

Project report on Factor Investing in Indian Equities

**Submitted towards partial fulfilment of the criteria
for award of PGPBABI by Great Lakes Institute of Management**

Submitted By

Group No. 8[Batch: October 2019, Bangalore]

Group Members

- 1. Shivkumar M**
- 2. Jithesh Janardhanan**
- 3. Vinod Kumar D**
- 4. Tarique Zaman**

Research Supervisor

Pranov Mishra

Great Lakes Institute of Management



ABSTRACT

Factor investing is a strategy that chooses securities on attributes that are associated with higher returns. A large amount of research has been conducted in factor-based investing models has been done globally. However, India remains nascent with this strategy of investment. Most of these existing approaches in India have focused on short term prediction using stocks' historical price and technical indicators. In this project, we prepared 10 years of annual financial data consisting of financial ratios cutting across macro-sectors such as Consumer, Financials, Manufacturing and Commodities. We investigated variety of machine learning algorithms: Logistic regression, Random Forest, Decision Trees & Linear Regression for prediction of growth category of firms and Earnings per share based on significant financial ratios. Model Performance measures like AUC, Sensitivity, Specificity, Accuracy were used to compare different models.

Our results show that Logistic Regression model achieves the best prediction results. Our findings demonstrate that machine learning models could be used to aid investors and analysts with decision making regarding to stock investment.

Key Words: Factor Based Investing, Logistic Regression, Random Forest, R, Tableau, Earnings Per Share, EPS

ACKNOWLEDGEMENT

We certify that the work done by us for conceptualizing and completing this project is original and authentic. We would like to express our sincere gratitude to our Research Advisor, Pranov Mishra for his guidance throughout the project.

We would also like to thank Md. Shoaib and Deepak Agarwal for their inputs and guidance on domains related to Indian Stock Market and Financial Risk Management. Last but not the least, we would like to express our gratitude to Maksud S, Vice-President ACE Equity for enabling the data collection process which is critical for this Data Science Project.

Date: 1st April 2020

Shivkumar M

Place: Bangalore

Jithesh Janardhanan

Vinod Kumar D

Tarique Zaman

CERTIFICATE OF COMPLETION

I hereby certify that the project titled ‘Factor Investing in Indian Equities’ was undertaken and completed under my supervision by Group No. 8 (Batch:October 2019, Bangalore) of Post Graduate Program in Business Analytics and Business Intelligence (PGPBABI).



Pranov Mishra

Date: 1st April 2021

Place: Bangalore

Table of Contents

Executive Summary	10
Chapter 1	12
Chapter 2	19
2.1 Banks and Financial Services	19
2.2 Consumer	37
2.3 Commodities	49
2.4 Manufacturing	66
Chapter 3: Final Recommendations and Conclusion	84
Bibliography	86
Annexure.....	87

LIST OF GRAPHS AND TABLES

Fig1 – High Level Overview of project

Fig2 – Model Summary

Fig3 – Factor Map

Table 1 - List of data points

Financial Services

Fig: FS1 - Adjusted EPS across firms

Fig: FS2- Adjusted EPS across Time Period

Fig: FS3- Adjusted Book Value across Firms

Fig: FS4 - Free Cash Flow per Share across Firms

Fig: FS5 -Adjusted DPS across Companies

Fig: FS6 - Return on Assets

Fig: FS7 - Return on Equity

Fig: FS8 - Return on Capital Employed

Fig: FS9 - Median / Average of Ratios across Time Period

Fig: FS10 - Median / Average of Ratios across Time Period

Fig: FS11- Random Forest Variable Importance Plot

Fig: FS12 - Logistic Regression FULL MODEL

Fig: FS13 - Model Output Based on Significant Variables

Fig: FS14 - Model Output Based on Significant Variables from Previous Model

Fig: FS15 - Stepwise Selection Model Output

Fig: FS16 - Variable Importance Plot

Fig: FS17 - Variable Importance Plot

Fig: FS18 - RPART Tree

Fig: FS19 - Performance Ratio of a company over years

Table 1 - List of data points

Table: FS1 - Significant Correlated Variables

Table: FS2 - Model Performance Measures of Linear/Ridge/Lasso/Elasticnet Models

Table: FS3 - Overall Model Evaluation Parameters

Table: FS4 – VIF

Table: FS5 - LR Summary

Table: FS6 - Tree Based Model Summary

Table: FS7 - Categories & Financial Ratios

Consumer

Fig: CN1 - Adjusted EPS Vs FCF/Adjusted BV/ FCF Per Share / DPS

Fig: CN2 - Adjusted EPS Vs P_CE/PATM/Dividend Yield/ Total Debt Mcap

Fig: CN3 - Logistic Regression FULL MODEL

Fig: CN4 - VIF values

Fig: CN5 - Model Output Based on Significant Variables

Fig: CN6 - Stepwise Selection Output using Blorr

Fig: CN7- Random Forest Model Output

Fig: CN8 - Random Forest Model with Top Variables Output
 Fig: CN9 - RPART Tree
 Table: CN1 - Significant Correlated Variables
 Table: CN2 - Model Performance Measures of Linear/Ridge/Lasso/Random Forest/SVM
 Table: CN3 - Variable Importance Table across Models
 Table: CN3 - Overall Model Evaluation Parameters
 Table: CN4 - VIF
 Table: CN5 - LR Summary
 Table: CN6 - Tree Based Model Summary
 Table: CN7 - Categories & Financial Ratios

Commodities

Fig: CM1- Median Ratios across time period
 Fig: CM2 - Median Adjusted EPS by firms
 Fig: CM3 - Median Adjusted EPS for the sector
 Fig: CM4 - Median Adjusted book value by firms
 Fig: CM5 - Median cash flow by firms
 Fig: CM6 - Median adjusted DPS by firms
 Fig: CM7 - Median ROA/ROE/ROCE by firms
 Fig: CM8 - Ratio trends against Adj. EPS
 Fig: CM9 - Correlation matrix
 Fig: CM10 - Variable Summary
 Fig: CM11 - Logistic Regression - full model
 Fig: CM12 - Model output - Logistic Rregression 2
 Fig: CM13 - Model output - Random Forest
 Fig: CM14 - rPart - Tree
 Fig: CM15 - Stage 3 variable distribution
 Fig: CM16 - Stage 3 variable Correlation
 Fig: CM17 - Stage 3 Linear Regression model output
 Fig: CM18 - Stage 3 Logistic Regression model output
 Fig: CM19 - Shree Cement trend of variables
 Table: CM1 - Model Performance Measures of
 Linear/Ridge/Lasso/Random Forest/SVM
 Table: CM2 - VIF - full model
 Table: CM3 - Logistic Regression- Model
 measures
 Table: CM4 -Random Forest- Model measures
 Table: CM5 - Tree based models - Model measures
 Table: CM6 - Stage 3 variable list
 Table: CM7 - Categories & Financial Ratios

Manufacturing

Fig: M1 - Univariate analysis of the data
 Fig: M2.1, 2.2, 2.3, 2.4 - Plots of Adjusted EPS by industry and company
 Fig: 3.1,3.2,3.3- EDA of various variables
 Fig: M7 - Logistic Regression based on all variables
 Fig: M9 - Logistic Regression post removing all multicollinear variables
 Fig: M10 - Classification tree
 Fig: M11 - Random Forest
 Fig: M12 - Boruta feature selection algorithm

- Fig: M13 - Variable selection for stage 3
Fig: M14 - Linear Regression model for predicting Adjusted EPS.
Fig: M16 - Correlation plot on the transposed data
Fig: M17 - Logistic regression on full data
Fig: M18 - Logistic Regression without multicollinear variables
Fig: M19 - Classification on transposed data
Fig: M20 - Random forest
Fig M21 - Summary of model metrics
Table Fig:M4 - Correlation analysis of variables
Table Fig: M5 - comparison of model metrics from Stage 1
Table Fig: M6 - Model metrics from the data to be modeled - Stage 2
Table Fig: M8 - Multicollinearity
Table Fig: M15 - Sample of data post building more features

ABBREVIATIONS

VR	Valuation Ratio
FR	Financial Ratio
FCF	Free Cash Flow
CF_Per_share	Cash Flow Per Share
BV	Book Value
EV	Enterprise Value
PAT	Profit After Taxes
ROCE	Return on Capital Employed
CE	Capital Employed
ROE	Return on Equity
ROA	Return on Assets
EPS	Earnings Per Share
PE	Price to Earnings
PB	Price to Book
PBIDTM	Profit before Interest Depreciation and Tax Margin
EBITM	Earning Before Interests Taxes and Management
PATM	Profit After Taxes Margin
DPS	Dividend Per Share

Executive Summary

Factor investing is a strategy that chooses securities on attributes that are associated with higher returns. The project titled “Factor Based Investing in Indian Equities” is aimed at driving the decision to invest in Indian Equities leveraging industry standard factors such as Value, Momentum, Size and Quality. However, Factor-Based Investment strategy does not take a holistic view as it does not consider the underlying financial ratios impact on the performance of a stock.

Earnings per Share is considered as an Industry Standard parameter to measure the performance of a company across a time-period as well as enabling comparison amongst companies.

This project is aimed at identifying the significant financial ratios that impacts company's Earnings per share and predicting the growth category to which the company belongs.

The key steps carried out to execute the project were: -

- Gathering data on primarily financial ratios related to Nifty 500 Companies
- Organize the data based on Macro-Sectors: Consumer, Banks and Financial Services, Commodities and Manufacturing
- Categorizing firms as Growth and Non-Growth firms based on the Annual Adjusted Earnings per Share
 - All Firms above the median Adjusted Earnings Per Share have been considered as Growth (1) and the ones below have been considered as Non-Growth (0)
 - Lagging the dataset by 1 Year, for example, the Growth Category of 2019 becomes the Y-Variable to predict for the Year 2018
- Leveraging Tableau and R to carry out exploratory data analysis
- Applying Statistical Techniques and Tools such as Logistical Regression, Tree Models and Linear Regression
 - Logistic Regression and Tree-Based Models such as Random Forest and RPART has been applied to predict the category to which the firm belongs to i.e. Growth or Non-Growth
 - Linear Regression has been applied to predict the impact of significant financial ratios on the Earnings Per Share
- The dataset is transposed with Company names as rows and columns having the significant variables from the previous step
 - The Significant variables derived from the Logistic Regression and Tree-Based models are then converted to Slope and Average. Slope is calculated based on the values of the significant variables from 2009-2018. Average over the last 3 Years has been considered from 2016-2018.
 - Logistic Regression model is applied on the Slope, Average and 2018 Variable Value to predict the growth category in 2019

For each Macro-Sector, the above steps were carried out to measure the impact of the variables on the Adjusted EPS. The key model performance measures used in Logistic Regression and Tree-Based Models are AUC, Sensitivity and Specificity and these were used to compare multiple models and arrive at the most optimum set of significant variables in the form of ratios.

The significant ratios were then tied back to the factors Value, Size, Momentum, Volatility and Quality.

There were certain common ratios driving the factors across Macro-Sectors and there were certain Macro-Sector specific ratios as well driving the factors. The report provides detailed breakdown of every Macro-Sector in terms of the exploratory data analysis, statistical techniques applied and recommendations around the key ratios that drive the Adjusted Earnings per Share and the industry standard investment factors.

Following graphic depicts the structure of the report and approach of the project: -

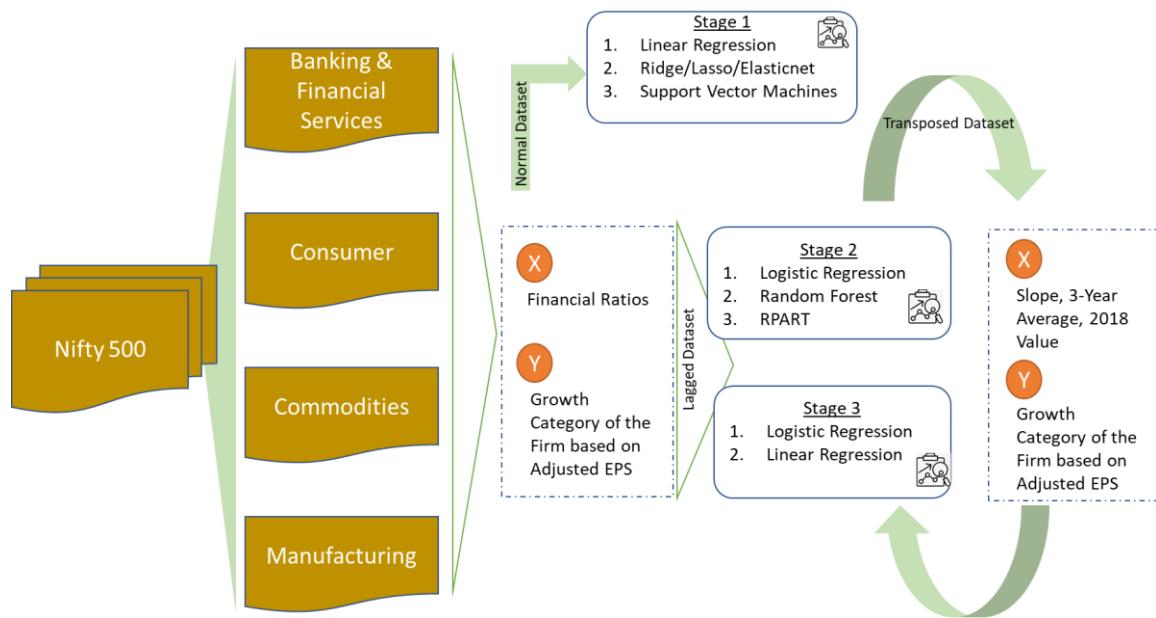


Fig1 – High level Overview of project

Chapter 1

Introduction

The project is an attempt to enhance the Factor-Based Investment Model. Factor investing is a strategy that chooses securities on attributes that are associated with higher returns.

The concept of using historical data and identifying factors that influence stock price or returns has been in the forefront of research to predict stock prices and maximize returns. Traditionally either fundamental or technical analysis were used to pick stocks but there was a need for a scalable and integrated approach to pick a set of stocks for maximum returns. It started with the Capital Pricing Asset Model (CAPM) and later the Arbitrage pricing theory which basically established the linear relationship to different indices and factors. The Fama-French 3 factor model expanded on this to initially prove that a combination of small cap, a high book to market ratio and risk-free returns over time, yields better returns. Two other factors (Momentum and Quality) were later added to explain the variation better. A lot of research and academic work has already been performed to predict stock prices / returns, however, the financial products in the market today limited to sector or style based.

The industry standard factors are Value, Size, Momentum, Volatility and Quality. However, Factor-Based Investment strategy does not take a holistic view, i.e. all the factors and the underlying variables and its impact on the Earnings per Share of a stock. It is this gap that the project is looking to address.

The main objective is to build a model which combines various factors and its underlying variables to predict the Earnings Per Share on an individual stock. Initially the study is limited to nifty stocks - National Stock Exchange (NSE) (top 500 stocks).

Objectives

Following are the 5 Steps to achieve the defined objectives: -

Step 1 – Collection of data: Collecting historical data on Nifty stocks:

Data points related to the Company Financial Results (historical financial results) collected are:

-

Free Cash flow	Price to Cash Flow Ratio	EBITM (%)	Payable days
Adjusted PE (x)	Free Cash Flow per Share	Pre Tax Margin(%)	Net Sales Growth(%)
PCE(x)	Price to Free Cash Flow	PATM (%)	Core EBITDA Growth(%)
Price / Book Value(x)	Free Cash Flow Yield	CPM(%)	EBIT Growth(%)
Dividend Yield(%)	Sales to cash flow ratios	ROA (%)	PAT Growth(%)
EV/Net Sales(x)	Earnings Per Share (Rs)	ROE (%)	Adj. EPS Growth(%)

EV/EBITDA(x)	Adjusted EPS (Rs.)	ROCE (%)	Total Debt/Equity(x)
EV/EBIT(x)	CEPS(Rs)	Asset Turnover(x)	Current Ratio(x)
EV/CE(x)	DPS(Rs)	Inventory Turnover(x)	Quick Ratio(x)
M Cap / Sales	Adj DPS(Rs)	Debtors Turnover(x)	Interest Cover(x)
High PE	Book Value (Rs)	Fixed Asset Turnover (x)	Total Debt/Mcap(x)
Low PE	Adjusted Book Value (Rs)	Sales(x)/Working Capital	
Net Sales	Tax Rate(%)	Fixed Capital/Sales(x)	
Profit After Tax	Dividend Pay Out Ratio(%)	Receivable days	
Cash Flow Per share	PBIDTM (%)	Inventory Days	

Table 1 - List of data points

[Refer to Data Dictionary in Annexure](#)

Step 2 – Data Pre-processing step

1. Perform a preliminary EDA to: -
 - Check for Outliers
 - Missing Data Check
 - Check for Correlation between predictor variables & between target variables and predictor variables.
2. Calculating the Growth Category of a firm based on the Growth in the Adjusted Earnings per Share of a firm.
3. Lagging the dataset by 1 Year
4. The dataset consists of 9 Macro Economic Sectors and the analysis and statistical techniques have been applied based on the following Macro-Groupings
 - a. Consumer Firms
 - b. Banking and Financial Services
 - c. Commodities
 - d. Manufacturing

Step 3 – Applying Statistical Techniques

Following Statistical tools and techniques have been used: -

- Tableau for EDA
- Modelling Techniques
 - Logistical Regression
 - Random Forest
 - RPART
 - Linear Regression

Data Sources

The data used for this study is of top 500 companies in National Stock Exchange of India (NSE). Nifty Index acts as one of the barometers for Indian Stock market health. The data used will be financial results of Nifty500 Companies.

The data used for this analysis is primarily sourced from [AceEquity](#) with help from one of the executives of company who is supporting us for this academic project. AceEquity is one of the top vendors in India for sourcing data intended for purposes of Wealth management, Portfolio Management, Investment Banking, Institutional Stock broking, Management consulting and Business Advisory.

As stated earlier that the data sourced comprises of financial results of Nifty500 companies.

Let us delve a little deeper into the data:

Nifty500 Company Financial Results data:

Key Variable definitions:

1. Free Cash Flow (FCF): this is cash that a company generates after accounting for outflows to support its operations. FCF is generally a measure of profitability and often a supplemental tool for analysis
Free Cash Flow per Share (FCF_Per_Share) is the FCF divided by the total number of shares outstanding
2. Cash Flow Per Share (CF_Per_share): is the after-tax earnings plus depreciation on a per share basis which serves as a measure of firm's financial strength
3. Book Value: generally, refers to a company's total assets minus its outstanding liabilities. It represents the total amount of equity worth to its shareholders after liquidating all its tangible assets and paying off all its liabilities. Its useful in determining the value of a company and is often expressed as book value per share
4. Adjusted Book Value – is a measure of the company's valuation after liabilities including off-balance sheet liabilities and assets adjusted to reflect its fair market value.

Generally, both Book Value and Adjusted book value are very highly correlated.

5. Enterprise Value (EV): is a measure of a company's total value it's often used as an alternative to market capitalization. EV includes market capitalization of a company but also short term and long-term debt as well as any cash on balance sheet. It's generally used for valuing a company for potential takeover.
6. EV/CE(%) (VR_CE): is a financial ratio between Enterprise Values and Capital employed expressed as a percentage.
7. Netsales: is the Gross Sales minus returns, allowances & discounts. Changes in Netsales do affect the gross profit of a company
8. Profit After Taxes (PAT): is the earning of a company after deducting all its taxes. PAT is watched closely by investors to see profit generating ability of the company
9. Return on Capital Employed (ROCE): this is a financial ratio that measures a company's profitability & efficiency with which the capital is used. A higher ROCE means the company has efficiently utilized its capital. Investors generally prefer companies with stable and gradually rising ROCE

$$\text{ROCE} = \frac{\text{EBIT}}{\text{Capital Employed}}$$

Where: EBIT is Earnings before Interest & Taxes

Capital Employed = Total Assets – Current Liabilities

10. Return on Equity (ROE): is a measure of financial performance & expressed as a percentage. It can be calculated by dividing net income by Shareholders equity. It is an important measure to see how effectively the company management is using company's assets to create profits.

$$\text{ROE} = \frac{\text{Net Income}}{\text{Average Shareholders' Equity}}$$

11. Return on Assets (ROA): is an indicator of how profitable a company is relative to its total assets. This gives an idea to investors how efficient is a company in using its assets to generate earnings. It's a useful ratio to compare similar companies or comparing a company to its earlier performance. Higher ROA is an indication of assets efficiency.

$$\text{ROA} = \frac{\text{Net Income}}{\text{Total Assets}}$$

12. Earnings Per Share (EPS): is an important measure which reflect how much money does a company make for each share of its stock. Its calculated by company's net profit divided by outstanding shares of its common stockholder.

$$\text{EPS} = \frac{\text{Net Profit}}{\text{End of Period Common Shares Outstanding}}$$

13. Debt Equity: is a ratio of company total liabilities divided by its shareholder equity. A higher debt to Equity ratio for a company generally indicates a risk to shareholders.

However, comparing Debt Equity ratios across industries is not ideal since they may vary by industry.

$$\text{Debt Equity} = \frac{\text{Total Liabilities}}{\text{Total Shareholders' Equity}}$$

14. Dividend Yield: is always expressed as a percentage. It's a financial ratio – Dividend by Price, essentially how much a company pays out in dividends each year relative to its stock price. Mature companies are most likely to pay dividends.

$$\text{Dividend Yield} = \frac{\text{Annual Dividends per Share}}{\text{Price per Share}}$$

15. Price to Earnings Ratio (PE): as the name suggests it is a ratio company share price to its earning per share. A **high PE** ratio could mean overvaluation or high growth expected by investors. A company with no PE ratio would mean that they have no earnings.

$$\text{PE Ratio} = \frac{\text{Market Value Per Share}}{\text{Earnings per Share}}$$

16. Adjusted PE: also known as CAPE ratio is a valuation measure that uses real earning per share (EPS) over a 10-year period to smooth out the fluctuations in corporate profits that occur over different periods of business cycle. This ratio is generally used to broad equity indices to whether market is undervalued or overvalued.

$$\text{Adjusted PE} = \frac{\text{Price}}{\text{Average earnings for 10 years adjusted for inflation}}$$

17. Price to Book Ratio (PB): it is a ratio company's market capitalization to its book value.

Calculated by dividing company stocks value per share by its book value per share. Typically, the market value of an equity is higher than the book value. PB ratios under 1 are considered as solid investments.

$$\text{PB Ratio} = \frac{\text{Market Value Per Share}}{\text{Book Value Per Share}}$$

18. Profit before Interest Depreciation and Tax Margin (PBIDTM(%)) : is a financial ratio which is calculated as Adjusted Gross profit + Interest by sales ratio, and then expressed as %

$$\text{PBIDTM(%) = } \frac{\text{Adjusted Gross Profit+ Interest}}{\text{Sales}} * 100$$

19. Earnings Before Interest Taxes and Management (EBITM (%)) : is a financial ratio which is calculated by revenue before interest and taxes and management expenses

20. Profit After Taxes Margin (PATM(%)): this is a profit margin ratio which shows the percentage of net sales that remains after deducting cost of goods sold and all other expenses including taxes.
21. Dividend Per Share (DPS): is sum of declared dividends issued to every share outstanding. This is an important metric to investors as it translates to direct income for shareholder.
22. Debtors Turn Over Ratio: is an accounting measure to quantify a company's effectiveness in collecting its receivables. A high debtor turnover ratio may indicate that a company's collection of accounts receivable is efficient and that the company has a high proportion of quality customers that pay their debts quickly
23. Asset Turn Over: is the ratio of total sales or revenue to average assets. This is used to understand how effectively companies are using their assets to generate sales. It is used by compare similar companies in the same sector or group.
24. Debt to Market Capitalization: Measurement of a company's financial leverage, calculated by taking the company's interest-bearing debt and dividing it by total capital. A higher debt-to-capital ratio means the riskier the company.

$$\text{Debt to Market Capitalization} = \frac{\text{Debt}}{\text{Debt} + \text{Shareholders Equity}}$$
25. Sales to Working Capital: a liquidity and activity ratio indicating the amount of money from sales, generated by a dollar of working capital investment. In other words, a ratio measuring the efficiency of company's working capital utilization in order to generate the certain level of sales.
26. Momentum: is the speed/velocity of price changes in a stock/security. Momentum shows the rate of change in price movement over a period of time to help investors determine the strength of a trend. Stocks that tend to move with the strength of momentum are called momentum stocks.

Data structure details

Following is the structure of the data collected: -

- Number of records – 4234 rows, 60 columns
- Number of Companies – 402
- Number of Industry – 37
- Number of Macroeconomic – 9
- Years – 10 Years Data (2009 – 2019)
- Missing Values - There are no missing values present

Data Preparation

The Dataset was re-structured and new variable is introduced based on the Adjusted EPS Growth. Earnings per share (EPS) is calculated as a company's profit divided by the outstanding shares of its common stock. The profits are adjusted to take into account extraordinary items to report Adjusted EPS. For example, the profits from the sale of a building needs to be deducted from the total profits or the losses from a fire in the same building needs to be included. Adjusted Earnings Per Share is a key measure of the profitability of a company and can also be used to compare the performance of firms along with comparison across accounting periods.

The Adjusted EPS for the next year has been considered as the predictor for the current year. For example, the 2019 Adjusted EPS value becomes the value to predict for the year 2018 based on the dependent variable. This new variable has been named as “Target_Growth_Lag”.

Chapter 2

2.1 Banks and Financial Services

Data Summary

Dimension – 583(rows) x 59 (variables) (including Target variable)

53 companies in the data set spread across 2 industry for Financial Services Sector

Exploratory Data Analysis

Adjusted EPS across Firms Belonging to Financial Services Macro-Economic Sector

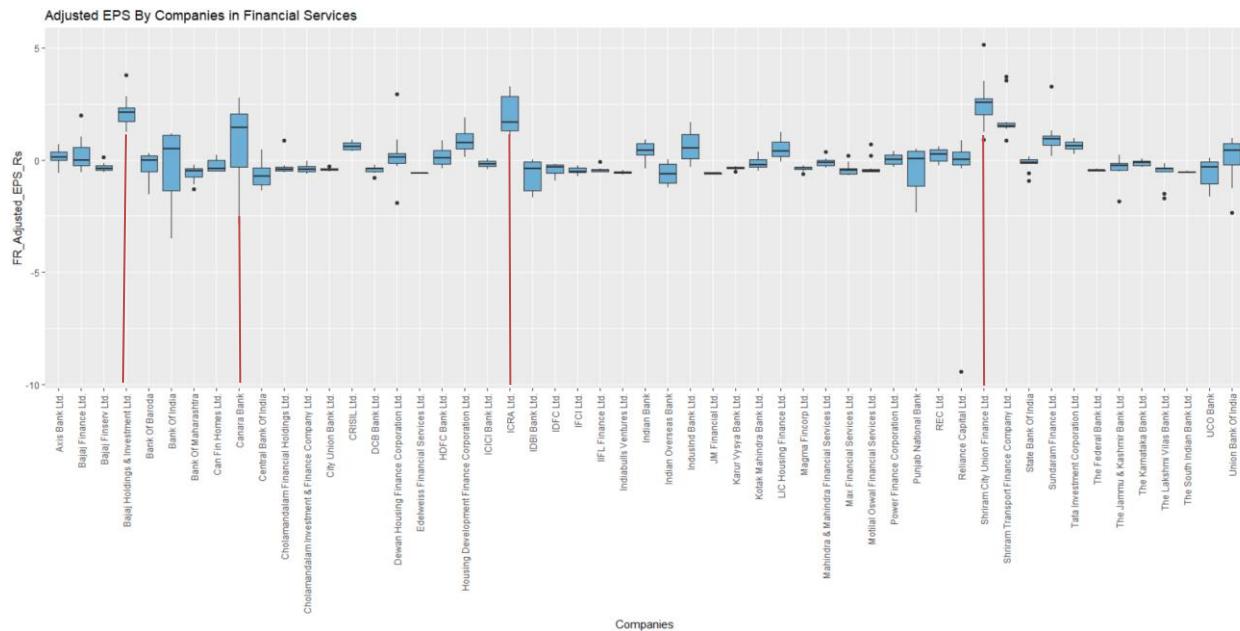


Fig: FS1 - Adjusted EPS across firms

The following Companies belonging to Financial Services Macro Sector have a higher Adjusted EPS compared to the rest. The number of outliers with respect to a single firm is on the lower side.

- Shriram City Union Finance LTD.
- ICRA
- Canara Bank
- Bajaj Holding and Investment LTD.

Rest of the companies have a median Adjusted EPS close to the average.

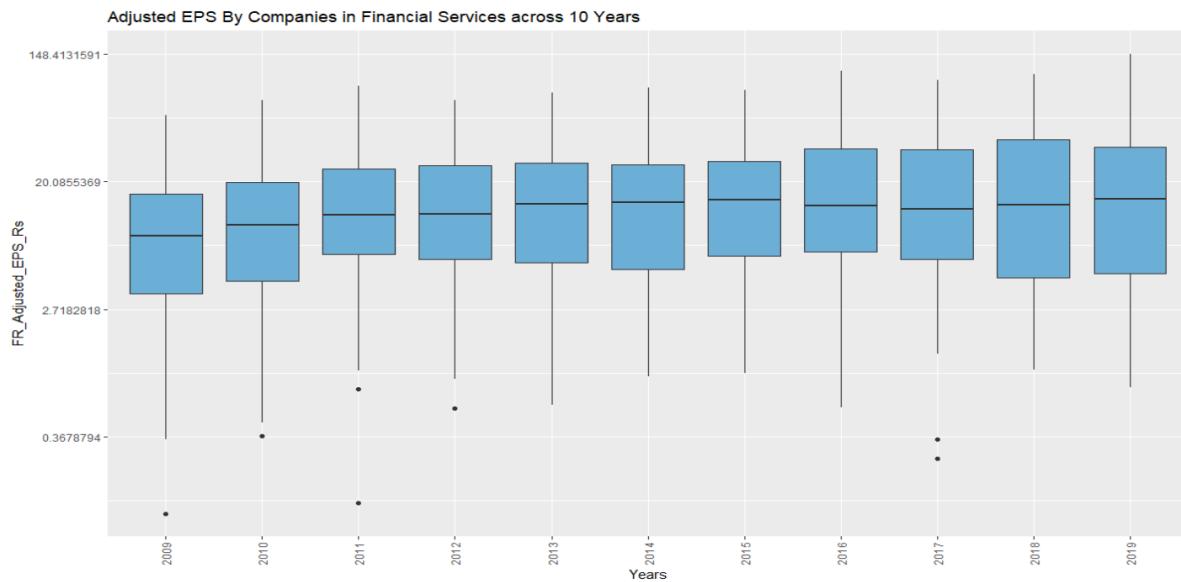


Fig: FS2- Adjusted EPS across Time Period

The Adjusted EPS across Companies belonging to the Financial Services Industry increased in the first two years and has stabilized over the last 8 years.

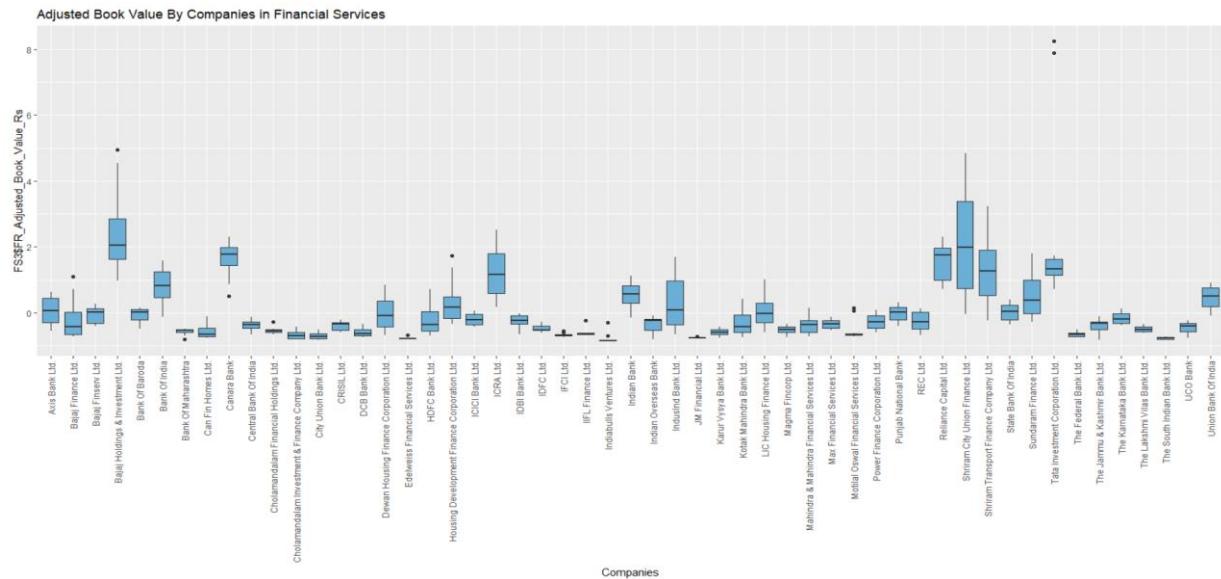


Fig: FS3- Adjusted Book Value across Firms

The Adjusted Book Value is higher for the following companies belonging to the Financial Services Macro Sector as compared to the rest: -

- Shriram City Union Finance LTD.
- ICRA
- Canara Bank
- Bajaj Holding and Investment LTD.
- Reliance Capital

Adjusted book value looks at the value of a company in terms of the current market values of its assets and liabilities. Specifically, book value concerns the total value of company assets minus the total value of company liabilities.

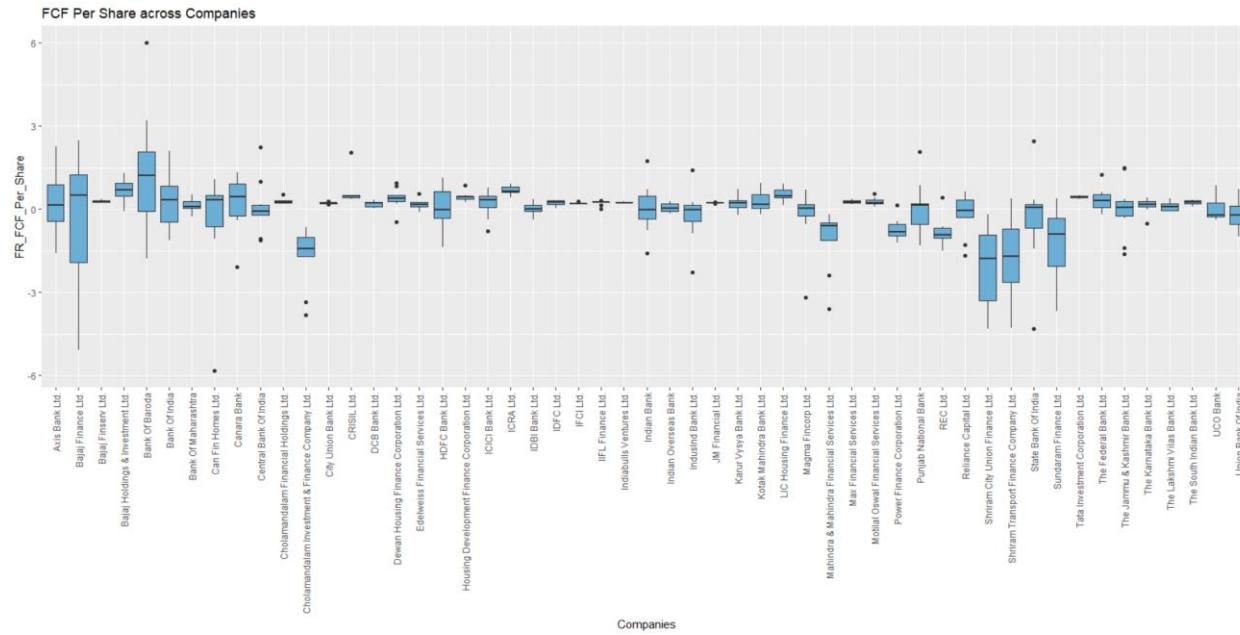


Fig: FS4 - Free Cash Flow per Share across Firms

Free cash flow per share (FCF) is a measure of a company's financial flexibility that is determined by dividing free cash flow by the total number of shares outstanding. Free Cash Flow per share is low for the following: -

- Shriram City Union Finance LTD
 - Shriram Transport Finance LTD.
 - Sundaram Finance LTD.

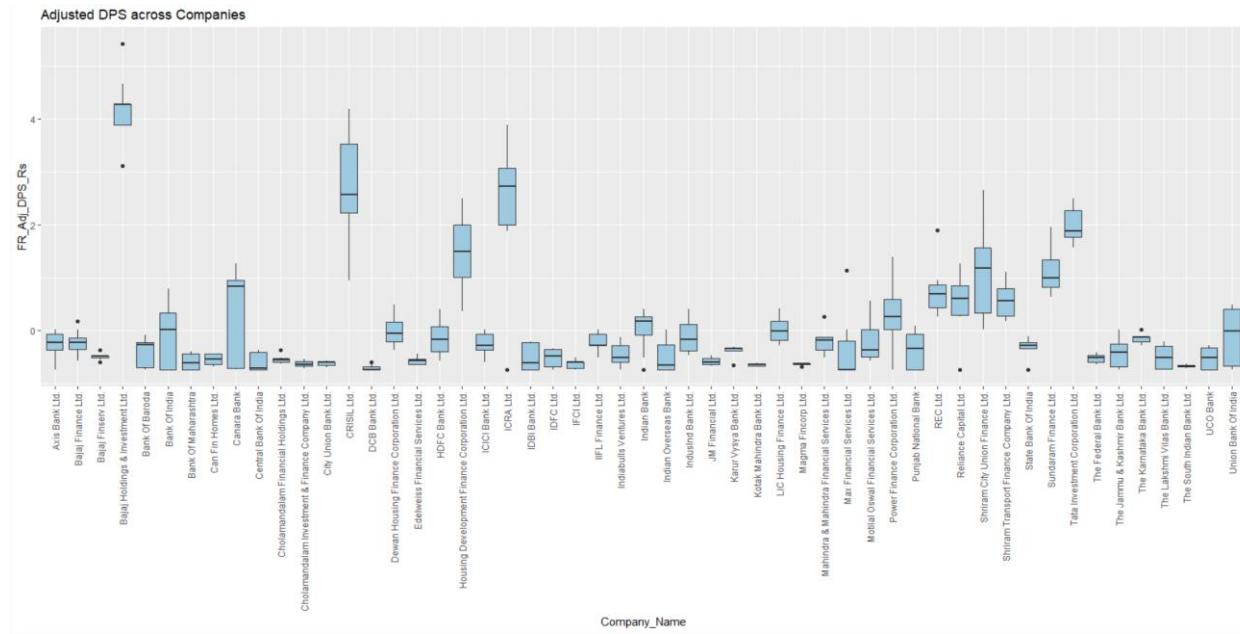


Fig: FS5 -Adjusted DPS across Companies

Adjusted DPS is high for the following: -

- Bajaj Holdings and Investments
- CRISIL
- ICRA
- HDFC
- Tata Investment Corporation
- Shriram City Union

The Return on Assets Return on Capital Employed and Return on Equity are correlated with each other. For example, CRISIL has a high ROA, ROE and ROCE.

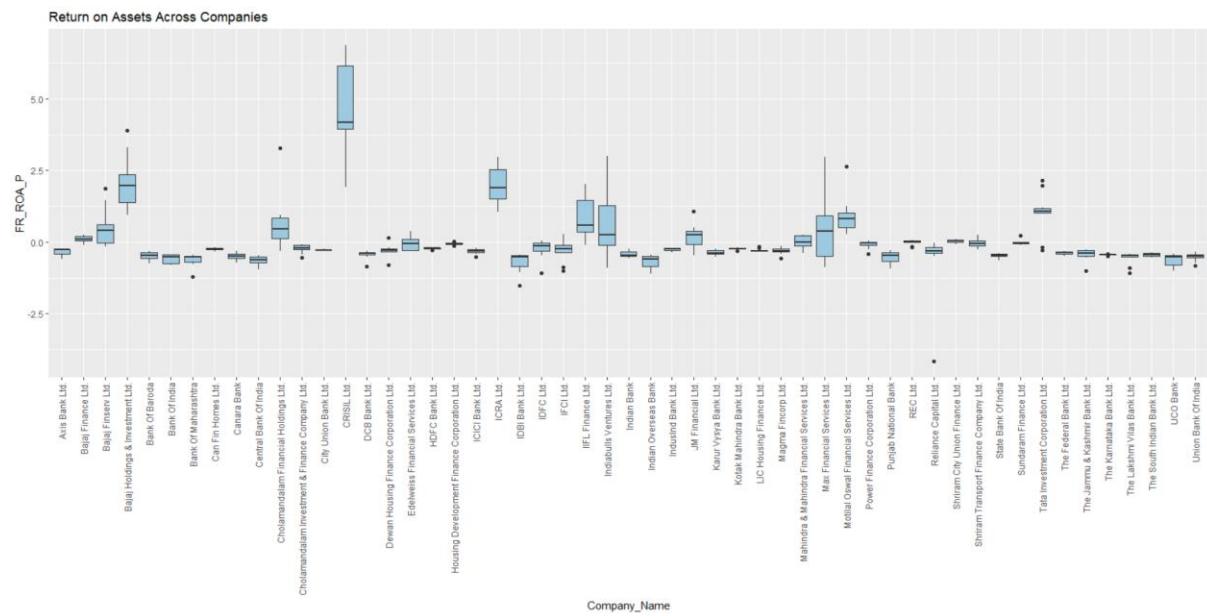


Fig: FS6 - Return on Assets

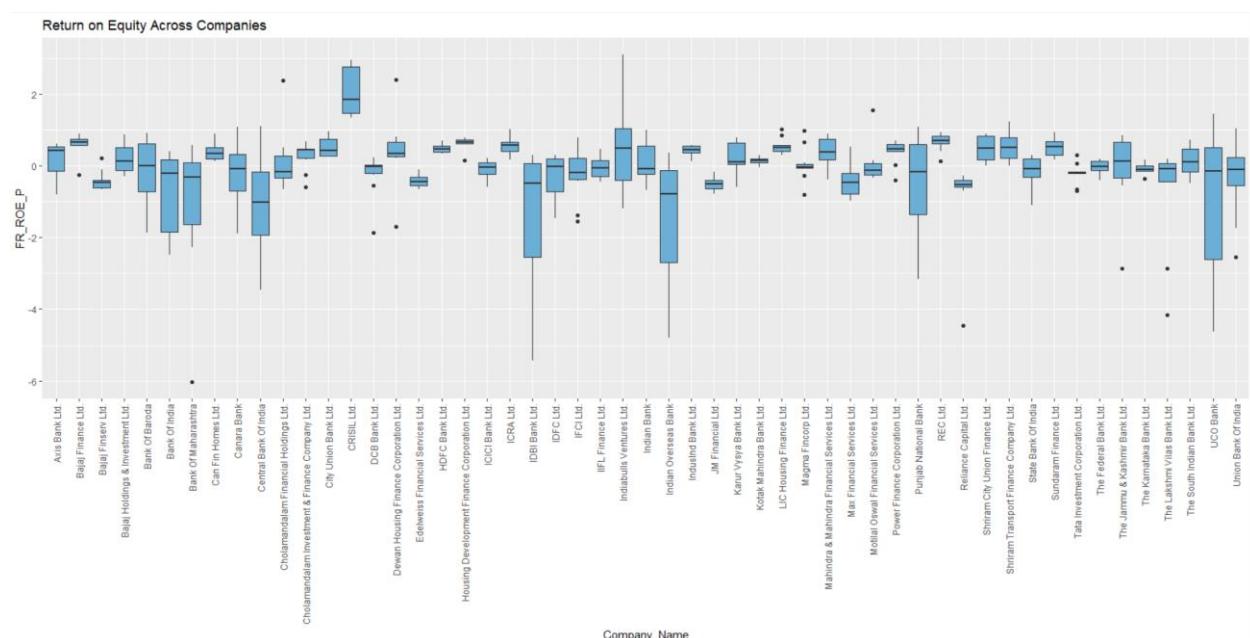


Fig: FS7 - Return on Equity

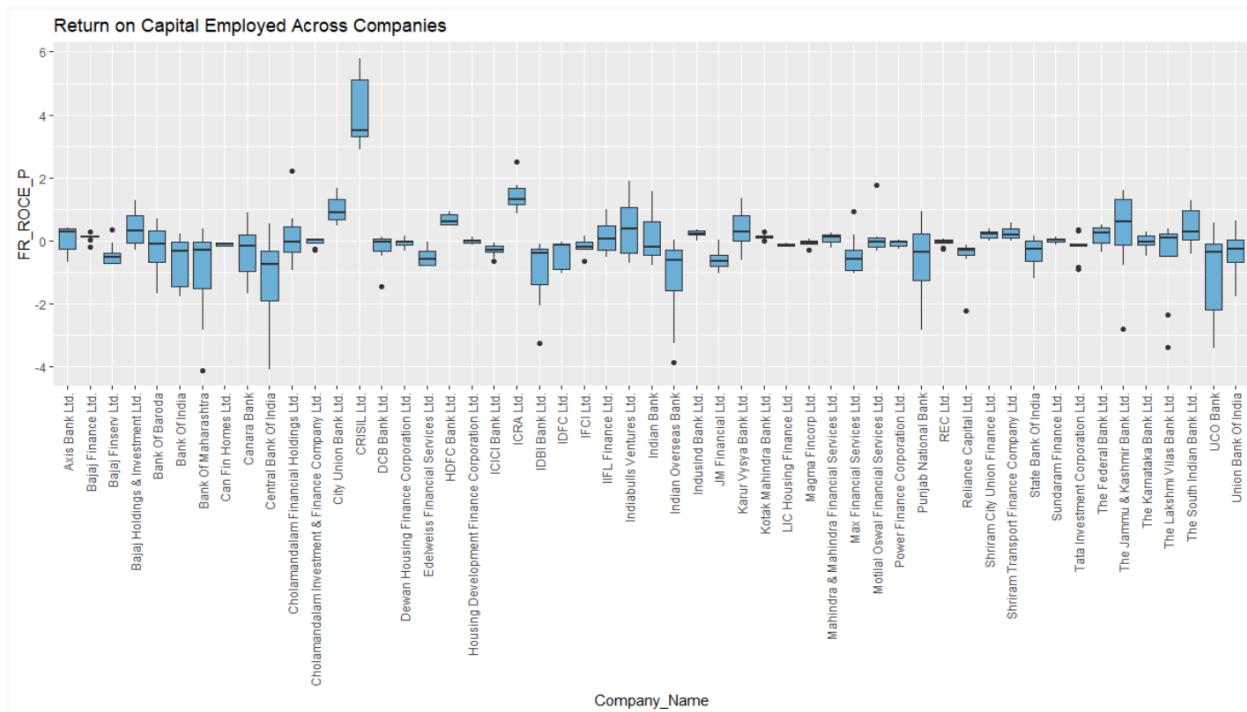


Fig: FS8 - Return on Capital Employed

Following charts give a view into the performance of banks and financial services firms based on different financial ratios categorized by Growth classification of the company. The charts indicate correlation between: -

- Return on Assets, Return on Capital Employed and Return on Equity
- Price to Free Cash Flow, Free Cash Flow Yield
- Dividend Yield, Dividend Payout Ratio

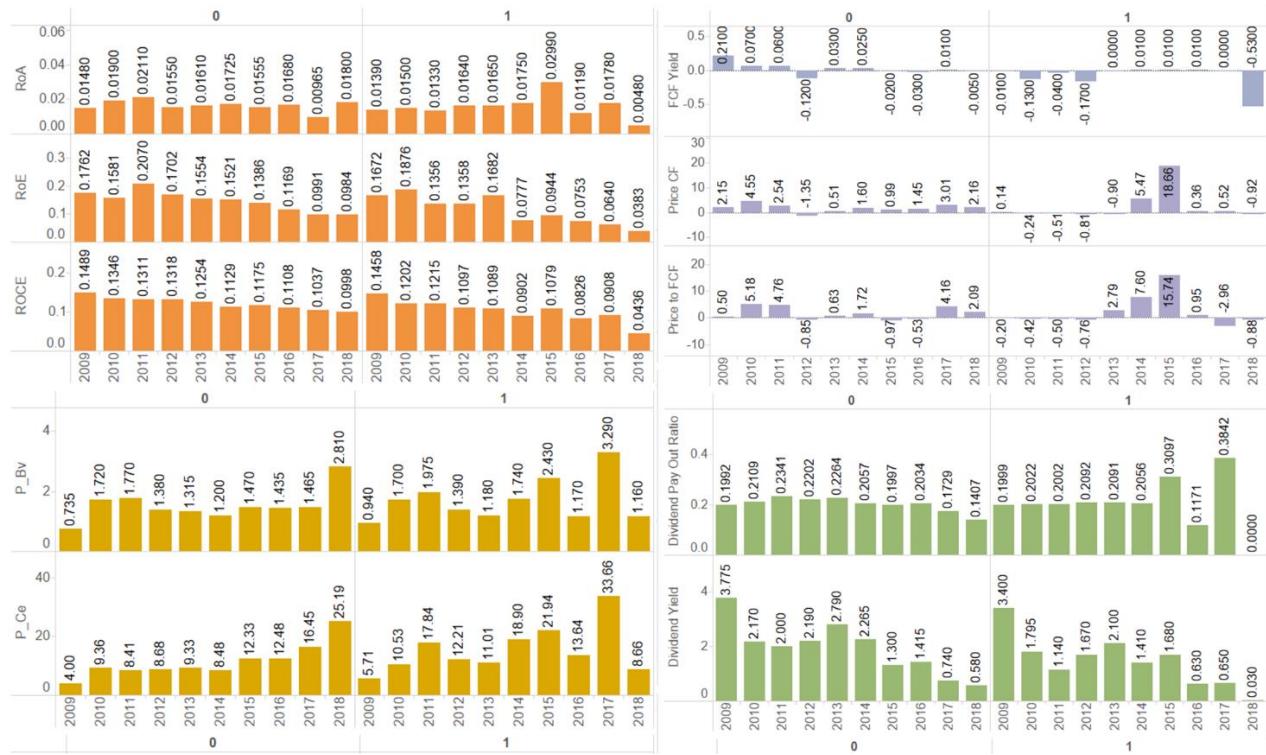


Fig: FS9 - Median / Average of Ratios across Time Period

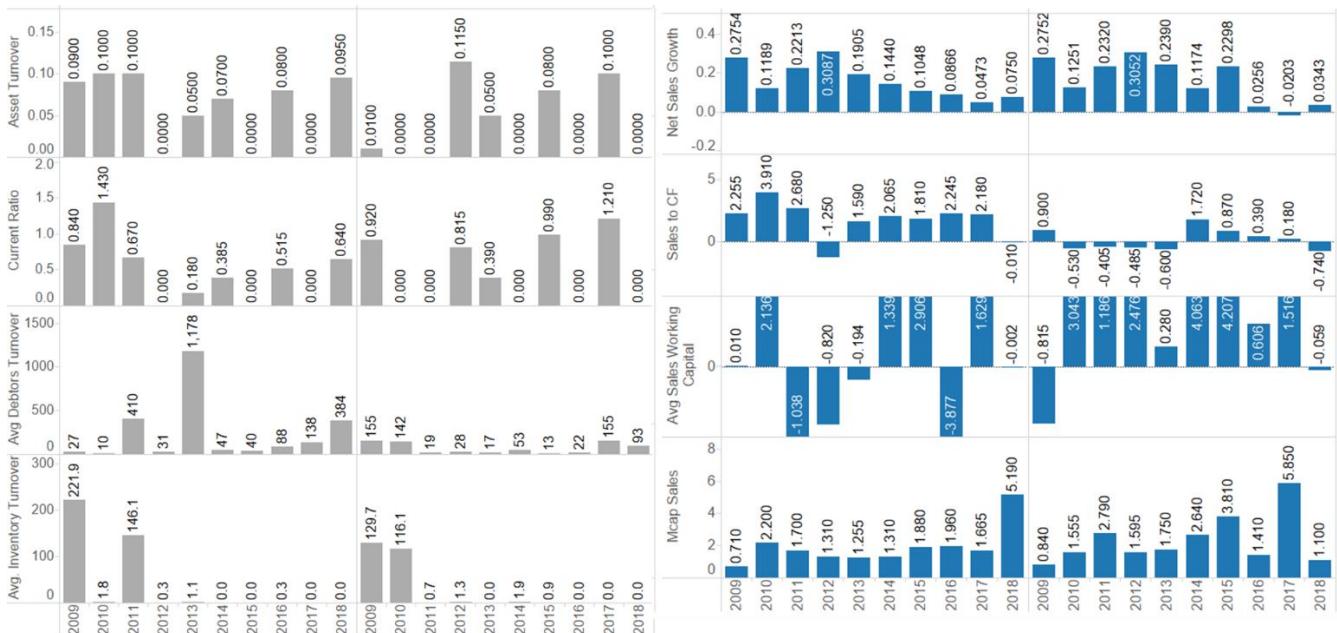


Fig: FS10 - Median / Average of Ratios across Time Period

Correlation Between Variables

Following table provides the correlation between select variables and its p-values to determine the significance of the correlation: -

First Variable	Second Variable	Correlation value	p value
VR_Adjusted_PE_	VR_P_CE	0.70963	0
VR_Adjusted_PE_	VR_P_BV	0.56507	0
VR_Adjusted_PE_	VR_Mcap_Sales	0.80765	0
VR_Adjusted_PE_	VR_High_PE	0.73010	0
VR_Adjusted_PE_	VR_Low_PE	0.58005	0
VR_Adjusted_PE_	FR_Price_CF_Ratio	0.56126	0
VR_Adjusted_PE_	NDP_PE	0.60538	
VR_Adjusted_PE_	NDP_Price_BV	0.58945	0
VR_P_CE	VR_Mcap_Sales	0.68842	0
VR_P_CE	VR_High_PE	0.62724	0
VR_P_CE	FR_Price_CF_Ratio	0.51432	0
VR_P_BV	VR_Mcap_Sales	0.57032	0
VR_P_BV	VR_Low_PE	0.55976	0
VR_P_BV	FR_ROA_P	0.56286	0
VR_P_BV	FR_Total_Debt_Mcap	-0.25912	1.404E-09
VR_P_BV	NDP_Price_BV	0.95197	0
VR_Mcap_Sales	VR_High_PE	0.71050	0
VR_Mcap_Sales	VR_Low_PE	0.55312	0
VR_Mcap_Sales	FR_Price_CF_Ratio	0.53145	0
VR_Mcap_Sales	NDP_Price_BV	0.59394	0
VR_High_PE	VR_Low_PE	0.52841	0
VR_High_PE	FR_Price_CF_Ratio	0.54154	0
VR_High_PE	NDP_Price_BV	0.51790	0
VR_Low_PE	FR_Total_Debt_Mcap	-0.25609	2.206E-09
VR_Low_PE	NDP_Price_BV	0.57535	0
FR_Price_CF_Ratio	FR_Price_to_FCF	0.57990	0
FR_PBIDTM_P	FR_EBITM_P	0.75180	0
FR_PBIDTM_P	FR_Pre_Tax_Margin_P	0.83258	0
FR_PBIDTM_P	FR_PATM_P	0.83100	0
FR_PBIDTM_P	FR_CPM_P	0.82725	0
FR_EBITM_P	FR_Pre_Tax_Margin_P	0.79067	0
FR_EBITM_P	FR_PATM_P	0.80213	0
FR_EBITM_P	FR_CPM_P	0.79310	0
FR_EBITM_P	FR_Asset_Turnover	-0.33301	3.432E-15
FR_Pre_Tax_Margin_P	FR_PATM_P	0.99129	0
FR_Pre_Tax_Margin_P	FR_CPM_P	0.99085	0
FR_PATM_P	FR_CPM_P	0.99890	0
FR_ROA_P	FR_ROE_P	0.50270	0
FR_ROA_P	FR_ROCE_P	0.69631	0
FR_ROA_P	FR_Asset_Turnover	0.63994	0
FR_ROA_P	NDP_Price_BV	0.51895	0
FR_ROE_P	FR_ROCE_P	0.83113	0
FR_Core_EBITDA_Growth_P	FR_EBIT_Growth_P	0.74223	0
FR_Total_Debt_Equity	FR_Total_Debt_Mcap	0.69160	0
FR_Current_Ratio	FR_Quick_Ratio	0.99994	0
NDP_PE	NDP_Price_BV	0.50347	0

Table: FS1 - Significant Correlated Variables

Statistical Tools and Techniques

The analysis and modelling were divided into three stages as a part of the project: -

- Stage 1: Statistical Analysis of the variables and its impact on the Adjusted EPS for the current Year
- Stage 2: Statistical analysis of the Variables and its role in PREDICTING the Growth Category of a firm for the next year
- Stage 3: Based on the significant variables derived from the Stage 2, determining the impact of the time, leveraging slope, 3-Year average and current year's data on the Growth of the firm for the next year

Stage 1: Statistical Analysis of the variables and its impact on the Adjusted EPS for the current Year

Following statistical techniques were leveraged for Stage 1: -

- Linear Regression
- Lasso
- Ridge
- Elasticnet
- Random Forest
- Support Vector Machine

Linear Regression/Lasso/Ridge/Elasticnet

The Significant Variables that have an impact on the Adjusted Earnings Per Share, based on different models and iterations run are as follows:

- Asset Turnover Ratio - measures the value of a company's sales or revenues relative to the value of its assets
- Return on Equity - measure of the profitability of a business in relation to the equity
- Free Cash Flow per Share - measure of a company's financial flexibility that is determined by dividing free cash flow by the total number of shares outstanding
- Adjusted Dividend Per Share - dividing the total dividends paid out by a business, including interim dividends, over a period of time by the number of outstanding ordinary shares issued
- Adjusted Book Value - measure of a company's valuation after liabilities—including off-balance sheet liabilities—and assets adjusted to reflect true fair market value
- Profit After Tax

Model Performance Measures

Model Types	RMSE
Full Model	0.044747
Model#2	1.003964
Model#3	1.002625
Model#4	0.804857
Lasso	1.000953
Ridge	0.9666798

Elasticnet	1.000573
------------	----------

Table: FS2 - Model Performance Measures of Linear/Ridge/Lasso/Elasticnet Models

Random Forest

Random Forest Model has been run on the variables and following is the output based on 50 Trees and mtry value of 8: -

Variable Importance Plot

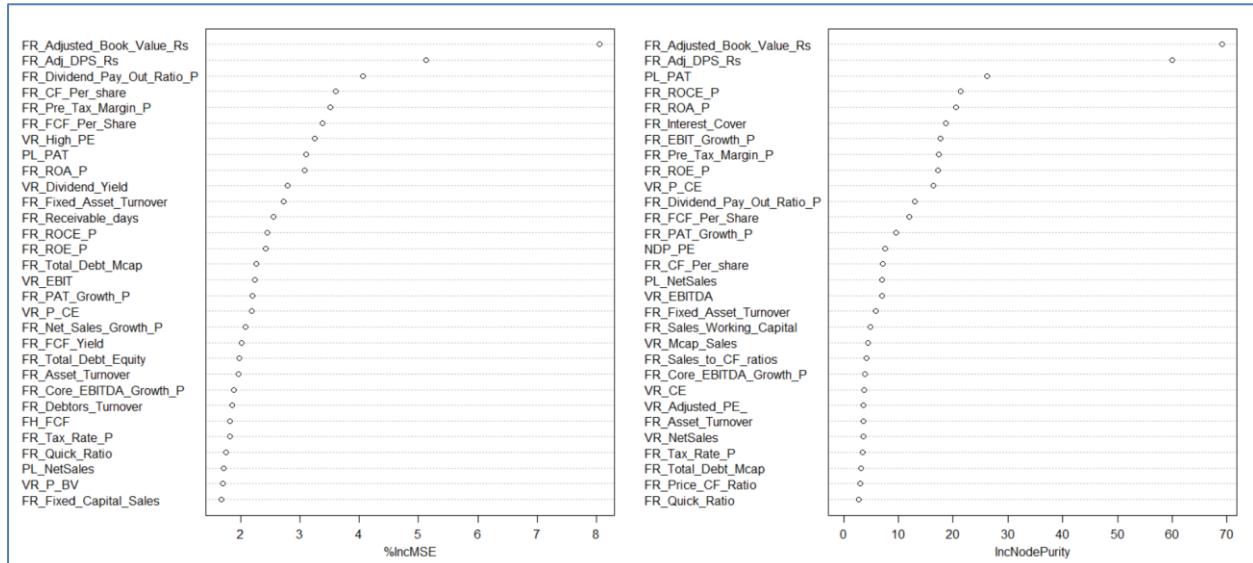


Fig: FS11- Random Forest Variable Importance Plot

The variable importance plot gives the following variables as important

- Adjusted Book Value
- Adjusted Dividend Per Share
- Cash Flow per Share
- Pre-Tax Profit Margin
- Profit after Tax
- Return on Asset
- Return on Equity
- Cash Flow per share

The RMSE score based on the Random Forest Model is **0.7265402**. The important variables got out of the Random Forest Model matches with the significant variables got out of the Linear Regression Model

Overall Model Evaluation Parameters

Model Types	RMSE
Full Model	0.044747
Model#2	1.003964
Model#3	1.002625
Model#4	0.804857

Lasso	1.000953
Ridge	0.9666798
Elasticnet	1.000573
Random Forest	0.7265402
SVM	0.7803777

Table: FS3 - Overall Model Evaluation Parameters

Stage 2: Statistical analysis of the Variables and its role in PREDICTING the Growth Category of a firm for the next year

The following models have been created to predict the Adjusted EPS Growth for the next year.

- Logistic Regression
 - Stepwise Selection Method
- Tree Models
 - Random Forest
 - Rpart
- Linear Regression

Logistic Regression

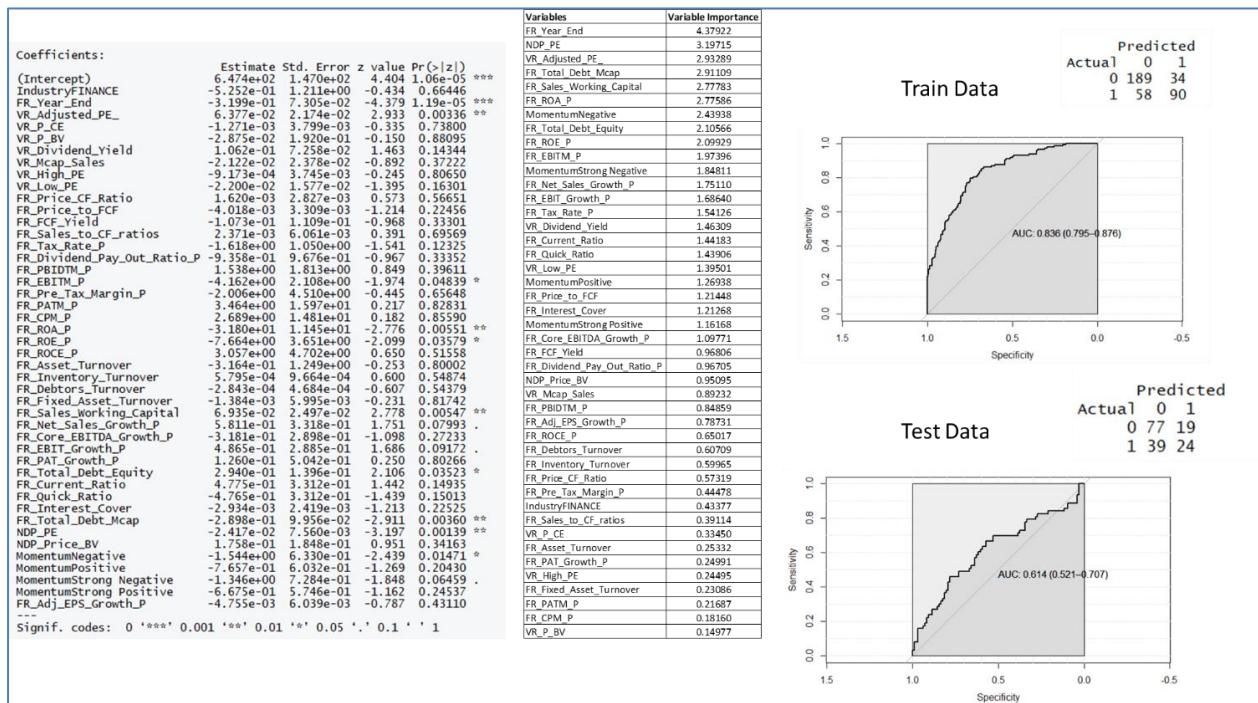


Fig: FS12 - Logistic Regression FULL MODEL

The significant variables based on the Full Model are as follows: -

- Year
- Adjusted Price to Equity Ratio
- Return on Assets
- Return on Equity

- Sales to Working Capital Ratio
- Net Sales Growth in percentage
- Total Debt Equity Ratio
- Total Debt to Market Capitalization Ratio
- Negative and Strong Negative Momentum

Multi-Collinearity

There exists a high degree of multi-collinearity between variables as indicated by the Variable Inflation Factor output given below: -

	GVIF <dbl>	Df	GVIF^(1/(2*Df)) <dbl>
FR_Current_Ratio	38021.413597	1	194.990804
FR_Quick_Ratio	38016.559986	1	194.978358
FR_PATM_P	1290.662244	1	35.925788
FR_CPM_P	1111.911313	1	33.345334
FR_Pre_Tax_Margin_P	127.809284	1	11.305277
VR_Adjusted_PE_	27.979867	1	5.289600
FR_PBIDTM_P	27.258629	1	5.220980
Industry	21.536183	1	4.640709
FR_EBITM_P	19.016746	1	4.360819
VR_Mcap_Sales	14.692884	1	3.833130
NDP_Price_BV	11.676911	1	3.417149
NDP_PE	9.780386	1	3.127361
VR_P_BV	9.753170	1	3.123007
FR_Total_Debt_Equity	9.191044	1	3.031673
FR_ROE_P	8.494855	1	2.914593
FR_ROA_P	7.532667	1	2.744570
FR_ROCE_P	7.419154	1	2.723812
VR_Low_PE	6.861250	1	2.619399
Momentum	4.076031	4	1.192009
VR_High_PE	3.851522	1	1.962529
FR_Core_EBITDA_Growth_P	3.819301	1	1.954403
FR_Total_Debt_Mcap	3.444349	1	1.855896
FR_Dividend_Pay_Out_Ratio_P	3.386313	1	1.840194
FR_EBIT_Growth_P	3.243554	1	1.800987
FR_PAT_Growth_P	3.083206	1	1.755906
FR_Year_End	2.600012	1	1.612455
FR_Price_CF_Ratio	2.581570	1	1.606726
FR_Asset_Turnover	2.477389	1	1.573972
VR_Dividend_Yield	2.372868	1	1.540412
VR_P_CE	2.251724	1	1.500574
FR_Price_to_FCF	2.024776	1	1.422946
FR_Fixed_Asset_Turnover	1.922517	1	1.386549
FR_Interest_Cover	1.864330	1	1.365405
FR_Tax_Rate_P	1.763157	1	1.327839
FR_Inventory_Turnover	1.738406	1	1.318486
FR_Sales_to_CF_ratios	1.523982	1	1.234497
FR_Net_Sales_Growth_P	1.384094	1	1.176475
FR_FCF_Yield	1.193783	1	1.092604
FR_Sales_Working_Capital	1.193585	1	1.092513
FR_Debtors_Turnover	1.139129	1	1.067300

Table: FS4 - VIF

Model based on the Significant Variables from the Full Model

Variables related to Year and Industry, though significant, have not been considered while building this model. The correlation between Year and Momentum is significant, hence the Year variable has not been included.

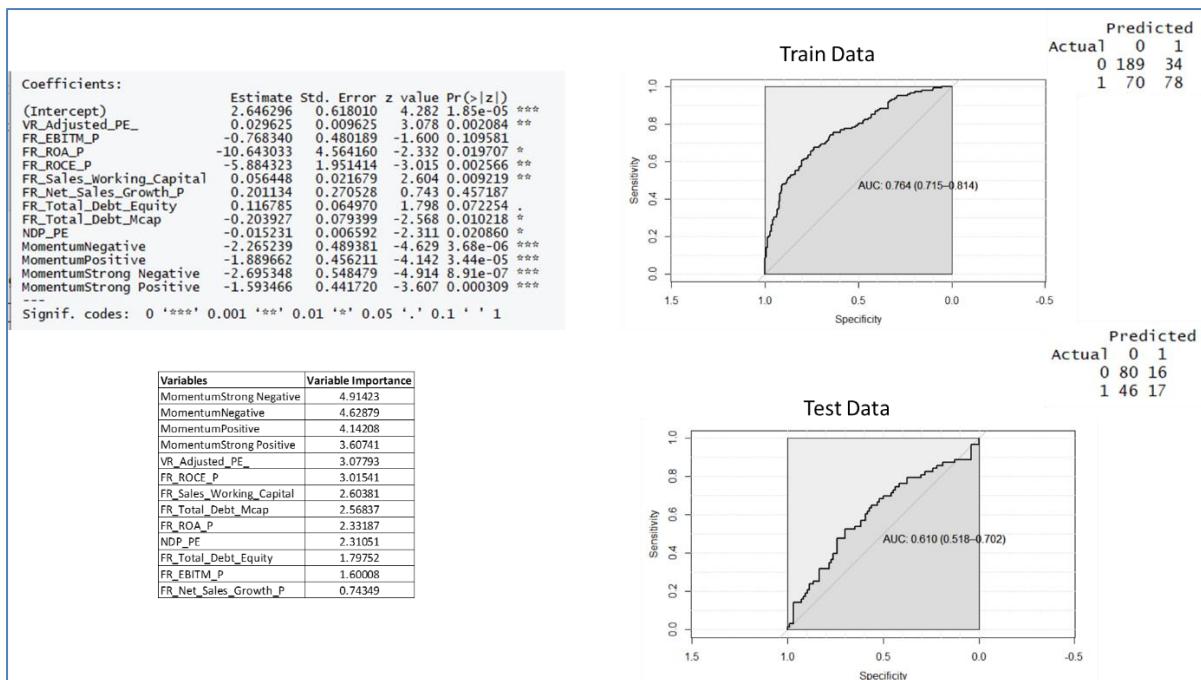


Fig: FS13 - Model Output Based on Significant Variables
Model based on the significant variables from the previous model

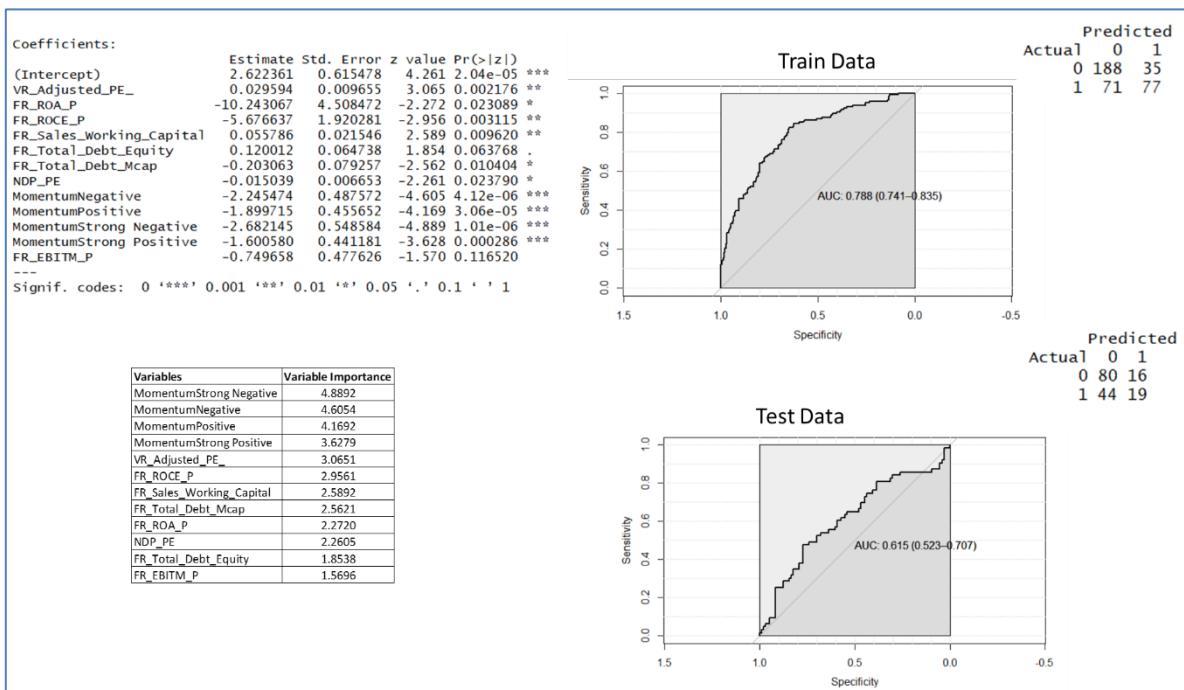


Fig: FS14 - Model Output Based on Significant Variables from Previous Model

Stepwise Selection Method

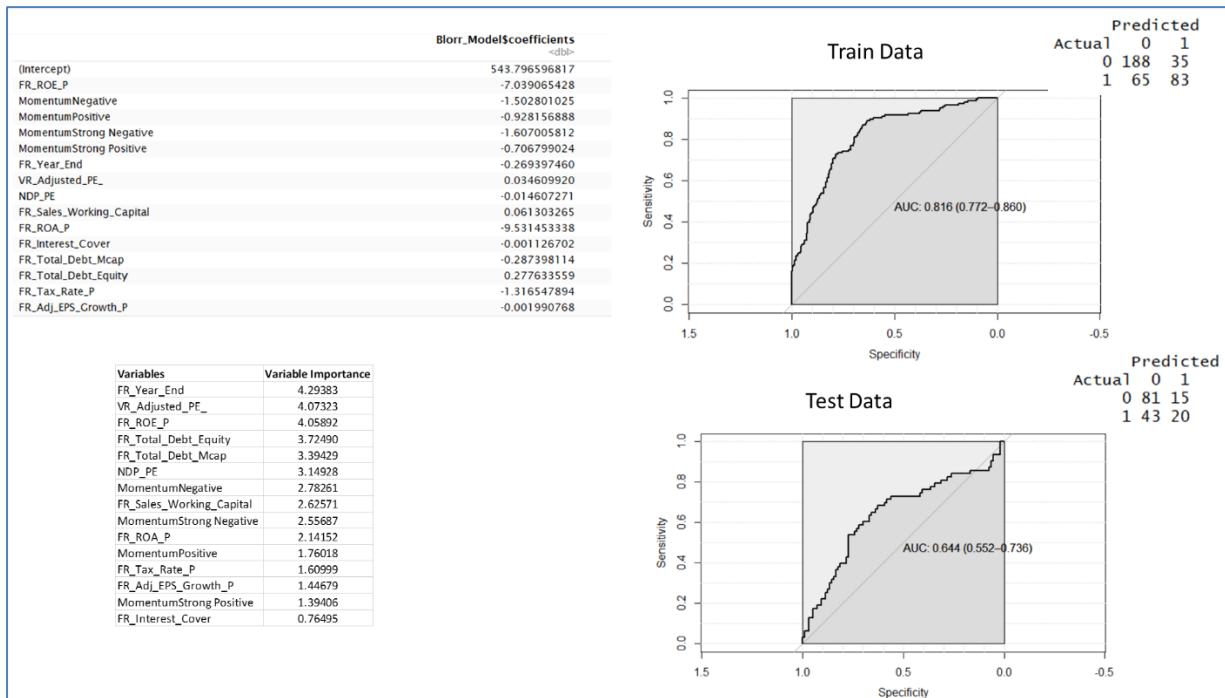


Fig: FS15 - Stepwise Selection Model Output
Logistic Regression Summary

	Full Model		Model#1		Model#2		Blorr(Full Model)	
	Train	Test	Train	Test	Train	Test	Train	Test
AUC	83.40%	61%	76.40%	61%	78.80%	61.50%	81.50%	64%
Sensitivity	61%	37%	53%	27%	52%	30%	56%	32%
Specificity	84%	81%	85%	83%	84%	83%	84%	84%

Table: FS5 - LR Summary

The significant variables based on the above models that will help predicting the growth category of a firm are as follows: -

- Momentum
- Adjusted Price Equity Ratio
- Return on Capital Employed
- Sales to Working Capital Ratio
- Total Debt to Market Capitalization Ratio

Random Forest

Year Variable followed by Growth in Profit After Tax will have the highest impact on the accuracy, if excluded.

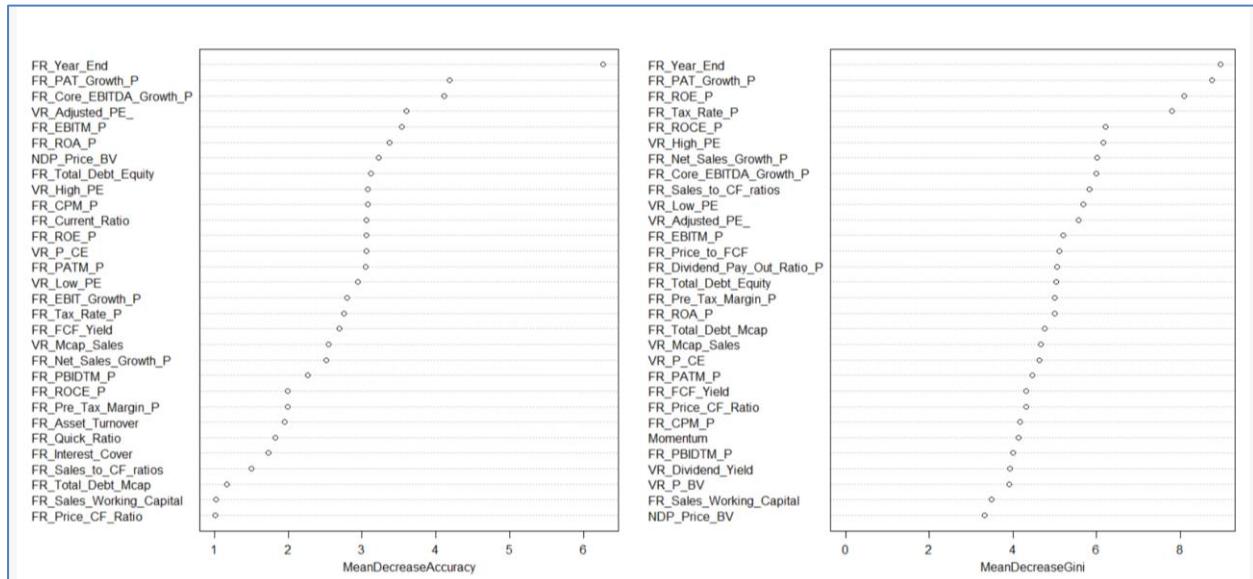


Fig: FS16 - Variable Importance Plot

Year and Growth in PAT can be dropped, since Year has a high degree of correlation with Momentum and Growth in Profit After Tax has a high degree of correlation with Adjusted Earnings Per Share.

Variables such as Total Debt to Market Capitalization, Adjusted Price Equity and Return on Capital Employed features amongst the topmost variables that have a high impact on accuracy.

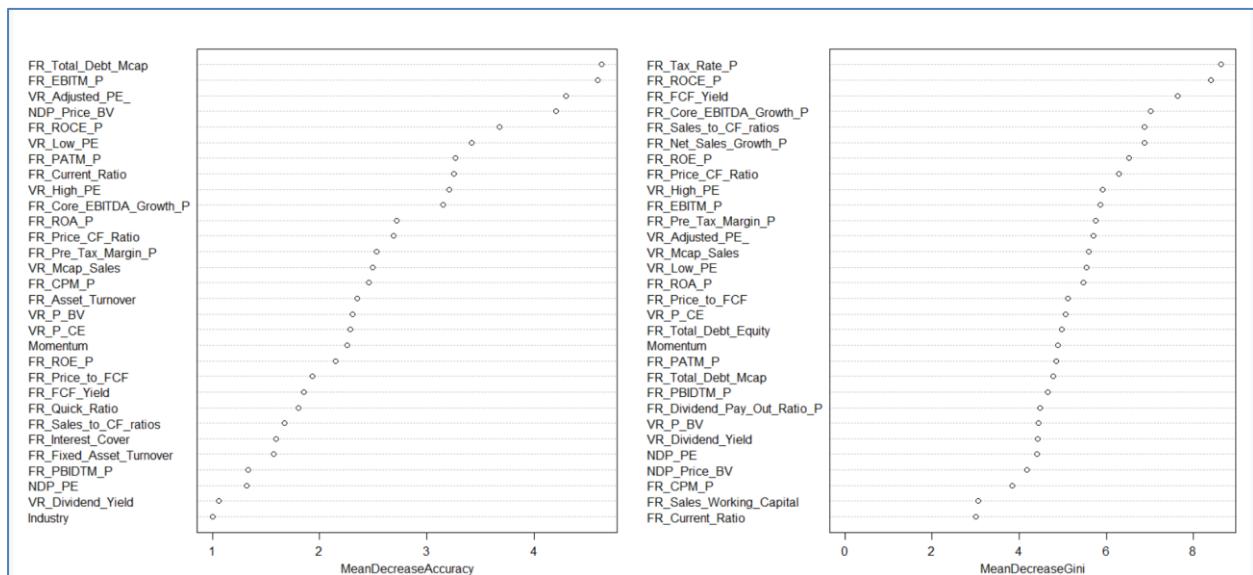


Fig: FS17 - Variable Importance Plot

RPART

Following decision tree is formed based on the following parameters

- Complexity Parameter, Cp = 0 -> The value at which the relative error is the least

	CP	nsplit	rel error	xerror	xstd
1	0.148649	0	1.00000	1.00000	0.063729
2	0.094595	1	0.85135	1.01351	0.063869
3	0.054054	2	0.75676	0.93243	0.062903
4	0.040541	3	0.70270	0.89865	0.062412
5	0.026000	4	0.66216	0.91892	0.062713

- The minimum number of observations in a node (minsplit) considered is 60
- The minimum number of observations in any terminal <leaf> node considered is 20

The first node split is based on *Return on Equity* followed by *Year*, *Sales to Cash Flow Ratio*, *Price to Cash Flow Ratio*, *Asset Turnover*

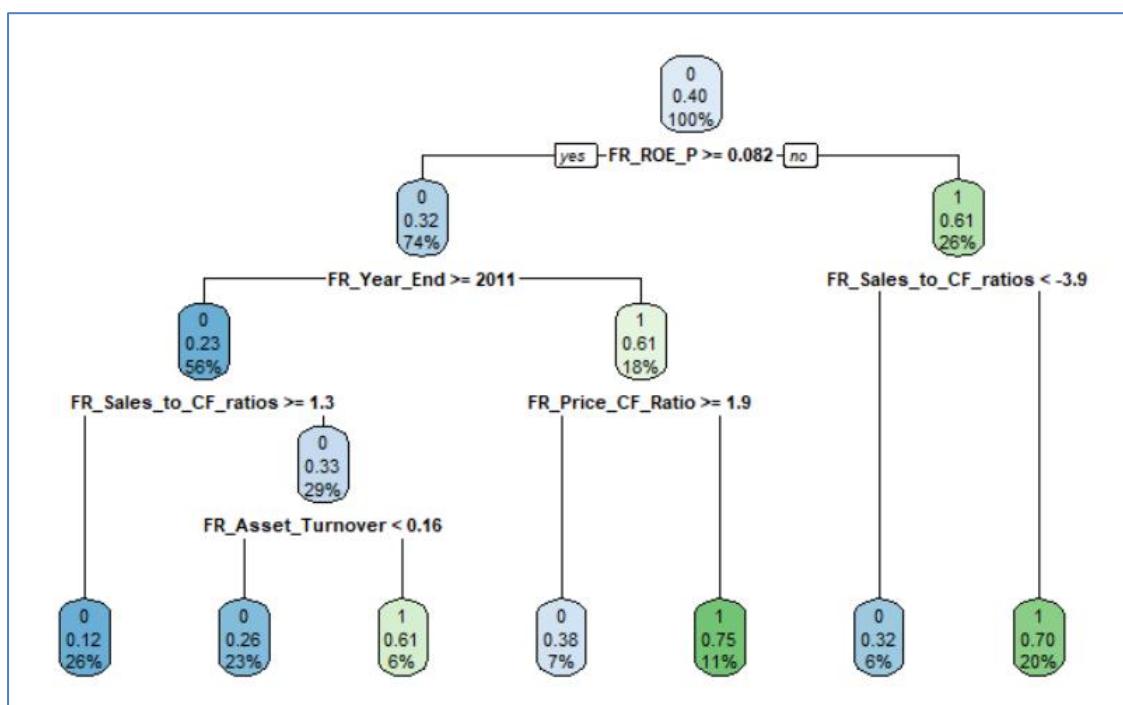


Fig: FS18 - RPART Tree

Tree-Based Models Summary

	rpart		Random Forest	
	Train	Test	Train	Test
Accuracy	75%	60%	100	66%
Sensitivity	66%	41%	100	38%
Specificity	81%	72%	100	84%

Table: FS6 - Tree Based Model Summary

Linear Regression

Linear Regression model applied on a lagged dataset having Adjusted EPS in Rs for the next year as the target variable has pointed out the following ratios as significant: -

- Return on Equity
- Earnings Before Interest and Tax Margin
- Profits After Tax Margin
- Asset Turnover Ratio
- Total Debt to Equity Ratio
- Total Debt to Market Capitalization

However, Linear Regression Model has an Adjusted R² of around 20% which is on the lower side.

Stage 3: Based on the significant variables derived from the Stage 2, determining the impact of the time, leveraging slope, 3-Year average and current year's data on the Growth of the firm for the next year

The dataset has been transposed for the various bank and financial services firms based on the significant variables which have identified from the Logistic Regression models. The following variables have been included in the transposed dataset: -

- Return on Assets
- Return on Capital Employed
- Return on Equity
- Sales to Working Capital Ratio
- Total Debt to Market Capitalization Ratio
- Total Debt Equity Ratio
- Dividend Payout Ratio
- Adjusted Price Equity Ratio

Data for the above variables is available from 2009 to 2019. The following has been calculated for the above variables to form the dataset: -

- Slope of the values from 2009 to 2018
- Average of the last 3 years of data – 2016 to 2018
- Last year value - 2018

The dataset was modified to have the dependent variable as the Target Growth for the year 2019 calculated based on the Adjusted Earnings of Share.

The number of records in the dataset is reduced to 53 observations with 26 variables. The significant variables based on Logistic Regression model are as follows: -

- Average of Return on Capital Employed / Return on Equity for 3 Years
- Average of Sales to Working Capital Ratio for 3 Years

Slope of the variables calculated over 10 Years, is not significant in predicting the Growth category of a firm. Average of the ratios and the ratios in the current year have a higher significance than slope of ratios over the last 10 years. Thus, the horizon of analysis of ratios for long term investment does not go beyond the last 3 years of performance.

Findings Linked to Project Objectives

The industry standard factors are Value, Size, Momentum, Volatility and Quality. However, Factor-Based Investment strategy does not take a holistic view, i.e. all the factors and the underlying variables and its impact on the Adjusted Earnings per Share of a stock. The objective of the project was to determine key factors and its underlying variables that will help in making investment decisions.

Financial Ratios and its performance are an indicator of the health of any stock. Financial Ratios considered in the dataset belong to different categories: -

- Valuation Ratios
- Growth Ratios
- Liquidity Ratios
- Performance Ratios

Based on the statistical analysis performed in the dataset related to Financial Services, following key ratios map to the ratio categories which in turn contribute to Factors – Value, Size, Momentum, Volatility and Quality: -

Categories	Ratios
Valuation Ratios	Adjusted Price to Equity
Performance Ratios	Return on Assets Return on Capital Employed
Liquidity Ratios	Debt to Market Capitalization Sales to Working Capital

Table: FS7 - Categories & Financial Ratios

Thus, Adjusted Price to Equity ratio will be key driver for the VALUE of a stock in Financial Services. Whereas Return on Assets and Return on Capital Employed, Debt to Market Capitalization and Sales to Working Capital Ratios will be the key driver for the QUALITY of a stock in Financial Services. MOMENTUM will be standard factor driving all investments and may not be specifically linked to a ratio. The average performance of certain ratios such as Return on Equity over the last 10 Years and the ratios in the most recent previous year could potentially be driving MOMENTUM of a stock.

Recommendations & Conclusions

As an investor or a fund manager looking to make long-term investments in firms, the following factors need to be analysed: -

- **Return on Assets** is a profitability ratio that provides how much profit a company is able to generate from its assets. *If the example of Yes Bank is considered, the Earnings per Share of Yes Bank dropped from Rs.48.01 in 2015 to a Rs.7.43 in 2020, while the Return on Assets dropped from 1.64 in 2015 to a 0.50 in 2020. This proves that ROA does have a significant impact on Adjusted EPS*

Parameters	MAR'19 (₹ Cr.)	MAR'18 (₹ Cr.)	MAR'17 (₹ Cr.)	MAR'16 (₹ Cr.)	MAR'15 (₹ Cr.)
Earnings Per Share (Rs)	7.43	18.34	72.95	60.39	48.01
Performance Ratios:					
ROA(%)	0.50	1.60	1.75	1.68	1.64
ROE(%)	6.53	17.67	18.58	19.94	21.33
ROCE(%)	7.19	11.58	14.08	13.33	13.44

Source: https://www.ndtv.com/business/stock/yes-bank-ltd_yesbank/financials-historical-ratio

Fig: FS19 - Performance Ratio of a company over years

- **Return on Capital Employed** is the primary measure of how efficiently a company utilizes all available capital to generate additional profits. ROCE, ROA and ROE are analysed together, and this supports the multi-collinearity between the two variables. ROCE provides a better indication of financial performance for companies with significant debt.
- **Adjusted Price to Equity Ratio** is calculated by dividing a company's stock price by the average of the company's earnings for the last ten years, adjusted for inflation.
- **Sales to Working Capital Ratio** is a ratio measuring the efficiency of company's working capital utilization in order to generate the certain level of sales
- **Total Debt to Market Capitalization Ratio** a solvency measure that shows the proportion of debt a company uses to finance its assets, relative to the amount of equity used for the same purpose. A higher ratio result means that a company is more highly leveraged, which carries a higher risk of insolvency.

Apart from the above factors, Momentum of the stock over the last one year will be another factor to assess. Positive or Negative Momentums stocks tend to continue with the same momentum the following years as well. Average of the ratios and the ratios in the current year have a higher significance than slope of ratios over the last 10 years. Thus, the horizon of analysis of ratios for long term investment should not go beyond the last 3 years of performance.

2.2 Consumer

Data Summary

Dimension – 809(rows) x 59 (variables) (including Target variable)
 74 companies in the data set spread across 8 industry for Consumer Sector

Exploratory Data Analysis

Univariate Analysis: on the scaled data would show that majority variables are not normally distributed which is due two reasons 1. Data has outliers on the higher side 2. Number of observations (note – the graph for the univariate analysis is included in Appendix)

Bivariate Analysis:

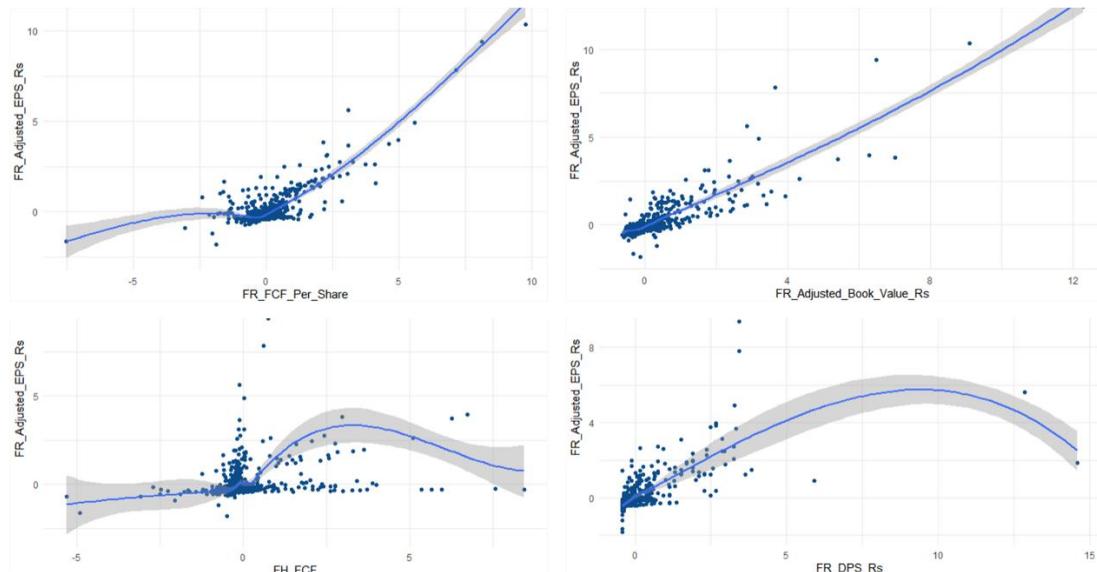


Fig: CN1 - Adjusted EPS Vs FCF/Adjusted BV/ FCF Per Share / DPS

Companies having a higher book value seems to have a higher Adjusted Earnings per share

Following are top companies with higher EPS & high book value are

- Eicher Motors Ltd.
- Maruti Suzuki Ltd.
- Page Industries Ltd.
- Vardhman Textiles Ltd.

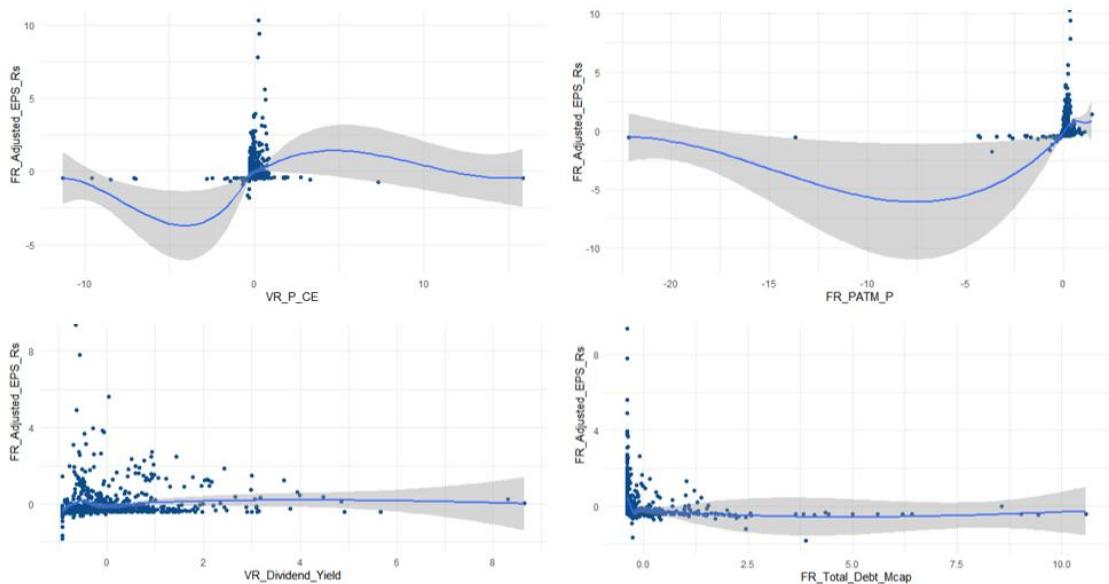


Fig: CN2 - Adjusted EPS Vs P_CE/PATM/Dividend Yield/ Total Debt Mcap

The above variables Dividend Yield, Profit After Tax Margin, Total Debt Market Capital have a weak correlation with Earnings Per Share

Following charts give a view into the consumer macroeconomic sectors based on different financial ratios categorized by Growth classification of the company. The charts indicate correlation between: -

- Return on Assets, Return on Capital Employed and Return on Equity
- Price to Free Cash Flow, Free Cash Flow Yield
- Dividend Yield, Dividend Payout Ratio
- Quick Ratio, Current Ratio

Correlation :

- Running a quick correlation reveals some variables are correlated with the target variable which is observed in the above plots as well
- Below tables reveals some variabels which have high correlation with target varaibles

- Correlation also reveals some variables which are highly correlated amongst themselves. We will however not remove them now but remove them based on the VIF check on the models which should show the multicollinearity

Correlation with Target Variable	
	FR_Adjusted_EPS_Rs
FR_FCF_Per_Share	0.81
FR_Adj_DPS_Rs	0.67
FR_Book_Value_Rs	0.81
FR_Adjusted_Book_Value_Rs	0.89
FH_FCF	0.26
VR_Adjusted_PE_	-0.02
VR_P_CE	0.04
VR_Dividend_Yield	0.06
FR_Total_Debt_Mcap	-0.13
PL_NetSales	0.21
PL_PAT	0.26
FR_ROA_P	0.37

Variables with High Correlation	
	FR_EBITM_P
FR_Pre_Tax_Margin_P	0.95
FR_PATM_P	0.94
FR_CPM_P	0.94
FR_Adj_EPS_Growth_P	0.9972

Table: CN1 - Significant Correlated Variables

Statistical Tools and Techniques

Stage I: Regression Model building approach to assess which independent variables has the most significant impact on Adjusted EPS for current year

Following statistical techniques were leveraged for Stage 1: -

- Linear Regression
- Lasso
- Ridge
- Random Forest
- Support Vector Machine

Linear Regression/ Lasso/ Ridge

Model Performance Measures

Model Types	RMSE
Linear Regression	0.3533869
Lasso	0.3475785
Ridge	0.3528797
Random Forest	0.2166468
SVM	0.3681683

Table: CN2 - Model Performance Measures of Linear/Ridge/Lasso/Random Forest/SVM

Below table has the list of important variables for various models which has an effect on Adjusted EPS

	Variable of Importance		
Linear Regression Model	Ridge Model	Lasso Model	Random Forest
Adjusted Book Value	Adjusted Book Value	Adjusted Book Value	Adjusted Book Value
Free Cash Flow Per_Share			
Enterprise Value by Capital Employed (%)	Enterprise Value by Capital Employed (%)	Enterprise Value by Capital Employed (%)	Return on Capital Employed (ROCE)
Return on Capital Employed (ROCE)	Return on Capital Employed (ROCE)	Return on Capital Employed (ROCE)	Return on Equity (ROE)
Total Debt Equity	Total Debt Equity	Total Debt Equity	Enterprise Value by Capital Employed (%)
Price by Book Value	Debtors Turn Over	Price by Book Value	Debtors Turn Over
Return on Equity (ROE)	Price by Book Value	Netsales Growth (%)	Free Cash Flow Yield
Debtors Turn Over	Netsales Growth (%)	Debtors Turn Over	Price by Book Value
Dividend Yield	Dividend Yield	Return on Equity (ROE)	Netsales Growth (%)
Netsales Growth (%)	Free Cash Flow Yield	Dividend Yield	Total Debt Equity
Free Cash Flow Yield	Return on Equity (ROE)	Free Cash Flow Yield	Dividend Yield

Table: CN3 - Variable Importance Table across Models

Statistical Tools and Techniques

Stage II – Classification Problem: In this stage the predictor variable/ target variable was converted to a class – Growth Companies labeled as ‘1’ and Non- Growth Companies labeled as ‘0’. This classification was based on the calculating the 60th Percentile of EPS growth % of each Industry in a macro-economic sector.

After experimenting with median, 75th & 90th percentile, use of 60th percentile was finalized as it has the best results and distribution of companies in terms of proportions.

The target variable was set in the dataset with a lag of one year to set it up predicting which class the company belonged to in next year

The following classification models were created to assess the key variables impacting Growth/ Non-Growth companies

- Logistic Regression
 - Blorr
- Random Forest
- Rpart

Logistic Regression

Full Model :

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)_PE_	-6.421e-01	5.306e-01	-1.210	0.22622
VR_Adjusted_PE_	-3.888e-04	8.157e-04	-0.477	0.63362
VR_CE	-7.403e-02	7.999e-02	-0.926	0.35470
VR_Dividend_Yield	3.114e-04	6.637e-04	0.469	0.63896
VR_EBIT	-7.082e-04	2.400e-03	-0.295	0.76798
FR_Adjusted_Book_Value_Rs	1.585e-03	1.666e-03	0.950	0.34194
VR_CE	2.581e-01	9.496e-02	2.718	0.00657 **
VR_Mcap_Sales	-2.544e-02	2.508e-02	-1.014	0.31056
VR_High_PE	1.431e-04	1.881e-03	0.076	0.93937
VR_Low_PE	2.460e-04	6.259e-03	0.382	0.69432
PR_Accruals_CF_Ratio	1.379e-03	1.346e-03	1.046	0.30507
PR_Price_to_FCF	-3.306e-04	3.577e-04	-0.924	0.35538
PR_FCF_Yield	8.018e-04	4.288e-01	1.870	0.06150
PR_Sales_to_CF_ratios	-1.212e-03	2.250e-03	-0.539	0.58992
FR_Adjusted_EPS_Rs	-1.270e-02	7.999e-03	-1.588	0.11234
FR_Tax_Rate_P	-1.816e-01	4.758e-01	-0.382	0.70265
FR_Dividend_Pay_Out_Ratio_P	1.462e-01	2.469e-01	0.592	0.55383
PR_ROAL_P	-1.530e-01	1.496e-01	-1.026	0.30995
PR_ROA_P	-7.210e-02	1.976e-02	-0.340	0.61477 *
PR_ROE_P	9.219e-01	9.750e-01	0.956	0.33917
PR_Asset_Turnover	3.609e-01	2.116e-01	1.706	0.08806
PR_Inventory_Turnover	1.324e-01	1.520e-04	0.871	0.38353
PR_Debtors_Turnover	-2.569e-03	2.252e-03	-1.141	0.25391
PR_Fixed_Asset_Turnover	-4.398e-03	2.712e-03	-1.622	0.10485
PR_Sales_Working_Capital	3.336e-04	4.500e-04	0.741	0.45843
PR_Adjusted_Capital_Sales	2.000e-01	3.707e-01	0.525	0.59929
PR_Accruals_Growth_P	8.827e-01	5.969e-01	1.393	0.13003
PR_Core_EBITDA_Growth_P	8.476e-05	6.656e-02	0.002	0.99815
PR_EBIT_Growth_P	-2.387e-01	1.374e-01	-1.737	0.08240
PR_Total_Debt_Equity	2.690e-01	1.540e-01	1.747	0.08067
PR_Current_Ratio	2.089e-01	1.143e-01	1.827	0.06767
PR_Interest_Cover	-1.327e-05	1.227e-04	-0.108	0.91386
PR_Total_Debt_Mcap	1.167e-01	1.495e-01	0.780	0.43511
NPV	1.356e-03	2.800e-03	0.484	0.62822
PR_Adj_EPS_Growth_P	3.393e-01	2.929e-01	1.157	0.27777
Momentum_Negative	-9.121e-01	3.394e-01	-2.688	0.00719 **
Momentum_Positive	-8.297e-01	2.493e-01	-3.326	0.00088 ***
Momentum_Strong_Negative	-6.764e-01	3.350e-01	-2.019	0.04346 *
StockSplit1	-1.140e-01	5.115e-01	-0.223	0.82361

Signif. codes: 0 **** 0.001 *** 0.01 ** 0.05 * 0.1 . ' 1 . '				

Top Variable Importance
Momentum_Positive
VR_CE
Momentum_Negative
FR_ROA_P
Momentum_Strong_Negative
VR_P_BV
FR_FCF_Yield
FR_Current_Ratio
FR_Total_Debt_Equity
FR_EBIT_Growth_P
FR_Asset_Turnover
FR_Fixed_Asset_Turnover
FR_Adjusted_EPS_Rs
FR_Net_Sales_Growth_P
FR_Debtors_Turnover
FR_EBITM_P
VR_Mcap_Sales
FR_ROE_P
FR_Adjusted_Book_Value_Rs
VR_Dividend_Yield

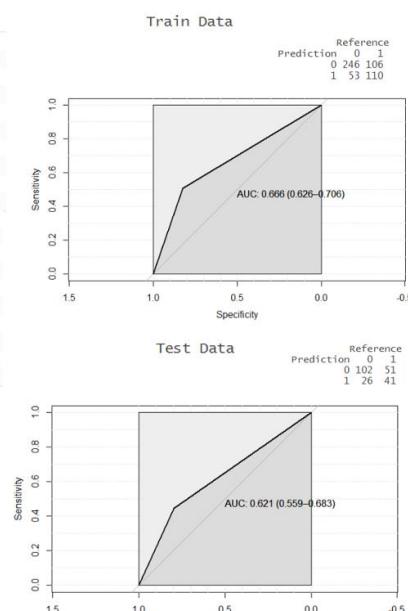


Fig: CN3 - Logistic Regression FULL MODEL

The key variables for the full model are

- Momentum
- Enterprise Value by Capital Employed (%)
- Return on Assets
- Price by Book Value

- Free Cash flow Yield
- Current Ratio
- EBIT Growth
- Assets Turnover
- Total Debt Equity

Multi-Collinearity

There exists a high degree of multi-collinearity between variables as indicated by the Variable Inflation Factor output given below: -

FR_Sales_Working_Capital	1.031472	StockSplit1	1.054627	FR_Interest_Cover	1.115531	FR_Debtors_Turnover	1.168016	FR_Adj_EPS_Growth_P	1.226896
Momentum_Negative	1.246840	VR_EBITDA	1.267599	FR_Net_Sales_Growth_P	1.284173	FR_FCF_Yield	1.295060	FR_Price_to_FCF	1.296909
Momentum_Strong_Negative	1.369528	VR_Adjusted_PE_	1.375534	NDP_PE	1.535386	FR_Tax_Rate_P	1.582250	FR_Dividend_Pay_Out_Ratio_P	1.618586
VR_Dividend_Yield	1.636403	VR_High_PE	1.649519	VR_P_CE	1.686545	FR_Core_EBITDA_Growth_P	1.939875	FR_EBIT_Growth_P	2.032163
VR_Low_PE	2.283912	FR_Inventory_Turnover	2.620799	FR_Asset_Turnover	3.147141	FR_Total_Debt_Equity	5.011733	FR_Fixed_Capital_Sales	5.413850
FR_Adjusted_Book_Value_Rs	6.386172	FR_Adjusted_EPS_Rs	7.086222	FR_Current_Ratio	8.324705	FR_Quick_Ratio	9.423297	FR_ROE_P	11.090910
FR_Price_CF_Ratio	11.995766	FR_ROCE_P	12.393166	FR_ROA_P	13.631545	VR_CE	14.291518	VR_P_BV	15.365813
FR_EBITM_P	81.661819	FR_PBIDTM_P	91.473027	FR_Pre_Tax_Margin_P	169.407256	FR_CPM_P	201.148443	FR_PATM_P	269.062296

Fig: CN4 - VIF values

Variables related to Year and Industry were not considered while building this model. There is correlation between Year and Momentum.

Model with Significant variables:

```
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.2118123 0.2014655 -1.051 0.2931
VR_Adjusted_PE_-0.0003242 0.0003988 -0.813 0.4163
VR_P_BV -0.0363886 0.0235318 -1.546 0.1220
VR_CE 0.1057374 0.0652555 1.620 0.1052
FR_FCF_Yield 0.5138622 0.2636609 1.949 0.0513 .
FR_ROA_P -5.9509527 1.4115802 -4.216 2.49e-05 ***
FR_Total_Debt_Equity 0.1704160 0.0941994 1.809 0.0704 .
FR_Current_Ratio 0.1636040 0.0813021 2.012 0.0442 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Dispersion parameter for binomial family taken to be 1
Null deviance: 700.51 on 514 degrees of freedom
Residual deviance: 660.11 on 507 degrees of freedom
AIC: 676.11
Number of Fisher Scoring iterations: 5
```

Top Variable Importance	
FR_ROA_P	
FR_Current_Ratio	
FR_FCF_Yield	
FR_Total_Debt_Equity	
VR_CE	
VR_P_BV	
VR_Adjusted_PE_-	

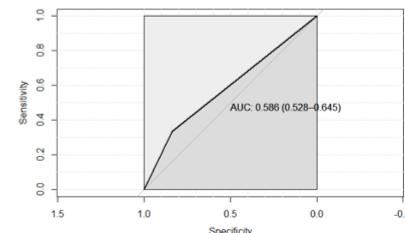
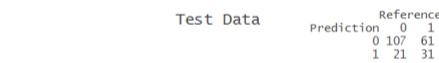
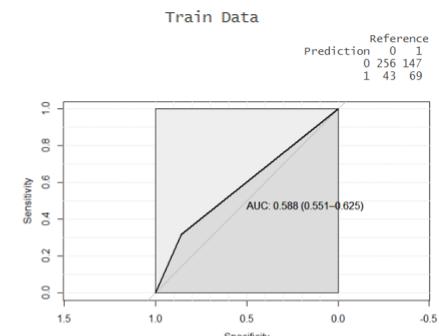


Fig: CN5 - Model Output Based on Significant Variables

Stepwise Selection Method using Blorr:

```
Coefficients:
(Intercept)          FR_ROA_P        Momentum.Positive
-0.408404           -7.808241       -0.794449
FR_Total_Debt_Mcap  0.303605        FR_Asset_Turnover
                     0.415985        FR_Current_Ratio
Momentum.Strong.Negative
-0.692100           Momentum.Negative
                     -0.791902        FR_Fixed_Asset_Turnover
FR_EBIT_Growth_P   -0.165547        0.709130        0.209027
                     VR_CE           0.081084        0.964085
FR_Adjusted_EPS_Rs -0.005661

Degrees of Freedom: 514 Total (i.e. Null); 501 Residual
Null Deviance: 700.5
Residual Deviance: 620.9      AIC: 648.9
```

Top Variable Importance

```
FR_ROA_P
Momentum.Positive
FR_Total_Debt_Mcap
FR_Asset_Turnover
FR_Current_Ratio
Momentum.Strong.Negative
Momentum.Negative
FR_Fixed_Asset_Turnover
FR_EBIT_Growth_P
FR_FCF_Yield
FR_Net_Sales_Growth_P
FR_Adjusted_EPS_Rs
VR_CE
```

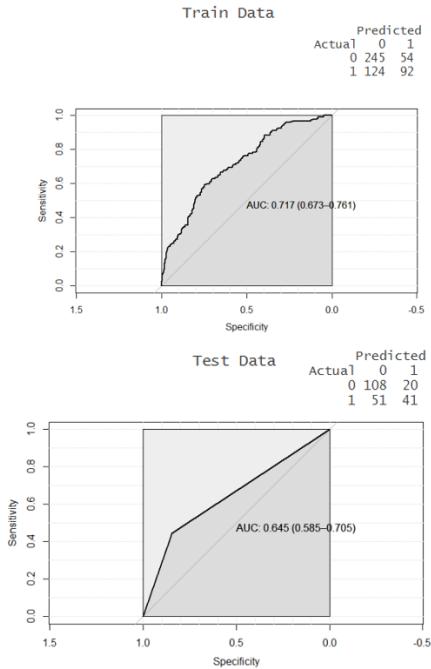


Fig: CN6 - Stepwise Selection Output using Blorr

Logistic Regression Summary

Model	Train				Test			
	Sensitivity	Specificity	Accuracy	AUC	Sensitivity	Specificity	Accuracy	AUC
Logistic Full Model	0.5093	0.8227	0.691	0.666	0.4457	0.7969	0.65	0.621
Logistic Model#1	0.4491	0.8227	0.666	0.636	0.4348	0.7969	0.646	0.616
Logistic Model#2	0.3194	0.8562	0.631	0.588	0.337	0.8359	0.627	0.586
Logistic Model(Blorr)	0.4259	0.8194	0.654	0.717	0.44565	0.8438	0.677	0.645

Table: CN5 - LR Summary

The significant variables based on the above models that will help predicting the growth category of a firm are as follows: -

- Momentum
- Return on Assets
- Price by Book Value
- Free Cash flow Yield
- Current Ratio
- EBIT Growth
- Enterprise Value by Capital Employed (%)
- Assets Turnover
- Total Debt Equity

- Adjusted EPS

Random Forest Model:

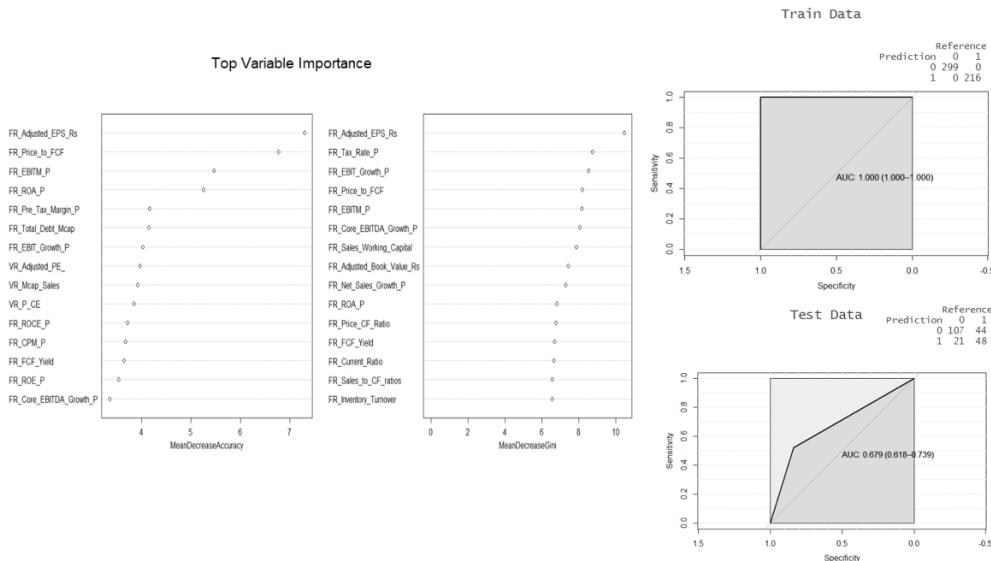


Fig: CN7- Random Forest Model Output

In above Adjusted EPS is the variable that has highest impact in accuracy followed by Price to Free Cash Flow, EBITM %, Return on Assets.

The full model is overfit

Random Forest Model with top variables:

Creating a model with only top variables does bring down the overfitting compared to full model

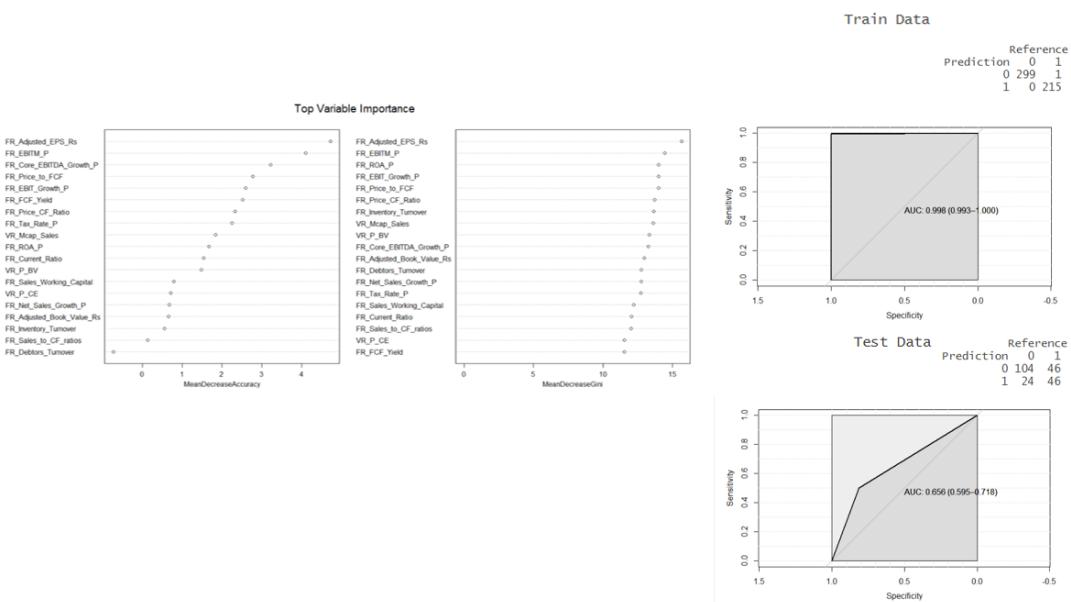


Fig: CN8 - Random Forest Model with Top Variables Output

RPART Model:

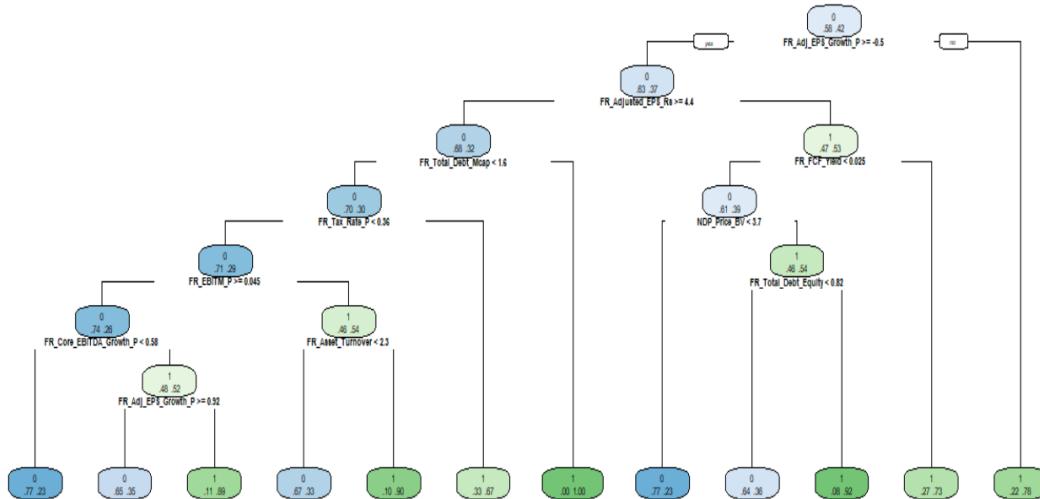


Fig: CN9 - RPART Tree

- The split in Decision Tree is based on Adjusted EPS followed by Free Cash Flow Yield, Total Debt Market Capitalization, Price to Book Value Ratio, Tax Rate

Tree Based Model summary

Model	Train			AUC	Test			AUC
	Sensitivity	Specificity	Accuracy		Sensitivity	Specificity	Accuracy	
Random Forest	1	1	1	1	0.4783	0.8047	0.668	0.641
Random Forest	1	1	1	0.998	0.5	0.8125	0.682	0.656
Decision Tree	0.588	0.8829	0.759	0.735	0.4565	0.7578	0.632	0.607

Table: CN6 - Tree Based Model Summary

The Variables that come out to be on top for Tree based model:

- Adjusted EPS
- Return on Assets
- Free Cash Flow Yield
- Price to Free Cash Flow ratio
- Price to Book Value Ratio
- EBITM

Stage III: Based on the significant variables derived from the Stage 2, determining the impact of the time by leveraging slope – Change in the variables compared to change in time, 3-Year average of the variables and last year's value of the significant variables. The target variable is class of the firm (Growth '1' or Non-Growth '0') for the next year (2019)

The dataset has been transposed for all the companies in consumer macro-economic sector based on the top significant variables which have identified from the various models. The following variables have been included in the transposed dataset: -

- Enterprise Value by Capital Employed (%)
- Price to Book Value
- Free Cash flow Yield
- Assets Turnover
- EBIT Growth
- Return on Assets
- Adjusted EPS
- EBITM
- Current Ratio

Note: Momentum was dropped from selection in preparation of this dataset since slope or average of past years cannot be calculated. But it is a key variable in most models in stage II

Data for the above variables is available from 2009 to 2019. The following has been calculated for the above variables to form the dataset: -

- Slope of the values from 2009 to 2018
- Average of the last 3 years of data – 2016 to 2018
- Last year value - 2018

The dataset was modified to have dependent variable as the Target Growth for year 2019 calculated based on the Adjusted Earnings of Share.

The number of records in the dataset is reduced to 74 observations with 24 variables. The significant variables based on Logistic Regression model are as follows: -

- Enterprise Value by Capital Employed (%) of last year
- 3-year average Enterprise Value by Capital Employed (%)
- Assets Turnover of last year
- EBIT Growth of last
- 3-year average of EBIT Growth
- 3-Year average of Return on Assets

Slope of the variables is not significant in predicting the Growth category of a company. 3-Year average and last year values seem to have more importance in growth of a company.

Findings Linked to Project Objectives

The factors categorization by the industry is spread across Valuation, Size, Momentum & Quality. However, Factor based investing does not take a holistic view of all the factors and its impact on Earning per share(EPS). The objective was to understand what are the factors that affect the EPS of Indian Companies on Stock market & how it can help in investment decision.

Financial Ratios and its performance are an indicator of the health of any stock. Financial Ratios considered in the dataset belong to different categories: -

- Valuation Ratios
- Growth Ratios
- Liquidity Ratios
- Performance Ratios

Based on the statistical analysis performed in the dataset related to the Consumer sector, following key financial ratios mapped to ratio categories:

Categories	Ratios
Valuation Ratios	Price to Book Value
Performance Ratios	Return on Assets Enterprise Value by Capital Employed (%)
Liquidity Ratios	Free Cash flow Yield
Growth Ratio	EBIT Growth

Table: CN7 - Categories & Financial Ratios

Thus, for Consumer Sector, Price to Book Value Ratio will be key driver for the VALUE factor. Whereas Return on Assets and Enterprise Value by Capital Employed (%), will be the key driver for the QUALITY factor. MOMENTUM will be standard factor driving all investments and may not be specifically linked to a ratio. The average performance of certain ratios such as Return on Equity over the last 10 Years and the ratios in the most recent previous year could potentially be driving MOMENTUM of a stock

The average performance of past years for above mentioned ratios will certainly be a key to picking growth stocks.

Recommendation & Conclusions

As an investor or fund manager looking at long-term investment in consumer firms below factor need to be analyzed: -

- **Price to Book Value** is a valuation ratio of company's market capitalization to its book value, companies with PB ratio of less than 1 are considered solid growth investments
- **Return on Assets** is profitability ratio that provides how much profit a company can generate from its assets
- **Enterprise values of Capital Employed** is a ratio between enterprise value and capital employed
- **Free Cash Flow Yield** is the cash that company generates after accounting for outflow to support its operations. It's generally used as a measure of profitability
- **EBIT Growth** is the growth in earnings before Interests and taxes

For example, in case of Asian Paints from below example we can see that the Earnings Per Share has been steadily rising which can be attributed to significant increase in liquidity ratio related to cash flow and also the enterprise value/ EBIT ratio which has increased steadily over the years.

Parameters	MAR'19 (₹ Cr.)	MAR'18 (₹ Cr.)	MAR'17 (₹ Cr.)	MAR'16 (₹ Cr.)	MAR'15 (₹ Cr.)
Operational & Financial Ratios:					
Earnings Per Share (Rs)	22.26	19.75	18.78	16.92	13.84
Valuation Parameters:					
EV / CE(x)	10.87	9.27	9.90	9.54	10.66
EV / EBIT(x)	44.58	37.16	38.29	34.24	39.47
Liquidity Ratios:					
Cash Flow from operating activities	2,184.14	2,136.09	1,420.04	1,979.67	1,143.60

Source: https://www.ndtv.com/business/stock/asian-paints-ltd_asianpaint/financials-historical-ratio

Apart from above factors Momentum of stock over the last 3 years will be another factor to assess. Positive or Negative Momentum stocks tend to continue with the same momentum the following year as well. Last 3-year average ratios have a higher significance than slope of past 10 years. Hence, the horizon of analysis of ratios should not be beyond last 3 years.

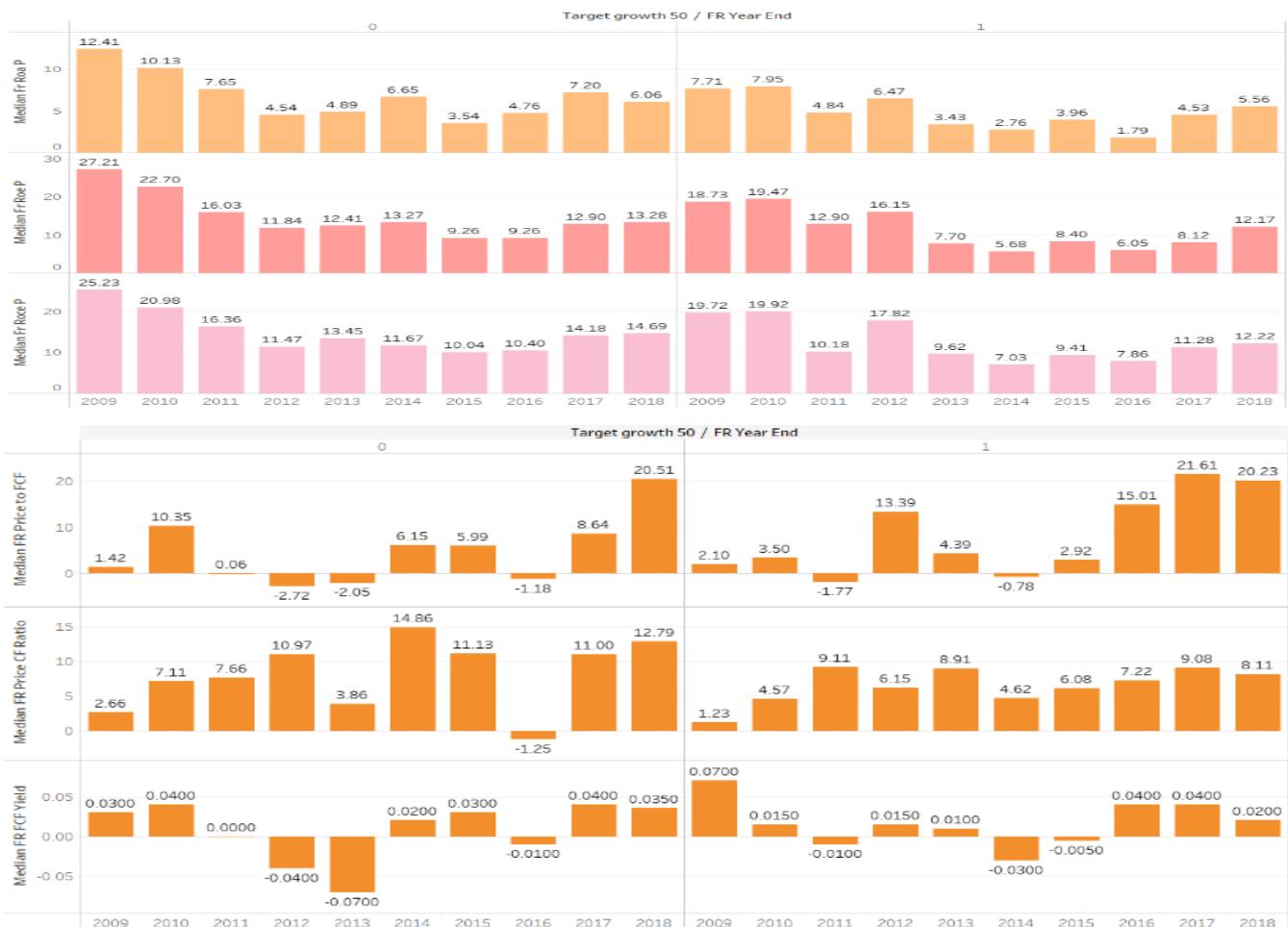
2.3 Commodities Data Summary

Dimension – 583(rows) x 59 (variables) (including Target variable)
51 companies in the data set spread across 8 industry for Commodities Sector

Exploratory Data Analysis

Following charts give a view into the performance of commodities macroeconomic sector firms based on different financial ratios categorized by Growth classification of the company. The charts indicate correlation between: -

- Return on Assets, Return on Capital Employed and Return on Equity
- Price to Free Cash Flow, Free Cash Flow Yield
- Dividend Yield, Dividend Payout Ratio



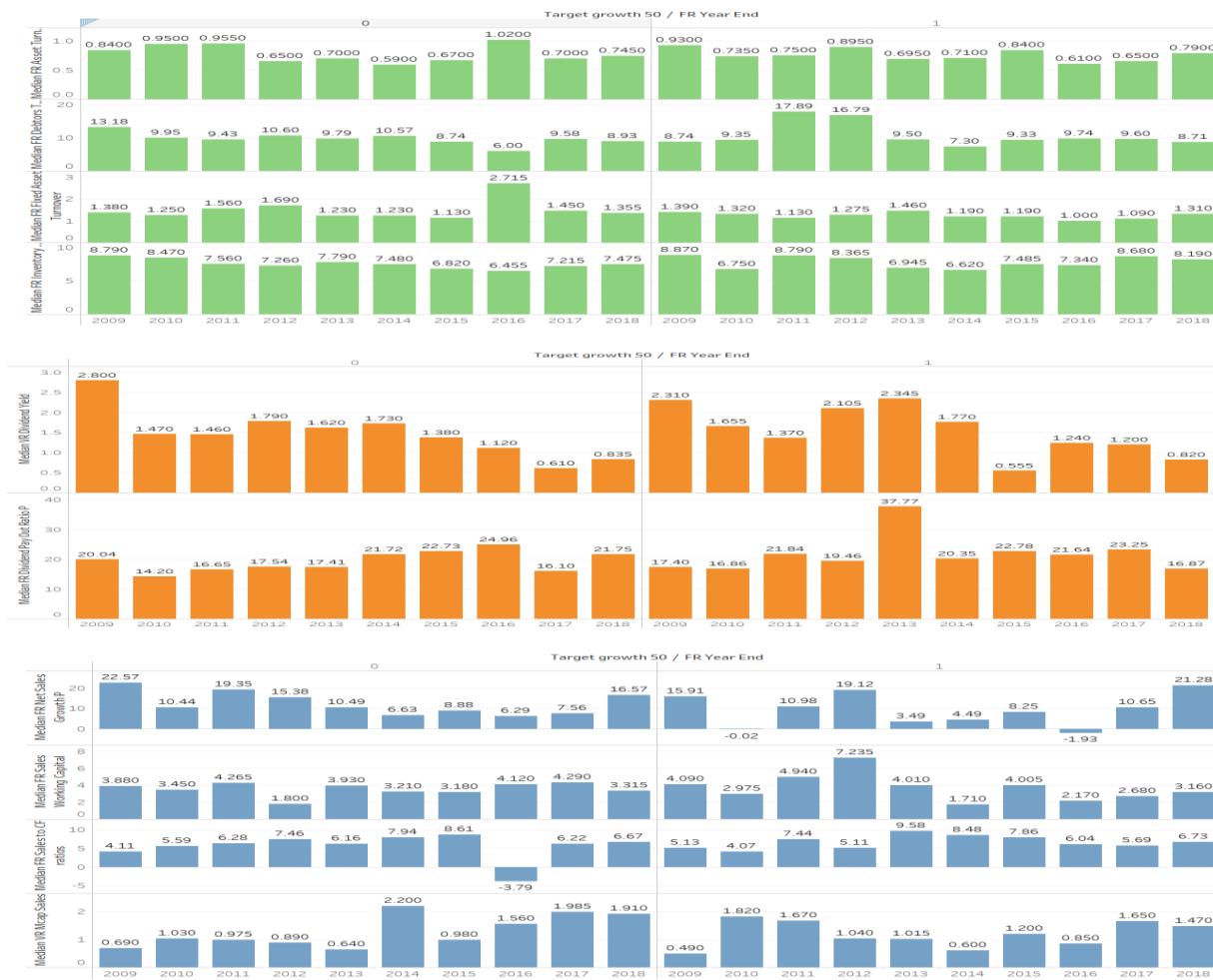


Fig: CM1- Median Ratios across time period

Adjusted EPS across Firms Belonging to Commodities Macro-Economic Sector

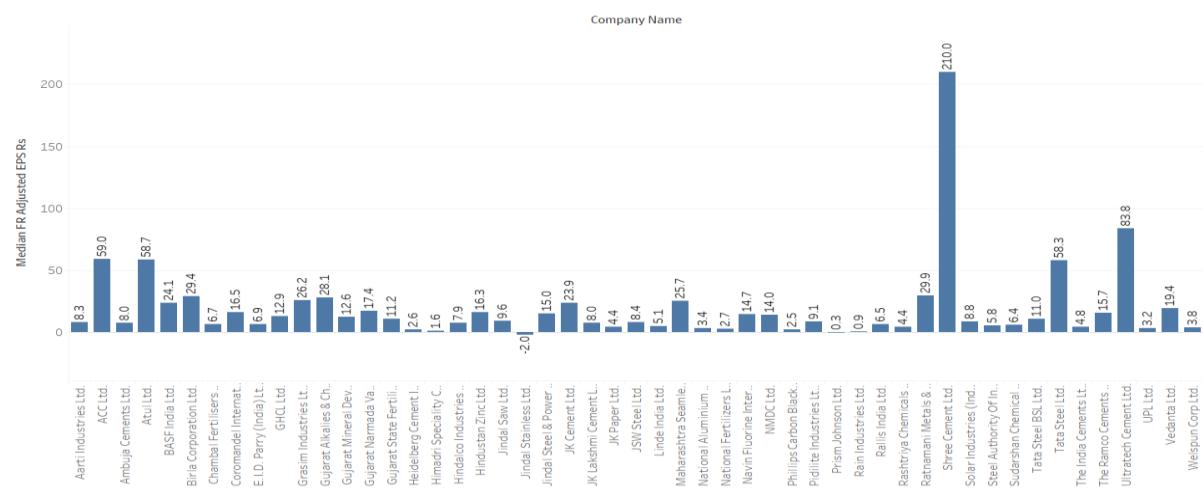


Fig: CM2 - Median Adjusted EPS by firms

Shree Cement, Ultratech Cement, ACC Ltd., Atul Ltd., and Tata steel has more earnings per share than other companies within the commodities sector

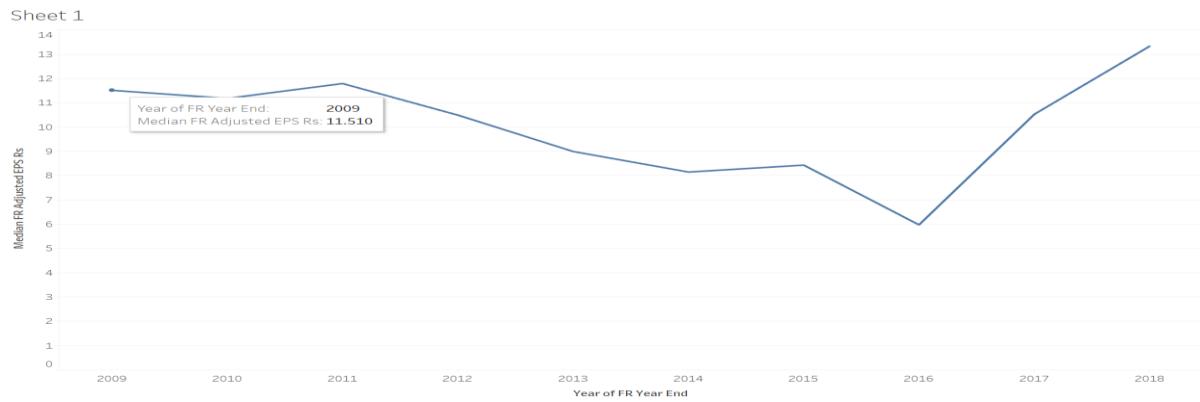


Fig: CM3 - Median Adjusted EPS for the sector

The adjusted EPS had dropped significantly in 2016 for the companies in Commodities sector and started to pick up again afterwards.

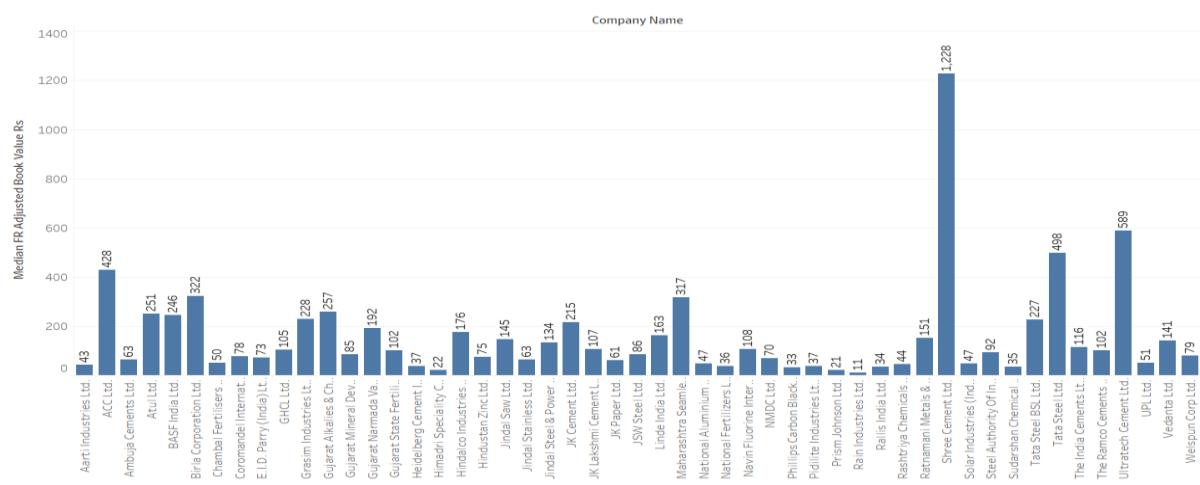


Fig: CM4 - Median Adjusted book value by firms

The adjusted book value follows the similar trend as the adjusted EPS as Shree Cement, Ultratech Cement, ACC Ltd., Atul Ltd., and Tata steel are leading the chart. The adjusted book value shows the health of a company in terms of assets and liabilities.

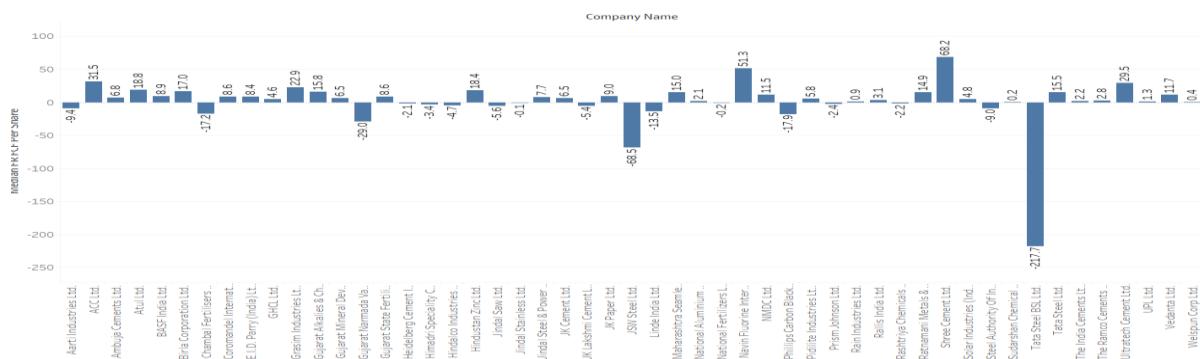


Fig: CM5 - Median cash flow by firms

The free cashflow is generally low for steel firms (Tata Steel and JSW steel)

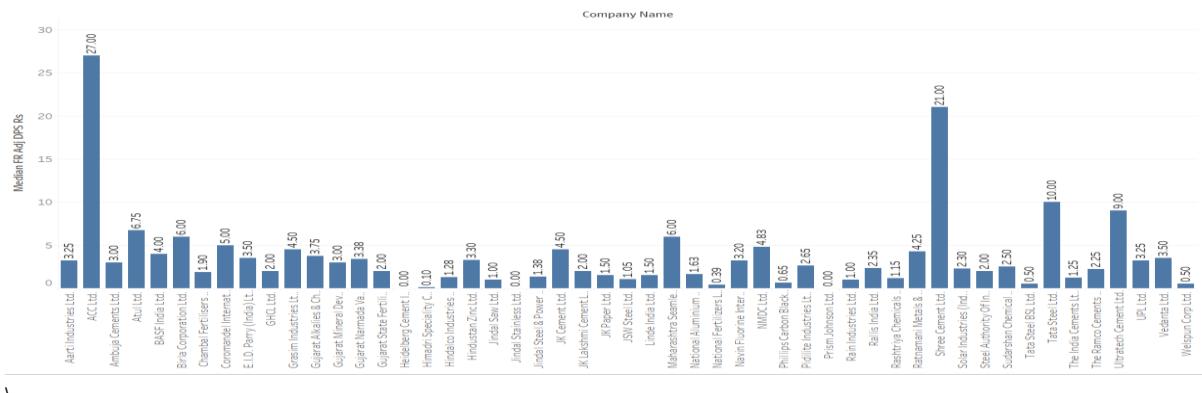


Fig: CM6 - Median adjusted DPS by firms

Median value of adjusted DPS is highest for Acc Ltd., followed by Shree Cement.



Fig: CM7 - Median ROA/ROE/ROCE by firms

As we can expect, the Return on Assets, Return on Capital Employed and Return on Equity are correlated with each other from the above chart generally.

Variables with good correlation with the adjusted EPS growth

The variables like Pre tax margin, PATM, CPM, ROE, EBITDA and ROA seems to have good correlation with the adjusted EPS growth from the above charts

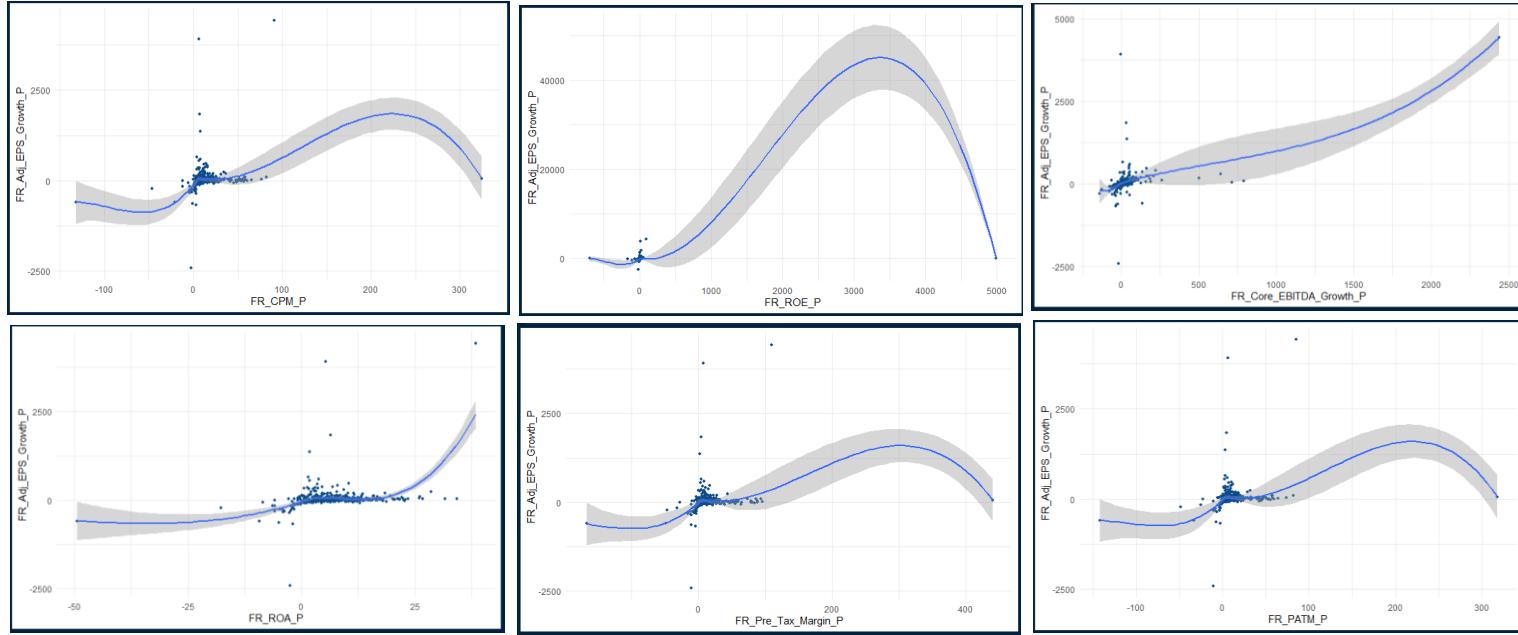


Fig: CM8 - Ratio trends against Adj. EPS

- **Correlation in the data**

The below correlation matrix shows the correlation between the variables in the model. There is a fair bit of correlation in the data between few variables which are dealt in the modelling stages by removing the highly collinear variables

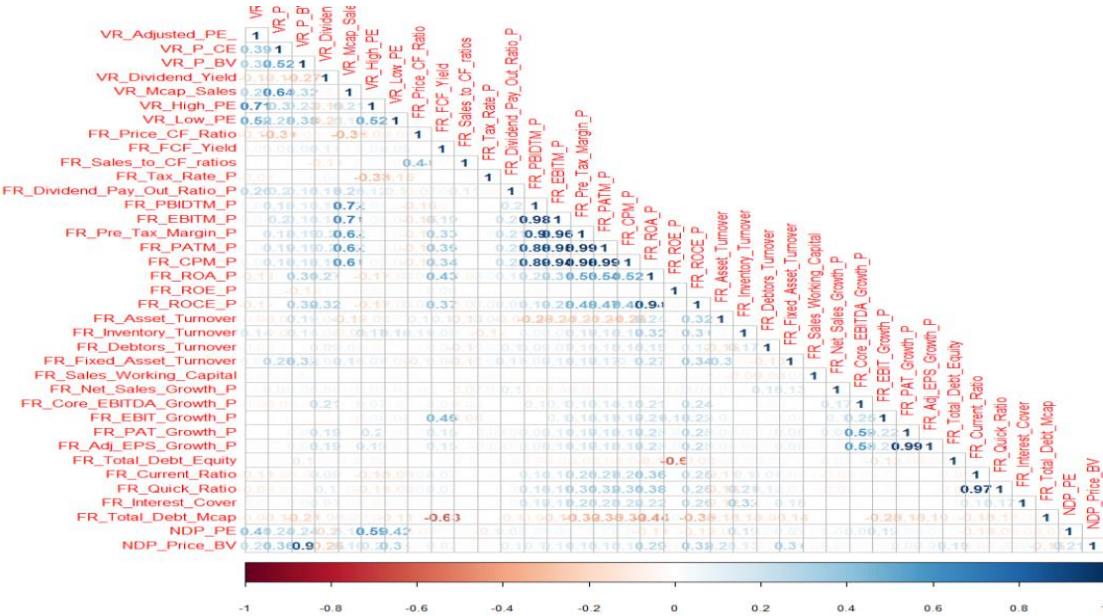


Fig: CM9 - Correlation matrix

Analysis

The analysis and modelling were divided into three stages as a part of the project: -

- Stage 1: Statistical Analysis of the variables and its impact on the Adjusted EPS for the current Year
- Stage 2: Statistical analysis of the Variables and its role in INTERPRETING the Growth Category of a firm for the next year.
- Stage 3: Based on the significant variables derived from the Stage 2, determining the impact of the previous years, leveraging slope, 3-Year average and current year's data on the Growth of the firm for the next year

Stage 1 Analysis: Statistical Analysis of the variables and its impact on the Adjusted EPS for the current Year

Following statistical techniques were leveraged for Stage 1: -

- Linear Regression
- Lasso
- Ridge
- Random Forest
- Support Vector Machine

Lasso regression output came out to be the best regularised model out of all the models run based on the RMSE values given below for the different model outputs

	LR	Ridge	Lasso	RF	SVM
RMSE	0.59	0.44	0.35	0.47	0.46

Table: CM1 - Model Performance Measures of Linear/Ridge/Lasso/Random Forest/SVM

The below variables were coming out to be significant out of the models run to predict the rupee value of adjusted EPS.

1. FR_Adjusted_Book_Value_Rs
2. FR_FCF_Yield
3. FR_EBIT_Growth_P
4. FR_Total_Debt_Mcap
5. FR_FCF_Per_Share
6. FR_Fixed_Capital_Sales
7. FR_Adj_DPS_Rs

Stage 2 Analysis: Statistical analysis of the Variables and its role in PREDICTING the Growth Category of a firm for the next year.

The following models have been created to predict the Adjusted EPS Growth for the next year.

- Logistic Regression
- Random Forest
- Rpart

Modelling Assumptions

Below are the summary stats of the data for Commodities sector

VR_Adjusted_PE_	VR_P_CE	VR_P_BV	VR_Dividend_Yield	VR_Mcap_Sales	VR_High_PE
Min. : 0.000	Min. :-164.910	Min. :-4.470	Min. : 0.000	Min. : 0.0500	Min. : 0.00
1st Qu.: 6.713	1st Qu.: 4.987	1st Qu.: 0.850	1st Qu.: 0.590	1st Qu.: 0.5175	1st Qu.: 10.67
Median : 12.490	Median : 9.135	Median : 1.545	Median : 1.415	Median : 1.0450	Median : 20.45
Mean : 23.501	Mean : 12.662	Mean : 2.206	Mean : 1.972	Mean : 2.7512	Mean : 44.45
3rd Qu.: 24.602	3rd Qu.: 16.485	3rd Qu.: 2.990	3rd Qu.: 2.553	3rd Qu.: 2.2850	3rd Qu.: 39.65
Max. : 700.490	Max. : 381.530	Max. :13.520	Max. :21.260	Max. :209.3000	Max. :727.13
VR_Low_PE	VR_Price_CF_Ratio	VR_FCF_Yield	VR_Sales_to_CF_ratios	VR_Tax_Rate_P	
Min. : 0.000	Min. :-1095.9600	Min. :-26.2100	Min. :-1566.860	Min. :-4118.88	
1st Qu.: 3.875	1st Qu.: 2.5375	1st Qu.: -0.0700	1st Qu.: 4.110	1st Qu.: 21.73	
Median : 7.785	Median : 6.9650	Median : 0.0200	Median : 6.410	Median : 28.70	
Mean : 13.006	Mean : 0.5821	Mean : -0.1082	Mean : 2.544	Mean : 18.40	
3rd Qu.: 16.858	3rd Qu.: 15.0225	3rd Qu.: 0.0700	3rd Qu.: 10.092	3rd Qu.: 33.09	
Max. : 345.160	Max. : 478.6000	Max. : 4.3900	Max. : 632.800	Max. : 236.74	
FR_Dividend_Pay_Out_Ratio_P	FR_PBIDTM_P	FR_EBITM_P	FR_Pre_Tax_Margin_P	FR_PATM_P	
Min. :-709.74	Min. :-5.25	Min. :-131.180	Min. :-167.400	Min. :-142.570	
1st Qu.: 11.26	1st Qu.: 12.76	1st Qu.: 8.562	1st Qu.: 5.088	1st Qu.: 3.728	
Median : 19.11	Median : 17.50	Median : 13.035	Median : 10.125	Median : 7.430	
Mean : 23.06	Mean : 22.66	Mean : 17.964	Mean : 13.736	Mean : 10.120	
3rd Qu.: 30.19	3rd Qu.: 23.09	3rd Qu.: 18.235	3rd Qu.: 16.262	3rd Qu.: 11.992	
Max. : 354.31	Max. : 732.73	Max. : 726.050	Max. : 439.580	Max. : 317.800	
FR_CPM_P	FR_ROA_P	FR_ROE_P	FR_ROCE_P	FR_Asset_Turnover	
Min. :-132.310	Min. :-49.480	Min. :-690.270	Min. :-63.160	Min. : 0.0100	
1st Qu.: 7.428	1st Qu.: 2.850	1st Qu.: 6.435	1st Qu.: 8.195	1st Qu.: 0.5500	
Median : 11.355	Median : 5.715	Median : 12.375	Median : 13.365	Median : 0.7400	
Mean : 14.252	Mean : 6.685	Mean : 21.019	Mean : 15.308	Mean : 0.8237	
3rd Qu.: 16.465	3rd Qu.: 10.217	3rd Qu.: 20.253	3rd Qu.: 21.000	3rd Qu.: 1.0700	
Max. : 324.480	Max. : 38.410	Max. : 4988.020	Max. : 72.060	Max. : 2.7400	
FR_Inventory_Turnover	FR_Debtors_Turnover	FR_Fixed_Asset_Turnover	FR_Sales_working_Capital		
Min. : 0.000	Min. : 0.000	Min. : 0.280	Min. :-2771.390		
1st Qu.: 5.647	1st Qu.: 5.883	1st Qu.: 0.930	1st Qu.: -1.195		
Median : 7.655	Median : 9.635	Median : 1.280	Median : 3.580		
Mean : 9.025	Mean : 18.392	Mean : 1.769	Mean : -0.446		
3rd Qu.: 10.842	3rd Qu.: 23.688	3rd Qu.: 1.990	3rd Qu.: 7.070		
Max. : 33.390	Max. : 154.720	Max. : 16.230	Max. : 923.690		
FR_Net_Sales_Growth_P	FR_Core_EBITDA_Growth_P	FR_EBIT_Growth_P	FR_PAT_Growth_P	FR_Adj_EPS_Growth_P	
Min. :-89.060	Min. :-141.26	Min. :-1853.92	Min. :-2123.620	Min. :-2414.19	
1st Qu.: 1.468	1st Qu.: -10.17	1st Qu.: -16.38	1st Qu.: -22.172	1st Qu.: -24.14	
Median : 10.540	Median : 8.66	Median : 7.85	Median : 6.595	Median : 5.15	
Mean : 16.188	Mean : 21.84	Mean : 24.41	Mean : 33.162	Mean : 30.75	
3rd Qu.: 20.782	3rd Qu.: 26.73	3rd Qu.: 32.41	3rd Qu.: 42.197	3rd Qu.: 42.21	
Max. : 1115.550	Max. : 2437.16	Max. : 1960.41	Max. : 4273.460	Max. : 4426.15	
Target_growth_50	FR_Total_Debt_Equity	FR_Current_Ratio	FR_Quick_Ratio	FR_Interest_Cover	
0:246	Min. :-61.9300	Min. : 0.1100	Min. : 0.0400	Min. : -8.450	
1:262	1st Qu.: 0.1500	1st Qu.: 0.9775	1st Qu.: 0.6375	1st Qu.: 2.248	
	Median : 0.6050	Median : 1.3900	Median : 0.9700	Median : 4.940	
	Mean : 0.7363	Mean : 1.8214	Mean : 1.3647	Mean : 76.748	
	3rd Qu.: 1.1325	3rd Qu.: 2.0050	3rd Qu.: 1.5100	3rd Qu.: 16.118	
	Max. : 59.4000	Max. :13.2900	Max. :12.7500	Max. : 7271.070	
FR_Total_Debt_Mcap	NDP_PE	NDP_Price_BV	Momentum		
Min. : 0.000	Min. : 0.000	Min. : 0.0000	: 50		
1st Qu.: 0.060	1st Qu.: 7.285	1st Qu.: 0.9175	Negative : 67		
Median : 0.360	Median : 12.975	Median : 1.6500	Positive : 127		
Mean : 1.228	Mean : 24.234	Mean : 2.3209	Strong Negative: 97		
3rd Qu.: 1.040	3rd Qu.: 26.578	3rd Qu.: 3.0150	Strong Positive:167		
Max. : 59.420	Max. : 606.300	Max. : 16.9200			

Fig: CM10 - Variable Summary

- The summary statistics shows that the data set is greatly normal as the median and mean values of the variables greatly align for majority of the variables. The rest of variables are assumed to be normally distributed on a larger dataset in the real-world scenario.
- There is multi collinearity amongst the variables which are erased by removing the high VIF variables in the model in the analysis stage.
- The outliers are extreme values and are included in the data set.

Logistic Regression on full model

Model Output

Coefficients:

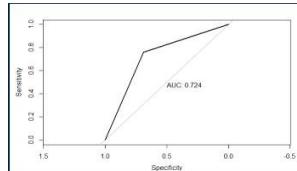
	estimate	std.	error	z value	Pr(> z)
(Intercept)	3.5807636	1.0098061	1.360	0.11558	
VR_Adjusted_PE_	-0.0133944	0.0191077	0.701	0.48331	
VR_P_C	-0.0040907	0.0135469	-0.302	0.76268	
VR_P_BV	-0.3466233	0.2822354	-1.228	0.21040	
VR_Dividend_Yield	0.0114688	0.0918441	0.125	0.90062	
VR_Ncap_sales	0.2248960	0.1178257	1.908	0.05620	*
VR_High_PE	-0.0090888	0.0041985	2.185	0.03040	*
VR_Low_PE	-0.0112084	0.0040946	-0.139	0.89825	
FR_Dividend_CF_Ratio	-0.0004442	0.0034444	0.037	0.93777	
FR_P_E_Yield	0.1247593	0.2665886	0.615	0.51218	
FR_Sales_to_CF_ratio	-0.0033748	0.0022861	-1.489	0.13642	
FR_TAX_RATE_P	0.0035753	0.0058661	0.609	0.54221	
FR_Dividend_Pay_Out_Ratio_P	0.0192874	0.0085304	2.256	0.02409	*
FR_PBDTM_P	0.1485148	0.1128518	1.316	0.18817	
FR_EBITM_P	-0.2762098	0.1284717	-2.150	0.03156	*
FR_Pre_Tax_Margin_P	0.1744461	0.0827646	2.104	0.03537	*
FR_PATM_P	-0.1221087	0.1766820	-0.691	0.48949	
FR_PME_P	0.0052054	0.0088665	0.616	0.50868	
FR_RBC_P	-0.2221503	0.1128518	-1.977	0.28115	
FR_RBC_P	-0.0271894	0.0215726	-1.260	0.20754	
FR_BDCE_P	0.0419000	0.0313709	0.812	0.41652	
FR_Asset_Turnover	-1.2187858	0.5922082	-2.058	0.03059	*
FR_Inventory_Turnover	0.0252088	0.0416247	0.609	0.54477	
FR_debtors_Turnover	0.0004520	0.0083486	0.056	0.95673	
FR_Fixed_Asset_Turnover	0.2645098	0.1713216	1.344	0.12260	
FR_sales_working_capital	0.0007762	0.0084660	1.036	0.30544	
FR_net_sales_growth	0.0107173	0.0051557	0.638	0.52253	
FR_EBIT_to_total_debt_P	0.0001818	0.0049016	0.054	0.95112	
FR_EBIT_Growth_P	-0.0008411	0.0058841	-0.951	0.34139	
FR_PAT_Growth_P	-0.0049359	0.0068321	-0.722	0.47001	
FR_Adj_EPS_Growth_P	0.0045682	0.0066262	0.689	0.49056	
FR_Total_debt_equality	0.3637452	0.2346401	1.428	0.15116	
FR_Current_Ratio	-0.0766397	0.40686936	-0.168	0.87011	
FR_Quick_Ratio	0.2527955	0.3486397	0.461	0.64400	
FR_Interest_Cover	-0.0097739	0.0054904	-1.577	0.11472	
FR_Total_debt_VCap	-0.1734476	0.2079959	-1.277	0.30865	
NDP_P	-0.0117735	0.0721170	-2.118	0.03117	*
NDP_Price_BV	0.2661263	0.2549780	1.044	0.20661	
Momentumnegative	-2.0240839	0.6422273	-3.152	0.00102	**
Momentumpositive	-0.9691851	0.5863149	-1.653	0.09833	.
'MomentumStrong negative'	-3.2057167	0.6557769	-4.888	1.02e-06	***
'Momentumstrong positive'	-1.1205778	0.5330065	-2.102	0.03552	*

Variable Importance

Variables	Variable Importance
'MomentumStrong_Negative'	4.88842399
MomentumNegative	3.11662391
FR_Dividend_Pay_Out_Rate	2.21647765
VR_High_PE	2.1647785
FR_BDITM_P	2.14906728
NDP_P	2.11806728
FR_Per_Tax_Margin_P	2.10411409
'MomentumStrong_Positive'	2.1023719
FR_Asset_Turnover	2.05806111
VR_Mcap_Sales	1.90871807
MomentumPositive	1.65310116
FR_Interest_Cover	1.57733374
FR_Fixed_Asset_Turnover	1.54935708
FR_Sales_to_CF_ratio	1.48934977
FR_Total_Debt_Equity	1.42846774
PRBDITM_P	1.31601626
FR_Total_Debt_Mcap	1.25727223
VR_B_P	1.22813359
FR_RBC_P	1.17733712
NDP_Price_BV	1.0417267
FR_Sales_Working_Capita	1.03483772
FR_EBIT_Growth_P	0.95142014
FR_ROCE_P	0.81247378
FR_PAT_Growth_P	0.72245799
FR_Adjusted_PE	0.70099316
FR_PATM_P	0.69121173
FR_Alt_EPS_Growth_P	0.68941102
FR_FCF_Yield	0.6552036
FR_Price_CF_Ratio	0.64234004
FR_Net_Sales_Growth_P	0.63942123
FR_Tax_Rate_P	0.60948024
FR_Inventory_Turnover	0.60562104
VR_Low_PE	0.53909759
FR_Quick_Ratio	0.46085173
VR_P_C_E	0.3019658
FR_Current_Ratio	0.25176776
VR_Dividend_Yield	0.13407204
FR_CPM_P	0.07567824
FR_Debtors_Turnover	0.05425408
FR_Cost_EBITDA_Growth	0.05351798

Train Data

Predicted	Actual	0	1
	0	119	54
	1	44	140



Test Data

Predicted	Actual	0	1
	0	42	31
	1	23	55

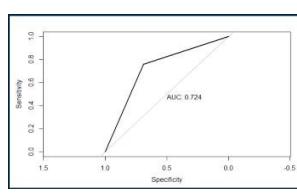


Fig: CM11 - Logistic Regression - full model

The top significant variables from the LR full model are-

1. Momentum
2. Dividend payout ratio
3. Price to Equity Ratio
4. Earnings before Tax
5. Pre-tax margin
6. Asset Turnover ratio
7. Market Capitalization to Sales ratio
8. Interest cover ratio
9. Fixed asset turnover ratio
10. Sales to Cash flow ratio

But there is high amount of multi collinearity in the data from the VIF values of the model outputs as given below.

Variables	VIF Value	Variables	VIF Value	Variables	VIF Value
FR_EBITM_P	408.5	FR_ROCE_P	15.2	FR_Fixed_Asset_Turnover	2.5
FR_PBIDTM_P	321.2	FR_Total_Debt_Mcap	12.6	FR_Dividend_Pay_Out_Ratio_P	2.3
FR_PATM_P	287.9	FR_Total_Debt_Equity	11.1	FR_Net_Sales_Growth_P	2.2
FR_CPM_P	187.4	VR_Adjusted_PE_	10.8	FR_Price_CF_Ratio	2.1
FR_PAT_Growth_P	124.2	VR_Low_PE	6.1	VR_P_CE	2.0
FR_Adj_EPS_Growth_P	118.7	NDP_PE	5.1	FR_Sales_to_CF_ratios	1.8
FR_Pre_Tax_Margin_P	110.8	'MomentumStrong Positive'	4.3	FR_Core_EBITDA_Growth_P	1.6
FR_Quick_Ratio	41.8	MomentumPositive	4.1	FR_Debtors_Turnover	1.4
FR_Current_Ratio	37.8	VR_High_PE	3.6	FR_EBIT_Growth_P	1.4
VR_P_BV	23.3	'MomentumStrong Negative'	3.3	FR_FCF_Yield	1.3
FR_ROA_P	22.9	FR_Asset_Turnover	3.2	FR_Tax_Rate_P	1.3
VR_Mcap_Sales	19.6	MomentumNegative	3.0	FR_Interest_Cover	1.2
NDP_Price_BV	19.0	FR_Inventory_Turnover	2.7	FR_Sales_Working_Capital	1.1
FR_ROE_P	15.2	VR_Dividend_Yield	2.6		

Table: CM2 - VIF - full model

Model based on selected variables from the full model.

A logistic model is built after removing the high collinear variables from the full model. The model output is as shown below.

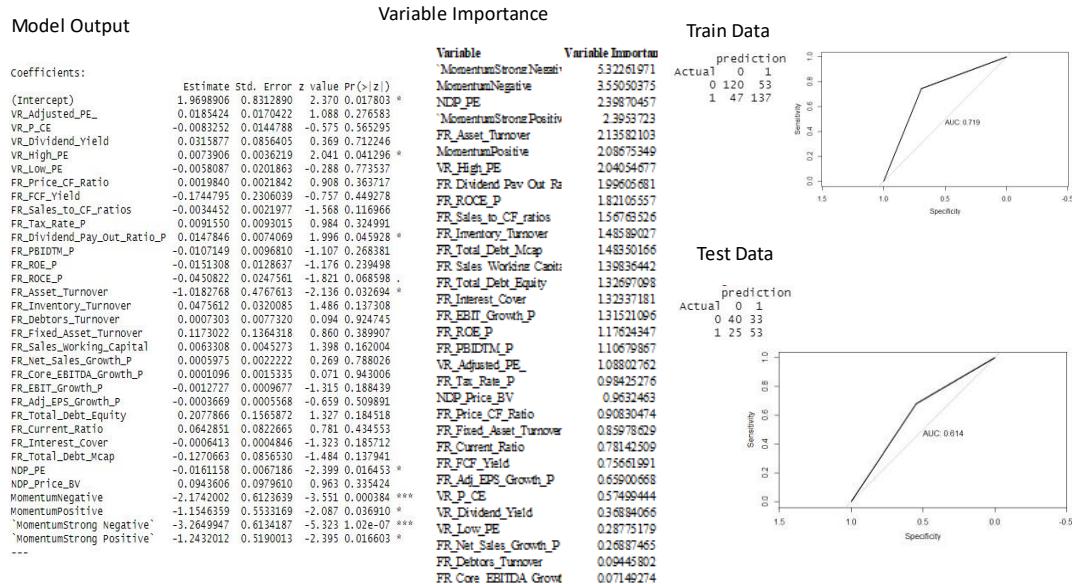


Fig: CM12 - Model output - Logistic Regression 2

The top significant variables from the LR full model are-

- 1. Momentum
- 2. NDP_PE
- 3. Asset Turnover ratio
- 4. VR_High_PE
- 5. Dividend payout ratio FR_ROCE_P
- 6. Sales to Cash flow ratio
- 7. Inventory turnover ratio
- 8. Debt to Market cap ratio
- 9. Sales to working capital ratio

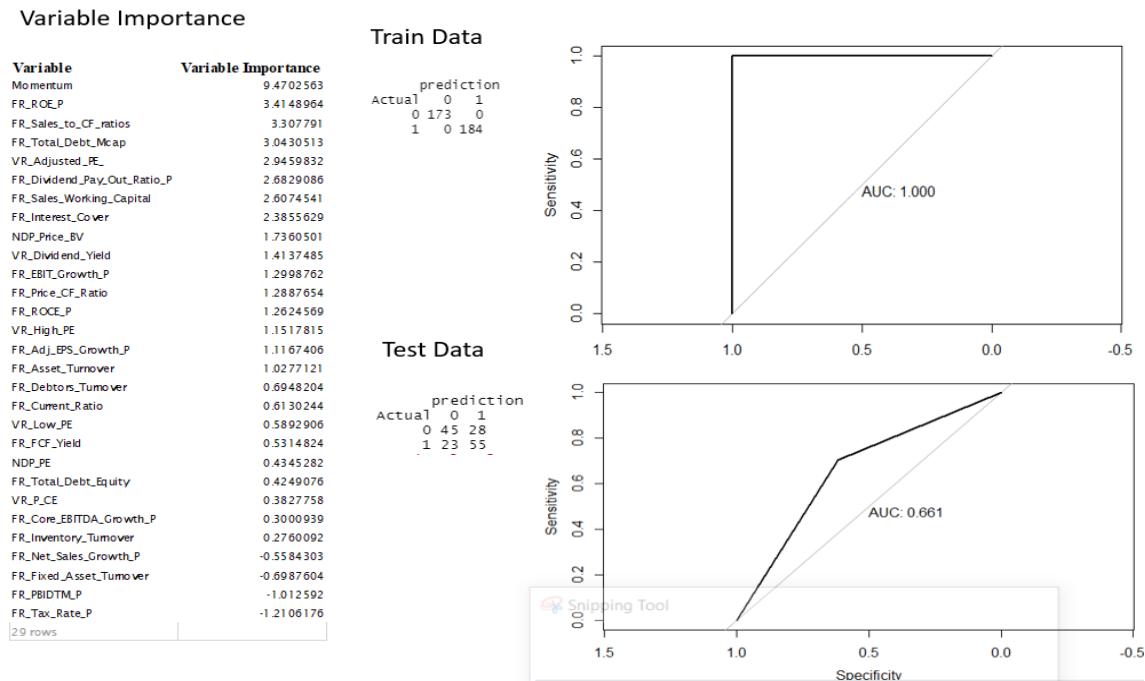
Logistic Regression Summary

	Full Model		Model with selected variables	
	Train	Test	Train	Test
Accuracy	73%	62%	72%	62%
Sensitivity	76%	71%	74%	68%
Specificity	69%	58%	69%	55%
AUC	72%	72%	72%	61%

Table: CM3 - Logistic Regression- Model measures

Random Forest model on the selected variables after removing multi collinearity.

A random forest model was built on the selected variables after removing the multi



collinearity.

Fig: CM13 - Model output - Random Forest

The top variables that came out to be important in the prediction are as below:

- 1. Momentum
- 2. Return on Equity
- 6. Dividend pay-out ratio
- 7. Sales to working capital ratio

- 3. Sales to Cash flow ratio
- 4. Total debt to Market Cap
- 5. VR_Adjusted_PE_
- 8. FR_Interest_Cover
- 9. NDP_Price_BV
- 10. VR_Dividend_Yield

	RF with Selected Variables	
	Train	Test
Accuracy	100%	66%
Sensitivity	100%	71%
Specificity	100%	62%
AUC	100%	66%

Table: CM4 -Random Forest- Model measures

Decision Tree model

The model outputs on the selected variables as predictors are as given below

```
Variables actually used in tree construction:
[1] FR_Asset_Turnover          FR_Dividend_Pay_Out_Ratio_P   FR_FCF_Yield
[4] FR_ROE_P                   FR_Sales_to_CF_ratios        FR_Sales_Working_Capital
[7] FR_Tax_Rate_P              Momentum                         VR_Adjusted_PE_
[10] VR_Dividend_Yield         VR_Low_PE

Root node error: 173/357 = 0.48459
```

Following are the considerations to arrive at the best split

cp = 0.01, minsplit = 20 and minbucket = 10

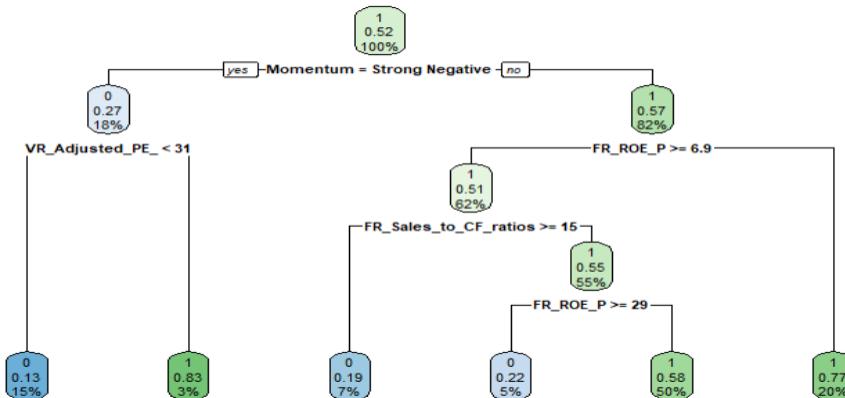


Fig: CM14 - rPart – Tree

The most important variable in separating the classes as per the tree model is Momentum which is coming out to be the most significant in the other models too. This is followed by, VR_Adjusted_PE, FR_ROE_P, Sales to cash flow ratio

Tree based Models' performance.

	RF with Selected Variables		rPart	
	Train	Test	Train	Test
Accuracy	100%	66%	69%	64%
Sensitivity	100%	71%	91%	71%
Specificity	100%	62%	46%	57%
AUC	100%	66%	69%	64%

Table: CM5 - Tree based models - Model measures

Stage 3 Analysis

A stage 3 analysis was conducted on the final set of most important variables out of all models built on the original data set.

Data Preparation for Stage 3

For stage 3, all independent variables are derived from the original data set variables which were coming as most important variables in the stage 2 analysis. The final set of 8 variables which were coming out to be the most important variables in determining the adjusted EPS of the next year are.

1. Momentum
2. FR_ROE
3. FR_Sales_to_CF_ratios
4. FR_Total_Debt_Mcap
5. VR_Adjusted_PE
6. FR_Dividend_Pay_Out_Ratio_P
7. FR_Sales_Working_Capital
8. FR_Interest_Cover

The derived variables were constructed based on the the following considerations.

- The 10-year slope of the above 7 variables except momentum
- Last 3-year average value of the above 8 variables except momentum
- 2018 value of the above 8 significant variables except momentum

This makes a total of 21 independent variables in addition to the 2018 momentum value which makes it 22 total independent variables. 2 approaches were taken in the 3rd stage of analysis, with a sample size of 51.

1. A regression analysis with above mentioned 22 independent variables and 2019 FR_Adjusted_EPS_Rs as the dependent variable.
2. A classification modelling with above mentioned 22 variables as predictor variables and 2019 EPS growth variable as the dependent variable.

Regression Model		Classification Model	
Independent variables	Dependent variable	Independent variables	Dependent variable
2018_Momentum	FR_Adjusted_EPS_Rs	2018_Momentum	2019 Adjusted EPS growth
FR_ROE_P_Slope		FR_ROE_P_Slope	
FR_ROE_P_3 yr avg		FR_ROE_P_3 yr avg	
FR_ROE_P_2018 value		FR_ROE_P_2018 value	
FR_Sales_to_CF_ratios_Slope		FR_Sales_to_CF_ratios_Slope	
FR_Sales_to_CF_ratios_3 yr avg		FR_Sales_to_CF_ratios_3 yr avg	
FR_Sales_to_CF_ratios_2018 value		FR_Sales_to_CF_ratios_2018 value	
FR_Total_Debt_Mcap_Slope		FR_Total_Debt_Mcap_Slope	
FR_Total_Debt_Mcap_3 yr avg		FR_Total_Debt_Mcap_3 yr avg	
FR_Total_Debt_Mcap_2018 value		FR_Total_Debt_Mcap_2018 value	
VR_Adjusted_PE_Slope		VR_Adjusted_PE_Slope	
VR_Adjusted_PE_3 yr avg		VR_Adjusted_PE_3 yr avg	
VR_Adjusted_PE_2018 value		VR_Adjusted_PE_2018 value	
FR_Dividend_Pay_Out_Ratio_P_Slope		FR_Dividend_Pay_Out_Ratio_P_Slope	
FR_Dividend_Pay_Out_Ratio_P_3 yr avg		FR_Dividend_Pay_Out_Ratio_P_3 yr avg	
FR_Dividend_Pay_Out_Ratio_P_2018 value		FR_Dividend_Pay_Out_Ratio_P_2018 value	
FR_Sales_Working_Capital_Slope		FR_Sales_Working_Capital_Slope	
FR_Sales_Working_Capital_3 yr avg		FR_Sales_Working_Capital_3 yr avg	
FR_Sales_Working_Capital_2018 value		FR_Sales_Working_Capital_2018 value	
FR_Interest_Cover_Slope		FR_Interest_Cover_Slope	
FR_Interest_Cover_3 yr avg		FR_Interest_Cover_3 yr avg	
FR_Interest_Cover_2018 value		FR_Interest_Cover_2018 value	

Table: CM6 - Stage 3 variable list

The below chart shows the distribution of variables.

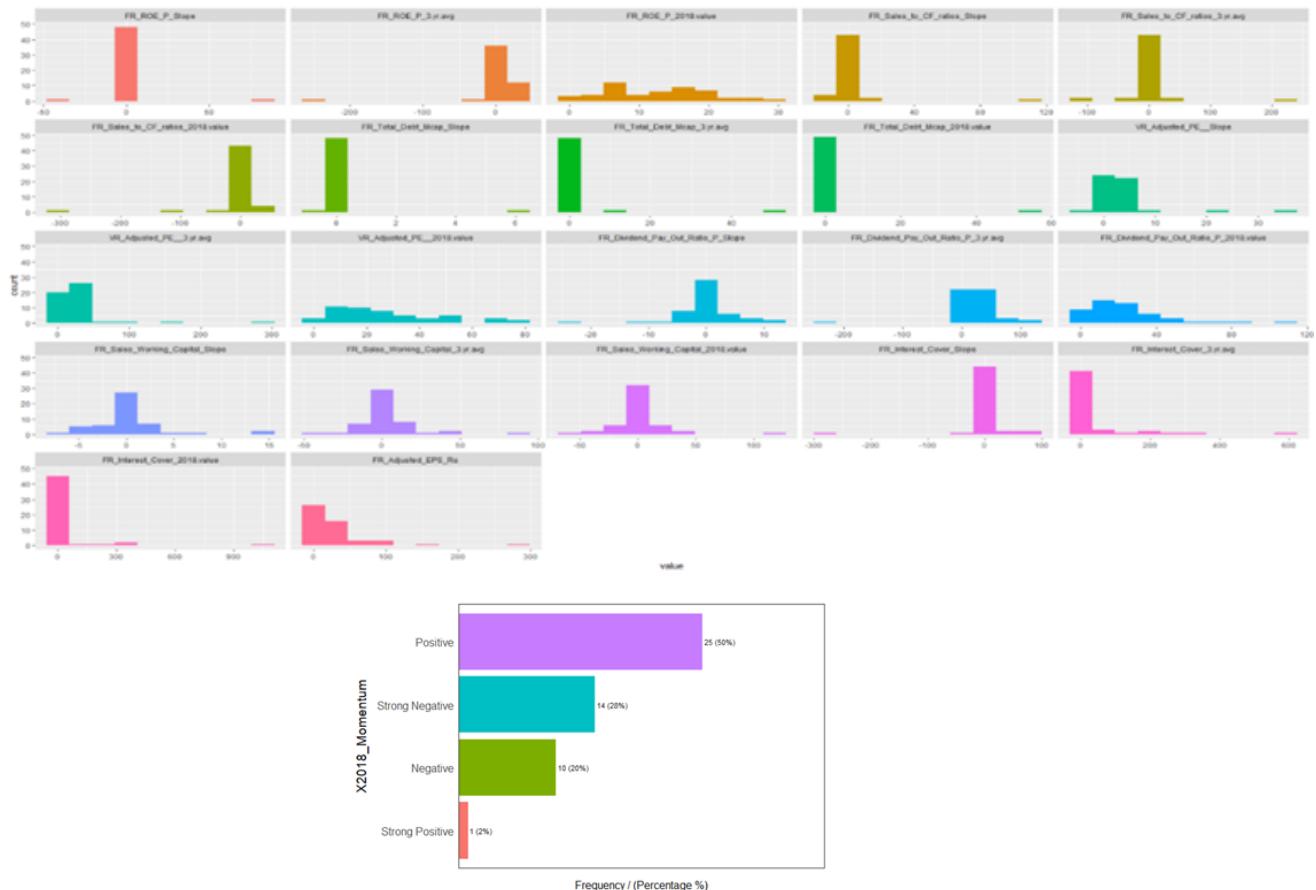


Fig: CM15 - Stage 3 variable distribution

The sample size is less (51), the variables from the above are mostly normally distributed and since they are financial growth slope and previous year averages,

Mean and median values of most variables fall in the similar range and other variables are assumed to be normally distributed for a larger data set. The extreme values are considered into the model.

There is fair amount of correlation and multicollinearity in the data between variables in the model. But majority of it is uncorrelated.

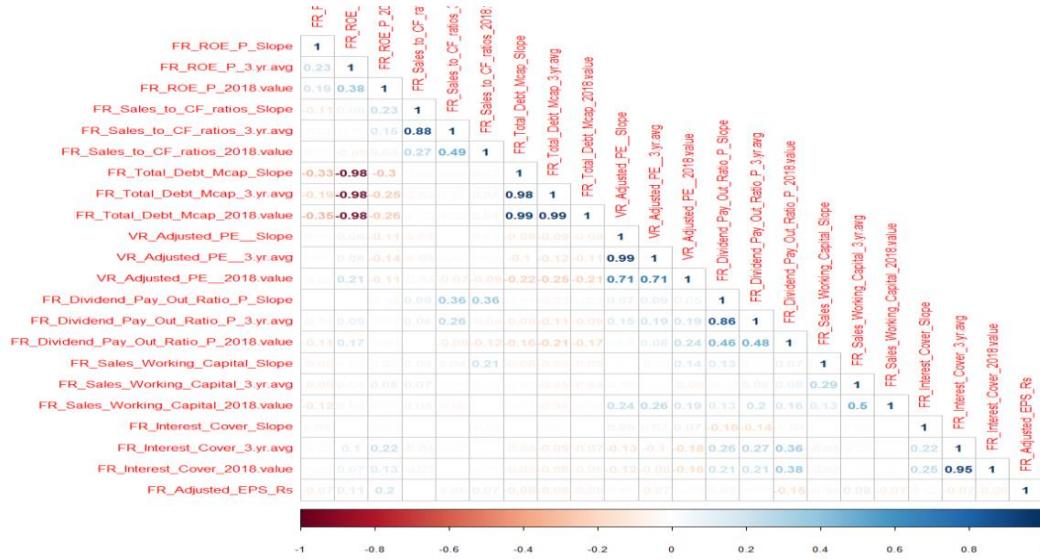


Fig: CM16 - Stage 3 variable Correlation

The assumptions of a regression or classification model are met or assumed to be met for the data set and we carry out Linear regression with all the variables as predictors and “Adjusted EPS in Rupees” for the year 2019 is considered as the dependent variable under the assumption that, the EPS value can be predicted using the significant variables (Considered as “FACTORS”) from the stage 2.

Coefficients:		Estimate	Std. Error	t value	Pr(> t)
(Intercept)		-63.2918	64.2327	-0.985	0.34244
X2018_MomentumPositive		70.1574	25.4476	2.757	0.01632 *
`X2018_MomentumStrong_Negative`		54.8856	32.2300	1.703	0.11235
`X2018_MomentumStrong_Positive`		4.8464	70.3471	0.069	0.94612
FR_ROE_P_Slope		-27.9436	11.9696	-2.335	0.03625 *
FR_ROE_P_3.yr.avg		5.8785	3.1394	1.873	0.08380
FR_ROE_P_2018.value		1.4539	2.5820	0.563	0.58297
FR_Sales_to_CF_ratios_Slope		-4.0178	4.1226	-0.975	0.34755
FR_sales_to_CF_ratios_3.yr.avg		1.9303	2.0467	0.943	0.36282
FR_sales_to_CF_ratios_2018.value		-0.5393	0.8704	-0.620	0.54618
FR_Total_Debt_Mcap_Slope		111.1740	135.0614	0.823	0.42527
FR_Total_Debt_Mcap_3.yr.avg		-71.4612	41.3683	-1.727	0.10775
FR_Total_Debt_Mcap_2018.value		62.6203	42.7001	1.467	0.16628
VR_Adjusted_PE_Slope		-2.2567	21.7007	-0.104	0.91876
VR_Adjusted_PE_3.yr.avg		0.3128	2.8220	0.111	0.91344
VR_Adjusted_PE_2018.value		-0.6238	1.0580	-0.590	0.56554
FR_Dividend_Pay_Out_Ratio_P_Slope		9.1165	11.9424	0.763	0.45887
FR_Dividend_Pay_Out_Ratio_P_3.yr.avg		0.4318	1.7940	0.241	0.81354
FR_Dividend_Pay_Out_Ratio_P_2018.value		-0.9881	1.0010	-0.987	0.34159
FR_sales_working_capital_Slope		0.3829	3.2799	0.117	0.90884
FR_sales_working_capital_3.yr.avg		1.4014	0.7198	1.947	0.07347
FR_sales_working_capital_2018.value		-0.7564	0.4974	-1.521	0.15232
FR_Interest_Cover_Slope		2.7757	1.5524	1.788	0.09708
FR_Interest_Cover_3.yr.avg		-1.5339	0.4815	-3.186	0.00716 **
FR_Interest_Cover_2018.value		0.5574	0.2089	2.668	0.01933 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 46.52 on 13 degrees of freedom
Multiple R-squared: 0.707, Adjusted R-squared: 0.1661
F-statistic: 1.307 on 24 and 13 DF, p-value: 0.3135

Fig: CM17 - Stage 3 Linear Regression model output

The adjusted R squared value is very less (16%) and the only significant variables for the model are Momentum, ROS_Slope, Interest Cover 3-year avg. and 2018 value. The model gives poor outputs with a high P value (0.31). The model is not statistically significant at 90% confidence level. We proceed with a Logistic regression with the same set of predictor variables and “Target growth for 2019” as the dependent variable. The outputs for the model are as given below

```
Coefficients:
(Intercept) -1.737e+02 3.024e-05 -0.001 1.000
X2018_MomentumPositive -3.988e+01 1.262e-05 0.000 1.000
`X2018_MomentumStrong_Negative` 9.393e+01 1.224e-05 0.001 0.999
`X2018_MomentumStrong_Positive` -4.747e+01 4.741e-05 0.000 1.000
FR_ROE_P_Slope -1.136e+01 2.760e-04 0.000 1.000
FR_ROE_P_3.yr.avg -2.721e+01 3.849e-04 -0.001 0.999
FR_ROE_P_2018.value 3.388e+01 3.676e-04 0.001 0.999
FR_Sales_to_CF_ratios_Slope -2.036e+00 2.181e-04 0.000 1.000
FR_Sales_to_CF_ratios_3.yr.avg -6.589e+00 1.377e-04 0.000 1.000
FR_Sales_to_CF_ratios_2018.value 3.406e+00 5.405e-03 0.001 0.999
FR_Total_debt_Mcap_Slope 2.189e+02 1.272e-06 0.000 1.000
FR_Total_debt_Mcap_3.yr.avg -1.982e+01 1.856e-05 0.000 1.000
FR_Total_debt_Mcap_2018.value 2.472e+01 3.716e-05 0.000 1.000
VR_Adjusted_PE_Slope -8.737e+00 1.263e-05 0.000 1.000
VR_Adjusted_PE_3.yr.avg -5.597e-01 1.739e-04 0.000 1.000
VR_Adjusted_PE_2018.value 3.425e+00 5.125e-03 0.001 0.999
FR_Dividend_Pay_Out_Ratio_P_slope -1.871e+01 4.485e-04 0.000 1.000
FR_Dividend_Pay_Out_Ratio_P_3.yr.avg 1.034e+01 9.319e-03 0.001 0.999
FR_Dividend_Pay_Out_Ratio_P_2018.value -5.357e+00 5.769e-03 -0.001 0.999
FR_Sales_working_capital_slope -1.875e+01 2.964e-04 -0.001 0.999
FR_Sales_working_capital_3.yr.avg -1.156e+01 1.071e-04 -0.001 0.999
FR_Sales_working_capital_2018.value 7.352e+00 6.423e-03 0.001 0.999
FR_Interest_cover_Slope 1.351e+00 1.361e-03 0.001 0.999
FR_Interest_cover_3.yr.avg -4.782e+00 3.662e-03 -0.001 0.999
FR_Interest_cover_2018.value 2.765e+00 2.259e-03 0.001 0.999

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4.9795e+01 on 35 degrees of freedom
Residual deviance: 3.5040e-09 on 11 degrees of freedom
AIC: 50

Number of Fisher scoring iterations: 25
```

Fig: CM18 - Stage 3 Logistic Regression model output

This model again gives a poor output with no significant variables. So, we come to the following assumptions based on the above model outputs.

- There is no statistical evidence to assume that the previous year trends of the “FACTORS” from the stage2 impact the Adjusted EPS or EPS growth
- There is no significant relationship between the previous years’ average values of the “FACTORS” and the Adjusted EPS or EPS growth
- The Previous year “FACTOR” values do not significantly impact the current year EPS or EPS growth

Findings Linked to Project Objectives

Based on the statistical analysis performed in the dataset related to the Commodities sector, following key financial ratios mapped to ratio categories:

Categories	Ratios
Valuation Ratios	Adjusted Price to Equity
Performance Ratios	Momentum Return on Equity. Interest cover Dividend payout ratio
Liquidity Ratios	Debt to Market Capitalization Sales to Working Capital Sales to cashflow ratio

Table: CM7 - Categories & Financial Ratios

Recommendations & Conclusions

Based on the quantitative research done on the Commodities macroeconomic sector companies, the following are the major conclusions and recommendations derived

For any investor who is investing the stock market in a commodities macroeconomic sector firm, before making any investment decisions, they should consider the following factors from the previous years' performance of the company.

- **Momentum:** The momentum of the stocks over the years is a key determining factor for investment decisions

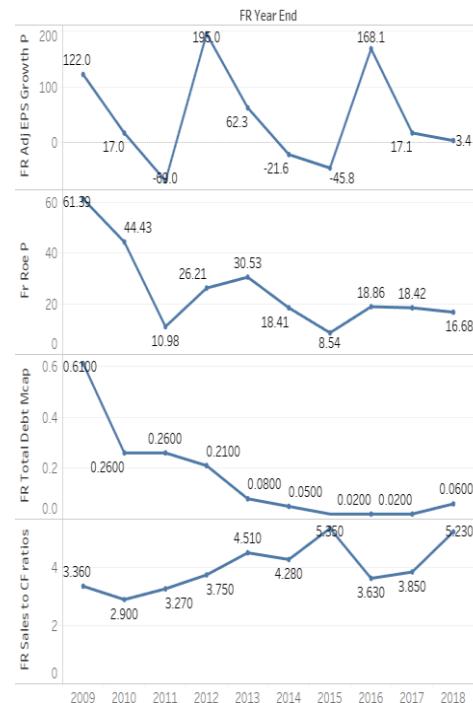


Fig: CM19 - Shree Cement trend of variables

- **Return on Equity:** Which is a directly correlated variable to the momentum of the stocks of any company. **Shree Cement** which a firm from the commodity macroeconomic sector shows the following trend for *Adjusted EPS growth, Return on Equity, Debt to market Cap ratio, and sales Free cash flow ratio*. We can clearly observe that these ratios have a high rate of correlation with the trend of Adjusted EPS growth
- **Total Debt to Market Capitalization Ratio** a solvency measure that shows the proportion of debt a company uses to finance its assets, relative to the amount of equity used for the same purpose. A higher ratio result means that a company is more highly leveraged, which carries a higher risk of insolvency.
- **Sales to Working Capital Ratio** is a ratio measuring the efficiency of company's working capital utilization to generate the certain level of sales. This factor significantly impacts the EPS growth of a company from Commodities sector.
- The **Interest Coverage Ratio** is used to determine how easily a company can pay its interest expenses on outstanding debt. The ratio is the company's earnings before interest and taxes (EBIT) to the company's interest expenses for the same period. This ratio is related to the performance factor of a company and significantly impacts the growth of a commodities macroeconomic company.
- The **Cash Flow to Sales Ratio** determines the ability of a business to generate cash flow in proportion to its sales volume. This can determine the growth of a company significantly.
- **Dividend pay-out Ratio** is a direct indicator of the financial health of a company in terms of earnings for the investors.

- **Adjusted Price to Equity Ratio** is calculated by dividing a company's stock price by the average of the company's earnings for the last ten years, adjusted for inflation. This ratio is significant in predicting the growth of a commodities macroeconomic sector company.

2.4 Manufacturing

Data Summary

Dimension – 761(rows) x 59 (variables) (including Target variable)
 70 companies in the data set spread across 5 industry for Manufacturing Sector

Exploratory Data Analysis:

Univariate Analysis: Univariate Analysis was performed using the data-explorer package on the full data

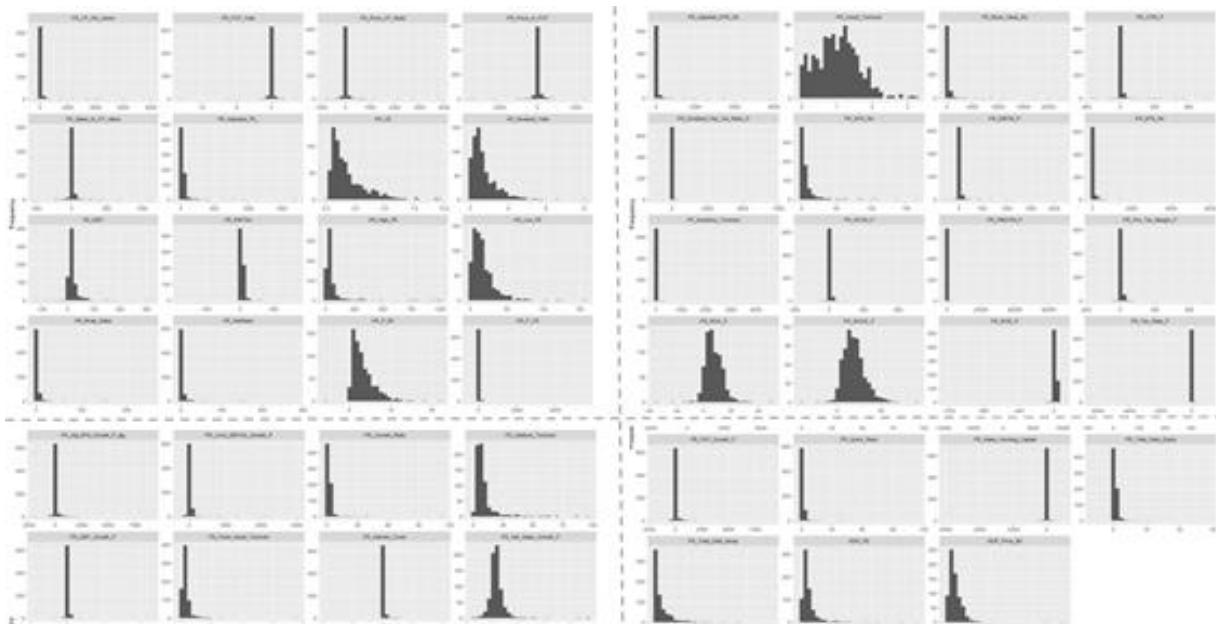


Fig: M1- Univariate Analysis of the data

As seen from Fig: M1 there are violations of normality from the histogram plots above for a lot of variables under consideration, but we are going to include these in the model building without any treatment. Perspective of Adjusted EPS by industry and companies is provided under Fig: M3

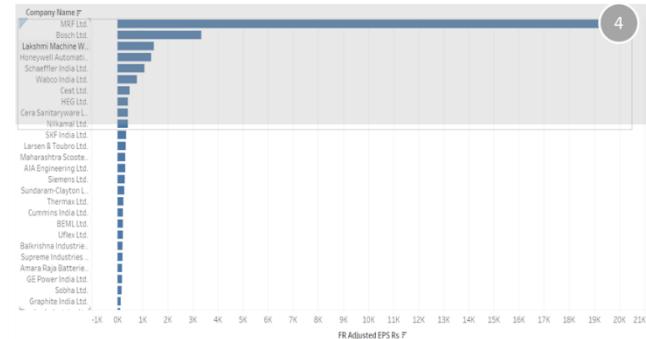
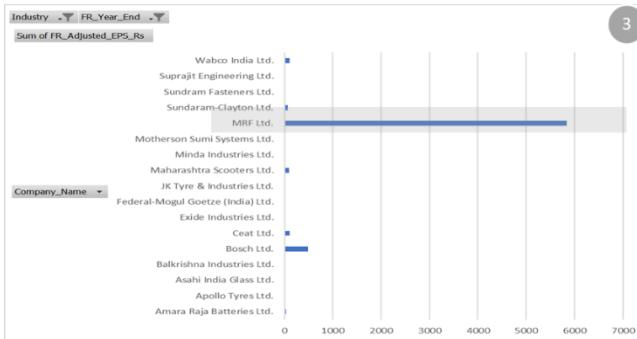
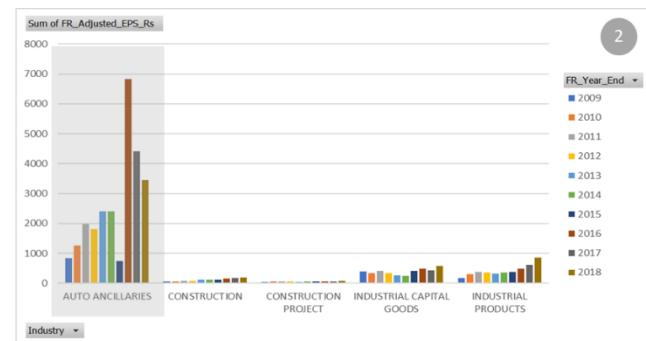
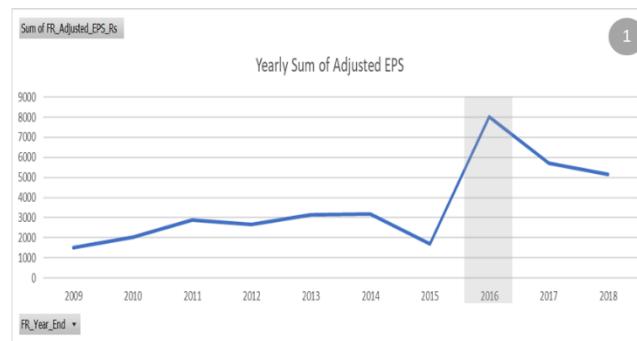


Fig: M2.1, 2.2, 2.3, 2.4

- There is a level shift observed post 2015 and although there is a small dip in the Adjusted EPS value in 2017 and 2018 it is higher than the previous years. (Fig: M2.1)
- This surge is led by the Auto Ancillaries Industry type (Fig:M2.2)
- A further drill down reveals that MRF was the one company that led to this level shifting (Fig: M2.3)
- A view of the top 10 companies in terms of Adjusted Earnings per share reveals that MRF's value is greater by a high margin, followed by Bosch and Lakshmi Works Ltd. (Fig: M2.4)

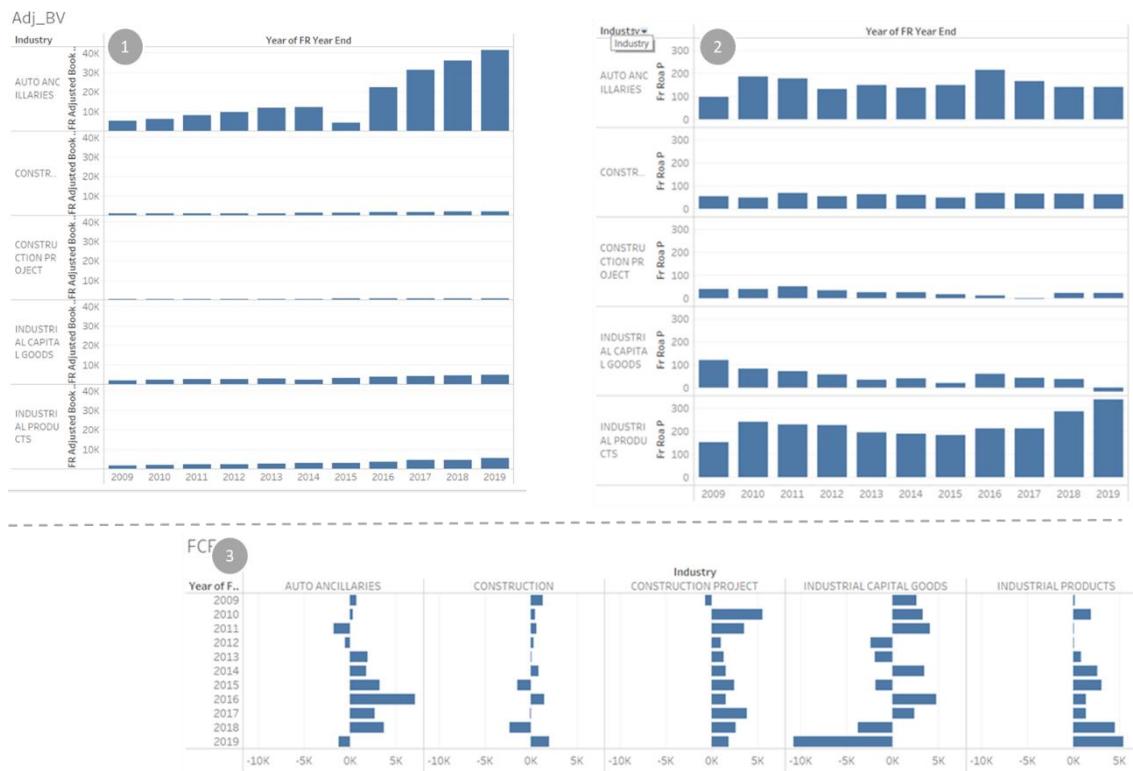


Fig: M3.1,3.2,3.3- EDA of various variables

- Furthermore, the FR_Adjusted_Book_Value_Rs for Auto Ancillaries grew overall greater over the years than any other industry type (Fig: M3.1)
- While FR_ROA_P increased through the years for Industrial products, whereas others were fairly stable. (Fig: M3.2)
- FH_FCF had one big fall in 2019 mainly for Industrial Capital goods, larger than any other categories when the average values were consistent. (Fig M3.3)

Correlation Analysis

A Correlation was run on the numeric values of the dataset. The results are displayed by High Correlation, Medium and Low Correlation values

Variable 1	Variable 2	Value	Variable 1	Variable 2	Value
FR_PATM_P	FR_CPM_P	0.9994	FR_Fixed_Asset_Turnover	FR_Quick_Ratio	0.5787
FR_EPS_Rs	FR_Adjusted_EPS_Rs	0.9989	VR_Adjusted_PE	VR_High_PE	0.5504
FR_Pre_Tax_Margin_P	FR_PATM_P	0.9984	FR_Asset_Turnover	FR_ROA_P	0.5358
FR_Current_Ratio	FR_Quick_Ratio	0.9961	FR_DPS_Rs	FR_Book_Value_Rs	0.5023
FR_CF_Per_share	FR_EPS_Rs	0.9648	VR_EBITDA	FR_ROE_P	0.4949
VR_NetSales	VR_Mcap_Sales	0.9607	NDP_PE	FR_Quick_Ratio	0.4941
FR_Book_Value_Rs	FR_CF_Per_share	0.9128	VR_EBIT	VR_Adjusted_PE	0.4817
FR_Dividend_Pay_Out_Ratio_P	FR_Adj_EPS_Growth_P_lag	0.8760	FR_Price_CF_Ratio	FR_Sales_to_CF_ratios	0.4762
FR_Adj_EPS_Growth_P_lag	FR_Dividend_Pay_Out_Ratio_P	0.8760	VR_Low_PE	NDP_PE	0.4504
FR_ROA_P	FR_ROCE_P	0.8389	FR_Core_EBITDA_Growth_P	FR_Dividend_Pay_Out_Ratio_P	0.4439
FR_Adj_EPS_Growth_P_lag_50_pc	FR_Adj_EPS_Growth_P_lag_60_pc	0.8193	FR_Total_Debt_Equity	FR_Total_Debt_Mcap	0.3388
VR_P_BV	VR_CE	0.8060	FR_EBIT_Growth_P	FR_ROE_P	0.3268
FR_EBITM_P	VR_NetSales	0.8051	FR_Net_Sales_Growth_P	FR_Core_EBITDA_Growth_P	0.3150
FR_Adj_EPS_Growth_P_lag_75_pc	FR_Adj_EPS_Growth_P_lag_60_pc	0.7050	VR_P_CE	VR_EBIT	0.2436
NDP_Price_BV	VR_P_BV	0.6703	Mo_Positive	VR_CE	0.2309
FR_Adj_EPS_Growth_P_lag_90_pc	FR_Adj_EPS_Growth_P_lag_75_pc	0.5833	FR_PBIDTM_P	VR_High_PE	0.2265

FR_PATM_P	FR_Quick_Ratio	0.5787
VR_High_PE	VR_High_PE	0.5504
FR_ROA_P	FR_ROA_P	0.5358
FR_Book_Value_Rs	FR_Book_Value_Rs	0.5023
FR_ROE_P	FR_ROE_P	0.4949
FR_Quick_Ratio	FR_Quick_Ratio	0.4941
VR_Adjusted_PE	VR_Adjusted_PE	0.4817
FR_Sales_to_CF_ratios	FR_Sales_to_CF_ratios	0.4762
NDP_PE	NDP_PE	0.4504
FR_Dividend_Pay_Out_Ratio_P	FR_Dividend_Pay_Out_Ratio_P	0.4439
FR_Total_Debt_Equity	FR_Total_Debt_Mcap	0.3388
FR_EBIT_Growth_P	FR_ROE_P	0.3268
FR_Net_Sales_Growth_P	FR_Core_EBITDA_Growth_P	0.3150
VR_P_CE	VR_EBIT	0.2436
Mo_Positive	VR_CE	0.2309
FR_PBIDTM_P	VR_High_PE	0.2265
FR_PATM_P	VR_High_PE	0.2144
VR_Dividend_Yield	FR_Total_Debt_Mcap	0.2124
FR_Interest_Cover	FR_DPS_Rs	0.1616
Mo_Strong_Negative	VR_High_PE	0.1613
FR_Inventory_Turnover	FR_Fixed_Asset_Turnover	0.1552
FR_Price_to_FCF	VR_Low_PE	0.1252
Mo_Strong_Positive	FR_ROA_P	0.1227
FR_Debtors_Turnover	FR_EBIT_Growth_P	0.1142
FR_FCF_Yield	FR_ROA_P	0.1086
Mo_Negative	FR_Dividend_Pay_Out_Ratio_P	0.0885
FR_Tax_Rate_P	FR_ROCE_P	0.0613
FR_Sales_Working_Capital	VR_P_BV	0.0456

Fig: M4

- The Fig: M5 is divided into 3 parts, variable pairs having high, medium and low correlation. For display purposes, only the highest variable correlation is listed.
- There may be multicollinearity For example, FR_EPS_Rs is highly correlated with FR_Adjusted_EPS_Rs and FT_EPS_Rs is highly correlated with FR_CF_Per_Share.
- Further analysis to remove multicollinearity will be done in the modeling exercise.

Statistical Tools and Techniques

The analysis and modelling were divided into three stages as a part of the project: -

- Stage 1: Statistical Analysis of the variables and its impact on the Adjusted EPS for the current Year
- Stage 2: Statistical analysis of the Variables and its role in PREDICTING the Growth Category of a firm for the next year
- Stage 3: Based on the significant variables derived from the Stage 2, determining the impact of the time, leveraging slope, 3-Year average and current year's data on the Growth of the firm for the next year

Stage 1: Statistical Analysis of the variables and its impact on the Adjusted EPS for the current Year

Following statistical techniques were leveraged for Stage 1: -

- Linear Regression
- Lasso

- Ridge
- Random Forest
- Support Vector Machine
- Other feature selection algorithms like Boruta, Genetic Algorithms and Simulated Annealing.

Model Types	RMSE
Full Model	1.99
Linear Regression Model (Variables selected with Blackbox approach)	.26
Linear Regression Model post selecting significant variables	.26
Random Forest	.87

Fig: M5

Observations (based on Fig: M5):

- The linear regression model explains the variation in the target variable better, therefore it's the model of choice here.
- The model with variables selected using the black box approach and via the traditional methods perform on par.

Final List of variables that were significant.

- Earnings before interest, tax, depreciation, amortization
- Dividend per share
- Adjusted Book Value
- Profit After Tax
- Price to Earnings Ratio
- Receivable Days
- Payable days

Stage 1 Final Conclusion: Owing to simplicity in process, compute and explainability of the model, the linear regression model with significant variables selected using statistical significance is the final choice.

Stage 2: Statistical analysis of the Variables and its role in PREDICTING the Growth Category of a firm for the next year

The following models have been created to predict the Adjusted EPS Growth for the next year.

- Logistic Regression
- Random Forest
- Rpart
- Boruta

The Y variable based on 60th percentile of the EPS Growth percentage was selected based on model metrics (Fig M6) using test and train data.

Model	Train				Test			
	AUC	Sensitivity	Specificity	Accuracy	AUC	Sensitivity	Specificity	Accuracy
Logit	63%	66%	67%	67%	68%	78%	71%	73%
Rpart	77%	81%	79%	79%	59%	54%	66%	62%
RF	100%	100%	100%	100%	72%	75%	74%	75%

Fig: M6

Model Outputs

Logistic Regression based on all variables.

Model Output

```

Deviance Residuals:
    Min      1Q Median      3Q     Max
-2.3303 -0.8751 -0.4750  0.9630  2.4334

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -9.976e-01  8.770e-01 -1.138  0.255292
VR_Adjusted_PE_ 6.770e-02  7.899e-03  0.889  0.395652
VR_Cap_Sales 3.996e-02  1.979e-02  1.020  0.062936
VR_P_Cap 9.221e-02  1.903e-01  4.844  0.628052
VR_Dividend_Yield -1.436e-01  1.408e-01 -1.020  0.307760
VR_NetSales 4.886e-02  2.020e-02  2.414  0.047442 *
VR_Earnings 5.312e-03  7.924e-03  0.645  0.488826
VR_EBIT 2.331e-03  9.311e-03 -0.250  0.802327
VR_CE -1.746e-01  2.727e-01 -0.640  0.522057
VR_Mcap_Sales 6.801e-02  1.979e-02  3.457  0.000446 *
VR_High_PE 1.022e-02  1.457e-02  0.701  0.483072
VR_Low_PE -2.828e-02  1.260e-02 -2.241  0.024794 =
FR_CF_Per_share 2.286e-04  9.652e-04  0.231  0.812752
FR_Cap_to_CF_Ratio -1.414e-01  1.370e-01  1.024  0.318838
FR_Price_to_CFC 1.023e-05  3.522e-04 -0.029  0.976910
FR_FCF_Yield -1.550e-01  1.370e-01 -1.113  0.909971
FR_Sales_to_CF_ratios 1.846e-03  2.140e-03  0.865  0.388222
FR_Adjusted_EPS_Rs 2.913e-02  1.451e-02  2.023  0.127139
FR_DPS_Rs 2.213e-02  1.451e-02  1.523  0.127139
FR_Book_Value_Rs 5.058e-06  2.348e-06  0.022  0.982713
FR_Ax_Rate_P 2.090e-04  2.689e-03  0.078  0.938055
FR_Dividend_Pay_Out_Ratio_P 1.336e-04  2.020e-04  0.212  0.850585 =
FR_EBITM_P 5.414e-03  3.673e-02  0.147  0.882822
FR_Pre_Tax_Margin_P 2.594e-02  6.720e-02  0.387  0.699515
FR_ROE_P -2.020e-01  1.923e-01  1.023  0.318838
FR_CPM_P 1.533e-01  7.962e-02  1.926  0.054098 =
FR_ROA_P -1.791e-01  6.614e-02 -2.709  0.006778
FR_ROE_P -9.170e-02  7.793e-03 -1.187  0.300006
FR_DCF_P 4.674e-02  2.753e-02  1.695  0.090001 =
FR_Asset_Turnover 1.021e+00  3.497e-01  2.920  0.003504 ==
FR_Inventory_Turnover -2.402e-03  4.023e-03 -0.597  0.550553
FR_Debtors_Turnover 9.715e-02  1.025e-02  0.941  0.359722
FR_Fixed_Asset_Turnover 5.444e-02  4.028e-02  1.352  0.166482
FR_Sales_Working_Capital -6.175e-04  9.634e-04 -0.643  0.521557
FR_Net_Sales_Growth_P 1.862e-03  4.132e-03  0.451  0.652248
FR_EBIT_Growth_P 8.460e-04  1.152e-04  0.369  0.565757
FR_EBIT_Growth_DL_Growth_P 4.860e-04  8.544e-04  0.369  0.569478
FR_PAT_P 5.788e-04  7.756e-04 -0.743  0.455508
FR_Total_Debt_Equity 9.282e-02  1.121e-01  0.828  0.407801
FR_Debt_Equity_Ratio 2.270e-02  3.121e-02  0.704  0.300073
FR_Quick_Ratio -2.549e-01  3.666e-01 -0.695  0.366970
FR_Interest_Cover 8.732e-05  5.283e-01  0.165  0.868721
FR_Total_Debt_Mcap -6.270e-02  1.897e-01 -0.331  0.741016
NDP_BV 8.210e-02  9.392e-02  0.874  0.141414
NDP_Price_BV -6.309e-02  1.252e-01 -0.504  0.614342
Mo_Positive -7.040e-01  4.682e-01 -1.504  0.132664
Mo_Strong_Positive' -6.180e-00  4.427e-01 -1.380  0.167674
Mo_Negative -2.600e-00  2.160e-01  1.200  0.090000 ===
Mo_Strong_Negative' -1.018e+00  4.991e-01 -2.040  0.041319 =
```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 654.41 on 484 degrees of freedom
Residual deviance: 513.26 on 435 degrees of freedom
AIC: 613.26
Number of Fisher Scoring iterations: 9

```

##### Variables and VIF

| Variable             | VIF     |
|----------------------|---------|
| VR_NetSales          | 1511.63 |
| FR_PATM_P            | 983.74  |
| VR_Mcap_Sales        | 939.99  |
| FR_CPM_P             | 678.14  |
| FR_Pre_Tax_Margin_P  | 394.08  |
| FR_EBITM_P           | 250.63  |
| FR_Quick_Ratio       | 204.78  |
| FR_Current_Ratio     | 196.31  |
| FR_CF_Per_share      | 26.54   |
| VR_P_BV              | 20.34   |
| FR_Adjusted_EPS_Rs   | 16.41   |
| FR_Book_Value_Rs     | 11.83   |
| FR_CE                | 11.65   |
| NDP_Price_BV         | 10.87   |
| FR_ROA_P             | 10.66   |
| FR_ROCE_P            | 7.87    |
| FR_ROE_P             | 4.42    |
| NDP_PE               | 4.28    |
| VR_Adjusted_PE_      | 3.84    |
| 'Mo_Strong Positive' | 3.62    |

##### Confusion Matrix

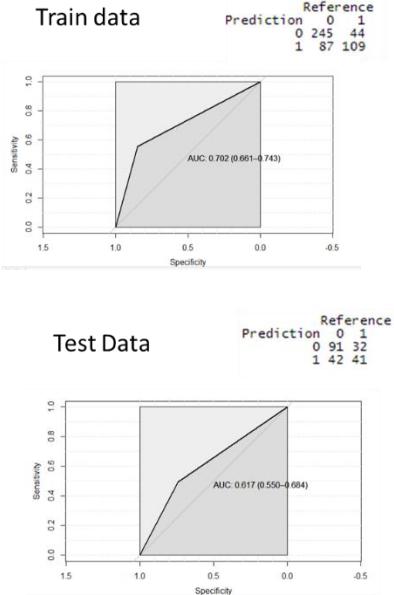


Fig: M7

The significant variables based on the Full Model are as follows (Fig M7): -

- Mo\_Negative,
- FR\_Asset\_Turnover
- FR\_ROA\_P
- 'Mo\_Strong Negative'
- FR\_PATM\_P
- VR\_Low\_PE
- VR\_Mcap\_Sales
- VR\_NetSales

##### Multi-Collinearity:

- There exists a high degree of multicollinearity based on the Variation inflation factor (VIF) as shown in the sample below (Fig: M8)
- Next Step is to recursively remove one variable from highest variable pair and re-run the model until the VIF ~10

- Using the above process, VR\_NetSales, FR\_PATM\_P, FR\_Pre\_Tax\_Margin\_P, FR\_Quick\_Ratio, FR\_CF\_Per\_share, VR\_P\_BV were removed.
- The table on the 2<sup>nd</sup> half of Fig M8 shows the final VIFs post removing all the multicollinear variables.

| Variable             | VIF      | Variable                | VIF  |
|----------------------|----------|-------------------------|------|
| VR_NetSales          | 1,511.63 | VR_Mcap_Sales           | 8.78 |
| FR_PATM_P            | 983.74   | FR_ROA_P                | 8.06 |
| VR_Mcap_Sales        | 939.99   | FR_Book_Value_Rs        | 7.46 |
| FR_CPM_P             | 678.14   | FR_Adjusted_EPS_Rs      | 6.87 |
| FR_Pre_Tax_Margin_P  | 394.08   | NDP_Price_BV            | 6.21 |
| FR_EBITM_P           | 250.63   | VR_CE                   | 5.80 |
| FR_Quick_Ratio       | 204.78   | FR_EBITM_P              | 5.60 |
| FR_Current_Ratio     | 196.31   | FR_ROCE_P               | 5.58 |
| FR_CF_Per_share      | 26.54    | NDP_PE                  | 4.24 |
| VR_P_BV              | 20.34    | VR_Adjusted_PE_         | 4.09 |
| FR_Adjusted_EPS_Rs   | 16.41    | FR_Current_Ratio        | 3.74 |
| FR_Book_Value_Rs     | 11.83    | `Mo_Strong Positive`    | 3.56 |
| VR_CE                | 11.65    | Mo_Positive             | 3.40 |
| NDP_Price_BV         | 10.87    | FR_ROE_P                | 3.15 |
| FR_ROA_P             | 10.66    | `Mo_Strong Negative`    | 2.70 |
| FR_ROCE_P            | 7.87     | FR_Price_CF_Ratio       | 2.63 |
| FR_ROE_P             | 4.42     | FR_CPM_P                | 2.63 |
| NDP_PE               | 4.28     | FR_Core_EBITDA_Growth_P | 2.62 |
| VR_Adjusted_PE_      | 3.84     | Mo_Negative             | 2.46 |
| `Mo_Strong Positive` | 3.62     |                         |      |

Fig: M8

## Logistic Regression post removing all multicollinear variables

### Model Output

```

Deviance Residuals:
 Min Q1 Median Q3 Max
-2.1003 -0.9362 -0.6639 1.1735 2.1725

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.814e-02 6.306e-01 -0.092 0.92654
Mo_Negative -1.531e+00 4.939e-01 -3.100 0.00193 ***
`Mo_Strong Negative` -1.113e+00 4.706e-01 -2.365 0.01804 ***
FR_ROA_P -1.569e-01 5.163e-02 -3.039 0.00237 ***
FR_Asset_Turnover 5.169e-01 2.132e-02 2.437 0.05075
`Mo_Strong Positive` -2.177e-01 1.832e-01 -1.176 0.08621 *
Mo_Positive -8.217e-01 4.531e-01 -1.814 0.06974
FR_ROCE_P 4.912e-02 2.033e-02 2.416 0.01569 **
FR_EBITM_P -7.408e-03 3.105e-03 -2.386 0.01702 **
FR_Price_CF_Ratio -1.180e-03 8.740e-04 -1.348 0.15038
VR_Mcap_Sales 3.220e-02 2.398e-02 1.438 0.15038
FR_Core_EBITDA_Growth_P 3.656e-03 1.234e-03 2.963 0.00305 ***
NDP_Price_BV -2.160e-01 8.606e-02 -2.510 0.01208 **
FR_Current_Ratio 1.072e-02 4.554e-02 0.235 0.81385
VR_Low_PE 7.188e-03 2.033e-02 0.359 0.73955
FR_Fixed_Asset_Turnover -3.636e-02 3.564e-02 -0.388 0.69800
FR_ROE_P 4.178e-03 1.663e-02 0.251 0.80165
FR_Dividend_Pay_Out_Ratio_P 4.830e-03 5.147e-03 0.938 0.34800
VR_P_CE 1.772e-02 8.180e-03 2.120 0.03398 **
VR_Adjusted_P_E 5.186e-03 2.098e-02 0.253 0.69735
VR_PBIIDTM_P -5.585e-05 2.949e-05 -1.894 0.05825 *
VR_Adjusted_P_E_ 3.358e-03 4.680e-03 0.717 0.47308
NDP_P_E 3.996e-03 5.981e-03 0.668 0.50407
FR_Sales_to_CF_ratios 1.631e-03 2.334e-03 0.699 0.48464
VR_High_P_E 3.167e-03 1.032e-02 0.307 0.75898
FR_Debtors_Turnover -3.846e-04 5.799e-04 -0.663 0.50719
FR_Sales_Working_Capital -9.733e-03 1.130e-01 -0.086 0.93149
VR_Dividend_Yield -1.109e-03 2.218e-03 -0.504 0.69735
FR_Tax_Rate_P 2.286e-03 1.032e-03 0.267 0.44330
FR_Asset_Turnover 4.659e-04 4.595e-04 1.014 0.31068
VR_High_P_E 1.483e-03 1.390e-03 1.067 0.28614
FR_Book_Value_Rs -3.506e-04 4.245e-04 -0.826 0.40878
VR_CE 1.119e-01 1.760e-01 0.634 0.52580
FR_Total_Debt_Equity -5.231e-03 9.663e-03 -0.541 0.58628
FR_DPS_Rs 5.428e-05 1.604e-04 0.339 0.73498
FR_Interest_Cover 1.776e-03 2.207e-03 0.805 0.42088
FR_Adjusted_EPS_Rs -5.731e-04 7.136e-03 -0.731 0.46099
FR_GIT_Growth_P 3.616e-03 1.142e-03 1.368 0.27279
FR_Total_Debt_Mcap 6.592e-02 1.922e-01 0.343 0.73163
FR_EBIT_Growth_P -6.257e-04 8.941e-04 -0.700 0.48404
FR_FCF_Yield 1.452e-01 2.312e-01 0.628 0.52994
FR_Net_Sales_Growth_P -5.138e-03 4.148e-03 -1.239 0.21544

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Dispersion parameter for binomial family taken to be 1

Null deviance: 654.41 on 484 degrees of freedom
Residual deviance: 565.31 on 441 degrees of freedom
AIC: 653.31

Number of Fisher Scoring iterations: 7

```

### Top 20 Variables

| 60pc                        | Overall |
|-----------------------------|---------|
| Mo_Negative                 | 4.13    |
| `Mo_Strong Negative`        | 3.74    |
| FR_ROA_P                    | 3.38    |
| FR_Asset_Turnover           | 3.26    |
| `Mo_Strong Positive`        | 2.75    |
| Mo_Positive                 | 2.64    |
| FR_ROCE_P                   | 2.28    |
| FR_EBITM_P                  | 2.20    |
| FR_Price_CF_Ratio           | 1.92    |
| VR_Mcap_Sales               | 1.88    |
| FR_Core_EBITDA_Growth_P     | 1.85    |
| NDP_Price_BV                | 1.69    |
| FR_Current_Ratio            | 1.68    |
| VR_EBITDA                   | 1.65    |
| FR_Fixed_Asset_Turnover     | 1.64    |
| FR_ROE_P                    | 1.52    |
| FR_Dividend_Pay_Out_Ratio_P | 1.48    |
| VR_P_CE                     | 1.34    |
| VR_Low_PE                   | 1.27    |
| FR_PBIIDTM_P                | 1.21    |

### Confusion Matrix

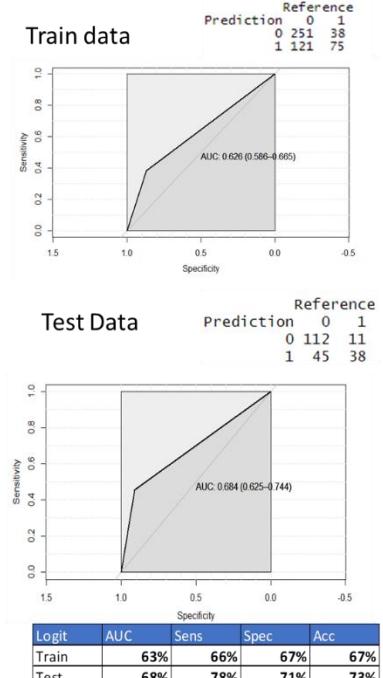


Fig: M9

The significant variables post removing multicollinear variables as per Fig: M9 is as follows

- Mo\_Negative
- `Mo\_Strong Negative`
- FR\_Asset\_Turnover
- FR\_Dividend\_Pay\_Out\_Ratio\_P
- VR\_Low\_PE
- VR\_Mcap\_Sales
- VR\_NetSales

Also there is an improvement in predicting on the test data on all model metrics. This implies that there is no overfitting of the model (Fig: M9)

## Classification Tree

### Model Output

```
Variables actually used in tree construction:
[1] FR_Dividend_Pay_Out_Ratio_P FR_EBITM_P
[5] FR_Price_to_FCF FR_ROA_P
[9] FR_Total_Debt_Mcap NDP_Price_BV

FR_Fixed_Asset_Turnover FR_PBIDTM_P
FR_ROCE_P FR_Tax_Rate_P
VR_Low_PE

Root node error: 196/485 = 0.40412
n= 485

CP nsplit rel error xerror xstd
1 0.229592 0 1.00000 1.00000 0.055138
2 0.033163 1 0.77041 0.88265 0.053824
3 0.030612 3 0.70408 0.93878 0.054521
4 0.020408 5 0.64286 0.93367 0.054463
5 0.012755 9 0.54592 0.97449 0.054899
6 0.010204 11 0.52041 0.99490 0.055092
7 0.010000 12 0.51020 0.99490 0.055092
```

### Top 20 Variables

| Variable Name               | Score |
|-----------------------------|-------|
| FR_ROA_P                    | 50.99 |
| FR_Tax_Rate_P               | 40.60 |
| FR_EBITM_P                  | 39.62 |
| FR_Pre_TaxMargin_P          | 33.87 |
| FR_CPM_P                    | 31.92 |
| VR_P_CE                     | 16.57 |
| FR_Dividend_Pay_Out_Ratio_P | 15.10 |
| FR_ROE_P                    | 14.88 |
| VR_AdjustedPE_              | 12.52 |
| FR_Interest_Cover           | 9.42  |
| FR_CF_Per_Share             | 6.03  |
| FR_Price_CF_Ratio           | 5.92  |
| FR_Sales_Working_Capital    | 5.75  |
| VR_EBIT                     | 5.57  |
| VR_LowPE                    | 5.34  |
| FR_ROCE_P                   | 5.05  |
| FR_Book_Value_Rs            | 3.78  |
| FR_Total_Debt_Mcap          | 2.82  |
| FR_DPS_Rs                   | 2.51  |
| FR_Inventory_Turnover       | 2.27  |

### Confusion Matrix

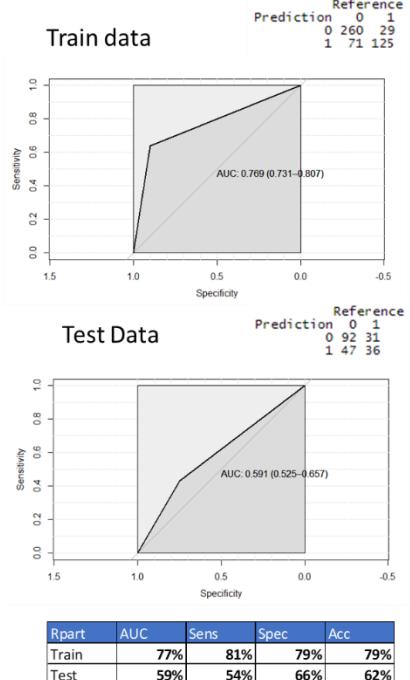


Fig: M10

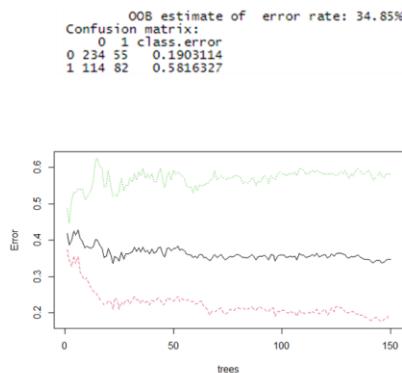
The model was run with minsplit= 20 and minbucket =10 and maxdepth =20. Some of the significant variables are as follows (Fig: M10)

- FR\_ROA\_P
- FR\_Tax\_Rate\_P
- FR\_EBITM\_P
- FR\_Pre\_TaxMargin\_P
- FR\_CPM\_P
- VR\_P\_CE
- FR\_Dividend\_Pay\_Out\_Ratio\_P
- FR\_ROE\_P
- FR\_AdjustedPE\_
- FR\_Interest\_Cover
- FR\_CF\_Per\_Share
- FR\_Price\_CF\_Ratio
- FR\_Sales\_Working\_Capital
- VR\_EBIT
- VR\_LowPE
- FR\_ROCE\_P
- FR\_Book\_Value\_Rs
- FR\_Total\_Debt\_Mcap
- FR\_DPS\_Rs
- FR\_Inventory\_Turnover

Also as seen from the model metrics there is a fair bit of difference in the train and test model metrics, but this does not imply overfitting of the model. (Fig: M10)

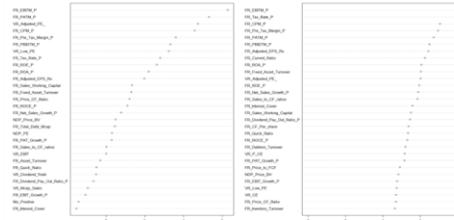
## Random Forest

### Model Output



### Top 20 Variables and Variable importance plot

| Variable Name            | 0        | 1        |
|--------------------------|----------|----------|
| FR_Tax_Rate_P            | 5.270641 | 5.270641 |
| FR_EBITM_P               | 3.060246 | 3.060246 |
| FR_Adjusted_EPS_Rs       | 3.012169 | 3.012169 |
| FR_PATM_P                | 2.72488  | 2.72488  |
| FR_Pre_Tax_Margin_P      | 2.688084 | 2.688084 |
| FR_Fixed_Asset_Turnover  | 2.651835 | 2.651835 |
| FR_ROA_P                 | 2.554999 | 2.554999 |
| FR_PBIDTM_P              | 2.283482 | 2.283482 |
| FR_ROE_P                 | 2.13324  | 2.13324  |
| FR_EBIT_Growth_P         | 1.943503 | 1.943503 |
| FR_CPM_P                 | 1.916333 | 1.916333 |
| VR_P_CE                  | 1.906015 | 1.906015 |
| VR_Adjusted_PE           | 1.905132 | 1.905132 |
| FR_Total_Debt_Equity     | 1.828213 | 1.828213 |
| FR_Price_CF_Ratio        | 1.790613 | 1.790613 |
| FR_ROCE_P                | 1.763503 | 1.763503 |
| VR_EBIT                  | 1.597859 | 1.597859 |
| FR_Sales_Working_Capital | 1.508883 | 1.508883 |
| VR_Mcap_Sales            | 1.392884 | 1.392884 |
| VR_High_PE               | 1.377306 | 1.377306 |



### Confusion Matrix

|            |  | Reference                |       |
|------------|--|--------------------------|-------|
|            |  | Prediction               | 0 1   |
| Train data |  | 0                        | 289 0 |
|            |  | 1                        | 0 196 |
|            |  | Sensitivity              |       |
|            |  | 1.0                      | 0.0   |
|            |  | 0.5                      | 0.5   |
|            |  | 0.0                      | 1.0   |
|            |  | AUC: 1.000 (1.000-1.000) |       |
|            |  | Specificity              | -0.5  |
| Test Data  |  | Reference                |       |
|            |  | Prediction               | 0 1   |
|            |  | 0                        | 289 0 |
|            |  | 1                        | 0 196 |
|            |  | Sensitivity              |       |
|            |  | 1.0                      | 0.0   |
|            |  | 0.5                      | 0.5   |
|            |  | 0.0                      | 1.0   |
|            |  | AUC: 0.716 (0.655-0.777) |       |
|            |  | Specificity              | -0.5  |

Fig: M11

The random forest model was run with the ntree=150 and mtry=10. Other models with combination of ntree = 250 and mtry =5 and 15 were tried. This model had relatively better performance.( Fig: M11)

The variable importance from this model were (top 10)

- FR\_Tax\_Rate\_P
- FR\_EBITM\_P
- FR\_Adjusted\_EPS\_Rs
- FR\_PATM\_P
- FR\_Pre\_Tax\_Margin\_P
- FR\_Fixed\_Asset\_Turnover
- FR\_PBIDTM\_P
- FR\_ROE\_P
- FR\_ROA\_P
- FR\_EBIT\_Growth\_P

## Boruta Feature Selection Algorithm

### Model Output

```
[1] "finalDecision" "ImpHistory" "pValue" "maxRuns"
[2] "VR_EBITDA" "VR_Mcap_Sales" "VR_Adjusted_PE_"
[3] "VR_EBIT" "VR_P_CE" "VR_C_E"
[4] "VR_LowPE" "VR_HighPE" "FR_Price_CF_Ratio"
[5] "FR_Tax_Rate_P" "FR_PATM_P" "FR_Dividend_Pay_Out_Ratio_P"
[6] "FR_ROE_P" "FR_EBITM_P" "FR_Pre_Tax_Margin_P"
[7] "FR_PATM_Growth_P" "FR_EBITM_Growth_P"
[8] "VR_Adjusted_P_E" "VR_P_CE" "VR_HighPE"
[9] "VR_Mcap_Sales" "VR_EBITDA" "VR_LowPE"
[10] "VR_Adjusted_EPS_Rs" "FR_EBITM_P" "FR_PATM_P"
[11] "FR_EBITM_P" "FR_PATM_P" "FR_CPM_P"
[12] "FR_Sales_Working_Capital" "FR_EBIT_Growth_P"
[13] "NDP_P_E" "FR_ROE_P" "FR_PATM_P"
[14] "FR_ROA_P" "FR_EBITM_P" "FR_CPM_P"
[15] "FR_EBITM_P" "FR_PATM_P" "FR_Sales_Working_Capital"
[16] "VR_EBITDA" "VR_HighPE" "VR_LowPE"
[17] "VR_Mcap_Sales" "VR_EBITDA" "FR_Dividend_Pay_Out_Ratio_P"
[18] "VR_EBIT" "VR_C_E" "FR_PATM_P"
[19] "VR_LowPE" "VR_HighPE" "FR_EBITM_P"
[20] "FR_Dividend_Pay_Out_Ratio_P" "FR_PATM_P" "FR_CPM_P"
[21] "FR_PATM_P" "FR_EBITM_P" "FR_Sales_Working_Capital"
[22] "FR_ROE_P" "FR_EBITM_P" "FR_PATM_P"
[23] "FR_EBITM_P" "FR_PATM_P" "FR_CPM_P"
[24] "FR_Sales_Working_Capital" "FR_EBIT_Growth_P"
[25] "VR_EBITDA" "VR_HighPE" "VR_LowPE"
[26] "VR_Mcap_Sales" "VR_EBITDA" "FR_Dividend_Pay_Out_Ratio_P"
[27] "VR_EBIT" "VR_C_E" "FR_PATM_P"
[28] "VR_LowPE" "VR_HighPE" "FR_EBITM_P"
[29] "FR_Dividend_Pay_Out_Ratio_P" "FR_PATM_P" "FR_CPM_P"
[30] "FR_PATM_P" "FR_EBITM_P" "FR_Sales_Working_Capital"
[31] "FR_ROE_P" "FR_EBITM_P" "FR_PATM_P"
[32] "FR_EBITM_P" "FR_PATM_P" "FR_CPM_P"
[33] "FR_Sales_Working_Capital" "FR_EBIT_Growth_P"
[34] "VR_EBITDA" "VR_HighPE" "VR_LowPE"
[35] "VR_Mcap_Sales" "VR_EBITDA" "FR_Dividend_Pay_Out_Ratio_P"
[36] "VR_EBIT" "VR_C_E" "FR_PATM_P"
[37] "VR_LowPE" "VR_HighPE" "FR_EBITM_P"
[38] "FR_Dividend_Pay_Out_Ratio_P" "FR_PATM_P" "FR_CPM_P"
[39] "FR_PATM_P" "FR_EBITM_P" "FR_Sales_Working_Capital"
[40] "FR_ROE_P" "FR_EBITM_P" "FR_PATM_P"
[41] "FR_EBITM_P" "FR_PATM_P" "FR_CPM_P"
[42] "FR_Sales_Working_Capital" "FR_EBIT_Growth_P"
[43] "VR_EBITDA" "VR_HighPE" "VR_LowPE"
[44] "VR_Mcap_Sales" "VR_EBITDA" "FR_Dividend_Pay_Out_Ratio_P"
[45] "VR_EBIT" "VR_C_E" "FR_PATM_P"
[46] "VR_LowPE" "VR_HighPE" "FR_EBITM_P"
[47] "FR_Dividend_Pay_Out_Ratio_P" "FR_PATM_P" "FR_CPM_P"
[48] "FR_PATM_P" "FR_EBITM_P" "FR_Sales_Working_Capital"
[49] "FR_ROE_P" "FR_EBITM_P" "FR_PATM_P"
[50] "FR_EBITM_P" "FR_PATM_P" "FR_CPM_P"
[51] "FR_Sales_Working_Capital" "FR_EBIT_Growth_P"
[52] "VR_EBITDA" "VR_HighPE" "VR_LowPE"
[53] "VR_Mcap_Sales" "VR_EBITDA" "FR_Dividend_Pay_Out_Ratio_P"
[54] "VR_EBIT" "VR_C_E" "FR_PATM_P"
[55] "VR_LowPE" "VR_HighPE" "FR_EBITM_P"
[56] "FR_Dividend_Pay_Out_Ratio_P" "FR_PATM_P" "FR_CPM_P"
[57] "FR_PATM_P" "FR_EBITM_P" "FR_Sales_Working_Capital"
[58] "FR_ROE_P" "FR_EBITM_P" "FR_PATM_P"
[59] "FR_EBITM_P" "FR_PATM_P" "FR_CPM_P"
[60] "FR_Sales_Working_Capital" "FR_EBIT_Growth_P"]

```

### Selected variables and importance plot

| Variables           | meanImp     | decision  |
|---------------------|-------------|-----------|
| FR_Tax_Rate_P       | 11.00752567 | Confirmed |
| FR_PATM_P           | 9.794832428 | Confirmed |
| FR_EBITM_P          | 9.61914521  | Confirmed |
| FR_ROA_P            | 8.326324593 | Confirmed |
| FR_CPM_P            | 8.108092265 | Confirmed |
| FR_Pre_Tax_Margin_P | 7.100379799 | Confirmed |
| FR_Adjusted_EPS_Rs  | 6.955138615 | Confirmed |
| FR_PBIDTM_P         | 6.28577316  | Confirmed |
| VR_Adjusted_PE_     | 5.426861915 | Confirmed |

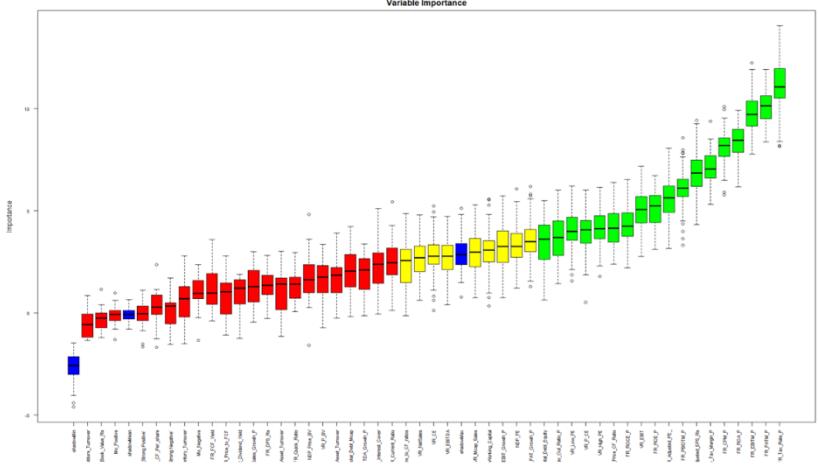


Fig: M12

Boruta is a feature selection algorithm based on a random forest classifier that uses randomized copies of the features and compares the best of the randomized copies to the real variable to arrive at a variable list determined by the mean decrease accuracy metric

A roughfix was performed on the tentative variables and finally confirmed features were extracted from the algorithm (Fig M12). Top 10 listed below.

- FR\_Tax\_Rate\_P
- FR\_PATM\_P
- FR\_EBITM\_P
- FR\_ROA\_P
- FR\_CPM\_P
- FR\_Pre\_Tax\_Margin\_P
- FR\_Adjusted\_EPS\_Rs
- FR\_PBIDTM\_P
- VR\_Adjusted\_PE\_
- VR\_EBIT

### Final Variable Selection process

Significant variables from the logit, classification tree, random forest and Boruta were analyzed and the top 10 variables that were selected in all models were finalized by a simple scoring mechanism.

#### Selected Variables

| Full_Vars           | Logit | Rpart | RF | Boruta | Score |
|---------------------|-------|-------|----|--------|-------|
| FR_EBITM_P          | 1     | 1     | 1  | 1      | 100%  |
| FR_ROA_P            | 1     | 1     | 1  | 1      | 100%  |
| FR_Pre_Tax_Margin_P | 0     | 1     | 1  | 1      | 75%   |
| FR_Tax_Rate_P       | 0     | 1     | 1  | 1      | 75%   |
| FR_PATM_P           | 0     | 0     | 1  | 1      | 50%   |
| FR_CPM_P            | 0     | 1     | 0  | 1      | 50%   |
| FR_Adjusted_EPS_Rs  | 0     | 0     | 1  | 1      | 50%   |
| FR_ROE_P            | 0     | 1     | 1  | 0      | 50%   |
| VR_Adjusted_PE_     | 0     | 1     | 0  | 1      | 50%   |
| FR_PBIDTM_P         | 0     | 0     | 1  | 1      | 50%   |

#### Not selected based on score

| Full_Vars                   | Logit | Rpart | RF | Boruta | Score |
|-----------------------------|-------|-------|----|--------|-------|
| VR_Mcap_Sales               | 1     | 0     | 0  | 0      | 25%   |
| FR_ROCE_P                   | 1     | 0     | 0  | 0      | 25%   |
| 'Mo_Strong Positive'        | 1     | 0     | 0  | 0      | 25%   |
| Mo_Positive                 | 1     | 0     | 0  | 0      | 25%   |
| FR_Asset_Turnover           | 1     | 0     | 0  | 0      | 25%   |
| 'Mo_Strong Negative'        | 1     | 0     | 0  | 0      | 25%   |
| Mo_Negative                 | 1     | 0     | 0  | 0      | 25%   |
| FR_Fixed_Asset_Turnover     | 0     | 0     | 1  | 0      | 25%   |
| FR_Price_CF_Ratio           | 1     | 0     | 0  | 0      | 25%   |
| VR_EBIT                     | 0     | 0     | 0  | 1      | 25%   |
| VR_P_CE                     | 0     | 1     | 0  | 0      | 25%   |
| FR_Dividend_Pay_Out_Ratio_P | 0     | 1     | 0  | 0      | 25%   |
| FR_Interest_Cover           | 0     | 1     | 0  | 0      | 25%   |
| FR_EBIT_Growth_P            | 0     | 0     | 1  | 0      | 25%   |

Fig: M13

The selected variables for phase 3 are:

- FR\_EBITM\_P : Earning Before Interests Taxes and Management (%)
- FR\_ROA\_P: Return on Assets (%)
- FR\_Pre\_Tax\_Margin\_P: Margin Pre-Tax (%)
- FR\_Tax\_Rate\_P: Tax Rate (%)
- FR\_PATM\_P: Profit After Taxes Margin (%)
- FR\_CPM\_P
- FR\_Adjusted\_EPS\_Rs: Adjusted Earning per Share (Rs)
- FR\_ROE\_P: Return on Equity (%)
- VR\_Adjusted\_PE: Price to Earnings ratio
- FR\_PBIDTM\_P: Profit before Interest Depreciation and Tax Margin

### Linear Regression

Linear Regression model applied on a lagged dataset having Adjusted EPS in Rs for the next year as the target variable has pointed out the following ratios as significant:

- FR\_ROA\_P\_18: Return on Assets for the year 2018
- FR\_Adjusted\_EPS\_Rs\_18 : Adjusted Earning per share for the year 2018
- FR\_ROE\_P\_Avg: : Return on Equity averaged over a period of 10 years.

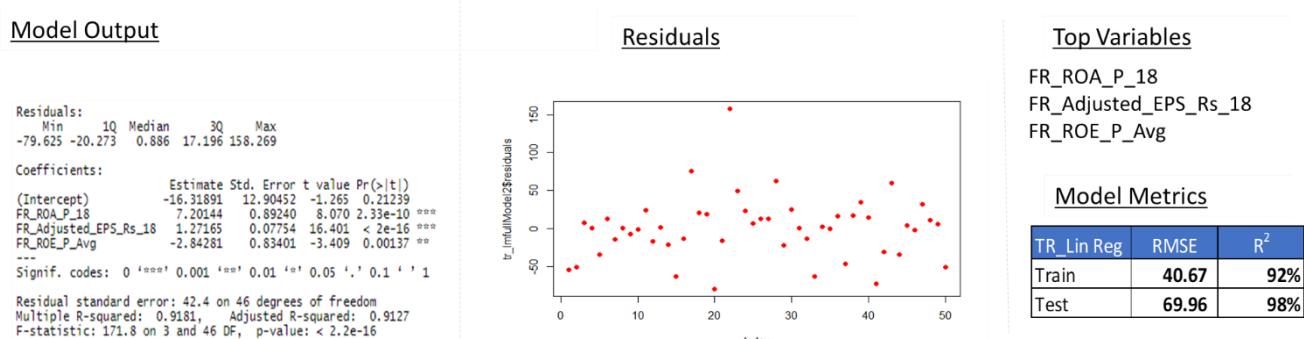


Fig: M14

The model's prediction power is good given the small dataset. The residuals are random so it implies the model is not overfitting. Model metrics (RMSE and R2) are consistent across train and test data.

Conclusion: The Return on assets and Equity for the year 2018 and averaged out Return on Equity explains the variation in Adjusted Earning per Share (Rs) in ~90% of the cases. This model can be used to predict the Adjusted Earning per Share (Rs) given previous years financial ratios.

**Stage 3:** Using the values of 2009 to 2018 of the selected variables from Stage 2, constructed additional features using slope, average and last previous value for a growth v/s non growth

classification. Using the 10 variables from stage 2, 30 features were created by company name for further analysis and modeling.

A sample of the variables are as shown in Fig M15

| Company Name              | FR_EBITM_P_Slope | FR_EBITM_P_Avg | FR_EBITM_P_18 | FR_ROA_P_Slope | FR_ROA_P_Avg | FR_ROA_P_18 |
|---------------------------|------------------|----------------|---------------|----------------|--------------|-------------|
| AIA Engineering Ltd.      | 1.65             | 24.58          | 28.52         | 0.78           | 16.737       | 15.55       |
| Amara Raja Batteries Ltd. | 0.08             | 12.57          | 11.54         | 0.16           | 15.919       | 12.16       |
| Apollo Tyres Ltd.         | 0.55             | 9.57           | 9.54          | 0.06           | 7.303        | 4.73        |
| Asahi India Glass Ltd.    | 1.21             | 8.28           | 15.38         | 0.91           | 0.895        | 6.22        |
| Astral Poly Technik Ltd.  | (0.09)           | 10.80          | 12.34         | (0.32)         | 9.652        | 8.66        |

Fig: M15

### Correlation Plot

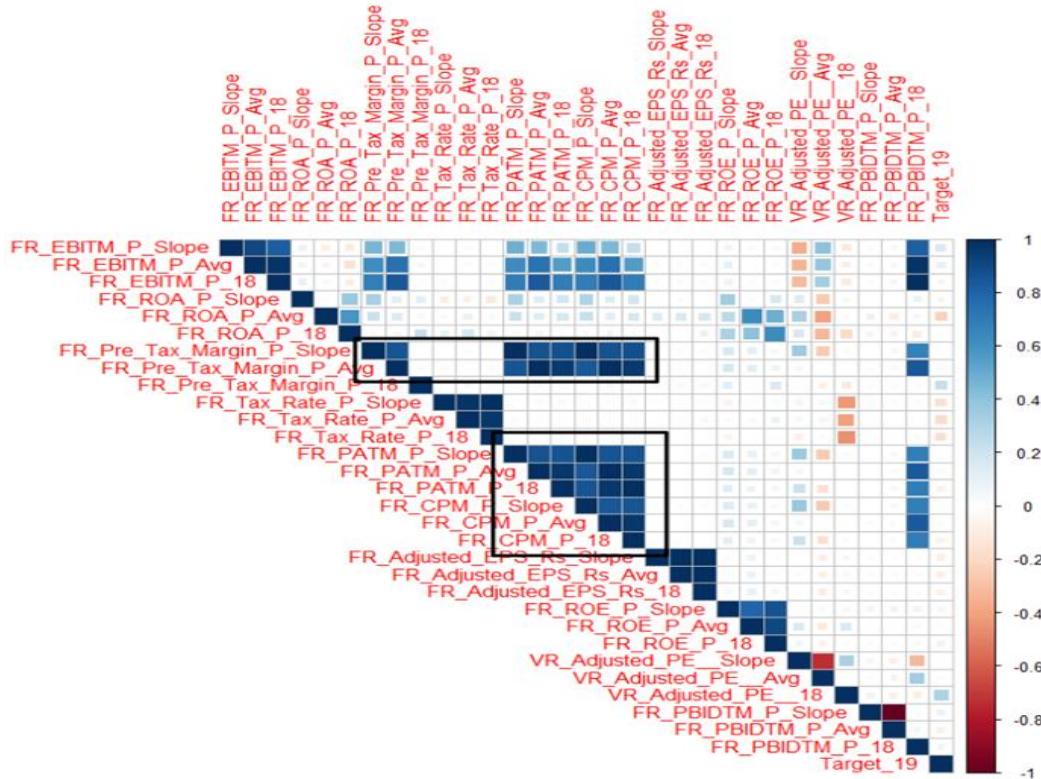


Fig: M16

A high level of collinearity is observed between multiple variables that indicates multicollinearity (Indicated by the box in Fig: M16). Removing multicollinear variables will be handled in the modeling exercise.

## Model Outputs

### Logistic regression on full data

#### Model Output

```

Deviance Residuals:
 Min Q1 Median Q3 Max
-8.49 0.00 0.00 0.00 8.49

Coefficients:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) 5.261e+14 4.359e-07 12070135 <2e-16 ***
FR_EBITM_P_Slope -2.040e-14 1.672e-07 -12200387 <2e-16 ***
FR_EBITM_P_Avg -4.146e-14 5.535e-06 -74914525 <2e-16 ***
FR_PATM_P_Slope 4.147e-14 1.672e-07 25202777 <2e-16 ***
FR_ROA_P_Slope -9.180e-14 3.329e-07 -27575338 <2e-16 ***
FR_ROA_P_Avg -6.895e-14 8.638e-06 -79860068 <2e-16 ***
FR_ROA_P_18 5.772e-14 8.616e-06 66987875 <2e-16 ***
FR_Pre_Tax_Margin_P_Slope 1.422e-14 4.281e-06 32380383 <2e-16 ***
FR_Pre_Tax_Margin_P_Avg 9.214e-12 1.617e-05 56988562 <2e-16 ***
FR_Tax_Rate_P_Slope -2.817e-14 4.795e-06 -58758534 <2e-16 ***
FR_Tax_Rate_P_Avg -2.140e-13 6.803e-05 -31455706 <2e-16 ***
FR_PATM_P_18 1.152e-13 3.238e-06 35789921 <2e-16 ***
FR_PATM_P_Slope 1.147e-15 1.298e-06 8840125 <2e-16 ***
FR_PATM_P_Avg 4.260e-14 2.240e-07 19012935 <2e-16 ***
FR_PATM_P_18 -1.039e-15 2.533e-07 -40996044 <2e-16 ***
FR_CPM_P_Slope 4.263e-14 9.398e-06 45269988 <2e-16 ***
FR_CPM_P_Avg 4.864e-14 1.120e-06 23864649 <2e-16 ***
FR_CPM_P_18 -5.186e-14 1.982e-07 -26166551 <2e-16 ***
FR_Adjusted_EPS_Rs_Slope 4.896e-13 2.384e-06 20540063 <2e-16 ***
FR_Adjusted_EPS_Rs_Avg -5.037e-13 9.411e-05 -53477671 <2e-16 ***
FR_Adjusted_EPS_Rs_18 3.147e-13 3.557e-06 35789921 <2e-16 ***
FR_ROE_P_Slope -8.054e-14 1.413e-07 -35981313 <2e-16 ***
FR_ROE_P_Avg -4.232e-13 3.557e-06 -11898837 <2e-16 ***
FR_ROE_P_18 3.576e-13 4.055e-06 882352 <2e-16 ***
VR_Adjusted_PE_Slope 4.060e-13 2.672e-06 16421909 <2e-16 ***
VR_Adjusted_PE_Avg 2.296e-13 3.112e-06 37322803 <2e-16 ***
VR_Adjusted_PE_18 2.396e-13 6.151e-05 37322803 <2e-16 ***
FR_PBIDTM_P_Slope -3.784e-14 1.170e-07 -32352624 <2e-16 ***
FR_PBIDTM_P_Avg -1.842e-14 5.699e-06 -32321096 <2e-16 ***
FR_PBIDTM_P_18 1.871e-14 1.233e-07 15174699 <2e-16 ***
--
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 87.148 on 69 degrees of freedom
Residual deviance: 720.873 on 39 degrees of freedom
AIC: 782.87
Number of Fisher Scoring iterations: 15

```

#### Multicollinearity - VIF

|                           |              |
|---------------------------|--------------|
| FR_EBITM_P_Slope          | 1.286767e+03 |
| FR_EBITM_P_Avg            | 3.709569e+03 |
| FR_EBITM_P_18             | 1.643129e+04 |
| FR_ROA_P_Slope            | 9.286659e+00 |
| FR_ROA_P_Avg              | 2.747268e+01 |
| FR_ROA_P_18               | 6.597560e+01 |
| FR_Pre_Tax_Margin_P_Slope | 1.819507e+03 |
| FR_Pre_Tax_Margin_P_Avg   | 3.548672e+03 |
| FR_Pre_Tax_Margin_P_18    | 2.244683e+00 |
| FR_Tax_Rate_P_Slope       | 6.299196e+02 |
| FR_Tax_Rate_P_Avg         | 4.746236e+01 |
| FR_Tax_Rate_P_18          | 4.771605e+02 |
| FR_PATM_P_Slope           | 9.168748e+03 |
| FR_PATM_P_Avg             | 2.625952e+04 |
| FR_PATM_P_18              | 3.331580e+04 |
| FR_CPM_P_Slope            | 4.491796e+03 |
| FR_CPM_P_Avg              | 2.525713e+04 |
| FR_CPM_P_18               | 2.068348e+04 |
| FR_Adjusted_EPS_Rs_Slope  | 1.788219e+02 |
| FR_Adjusted_EPS_Rs_Avg    | 9.801682e+02 |
| FR_Adjusted_EPS_Rs_18     | 6.047748e+02 |
| FR_ROE_P_Slope            | 4.112471e+01 |
| FR_ROE_P_Avg              | 8.857383e+01 |
| FR_ROE_P_18               | 1.689032e+02 |
| VR_Adjusted_PE_Slope      | 2.030497e+01 |
| VR_Adjusted_PE_Avg        | 6.368624e+00 |
| VR_Adjusted_PE_18         | 7.749218e+00 |
| FR_PBIDTM_P_Slope         | 1.206952e+06 |
| FR_PBIDTM_P_Avg           | 1.208243e+06 |
| FR_PBIDTM_P_18            | 1.291529e+04 |

Fig: M17

A high level of multicollinearity is indicated by high VIF values in Fig M17. Recursively remove the variables with the high VIF and re-run the model on train and test data

### Logistic Regression without multicollinear variables

#### Model Output

```

Deviance Residuals:
 Min Q1 Median Q3 Max
-1.9869 -0.6488 -0.2884 0.3559 2.1268

Coefficients:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.968227 1.759044 -1.119 0.2632
FR_ROA_P_Avg -0.271573 0.145644 -1.863 0.0622 .
VR_Adjusted_PE_18 0.069432 0.033811 2.054 0.0400 *
FR_ROE_P_18 0.147452 0.062135 2.373 0.0176 *
VR_Adjusted_PE_Avg -0.059100 0.037849 -1.361 0.1184
FR_ROA_P_Slope -1.408183 1.185190 -1.184 0.2986
FR_PATM_P_Slope 0.003487 0.008048 0.433 0.6468
FR_Pre_Tax_Margin_P_18 -0.003487 0.008048 0.433 0.6468
FR_Tax_Rate_P_18 -0.165491 0.114157 -1.450 0.1471
FR_PBIDTM_P_18 -0.008277 0.010902 -0.434 0.6646
VR_Adjusted_PE_Slope -0.060250 0.109815 -0.549 0.5832
FR_ROE_P_Slope -0.079130 0.602458 -0.131 0.8955
FR_Adjusted_EPS_Rs_Avg -0.002944 0.011785 -0.250 0.8027
FR_Pre_Tax_Margin_P_Slope 0.138368 0.217670 0.636 0.5250
--
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 62.687 on 49 degrees of freedom
Residual deviance: 39.630 on 36 degrees of freedom
AIC: 67.63
Number of Fisher Scoring iterations: 10

```

#### Top Variables

| Variable Name             | Overall  |
|---------------------------|----------|
| FR_ROE_P_18               | 2.373087 |
| VR_Adjusted_PE_18         | 2.053538 |
| FR_ROA_P_Avg              | 1.86464  |
| VR_Adjusted_PE_Avg        | 1.561476 |
| FR_Tax_Rate_P_Slope       | 1.449683 |
| FR_ROA_P_Slope            | 1.183563 |
| FR_Pre_Tax_Margin_P_Slope | 0.635677 |
| VR_Adjusted_PE_Slope      | 0.548654 |
| FR_PBIDTM_P_18            | 0.433542 |
| FR_Pre_Tax_Margin_P_18    | 0.433299 |
| FR_Adjusted_EPS_Rs_Avg    | 0.249797 |
| FR_ROE_P_Slope            | 0.131345 |
| FR_PBIDTM_P_Slope         | 0.074521 |

#### Confusion Matrix

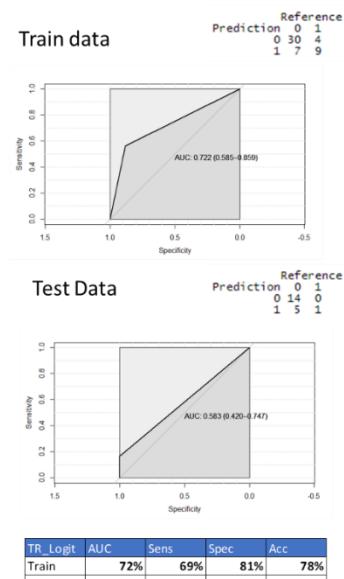


Fig: M18

After running the model without multicollinear variables the model performed well with an AUC score of 72% and 58% on train and test data respectively. Accuracy across train and test has been consistent as well. (Fig M18)

The significant variables with P value <.05 are as follows.

- VR\_Adjusted\_PE\_18
- FR\_ROE\_P\_18

### Classification tree

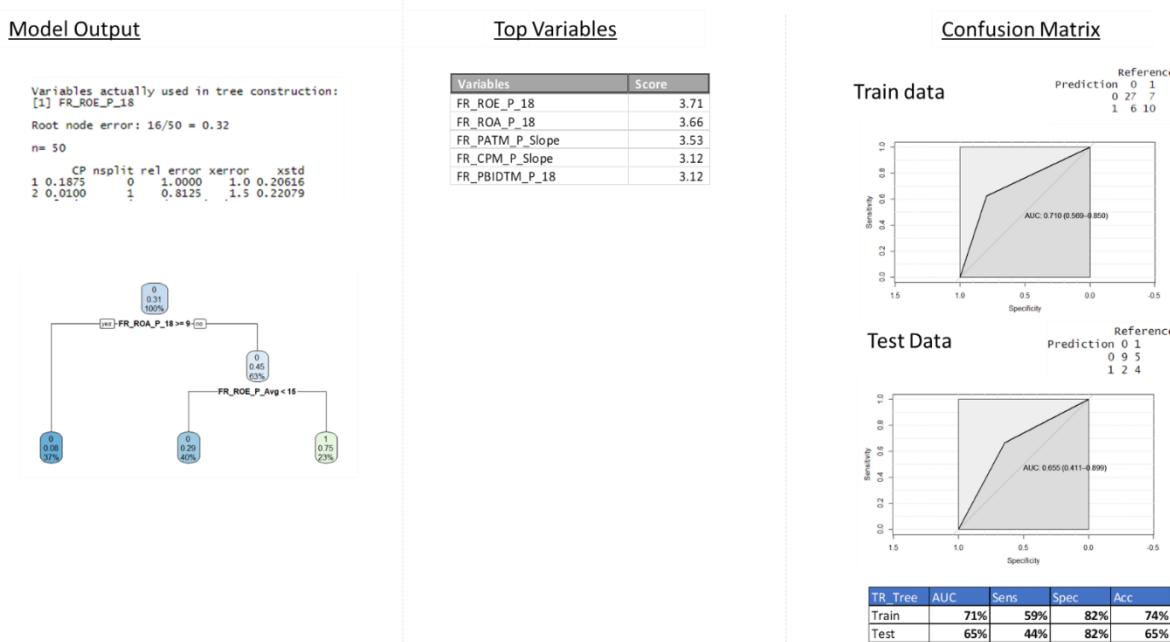


Fig: M19

The classification tree was run with minsplit =20 and minbucket =10. From the model outputs (Fig M19), the model performs well across train and test data shown as per the accuracy and AUC scores.

The significant variables from this model are:

- |                           |                |                     |
|---------------------------|----------------|---------------------|
| • FR_PATM_P_18            | • FR_ROA_P_18  | • FR_ROE_P_Avg      |
| • FR_PBIDTM_P_Slope       | • FR_ROA_P_Avg | • FR_Tax_Rate_P_18  |
| • FR_Pre_Tax_Margin_P_Avg | • FR_ROE_P_18  | • VR_Adjusted_PE_18 |

## Random Forest

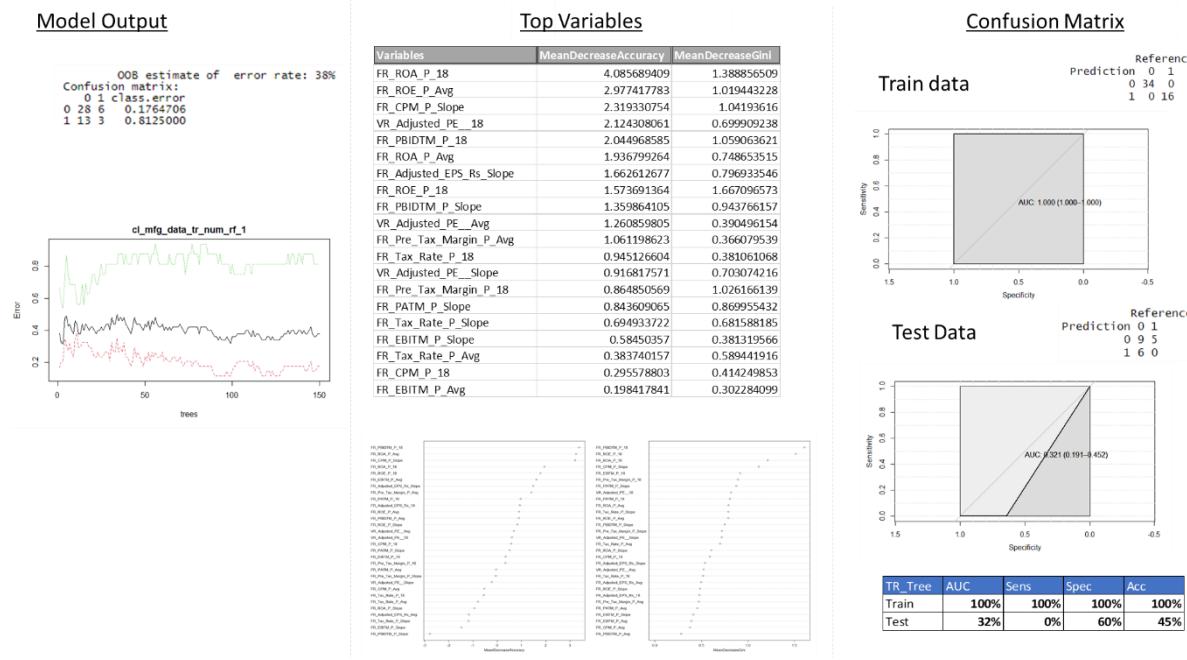


Fig: M20

The random forest model was run with ntree=150 and mtry=5. Other combinations of this were tried but it led to similar / worse results hence these parameters are chosen.

On the train data, the model is overfitting as the model metrics are 100%, and very low on the test data. Taking the model metrics into account, this model was discarded. (Fig M20)

Summary of models from this exercise

| TR_Logit | AUC        | Sens        | Spec       | Acc        |
|----------|------------|-------------|------------|------------|
| Train    | <b>72%</b> | <b>69%</b>  | <b>81%</b> | <b>78%</b> |
| Test     | <b>58%</b> | <b>100%</b> | <b>74%</b> | <b>75%</b> |

| TR_Tree | AUC        | Sens       | Spec       | Acc        |
|---------|------------|------------|------------|------------|
| Train   | <b>71%</b> | <b>59%</b> | <b>82%</b> | <b>74%</b> |
| Test    | <b>65%</b> | <b>44%</b> | <b>82%</b> | <b>65%</b> |

| TR_Tree | AUC         | Sens        | Spec        | Acc         |
|---------|-------------|-------------|-------------|-------------|
| Train   | <b>100%</b> | <b>100%</b> | <b>100%</b> | <b>100%</b> |
| Test    | <b>32%</b>  | <b>0%</b>   | <b>60%</b>  | <b>45%</b>  |

Fig M21

Due to the overall better metrics, the Logit model is chosen in this exercise.

## Findings Linked to Project Objectives

Based on the statistical analysis performed in the dataset related to the Manufacturing sector, following key financial ratios mapped to ratio categories: -

| Categories         | Ratios                               |
|--------------------|--------------------------------------|
| Valuation Ratios   | Adjusted Price to Earnings           |
| Performance Ratios | Return on Assets<br>Return on Equity |

Thus, Adjusted Price to Equity ratio will be key driver for the VALUE of a stock in Financial Services. Whereas Return on Assets and Return on Capital Employed, Debt to Market Capitalization and Sales to Working Capital Ratios will be the key driver for the QUALITY of a stock in Financial Services. MOMENTUM will be standard factor driving all investments and may not be specifically linked to a ratio. The average performance of certain ratios such as Return on Equity over the last 10 Years and the ratios in the most recent previous year could potentially be driving MOMENTUM of a stock.

## Recommendations & Conclusions

As an investor or a fund manager looking to make long-term investments in firms in the manufacturing sector, there are broadly two approaches one can adopt: -

1. Using the financial ratios of the current year, predict the Adjusted Earning per share for the next year then calculate the growth from this year and decide to buy, keep, or sell. The Return on Assets (%) and Adjusted EPS (Rs) from the current year along with Return on Equity averaged over past periods can be used to predict the Adjusted EPS for the next year. The fact that current year values impact the next year indicated Momentum based models to be effective, coupled with long term Return on Equity which means a company is making the most of the shareholders investments.
2. Using financial ratios classify companies as growth v/s non growth.

**Return on Assets** is a profitability ratio that provides how much profit a company is able to generate from its assets. For example, a study of the ROA values for Cummins India Ltd from 2016 to 2019 shows a slightly declining trend. The impact is a lower Earning per share in 2020.



Source: <https://www.moneycontrol.com/india/stockpricequote/engineers/cumminsindia/CI02>

- **Return on Equity:** Is the ratio of Net Income of the company to the Average shareholders equity. A strong ROE in the present year or over the years can indicate strong growth as it indicates the company is able to generate value for the shareholders.
- **Price to Earnings Ratio:** It is defined as the company's current price in relation to the earnings per share. This is related to the EPS and hence a strong P/E ratio over time or current year can be expected to translate directly into earning per share.

Apart from the above factors, Momentum of the stock (indicated by variables that are either the slope of the current value of the current year) over the last one year will be an important factor to assess. Positive or Negative Momentum stocks tend to continue with the same momentum the following year as well. Average of the ratios and the ratios in the current year have a higher significance than slope of ratios over the last 10 years. Thus, the horizon of analysis of ratios for long term investment should not go beyond the last 3-5 years of performance.

## Chapter 3: Final Recommendations and Conclusion

Following table summarizes the model output of Classification and Regression techniques applied on the dataset across macro-sectors: -

| Project Stage | Models                           | Macro-Sectors        | CLASSIFICATION MODELS |      |          |      |             |      |               |      |
|---------------|----------------------------------|----------------------|-----------------------|------|----------|------|-------------|------|---------------|------|
|               |                                  |                      | BFS                   |      | Consumer |      | Commodities |      | Manufacturing |      |
|               |                                  | Performance Measures | Train                 | Test | Train    | Test | Train       | Test | Train         | Test |
| Stage 2       | Logistic Regression (Best Model) | AUC                  | 82%                   | 64%  | 72%      | 65%  | 72%         | 61%  | 63%           | 68%  |
|               |                                  | Sensitivity          | 56%                   | 32%  | 43%      | 45%  | 73%         | 68%  | 66%           | 78%  |
|               |                                  | Specificity          | 84%                   | 84%  | 82%      | 84%  | 69%         | 55%  | 67%           | 71%  |
| Stage 2       | Random Forest                    | AUC                  | 100%                  | 74%  | 100%     | 66%  | 100%        | 66%  | 100%          | 72%  |
|               |                                  | Sensitivity          | 100%                  | 38%  | 100%     | 50%  | 100%        | 71%  | 100%          | 75%  |
|               |                                  | Specificity          | 100%                  | 84%  | 100%     | 81%  | 100%        | 62%  | 100%          | 74%  |
| Stage 2       | Decision Tree                    | AUC                  | 79%                   | 70%  | 74%      | 61%  | 69%         | 64%  | 77%           | 59%  |
|               |                                  | Sensitivity          | 66%                   | 41%  | 59%      | 46%  | 91%         | 71%  | 81%           | 54%  |
|               |                                  | Specificity          | 81%                   | 72%  | 88%      | 76%  | 46%         | 57%  | 79%           | 66%  |
| Stage 3       | Logistic Regression (Best Model) | AUC                  | 80%                   | 56%  | 93%      | 65%  | 78%         | 56%  | 72%           | 58%  |
|               |                                  | Sensitivity          | 83%                   | 50%  | 90%      | 50%  | 79%         | 50%  | 69%           | 100% |
|               |                                  | Specificity          | 73%                   | 75%  | 95%      | 80%  | 76%         | 38%  | 81%           | 74%  |

| Project Stages | Models  | REGRESSION MODELS |                      |       |          |       |             |       |               |       |
|----------------|---------|-------------------|----------------------|-------|----------|-------|-------------|-------|---------------|-------|
|                |         | Macro-Sectors     | BFS                  |       | Consumer |       | Commodities |       | Manufacturing |       |
|                |         |                   | Performance Measures | Train | Test     | Train | Test        | Train | Test          | Train |
| Stage 2        | Stage 2 | Linear Regression | RMSE                 | 22    | 28       | 22    | 27          | 29    | 35            | 27    |
|                |         |                   | Adjusted R2          | 13%   |          | 86%   |             | 89%   |               | 86%   |
| Stage 3        | Stage 3 | Linear Regression | RMSE                 | 17    | 84       | 13    | 22          | 39    | 37            | 40    |
|                |         |                   | Adjusted R2          | 56%   |          | 92%   |             | 34%   |               | 91%   |

Fig 2 – Model Summary

Sensitivity is a key model performance measure to consider as it is important for the model to predict the growth category of a firm (positive class -1) more accurately.

Random Forest overfits the Train Dataset across macro-sectors based on Sensitivity. Decision Tree techniques based on **RPART model** are giving out **better Sensitivities** for **BFS, Consumer and Commodities**, while **Logistic Regression** is giving out **better Sensitivities** across **Commodities and Manufacturing**.

Financial Ratios and its performance are an indicator of the health of any stock. Financial Ratios considered in the dataset belong to different categories: -

- Valuation Ratios
- Growth Ratios
- Liquidity Ratios
- Performance Ratios

Based on the detailed statistical analysis performed in the dataset related to every Macro-Sector, following represents the “Factor Map” of the Investment Strategy. This is arrived based on the statistically significant financial ratios for every Macro-Sector. The Factor Map will help investors and analysts in focusing on the appropriate firms belonging to every Macro-Sector.

## Application of the Factor Map

Traditionally in India the Mutual Fund managers are driven more by qualitative factors such as government policy changes, possible merger acquisition and recency metric. The factor map given below derived from the above models can be leveraged by Mutual Fund managers as a quantitative checklist that will help them build a high yielding portfolio of stocks.

| FINANCIAL RATIOS                                | BFS | CON | COMM | MFG |
|-------------------------------------------------|-----|-----|------|-----|
| <b>VALUATION RATIOS</b>                         |     |     |      |     |
| <i>Adjusted Price to Equity</i>                 | ✓   |     | ✓    | ✓   |
| <i>Price to Book Value</i>                      | ✓   | ✓   |      | ✓   |
| <i>Price to Earnings</i>                        |     |     |      | ✓   |
| <b>PERFORMANCE RATIOS</b>                       |     |     |      |     |
| <i>Return on Assets</i>                         | ✓   | ✓   |      | ✓   |
| <i>Return on Capital Employed</i>               | ✓   |     |      | ✓   |
| <i>Return on Equity</i>                         |     |     | ✓    | ✓   |
| <i>Enterprise Value by Capital Employed (%)</i> |     | ✓   |      |     |
| <i>Dividend Payout</i>                          |     |     | ✓    |     |
| <i>PBIT Margin</i>                              |     |     |      | ✓   |
| <b>LIQUIDITY RATIOS</b>                         |     |     |      |     |
| <i>Sales to Working Capital</i>                 | ✓   |     | ✓    |     |
| <i>Free Cash flow Yield</i>                     |     | ✓   |      |     |
| <i>Sales to Cashflow</i>                        |     |     | ✓    |     |
| <i>Interest Coverage</i>                        |     |     | ✓    |     |
| <i>Debt to Market Capitalization</i>            | ✓   |     | ✓    |     |
| <b>GROWTH RATIOS</b>                            |     |     |      |     |
| <i>EBIT Growth</i>                              |     | ✓   |      |     |
| <b>MOMENTUM</b>                                 |     |     |      |     |
| <i>Momentum</i>                                 | ✓   | ✓   | ✓    | ✓   |

Fig 3 – Factor Map

## Bibliography

- Fischer Black, Michael C. Jensen, Myron Scholes, The Capital Asset Pricing Model: Some Empirical Tests (1972)
- Stephen A. Ross, The arbitrage theory of capital asset pricing (1976)
- E. F. Fama, K. R. French, The Cross-Section of Expected Stock Returns (1992)
- Wen Long, Zhichen Lu , Lingxiao Cui, Deep learning-based feature engineering for stock price movement prediction (2018)
- Nikola Milosevic, Equity forecast: Predicting long term stock price movement using machine learning (2016)
- Yuxuan Huang, Machine Learning for Stock Prediction Based on Fundamental Analysis (2019)
- Fama French Three Factor Model  
<https://www.investopedia.com/terms/f/famaandfrenchthreefactormodel.asp>
- Size, Value and Momentum in Indian equities  
<https://journals.sagepub.com/doi/pdf/10.1177/0256090917733848>
- Corporate Finance & Accounting – terms  
<https://www.investopedia.com/terms/r/roce.asp>
- An Overview of Factor Investing - Fidelity Investments  
[https://www.fidelity.com/bin-public/060/www\\_fidelity\\_com/documents/brokerage/overview-factor-investing.pdf](https://www.fidelity.com/bin-public/060/www_fidelity_com/documents/brokerage/overview-factor-investing.pdf)
- Foundations of factor investing – MSCI  
<https://www.msci.com/www/research-paper/foundations-of-factor-investing/016381488>
- Correlation and Regression with R - Contributing Authors:Ching-Ti Liu, PhD, Associate Professor, Biostatistics ; Jacqueline Milton, PhD, Clinical Assistant Professor, Biostatistics ;Avery McIntosh, doctoral candidate
- Deep learning-based feature engineering for stock price movement prediction - Wen Long, Zhichen Lu \*; Lingxiao Cui

## Annexure

### 1. Data Dictionary –

| <b>Field Name used in Code</b> | <b>Original Field name</b>   |
|--------------------------------|------------------------------|
| Company_Name                   | Company_Name                 |
| Industry                       | Industry                     |
| MacroEconomicSector            | MacroEconomicSector          |
| FR_Year_End                    | FR_Year_End                  |
| FH_FCF                         | FH_Free Cash flow            |
| VR_Adjusted_PE_                | FR_Adjusted PE (x)           |
| VR_P_CE                        | FR_PCE(x)                    |
| VR_P_BV                        | FR_Price / Book Value(x)     |
| VR_Dividend_Yield              | FR_Dividend Yield(%)         |
| VR_NetSales                    | FR_EV/Net Sales(x)           |
| VR_EBITDA                      | FR_EV/EBITDA(x)              |
| VR_EBIT                        | FR_EV/EBIT(x)                |
| VR_CE                          | FR_EV/CE(x)                  |
| VR_Mcap_Sales                  | FR_M Cap / Sales             |
| VR_High_PE                     | FR_High PE                   |
| VR_Low_PE                      | FR_Low PE                    |
| PL_NetSales                    | PL_Net Sales                 |
| PL_PAT                         | PL_Profit After Tax          |
| FR_CF_Per_share                | FR_Cash Flow Per share       |
| FR_Price_CF_Ratio              | FR_Price to Cash Flow Ratio  |
| FR_FCF_Per_Share               | FR_Free Cash Flow per Share  |
| FR_Price_to_FCF                | FR_Price to Free Cash Flow   |
| FR_FCF_Yield                   | FR_Free Cash Flow Yield      |
| FR_Sales_to_CF_ratios          | FR_Sales to cash flow ratios |
| FR_EPS_Rs                      | FR_Earnings Per Share (Rs)   |
| FR_Adjusted_EPS_Rs             | FR_Adjusted EPS (Rs.)        |
| FR_CEPS_Rs                     | FR_CEPS(Rs)                  |
| FR_DPS_Rs                      | FR_DPS(Rs)                   |
| FR_Adj_DPS_Rs                  | FR_Adj DPS(Rs)               |
| FR_Book_Value_Rs               | FR_Book Value (Rs)           |
| FR_Adjusted_Book_Value_Rs      | FR_Adjusted Book Value (Rs)  |
| FR_Tax_Rate_P                  | FR_Tax Rate(%)               |
| FR_Dividend_Pay_Out_Ratio_P    | FR_Dividend Pay Out Ratio(%) |

|                          |                             |
|--------------------------|-----------------------------|
| FR_PBIDTM_P              | FR_PBIDTM (%)               |
| FR_EBITM_P               | FR_EBITM (%)                |
| FR_Pre_Tax_Margin_P      | FR_Pre Tax Margin(%)        |
| FR_PATM_P                | FR_PATM (%)                 |
| FR_CPM_P                 | FR_CPM(%)                   |
| FR_ROA_P                 | FR_ROA (%)                  |
| FR_ROE_P                 | FR_ROE (%)                  |
| FR_ROCE_P                | FR_ROCE (%)                 |
| FR_Asset_Turnover        | FR_Asset Turnover(x)        |
| FR_Inventory_Turnover    | FR_Inventory Turnover(x)    |
| FR_Debtors_Turnover      | FR_Debtors Turnover(x)      |
| FR_Fixed_Asset_Turnover  | FR_Fixed Asset Turnover (x) |
| FR_Sales_Working_Capital | FR_Sales(x)/Working Capital |
| FR_Fixed_Capital_Sales   | FR_Fixed Capital/Sales(x)   |
| FR_Receivable_days       | FR_Receivable days          |
| FR_Inventory_Days        | FR_Inventory Days           |
| FR_Payable_days          | FR_Payable days             |
| FR_Net_Sales_Growth_P    | FR_Net Sales Growth(%)      |
| FR_Core_EBITDA_Growth_P  | FR_Core EBITDA Growth(%)    |
| FR_EBIT_Growth_P         | FR_EBIT Growth(%)           |
| FR_PAT_Growth_P          | FR_PAT Growth(%)            |
| FR_Adj_EPS_Growth_P      | FR_Adj. EPS Growth(%)       |
| FR_Total_Debt_Equity     | FR_Total Debt/Equity(x)     |
| FR_Current_Ratio         | FR_Current Ratio(x)         |
| FR_Quick_Ratio           | FR_Quick Ratio(x)           |
| FR_Interest_Cover        | FR_Interest Cover(x)        |
| FR_Total_Debt_Mcap       | FR_Total Debt/Mcap(x)       |