DEVELOPMENT OF CREDIT RISK SCORECARD MODEL FOR NBFC'S IN INDIA

By:

Balam Sri Venkata Sravan

Roll.No:190A3010048

MBA (2019-2021)

Name of the Institute:

BML Munjal University

BML MUNJAL UNIVERSITY ™

FROM HERE TO THE WORLD

Faculty Guide: Dr. Jaskiran Arora

BML Munjal University

Asst. Dean (SOM)

MBA

Company Guide: Mr. Kingshuk Roy

ICRA Analytics Ltd

Senior Analyst

Noida

DEVELOPMENT OF CREDIT RISK SCORECARD MODEL FOR NBFC'S IN INDIA

By:

Balam Sri Venkata Sravan

Roll.No:190A3010048

MBA (2019-2021)

Name of the Organization:

ICRA ANALYTICS LTD.

Company Guide: Mr. Kingshuk Roy

ICRA Analytics Ltd

Senior Analyst

Noida

Faculty Guide: Dr. Jaskiran Arora

BML Munjal University

Asst. Dean (SOM)

MBA

ICRA ANALYTICS
A Group ICRA Company

ICRA ANALYTICS LIMITED
(Formerly known as ICRA Online Limited)

CERTIFICATE FROM ORGANIZATION

TO WHOM IT MAY CONCERN

This is to certify that Balam Sravan has undergone a training in our organization from July 13, 2020 to August 31, 2020, under the guidance of Kingshuk Roy.

During his tenure, he was engaged in developing a credit risk scorecard for NBFCs in India based on financial ratios Project in which he undertook the responsibility of:

- collection and preparation of data regarding NBFCs and develop a model based on machine learning algorithm.

His performance has been good, and we wish him all the best

in his future endeavors. ICRA Analytics Limited

Ishita Basu

Ishita Basu

Principal Associate - Human Resources

CERTIFICATE FROM FACULTY

This is to certify that the Project Work entitled “DEVELOPMENT OF CREDIT RISK SCORECARD MODEL FOR NBFC’s IN INDIA” in ICRA Analytics Ltd. Noida, New Delhi, is the work done by Balam Sravan under my guidance and supervision for the summer internship program offered by BML Munjal University.

To the best of my knowledge the work done in the project is by the candidate himself, completed the work in the due time given, fulfils all the rules and regulations relating to the summer internship of the institute. The content is up to the standard both in respect to content and language for being referred to examiner.

Date: 30-09-2020

Dr. Jaskiran Arora
Asst. Dean (SOM)

DECLARATION

I hereby declare that the project "DEVELOPMENT OF CREDIT RISK SCORECARD MODEL FOR NBFC's IN INDIA" at ICRA Analytics Ltd. is my own work. This project is an original work authored by me and contains no material that was published or written in the past researches by other authors. The material, data and analysis used in this project is prepared, collected and analyzed during the proceedings with the project. There are some references from past that have been used to thesis and due acknowledgement has been made for the same.

Balam Sri Venkata Sravan
MBA 2019-21
BML Munjal University

ACKNOWLEDGEMENT

I would like to extend our gratitude towards everyone who has played a crucial role in helping us to complete this internship. We would like to thank ICRA Analytics Ltd. for giving me this platform of experiential learning and providing the opportunity of interning with them. I would also like to thank our mentor Mr. Kingshuk Roy, Senior Analyst – ICRA Analytics LTD, who rendered continuous help, taught the needed concepts and backed with his moral support for completing this project. It is because of him and his team of Mr. Puneet Gupta, Ms. Goldy Bajaj and Mr. Sourav Das that the motivation and confidence in the approach of doing an effective project.

I would like to thank my faculty guide Dr. Jaskiran Arora who helped with the much-needed resources and support that were needed for the completion of this project.

It was a wonderful experience doing this project and its success would never have been possible without the support of all of them.

ABSTRACT

Credit risk scorecards are used to measure the credit worthiness of different companies, so that there can be an appropriate measure, based on numerous data, when these companies apply for a loan, to reduce their risk of default. In this project, which focuses on NBFCs in the Indian subcontinent, the authors have carried out the model building process for creating a Financial Risk Scorecard using financial ratios of the companies. After the collection and cleaning of data has been carried out, Logistic Regression is used to test and train the data according to their Weight of Evidence and Risk Ranking. After the process of model fitting the model validation has been carried out through Confusion Matrix and ROC curve to finalize the eligibility of the model used.

LIST OF TABLES & ILLUSTRATIONS

LIST OF FIGURES

LIST OF ABBREVIATIONS:

Abbreviation	**Full Form**
WOE	Weight of Evidence
NBFC	Non-Banking Financial Company
IV	Information Value
ROA	Return on Assets
ROC	Receiver Operating Characteristic Curve
AUC	Area Under Curve
EBITDA	Earnings before Interest, Tax, Depreciation and Amortization
EV	Enterprise Value
ICR	Interest Coverage Ratio
AUC	Area Under Curve
ROC - Curve	Receiver Operating Characteristic Curve

Table of Contents

1. INTRODUCTION

1.1 NBFCs

Non-Banking Financial companies are registered under the Companies Act 1956. According to the definition provided by the Companies Act the non-banking financial company are companies that provide financial services similar to what a bank provides but without holding any banking license. NBFCs primarily deal in the acquisition of stocks, debenture, and securities issued by the government or local authorities. They also deal in the business of advances and loans.

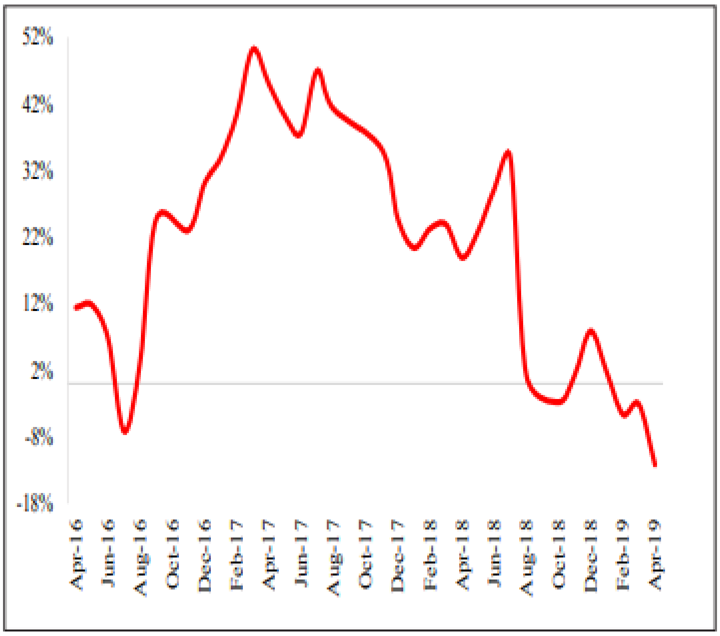
NBFCs perform lending functions to the public and cannot accept deposits from the public. NBFCs contribute largely to the economy by lending to infrastructure projects, small businesses and Housing projects and are monitored by RBI under RBI Act 1934. NBFCs borrow money from Banks or sell commercial papers to mutual funds to raise money. NBFC does not include financial institutions that engage in agricultural, industrial activity, purchase or sale of any good or providing any services and sale/purchase/construction of immovable property. Even though Banks are more regulated with clear sets of rules and have less rates of interest, borrowers prefer NBFCs as they disburse loan in no time. NBFCs are sometimes also referred to as Shadow Banks as they are a group of financial intermediaries which are not subject to regulatory oversight.

NBFCs are facing a liquidity crunch i.e. they have no money to lend or are facing difficulties in raising funds. As we know, the NBFCs typically borrow money from banks or sell commercial papers to mutual funds to raise money. They then lend this money to small and medium enterprises. When NBFCs do not have money to lend, the economic growth is hit – which reduces the credit flow to the economy. This in turn causes many borrowers to default on loans.

This leads us to the point that the NBFCs are fighting for survival as they were able to raise just 9% year-to-date in FY20 as compared to 30% in FY18. This has further been accentuated by high NPAs and stressed corporates. There were two main reasons that brought the NBFCs down to this situation –

a) The flawed business model of NBFCs which relied on raising short-term funds which were then lent out as long-term loans led to asset-liability mismatch.

b) Defaults by some firms of the IL&FS group put banks, mutual funds and the investors in a state of panic. This resulted in them refusing to give money to NBFCs.



Source: Estimated from SEBI data.

Figure 1: Liquidity crunch in NBFCs

The defaults by IF&LS, followed by DSP Mutual Fund dumping around ₹300 crore worth of commercial papers of Dewan Housing Finance Limited (DHFL) at a discounted rate made the investor doubt the condition of this sector and speculations of the downgrading of ratings of these companies started arising. This resulted in large NBFCs like Housing Development and Finance Corporation (HDFC) and Bajaj Finance's market capital eroded by around ₹18,600 crore and ₹ 13,800 crore respectively. These developments made investors nervous and their market capital of the NBFCs depleted. The year 2020 proved to increase the woes of NBFCs as their decline in non-bank credit growth, which started in second half of fiscal 2019, was amplified by the pandemic. The impact of economic slowdown, which was expected to be gradual, proved to be more immediate and debilitating. This provided us with the idea to check the credit worthiness of these NBFCs through by building a credit scorecard and gauging their credit worthiness.

1.2 Credit Scorecard

A Credit Scorecard is a formula that uses data elements, or variables to determine a threshold of risk tolerance. The score that credit lending institutions like banks obtain through this credit models assesses the credit history and payment performance, hence the credit worthiness of an organization. It tells us whether an organization is eligible for lending a certain amount as loan based on their financial strength, payment behavior, company age and size and a few other factors taken into consideration while perusing their application.

In this project we have tried to find out the credibility of the NBFC sector 2 years after the crisis in the sector started to emerge. To do that we have used a Financial Risk Scorecard which is used to gauge the credit worthiness of companies that approach for loans. For our project we are focusing on Non-Banking Financial Companies (NBFCs) and are building a scorecard model to find the eligibility of the companies through their Financial Ratios like their EPS, Book Value, Asset Turnover Ratio, EV/EBITDA etc.

1.3 Financial Ratios:

A financial ratio or accounting ratio is a relative magnitude of two selected numerical values taken from an enterprise's financial statement. Often used in accounting, there are many standard ratios used to try to evaluate the overall financial condition of a corporation/organization.

In order to find the financial strength of these NBFC companies we generally take help of Key Financial Ratios like their Asset Turnover Ratio, Interest Coverage Ratio, EV/EBITDA, EPS and some other key ratios. In the case of the NBFCs that have been taken for analysis these main financial ratios have been considered and keeping the Interest Coverage Ratio as the dependent variable we have developed a model which could predict the credit score of an NBFC and measure its credit worthiness.

2. REVIEW OF LITERATURE

Shadow banks are a group of financial intermediaries like activities performed by non-banking finance companies (NBFCs), which are not subject to regulatory oversight. NBFCs play a critical role in ensuring availability of loans and they disburse loans at a faster rate than banks making them a more effective option for easier and faster loans.
Srinivas Gumparthi SSn in his paper "Risk Assessment Model for Assessing NBFCs' (Asset Finance) Customers" said that NBFCs form an integral part of the Indian Financial system and the history of NBFC Industry in India is a story of under-regulation followed by over-regulation(Ssn & Industry, 2010). Srinivas says that policy makers have swung from one extreme position to another in their attempt to set controls and then restrain them so that they do not curb the growth of the industry. The key findings in his report, in which he identified 28 parameters to measure the risk associated with the customer, suggested that the qualitative measures (holistic view by a bank at its overall portfolio, deciding the lending limits to a sector, setting up the broad policies and procedures) interest the banks more than the quantitative measures (managing the credit risk by using quantitative tools and techniques such as ratio analysis to indicate the magnitude of risk and expected returns).

In his working paper "Financial Fragility in Retail – NBFCs" Prof. V. Ravi Anshuman examines the Financial Fragility of the Retail Non-Banking Financial Companies sector. He was successful in showing that the liquidity crunch in Retail NBFCs stemmed from their over-dependence on "short-term wholesale funding from Liquid Debt Mutual Funds (LDMFs) and the low level of high quality liquid investments in the LDMF sector". This kind of reliance worked well sometimes but during the times of stress it creates a significant amount of short-term debt rollover problems (Anshuman & Sharma, 2020). Prof. Anshuman pointed out that the transmission of risk from Retail-NBFCs to the LDMF sector was the key reason for the inability of Retail-NBFCs to roll over commercial papers. Due to the anticipated defaults by these Retail-NBFCs the mutual fund investors exited from the LDMF sector which led to the shortage of funding within the sector.

Mr. Vijaya Kittu Manda and Dr. P Sai Rani in their journal “Crisis in Indian Non-Banking Finance Company (NBFC) Sector” said that the problems in the NBFC sector, which seemed to be a liquidity crisis were turning into a solvency issue. It was noticed that major NBFCs which were backed by reputed promoters and investors were going out of business. One of the central problems that led to the liquidity issues in the industry is the asset-liability mismatch and the Housing Finance Companies (HFCs) as well as the Asset Management Industry were found to be more vulnerable and got highly hurt by the crisis (Manda & Rani, 2019). Mr. Vijaya and Dr. Sai Rani found out through their research that the regulator intervened timely during the time of the crisis to handle it but they gave much more importance to regulatory approach than creating a liquidity window for NBFCs. Mr. Vijaya and Dr. Sai Rani also put forward a point that NBFCs require tools to check the quality of their underlying assets and RBI or the Government should create a means to fund or something similar to bring in “short-term liquidity” at short notice.

Risk aversion in debt markets has heightened to an extent that the market has lost its ability to make a distinction across NBFCs, bracketing all of them in the same risk category, irrespective of the underlying nature of their assets and liabilities. The NBFC sector in reality is a very heterogeneous and constitutes different types of companies with different business models addressing very different underlying borrower segments.

NBFCs can be divided into three segments—asset financing, personal loans and business loans. One of the most prominent asset financing NBFCs are commercial vehicle financiers. The remaining NBFCs provide a range of personal and business loans with widely varying business models.

There are numerous drivers of the current risk aversion for NBFCs. The first one is related to short-term funding being used to finance long-term assets—an asset liability mismatch (ALM). NBFCs have been maintaining high credit growth at a time banks have gone slow on lending, owing to ongoing asset quality concerns and corrective actions. Some of these NBFCs have been running a large asset-liability mismatch. When liabilities far exceed assets in any period of time, it may lead to liquidity stress for the NBFC.

In the case of micro-finance, the average loan tenure ranges from eight to nine months. For commercial vehicle finance, it is 16 to 18 months and for small business finance, it is 12 to 16 months. Thus, the duration for these businesses is very short on the side of assets.

On the side of the liabilities, the duration is either similar to the asset side, or is longer, and generally ranges from one to two years. Hence, the small to mid-sized NBFCs run a positive ALM mismatch. A graphical representation of the same has been shown below.

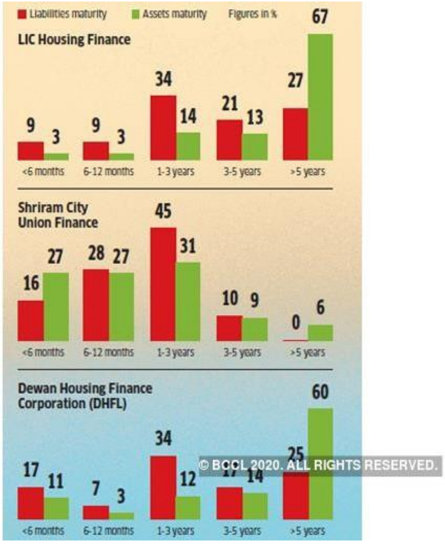


Figure 2: Asset Liability mismatch (Source - Emkay Research. Data as on March 2018)

Another cause of current risk aversion towards NBFCs has to do with refinancing of short-term capital market borrowings. This concern is linked to the ALM issue as smooth rollover of shorter duration liabilities when assets are of longer duration is key for business continuity. Asset Liability Mismatch can occur when the financials of a company's assets and liabilities do not match. NBFCs borrow short term but lend on a long term basis. This risky strategy works when the interest rates are under control but proves to be hurtful in case the interest rates rise.

Recently, the share of commercial paper funding to NBFCs subscribed by mutual funds has shown an increasing trend. Commercial paper funding (of up to 90 days) for NBFCs from mutual funds was increased from ₹ 50,000 crore in March 2016 to ₹ 1.2 trillion in September 2018. The increase in the market borrowings came in sharp focus recently as mutual funds have withdrawn from the market for NBFC paper. It was estimated that NBFC and HFC debt of about ₹ 2.5 trillion was due for roll-over by early 2019. This led to heightened refinancing risk for those who are dependent on such funding.

Apart from the ones mentioned above, another cause for concern has to do with asset quality. This pertains to NBFC exposure to the real estate sector—either as builder funding or loan against property (LAP). Builder funding does not exist in the loan book of NBFCs. In case of LAP portfolio of affordable housing financiers or small business loan financiers, the asset quality continues to be above par as the sourcing of these loans has been outside the ultra-competitive urban LAP market. More than that, the property underlying the LAP loans is typically self-occupied and not purchased for investment purposes.

Some of the effects of the risks associated with NBFCs after the IF&LS defaults have been shown below.

The credit ratings of NBFCs have downgraded after the default in IL&FS bonds sparked a liquidity squeeze that spread to other NBFCs. In the table below we can see the particular dates on which the credit ratings of NBFCs were revised and their revised ratings as well. It can be noticed that on the 8th of August in 2018 the credit rating of IL&FS fell from 'AAA' to 'AA+' which fell further down to 'D' on 17th September of the same year.

Company	Date	Earlier credit rating	Revised credit rating
IL&FS	08-Aug-18	AAA	AA+
IL&FS	17-Sep-18	AA+	D
Reliance Home Finance	27-Apr-19	BBB+	D
Reliance Commercial Finance	27-Apr-19	BBB+	D
DHFL	11-May-19	A3+	A4+
Reliance Capital	19-May-19	A	BBB
DHFL	05-Jun-19	A4+	D

Table 1: Companies and their Credit Ratings (Source – Rating agencies)

The asset quality of NBFCs has been constantly deteriorating in recent years. The portion of non-performing assets as a percentage of the loan book has steadily increased, while capital adequacy has declined. As can be seen from the chart below, the Non-performing assets have been on a constant rise year after year and have been from just 1.4% in FY14 to an all-time high of 4.4% in FY17.

Period	Gross NPA (%)	Net NPA (%)	Capital Adequacy Ratio (%)
FY14	2.6	1.4	27.5
FY15	4.1	2.5	26.2
FY16	4.5	2.5	23.6
FY17	6.1	4.4	22.1
FY18	5.8	3.8	22.8
Sep-18	6.5	NA	21

Table 2: Asset Quality of NBFCs (Source – SBI Research)

Several NBFC shares have taken a severe beating. Ten biggest losers have collectively seen market capital erosion of nearly Rs 1 lakh crore in the year 2018-2019. Given below is the data for some of these NBFCs.

Company	MCAP	PBV	CMP	1 Year price change (%)
Indiabulls Housing Finance	23,746	1.77	555	-53.5
Indiabulls Ventures	15,389	7.94	252	-47.7
Edelweiss Financial Services	15,895	2.38	170	-44.6
ICICI Securities	6,971	8.36	216	-39.8
Shriram City Union Finance	9,000	1.57	1,364	-38.3
Aditya Birla Capital	19,780	2.29	90	-36.0
JM Financial	6,887	1.56	82	-35.7
Housing & Urban Development Corp	7,727	0.79	39	-35.1

Table 3: Market Capital change in Companies (Source - Capitaline)

Credit Risk Rating of NBFCs is carried out through their Risk Scorecards. In these scorecards the quantitative factors such as financial ratios which include Asset quality, Capital Adequacy Ratio and Liquidity ratios are taken into account. Other than the above stated ratios the Earning Quality of the NBFCs are also taken into account to evaluate their credit risk. Qualitative measures such as Ownership, Risk management and Market presence have also been critical in the rating process.

Despite fundamental differences in the NBFCs, all NBFCs are today facing the brunt of a liquidity squeeze. A prolonged drying up of credit has started impacting them and their borrowers. Over the years, NBFCs have played an important role in providing growth capital to various sectors of the economy. What is required is a concerted effort across stakeholders to prevent a market contagion that can cut off the critical supply of capital for the development of the nation.

For the purpose of calculating the credit-worthiness of the NBFCs we have built a model which can be used by Credit Rating agencies to gauge the chance of default that different NBFCs pose when they apply for loans.

3. PURPOSE OF RESEARCH

The main purpose of the project is to create a model that determines the risk score for NBFCs in India based on financial ratios. We know that financial ratios for a company are the key information to determine the risk of the credit default our main agenda is to create a predictive model using these key financial ratios, and determine the variables which will be able to explain the default risk for NBFC. Through this model we validate whether the desired target variable (Interest Coverage Ratio) is predicted by the features. In order to build the model, the Financial statements of the NBFC companies are studied and the ratios which explain the credit default risk are identified. Used the model to find out the Independent variables that have the most impact on the dependent variable. This can be used to identify the key factors that have the most effect on an NBFCs Interest Coverage Ratio, which in turn will affect their credit worthiness.

To find out the most important metric to predict the target variable. As there are many statistical methods and algorithms, explored the best possible statistical method (or) algorithm to build a predictive model.

4. METHODOLOGY:

The different types of Scorecards:

1. *Application Scorecard* – It quantifies the risk associated with NBFCs applying for loans, by evaluating the social, demographic, financial and other data collected at the time of their application.
2. *Behavioral Scorecard* – This type of scorecard quantifies the NBFCs behavior by evaluating their previous credit performances and defaults that they have committed.
3. *Collection Scorecard* – Collection scorecard quantifies the probability of recovery of the outstanding balance for those accounts in collections. Collection scorecard statistically estimates the debtor's willingness to pay and helps to define what actions should be taken to increase collections.

Amongst these we used an application score card which is based on the interest coverage ratio of NBFC companies. Based on these data, the methodology we adopted was to develop a predictive model using Logistic regression Interest Coverage Ratio (ICR) being as dependent variable and remaining ratios as independent variables. In order to build a logistic regression model, we have gone through Data Exploration, Data Transformation and have split the data into Train set and Test set using a stratified random sampling technique.

4.1 Logistic Regression:

Logistic Regression is a statistical model which use the logistic function to model a binary dependent variable. Logistic regression should have its Dependent variable in a binary form (yes/no, Good/bad). Based on that we derived a logistic regression equation to fit the data into the model and train data to predict future values. To perform Logistic Regression with the data we transformed Interest Coverage Ratio into Binary variable. We had categorized the Interest Coverage Ratio (ICR) of companies greater than 1.5 as Good (1), and the ones that have an Interest Coverage Ratio (ICR) less than 1.5 as Bad (0). After categorizing the values Logistic Regression was carried out. After doing so we divided the data into training and testing data for evaluating the effectiveness of our model.

5. DATA COLLECTION & DATA ANALYSIS:

5.1 Data Collection:

To build a predictive model that predict credit risk of an NBFC company we collected financial statements for past 5 years of NBFC's from an open source data base (Money Control). We got a number of 152 companies Year-on-Year which were used for analysis. We used the pre-calculated key financial ratios for data analysis and model building.

5.2 Data Analysis:

The Data consisted of 760 rows and 38 columns as the data file loaded. In Data cleaning process filtered out the columns and rows which are having missing values more than 20%

So, after dropping the missing values of more than 20%, imputed the missing values by using mean of past years for the current year. After cleaning and treatment of missing values we had a data set of 355 rows and 30 columns.

As the data was in form of ratios and after the necessary Exploratory Data Analysis (EDA). For binning we calculated risk ranking bins using Weight of Evidence (WOE) and Information Value metrics (IV).

To build a predictive model we chosen to go with logistic regression based on the data.

So, with 26 independent variables and 1 dependent variable

$$\text{WOE} = \ln\left(\frac{\text{Distribution of Goods}}{\text{Distribution of Bads}}\right)$$

ln: natural log

Fig3: Weight of Evidence

$$IV = \sum (\text{\% of non-events} - \text{\% of events}) * WOE$$

Fig4: Information value

After the Binning we did a correlation check between the dependent and independent variables to find out for any multi-collinearity.

Binning is a way to group a number of more or less continuous values into a smaller number of "bins". For example, if you have data about a group of people, you might want to arrange their ages into a smaller number of age intervals.

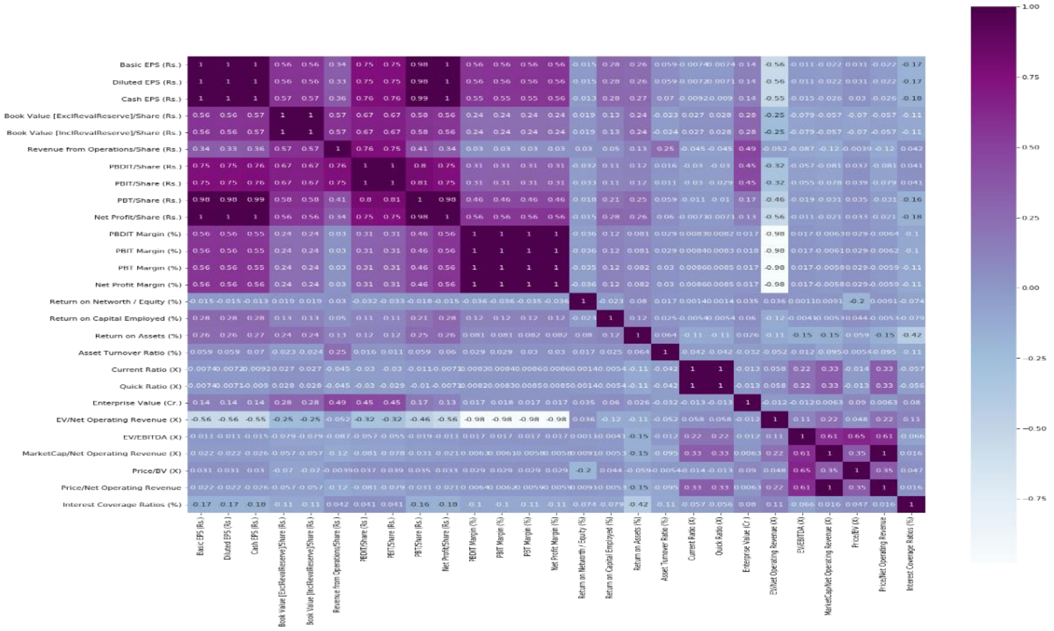


Fig5: Heat-map (Correlation)

From the above figure we can see the variables with multi-collinearity are shaded dark and one with lesser are bright. To treat this, we used Recursive step wise Logistic regression in order to eliminate multi-collinearity between variables and only significant variables will be included in model.

6. MODEL BUILDING AND MODEL VALIDATION:

6.1 Model Building

After data transformation, we opted to build model using Logistic regression, Interest Coverage Ratio (ICR) as dependent variable and the remaining 26 variables as independent variables.

To split the data into train and test set, we used a stratified random sampling to divide the data set into train set and test set of 80-20. That is 80% of data is divided as Train set and 20% is test set. After splitting the data into train and test sets the train set comprises of 284 rows and 26 columns, test data comprises of 71 rows and 26 columns. Each data set is divided as X and y in which X is the independent variables data frame and y is the dependent variable data frame.

```
X.shape, y.shape
Out[533]:
((355, 26), (355,))
```

Fig6: Dimensions of Train & Test data

As observed multi-collinearity between few variables we used a recursive (or) a step-wise logistic regression which helps us to eliminate variables with multi-collinearity and give the output with only significant variables, thereby increasing model validity.

Observing the p-statistic value we omit the variables whose significance levels are greater than alpha (0.05) and do an iteration of implementation of logistic regression and see until all the p-values are less than alpha (0.05). From the output we subset the variables, those are significant into a separate data frame and then fit it in the model.

```
Optimization terminated successfully.
         Current function value: 0.485234
         Iterations 6
                              Results: Logit
========================================================================
======
Model:              Logit                        Pseudo R-squared: 0.20
1
Dependent Variable: Interest Coverage Ratios (%) AIC:              350.
5161
Date:               2020-08-31 19:06             BIC:              362.
1325
No. Observations:   355                          Log-Likelihood:   -172
.26
Df Model:           2                            LL-Null:          -215
.57
Df Residuals:       352                          LLR p-value:      1.54
72e-19
Converged:          1.0000                       Scale:            1.00
00
No. Iterations:     6.0000
------------------------------------------------------------------------
------
                           Coef.   Std.Err.     z     P>|z|    [0.025
0.975]
------------------------------------------------------------------------
------
Return on Assets (%)      -0.6116    0.1020  -5.9980  0.0000  -0.8115
-0.4118
Enterprise Value (Cr.)    -0.3092    0.0873  -3.5417  0.0004  -0.4802
-0.1381
EV/EBITDA (X)              0.3665    0.1016   3.6069  0.0003   0.1673
0.5656
========================================================================
======
```

Fig7: Significant variables Logistic Regression

From the above, only three variables are significant in predicting the Credit risk score of NBFC and the model equation is:

ICR = -0.61 * ROA – 0.30 * EV + 0.36 * EV/EBITDA

ICR (Interest Coverage ratio) is the Predicted variable using ROA (Return on Assets), EV (Enterprise Value), EV/EBITDA. The model equation explains that for every unit increase in the dependent variable the ROA & EV are decreasing 0.61 and 0.30 times respectively and EV/EBITDA is increasing 0.36 times and vice-versa.

```
Logist
,
                    intercept
                    multi_class='auto',
                    random_state=None, solver='lbfg
,
                    warm_start=False)
```

Fig8: Model Fitting

6.2 Model validation:

Model validation is the process of confirming the outputs of a statistical model have enough relevance with what the actual output is. For this model we used the following Model validation techniques. The accuracy scores of the train and test data set are as follows.

```
accuracy_score on test dataset : 0.8732394366197183
```

Fig9: Accuracy Score - Train data

The accuracy score on the train data is 87% for the target variable and the predicted value for target variable.

```
accuracy_score on traindata
```

Fig10: Accuracy Score - Test data

The accuracy score on the test data is 87% for the target variable and the predicted value for target variable.

The accuracy score depicts the True positive and True negative rate in comparison to all the values in classification matrix. The higher the accuracy score the better the model is and for a classification model just accuracy cannot interpret the best fit of model, did the other validation metrics.

6.2.1 Confusion Matrix:

A confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known.

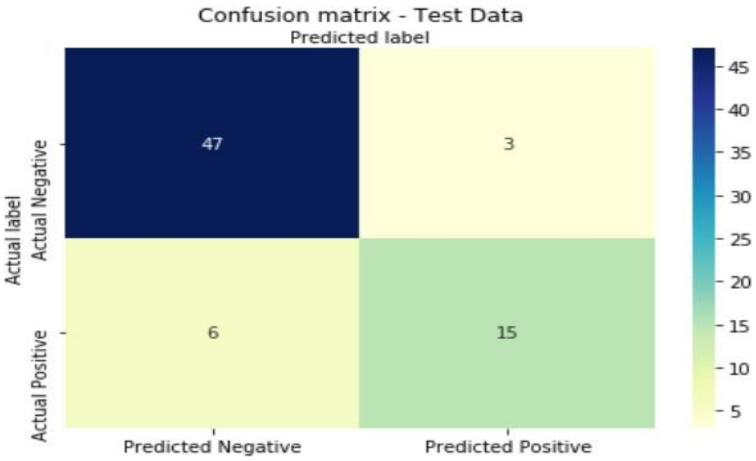


Fig11: Confusion Matrix- Test data

Above figure depicts the predicted test data with following values:

True positive: 15, True Negative: 47, Type 1 Error: 3, Type 2 Error: 6

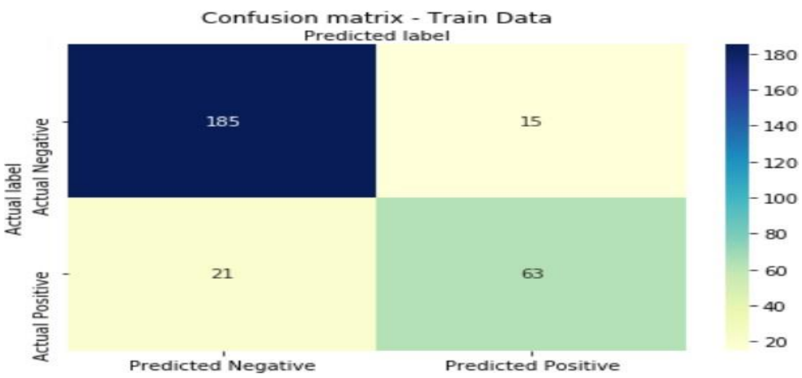


Fig12: Confusion Matrix- Train data

Above figure depicts the predicted train data with following values:

True positive: 63, True Negative :185, Type 1 Error: 21, Type 2 Error: 15

6.2.2. Classification Report:

Along with the confusion matrix we calculated the specificity, sensitivity, precision values of the test and train data set also.

```
Sensitivity: 0.7142857142857143
Specificity: 0.94
Type2: 0.2857142857142857
Type1: 0.06
```

Fig13: Valuation values- Test data

From the above table we can see that the specificity is 94%, for the test data which is a good sign about the model. And Type 1 error is 6% and Type 2 error is 28%, According to the data samples we have these values are okay with the model.

```
Sensitivity:  0.8076923076923077
Specificity:  0.8980582524271845
Type2:  0.19230769230769232
Type1:  0.10194174757281553
```

Fig14: Valuation values- Train data

From the above table we can see that the specificity is 90%, Mis-Classification is 13% for the train data which is a good sign about the model.

From above validation metric we can see both the sensitivity and specificity of both the train and test data are high, which means the values predicted are matching actual observed values, the model is valid.

6.2.3 ROC Curve:

A ROC curve plots the performance of a binary classifier under various threshold settings, this is measured by true positive rate and false positive rate. If your classifier predicts “true” more often, it will have more true positives (good) but also more false positives (bad). If your classifier is more conservative, predicting “true” less often, it will have fewer false positives but fewer true positives as well. The ROC curve is a graphical representation of this trade-off.

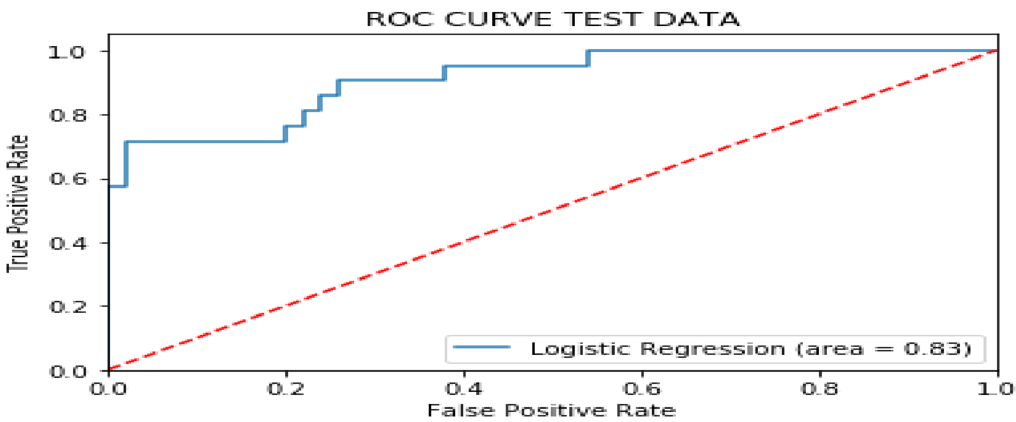


Fig15: ROC Curve- Test data

The Area under curve is above 0.83 that means the model is acceptable and distinguishes between positive and negative classes.

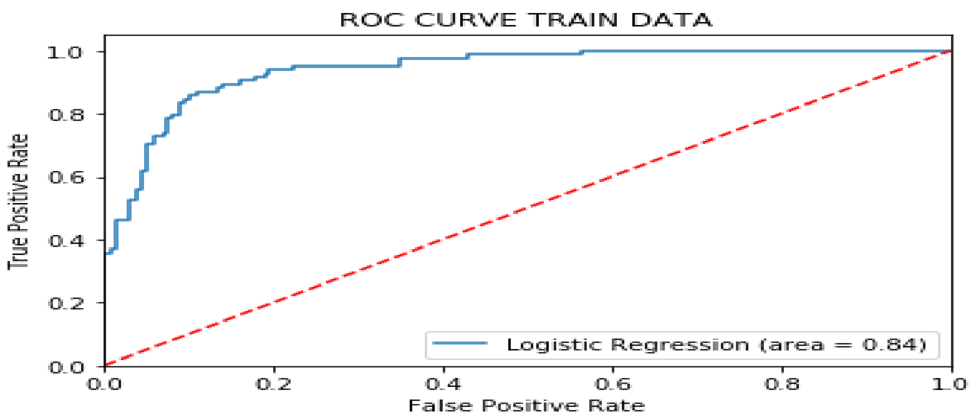


Fig16: ROC Curve - Test data

The Area under curve is above 0.84 that means the model is acceptable and distinguishes between positive and negative classes.

6.2.4 Gini Score:

The Gini index, or Gini coefficient, is a measure of the distribution of income across a population developed by the Italian statistician Corrado Gini in 1912. The coefficient ranges from 0 (or 0%) to 1 (or 100%), with 0 representing perfect equality and 1 representing perfect inequality.

The formula to compute Gini score is = (AUC*2) – 1

```
AUC Score: 0.8271428571428572
Gini score: 0.6542857142857144
```

Fig17: Gini Score – Test data

From the above observed Gini value 0.65 means that the model is good and also classify the country wealth index to good.

```
AUC Score: 0.8375
Gini score: 0.675
```

Fig18: Gini Score – Train data

From the Area under curve (AUC) of both train and test data the values are 82% and 83% respectively. For a classification model if an AUC value is close to 1 it is a good classifier and close to 0.5 is a bad classifier. So, from the above result the classification model used is good model. The ROC curve plotted between the True positive values and False positive values which are classified by the thresholds created i.e. greater than 1.5 good else bad.

7. CONCLUSIONS:

From the above validation metrics and values, we can conclude that our model is neither over-fitting nor under-fitting. Having a deep look into the Type 1 and Type 2 error values for both training and testing data set, we can see that the model is valid. After the model implementation we can observe that only "Return on Assets, Enterprise Value, EV/EBITDA "are significant variables and these are used for the model fitting. From the Area Under Curve (AUC) and Gini coefficients of both train and test data we can say that the model is good and can be used for prediction.

After the Data Analysis we found the insights for the research questions:

Since the target variable (Interest Coverage Ratio) is continuous we had classified it into categories of Good and Bad. Because for any NBFC company the ICR can explain the weight of borrowing and pay back. Since the features and target are in categories, used logistic regression to predict the target variable. It was evident that all the variables were not showing significant effect on the target variable except three variables (Return on Assets, Enterprise Value, EV/EBITDA). From this project we can say that ROA, EV, EV/EBITDA are the three main factors of a NBFC company to take into consideration while calculating the credit risk. From the model predicted equation we can say that The NBFC companies should keep an eye on the Return of Assets, Enterprise value as they go on decreasing the Credit Risk for the NBFC will also increase. And at same time the EV/EBITDA margins are to be maintained as they increase the Credit risk also increases.

From this project we can conclude that an NBFC company's Risk is mostly with the Return on Assets (ROA), Enterprise Value (EV) and EV/EBITDA ratios. A financial institute should have a strict check on these ratios of any company before they issue loan to prevent themselves from making a default payment.

8. LIMITATIONS OF THE PROJECT:

The main limitation for this project is:

- Data Sample: The sample data which we have collected is of dimensions of (760 * 38). But due to data consistency issues and missing values the unnecessary elements are omitted, which resulted in a very small sample of dimensions (355*30).
- Model Building: Due to the very small sample the model cannot predict accurately for huge amount of data.

9. REFERENCES:

1. https://www.datacamp.com/community/tutorials/feature-selection-python
2. https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python
3. https://medium.com/@prashant.nair2050/hands-on-outlier-detection-and-treatment-in-python-using-1-5-iqr-rule-f9ff1961a414
4. https://www.google.com/amp/s/www.geeksforgeeks.org/interquartile-range-and-quartile-deviation-using-numpy-and-scipy/amp/
5. https://www.analyticsvidhya.com/blog/2015/11/beginners-guide-on-logistic-regression-in-r/
6. https://docs.tibco.com/pub/sfire-dsc/6.5.0/doc/html/TIB_sfire-dsc_user-guide/GUID-07A78308-525A-406F-8221-9281F4E9D7CF.html#:~:text=IV%20%2D%20A%20numerical%20value%20that,the%20binary%20dependent%20variable%20y.&text=WOE%20%2D%20Closely%20related%20to%20the,value%20of%20the%20Dependent%20Variable.
7. https://multithreaded.stitchfix.com/blog/2015/08/13/weight-of-evidence/
8. https://www.datavedas.com/model-evaluation-in-python/
9. https://www.listendata.com/2019/07/KS-Statistics-Python.html
10. https://machinelearningmastery.com/repeated-k-fold-cross-validation-with-python/
11. https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/
12. https://stackoverflow.com/questions/31417487/sklearn-logisticregression-and-changing-the-default-threshold-for-classification
13. https://www.listendata.com/2019/07/KS-Statistics-Python.html#:~:text=By%20using%20scipy%20python%20library,take%20probability%20scores%20against%20events.
14. https://towardsdatascience.com/building-a-logistic-regression-in-python-step-by-step-becd4d56c9c8
15. https://stackoverflow.com/questions/43676905/how-to-calculate-cohens-kappa-coefficient-that-measures-inter-rater-agreement