2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress

Application of Logistic Regression in assessing Stock Performances

Usha Ananthakumar
Shailesh J. Mehta School of Management
IIT Bombay
Mumbai, India

Ratul Sarkar
Shailesh J. Mehta School of Management
IIT Bombay
Mumbai, India

***Abstract*—Stock market prediction pertains to predicting the future value of stock of a company or other financial instrument traded on an exchange. Though the successful prediction of a stock's future price is a complex phenomenon, even a reasonable prediction would yield significant profit and this study attempts to address this unpredictability with the help of a data mining technique. In this study, the companies listed on BSE SENSEX are chosen as representative set of companies that are most actively traded. Logistic Regression is used on various important financial ratios of these companies and certain macro financial variables to analyze which ratios are important and how they are affecting the stock prices. The proposed model results in better classification accuracy when compared to a similar study in the literature.**

Keywords—financial ratios; classification accuracy; ROC curve; multicollinearity

I. INTRODUCTION

To make a good investing decision or while exploring the investing options in stock market, it is important for the shareholders and potential investors to analyze the relevant financial information available with them. But at the same time, foreseeing the future stock performance is very difficult and complicated and as such there is no tool available in the market to predict future stock performance. But to an extent, stock performance can be analyzed on the basis of financial indicators that we can get from the annual report of the company. Company's annual report consists of huge amount of data that can be converted into key financial ratios that can help investors to do the analysis. Different users such as management bankers, creditors, shareholders and researchers use these financial ratios to project future stock price trends. The study of financial ratios emerged as a new discipline after stock market crashes in the 1990s and early 2000s in the United States and parts of Europe and southern Asia. The level of importance given to financial ratios differs from one country to another. Thus, selecting appropriate ratios is very crucial in increasing the prediction success rate.

In stock performance literature, very little attention has been given in the past to the Indian stock market. But there is a shift towards generating awareness on the market due to its massive growth and its increasing potential for global investors. As a result of stock market's growing importance, more attention has been given to studies concerning different classification techniques for measuring stock performance. A number of research papers predict stock performance as well as pricing of the stock index across the globe. Harvey [1] observes that emerging market returns are usually more predictable than developed market returns because emerging market returns are more likely to be influenced by local information than developed markets.

Model for prediction of financial bankruptcy of companies by using logistic regression is given in [2] where it is shown that there are significant differences among the means of some of the financial ratios in two solvent and insolvent groups and hence can predict financial crisis of the company which can help companies to change its strategy to prevent crisis. de Oliveira, Nobre & Zarate [3] have used ANN to predict the behavior and trends of stocks and have also identified the variables that drive stock prices. A Bayesian regularized artificial neural network has been proposed by [4] to forecast the movement of stock prices and is shown to have improved prediction quality and generalization. Han, Ma & Yu [5] have modeled financial prediction by introducing factor analysis into logistic regression. In their study, they chose 32 financial ratios, nearly covering all financial aspects of a company, which helps the model reflect the corporate finance situations comprehensively, carried out the factor analysis on the financial ratios and then selected certain factor variables according to contribution rates to carry out logistic regression. Chen, Chen & Ye [6] studied about the stock price prediction problem by using techniques like Multinomial Logistic Regression, K-nearest neighbours algorithm, Linear Discriminant Analysis, Quadratic Discriminant Analysis, and Multiclass Support Vector Machine, and compared their performance based on the results of test accuracy, robustness, and run time efficiency. Similarly, various classifiers including several ensemble methods have been studied by [7] and ensemble methods are seen to be good alternatives for single classifiers in predicting stock price direction. Recently, the ability of ANN in forecasting the daily NASDAQ exchange rate has been investigated by [8].

The motivation for the current study is due to [9] where they have used the binary logistic regression model to determine the factors that significantly affect the performance of a company in the stock market. The outcome of this study

978-1-5386-1956-8/17 $31.00 © 2017 IEEE
DOI 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.199

helps the investor to form an opinion about the shares to be invested. They observed that eight financial ratios can classify companies up to a 74.6% level of accuracy into two categories ("good" or "poor"), based on their rate of return. In the study, they concluded that ratio methods have the capability to reveal maximum information content, if variables are chosen very carefully with regard to the purpose at hand. But in the above study, they had only considered ratios which were normally distributed. As this criterion is not mandatory in case of Logistic regression, we tried to consider more reasonable ratios that could be used to build the model. In the present study, we have tried to use the insights derived from the previous studies to come up with a more accurate and useful model for predicting stock prices using binary Logistic Regression and have used various ratios that affect the stock prices at a considerable higher rate. We have also considered qualitative measures like Revenue per employee, corporate actions like stock split and bonus announcement and macro financial factors as additional variables in our study.

Section II presents details about logistic regression model and certain measures of accuracy. Section III presents the data analysis along with the results obtained. Further we present a comparative analysis of our model against the model given in [9] in Section IV and conclusions are presented in Section V.

II. Logistic Regression

In statistical modeling, regression analysis is a statistical process for estimating the relationships among variables. It includes techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors'). Simple linear regression or multiple linear regression is applicable when the relationship between variables is assumed to be linear. However, when the response variable is binary, it is unrealistic to model using linear regression as it allows dependent variable to take values greater than 1 or less than 0. The logistic regression model is a type of generalized linear model that extends the linear regression model by linking the range of real numbers to the 0-1 range. Logistic regression has become an important tool in machine learning to classify incoming data on the basis of historical data. This has made this tool quite popular on various applications involving large data.

A. Logistic Regression Model

The difference between logistic and linear regression is reflected both in the choice of a parametric model and in the assumptions. Once this difference is accounted for, the methods employed in an analysis using logistic regression follow the same general principles used in linear regression. As stated by [10], the techniques used in linear regression analysis motivate the approach to logistic regression.

Let Y denote the response variable that can take value either 0 or 1.The expected response is just the probability that the response variable takes on value 1, say $\pi(x)$.

The logistic response function has the form

$$E(Y|x) = \pi(x) = \exp(\beta_0 + \beta_1 x) / \{1 + \exp(\beta_0 + \beta_1 x)\} \quad (1)$$

The logit transformation of $\pi(x)$ is given by

$$\eta = \beta_0 + \beta_1 x = \ln[\pi(x) / \{1 - \pi(x)\}] \quad (2)$$

This serves as a link function between probability and linear expression on the right hand side. The ratio $\pi(x)/\{1- \pi(x)\}$ is called the odds. Maximum likelihood method is used to find estimates of β_0 and β_1, given by $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively. After estimating the coefficients, relevant hypothesis testing is carried out to determine whether the model is adequate and whether the independent variables in the model are significantly related to the outcome variable. Odds ratio is given by

$$\text{Odds}_{xi+1} / \text{Odds}_{xi} = e^{\beta_1} \quad (3)$$

The odds ratio can be interpreted as the estimated increase in odds of achieving success associated with one unit change in the value of the predicted variable.

B. Classification Accuracy and ROC curve

Though we predict the probability of categorical outcome, it is most often used for classification. The probability of belonging to class 1 is compared with a threshold called cutoff value. If the probability is above cutoff, the case is classified as belonging to class 1 or otherwise to class 0. This results in the confusion matrix given by Table I.

Estimated misclassification rate is = (FP + FN)/GT

Overall accuracy is estimated by accuracy = 1 – Estimated Misclassification Rate

However, when two classes are asymmetric, where it is more important to predict the correct membership of a particular class, this accuracy metric is not an appropriate measure for evaluating the classifier. Suppose the important class is 1, some of the well-known accuracy measures are:

- The Sensitivity of the classifier is its ability to detect the important class member correctly. This is measured by True Positive Rate given by

$$TPR = \text{Sensitivity} = TP / P_c$$

- The Specificity of the classifier is its ability to classify class 0 members correctly. This is measured as True Negative Rate given by

$$TNR = \text{Specificity} = TN / N_c$$

It is useful to plot these measures against various thresholds in order to assess the performance of a classifier.

Table I. Confusion Matrix

		Predicted result		
		1	0	
Actual result	1	TP	FN	Pc
	0	FP	TN	Nc
				GT

A Receiver Operating Characteristic curve, or ROC curve, is a graphical plot formed by Sensitivity over (1 – Specificity) at various threshold settings. Any point in ROC space corresponds to the performance of a classifier for a specific threshold. The ROC curve is quite useful because it provides visual representation of relative tradeoff between the benefits reflected by True Positives and cost reflected by False Positives. Further, this plot is very effective while comparing the performance of different classifiers. In order to assess the performance of different classifiers, generally Area Under the Curve (AUC) is used as an evaluation criterion. Hence more the area under the curve, more is the significance or accuracy of the model.

III. Data Analysis

In our study, the companies with large market capitalizations have been considered, of which, most of these companies are part of the BSE SENSEX index. The financial data used in this analysis was collected from the Web link www.capitaline.com. The sample consists of 100 companies that are most actively traded on Indian Stock exchange and financial ratios and stock prices for calculating stock return were then calculated for a period of four years from 2012 to 2015. We had removed the data for eight banking companies due to difference in their balance sheet format. Thus the study consists of a sample size of 368 distinct observations corresponding to 92 different companies for a period of four years.

A. Data

To start with the model, first a method is required for classifying a company as a "good" or "poor" investment choice for a given year. Although there is no definitive method for defining a market investment as "good" or "poor," in this study we use a method that is simple and objective – namely, if the value of a company's stock over a given year rose above market return, it is classified as a "good" investment option; otherwise, it is classified as a "poor" investment option. Here, the BSE (Index of Bombay Stock Exchange) return has been taken as proxy for market return. To obtain the return at the end of each financial year, the March ending prices were used for each year.

The return was calculated using the following formula:

$$\textbf{Return of stock} = \frac{P_t - P_{t-1}}{P_{t-1}}$$

Where,
P_t = Price at year t
P_{t-1} = Price at year t-1

$$\textbf{Market Return} = \frac{BSE(t) - BSE(t-1)}{BSE(t-1)}$$

where BSE(t) = BSE at year t, and BSE(t-1) = BSE at year t-1.

For building the model, 26 financial ratios were taken for analysis of which two were categorical as given in Table II.

Due to the nature of data, there was a huge possibility for multicollinearity as there could be some inter dependencies among different independent variables which is briefly explained in the next section.

B. Multicollinearity

Multicollinearity is a statistical phenomenon in which predictor variables in a regression model are highly correlated. It is not uncommon when there are a large number of covariates in the model. Multicollinearity can cause unstable estimates and inaccurate variances which affects confidence intervals and hypothesis tests. Mathematically, multicollinearity can mainly be detected with the help of tolerance and its reciprocal, called *variance inflation factor* (VIF). By definition, tolerance of any specific explanatory variable is

$$\textbf{Tolerance} = \textbf{Tol} = 1 - R^2$$

where R^2 is the coefficient of determination for the regression of that explanatory variable on all remaining independent variables. The variance inflation factor is defined as the reciprocal of tolerance as

$$\textbf{VIF} = 1 / \textbf{Tol} = 1/(1 - R^2)$$

Table II. List of Independent Variables

Name of Variable	Description	Name of Variable	Description
Var1	Book Value per share	**Var14**	Revenue per Employee
Var2	Dividend / Profit After Tax	**Var15**	Operating Cash Flow ratio
Var3	Price to earnings ratio	**Var16**	Sales to Cash Flow
Var4	Price to sales ratio	**Var17** *	Stock Split
Var5	Current Ratio	**Var18** *	Bonus Announcement
Var6	Profit Margin	**Var19**	Net Sales increase (%)
Var7	Return on Assets	**Var20**	Cash earnings per share
Var8	Return on Capital Employed	**Var21**	Price / Cash Earning
Var9	Return on Equity	**Var22**	PBITDS
Var10	Debt to Asset ratio	**Var23**	Sales / Net Assets
Var11	Debt to equity	**Var24**	Price / Book Value
Var12	Interest Coverage ratio	**Var25**	Inflation
Var13	Fixed Asset Turnover ratio	**Var26**	GDP Change ratio

*Categorical Variables

The variance inflation factor VIF shows us how much the variance of the coefficient estimate is being inflated by multicollinearity. The square root of VIF tells us how much larger the standard error is, compared with what it would be if that variable were uncorrelated with the other explanatory variables in the equation.

As logistic regression technique is used in the current study, to deal with the problem of multicollinearity in our Logistic Regression model, we have followed the approach given in [11]. In their study, they suggested that:

- In some situations, when no pair of variables is highly correlated, but several variables are involved in interdependencies, it is better to use multicollinearity diagnostic statistics produced by linear regression analysis.
- Use the dependent variable from logistic regression analysis or any other variable that is not one of the explanatory variables, as a dependent variable in the linear regression.
- Values of Vif exceeding 10 are often regarded as indicating multicollinearity, but in weaker models, which is often the case in logistic regression; values above 2.5 may be a cause for concern.

The results thus obtained are given in Table III and it can be observed that 6 variables have VIF greater than 2.5. Again, linear regression was carried out by removing those variables which have VIF value greater than 2.5 one by one in decreasing order. On removal of var11 and var13, var7 was the only variable with VIF greater than 2.5. Hence, we finally check the data by removing var7 along with var11 and var13. The results are given in Table IV showing variables with 2.39 as maximum VIF value.

C. *Model Equation*

Logistic regression was carried out on the data after removing var7, var11 and var13 from those 26 variables. Further, Backward Elimination method was employed resulting in the following equation :-

$$Z = 3.0987 - 0.0154*ROE + 0.9551*D/A - 0.00661*Sales/Cash - 0.000061*PBITD/S + 0.0352*P/BV - 0.3048*Inflation - 3.4803*GDP$$

where
$Z = \log (p/1-p)$
and 'p' is the probability that the outcome is good.

The estimates of various parameters of the logistic regression model of the stock price return performance are summarized in Table V and the odds ratio estimates are summarized in Table VI.

D. *Selecting Cutoff value*

Having obtained the model equation, in order to classify the companies, we need to have a suitable cutoff value. The selection of cutoff should be in such a manner that cases of misclassifying "Poor" companies into "Good" companies should be minimum. Neter, Wasserman, Nachtsheim and Kutner [12] suggest following ways to select a cutoff value for predicting:

- Determine a cutoff value that will give the best predictive fit for the sample data. This is usually determined through trial and error.
- Select a cutoff value that will separate the sample data into a specific proportion of the two states, based on a prior known proportion split in the population.

Table III. Linear Regression with VIF values

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1.15342	0.1314	8.78	<.0001	0
var1	0.0001188	0.000252	0.47	0.6377	1.5134
var2	0.0043	0.00601	0.72	0.4746	1.1769
var3	4.83E-06	0.000158	0.03	0.9757	1.174
var4	1.07E-05	7.32E-06	1.46	0.1463	1.0359
var5	-0.02193	0.01961	-1.12	0.2643	1.5389
var6	0.04351	0.02572	1.69	0.0917	12.381
var7	0.30284	0.26467	1.14	0.2533	2.8949
var8	0.02141	0.01697	1.26	0.2079	1.641
var9	-0.00258	0.00129	-2	0.0463	1.5248
var10	0.39261	0.20546	1.91	0.0569	4.5141
var11	-0.02442	0.03022	-0.81	0.4195	4.6518
var12	-2.65E-06	5.08E-06	-0.52	0.602	1.1925
var13	-0.04047	0.02257	-1.79	0.0738	12.586
var14	-1.38E-05	5.01E-05	-0.28	0.7824	1.0802
var15	0.00842	0.02452	0.34	0.7316	1.9876
var16	-0.00163	0.00086077	-1.9	0.0585	1.4076
var17	-0.15788	0.16392	-0.96	0.3361	1.1419
var18	0.04779	0.1255	0.38	0.7036	1.1667
var19	-2.68E-06	4.51E-06	-0.6	0.5522	1.0695
var20	-6.56E-06	6.34E-06	-1.03	0.3017	2.6092
var21	0.0002137	0.000365	0.59	0.5585	1.6259
var22	-8.32E-06	5.16E-06	-1.61	0.1075	2.6681
var23	-0.000699	0.00219	-0.32	0.7499	1.6394
Var24	0.00381	0.00468	0.81	0.416	2.2242
Var25	-0.0686	0.01267	-5.42	<0.0001	1.2371
Var26	-0.78599	0.18177	-4.32	<0.0001	1.2026

In our study, we applied the Trial and Error method to come up with the cutoff value. To narrow down the range of cutoff values, scatterplot between Predicted Probability and Actual Result was used. From the scatter plot, it was observed that concentration of "Good" Companies classified as "Good" companies was more for predicted probability exceeding 0.6. Thus, various cutoff values around 0.6 was tried and among them 0.65 was chosen as this resulted in minimum misclassification. Using this cutoff value, any company whose score is greater than or equal to 0.65 would be predicted to be a good performing company, and any company with a score less than 0.65 would be classified as a poor performing company.

Table IV. Linear Regression after removing var7, var11 & var13

Parameter Estimates					
Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1.16646	0.12553	9.29	<.0001	0
var1	0.000111	0.000253	0.44	0.6613	1.5069
var2	0.00527	0.00598	0.88	0.379	1.15451
var3	9.3E-05	0.000154	0.6	0.5478	1.10736
var4	9.75E-06	7.33E-06	1.33	0.184	1.02942
var5	-0.02348	0.01796	-1.31	0.1919	1.27814
var6	-0.00218	0.00766	-0.28	0.7762	1.08794
var8	0.01834	0.01689	1.09	0.2782	1.60972
var9	-0.00276	0.00124	-2.21	0.0275	1.40181
var10	0.19831	0.13669	1.45	0.1477	1.97963
var12	-2.9E-06	5.10E-06	-0.57	0.5698	1.18758
var14	-2.39E-06	4.99E-05	-0.05	0.9618	1.06575
var15	0.01258	0.02285	0.55	0.5823	1.71025
var16	-0.00174	0.000859	-2.03	0.0432	1.3893
var17	-0.16344	0.16338	-1	0.3178	1.124
var18	0.07172	0.1257	0.57	0.5687	1.15979
var19	-2.57E-06	4.49E-06	-0.57	0.5666	1.0509
var20	-6.06E-06	5.76E-06	-1.05	0.2933	2.13353
var21	0.000328	0.000347	0.95	0.3447	1.45835
var22	-8E-06	4.90E-06	-1.63	0.1036	2.39001
var23	-0.00057	0.00219	-0.26	0.7931	1.61546
var24	0.00756	0.00354	2.14	0.0334	1.26326
var25	-0.06652	0.01263	-5.27	<.0001	1.2186
var26	-0.79219	0.18181	-4.36	<.0001	1.19212

Table V. Logistic Regression (Backward Elimination)

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-square	Pr > Chisq
Intercept	1	3.0987	0.5764	28.9048	<.0001
ROE	1	-0.0154	0.00567	7.5203	0.0061
D/A	1	0.9551	0.5220	3.3484	0.0673
Sales/Cash	1	-0.00661	0.00373	3.1432	0.0762
Price / BV	1	-0.000064	0.000018	12.2644	0.0004
PBITD/Sales	1	0.035	0.0185	3.6189	0.0571
Inflation	1	-0.3048	0.0602	25.6684	<.0001
GDP	1	-3.4803	0.8904	15.2782	<.0001

Table VI. Odds Ratio Estimate

Odds Ratio Estimate			
Effect	Point Estimate	95% Wald Confidence Limits	
ROE	0.985	0.974	0.996
D/A	2.599	0.934	7.230
Sales/Cash	0.994	0.987	1.001
Price / BV	1.000	1.000	1.000
PBITD/Sales	1.036	0.999	1.074
Inflation	0.737	0.655	0.830
GDP	0.031	0.005	0.0176

E. Classification Accuracy of the Model

The classification table shown in Table VII helps to assess the performance of the model by cross-tabulating the observed response categories with the predicted response categories.

It is clear from Table VII that the poor companies have a 74% correct classification rate, whereas good companies have a 72.94% correct classification rate. Overall, correct classification was observed in 71.2% of original grouped cases.

IV. Comparative Analysis

In this section, we compare our model with that of work given in [9] (henceforth named as Model X in this section). However, Model X was based on only 116 year wise observations from 2005-2008 and our model has 368 year wise observations from 2012-2015. Accordingly, we carried out comparative analysis with model X against our model and considered only those eight variables mentioned in their work for the time period that we have considered in our study.

A. Model & Classification Accuracy

When we applied full model fit, Logistic Regression with variables specified in earlier model, which were Book Value per share (var1), Price to earnings ratio (var3), Net Sales Increase percentage (var19), Cash Earning per share (var20), Price to cash earning (var21), PBITDS (var22), Sales to Net Asset (var23) and Price to Book value (Var24), it is observed

that out of seven variables, only Var22 and Var24, i.e., PBITDS and Price/Book Value are significant at 10% level of significance. Also, the model so obtained can classify correctly up to 62.77% as shown in Table VIII at the cutoff value of 0.56 which is computed by the method discussed in Section III.D.

In terms of classification accuracy, our model is better as compared to that of Model X. ROC curves for Model X and our model are shown in Fig. 1. From Fig.1, it can be clearly observed that our model is performing better than Model X at all the thresholds. This is also indicated by higher value of AUC for our model namely, 0.7487 when compared to AUC value 0.6427 of Model X.

V. CONCLUSION

This study used the binary logistic regression model to determine the financial ratios that significantly affect the performance of a company in the stock market. As various ratios could be interdependent, the model proposed in this study effectively deals with the problem of multicollinearity.

Table VII. Classification Accuracy of the current model

Result	result_p		
	0	1	Total
0	111	39	150
1	67	151	218
Total	178	190	368

Table VIII. Classification Accuracy of Model X

Result	result_p		
	0	1	Total
0	59	91	150
1	46	172	218
Total	105	263	368

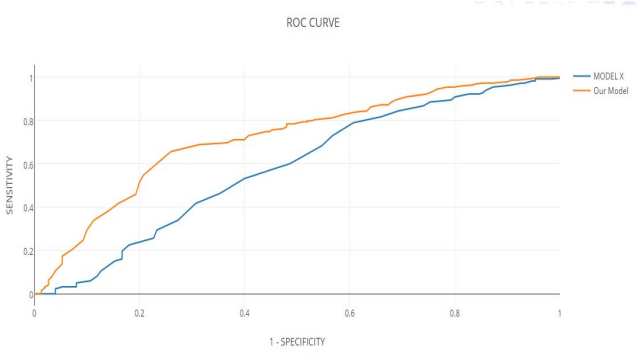

Fig. 1 ROC curves of our model & Model X

It is found that seven financial ratios can classify companies up to a 71.2% level of accuracy into two categories ("good" or "poor"), based on their rate of return. The seven financial ratios are **ROE, Debt to Asset ratio, Sales to cash ratio, PBITDS, Price to Book Value, Inflation & GDP.**

When evaluated from the investors' point of view, we conclude that it is possible to predict out-performing shares by examining these ratios. Though various methods are available for data processing for analysis, we derive insights from this study that ratio methods have the capability to reveal maximum information content, if variables are chosen very carefully with regard to the purpose at hand. Also, when we compared our model with the model given in [9], only two variables out of eight as considered by them came out to be significant in the model. These two variables were found to be common in both the models and also we achieved better classification accuracy and hence we can conclude that the ratios identified by our Logistic Regression model are more important for stock performance prediction. The results of this study help the investor to form an opinion about the shares to be invested.

REFERENCES

[1] C.R. Harvey, Predictable risk and returns in emerging markets, The Review of Financial Studies, vol. 8, pp. 773–816, 1995.

[2] H. Mohammad and P. Nasim, "The Presentation of Financial Crisis Forecast Pattern (Evidence from Tehran Stock Exchange)", International Journal of Finance and Accounting, vol. 1(6), pp. 142-147, 2012.

[3] F.A. de Oliveira, C.N. Nobre and L.E. Zarate, "Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index – Case study of PETR4, Petrobras, Brazil", Expert Systems with Applications, vol. 40, pp. 7596-7606, 2013.

[4] J.L. Ticknor, "A Bayesian regularized artificial neural network for stock market forecasting", Expert Systems with Applications, vol. 40, pp. 5501 – 5506, 2013.

[5] D. Han, L. Ma and C. Yu, "Financial prediction: Application of Logistic Regression with factor Analysis", 4th International Conference on Wireless Communications, Networking and Mobile Computing, Dalian, pp. 1-4, 2008.

[6] J. Chen, M. Chen and N. Ye, "Forecasting the Direction and Strength of Stock Market Movement", Stanford University Technical Report, 2013.

[7] M. Ballings, D.Van den Poel, N. Hespeels, R. Gryp, "Evaluating multiple classifiers for stock price direction prediction", Expert systems with applications, vol. 42, pp. 7046 – 7056, 2015.

[8] A. H. Moghaddam, M.H. Moghaddam and M. Esfandyari, "Stock market index prediction using artificial neural network", Journal of Economics, Finance and Administrative Science, vol. 21, pp. 89 – 93, 2016.

[9] A. Dutta, G. Bandyopadhyay and S. Sengupta, "Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression", International Journal of Business and Information, vol. 7 (1), pp. 105-136, 2012.

[10] D.W. Hosmer, S. Lemeshow and R.X. Sturdivant, Applied Logistic Regression, Wiley, 3rd edition, 2013.

[11] H. Midi, S.K. Sarkar and S. Rana, "Collinearity diagnostics of binary logistic regression model", Journal of Interdisciplinary Mathematics, vol. 13 (3) , pp. 253-267, 2013

[12] J. Neter, W. Wasserman, C.J. Nachtsheim; and M.H. Kutner, Applied Linear Regression Models, 3rd edition, Chicago: Irwin, 1996.