

International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

DETECTION OF OUTLIERS IN STOCK MARKET USING REGRESSION ANALYSIS

Dr. Pankaj Nagar¹, Gurjeet Singh Issar²

¹Asstt. Professor, Department of Statistics, University of Rajasthan, Jaipur, India. ²Asstt. Professor, Department of Computer Science, CIITM, Jaipur & Research Scholar-Computer Science, Jagannath University, Rajasthan

Abstract: By comparing historical data of trading like daily Open, High, Low, Close, Volume, Number of Trades, Turnover, Delivery percentage etc. of a particular stock with its Peer Group companies and Non Peer Group companies stocks for a particular period, we can find some unusual observations which are also known as outliers. In this paper we have tried to detect the observations, which are very different from the other observations using a Data Mining Technique for Outlier Detection-"Multiple Linear Regression Analysis".

Key Words: Outliers, Stock Market, Peer Group Analysis, Regression, ANOVA

I. Introduction

To find outliers is a very important task in Data Mining. In Stock Market, outlier detection plays very important role in detecting fraud. The main objective of Data Mining is to search for a general pattern for the input data. According to Barnett & Lewis, 1994, Novelty detection, or so-called outlier detection, is the identification of "novel" or "unknown" events that an expert system is not aware of during training or testing. Outliers may indicate abnormal running conditions and lead to significant performance degradation. An outlier is one that appears to obviously deviate from the others of the sample in which it occurs or an observation which appears to be inconsistent with the remainder of the dataset. According to Aggarwal and Yu (2001), outliers may be considered as noisy points lying outside a set of defined clusters or may be defined as points that lie outside of the set of clusters but are also different from the noise. According to He, Xu, Huang, Deng, 2004, Coderre, 2009, Aggarwal & Guojun, 2003, outlier analysis attempts to find the rare class whose behavior is very exceptional when compared to the rest of input data. Many techniques are proposed to detect outliers, drawn from Statistics, Computer Science or Machine Learning. Hodge and Austin (2004) reviewed some fundamental approaches to solve the problem of Outlier Detection. This technique is usually named as Novelty Detection since it aims to define the boundary of normality instead of estimating the density of the dataset. In addition to the surveillance of stock price changes, Anomaly Detection Techniques have been applied to various fields such as Network Intrusion Detection (Naiman, 2004; Scott, 2004), Financial Fraud Detection (Juszczak, Adams, Hand, Whitrow, &Weston, 2008), Fault Detection (Chen, Martin, & Montague, 2009; Martins, Pires, & Amaral, 2011; Yiakopoulos, Gryllias, & Antoniadis, 2011).

There are some techniques which can be applied for the Security Fraud Detection System based on Outlier Analysis, for example Fuzzy Set Analysis, Bayesian Approach, Pattern Matching Techniques and Data Mining Techniques. The following Six algorithms are the most commonly used software techniques in Data Mining Applications-Neural Networks, Decision Trees (Expert Systems), Genetic Algorithms, Regression Analysis, Statistical Methods and Data Visualization. In the current study we have used **Multiple Linear Regression Analysis** to detect outliers. To accomplish this study we have used trial edition of Minitab-15 for evaluation purpose.

Linear Regression assumes that a linear relationship exists between the input data and the output data. The common formula for a linear relationship is used in this model:

$$y = c_0 + c_1 x_1 + \ldots + c_n x_n$$

In the above equation there are n input variables, which are predictors or regressors and one output variable (y, the variable which is to be predicted). The constants c_0,c_1,\ldots,c_n , are the regression coefficients which is computed by the method, known as principle of Least Square, and computed during the modeling process when processed on a statistical software like Minitab-15 or SPSS etc. This is called **Multiple Linear Regression** because there is more than one predictor (**Margaret H. Dunham, 2005**). Regression Analysis is a statistical methodology that is most often used for numeric prediction (**Jiawei Han and Micheline Kamber**,

2010). Following are the results of Outlier observations when applied on Minitab-15 to find the linear regression equation for Closing Price (dependent variable) with Opening Price and SHL (Spread between High Price and Low Price):

II. Observations

Table-1: Peer Group Companies (Type- Other Telecom Services) Regression Analysis: Close Price versus Open Price, SHL

	ession Analysis. C		egression Analys	,				
	Predictor	Coef	SE Coef	T Value	P Va	lue		
	Constant	2.4427	0.3662	6.67	0.0	00		
Results for: GTL Ltd	Open Price	1.00190	0.00183	547.76	0.0	00		
The regression equation is	SHL	-0.87134	0.01911	-45.60	0.0	00		
Close $Price = 2.44 + 1.00$ Open Price -		S = 4.24900	R-Sq = 99.9%	R-Sq(adj) = 99.9	%			
0.871 SHL			ANOVA Tab					
	Source	DF	SS	MS	F Ratio	P Value		
	Regression	2	5443905	2721952	150767.18	0.000		
	Residual Error	246	4441	18				
	Total	248	5448346					
		R	egression Analys	is Table				
	Predictor	Coef	SE Coef	T Value				
Describe from Open abile Clabel I tol	Constant	-0.0061	0.4797	-0.01				
Results for: Onmobile Global Ltd The regression equation is	Open Price	0.979136	0.005590	175.17	F Ratio 150767.18 0.990 0.990 0.000 0.000 0.000 0.000 0.000 5% F Ratio 25276.71 P Value 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.512 P Value 0.72% F Ratio P Value 0.727 0.000 3% F Ratio F Ratio P Value 0.000 3% F Ratio P Value 0.584			
Close Price = $-0.006 + 0.979$ Open Price	SHL	0.26305	0.06728	3.91		00		
+ 0.263 SHL		S = 3.87681		R-Sq(adj) = 99.5	%			
		r	ANOVA Tab		n	1		
	Source	DF	SS	MS		P Value		
	Regression	2	759802	379901	25276.71	0.000		
	Residual Error	246	3697	15				
	Total	248	763499					
_			egression Analys			_		
_	Predictor	Coef	SE Coef	T Value				
Results for: Quadrant Televentures	Constant	0.39794	0.07372	5.40				
Limited	Open Price	0.89174	0.01920	46.44				
The regression equation is Close Price = $0.398 + 0.892$ Open Price	SHL	0.05182	0.07892	0.66		12		
Close $Price = 0.398 + 0.892$ Open Price	S = 0.182469 R-Sq = 90.3% R-Sq(adj) = 90.2% ANOVA Table							
+ 0.0518 SHL	C	DE	ANOVA 1at	MS	E Datia	P Value		
-	Source	DF	76.380	38.190				
-	Regression	2		0.033	1147.02	0.000		
	Residual Error Total	246 248	8.191 84.571	0.055				
	Totai	-	egression Analys	ia Tabla				
_	Predictor	Coef	SE Coef	T Value	D V.	luo		
	Constant	0.421	1.207	0.35				
	Open Price	1.00197	0.00844	118.68				
Results for: Tulip Telecom Ltd	SHL	-0.26682	0.05924	-4.50				
The regression equation is	SIL		R-Sq = 98.3%			00		
Close Price = $0.42 + 1.00$ Open Price -		5 = 2.70114	ANOVA Tat		70			
0.267 SHL	Source	DF	SS	MS	F Ratio	P Value		
	Regression	2	123577	61789		0.000		
	Residual Error	246	2157	9				
	Total	248	125734					
		R	egression Analys	is Table				
	Predictor	Coef	SE Coef	T Value	P Va	lue		
	Constant	0.01149	0.02093	0.55				
Results for: Nutek India Ltd.	Open Price	0.994395	0.006504	152.88				
The regression equation is	SHL	-0.17608	0.04808	-3.66	0.0	00		
Close $Price = 0.0115 + 0.994$ Open Price		S = 0.212468		R-Sq(adj) = 99.4	1%			
- 0.176 SHL			ANOVA Tak					
Ē	Source	DF	SS	MS	F Ratio	P Value		
Ē	Regression	2	2018.9	1009.4	22361.31	0.000		
Γ	Residual Error	246	11.1	0.0				

Coef:-CoefficientSE Coef:-Standard Error CoefficientDF:-Degrees of Freedom

SS:-MS:-S:-

Sum of Square Mean Sum of Square Standard Error R-Sq:-R SquareR-Sq(adj):-Adjusted R SquareSHL:-Spread(High-Low)

Regression Analysis Table P Value Predictor Coef SE Coef T Value Constant 121.31 33.60 3.61 0.000 Open Price 0.96512 0.01184 81.49 0.000 Results for: Infosys SHL -0.45743 0.08139 0.000 -5.62 The regression equation is Close Price = 121 + 0.965 Open Price -S = 38.9343 R-Sq = 96.6% R-Sq(adj) = 96.5% 0.457 SHL ANOVA Table Source DF SS MS F Ratio P Value 10512382 5256191 3467.43 0.000 Regression 2 Residual Error 246 372906 1516 248 10885288 Total **Regression Analysis Table** Predictor Coef SE Coef T Value P Value Constant 8.109 3.398 2.39 0.018 Open Price 0.95333 0.01932 49.33 0.000 Results for: NTPC 0.09764 0.448 -0.07414-0.76 The regression equation is SHL R-Sq(adj) = 90.8% Close Price = 8.11 + 0.953 Open Price -S = 2.63499 R-Sq = 90.8% 0.0741 SHL ANOVA Table DF MS F Ratio P Value Source SS 16929.2 8464.6 1219.13 0.000 2 Regression Residual Error 246 1708.0 6.9 Total 248 18637.2 **Regression Analysis Table** Predictor Coef SE Coef T Value P Value Constant 49.78 17.50 2.84 0.005 Open Price 0.989625 0.008021 123.37 0.000 Results for: SBI 0.08274 0.000 The regression equation is SHL -0.53754-6.50Close Price = 49.8 + 0.990 Open Price -S = 40.9299 R-Sq = 98.4% R-Sq(adj) = 98.4%0.538 SHL ANOVA Table Source DF SS MS F Ratio P Value 25634555 12817278 7650.94 0.000 Regression 2 Residual Error 246 412113 1675 248 Total 26046668 **Regression Analysis Table** Predictor Coef SE Coef T Value P Value Constant -0.163 2.547 -0.06 0.949 Open Price 0.991416 0.007036 140.91 0.000 **Results for: Jet Airways** The regression equation is SHL 0.10949 0.07569 1.45 0.149 Close Price = -0.16 + 0.991 Open Price S = 11.3830 R-Sq = 98.8% R-Sq(adj) = 98.8%+ 0.109 SHL ANOVA Table DF MS F Ratio P Value Source SS 0.000 2711338 1355669 10462.52 2 Regression Residual Error 246 31875 130 248 2743213 Total **Regression Analysis Table** Coef P Value Predictor SE Coef T Value Constant 4.128 2.018 2.04 0.042 Open Price 0.96522 0.01382 69.86 0.000 **Results for: Ambuja Cement** 0.22886 SHL 0.09864 2.32 0.021 The regression equation is S = 2.96408 R-Sq = 95.4% R-Sq(adj) = 95.4% Close Price = 4.13 + 0.965 Open Price + 0.229 SHL ANOVA Table DF MS F Ratio P Value Source SS 45104 22552 0.000 Regression 2 2566.86 Residual Error 246 2161 9 Total 248 47265

Table-2: Non Peer Group (Group-A Companies) Regression Analysis: Close Price versus Open Price, SHL

Coef:-Coefficient SS:-Sum of Square R-Sq:-R Square SE Coef:- Standard Error Coefficient MS:-R-Sq(adj):-Adjusted R Square Mean Sum of Square DF:-SHL:-Spread(High-Low) Degrees of Freedom S:-Standard Error

			Regression Analy		n		
	Predictor	Coef	SE Coef	T Value			
	Constant	0.6090	0.2410	2.53			
Results for: Alok Industries Ltd.	Open Price	0.97956	0.01115	87.88			
The regression equation is	SHL	-0.30669	0.07930	-3.87	0.0	00	
Close Price = $0.609 + 0.980$ Open Price -	S = 0.564197 R-Sq =	= 97.0% R-Sq(adj)			0.012 0.000 0.000 F Ratio P V 3940.39 0.0 P Value 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.023 0.000 0.000 0.000 F Ratio P V 12150.10 0.0 F Ratio P V 12150.10 0.0 F Ratio P V 0.927 0.000 0.000 0.000 0.000 0.000 0.000 0.000 F Ratio P V I1222.74 0.0 P Value 0.989 0.000 0.000 0.000 0.000 P Value 0.989 P Value 0.989 P Value 0.989 P Value P Value P Value P Value		
).307 SHL		-	ANOVA Ta			1	
	Source	DF	SS	MS		P Valu	
	Regression	2	2508.6	1254.3	3940.39	0.000	
	Residual Error	246	78.3	0.3			
	Total	248	2586.9				
		1	Regression Analy		n		
	Predictor	Coef	SE Coef	T Value			
	Constant	6.959	1.908	3.65			
Results for: Praj Industries Ltd	Open Price	0.90594	0.02507	36.14			
The regression equation is	SHL	0.03527	0.07544	0.47	0.6	41	
Close Price = $6.96 + 0.906$ Open Price +	S = 2.06277 R-Sq =	84.7% R-Sq(adj)					
0.0353 SHL		-	ANOVA Ta		r		
	Source	DF	SS	MS		P Valu	
	Regression	2	5810.9	2905.5	682.83	0.000	
	Residual Error	246	1046.7	4.3		L	
	Total	248	6857.6				
			Regression Analy				
	Predictor	Coef	SE Coef	T Value			
	Constant	0.8144	0.3554	2.29		-	
Results for: Shree Renuka Sugar	Open Price	1.01367	0.00665	152.35			
The regression equation is	SHL	-0.78296	0.07055	-11.10	0.01: 0.000 0.000 0.000 0.000 3940.39 P Val 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.64 F Ratio 682.83 0.000 0.000 0.000 0.000 0.000 F Ratio 12150.10 P Val 0.92' 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000	00	
Close Price = $0.814 + 1.01$ Open Price -	S = 1.61605 R-Sq =	99.0% R-Sq(adj)					
0.783 SHL		-	ANOVA Ta		r		
	Source	DF	SS	MS		P Valu	
	Regression	2	63462	31731	12150.10	0.000	
	Residual Error	246	642	3			
	Total	248	64105				
			Regression Analy		r		
	Predictor	Coef	SE Coef	T Value			
	Constant	-0.00392	0.04304	-0.09			
Results for: Southern Ispat and	Open Price	0.977647	0.007257	134.73			
Energy Ltd	SHL	0.26051	0.07629	3.41	0.0	01	
The regression equation is	S = 0.239944 R-Sq =	= 98.9% R-Sq(adj)					
Close Price = $-0.0039 + 0.978$ Open		-	ANOVA Ta		r		
Price + 0.261 SHL	Source	DF	SS	MS		P Valu	
	Regression	2	1292.26	646.13	11222.74	0.000	
	Residual Error	246	14.16	0.06			
	Total	248	1306.42				
			Regression Analy		r		
	Predictor	Coef	SE Coef	T Value			
	Constant	-0.023	1.673	-0.01			
Doculta for uflor	Open Price	0.98504	0.01047	94.10			
Results for: uflex The regression equation is	SHL	0.21318	0.07582	2.81	0.0	05	
The regression equation is Close Price = $-0.02 + 0.985$ Open Price	S = 5.28412 R-Sq =	97.5% R-Sq(adj)					
		-	ANOVA Ta				
⊥ 0.213 SHI	Source	DF	SS	MS		P Valu	
+ 0.213 SHL		-	267897	133949	4797.26	0.000	
+ 0.213 SHL	Regression	2	201071			0.000	
+ 0.213 SHL	Regression Residual Error	246	6869	28		0.000	

Table-3: Non Peer Group (Group-B Companies) **Regression Analysis: Close Price versus Open Price, SHL**

Standard Error Table-4: Outliers in Peer Group (Type- Other Telecom Services)

SHL:-

Spread(High-Low)

Company Name	No. of Outliers	Outlier Observation Numbers
GTL Ltd	13	7, 37, 52-55, 57, 63, 83, 110, 116, 200, 220
Onmobile Global Ltd	22	1-20, 23, 88
Quadrant Televentures Ltd	18	3, 4, 6, 11, 45, 46, 73, 131, 133, 145, 146, 196, 204, 206, 218, 220, 221, 235
Tulip Telecom Ltd	21	3, 5, 22, 40, 54, 57, 62, 82, 101, 112, 114, 116, 139, 141, 176, 204, 212, 214, 220, 222, 223
Nutek India Ltd	30	1-11, 13, 18, 20, 24, 27, 34, 38, 45-47, 50, 61, 62, 66-68, 88, 93, 133
		Maan Number of Outliers 20.8

Mean Number of Outliers= 20.8

DF:-

Degrees of Freedom

Company Name	No. of Outliers	Outlier Observation Numbers
Infosys	13	2, 8, 9, 58, 93, 96, 113, 124, 132, 158, 164, 181, 195
NTPC	12	25, 41, 58, 67, 115, 131, 157, 166, 175, 177, 200, 239
SBI	14	19, 20, 30, 58, 90, 117, 132, 150, 166, 203, 217, 219, 222, 226
Jet- Airways Ltd	11	21, 23, 54, 58, 78, 94, 95, 161, 212, 213, 222
Ambuja Cement Ltd	13	16, 21, 89, 90, 108, 110, 115, 143, 160, 166, 171, 208, 222
	Mean Nu	mber of Outliers-126

Table-5:	Outliers in	Non Pe	er Group	(Group-A	companies)
I GOIC CI	o a mor b m	11011 1 0	or or oup	(Oroup II	companies)

Mean Number of Outliers= 12.6

Table-6: Outliers in Non Peer Group (Group-B companies)

Company Name	No. of Outliers	Outlier Observation Numbers
Alok Industries Ltd	12	3, 4, 8, 13, 44, 54, 94, 96, 151, 170, 210, 226
Praj Industries Ltd	12	78, 79, 86, 102, 119, 121, 153, 165, 205, 206, 212, 222
Shree Renuka Sugar	16	1, 27, 61, 64, 72, 75, 78, 90, 94, 98, 118, 119, 125, 152, 195, 217
Southern Ispat and Energy Ltd	15	8, 21, 30, 37, 38, 45-49, 96, 98, 119-121
Uflex Ltd	20	9, 10, 15, 17, 18, 20, 21, 37, 44, 54, 57, 96, 99, 101, 121, 129, 139, 152, 153, 192

Mean Number of Outliers= 15

Table-7: ANOVA Table

	Sum of Squares	DF	Mean Square	F	Sig.
Between Companies of Peer	212244.754	4	53061.189		
Group (Type-Other Telecom Services)				15.911	0.000
Within Groups	330153.707	99	3334.886		
Total	542398.462	103			

Table-8: ANOVA Table

	Sum of Squares	DF	Mean Square	F	Sig.		
Between Companies of Non-Peer	14990.645	4	3747.661				
Groups (Group-A Companies)				0.775	0.546		
Within Groups	280393.069	58	4834.363				
Total	295383.714	62					

Table-9: ANOVA Table

	Sum of Squares	DF	Mean Square	F	Sig.	
Between Companies of Non-Peer Groups	56147.200	4	14036.800			
(Group-B Companies)				4.122	0.005	
Within Groups	238371.467	70	3405.307			
Total	294518.667	74				

III. Discussion

We can analyze that how a single dependent variable can be affected by the values of one or more independent variables. For example, in this study we can analyze how a stock's close price is affected by such factors as open price, high price and low price. In our model, we have taken Close Price as Response Variable and two predictors (i) Open Price (ii) Difference of High and Low Price. The resultant Outliers are data points that are more than some appropriate distance from a regression line that is estimated using all the other data points in the sample.

We have taken historical data from BSE website [17] for period of one year, from 1st April, 2011 to 30th March, 2012. Total 249 days trading data has been taken in this period for all the companies, excluding holidays and Non-Trading Days. In Peer Group, all the companies are from B category stocks. In Non Peer Group, we have randomly chosen 5 companies from A category stocks (For Group A) and 5 companies from B category stocks (For Group B) from different-different sectors. From table 4, we can see that in Peer Group, Nutek India Ltd. has the highest numbers of outliers. Earlier we have studied how Operators/Manipulators badly hammered share price of Nutek India Ltd., from IPO listing price Rs. 192 to below 1 rupee (Singh & Nagar, April 2012) and also found illegal intraday trading done by II/NII (Institutional Investors/Non Institutional Investors)[12] in the said company (Nagar & Singh, October 2012). If we compare Mean value of Outliers from Table 4, 5 and 6, we can see that Peer Group's mean value of outliers is 20.8 and it is higher than both Non Peer Groups, 12.6 from Non Peer Group A and 15 from Non Peer Group B.

As per the table 1, 2 and 3, it can be easily observed that in almost all cases the R-Square value is more than 90%, except the one case of Non Peer Group of category B company- Praj Industries Ltd. That shows the regression model corresponding to each and every company either of Peer Group, Non Peer Group of category A and Non Peer Group of category B, is a best fit model to predict the Closing Price on the basis of Opening Price and SHL.

Secondly the p-values with respect to all regression coefficients in each fitted regression equation, comes to be less than 0.05, which shows that the regression coefficients in all fitted regression equations are significant except the regression equations to be the best fit for the given situations.

The results in table 7, 8 & 9 are obtained by using SPSS (Statistical packages for Social Sciences) version 20. The ANOVA (Analysis Of Variance) Table-7 with respect to Table-4 shows that the number of outliers in Peer Group companies are not same but varying with respect to time. It means there could be multiple number of factors, that varies with respect to time, to generate the outliers of all these 5 Peer Group companies, at various time points (p-value <0.05). On the other hand the outliers in Non Peer Group companies are not more than 15 on an average. The ANOVA Table-8 with respect to Table-5 shows that the outlier observation numbers are not varying much similar between the companies as compared to time of occurrence of outlier(p-value >0.05). The ANOVA Table-9 constructed with respect to Table-6 shows that the observations occur at significantly varying time (p-value <0.05) but the variation, within the outlier observation numbers of each company, is insignificant.

IV. Conclusion

The main objective of the current study is to find outliers in historical data of stock market using Multiple Regression Analysis. We have analyzed and studied 15 stocks by collecting historical data of a certain period from BSE website. We have detected outliers stock wise and then compared these outliers with its Peer Group to find out whether the results are almost similar or different, and further compared the Peer Group with Non Peer Groups of Category A and Category B stocks and found that category A stocks has lesser outliers in comparison of category B stocks. As all the stocks from Peer Group are from category B stocks, so we have further compared Peer Group with Non Peer Group of category B stocks, and found still Peer Group stocks have more outliers.

V. References

- [1] Aggarwal, C. C., & Yu, P. S. (2001). Outlier detection for high dimensional data. In Proceedings of the ACM SIGMOD conference.
- [2] Barnett, V., & Lewis, T. (1994). Outliers in statistical data. New York: John Wiley & Sons.
- [3] Chen, T., Martin, E., & Montague, G. (2009). Robust probabilistic PCA with missing data and contribution analysis for outlier detection, Computational Statistics & Data Analysis, 53(10), 3706–3716.
- [4] David Coderre, Computer Aided Fraud Prevention and Detection, A Step by Step Guide, John Wiley and Sons, 2009.
- [5] He, Z., Xu, X., Huang, J. Z. & Deng, S.(2004). Mining class outliers: Concepts, algorithms and applications in CRM. Expert Systems with Applications, 27, 681–697.
- [6] Hodge, V. J., & Austin, J. (2004). Survey of outlier detection methodologies. Artificial Intelligence Review, 22, 85-126.
- [7] Juszczak, P., Adams, N. M., Hand, D. J., Whitrow, C., & Weston, D. J. (2008). Off-the-peg and bespoke classifiers for fraud detection, Computational Statistics & Data Analysis, 52(9), 4521–4532.
- [8] Martins, J. F., Pires, V. F., & Amaral, T. (2011). Induction motor fault detection and diagnosis using a current state space pattern recognition, Pattern Recognition Letters, 32(2), 321–328.
- [9] Naiman, D. Q. (2004). Statistical anomaly detection via httpd data analysis, Computational Statistics & Data Analysis, 45(1), 51–67.
- [10]Rajesh K. Aggarwal & Guojun Wu. Stock Market Manipulation Theory and Evidence, Working paper, Univ. of Michigan, March 11, (2003).
- [11]Yiakopoulos, C. T., Gryllias, K. C., & Antoniadis, I. A. (2011). Rolling element bearing fault detection in industrial environments based on a K-means clustering approach. Expert Systems with Applications, 38(3), 2888–2911
- [12]Secondary Market Advisory Committee of SEBI, "Discussion paper on Short Selling and Securities Lending and Borrowing", http://www.sebi.gov.in/commreport/rep40.pdf
- [13]Singh, Gurjeet and Nager, Pankaj(2012): A case Study on Nutek India Limited Regarding Deep Falling in Share Price, Researchers World--Journal of Arts, Science & Commerce, Vol.- III, Issue 2(3), April 2012.
- [14]Nager, Pankaj and Singh, Gurjeet(2012): An Analysis of Outliers For Fraud Detection in Indian Stock Market, Researchers World-Journal of Arts, Science & Commerce, Vol.- III, Issue 4(4), October 2012.
- [15] Margaret H. Dunham (2005), "Data Mining, Introductory and Advanced Topics", Pearson Education(Singapore).
- [16] Jiawei Han and Micheline Kamber, Second Edition (2010), "Data Mining, Concepts and Techniques", Morgan Kaufmann, An imprint of Elsevier (San Francisco).
- [17] Bombay Stock Exchange's Website www.bseindia.com