Probability of NPA for NBFC – An Application of Random Forest Model

Vishweswarsastry V N
Assistant professor- selection grade
Department of Commerce
Manipal academy of higher
education Bengaluru Campus
Manipal academy of
higher education Manipal

Dr K.R.PundareekaVittala
Professor & Research Co-coordinator
Department of Commerce
Presidency College (autonomous)
Bengaluru

Abstract

Financial Analytics changed task based processing of Inspecting, transforming the business problem to business solution through analytical models using machine learning languages. Objective of the paper isto study the application of financial analytics on, thirdly Judge the model efficiency through confusion matrix and Sensitivity analysis and finally to obtain a random forest model for predicting the NPA of selected NBFCs. The analytical model applied is random forest. The outcome of the paper is the accuracy of the model at 97.53% in the prediction of NPA and proved that the defaults are more at every 2nd and 3rd companies.

Keywords: Financial Analytics, DuPont, Confusion matrix, Random Forest

Introduction:

The use of financial analytics is growing rapidly in the business environment. Financial analytics has fundamentally changed task based processes particularly those tasks that provide inference, prediction and assurance to decision makers. Financial analytics is the process of inspecting, cleaning, transforming, and modelling Big Data to discover and communicate useful information and patterns, suggest conclusions, and support decision making. Financial analytics now pervades every sector and function of the global economy.

Review of literature:

G. Sharma, (1992), stated that the Credit Defaults Predictors and Risk Modelling. The research study was aimed identify the most predictive variables (personal and loan specific) affecting credit worthiness of retail loan borrowers of NBFC, To develop consumer credit risk models to predict credit defaults of retail loan borrowers of an NBFC using two statistical techniques: Logistic Regression and Discriminant Analysis and to compare the models developed, in terms of identifying predictive indicators, degree of accurate classification of defaulters and significance of model. Research methodology is descriptive, predictive method and the tools used are logistic regression and discriminant analysis. The author observed that both the model has given same results and it showed that NBFCs should rely more upon alternate sources of loan rather than increasing internal sources, NBFC should focus their resources upon retaining existing borrowers rather than exploring new borrowers, increase the proportion of medium term loans in their portfolio.

Ssn Srinivas & Gumparthi Industry, (2010), the author analysed the Risk Assessment Model for Assessing NBFCs' Customers. The study aimed to design and develop risk assessment model based on market forces for assessing NBFC Customers. The author has also studied various risk assessment models availing for asset financing .To analyse the risk assessment the researcher followed descriptive and analytical analytics and the tools used are score model. The author could find out the defaulters with at least two dues against their names and the author also suggested the model to the NBFCs.

Kalra,Rosy ,(2016), the author studied Performance Analysis of Non-Banking Financial Institutions. The study was aimed analyze the performance of selected Non-Banking Financial Companies and to study the financial structure of NBFC in India and to analyze the profitability of selected NBFC. The methodology used is descriptive study of yearly data of 20 NBFC for the period of 2006-2015 and the tools used are graphical representations and ratios through Microsoft excel, Correlation and regression through SPSS. The author found out that the performance of NBFCs is improving and the results indicate that NBFIs are the dominant market player of the financial sector.

Lavanya, B Maheshwari,(2018), Credit Risk Management in NBFCs- Comparative Analysis. The objective of the study was credit risk management system in non-banking sector; examine the credit risk and probability of defaults; to determine the solvency credit worthiness of the selected companies and to compare the level of credit risk management practices of the companies under study. The research methodology is descriptive, analytical and the tools used are Merton model for credit risk and ratio analysis. The author stated that the credit risk management is essential for the Non banking financial companies

Tandon Deepak & Raghava Vandhana(2013) analysed the sectoral loan Defaults in SMES at NBFCs Using Altman Z score. The author examined the strength of select Indian NBFCs with help of Altman Z score tool and to identify the difference between the Altman z scores of Indian NBFCs serving Auto, Housing, Microfinance, Gold and Infrastructure sector spanning from 2012-2016.iThe methodology of the study employs the data for NBFCs for the period spanning 5 years starting from 2012-2016 by using Altman Z score model. The study shows resulted that the housing and infrastructure sectors are the worst performers in India.

Gumparthi, Manickavasagam, & Ramesh, (2010) Credit Scoring Model for Ancillary sector. The author aimed to provide a robust credit scoring model which stands as proof for decision making and the author attempt to design and develop a credit scoring model specific for Indian Auto Ancillary sector. The study conducted was descriptive and the tool used was discriminate analysis. The credit scoring model developed by the author should be considered as a tool for the credit analyst but this model is not intended to replace the existing method, it's an effective supporting tool in measuring the creditworthiness of the lease applicants.

Panwar & Aggarwal, (2018) Economic Development in India with Special Reference to Non-Banking Financial Companies, the author studied the role of NBFC towards the growth of Indian Economy to determine the present scenario & future prospects of NBFCs As the time passes the existence of NBFCs will show great results in the development and growth of the economy.

Madane Nikhil & Nand Siddharth (2019) here the author explains the loan prediction analysis using Decision tree model. The author reviewed the credit scoring of mortgage loans and the criteria that cause applicant to be rejected. The author also implemented a loan prediction system that helps companies make the right decision to approve or rejects customer's loan requests. The study draws the following insight about loan approval that is, Credit history applications that do not pass guidelines are mostly not approved, probably because they are more likely not to pay back and Most of the time, low-income applicants are more likely to received approval, which makes sense, and those applicants are more likely to repay their loans.

S. Panda & K. Rao (2019) here the author studies the Customer Acquisition and Retention in Non-Banking Finance Companies (NBFC), the main objective of the paper was to study how NBFCs can increase customer base and its retention. The author also developed the strategies for increasing and retain the customer base at MuthootFincrop Ltd. The author concludes that by proving benefits to the loyal customers and by having incentives and referral programs a NBFC can improve in retaining significant number of employees.

Shollapur M R (2010) Planning and Pricing of Financial Services, Author studied on Perceptions and Practices of Non-Banking Finance Companies. The author aimed to trace the unique features of financial products offered by NBFCs, to study the product differentiation techniques adopted, to examine the customer satisfying strategies and the feedback mechanism, to study the objectives of the pricing policy and to trace cost effectiveness of the financial services offered. The author studied 50 NBFCs operating in Karnataka. The author concludes that in infusing a strategic perspective in marketing financial product offers inputs to new thinking, new approaches and new skills for acquiring competitive ability that can boost up NBFCs financial performance in the competitive financial service industry.

D Chethan (2012) here the author studied on Credit Appraisal of Home Loans at Sundaram BNP Paribas Home Finance Ltd. The objective of the study was to ensure the balance between the risk and affordability of credit and to handle to effects of high NPAs. The research was descriptive and analytical and the tools used for the study purpose in Ratio Analysis. From this study it has been found that the company Norms and Criteria helps to verify the customer's credit worthiness, repayment of loans and good track record due to which the company is able to improve its standards and financial results.

Devi Uma (2014) the author conducted a study on the Investment Strategies of NBFCs. The author aimed to study the funding sources, investment strategies of NBFCs and investment philosophy rooted in NBFCs. The study was descriptive and investigative in nature and the data was analyzed theoretically. Through this study the author found that the NBFCs growth had been constrained due to lack of adequate capital. A number of NBFCs have been issuing non-convertible debentures (NCDs) in order to increase their balance sheet liquidity.

Khandani, Kim, & Lo, (2010) conducted a study on Consumer credit-risk models via machine-learning algorithms. The author aimed to apply machine-learning techniques to construct nonlinear nonparametric forecasting models of consumer credit risk.iThe study was descriptive, analytical and predictive analytics and the tool used was Classification and Regression trees (CART) to construct the forecast models. The author found that machine-learning forecasts are considerably more adaptive, and are able to pick up the dynamics of changing credit cycles as well as the absolute levels of defaults rates.

Krishnamurthy,(2012) the study was conducted on mathematical modelling of expected default frequency. The primary objective of the study was to design a mathematical model that captures the essential factors in evaluating the credit worthiness of a company. The study was descriptive and analytical in nature, KMV approach was employed for developing the mathematical model and regression analysis tool was used using STATA. The author concludes saying by employing appropriate model at the appropriate stage of the credit appraisal process the probability of defaults can be minimized to a very large extent and barring calamities, the company can be certain that the lease contract does not carry any risk of defaults.

Mall, Panigrahi, & Thomas, (2019) in this research paper the author focuses on identifying the tiggers of credit defaults and also focuses on checking and predicting the financial solvency of the borrowers of non-banking financial companies and assigning the credit worthiness to the companies. Altman's Z-score model is used to find credit worthiness and Dypont technique is used to find the main causes of financial distress. The results of this research highlights that the borrowing companies having a lower return on equity (ROE) are prone to be in distress zone. The author suggests the financial organizations to use Altman's Z-Score model and DuPont identity before granting loans to corporate.

Research Gap

From the extensive literature review it is evident that the researches concentrated on the performance and the associated risk with the NBFCs. This study aims to apply the financial analytical tools to assess the performance and frauds in IndostarCapital Finance, Bajaj Finance, L and T Finance, M & M finance, shriram city and Muthoot finance

Statement of Problem

NBFCs tend to grow at a faster pace but the problem of increased loan defaults rates and non- performing assets make it difficult on the other end. The problem of the study is to analyse the profitability of Indostar Capital Finance ltd using the various financial tool and predict on the NPAs.

Objectives of the Study

1. To study the application of financial analytics on NBFCs
2. To Analyse the accuracy of the model through generating Confusion matrix and Sensitivity Levels.
3. To judge the performance of Random forest model for predicting the Non Performing Assets of selected NBFCs.

Formulation of Hypothesis

a. Null Hypothesis (HO): There is no effect of Financial Analytics on Non performing assets of selected NBFCs.

b. Alternative Hypothesis (Ha): There is effect of Financial Analytics on Non performing assets of selected NBFCs.

Data Analysis and Interpretation

Table 1: showing DuPont Analysis for the Selected NBFCs

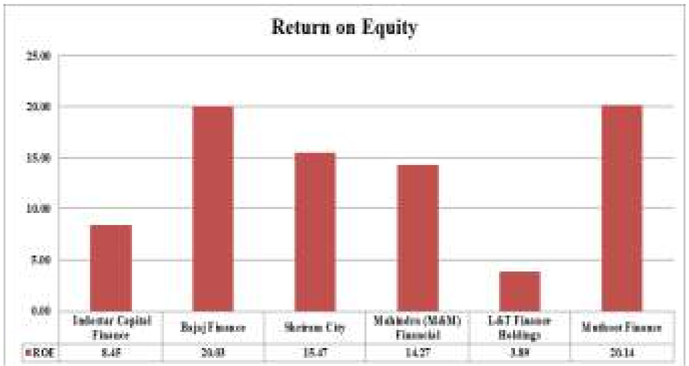
Performance	Indostar Capital Finance	Bajaj Finance	Shriram City	Mahindra (M&M) Financial	L&T Finance Holdings	Muthoot Finance
Net Income	255.13	3890.34	988.88	1557.06	339.81	1972.14
Sales	1177.17	17383.97	5778.78	8722.91	400.49	6878.21
Total Assets	12277.65	108499.87	29415.25	67077.99	10048.88	38068.7
Common Equity	3019.9	19425.78	6391.32	10908.03	8740.7	9792.72

Table 2: showing DuPont Analysis for the Selected NBFCs

DuPont Analysis	Indostar Capital Finance	Bajaj Finance	Shriram City	Mahindra (M&M) Financial	L&T Finance Holdings	Muthoot Finance
Profit Margin	0.22	0.22	0.17	0.18	0.85	0.29
Total Asset Turnover	0.10	0.16	0.20	0.13	0.04	0.18
Financial Leverage	4.07	5.59	4.60	6.15	1.15	3.89
Return on Equity	8.45	20.03	15.47	14.27	3.89	20.14

Graph 1. Representing the DuPont analysis of ROE of Selected NBFCs for the year 2019

Graph 2. Representing the Return on Equity of selected NBFCs for the year 2019



Interpretation: From the above table 1 it is evident that the net income and Total assets of Bajaj finance is highest compared to other above mentioned NBFCs and also the table explains the Profit margin of L and T highest compared to Indostar Capital Finance, Bajaj finance, Shriram Coty and Muthoot finance. The above Dupont analysis also proves that there is Financial leverage effect on its sources of capital proving its efficiency of repayment and it is noted that the financial Leverage is highest in case of Mahindra and Mahindra compared to other above mentioned NBFCs. As investor a metric of ROE is seen and Bajaj Finance is providing greater return on equity for its shareholders tp an extent of 20.03. the Graph 1 visualizes the profit margin, Total assets and Financial leverage whereas Graph 2 exclusively visualizes the effect of Dupont analysis on ROE for the investors.

Table 3: Showing the Confusion Matrix Of Training Data of Selected NBFCs

Prediction	Non Defaulters	Defaulters
Non Defaulters	27	0
Defaulters	2	52

Table 4: Showing the Accuracy matrix Of Training Data of Selected NBFCs

Accuracy	0.9753
95% CI	(0.9136, 0.997)
No Information Rate	0.6420
P-Value [Acc> NIR]	2.703E-13
Kappa	0.9455
Mcnemar's Test P-Value	0.4795
Sensitivity	0.9310
Specificity	1.0000
PosPred Value	1.0000
NegPred Value	0.9630
Prevalence	0.3580
Detection Rate	0.3333
Detection Prevalence	0.3333
Balanced Accuracy	0.9655
Positive' Class	Non Defaulters

Interpretation

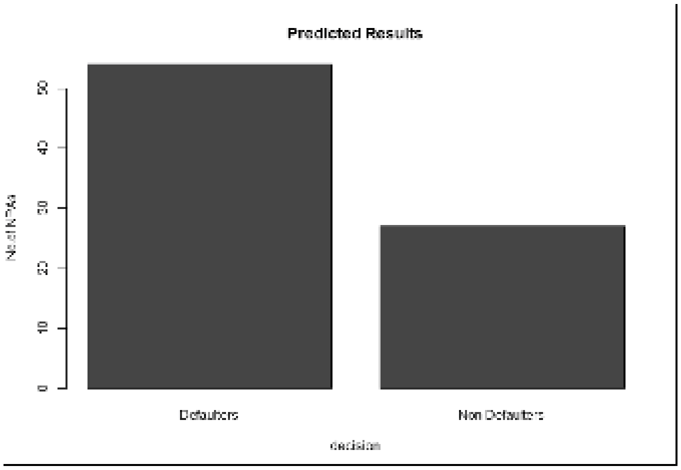
The above table 4 shows the confusion matrix with reference values and prediction values of Non Defaulters and Defaulters. It is seen that 27 numbers, the researcher is classified as Non Defaulters and the model has also classified as Non Defaulters so it is a correct classification. Similarly 52 numbers it is classified as Defaulters and there is only 2 misclassification between predicted non defaulters and actual defaulters. It is important to know the accuracy in the prediction and the model showed 97.53% accuracy for the data and also a sensitivity analysis which explains the model accuracy in the prediction of defaulters accurately is seen at 0.9310 0r 93.10% which is very attractive number when checked the accuracy of the variables.

Table 5: Showing the Predicted Training Summary 5 selected NBFCs Defaults

1	3	6	7
Non Defaulters	Non Defaulters	Non Defaulters	Non Defaulters
9	10	12	13
Non Defaulters	Defaulters	Defaulters	Defaulters
14	15	17	18
Defaulters	Defaulters	Defaulters	Non Defaulters
19	22	23	25
Defaulters	Defaulters	Defaulters	Defaulters
27	28	29	30
Non Defaulters	Non Defaulters	Non Defaulters	Non Defaulters
33	35	36	38
Defaulters	Defaulters	Defaulters	Defaulters
39	40	41	42
Defaulters	Defaulters	Defaulters	Non Defaulters
43	44	45	46
Defaulters	Defaulters	Defaulters	Defaulters
47	48	49	51
Defaulters	Defaulters	Non Defaulters	Defaulters
52	54	55	56
Defaulters	Defaulters	Defaulters	Non Defaulters
57	60	61	62
Defaulters	Defaulters	Defaulters	Non Defaulters
63	64	66	70
Defaulters	Defaulters	Non Defaulters	Defaulters
72	74	75	76
Non Defaulters	Non Defaulters	Defaulters	Non Defaulters
77	78	79	80
Defaulters	Non Defaulters	Defaulters	Defaulters
81	82	83	85
Defaulters	Defaulters	Defaulters	Defaulters
86	90	91	92
Non Defaulters	Defaulters	Defaulters	Defaulters
93	94	95	96
Defaulters	Non Defaulters	Defaulters	Non Defaulters
98	99	100	101
Non Defaulters	Defaulters	Defaulters	Defaulters
102	103	105	108
Defaulters	Defaulters	Defaulters	Non Defaulters
109	110	112	113
Defaulters	Defaulters	Defaulters	Defaulters
116			
Non Defaulters			

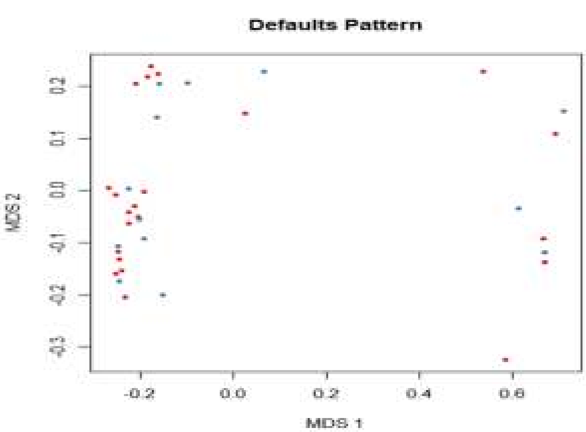
Interpretation: The above table 5 shows the training data of the NBFCs Predicted NPAs out of 116 Companies NPAs for 5 years is split in to 70% of training data, 30% of test the model. In the above table one can identify the prediction of defaulters and non defaulters from the training data which was split into 70% and clearly explains the visibility of default and non default which helps in the process of comparison with actuals which further visualises the accuracy of the model.

Graph 3 : Showing The Predicted Training Summary Of 116 NPAs of 5 selected NBFCs



Interpretation: The above graph 3 explains that the predicted values defaults of Indostar finance, L and T finance, Muthoot finance, M & M finance and Bajaj finance are higher for the selected 5 years than Non default rates. The Bar graph visualizes the same for the training data sets.

Graph 4: Representing the Multidimensional Scaling of Defaulters and Non Defaulters in Training Data set



Interpretation: Multidimensional scaling (MDS) is a multivariate data analysis approach that is used to visualize the similarity/dissimilarity between samples by plotting points in two dimensional plots. The iabove Graph 4 represents the Multidimensional Scaling for the training data set, it is predicted that the red colour dots represents the defaulters of the loans and blue colour dots represent the Non defaulters ,iit is been seen that the accuracy of the predicted data is high and the defaulters are more in the training data set.

Table 6 : Showing the Confusion Matrix Of Testing Data of Selected NBFCs

Prediction	Non Defaulters	Defaulters
Non Defaulters	8	0
Defaulters	0	27

Table 7: Showing the Accuracy Matrix of Testing Data

Accuracy	1.0000
95% CI	(0.9,1)
No Information Rate	0.7714
iP-Value [Acc> NIR]	0.0001136
Kappa	1.0000
Mcnemar's Test P -Value	NA
Sensitivity	1.0000
Specificity	1.0000
PosPred Value	1.0000
NegPred Value	1.0000
Prevalence	0.2286
Detection Rate	0.2286
Detection Prevalence	0.2286
Balanced Accuracy	1.0000
Positive' Class	Non Defaulters

Interpretation: The above table 6 shows the confusion matrix of test data which was given to check the accuracy of the model built with training data set and the model proved to very effective and also important to note the sensitivity levels which has also not shrinked proving the efficiency of the model. It is to be noted that the test data is a new data and a very minute data hence the sensitivity values are very high, but it is to be noted that the accuracy of the model is very efficient and any predicted values further will prove the same accuracy by reducing misclassifications.

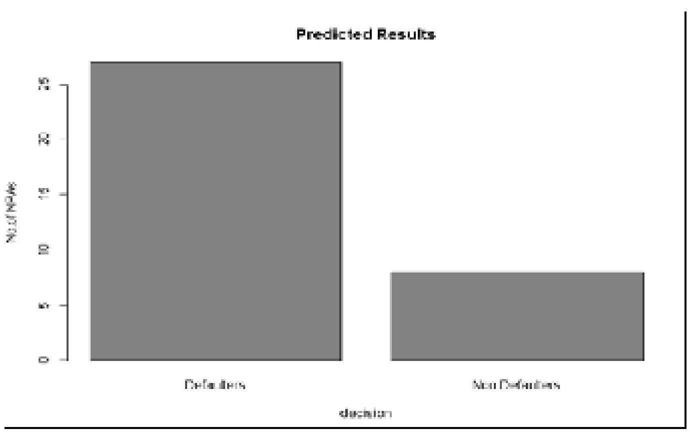
The above table 6 shows the confusion matrix of test data which was given to check the accuracy of the model built with training data set and the model proved to very effective and also important to note the sensitivity levels which has also not shrinked proving the efficiency of the model. It is to be noted that the test data is a new data and a very minute data hence the sensitivity values are very high, but it is to be noted that the accuracy of the model is very efficient and any predicted values further will prove the same accuracy by reducing misclassifications.

Table 8: Showing the testing Predicted Summary 5 selected NBFCs Defaults

2	4	5	8
Defaulters	Non Defaulters	Defaulters	Non Defaulters
11	16	20	21
Defaulters	Defaulters	Defaulters	Defaulters
24	26	31	32
Defaulters	Non Defaulters	Defaulters	Non Defaulters
34	37	50	53
Non Defaulters	Defaulters	Defaulters	Defaulters
58	59	65	67
Defaulters	Defaulters	Defaulters	Defaulters
68	69	71	73
Non Defaulters	Defaulters	Defaulters	Non Defaulters
84	87	88	89
Non Defaulters	Defaulters	Defaulters	Defaulters
97	114	106	107
Defaulters	Defaulters	Non Defaulters	
111	114	115	
Defaulters	Defaulters	Defaulters	

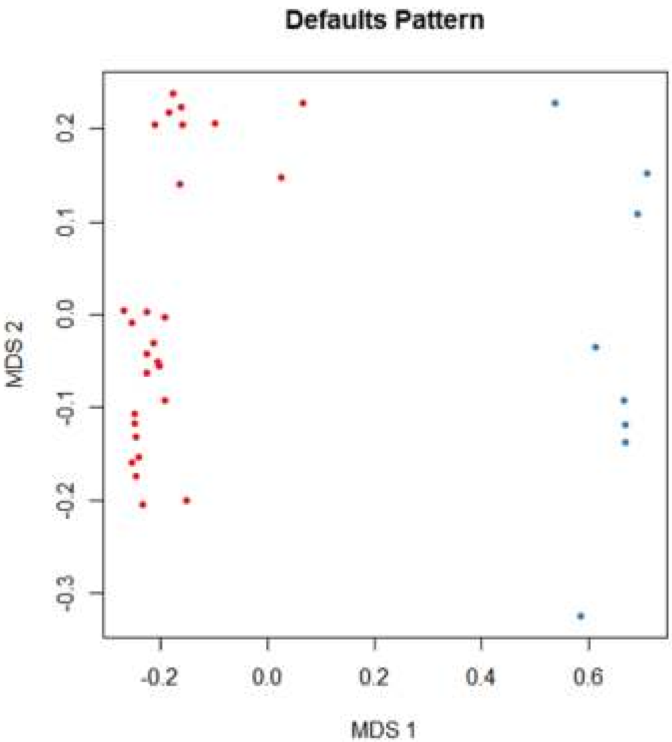
Interpretation- The above table 8 shows the testing data of the NBFCs NPAs for 5 years which was split in to 70% of training data, 30% for testing the model. In the above table one can identify the prediction of defaulters and non defaulters from the testing data as accurately as built through training the model which was split into 70% and clearly explains the visibility of default and non default which helps in the process of comparison with actuals which further visualises the accuracy of the model.

Graph 5.iShowing the Predicted Testing Summary Of NPAs of 5 selected NBFCs



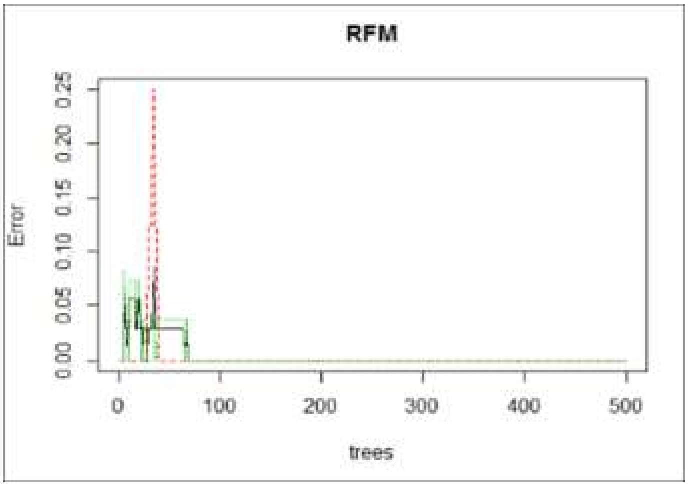
Interpretation: In the above graph 5 it shows the predicted testing summery NPAs of the Indostar Capital Finance, Bajaj Finance, Shriram City Union, M&M Financials, L&T Finance Holding and Muthoot Finance, this chart shows clearly shows Defaulters are more and Non Defaulters are less which was seen in the predicted data and also obtained the same with a new data of 30% out of 116 Companies NPA observation.

Graph 6: Representing the Multidimensional Scaling of Defaulters and Non Defaulters in Testing Data set



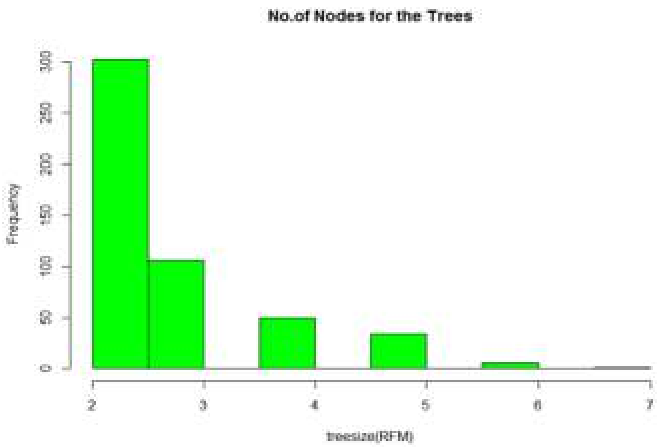
Interpretation-Multidimensional scaling (MDS) is a multivariate data analysis approach that is used to visualize the similarity/dissimilarity between samples by plotting points in two dimensional plots. Theiabove Graph 6 represents the Multidimensional Scaling for the testing data set, it is predicted that the red color dots represents the defaulters of the loans and blue color dots represent the Non defaultersiit is been seen that the accuracy of the predicted data is high and the defaulters are more in the testing data set.

Graph 7. Showing the Random Forest Plot of the testing and training data set.



Interpretation: From the above graph 7 it shows the plot of the random forest model where X axis contains trees and Y axis contains errors, it is seen that the out of bag error initially drops down and then becomes more or less constant, so it is not able to improve this error after about 100 trees.

Graph 8:.iShowing the Histogram of Random Forest model of the testing and training data set.



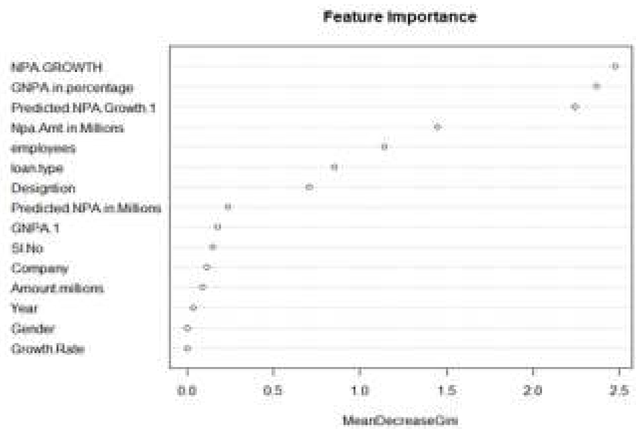
Interpretation: The above graph 8 shows the histogram of random forest model. The chart clearly shows more data scatter to the left side and one can understand 0 to 300. Accuracy of 2nd and 3rd Companies Default chances are more and in 4th Defaulters chances are little less when compared to 2 and 3 and 5th Companies accuracy of Defaulters chances is less when compared to 4th and 6th Defaulters accuracy is very very less. In this graph it shows that accuracy of more defaults in 2nd and 3rd Companies NPAs chances are more.

Table 9. Showing each feature Importance In Random Forest Model of the NPAs Prediction

MeanDecresseGini	
Sl.No	0.07444613
Year	0.02216133
Company	0.1306141
loan type	1.07706032
Employees	1.29860574
Gender	0.00000000
Designtion	0.87514325
Amount millions	0.10006219
NpaAmt in Millions	1.23854616
GNPA in percentage	2.19537457
GNPA 1	0.17647632
Growth Rate	0.00000000
NPA GROWTH	2.12374741
Predicted NPA Growth 1	2.35614695
Predicted NPA in Millions	0.15652981

Interpretation: From the above table 9, The MeanDecreaseGini measures the Gini importance = how important the features are over all splits done in the tree/forest - whereas for each individual split the Gini importance indicates how much the Gini criterion = "unequality/heterogeneity" was reduced using this split. So it makes sense to check how much features contributed to obtaining such homogeneous groups - which is the end is the MeanDecreaseGini. It is seen that Predicted NPA Growth, GNPA and Predicted NPA have played an important role and contributed most to obtaining such splits, so they are considered more important.

Graph 9.i. Showing The Each Variable Importance chart In Random Forest Model of the NPAs Prediction



Interpretation: The above Graph 9 showing the different variable importance for the variables used to predict the Less and Defaulters. MeanDecreaseGini, which is the average decrease in node impurity that results from splits over that variable. The Gini impurity index is only used for classification problem. In the regression the node impurity is measured by training set NPAs. These measures, calculated using the training set, are less reliable than a measure calculated on out-of-bag data.. The most important variable is Predicted NPAs Growth 1, GNPA in pergentage, NPA growth and less important variables are date of GNPA 1, Predicted NPA in Million, company, Loan Amounts, Sl.No, Year, gender, Growth rate 1.

Table 10: Showing the Single Tree of Randam Forest Model of Variable 45 of the Predicted NPAs of the selected NBFCs

Sl No.	Left Daughter	Right Daughter	Split Var	Split Point	Status	Prediction
1	2	3	Predicted.NPA.in.Mill	1178.685	1	<NA>
2	4	5	NPA -.Amt.in.Mill	319.500	1	<NA>
3	6	7	Sl.no	34.500	1	<NA>
4	0	0	<NA>	0.000	-1	Non Defaulters
5	0	0	<NA>	0.000	-1	Defaulters
6	8	9	Employees	67108831.000	1	<NA>
7	0	0	<NA>	0.000	-1	Defaulters
8	0	0	<NA>	0.000	-1	Defaulters
9	0	0	<NA>	0.000	-1	Non Defaulters

Interpretation: The above table 10 showing the single tree of random forest model variable of 45 it means the most important variable is Predicted NPAs in Millions, This has predicted the Non Defaulters and Defaulters, as per the prediction it can be said that all the selected NBFCs like Indostar Capital Finance, Bajaj Finance, Shriram City Union, M&M Financial, L&T holding and Muthoot Finance are all having Defaulters of loans when compared to the Non Defaulters.

Summary of Findings

1. It is been observed that DuPont analysis is a very powerful tool to analyse the quality of a business and its ability to generate capital. It analyses some of the most crucial factors such as what is causing the rise or decline in ROE. As per the Dupont analysis and reference to table 1 and graph 2 it is evident that the ROE of Muthoot Finance is highest having the value of 20.14 followed by 20.03, 15.47, 14.27,8.45 and 3,89 of Bajaj Finance, Shriram City, Mahindra and Mahindra Financials, Indostar finance and L and T finance holdings respectively.
2. The ROA of Muthoot Finance is high and it indicates that company is earning more money on less investment. The Debt/Equity Ratio of L&T Financials is much good then other company it can be said that the company is getting more of its financing by debts.
3. It's also observed that the exists positive relationship between the stock prices of Indostar, Shriram City Union, M&M Finance and L&T Finance which means that the risk is high and the diversification is not possible and it is advisable to trade the stocks Individually.
4. It is evident that in table 3 confusion matrix 27 companies, the researcher is classified as Non Defaulters and the model has also classified as Non Defaulters proving the accuracy of classification. Similarly 52 numbers is classified as Defaulters and there is only 2 misclassification between predicted non defaulters and actual defaulters. It is important to know the accuracy in the prediction and the model showed 97.53% accuracy for the data and also a sensitivity analysis which explains the model accuracy in the prediction of defaulters accurately is seen at 0.9310 0r 93.10% which is very attractive number when checked the accuracy of the variables.
5. Table 6 and 7 shows the confusion matrix of test data which was given to check the accuracy of the model built with training data set and the model proved to very effective and also important to note the sensitivity levels which has also not shrinked proving the efficiency of the model. It is to be noted that the test data is a new data and a very minute data hence the sensitivity values are very high, but it is to be noted that the accuracy of the model is very efficient and any predicted values further will prove the same accuracy by reducing misclassifications.
6. From the table 8 of the testing data of the NBFCs NPAs for 5 years which was split in to 70% of training data, 30% for testing the model one can identify the prediction of defaulters and non defaulters from the testing data as accurately as built through training the model which was split into 70% and clearly explains the visibility of default and non default which helps in the process of comparison with actuals which further visualises the accuracy of the model.
7. Graph 8 shows the plot of the random forest model where X axis contains trees and Y axis contains errors, it is seen that the out of bag error initially drops down and then becomes more or less constant, so it is not able to improve this error after about 100 trees.
8. Table 9 indicates The MeanDecreaseGini measures the Gini importance is equal to how important the features are over all splits done in the tree/forest - whereas for each individual split the Gini importance indicates how much the Gini criterion = "unequality/heterogeneity" was reduced using this split. So it makes sense to check how much features contributed to obtaining such homogeneous groups - which is the end is the MeanDecreaseGini. It is seen that Predicted NPA Growth, GNPA and Predicted NPA have played an important role and contributed most to obtaining such splits, so they are considered more important.

9. Graph 9 shows the histogram of random forest model. The chart clearly shows more data scatter to the left side and one can understand 0 to 300 accuracy data that of 2nd and 3rd Companies Default chances are more and in 4th Defaulters chances are little less when compared to 2 and 3 and 5th Companies accuracy of Defaulters chances is less when compared to 4th and 6th Defaulters accuracy is very very less. In this graph it shows that accuracy of more defaults in 2nd and 3rd Companies NPAs chances are more.
10. Graph 10 shows the different variable importance for the variables used to predict the Less Defaulters. MeanDecreaseGini, which is the average decrease in node impurity that results from splits over that variable. The Gini impurity index is only used for classification problem. In the regression the node impurity is measured by training set NPAs. These measures, calculated using the training set, are less reliable than a measure calculated on out-of-bag data. The most important variable is Predicted NPAs Growth 1, GNPA in percentage, NPA growth and less important variables are date of GNPA 1, Predicted NPA in Million, company, Loan Amounts, Sl.No, Year, gender, Growth rate 1.
11. The outcome from Table 10 shows the single tree of random forest model variable of 45. it means the most important variable is Predicted NPAs in Millions, This has predicted the Non Defaulters and Defaulters, as per the prediction it can be said that all the selected NBFCs like Indostar Capital Finance, Bajaj Finance, Shriram City Union, M&M Financial, L&T holding and Muthoot Finance are all having Defaulters of loans when compared to the Non Defaulters.

Conclusion: Random forest as a model predicts accurately which overtakes the statistical P value analysis by its accuracy by effective better feature engineering and feature selection by researcher which is shown in the paper and the model accuracy is not decrease with a new set of data proving the acceptability of the model.

References

- D Chethan. (2012). *Experience Certificate To Whomsoever It May Concern*. (1947),388120.
- Devi, V. U. (2014). Investment Strategies in NBFCs. *IOSR Journal of Business and Management*, *16*(10), 116–121.https://doi.org/10.9790/487x- 16103116121
- Gumparthi, S., Manickavasagam, V., & Ramesh, M. (2010). *Credit Scoring Model for Auto Ancillary Sector*. *1*(4),362–373.
- Joshi, P. V. (2011). Efficiency Evaluation of Banking Sector in India Basedon Data Envelopment Analysis. *Indian Journal of Commerce & Management Studies*, (March 2011),31–40.
- Kalra, R. (2016). *Performance Analysis of Non-BankingFinancial Institutions*. *6*(11),1–14.
- Khandani, A. E., Kim, A. J., &Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking andFinance*, *34*(11), 2767–2787.https://doi.org/10.1016/j.jbankfin.2010.06.001
- Krishnamurthy, P. (2012). Mathematical Modelling of ExpectedDefault Frequency. *SSRN Electronic Journal*, (October 2011). https://doi.org/10.2139/ssrn.193839
- Kumar, R. B., &Ayeswarya, R. B. (2019). *A Study on CurrentLiquidity crunch faced by NBFC ' s & HFC ' s in India Index in Cosmos* (Vol.9).
- Lavanya, B., Maheshwari, Y., &Prof, A. (2018). *CREDIT RISK MANAGEMENT IN NBFCS-A COMPARITIVE ANALYSIS*. *5*(6),469–474.
- Madane Nikhil, N. S. (2019). *LOAN PREDICTION ANALYSIS USING*. *21*(14),214–221.
- Mall, S., Panigrahi, T. R., & Thomas, S. (2019). Predicting FinancialSolvency of Commercial Borrowers: The Case of Non-Banking Financial Companies. *Accounting and Finance Research*, *8*(3), 61. https://doi.org/10.5430/afr.v8n3p61
- Panda, S., & Rao, K. S. N. (2019). *Customer Acquisition and Retentionin Non-Banking Finance Companies (NBFC)*. (5),601–613.
- Panwar, M. S., & Aggarwal, K. (2018). *Economic Development in India with Special Reference to Non-*

Banking Financial Companies- A Review.VII(Xiv), 72–75.

- Sakyi, P. A., Ofoeda, I., Kyereboah-coleman, A., & Abor, J. Y. (2014). *Risk and performance of non-bank financial institutions Risk and performanceof non-bank financial institutions Patience AsamoahSakyi Isaac Ofoeda * Anthony Kyereboah-Coleman and Joshua Yindenaba Abor*. (July). https://doi.org /10.1504/IJFSM.2014.062289
- Sam Raju, R., & Xavier, M. (2016). *Role of Nbfcs – Golden Principles on Lending for Gold*. *4*(4),1–7.
- Sharma, G. (1992). *MANAGEMENT Consumer Credit Default Predictorsand Risk Modeling : A Case Study*.93–97.
- Shollapur, M. R. (2010). Planning And Pricing Of Financial Services: AStudy On Perceptions And Practices Of Non-Banking Finance Companies. *International Business & Economics Research Journal (IBER)*, *9*(9),141–154. https://doi.org/ 10.19030/iber.v9i9.633
- Srivastava, U., &Gopalkrishnan, S. (2015). Impact of big data analytics on banking sector: Learning for Indian Banks. *Procedia Computer Science*, *50*, 643–652.https://doi.org/10.1016/j.procs.2015.04.098
- Tandon Deepak, R. V. (2013). Journal of management studies.*IEEE Transactions on Engineering Management*, *EM-12*(3), 57–63. https://doi.org/ 10.1109/tem.1986.6447671