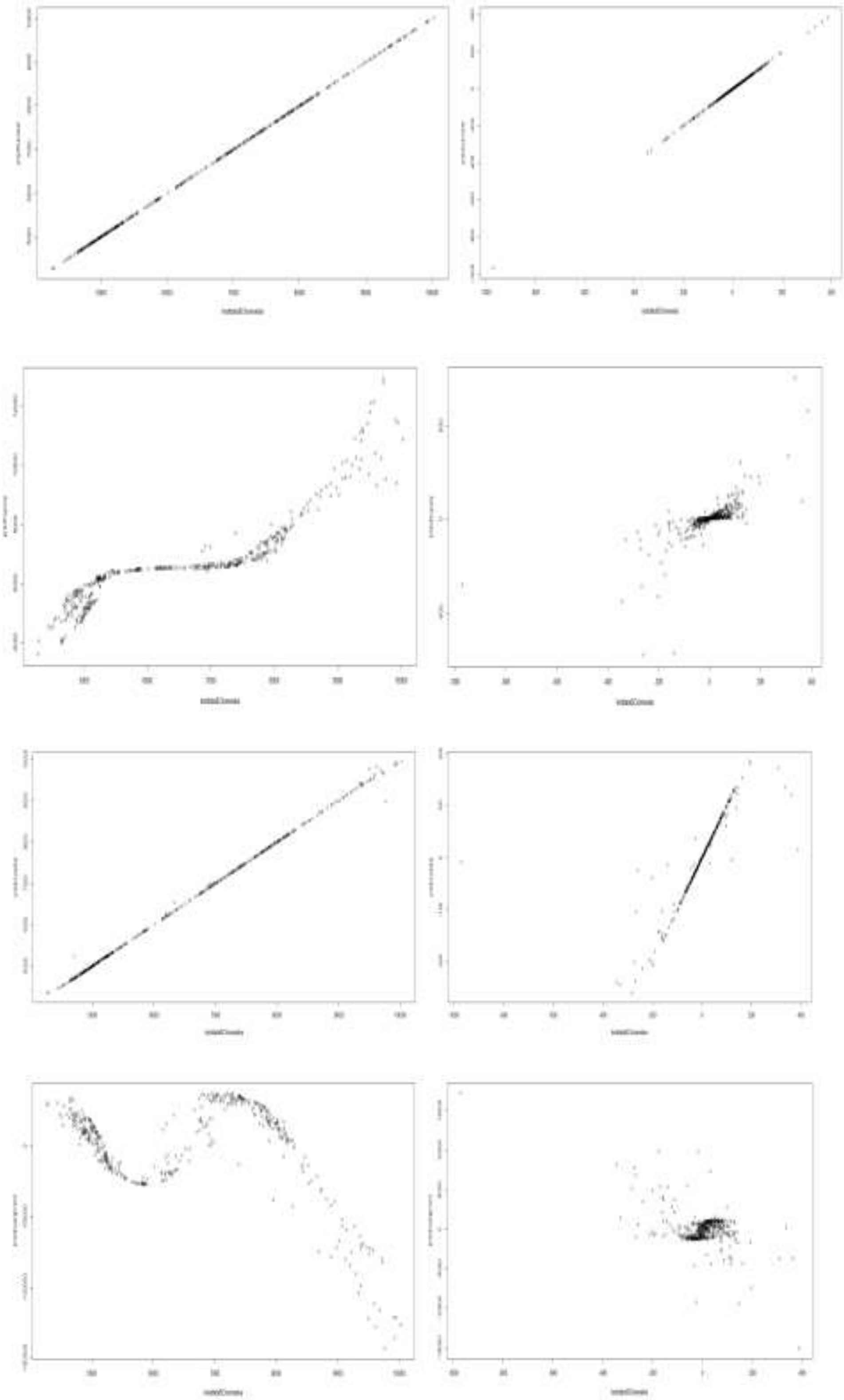
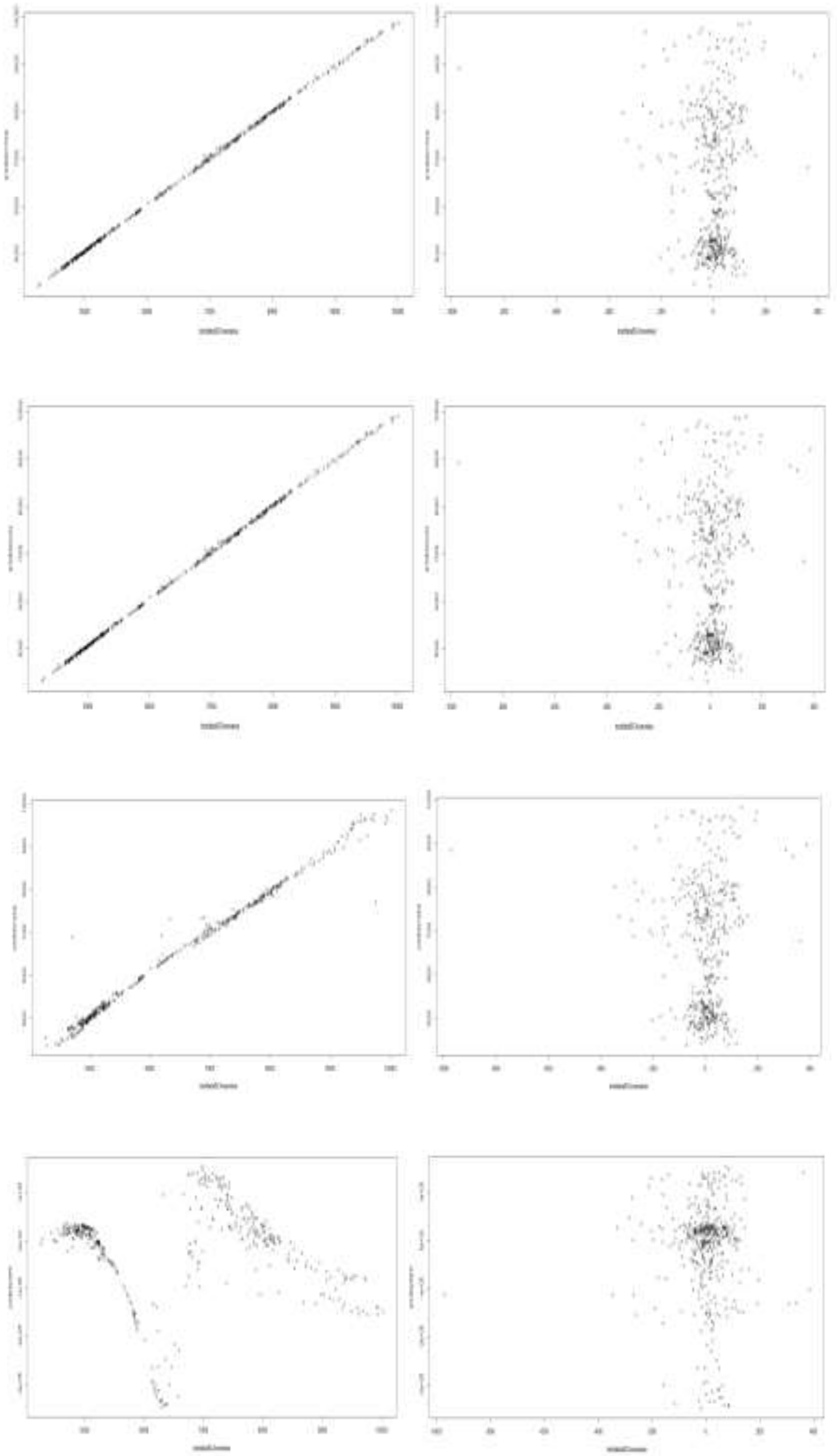


Figure 2: Support vector machine with type nu-regression and kernel Linear, Polynomial, Radial, Sigmoidal before and after data transformation

Figure 3 gives the Support vector machine with type eps-bsvr Regression and kernel Linear, Polynomial, Gaussian Radial Basis, Hyperbolic Tangent, Laplace, Bessel, ANOVA RBF before and after data transformation



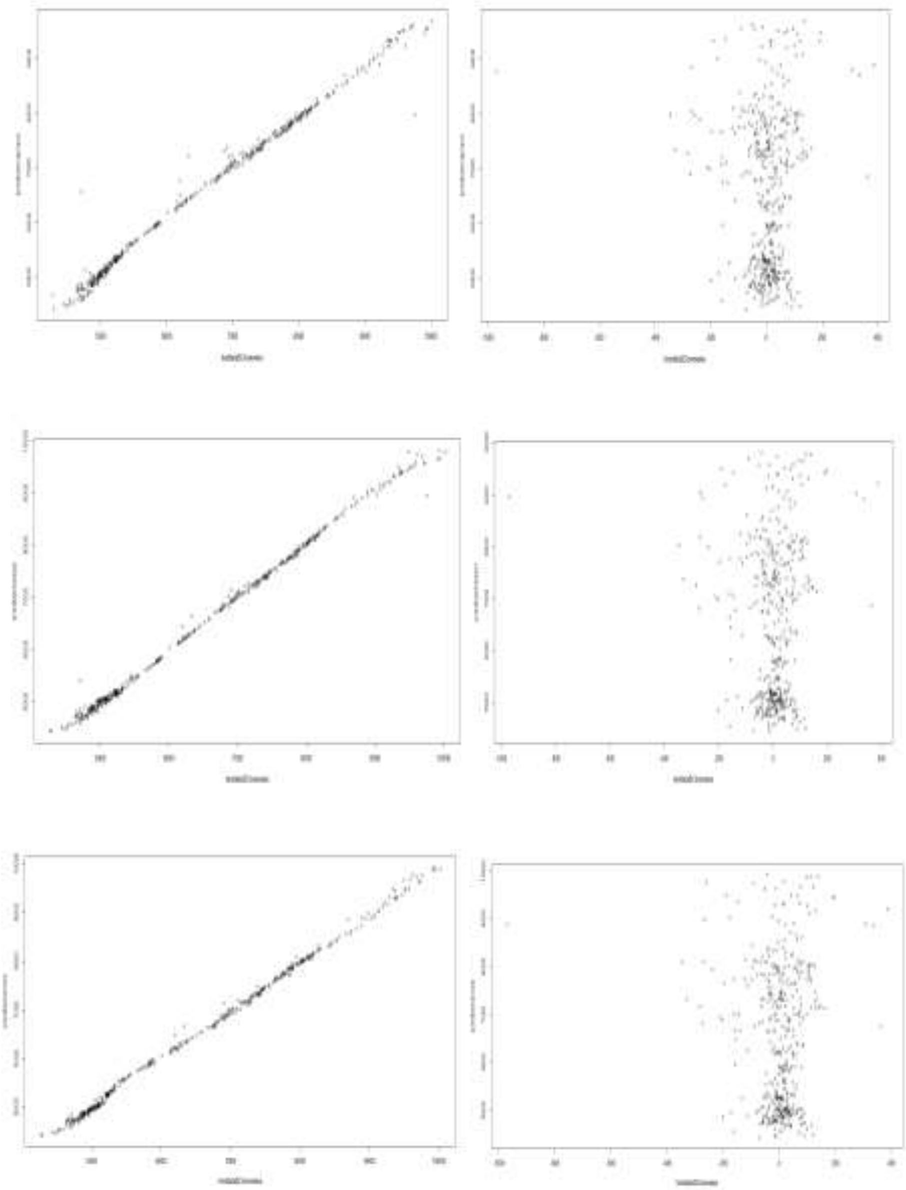


Figure 3: Support vector machine with type EPS-BSVR Regression and kernel Linear, Polynomial, Gaussian Radial Basis, Hyperbolic Tangent, Laplace, Bessel, ANOVA RBF before and after data transformation

4. Conclusion

For the experimental data setup, it is found that the correlation coefficient gives the best when the data were taken and transformed using the kernel, proceeded by training and testing. When the data were processed by taking the one-day difference in close value the correlation results obtained had the best goodness of fit. The methodology can be fine-tuned by optimizing the parameters using other optimization techniques in future for better accuracy in prediction and analysis.

References

[1] Kim, Kyoung-Jae.: Financial time series forecasting using support vector machines. In: Neuro computing. 55, 307 – 319 (2003)

[2] H. Ince, T.B., Trafalis.: Kernel principal component analysis and support vector machines for stock price prediction. In: 0-7803-8359-1/04/ 2004 IEEE. 2053-2058 (2004)

[3] Lucas, K. C., Lai, James, N. K., Liu.: Stock forecasting using support vector machine. In: Proceedings of the Ninth International Conference on Machine Learning and Cybernetics. 1607-1614 (2010)

[4] Smola, J., Alex, Scholkop. Bernhard: A tutorial on support vector regression. In: Statistics and Computing 8. 14. 199–222 (2004)

[5] Ahmad Kazema, Ebrahim Sharifia, Farookh Khadeer Hussainb, Morteza Saberic,Omar Khadeer Hussaind.: Support vector regression with chaos-based firefly algorithm for stock market price forecasting. In: Applied Soft Computing. 13, 947–958 (2013)

[6] Bruno Miranda Henrique, Vinicius Amorim Sobreiro, Herbert Kimura.: Stock price prediction using support vector regression on daily and up to the minute prices. The Journal of Finance and Data Science. 4(3), 183- 12. 201 (2018).