

Filtering out Stocks with High & Low Price-Earning Ratio: A Study on CNX NIFTY Large Midcap 250 Index using Data Mining Technique

Shuvashish Roy¹ and Rajib Bhattacharya²

¹Financial Advisor, Hazrat Khajar Bashir Unani Ayurvedic Medical College & Hospital Foundation, Jamalpur, Bangladesh. Email: shuvashishroy@gmail.com,

²Assistant Professor, Institute of Business Management, The National Council of Education Bengal, Kolkata, West Bengal, India. Email: rajb.conference@gmail.com

ABSTRACT :

Identifying stocks with highest potential for returns from the plethora of stocks in the market poses a crucial problem to the investors. The problem is increasingly getting aggravated due to the increasing complexity in the interrelations of the global macroeconomic variables and their combined effect on the financial performances of firms. A large volume of existing literature have tried to identify the factors which have high predictive accuracy and can filter out stocks which are better performers than the others. Most of the studies in this field has been done using the linear regression technique in various modified forms where either macro-economic factors or composite synthetic market or stock indicators or financial ratios or combination of any of these have been used as predictor variables to predict future returns from a stock. It has also been observed worldwide that investors look for stocks with low price-equity ratio in the anticipation of better returns in future. This study has taken a cue from this tendency of the market and has resorted to a different route. This study has used binary logistic regression to identify significant financial ratios to filter out the low Price-Earning stocks from the market as a whole across sectors. To test the consistency or otherwise of the identified significant financial ratios, this study has been done considering panel data for a time span of last five consecutive years for which annual reports are available i.e. from 31st March 2014 to 31st March 2018. This study has been done on the CNX NIFTY Large Midcap 250 index as a market proxy. The total debt to market capitalization, return on equity and fixed capital to sales ratios have been identified in this study to be statistically significant in segregating stocks into the high PE & low PE groups. The predictor variables have been able to explain 19.40% to 33% of the changes in the dependent variable. The diagnostic ability of the predictor variables, as measured by the area under curve of the receiver operating characteristic curve, have been in the satisfactory range from 0.747 to 0.838. The findings of this study may assist investors in filtering out high and low Price-earning stocks as per their respective requirements using financial ratios.

Key words : Price-earning Ratio, CNX NIFTY Large Midcap 250 Index, Binary Logistic Regression.

INTRODUCTION :

Using financial ratios to take investment decisions is a common practice among investors. Investors often try to identify high PE or low PE stocks to invest as per their respective investing philosophy. This paper attempts to use financial ratios of companies for the previous year to identify the low & high PE stocks. The benchmark for treating a PE ratio to be high or low may be set by the investor. This paper has considered the first decile of the distribution of the PE ratios of the CNX NIFTY Large Midcap 250 index to be the low PE stocks. Identifying investible stocks by predicting future returns thereon, has long remained one of the most researched topic in applied finance. Different statistical techniques from regression to time series analysis have been done on data from different capital markets across the globe. Various macro-economic variables, financial ratios, synthetic market indicators and various combinations of them have been used as predictor variables to predict future returns on stock. No single method has been found to be consistently effective for all stocks and in all markets. The approach in this study is a shift from the traditional endeavour to predict stock returns. Investors all over the globe tend to look for under-priced stocks characterized by low price-equity (PE) ratio in the anticipation of better returns from the stock in future. Thus there is a need for a tool which can identify the financial ratios which can divide the stocks in the market into dichotomous sections of low PE and high PE, the benchmark PE being decided by the investor. This study envisages to identify the financial ratios which are most prominent in determining whether a stock will be a low PE or a high PE one. Binary Logistic Regression technique has been used for the purpose.

SURVEY OF LITERATURE & IDENTIFICATION OF RESEARCH GAP:

Nissim & Penman (2001) found that the returns on common equity, returns on total assets and returns on net operating assets were useful in determining the valuation of common equity shares of organizations. Lewellen (2004) used linear regression to conclude that dividend yield was the prime predictor variable in predicting stock returns. Turk (2006) also used linear regression to test predictive power of financial ratios. He used data from 50 US companies and found that the ratios could not explain the variations in the stock prices as was evident from the low values of adjusted R^2 of models developed by him. Aono & Iwaisako (2010) constructed smoothed market price earnings ratios and examined their predictive ability by linear regression. They found that the empirical performance of the price earnings ratio in forecasting stock returns in Japan was generally weaker than both the price earnings ratio in comparable US studies and the price dividend ratio. Sehgal & Pandey (2010) used alternative price multiples for equity valuation purposes

in the Indian context. They considered data from 145 large companies from 13 prominent sectors. They generated price forecasts based on each multiple by regressing the historical prices on different value drivers and concluded that historical price to earnings (P/E) is the best price multiple for developing price forecasts in the Indian environment. Taani et al (2011) examined the effect of accounting information on earning per share (EPS) by using five categories of financial ratios. They considered a sample of 40 companies listed in the Amman Stock Market to measure the impact of financial ratios on EPS. Multiple regression method was used by them considering profitability, liquidity, debt to equity, market ratio, total assets, and cash flow from operation activities as independent variables. The results show that profitability ratio (ROE), Market ratio (PBV), cash flow from operation/sales, and leverage ratio (DER) has significant impact on earning per share. Siminica et al (2012) studied the statistical correlation between the return on assets and some measures of financial balance. They chose the return on assets as dependent variable. As independent variables, they selected 24 indicators, calculated for 40 Romanian companies listed on Bucharest Stock Exchange. They found that indicators were better predictors than financial ratios. Borhan & Mohamed (2014) explored the impact of financial ratios on the financial performance of a chemical company. Current ratio, quick ratio, debt ratio, debt equity ratio, operating profit margin and net profit margin were used as predictor variables. A multiple regression model was used. The results shows that Current ratio, Quick ratio, Debt ratio and Net profit margin had a positive relationship while debt-equity ratio and operating profit margin had a negative relationship with the company's financial performance. Inyama & Ugah (2015) examined the relationships between financial ratios and share price movements in Nigeria oil and gas sector. A sample of four oil and gas companies listed in the Nigeria Stock Exchange from 2002 to 2014 was used. Cross-sectional correlation analysis was used to establish the relationship. Share price was regressed using Earnings per share, Dividend per share, Net asset value per share and Dividend cover as predictor variables. They found that there was a weak and insignificant relationship between financial ratios and share movements. Wijesundera et al (2015) examined the ability of the historical financial information to predict stock returns in the Sri Lankan market. They applied ordinary least squares technique to estimate the predictive regressions in form of simple and multiple models of panel data sets. The results highlighted that ROE, EPS and MV/BV have a significant positive relationship with the stock returns. Arkan (2016) investigated the importance of financial ratios to predict stock price trends in emerging markets. Prediction power of 12 financial ratios was tested depending on data of 15 companies distributed over three sectors for the years 2005–2014 in the Kuwait capital market. Multiple regression technique was used by him. The results showed that the most effective ratios on the stock price for the industrial sector are ROA, ROE and net profit ratio. The most effective ratio on the stock price for the service sector was the ROA, ROE,

P/E and EPS ratio and the same for the investment sector. Chan (2016) explored the improvement in multiple-based valuations from using a composite of price to earnings (P/E) and price to book (P/B) ratios and firm-specific regression-based weights. His results supported that composite benchmark multiples lead to improved valuations over single multiples. Further improvement was achieved by incorporating firm characteristics to derive firm-specific regression-based weights. The unrestricted regression-weighted composite multiples perform better than other approaches in predicting year one to year three share prices. Kharatyan & Nunes (2016) investigated factors that affected return on equity (ROE). The study was conducted on a sample of 90 non-financial companies which were components of NASDAQ-100 index. The ordinary least squares method was used. Consequently, the study used eight ratios/indicators that are believed to have impact on ROE. They found out that the most relevant ratios that determine ROE are tax burden, interest burden, operating margin, asset turnover and financial leverage. Ping-Fu & Kwai-Yee (2016) examined the effectiveness of financial ratios of 17 firms of 50 HSI constituent stocks in Hong Kong Stock Exchange and using a multiple regression analysis technique. Predictor variables were price-to-sales, market-to-book, earnings per share, dividend yield, market capitalization etc., against the dependent variable of stock returns. Whilst the literature suggests that there is a clear relationship and dependence between these variables, the results of this research proved inconclusive. Ghajar & Reja (2017) investigated the effect of ratios measuring liquidity risks, Credit, capital adequacy & unexpected interest rate of commercial banks in the Stock Exchange between years 2011-2015 by using factor analysis technique and subsequent linear regression to identify 18 financial ratios to calculate the risks to be included in the regression model. Jagannathan & Suresh (2017) considered data from 500 stocks in the Standards and Poor's (S&P) 500 and checked the predictive accuracy of Price to Earnings (PE), Price to Sales (PS), Price to Earnings Growth (PEG) & Price to Book (PB) on actual stock price movements. Their results indicated that none of PE, PB, PS, PB ratios was a good predictor of stock price changes. Meric et al (2017) explored the relationship between stock prices and price earning & dividend yield by applying VAR technique. The output obtained by them suffered from inconsistency. Theogene et al (2017) analyzed the contribution of financial ratio analysis on decision making in commercial banks in Rwanda. They used regression analysis and found that financial ratios were effective in predicting organizational performance. Rm Musallam (2018) explored the relationship between financial ratios and market stock returns of twenty six listed firms in Qatar from 2009 to 2015 by using Weighted Least Square (WLS) technique. They found out that earnings per share, earnings yield ratio, and dividend yield ratio had significant and positive relationship with market stock returns while market to book value ratio, return on assets, return on equity, price to earnings ratio, dividends earnings ratio, and net profit margin had insignificant relationship with market stock returns.

The survey of literature revealed the dearth of studies linking financial ratios and categorization of stocks into high and low PE stock groups. This study has thus been triggered by the need for a study on applicability of using financial ratio in segregating stocks into low PE and high PE groups. Accordingly, binary logistic regression technique has been used in this study.

OBJECTIVE OF THE STUDY :

The prime objective of this study is to identify financial ratios which can segregate stocks into high and low Price-earning groups using binary logistic regression technique. Another objective of this study is to test whether the significant ratios so identified, remain the same over the years or change over time.

METHODOLOGY OF THE STUDY :

The CNX NIFTY Large Midcap 250 index has been selected as the market proxy as it the most broadly based index in the Indian capital market in the NSE family of indices. The CNX NIFTY Large Midcap 250 reflects the performance of a portfolio of 100 large cap and 150 mid cap companies listed on NSE, represented through the CNX NIFTY 100 and the CNX NIFTY Midcap 150 index respectively. The aggregate weight of large cap stocks and mid cap stocks is 50% each and are reset on a quarterly basis. The index has a base date of April 01, 2005 and a base value of 1000. Index consists of all Stocks forming part of NIFTY 100 and NIFTY Midcap 150 index. Aggregate weight of large cap stock and Midcap stocks are capped at 50% each. Index shall be reconstituted on a semi-annual basis along with NIFTY 100 and NIFTY Midcap 150 index. Weights of large cap and midcap stocks are rebalanced on a quarterly basis.

While selecting the financial ratios, the ratios selected by researchers and scholars in earlier studies have been considered. The ratios have been selected so as to avoid ratios which measure the same aspect and thus have high positive correlations among them. The following financial ratios (with their respective abbreviations used in this study) have been selected to be the predictor variables:

Table 1
Ratios used in identifying the significant financial ratios & their abbreviations

Ratio	Abbreviation	Description
Dividend Payout	DVPO	An indicator of percentage of Net Income distributed as dividends
Post-Tax Profit Margin	PATM	An indicator of post-tax profitability

Filtering out Stocks with High & Low Price-Earning Ratio: A Study on CNX NIFTY Large Midcap 250 Index using Data Mining Technique

Cash Profit Margin	CPRM	An indicator of cash profitability
Return on Assets %	ROAM	Net Income expressed s a percentage of total assets
Return on Equity %	ROEM	Net Income compared to Equity Shareholders' Funds
Return on Capital Employed %	ROCE	Net Operating Income compared to total capital employed by a business
Total Assets Turnover	TATO	An indicator of utilization of total assets in generating sales
Fixed Capital to Sales	FXCS	An indicator of efficiency in using fixed capital to generate sales
Total Debt to Market Capitalization	TDMC	A comparison between Total Debt and Total market value of outstanding shares

Source : Authors' own compilations and assignment of abbreviations

The Adjusted PE Ratio has been used as the dependent variable. The Adjusted PE ratio is one of the most widely used tools for stock selection. A high Adjusted PE ratio indicates that investors expect higher earnings. However, a stock with a high Adjusted PE ratio is not necessarily a better investment than one with a lower Adjusted PE ratio, as a high Adjusted PE ratio can indicate that the stock is being overvalued. Companies with high Price Earnings Ratio are often considered to be growth stocks. It means that investors have higher expectations for future earnings growth and are willing to pay more for them as it indicates a positive future performance. However, the disadvantage of high Adjusted PE is that growth stocks are often unpredictable, and this puts a lot of pressure on companies to do more to justify their higher valuation. Therefore, investing in growth stocks will more likely be a risky investment. Also, in some cases, it can even be interpreted as an overpriced stock.

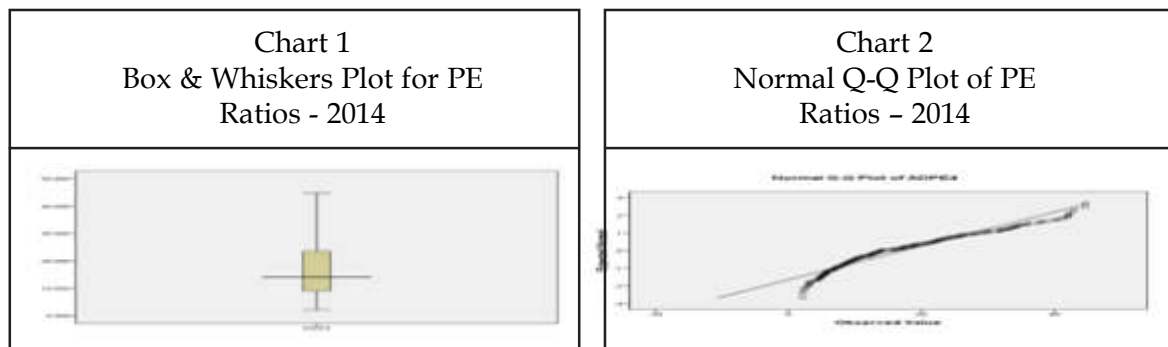
A lower Adjusted PE ratio means that investors are paying less per dollar of company earnings, and that it will take less time for the company to earn enough to buy back its shares. Thus in essence, When a company's stock has a low Adjusted PE ratio, it may indicate that the stock is undervalued. Investors can often buy undervalued stock at a discount and then profit when the price of that stock climbs.

However, a low Adjusted PE ratio sometimes reflects a genuine lack of growth potential. Companies with low Price Earnings Ratio are often considered to have undervalued stocks. This means that the price of their stock is relatively low. This lower pricing of stock attracts investors to buy their stock before the market corrects it. When it does so, investors make profits because of the higher stock price. The adjusted PE ratio

has been revealed to be the most appropriate one in the Indian markets (Sehgal & Pandey, 2010). The dependent variable has been taken to be lagged one by 1 year i.e. the dependent adjusted PE ratio for 2018 has been regressed using the predictor variables for 2017 and so on. The selected ratio for the constituent were collected for five consecutive years from year ended 31 March 2014 to year ended 31 March 2018 to capture the recency as well as the trajectory of the effects of the selected ratios in identifying the stocks into high PE and low PE categories. The ratio of all the years for the constituents of the selected index were obtained from Ace Equity Data Product v2 10.3.0. The Binary Logistic Regression has been carried out on the R Studio Platform, version 3.5.1. The R codes, as have been used in this study, have been provided in the Appendix.

The outliers in the PE ratios of the constitute companies were identified by plotting Box & Whiskers Plot of the PE ratios of the constituent stocks for each year. The outliers were excluded from the data set. Similarly stocks with negative and zero PE ratios were also excluded for all the five years under study. Thus the number of data points varied slightly from year to year. After excluding the negative and zero PE ratios and the outlier PE ratios, the distributions of PE ratios for all the five years were found to be more or less normal.

The box plots and the normal Q-Q plots for the PE ratios of the constituent companies of CNX NIFTY Large Midcap 250 index, for the years 2013-14 to 2017-18, are contained serially in Chart Nos. 1-10.



Filtering out Stocks with High & Low Price-Earning Ratio: A Study on CNX NIFTY Large Midcap 250 Index using Data Mining Technique

Chart 3
Box & Whiskers Plot for PE
Ratios - 2015

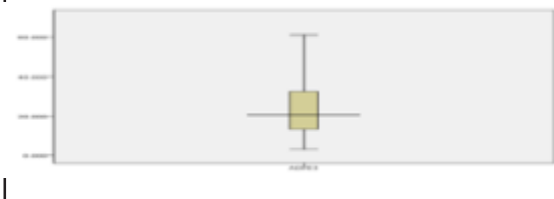


Chart 4
Normal Q-Q Plot of PE
Ratios - 2015

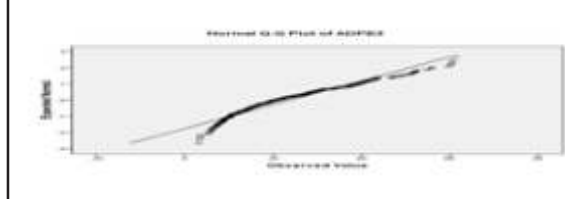


Chart 5
Box & Whiskers Plot for PE
Ratios - 2016

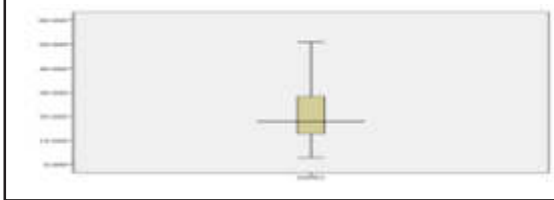


Chart 6
Normal Q-Q Plot of PE
Ratios - 2016

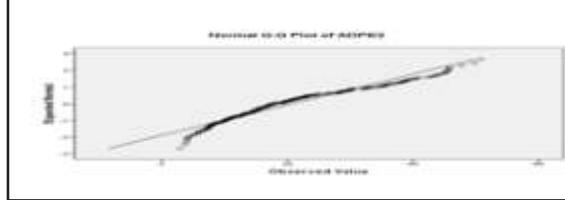


Chart 7
Box & Whiskers Plot for PE
Ratios - 2017

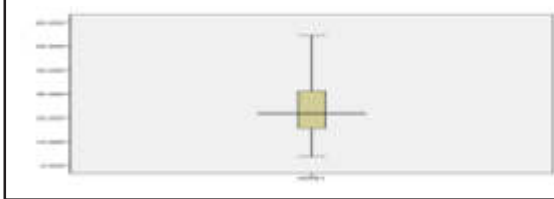
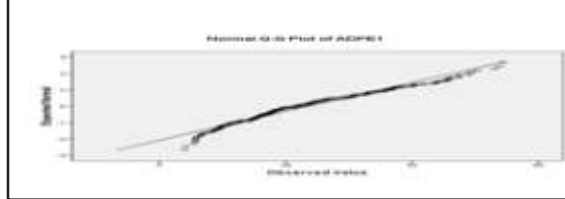
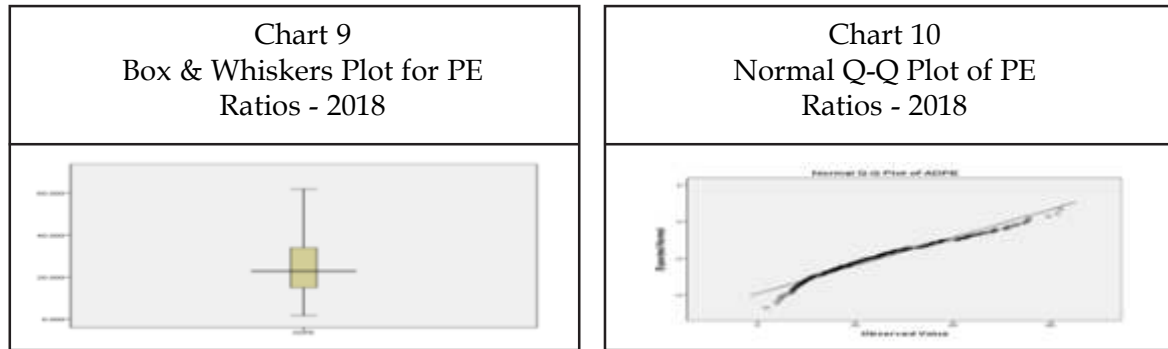


Chart 8
Normal Q-Q Plot of PE
Ratios - 2017





The year-wise summarized results from 2013-14 to 2017-18, of the Binary Logistic Regression have been appended below which principally constitutes Deviance Residuals, Significant Predictor Variables with Coefficients, ANOVA, Confidence Intervals, Odds Ratio with Confidence Intervals, Importance of the Identified Significant Predictor Variables, Wald's Test, Likelihood Ratio Test, Pseudo R^2 Tests (Adjusted McFadden, Cox & Snell, Nagelkerke and Corrected Akaike Information Criteria [AIC]), Hosmer & Lemeshow Test for Goodness of Fit and Receiver Operating Characteristic (ROC) Curve to show the Area under Curve (AUC).

Results of Binary Logistic Regression for the Year 2013-14

Table 2

Deviance Residuals					
Min	1Q	Median	3Q	Max	
-2.73738336	0.13657646	0.47200614	0.65250473	1.9319382	
Coefficients					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.17429417	0.5100278	0.34173464	0.7325506	
dvpo4	0.07001869	0.02269898	3.08466179	0.00203784	
tdmc4	-0.21878558	0.11221928	-1.94962555	0.05122077	
ANOVA					
	df	Deviance	Resid.df	Resid. Dev	Pr(>Chi)
NULL	NA	NA	127	131.8902124	NA
dvpo4	1	18.64445823	126	113.24575417	0.00001575
tdmc4	1	4.74927322	125	108.49648095	0.02931067

Filtering out Stocks with High & Low Price-Earning Ratio: A Study on CNX NIFTY Large Midcap 250 Index using Data Mining Technique

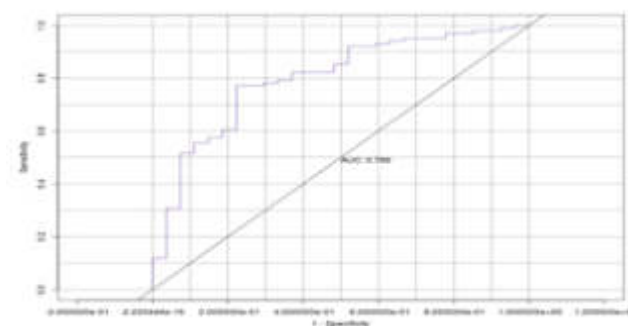
Table : 3

Confidence Intervals			Odds Ratio and Confidence Intervals			
	2.5 %	97.5 %		Odds Ratio	2.5%	97.5%
(Intercept)	-0.8327374	1.18560309	(Intercept)	1.19040569	0.43485728	3.27265994
dvpo4	0.02909638	0.11839712	dvpo4	1.07252822	1.02952382	1.12569106
tdmc4	-0.4709011	-0.02089446	tdmc4	0.80349399	0.62443933	0.97932231

Table : 4

		Variable Importance			
Pseudo R Square		Overall			
Overall		dvpo4		3.08466179	
		tdmc4		1.94962555	
McFadden	0.17737276	Wald Test			
Adj.McFadden	0.11671625	Resid.df	df	F	Pr(>F)
Cox.Snell	0.16703489	125	NA	NA	NA
Nagelkerke	0.25972058	127	-2	7.67292091	0.00071929
McKelvey.	0.99751247	Hosmer-Lemeshow Goodness of fit Test			
Zavoina		X squared		df	p value
Effron	0.1916698	8.69452306470127		8	0.368717265506618
Count	0.8046875	Likelihood Ratio Test			
Adj.Count	0.07407407	#df	LogLik	df	Chisq
AIC	114.49648095				Pr(>Chisq)
Corrected.AIC	114.69002934	3	-54.24824048	NA	NA
		1	-65.9451062	-2	23.39373145
					0.0000083
					1985507

Chart 11
AUC of ROC - 2013-14



Results of Binary Logistic Regression for the Year 2014-15

Table 5

Deviance Residuals					
Min	1Q	Median	3Q	Max	
-2.43330016	0.21515526	0.57874875	0.6963261	2.01187986	
Coefficients					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.17017304	0.62562427	1.87040864	0.06142709	
dvpo3	0.02261059	0.01488903	1.51860705	0.12886144	
roem3	-0.07551091	0.03879769	-1.94627334	0.05162191	
roce3	0.06019715	0.03818798	1.57633776	0.11494798	
tdmc3	-0.40587752	0.17516726	-2.31708558	0.02049907	
ANOVA					
	df	Deviance	Resid.df	Resid. Dev	Pr(>Chi)
NULL	NA	NA	129	142.81802011	NA
dvpo3	1	6.82191486	128	135.99610525	0.0090046
roem3	1	0.01343497	127	135.98267028	0.90772441
roce3	1	2.84866414	126	133.13400614	0.09144989
tdmc3	1	8.32340863	125	124.81059751	0.00391376

Table : 6

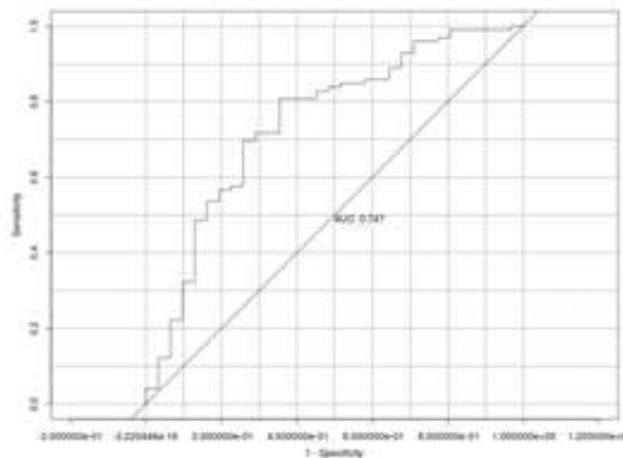
Confidence Intervals			Odds Ratio and Confidence Intervals		
	2.5 %	97.5 %		Odds Ratio	2.5% 97.5%
(Intercept)	-0.056028	2.39637408	(Intercept)	3.22255022	0.948947891 1.190562
dvpo3	-0.00657138	0.05179255	dvpo3	1.02286814	0.99685184 1.05658886
roem3	-0.15155299	0.00053117	roem3	0.92726961	0.85651616 1.00028293
roce3	-0.01464991	0.13504421	roce3	1.06204591	0.98800269 1.15044772
tdmc3	-0.74919904	-0.06255601	tdmc3	0.66639178	0.44807639 0.8918045

Filtering out Stocks with High & Low Price-Earning Ratio: A Study on CNX NIFTY Large Midcap 250 Index using Data Mining Technique

Table 7

Pseudo R Square		Likelihood Ratio Test				
	Overall	#df	LogLik	df	Chisq	Pr(>Chisq)
McFadden	0.12608649	5	62.40529876	NA	NA	NA
Adj.McFadden	0.04206348	1	-71.40901006	-4	18.0074226	0.00122998
Cox.Snell	0.12935297	Wald Test				
Nagelkerke	0.19403064	Resid.df	df	F	Pr(>F)	
McKelvey.	0.24602751	125	NA	NA	NA	
Zavoina		129	-4	3.13331626	0.01706783	
Effron	0.15048602	Variable Importance				
Count	0.78461538	Overall				
Adj.Count	0.09677419					
AIC	134.81059751					
Corrected.	135.29446848					
AIC						
		Hosmer-Lemeshow Goodness of fit Test				
		X squared		df	p value	
		14.5576504533579		8	0.0683392923892889	

Chart 12
AUC of ROC - 2014-15



Results of Binary Logistic Regression for the Year 2015-16

Table : 8

Deviance Residuals					
Min	1Q	Median	3Q	Max	
-2.16415795	-0.22755816	0.5653764	0.65962626	2.90260117	
Coefficients					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.3698821	0.49982975	4.74137859	0.00000212268848	
tato2	-0.51676039	0.29536586	-1.7495603	0.08019422	
tdmc2	-0.87370036	0.23611045	-3.70038842	0.00021527	
ANOVA					
	df	Deviance	Resid.df	Resid. Dev	Pr(>Chi)
NULL	NA	NA	129	149.41158868	NA
tato2	1	0.03280247	128	149.37878622	0.85627769
tdmc2	1	24.81614285	127	124.56264336	6.30670848e-7

Table : 9

Confidence Intervals			Odds Ratio and Confidence Intervals		
	2.5 %	97.5 %		Odds Ratio	2.5% 97.5%
(Intercept)	1.39023378	3.34953041	(Intercept)	10.6961311	4.189211023 0.23265063
tato2	-1.09566684	0.06214607	tato2	0.59644969	0.329653 1.07879523
tdmc2	-1.33646833	-0.41093239	tdmc2	0.41740414	0.24865401 0.62559946

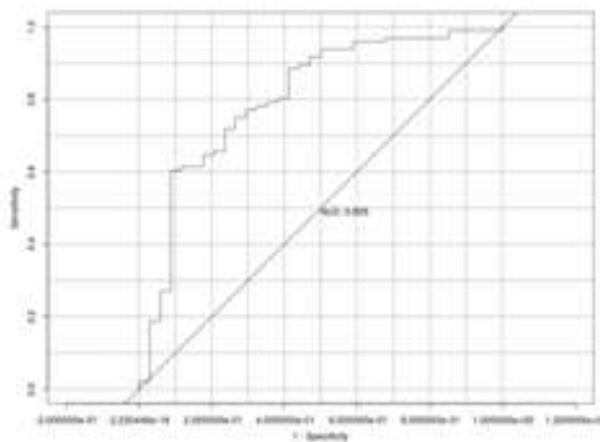
Table : 10

Pseudo R Square		Variable Importance			
Overall		Overall			
McFadden	0.16631203	tato2	1.7495603		
Adj.McFadden	0.11276866	tdmc2	3.70038842		
Cox.Snell	0.1739878	Wald Test			
Nagelkerke	0.25468543	Resid.df	df	F	Pr(>F)
McKelvey.	0.29895549	127	NA	NA	NA
Zavoina		129	-2	6.93133885	0.00138993

Filtering out Stocks with High & Low Price-Earning Ratio: A Study on CNX NIFTY Large Midcap 250 Index using Data Mining Technique

Effron	0.22333559	Likelihood Ratio Test				
Count	0.79230769	#df	LogLik	df	Chisq	Pr(>Chisq)
Adj.	0.20588235	3-6	2.28132168	NA	NA	NA
Count		1	-74.70579434	-2	24.84894532	0.0000040
AIC	130.56264336					1901922
Corrected.	130.75311955	Hosmer-Lemeshow Goodness of fit Test				
AIC		X squared		df	p value	
		8.9125036934861		8	0.349730778147645	

Chart : 13
AUC of ROC - 2015-16



Results of Binary Logistic Regression for the Year 2016-17

Table 11

Deviance Residuals				
Min	1Q	Median	3Q	Max
-2.24731162	0.40464222	0.55484835	0.63520787	2.45626552
Coefficients				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.40508051	0.75444671	4.51334796	0.0000063812199
roem1	-0.0338528	0.02247324	-1.5063599	0.13197479

tato1	-0.6014302	0.40542675	-1.4834499	0.13795491	
fxcs1	-0.6517597	0.41312493	-1.5776334	0.11464983	
tdmc1	-0.8835710	0.32863989	-2.6885689	0.0071759	
ANOVA					
	df	Deviance	Resid.df	Resid. Dev	Pr(>Chi)
NULL	NA	NA	129	138.00145458	NA
roem1	1	0.57473308	128	137.4267215	0.44838422
tato1	1	0.00767649	127	137.41904501	0.93018221
fxcs1	1	4.62587402	126	132.79317099	0.03149321
tdmc1	1	14.27050047	125	118.52267052	0.00015833

Table : 12

Confidence Intervals			Odds Ratio and Confidence Intervals		
	2.5 %	97.5 %		2.5%	97.5%
(Intercept)	1.92639214	4.88376889	(Intercept)	30.11672043	7.591152315
roem1	-0.07789955	0.01019395	roem1	0.9667138	0.9237982
tato1	-1.3960521	0.19319154	tato1	0.54802724	0.24260986
fxcs1	-1.46146969	0.15795026	fxcs1	0.52112793	0.19800417
tdmc1	-1.52769338	-0.23944866	tdmc1	0.41330436	0.19755625

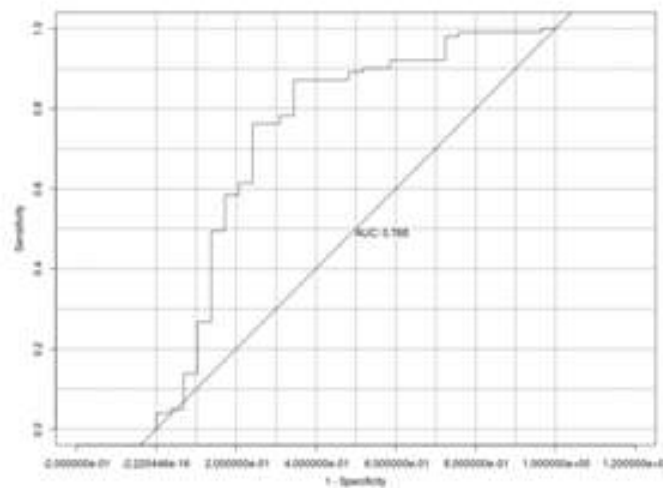
Table : 13

Pseudo R Square		Variable Importance			
	Overall				Overall
McFadden	0.14114912	roem1			1.50635998
Adj.McFadden	0.05419352	tato1			1.48344994
Cox.Snell	0.13915154	fxcs1			1.57763349
Nagelkerke	0.21274371	tdmc1			2.68856896
McKelvey.	0.27870145				
Zavoina		Likelihood Ratio Test			
Effron	0.17961488	#df	LogLik	df	Chisq
Count	0.82307692				Pr(>Chisq)
Adj.	0.20689655	5	-59.26133526	NA	NA
Count		1	-69.00072729	-4	19.47878406

Filtering out Stocks with High & Low Price-Earning Ratio: A Study on CNX NIFTY Large Midcap 250 Index using Data Mining Technique

AIC	128.52267052	Hosmer-Lemeshow Goodness of fit Test		
Corrected. AIC	129.00654149			
		X squared	df	p value
		8.03305459177085	8	0.430247924046507

Chart : 14
AUC of ROC - 2016-17



Results of Binary Logistic Regression for the Year 2017-18

Table 14

Deviance Residuals				
Min	1Q	Median	3Q	Max
-2.57821989	0.13834815	0.23127194	0.33440925	0.9419366
Coefficients				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.50435584	1.45818488	2.4032315	0.0162509
roem	-0.2352154	0.07512044	-3.13117698	0.00174107
roce	0.22379107	0.10532544	2.12475811	0.0336068
fxcs	-1.0441632	0.78657841	-1.32747504	0.18435159

ANOVA					
	df	Deviance	Resid.df	Resid. Dev	Pr(>Chi)
NULL	NA	NA	128	65.28353967	NA
Roem	1	0.77573486	127	64.5078048	0.37844875
Roce	1	13.05994843	126	51.44785638	0.00030168
FxcS	1	4.32406203	125	47.12379435	0.03757712

Table : 15

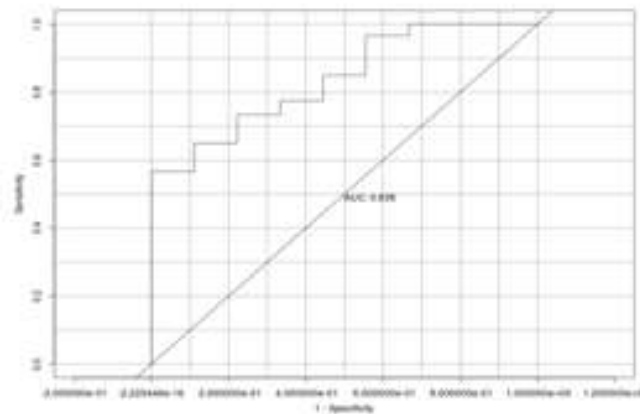
Confidence Intervals			Odds Ratio and Confidence Intervals		
	2.5 %	97.5 %		2.5%	97.5%
(Intercept)	0.64636599	6.36234569	(Intercept)	3.26001227	2.550289380
roem	-0.38244877	-0.08798204	roem	0.79040058	0.66775564
roce	0.01735701	0.43022514	roce	1.25080967	1.06659453
fxcs	-2.58582856	0.49750215	fxcs	0.35198624	0.05980211

Table : 16

Pseudo R Square		Variable Importance			
	Overall				Overall
McFadden	0.27816729	roem			3.13117698
Adj.McFadden	0.12498932	roce			2.12475811
Cox.Snell	0.13131371	fxcs			1.32747504
Nagelkerke	0.33064657	Wald Test			
McKelvey.	0.42529089	Resid.df	df	F	Pr(>F)
Zavoina		125	NA	NA	NA
Effron	0.26399195	128	-3	3.6895402	0.01379184
Count	0.95348837	Likelihood Ratio Test			
Adj. Count	0.33333333	#df	LogLik	df	Chisq
AIC	55.12379435	4	-23.56189717	NA	NA
Corrected. AIC	55.44637499	1	-32.64176983	-3	18.15974532
					0.00040771
		Hosmer-Lemeshow Goodness of fit Test			
		X squared	df	p value	
		3.85786833007306	8	0.869713903772775	

Filtering out Stocks with High & Low Price-Earning Ratio: A Study on CNX NIFTY Large Midcap 250 Index using Data Mining Technique

Chart : 15
AUC of ROC - 2017-18



In order to identify the low PE stocks, the first decile of the distribution of the PE ratios of the constituent companies of the CNX NIFTY Large Midcap 250 index, has been considered to be the low PE stocks. The low PE stocks have been assigned the value of “0” and the others have been assigned the value of “1”. This threshold of 10% of the range may be altered by the investor.

Binary logistic regression has been carried out for each year and the results have been interpreted from the extent of Goodness of fit using Hosmer Lemeshow Test, the extent of variation of the dependent variable explained by the predictor variables and overall accuracy of the identified significant financial ratios in correctly classifying the stocks into high & low PE groups. Moreover, the predictor variables making significant contributions to the accuracy of classification has also been traced for the years from 2013-14 to 2017-18.

FINDINGS OF THE STUDY :

The Cox and Snell R^2 as well as the Nagelkerke R^2 both have been calculated for the years as detailed below. As Nagelkerke R^2 is a betterment over the Cox & Snell R^2 , the results are interpreted on the basis of Nagelkerke R^2 . The minimum value of the corrected AIC of the final iteration for arriving at the binary logistic regression equation are also shown below.

Table : 17
Values of AIC and R² for the five years under study

Years	Corrected AIC	Cox & Snell R ²	Nagelkerke R ²
2014	114.69	0.167	0.259
2015	135.294	0.129	0.194
2016	130.753	0.173	0.255
2017	129.006	0.139	0.212
2018	55.446	0.131	0.330

Source : Authors' own calculations

It has been observed that the portion of variation of the dependent variable as could be explained by the predictor variables is maximum 33% which is in 2018. The extent of explained variation in the dependent variable undulated over the period from 2014 to 2018 within 19.40% and 33%. This observation reveals that other factors, not considered in this study, accounted for major portion of the unexplained variation of the dependent variable.

The goodness of fit of the regression equation was tested by the Hosmer Lemeshow Test, the results of which are appended below:

Table : 18
Results of the Hosmer Lemeshow Test

Year	Chi Square	Df	p Value at $\alpha = 5\%$
2014	8.694	8	0.368
2015	14.557	8	0.068
2016	8.912	8	0.349
2017	8.033	8	0.430
2018	3.857	8	0.861

Source : Authors' own calculations

The p-values being higher than the α value of 0.05 accepts the null and the goodness of fit of the regression equation is established.

The overall accuracy of the regression equation in correctly predicting low and high PE stocks has been estimated by plotting the Receiver Operating Characteristic (ROC)

curve and observing the Area under Curve (AUC). AUC provides an aggregate measure of performance across all possible classification thresholds. The area under curve (AUC) serves as an measure of accuracy or concordance index. The AUC remained between 0.74 and 0.84. The AUC remained more towards the higher end of 1. The diagnostic ability of the binary classifier has thus been satisfactory for all the years from 2013-14 to 2017-18.

Table : 19
AUC of ROC

Year	AUC of ROC
2014	0.799
2015	0.747
2016	0.805
2017	0.768
2018	0.838

Source : Authors' own calculations

The predictor variables in making significant contribution towards the predictive accuracy of the regression equation over the period from 2014 to 2018, is appended below

Table : 20
Predictor Variables having significant contribution to correct classification

2014	2015	2016	2017	2018
DVPO	DVPO	TATO	ROEM	ROEM
TDMC	ROEM	TDMC	TATO	ROCE
	ROCE		FXCS	FXCS
	TDMC		TDMC	

Source : Authors' own compilation

It has been observed from the table above that total debt to market capitalization has been a significant predictor variable in classifying stocks into high PE and low PE groups in four out of the five years covered under this study. Dividend payout was a significant contributor till 2015 after which it ceased to be so. Total Asset Turnover was a significant contributor in the two consecutive years 2016 & 2017. Return on equity was a significant predictor variable in three out of five years covered under the study. Return on Capital Employed turned out to be significant predictors abruptly in the years 2015 &

2018. Fixed Capital to Sales has emerged as significant predictor variables in the two recent years of 2017 & 2018. On summing up these observations, it transpires that Total Debt to Market capitalization, Return on Equity & Fixed Capital To Sales ratios are significant predictor variables in classifying stocks into high PE and low PE groups.

The findings above corroborates the fact that the selected financial ratios can be used to predict whether the stock will be a low PE stock or high PE stock next year with a limited accuracy rate. The significant financial ratio in contributing to the predictive accuracy, has been total debt to market capitalization, return on equity and fixed capital to sales, the first one being the most prominent of them.

CONCLUSION :

The findings of this study has led to the conclusion that using financial ratio to categorize stocks in the Indian market into low PE and high PE groups can lead to limited utility. The extent of variation in the dependent variable by the predictor variables has oscillated between 19.40% and 33% during the period 2013-14 to 2017-18. One of the reasons for such low percentage is that PE ratio of a stock is dependent on the market price which in turn is governed by the demand of that stock as supply is constant. The demand for stocks depend on numerous factors of which past results of the company measured by financial ratios form only a small part. Other macro-economic, legal, political and other factors also affect stock prices. The oscillating nature of the correct prediction of low PE stocks by financial ratios, indicate towards the undulating behavior of investors in taking invest or divestment decisions based on immediately past performances measured by financial ratios. In spite of such oscillations, the total debt to market capitalization ratio has consistently been a significant financial ratio in categorizing the stocks into low & high PE ratio groups. Thus it transpires from this study that financial leverage is consistently significant in determining the categorization of high & low PE stocks. Return on equity and fixed capital to sales have also been found to be important ratios considered by the investors. The diagnostic ability of the selected predictor variables, as evident from the AUC of ROC have been varying from 0.747 to 0.838. The values being closer to the upper limit of 1, indicates satisfactory classifying ability of the binary logistic equations.

RECOMMENDATIONS BASED ON THE INFERENCES DRAWN FROM THE STUDY :

Investors willing to predict underpriced low PE ratio stocks in the next year should focus on the total debt to market capitalization ratio, return on equity and also on fixed capital to sales. The low PE stocks have been found to be non sector-specific and have been found to be scattered over various sectors. Thus investors should do well to avoid mental attachment to any specific sector for having low PE ratio stocks.

SCOPE OF FUTURE RESEARCH :

This study has been done considering the adjusted PE ratio to be the dependent variable. Moreover this study has envisaged only the CNX NIFTY Large Midcap 250 index. However, various other valuation ratios e.g. Price to Book Value and Enterprise Value – based ratios may also be used as the dependent variable. In such situations, instead of going in for the lowest values of these ratios, one should go for highest values. The number of independent variables may be extended to other financial ratios also. This study may be repeated using Tjur R² (2009) which, as opined by some scholars, is preferred to Nagelkerke's R² as it satisfies Kvalseth's (1985) eight properties of a good R² in a better way. The results of this study may also be attempted to be validated by testing with other broad-based, strategy, thematic and sectoral indices also, not only in Indian markets but in global markets as well.

REFERENCES :

- Aono, Kohei., & Iwaisako, Tokuo. (2010). Forecasting Japanese stock returns with financial ratios and other variables. *Working Paper Series No. 63. Research Center For Price Dynamics, Institute of Economic Research, Hitotsubashi University, Japan.*
- Arkan, T. (2016). The importance Of financial ratios in predicting stock price trends: A case study in emerging markets. *Finanse, Rynki Finansowe, Ubezpieczenia*, 1 (79), 13–26.
- Borhan, Dr.Halimahton,. Mohamed, Rozita Naina,. (2014). The impact of financial ratios on the financial performance of a chemical company. *World Journal of Entrepreneurship, Management and Sustainable Development*, 10, (2), 154-160.
- Chan, Kelly, (2016). Equity valuation using benchmark multiples: an improved approach using regression-based weights. *Corporate Ownership & Control*, 13, (4).
- Ghajar Sarkolatheh, Milad. & Reza Modanlo Joibary, Ali. (2017). Investigating the relationship between the income response rate and the financial ratio with abnormal stock among the banks listed in Tehran stock exchange. *International Journal of Scientific Study*, 5 (4), 636-646.
- Inyama, Oliver Ike., Ugah, Helen,. (2015). Evaluation of the relationships between financial ratios and share price movements in Nigeria Oil and Gas (2002 – 2014), 3, (4), 367-375.
- Jagannathan, Uday Kumar., Suresh, N. (2017). A study on determinants of short term return on stocks in the s&p500. retrieved from <https://ssrn.com/Abstract=3029505>
- Kharatyan, Davit., Nunes, José Carlos Lopes. (2016). Financial ratios and indicators that determine return on equity. Retrieved from <http://www.aeca1.org/xviiencuentroaeca/comunicaciones/104a.pdf>

- Kvalseth, T.O. (1985). Cautionary note about R^2 . *The American Statistician*. 39, 279-285.
- Lewellen, Jonathan. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*. 74, (2), 209-235.
- Meriç, Ekrem., Kamisli, Melik. & Temizel, Fatih. (2017). Interactions among stock price and financial ratios: The case of Turkish banking sector. *Applied Economics and Finance*. 4, (6).
- Nissim, Doron. & H. Penman, Stephen. (2001). Ratio analysis and equity valuation: From research to practice. *Review of Accounting Studies*. 6, 109-154.
- Ping-Fu, Lai., Kwai-Yee, Cho. (2016). Relationships between stock returns and corporate financial ratios based on a statistical analysis of corporate data from the Hong Kong Stock Market. *Public Finance Quarterly*. 1, 110-123.
- Rm Musallam, Sami. (2018). Exploring the relationship between financial ratios and market stock returns. *Eurasian Journal of Business And Economics*. 2018, 11(21), 101-116.
- Sehgal, Sanjay., & Pandey, Asheesh. (2010). Equity valuation using price multiples: Evidence from India. *Asian Academy of Management Journal of Accounting And Finance*. 6, (1), 89-108.
- Siminica, Marian., Circiumaru, Daniel. & Simion, Dalia. (2012). The correlation between the return on assets and the measures of financial balance for Romanian companies. *International Journal of Mathematical Models and Methods in Applied Sciences*. 2, (6).
- Taani, Khalaf., Hasan Hamed Banykhaled, Mari'e. (2011). The effect of financial ratios, firm size and cash flows from operating activities on earnings per share: (An Applied Study: On Jordanian Industrial Sector). *International Journal of Social Sciences and Humanity Studies*, 3, (1), 2011.
- Theogene, Habimana., Mulegi, Tom. & Hosee, Niyompano. (2017). The contribution of financial ratios analysis on effective decision making in commercial banks. *International Journal of Management and Applied Science*. 3, (6)
- Tjur, T. (2009). Coefficients of determination in logistic regression models – A new proposal: The coefficient of discrimination. *The American Statistician*. 63, 366-372.
- Turk, Adam. (2006). The predictive nature of financial ratios. *The Park Place Economist*, XIV, 96-104.
- Wijesundera, A. A. V. I. D. A. S., Weerasinghe, T. P. C. R. Krishna., Gunawardena, M. M. D. & Peiris, H. R. I. (2015). Predictability of stock returns using financial ratios: Empirical evidence from Colombo stock exchange. *Kelaniya Journal of Management*. 4, (2).

APPENDIX :

R Codes used in applying Binary Logistic Regression technique

```
model<-glm(adpebin~dvpo4+tdmc4, family=binomial(link="logit"))
stepbinom<-step(model, direction="both")
model<-stepbinom
quantile((summary(model)$deviance.resid),c(0,0.25,0.5,0.75,1 ))
summary(model)$coefficients
summary(model)$dispersion
model$null.deviance
model$df.null
model$deviance
model$df.residual
model$aic
model$iter
anova(model,test='Chisq')
confint(model)
exp(cbind(odds_Ratio=coef(model), confint(model)))
caret::varImp(model)
lmtest::waldtest(model)
lmtest::lrtest(model)
BaylorEdPsych::PseudoR2(model)
hoslem<-ResourceSelection::hoslem.test(model$y, fitted(model), g=10)
pred<-predict(model, type="response")
roccurve<-pROC::roc(adpebin~pred)
plot(roccurve, print.auc=TRUE, print.auc.col="#000000", col="#4A148C", lty=1, lwd=1,
identity.col="#000000",identity.lty=1, identity.lwd=1, grid=TRUE, grid.col="#000000",
legacy.axes=TRUE)
```