Project: India Credit Risk Model

Table of Content

1. Project Objective

The objective of this project is to build the India Credit Risk model. The project was build using the logistic regression on the raw data available at raw-data.xlsx file. Raw data contains 52 different attributes such as: Net worth, Total Income, Shareholders fund, PBT as % of total income, PAT as % total income etc. The model which is being build here predicts whether the customer/firm/organization to whom the loan is extended by the bank will default or not default.

2. Exploratory Data Analysis

2.1 Basic Data Summary

The raw data is available in *raw-data.xlsx* file. Data given relates the characteristics or behavior of a company/customer who applies for the loan. Raw data contains 52 different attributes such as: Net worth, Total Income, Shareholders fund, PBT as % of total income, PAT as % total income etc. Attributes named Net worth next year is converted into a Default variable which is a binary variable containing '0' if the net worth next year is positive and '1' if the Net worth next year is negative. The Default rate is also calculated to be 6.8%.

Table-1: Basic data summary

```
Classes 'tbl_df', 'tbl' and 'data.frame':     3541 obs. of  52 variables:
 $ Num                                         : num  1 2 3 4 5 6 7 8 9 10 ...
 $ Networth Next Year                          : num  8890.6 394.3 92.2 2.7 ...
 $ Total assets                                : num  17512.3 941 232.8 2.7 ...
 $ Net worth                                   : num  7093.2 351.5 100.6 ...
 $ Total income                                : num  24965 1527 477 NA 1580 ...
 $ Change in stock                             : num  235.8 42.7 -5.2 NA -17 ...
 $ Total expenses                              : num  23658 1455 479 NA 1558 ...
 $ Profit after tax                            : num  1543.2 115.2 -6.6 NA ...
 $ PBDITA                                      : num  2860.2 283 5.8 NA 31 ...
 $ PBT                                         : num  2417.2 188.4 -6.6 NA ...
 $ Cash profit                                 : num  1872.8 158.6 0.3 NA ...
 $ PBDITA as % of total income                 : num  11.46 18.53 1.22 0 1.96 ...
 $ PBT as % of total income                    : num  9.68 12.33 -1.38 0 0.4 ...
 $ PAT as % of total income                    : num  6.18 7.54 -1.38 0 0.35 2.81
 $ Cash profit as % of total income            : num  7.5 10.38 0.06 0 0.75 ...
 $ PAT as % of net worth                       : num  23.78 38.08 -6.35 0  ...
 $ Sales                                       : num  24458 1504 476 NA 1575 ...
 $ Income from financial services              : num  158 4 1.5 NA 3.9 NA 7.3...
 $ Other income                                : num  297.2 15.9 0.2 NA 0.9 ...
 $ Total capital                               : num  423.8 115.5 81.4 0.5  ...
 $ Reserves and funds                          : num  6822.8 257.8 19.2 2.2 ...
 $ Deposits (accepted by commercial banks)     : logi  NA NA NA NA NA NA ...
```

```
$ Borrowings                                    : num  14.9 272.5 35.4 NA 1 ...
$ Current liabilities & provisions              : num  9965.9 210 96.8 NA ...
$ Deferred tax liability                        : num  284.9 85.2 NA NA 4.6 ...
$ Shareholders funds                            : num  7093.2 351.5 100.6 107.6...
$ Cumulative retained profits                   : num  6263.3 247.4 32.4 2.2 ...
$ Capital employed                              : num  7108.1 624 136 2.7  ...
$ TOL/TNW                                       : num  1.33 1.23 1.44 0 2.83 1.8
$ Total term liabilities / tangible net worth: num  0 0.34 0.29 0 1.59 0.37
$ Contingent liabilities / Net worth (%)        : num  14.8 19.2 45.8 0 34.9 ...
$ Contingent liabilities                        : num  1049.7 67.6 46.1 NA ...
$ Net fixed assets                              : num  1900.2 286.4 38.7 94.8 ...
$ Investments                                   : num  1069.6 2.2 4.3 NA 7.4 ...
$ Current assets                                : num  13277.5 563.9 167.5  ...
$ Net working capital                           : num  3588.5 203.5 59.6 215.8 ...
$ Quick ratio (times)                           : num  1.18 0.95 1.11 NA 1.41 0.48
$ Current ratio (times)                         : num  1.37 1.56 1.55 NA 2.54 1.27
$ Debt to equity ratio (times)                  : num  0 0.78 0.35 0 1.79 1.09
$ Cash to current liabilities (times)           : num  0.43 0.06 0.21 NA 0 0.11 NA
$ Cash to average cost of sales per day         : num  68.21 5.96 17.07 NA 0 ...
$ Creditors turnover                            : chr  "3.62" "9.8000000000000007"
$ Debtors turnover                              : chr  "3.85" "5.7" "5.07" "0" ...
$ Finished goods turnover                       : chr  "200.55" "14.21" "9.24"...
$ WIP turnover                                  : chr  "21.78" "7.49" "0.23" ...
$ Raw material turnover                         : chr  "7.71" "11.46" NA "0" ...
$ Shares outstanding                            : chr  "42381675" "11550000"...
$ Equity face value                             : chr  "10" "10" "10" "10" ...
$ EPS                                           : num  35.52 9.97 -0.5 0 7.91 ...
$ Adjusted EPS                                  : num  7.1 9.97 -0.5 0 7.91 ...
$ Total liabilities                             : num  17512.3 941 232.8 478.5 ...
$ PE on BSE                                     : chr  "27.31" "8.17" "NA" ...
```

2.2 Bivariate Analysis

The data represents the Bivariate Analysis as the dependent variable (i.e. Net worth next year) depends upon more than two variables.

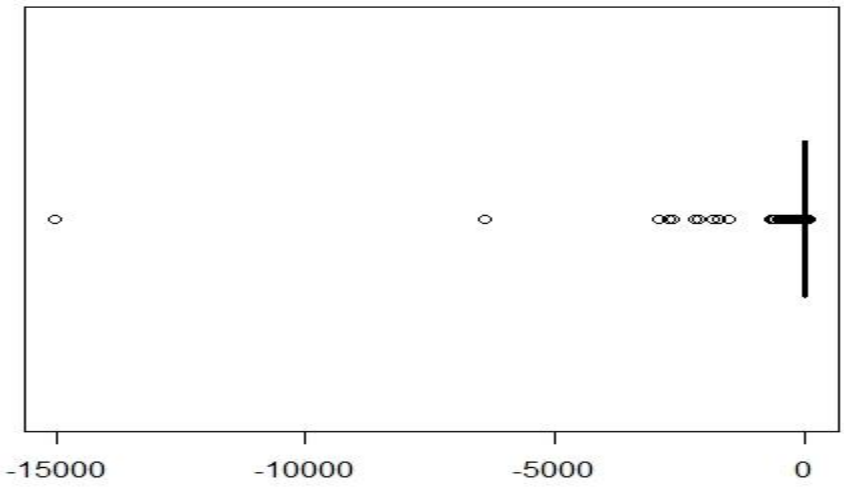
2.3 Summary of the data set

Table-2: Summary

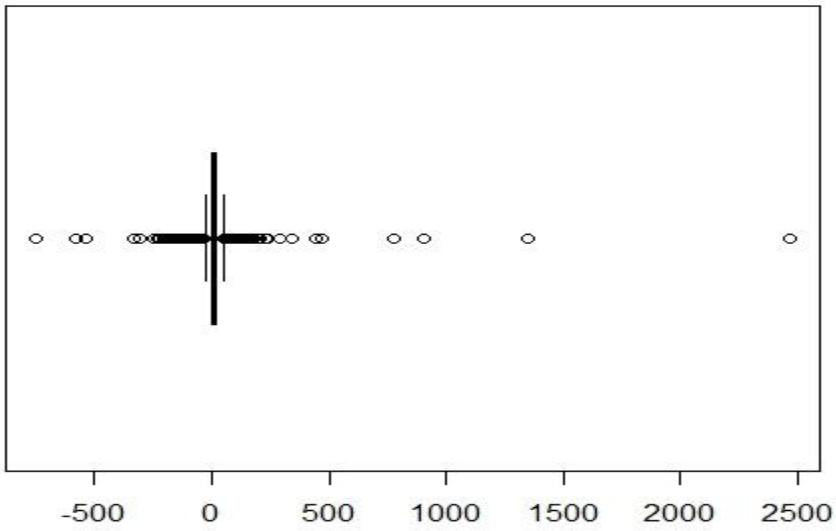
	Total assests	Net worth	Total income	Change in stock	Total expenses	PBT	PBDITA	Sales
Min	0.1	0.0	0.0	-3029.4	-0.1	-3894.8	-440.7	0.1
1st Qu	91.3	31.3	106.5	-1.8	95.8	0.70	6.9	112.7
Median	309.7	102.3	444.9	1.6	407.7	12.4	35.4	453.1
Mean	3443.4	1295.9	4582.8	41.49	4262.9	383.81	578.1	4549.5
3rd Qu	1098.7	377.3	1440.9	18.05	1359.8	71.97	150.2	1433.5
Max	1.1e06	6.1e05	2.4e06	1.4e04	2.3e06	1.4e05	2.1e05	2.4e06

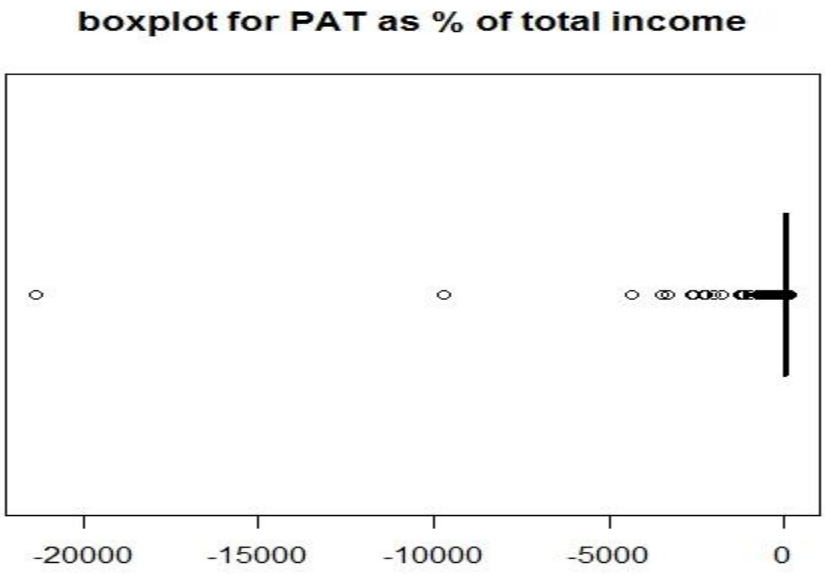
Result obtained from the five point summary of data set (Table-2) it is inferred that the data does contain missing values. (Note: summary of the data set contains 52 variables however in Table-2 only 8 variables were shown for the understanding of the data)

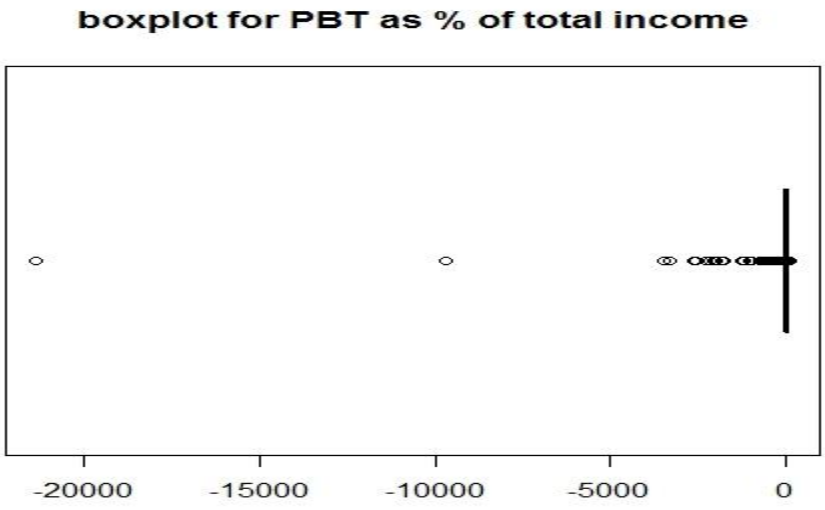
boxplot for Cash profit as % of total income



boxplot for PAT as % of net worth







2.4 Multi co-linearity

For checking whether the attributed show multi colinearity or not. The complete data set is divided into four categories i.e profit, leverage, liquidity and size. Then separately multi colinearity was tested among each group (Fig.1, Fig.2, Fig.3 and Fig.4).

Fig.1: Multi colinearity in size variables

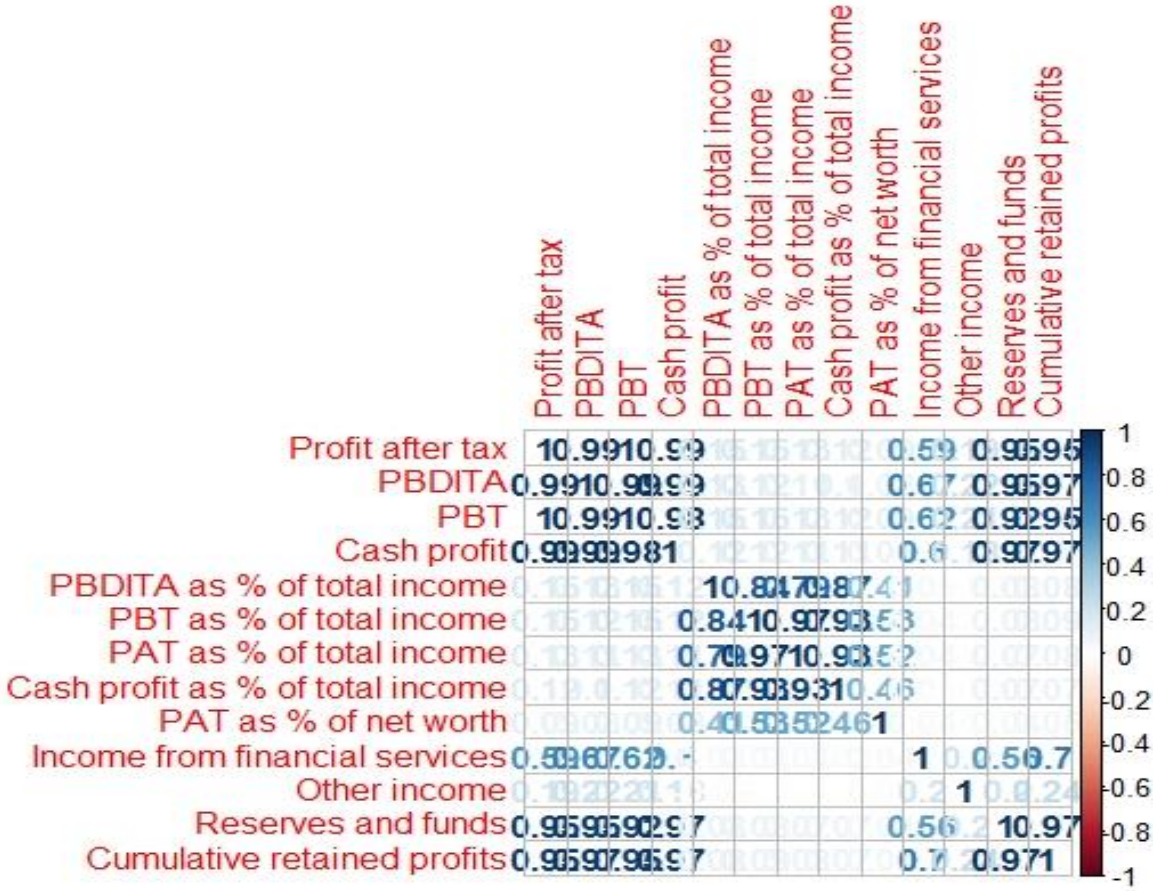


Fig.2: Multi colinearity in profit variable

Fig.3: Multi colinearity in leverage variable

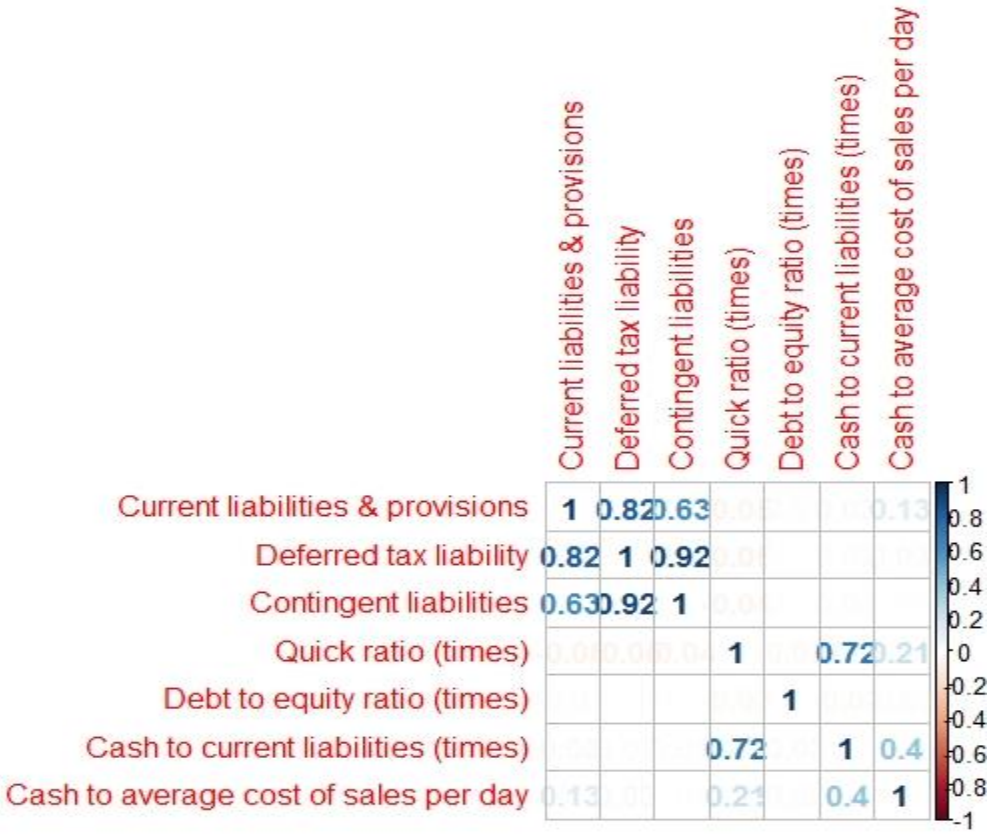


Fig.4: Multi colinearity in liquidity variable

3. Logistic Regression Model Building

3.1 Interpretation from Logistic Regression Model

- Logistic regression model was built by considering one variable from each of the four groups i.e. profit, size, leverage, liquidity.
- Subsequent models were built considering different variables.
- Finally a model (Default.model3) was obtained which contains the most important variables from each group. The most important variables were found to be Cash profit as % of total income, Total income, TOL/TNW, Cash to average cost of sales per day.
- The accuracy, sensitivity and specificity of model built is shown in Table-4.

Table-3: Confusion Matrix

	Default.prediction3	
	0	**1**
0	2988	150
1	78	121

Table-4: Logistic regression model performance

Accuracy	**Sensitivity**	**Specificity**
0.9316752	0.6080402	0.9521989

3.2 Model Validation

- Model validation has been done on the data set validation_data.xlsx
- Accuracy, sensitivity and specificity of the model validation is shown in Table-6

Table-5: Confusion Matrix for model validation

	Default.pred	
	0	**1**
0	3019	119
1	94	105

Table-4: Logistic regression model validation performance

Accuracy	Sensitivity	Specificity
0.9316702	0.5276382	0.9620778