# Finance and Risk Analytics PROJECT REPORT

(Milestone -I)

By:

SHARJIL SHAH

PGP – DSBA, Jan21’A

Great Learning

Table of Contents

[Finance and Risk Analytics PROJECT REPORT 1](#_Toc118657762)

[List of Tables 4](#_Toc118657763)

[Predicting Credit Risk 5](#_Toc118657764)

[Problem – I 5](#_Toc118657765)

[Introduction: 5](#_Toc118657766)

[Data Description: 6](#_Toc118657767)

[Sample of the dataset: 8](#_Toc118657768)

[Checking top 5 rows again after fixing messy column names for ease of use: 8](#_Toc118657769)

[Inference: 8](#_Toc118657770)

[1.1 Outlier Treatment 9](#_Toc118657771)

[1.2 Missing Value Treatment 12](#_Toc118657772)

[1.3 Transform Target variable into 0 and 1 14](#_Toc118657773)

[1.4 Univariate (4 marks) & Bivariate analysis with proper interpretation. (You may choose to include only those variables which were significant in the model building) 15](#_Toc118657774)

[Univariate Analysis: 15](#_Toc118657775)

[Inference: 19](#_Toc118657776)

[Bivariate Analysis: 19](#_Toc118657777)

[1.5 Train Test Split 22](#_Toc118657778)

[1.6 Build Logistic Regression Model (using statsmodel library) on most important variables on Train Dataset and choose the optimum cutoff. Also showcase your model building approach 24](#_Toc118657779)

[1.7 Validate the Model on Test Dataset and state the performance matrices. Also state Interpretation from the model 25](#_Toc118657780)

[Interpretation: 30](#_Toc118657781)

## Table of Figures

[Figure 1 - Dataset Sample 9](#_Toc118657786)

[Figure 2 - Fixing Messy Column - Dataset Sample 9](#_Toc118657787)

[Figure 3 - Outliers in data set 11](#_Toc118657788)

[Figure 4 - After Treating Outliers in data set 12](#_Toc118657789)

[Figure 5 - Missing values 13](#_Toc118657790)

[Figure 6 - Visual of missing values – After dropping 14](#_Toc118657791)

[Figure 7 - Target Variable - First and last 5 Rows 15](#_Toc118657792)

[Figure 8 - Descriptive Statistics Summary 16](#_Toc118657793)

[Figure 9 - Count plot of target variable Default 17](#_Toc118657794)

[Figure 10 - Important Feature variables Box Plot 17](#_Toc118657795)

[Figure 11 - Distribution of column with Displot & Box plot 18](#_Toc118657796)

[Figure 12 - Distribution of column with Displot & Box plot 19](#_Toc118657797)

[Figure 13 - Distribution of column with Displot & Box plot 19](#_Toc118657798)

[Figure 14 - Gross Sales Vs Net Sales 20](file:///E:\Great%20Learning\Capstone%20FRA\Sharjil_%20Shah%20Finance%20and%20Risk%20Analytics%20PROJECT%20REPORT.docx#_Toc118657799)

[Figure 15 - Net worth Vs Capital Employment 21](file:///E:\Great%20Learning\Capstone%20FRA\Sharjil_%20Shah%20Finance%20and%20Risk%20Analytics%20PROJECT%20REPORT.docx#_Toc118657800)

[Figure 16 - Net worth Vs Cost of Production 21](file:///E:\Great%20Learning\Capstone%20FRA\Sharjil_%20Shah%20Finance%20and%20Risk%20Analytics%20PROJECT%20REPORT.docx#_Toc118657801)

[Figure 17 - Correlation heat-map - Top significant Variables 22](#_Toc118657802)

[Figure 18 - Scaled Train dataset First 5 rows 24](#_Toc118657803)

[Figure 19 - Scaled Test dataset First 5 rows 24](#_Toc118657804)

[Figure 20 - Classification Report for the training set 27](#_Toc118657805)

[Figure 21 - Testing Set - Confusion matrix 28](#_Toc118657806)

[Figure 22 - SMOTE Testing Set - Confusion matrix 30](#_Toc118657807)

# List of Tables

[Table 1 - Data description of all variables of raw data 6](#_Toc118657722)

[Table 2 - Data description of all variables of raw data 7](#_Toc118657723)

[Table 3 - Confusion matrix for training set 25](#_Toc118657724)

[Table 4 - Confusion matrix for test set 26](#_Toc118657725)

[Table 5 - SMOTE Training Set - Classification Report 28](#_Toc118657726)

[Table 6 - SMOTE Training Set - Confusion matrix 28](#_Toc118657727)

[Table 7 - SMOTE Testing Set - Classification Report 29](#_Toc118657728)

# Predicting Credit Risk

## Problem – I

Businesses or companies can fall prey to default if they are not able to keep up their debt obligations. Defaults will lead to a lower credit rating for the company which in turn reduces its chances of getting credit in the future and may have to pay higher interests on existing debts as well as any new obligations. From an investor's point of view, he would want to invest in a company if it is capable of handling its financial obligations, can grow quickly, and is able to manage the growth scale.

A balance sheet is a financial statement of a company that provides a snapshot of what a company owns, owes, and the amount invested by the shareholders. Thus, it is an important tool that helps evaluate the performance of a business.

Data that is available includes information from the financial statement of the companies for the previous year (2015). Also, information about the Networth of the company in the following year (2016) is provided which can be used to drive the labeled field.

Explanation of data fields available in Data Dictionary, 'Credit Default Data Dictionary.xlsx'

### Introduction:

This assignment helps us to perform Outlier Treatment, Missing Value Treatment, Transform Target variable into 0 and 1, Univariate and Bivariate Analysis, Split data into Train & Test, Model Building Logistic Regression on most important variables on Train Dataset and choose the optimum cutoff, Model Validation is to be done on Test Dataset.

We have 3586 entries and 67 columns. The outcome of this assignment will suggesting investors good credit rating companies to invest their money.

### Data Description:



Table 1 - Data description of all variables of raw data



Table 2 - Data description of all variables of raw data

### Sample of the dataset:

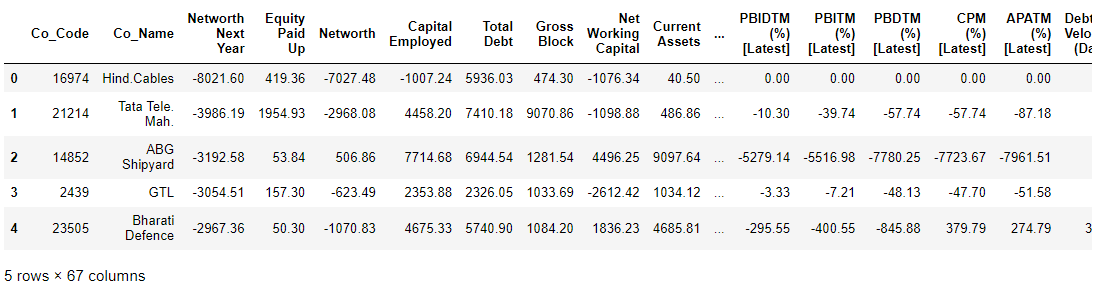


Figure 1 - Dataset Sample

### Checking top 5 rows again after fixing messy column names for ease of use:

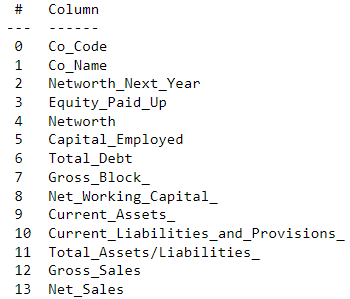
****

Figure 2 - Fixing Messy Column - Dataset Sample

### Inference:

The number of rows (observations) is = 3586 & The number of columns (variables) is = 67

• There is NO duplicate entry in the dataset.

• All the variables are numeric type except one variable (Co\_Name) which is object type.

• For our analysis, we can drop Co\_Code and Co\_Name variables.

• There are missing values in 13 of the variables. Missing values will be treated with either mean or median values of corresponding variables.

• There are outliers in the dataset. It will be treated for our analysis.

• The problem statement requires to predict “default” status of the company where the “Networth Next Year” of the company is used to drive the “default” field. The “default” is 1 when “Networth Next Year” is negative and it is 0 when “Networth Next Year” is positive.

• The “Default” field is created and added to the dataset based on the condition mentioned above. Subsequently “Networth Next Year” is not considered further as it became redundant.

## Outlier Treatment

Outliers are present in all of the independent variables. Outlier treatment is necessary for any regression model, whereas in Regression, outliers pull the regression line towards itself thereby affecting its slope. This distorts the reality and leads to faulty predictions.

1. Treating outlier by using Inter Quartile range for each of numerical column.
2. Values greater than Upper quartile range would be capped with 75% of quartile value
3. Values lesser than Lower quartile range would be capped with 25% of quartile value

* Q1 = 25th percentile, Q3 = 75th percentile - IQR = Q3 - Q1
* Outlier = any value which lies beyond 1.5 times of IQR from Q1 and Q3 on either side

• The outliers would be replaced with Upper Quartile values or lower. And post outlier treatment the numerical variables in boxplot:

#### Before Treating Outliers:

#### 

Figure 3 - Outliers in data set

#### After Treating Outliers:

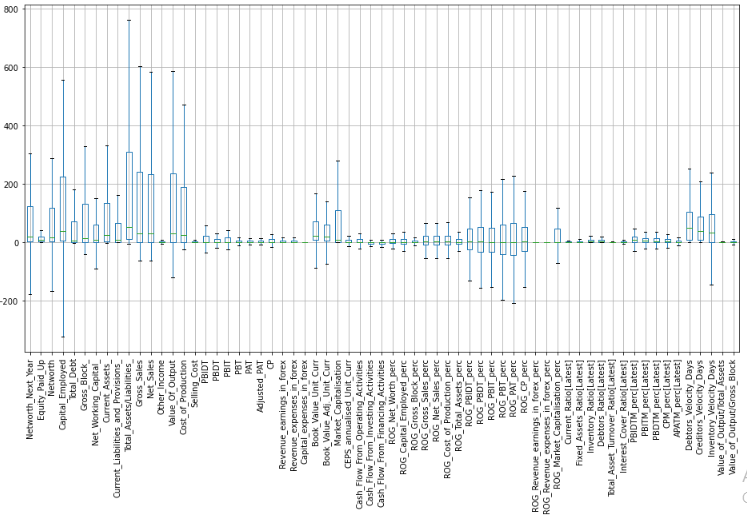


Figure 4 - After Treating Outliers in data set

#### Inference:

Given the fact that this is a financial data and the outliers might very well reflect the information which is genuine in nature. Since there is data captured for small, medium as well as large companies.

## Missing Value Treatment

There are some missing values in the dataset which is to be treated in the further steps. Given the size of the data set i.e. 3586 rows, there were not many missing values to start with.

#### Before imputing NULL Values :

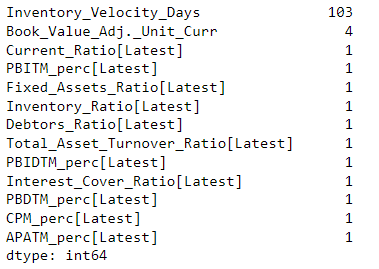


Figure 5 - Missing values

There were a total 118 missing records around 0.05% observed in the entire data. Null values were present in many columns, however significant number was present in "Inventory\_Vel\_Days" column. This is the one which we treated. There is a large number of zeros, these zeros don't add any predictive value. But it will cause "linear algebra error" when using StatsModel. Records with missing value in "Inventory\_Vel\_Days" column were imputed with the average value.

#### Columns with missing values:

(array([27, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 62], dtype=int64),)

Visually inspect the missing values in our data: A visual of all these missing values is give below - after dropping.

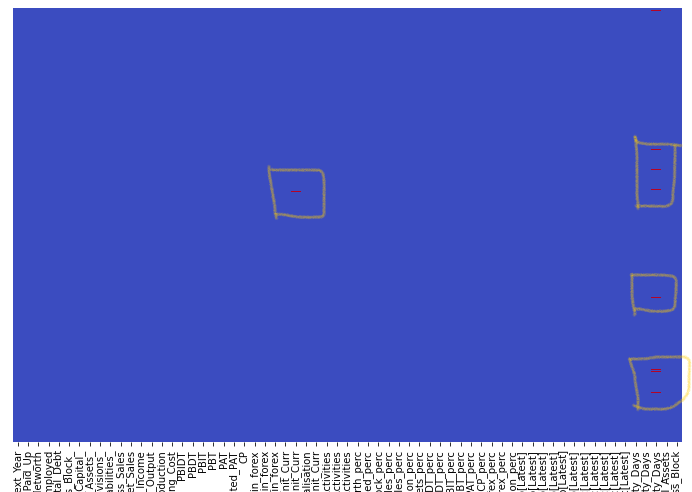


Figure 6 - Visual of missing values – After dropping

Networth\_Next\_Year 0

ROG\_Net\_Worth\_perc 0

ROG\_Gross\_Block\_perc 0

ROG\_Gross\_Sales\_perc 0

ROG\_Net\_Sales\_perc 0

ROG\_Cost\_of\_Production\_perc 0

ROG\_Total\_Assets\_perc 0

ROG\_PBIDT\_perc 0

ROG\_PBDT\_perc 0

ROG\_PBIT\_perc 0

ROG\_PBT\_perc 0

ROG\_PAT\_perc 0

ROG\_CP\_perc 0

ROG\_Revenue\_earnings\_in\_forex\_perc 0

ROG\_Revenue\_expenses\_in\_forex\_perc 0

ROG\_Market\_Capitalisation\_perc 0

Current\_Ratio[Latest] 0

Fixed\_Assets\_Ratio[Latest] 0

Inventory\_Ratio[Latest] 0

Debtors\_Ratio[Latest] 0

dtype: int64

No more missing values were present after treatment.

## Transform Target variable into 0 and 1

* There is no target variable defined – but since the objective is to build a model for investor to decode which company to invest in – the variable *Networth\_Next\_Year* could be used to transform into target variable.
* If the company’s *Networth\_Next\_Year* is positive and greater than 0 – then the company would continue to return good investment for investor and thus could be transformed as 0 – NON-DEFAULT
* If the company’s Networth\_Next\_Year is equal to 0 or less than it – then the company would likely not return a good investment to investor and transformed as 1 – DEFAULT.
* Overall distribution of defaults in the dataset:

0 3198

1. 388

* Which is to say about 11% of the companies from the dataset are likely to default and ones the investor could avoid investing in.

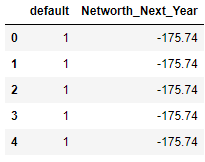
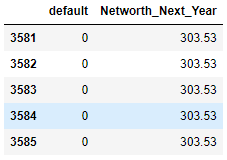
 

Figure 7 - Target Variable - First and last 5 Rows

## Univariate (4 marks) & Bivariate analysis with proper interpretation. (You may choose to include only those variables which were significant in the model building)

### Univariate Analysis:

Basic measures of Descriptive Statistics for the continuous variables.

We performed the descriptive summary for the company data. Since most of the column data is continuous, we can see the mean, standard deviation and percentile details for all the columns.

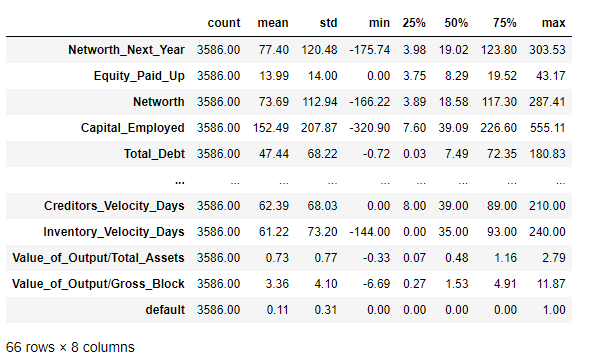


Figure 8 - Descriptive Statistics Summary

Target Variable Default:

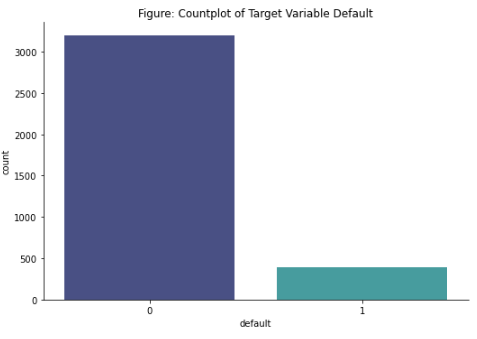


Figure 9 - Count plot of target variable Default

Some of Important Feature Box Plot

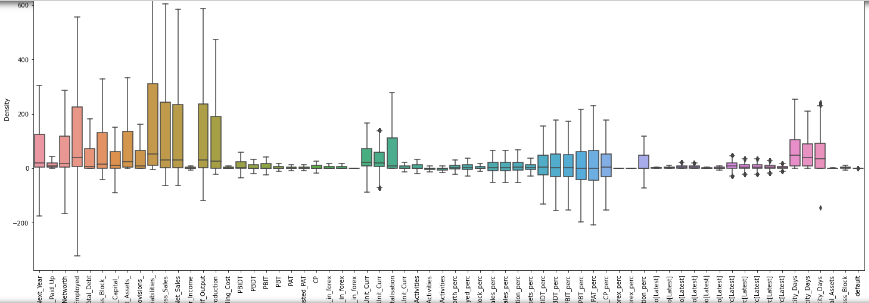
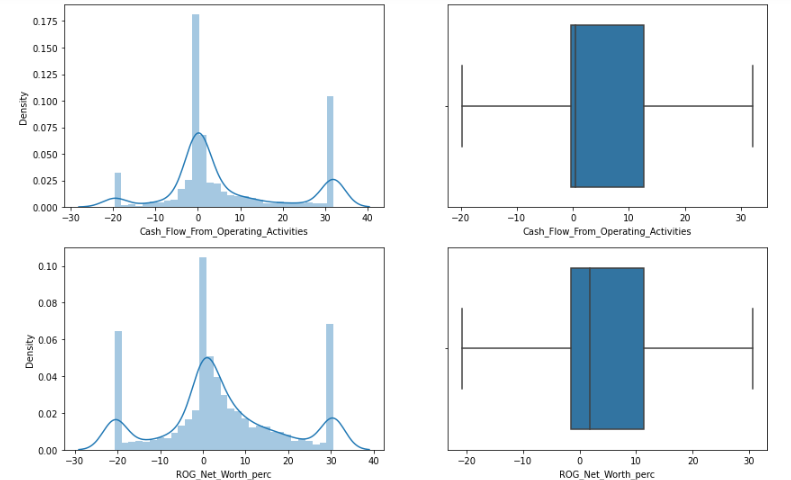
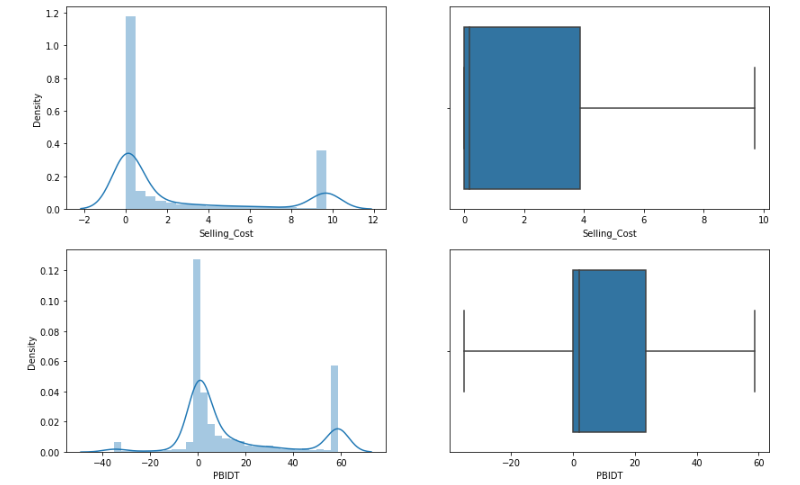


Figure 10 - Important Feature variables Box Plot

* Variable 'Total\_Asset\_To\_Liabilities still have some extreme values.
* Varible 'Capital\_Employes'still have some Extreme and Lower values.

#### Significate Scaled Feature Variables - Distribution of column with Displot & Box plot:

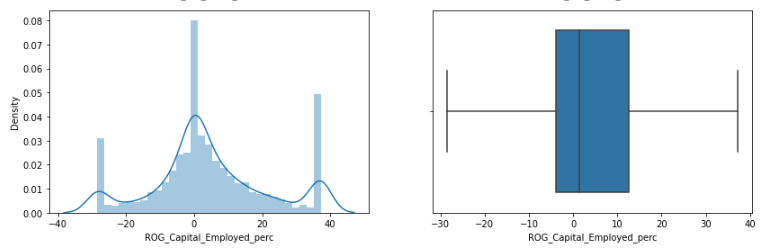
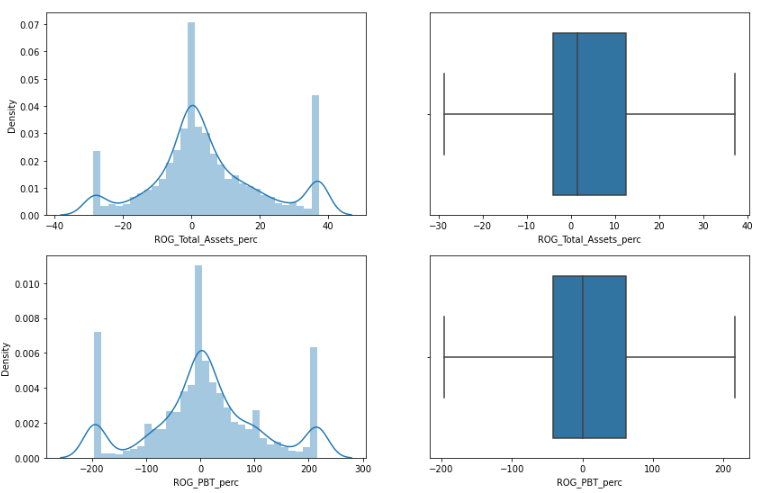
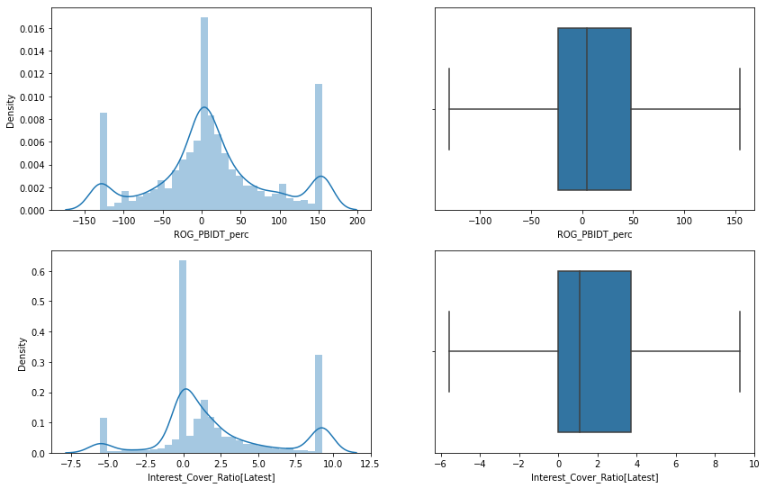


Figure 11 - Distribution of column with Displot & Box plot

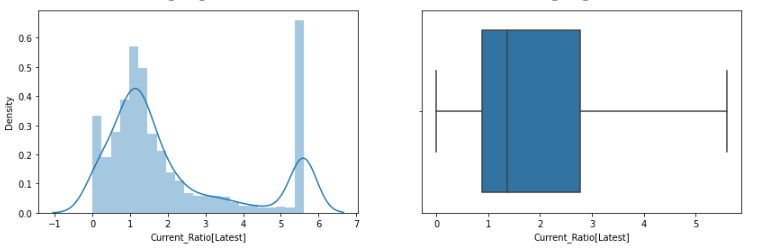
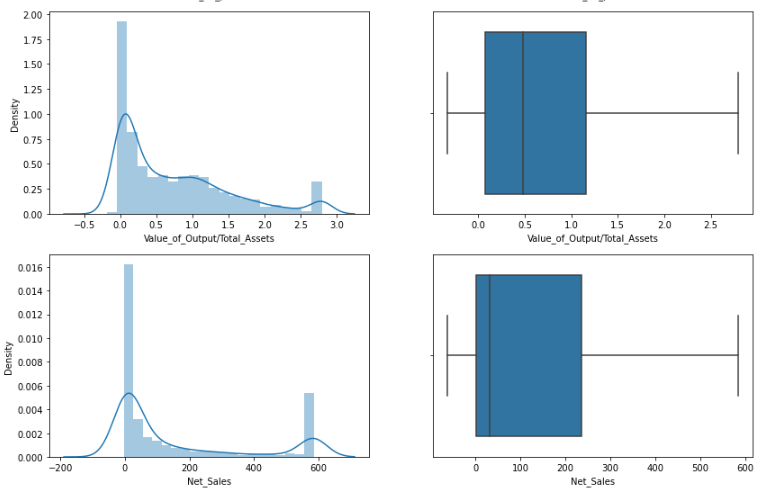
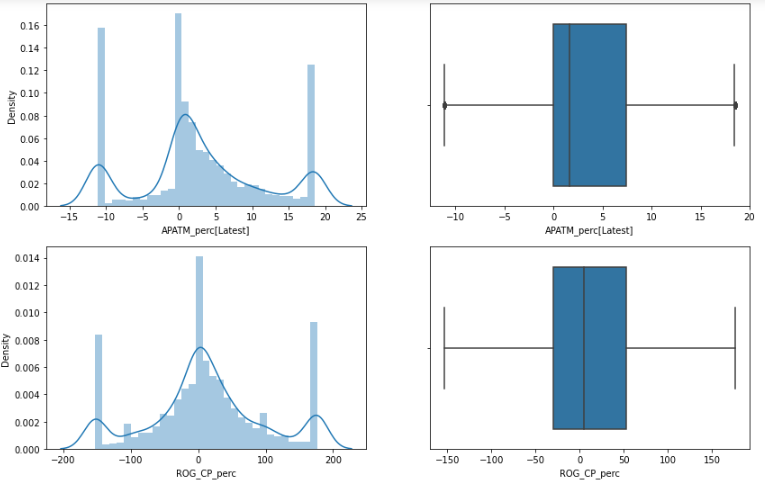


Figure 12 - Distribution of column with Displot & Box plot



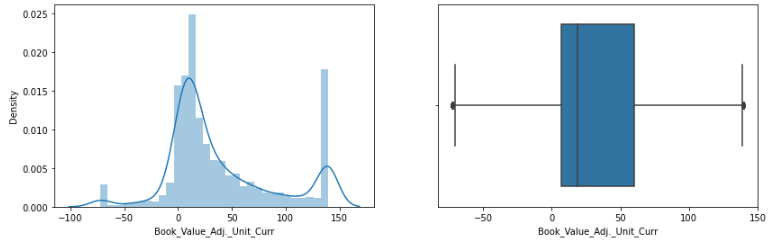


Figure 13 - Distribution of column with Displot & Box plot

### Inference:

1. ‘Selling Cost’ - has max companies around its mean. They have Right Skew with outliers on higher side.
2. ‘Cash Flow from Operating Activities’ - normal distribution with max companies lying around the mean.  
   ‘PBIDT’ - ‘Profit before Int Depreciation and Tax’ - max companies are around the mean with a prominent right skew. This indicates that there are still many companies with high PBIDT
3. ‘ROG Networth’, ‘ROG Capital Employed’, ‘ROG Total Assets’, ‘ROG PBIDT’, ‘ROG PBT (Profit Before Tax)’, ‘ROG CP’, ‘Current ratio Latest’, ‘Interest Cover Ratio Latest’, ‘Value of Output to Total Assets’, ‘Net Sales, ‘Book Value Adjusted’ - these variables have max density of companies around its mean with right skew. This indicates outliers on the higher side.
4. ‘APATM (After Tax Profit Margin)’ - has max density around its mean and a prominent left skew. This indicates that there are many companies have their Net Profit on the lower side of the distribution - Possible indication of default.
5. Mostly, it is observed that there are many companies with good margin and financials before tax and all other costs. But, after costs are considered, they slide to the lower half - Shows they need to work on their costs and bottom line.

### Bivariate Analysis:

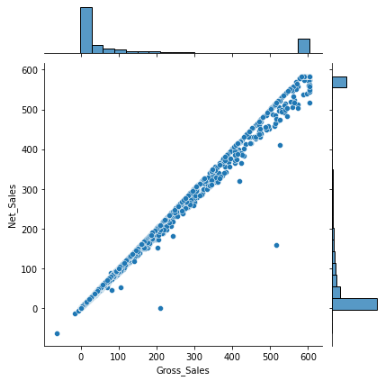
 There exists linear relationship between these two important variables.

Figure 14 - Gross Sales Vs Net Sales

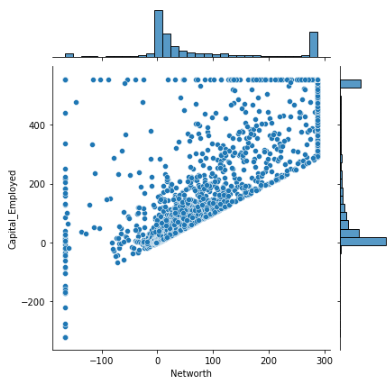
As the capital increases, net worth also increases, but in some cases, capital seems to be disbursed even for lesser net worth.

Figure 15 - Net worth Vs Capital Employment

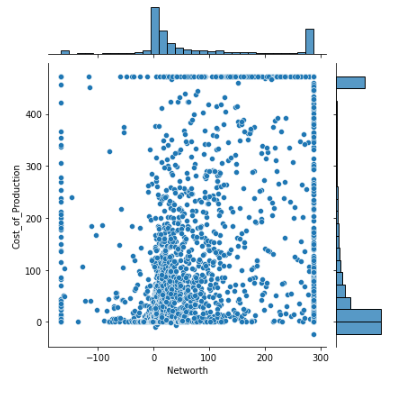
This plot is scattered and there exists no such relationship between these two variables. 

Figure 16 - Net worth Vs Cost of Production

#### Correlation Heatmap:

As per Regression Feature Elimination (RFE) we got Top important Variables and 1 Target Variable. These were significant in the model building.

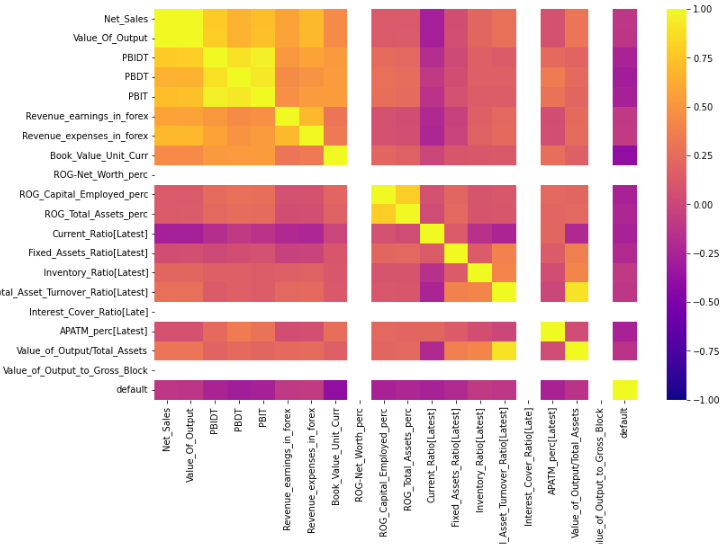


Figure 17 - Correlation heat-map - Top significant Variables

Correlation heat-map of top 19 predictors and 1 Target, used to get the best correlation is given above.

Fixed\_Asset\_Ratio(Latest) and Value\_of\_output\_to\_gross\_block & Total\_Asset\_Turnover\_Ration (Latest) and Value\_of\_output\_to\_Total\_Assets :- The above pairs of features show high correlation, it looks obvious as they seem derived or direct functions of each other

Target variable ‘default’ has high negative correlation with Book\_Value\_unit\_curr. This indicates as Book Value rises, Probability of Default falls.

#### Inferences from Univariate and Bi-variate analysis:

* Most of the variables have skewed distribution. But we will not treat those distribution by any kind of transformation or new features.
* All the variables have outliers. These outliers will be treated as we are going to apply Logistic regression to predict the outcome.
* Bi-variate analysis is performed on some of the important variables.
* We also performed multi variate analysis on the data to see if there are any correlation that are observed within the data.
* It is observed that there is negative correlation between variables.
* We observed that networth and networth next year were highly correlated. Apart from this, we also found various Rate of Growth variables were highly correlated.
* Also, we observed that there is positive correlation between variables CPM\_perc\_Latest, APATM\_perc\_Latest, PBIDTM\_perc\_Latest, PBITM\_perc\_Latest, PBDTM\_perc\_Latest.
* It is observed that there is positive correlation between variables Capital\_Employed and Current\_Assets, Current\_Assets and Total\_debt, Total\_debt and Capital\_Employed.
* Overall, high positive and negative correlation between variables can be seen above. This analysis tells us that there is a problem of collinearity with this data set.

## Train Test Split

* Split the data into Train and Test dataset in a ratio of 67:33 and use random\_state =42.
* We are splitting the data set as X (data which has independent variables) and y (data which has the predictor variable).
* Model Building is to be done on Train Dataset and Model Validation is to be done on Test Dataset.
* We use train\_test\_split function from scikit-learn library to split the data.
* Out of Total 3586 So after split, Train Set has 2402 observations & Test Set has 1184 observations.

#### Train Date First 5 Rows:

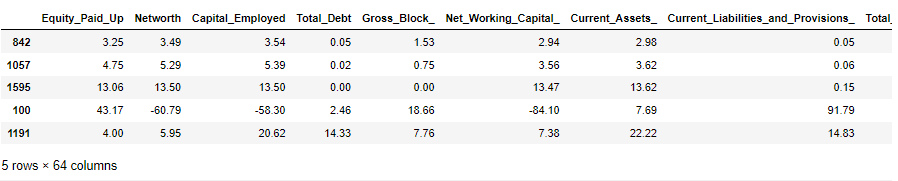


Figure 18 - Scaled Train dataset First 5 rows

#### Test Date First 5 Rows:

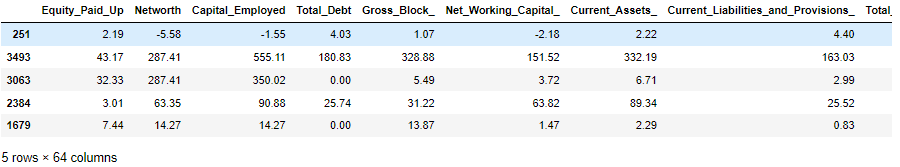


Figure 19 - Scaled Test dataset First 5 rows

#### Training Data Columns :

Index(['Equity\_Paid\_Up', 'Networth', 'Capital\_Employed', 'Total\_Debt',

'Gross\_Block\_', 'Net\_Working\_Capital\_', 'Current\_Assets\_',

'Current\_Liabilities\_and\_Provisions\_', 'Total\_Assets/Liabilities\_',

'Gross\_Sales', 'Net\_Sales', 'Other\_Income', 'Value\_Of\_Output',

'Cost\_of\_Production', 'Selling\_Cost', 'PBIDT', 'PBDT', 'PBIT', 'PBT',

'PAT', 'Adjusted\_PAT', 'CP', 'Revenue\_earnings\_in\_forex',

'Revenue\_expenses\_in\_forex', 'Capital\_expenses\_in\_forex',

'Book\_Value\_Unit\_Curr', 'Book\_Value\_Adj.\_Unit\_Curr',

'Market\_Capitalisation', 'CEPS\_annualised\_Unit\_Curr',

'Cash\_Flow\_From\_Operating\_Activities',

'Cash\_Flow\_From\_Investing\_Activities',

'Cash\_Flow\_From\_Financing\_Activities', 'ROG\_Net\_Worth\_perc',

'ROG\_Capital\_Employed\_perc', 'ROG\_Gross\_Block\_perc',

'ROG\_Gross\_Sales\_perc', 'ROG\_Net\_Sales\_perc',

'ROG\_Cost\_of\_Production\_perc', 'ROG\_Total\_Assets\_perc',

'ROG\_PBIDT\_perc', 'ROG\_PBDT\_perc', 'ROG\_PBIT\_perc', 'ROG\_PBT\_perc',

'ROG\_PAT\_perc', 'ROG\_CP\_perc', 'ROG\_Revenue\_earnings\_in\_forex\_perc',

'ROG\_Revenue\_expenses\_in\_forex\_perc', 'ROG\_Market\_Capitalisation\_perc',

'Current\_Ratio[Latest]', 'Fixed\_Assets\_Ratio[Latest]',

'Inventory\_Ratio[Latest]', 'Debtors\_Ratio[Latest]',

'Total\_Asset\_Turnover\_Ratio[Latest]', 'Interest\_Cover\_Ratio[Latest]',

'PBIDTM\_perc[Latest]', 'PBITM\_perc[Latest]', 'PBDTM\_perc[Latest]',

'CPM\_perc[Latest]', 'APATM\_perc[Latest]', 'Debtors\_Velocity\_Days',

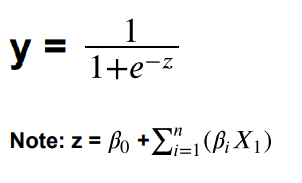
'Creditors\_Velocity\_Days', 'Inventory\_Velocity\_Days',

'Value\_of\_Output/Total\_Assets', 'Value\_of\_Output/Gross\_Block'],

dtype='object')

## Build Logistic Regression Model (using statsmodel library) on most important variables on Train Dataset and choose the optimum cutoff. Also showcase your model building approach

The equation of the Logistic regression by which we predict the corresponding probabilities and then go on predict a discrete target variable is



Build Logistic Regression Model (using statsmodel library) on most important variables on Train Dataset and choose the optimum cut-off. Also showcase your model building approach.

For model building, we try to approach Recursive Feature Elimination (RFE) and we want to select top 21 features (1/3rd of total feature variables) that would contribute to the model well.

We give weightage to each variable and based on the weightage; rankings are provided. For modelling, we will use Logistic Regression will recursive feature elimination.

#### Applying GridSearchCV for Logistic Regression :

grid\_search.best\_params\_ and grid\_search.best\_estimator\_ are as follows :

{'penalty': 'none', 'solver': 'lbfgs', 'tol': 0.0001}

LogisticRegression(max\_iter=10000, n\_jobs=2, penalty='none')

## Validate the Model on Test Dataset and state the performance matrices. Also state Interpretation from the model

We train the model and then validate the model in both the training and testing sets. We are plotting the confusion matrix and classification report for both sets.

We could see high precision and accuracy, but the recall seems to be less in the training data.

We need to improve the recall value as that would give us True Positives (TP), which in turn means that , we will correctly identify the defaulters accurately, because if we miss a defaulter, that would account to the bank paying higher interests to the existing debts and cash flow will not be regularized in the bank.

#### Confusion matrix and Classification Report for the training set:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | precision | recall | f1-score | support |
|  |  |  |  |  |
| 0 | 0.97 | 0.99 | 0.98 | 2142 |
|  |  |  |  |  |
| 1 | 0.88 | 0.75 | 0.81 | 260 |
|  |  |  |  |  |
| accuracy |  |  | 0.96 | 2402 |
| macro avg | 0.93 | 0.87 | 0.89 | 2402 |
| weighted avg | 0.96 | 0.96 | 0.96 | 2402 |

Table 3 - Confusion matrix for training set

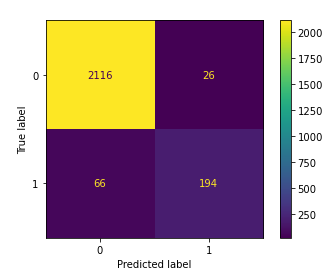


Figure 20 - Classification Report for the training set

We train the model and then validate the model in both the training and testing sets.

#### Confusion matrix and Classification Report for the test set :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | precision | recall | f1-score | support |
|  |  |  |  |  |
| 0 | 0.96 | 0.98 | 0.97 | 1056 |
| 1 | 0.81 | 0.7 | 0.75 | 128 |
|  |  |  |  |  |
| accuracy |  |  | 0.95 | 1184 |
| macro avg | 0.89 | 0.84 | 0.86 | 1184 |
| weighted avg | 0.95 | 0.95 | 0.95 | 1184 |

Table 4 - Confusion matrix for test set

We could see high precision and accuracy, but the recall seems to be less in the testing set.

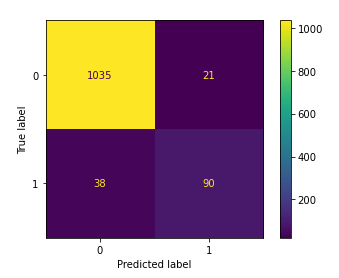


Figure 21 - Testing Set - Confusion matrix

Accuracy of over 95% was achieved, while precision, recall and f1 score were also very high at 96%, 98% and 97% respectively.

We could see for both the models, accuracy and precision (ratio of True Positive’s to the entire Positive’s) is on the higher side, but the recall seems to be downfall in both these sets.

This seems to be the case because we had an imbalanced dataset for our model.

* So, we have balanced our default values (the ratio of 0’s to 1’s is to be increased) in our dataset, we only had 11% of the defaults, we try to balance the dataset using SMOTE technique before fitting it in our model.
* After applying the SMOTE technique, we fit the model and predict our values in both training and testing sets.

#### Classification Report for the Training Set (SMOTE):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | precision | recall | f1-score | support |
|  |  |  |  |  |
| 0 | 0.94 | 0.93 | 0.93 | 2142 |
| 1 | 0.93 | 0.94 | 0.94 | 2142 |
|  |  |  |  |  |
| accuracy |  |  | 0.94 | 4284 |
| macro avg | 0.94 | 0.94 | 0.94 | 4284 |
| weighted avg | 0.94 | 0.94 | 0.94 | 4284 |

Table 5 - SMOTE Training Set - Classification Report

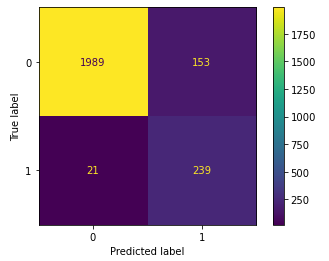


Table 6 - SMOTE Training Set - Confusion matrix

We could see that the recall has improved greatly, so the chances of identifying our defaulters has significantly improved and there is less chance of the model missing out on any potential default candidates/companies to our bank.

#### Classification Report for the Testing Set (SMOTE):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | precision | recall | f1-score | support |
|  |  |  |  |  |
| 0 | 0.99 | 0.93 | 0.96 | 1056 |
| 1 | 0.61 | 0.92 | 0.73 | 128 |
|  |  |  |  |  |
| accuracy |  |  | 0.93 | 1184 |
| macro avg | 0.8 | 0.92 | 0.84 | 1184 |
| weighted avg | 0.95 | 0.93 | 0.93 | 1184 |

Table 7 - SMOTE Testing Set - Classification Report

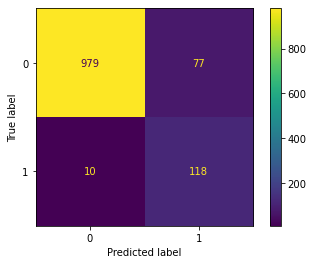


Figure 22 - SMOTE Testing Set - Confusion matrix

Accuracy of 92% and, precision, high recall and f1 score of 99% , 92% and 95% respectively were also observed on the test set.

Finally, we are able to achieve a descent recall value without overfitting. Considering the opportunities such as outliers, missing values and correlated features this is a fairly good model. It can be improved if we get better quality data where the features explaining the default are not missing to this extent. Of course, we can try other techniques which are not sensitive towards missing values and outliers.

### Interpretation:

This clearly indicates that the model which has been built is highly efficient and has been able to capture the correct variable for prediction. It has been proven to work on train as well as test data.

Re-call value for the testing & the training set are near and this model is the best suited to identify correct defaulters because of high-recall value in both sets.

Precision seem to be on a lower side for the sets because of the SMOTE technique as we try to create more values to balance the defaulter ratio.

But, in this model, recall seems to be an important factor as we stress on identifying the defaulters accurately.

From Multi-variate Analysis, we observed that many companies had good profit margins before considering taxes, interests and other costs. But once all costs are considered alongwith taxes and depreciation, majority of these companies slide to the bottom half in Profitability. These companies should focus on optimizing their bottom line.