CREDIT RISK OF BANK CUSTOMERS

- A Threshold analysis for Credit Worthiness of Bank Customers

Pradeep Panicker 8/2/20 Financial Risk Management

Table of Contents

PROJECT OBJECTIVE

Data Interpretation –

A dataset containing the CUSTOMER Financial Data, is imported into the Global environment of R Studio – for further interpretation.

Project Objective –

A CREDIT RISK MODEL is to be created to predict the Default Probability of 715 Customers.

1. EDA with Outlier Identification & Outlier Treatment.
2. Missing values treatment.
3. Multicollinearity check and solutions .
4. Use LOGISTIC Regression for Model building.
5. Optimize to the best model generation and validate on TEST Data.
6. MODEL Coefficients identification of significant factors.
7. Evaluate probability scores for DEFAULT and prepare docile gradation.

WORKING ENVIORNMENT SETUP

1. R Studio is setup for programming.
2. Both Working files- 'raw data' as 'tdata' & 'Validation data' as 'tsdata' is imported to R environment.
3. Necessary libraries are installed – DMwR; caTools; caret; rpart;dplyr;ROCR.

MISSING VALUES Treatment

The data set contains FINANCIAL DATA which has been submitted by various customers. Financial Data is usually a derivative of various basic performance parameters of the company. These basic performance parameters are not declared by the customer to the BANK and hence we do not access to such basic data.
Thereby, missing value imputation must be such that PURITY of the financial data IS NOT MANIPULATED.

Imputation of values by 'Median' / 'Mean' mode will get the financial data of each customer deviated to a greater extent.

A more statistical method – KNN Imputation – is being used wherein the missing values are generated by prediction – with relation to the nearest data values.

Train Data set ➔
Dim(tdata) = 3541 rows & 51 Columns(less col 'Deposits')
Missing values in original Train Data set – 14992 nos
Test Data set ➔
Dim(tsdata) = 715 rows & 51 Columns(less col 'Deposits')
Missing values in original Test Data set – 2786 nos

Data preparation for 'knn Imputation' –

- All variables are converted into 'numeric' since the method involves distance calculation between nearest data points.

Command - knnImputation(tsdata2, k=5) is used .

A single line command ensures all 'NA' values are replaced with the predicted values best suited as per each location of 'NA' data points.

The Value imputation was successful and files 'tdata12wona' & 'tsdata1wona' was created for further analysis.

OUTLIER TREATMENT

The data currently available has a huge volume of OUTLIERS.

Outliers had been identified visually through boxplots and the same quantified through manual calculation as well. Outliers identified is tabled as ANNEXURE- I of this report.

Being a Financial data, outlier treatment is preferred to be done through 'knn Imputation' or the least by 'Caping' method.

While attempting the 'knn imputation' method, if was found that the NAs substituted for the outlier values observed was so high that the prediction failed to be executed since 'absence of values was obstructing the distance calculation'.

Thereby, in this case the CAPING method was implied.

Also, being a FINANCIAL DATA, the Whiskers range was modified to accommodate the maximum data within the caping range ➔ the lower limit was kept at 0.05% above Minimum Value & the upper limit was kept at 0.95% below the Maximum Value.

This was defined by creating a function –

```
fun <- function(x){
  quantiles <- quantile( x, c(.05, .95 ) )
  x[ x < quantiles[1] ] <- quantiles[1]
  x[ x > quantiles[2] ] <- quantiles[2]
  x}
```

The above function was applied over the dataset – upon those variables that had outliers - such that the MINIMUM & MAXIMUM value was capped to remove outliers.

THE OUTLIER TREATMENT WAS SUCCESSFULLY ACCOMPLISHED OVER THIRTY-TWO VARIABLES WHICH INITIALLY HAD OUTLIERS.

REFER – ANNEXURE II FOR BOXPLOTS OF THREE VARIABLES SHOWING THE STATUS PRIOR TO OUTLIER TREATMENT & AFTER OUTLIER TREATMENT.

NEW VARIABLE/ GROUP CREATION – w.r.t. Financial Factor

For a comprehensive analysis, the complete set of 50 Variables[less 'Deposits' & NUM(which is customer id)] is formed into classified into FOUR major buckets –

- Profitability – 15 Variables
- Leverage – 05 Variables
- Liquidity – 14 Variables
- SIZE – 14 Variables.
- PE on BSE – evaluated independently.
- NETWORTH NEXT YEAR – Dependent Variable.

ANNEXURE -III appended will be the list of classified list.

MULTI-COLLENIARITY

For multicollinearity check, dataset with missing values imputed is utilized.

To have a more precise evaluation, INITIAL ATTEMPT was to check the multi-collinearity in-between the Financial Factor Buckets.

With this method, we could arrive as -

Leverage Bucket – The following variables were found to be independent within each other

1. Contingent liabilities / Net worth (%)
2. Total liabilities
3. Borrowings

Liquidity Bucket - The following variables were found to be independent within each other

1. Debt to equity ratio (times)
2. Cash to current liabilities (times)
3. Debtors turnover
4. Raw material turnover
5. WIP turnover
6. Cash to average cost of sales per day
7. Current liabilities & provisions
8. Finished goods turnover
9. Deferred tax liability
10. Contingent liabilities
11. Creditors turnover

Profitability Bucket - The following variables were found to be independent within each other

1. PAT as % of net worth
2. Reserves and funds
3. Cumulative retained profits
4. Income from financial services
5. Other income
6. Cash profit as % of total income
7. PAT as % of total income
8. PBDITA as % of total income
9. PBT as % of total income

SIZE Bucket - The following variables were found to be independent within each other

1. Shareholders' funds
2. Sales
3. Total capital
4. Equity face value
5. Shares outstanding
6. Change in stock
7. Investments

Further, the multicollinearity in between these THIRTY Selected variables is tested and the final list of variables for MODEL Building is arrived as below –

Criteria for selection / cut-off is => VIF VALUE Above '5' , variable is neglected.

vif(model16) – TWENTY-SIX Variables

`Debt to equity ratio (times)`	`Cash to current liabilities (times)`
1.131659	1.088015
`Contingent liabilities / Net worth (%)`	`Debtors turnover`
1.385335	1.015227
`Raw material turnover`	`WIP turnover`
1.001325	1.099242
`Cash to average cost of sales per day`	`Finished goods turnover`
1.180044	1.128658
`Creditors turnover`	`PAT as % of net worth`
1.021835	1.193069
`Other income`	`PBDITA as % of total income`
1.855106	1.915898
`Equity face value`	`Shares outstanding`
1.089558	3.470320
`Change in stock`	`PE on BSE`
1.284458	1.007684
`Net working capital`	`Deferred tax liability`
1.294176	2.783653
`Contingent liabilities`	`Income from financial services`
3.103772	2.615310
`Cash profit as % of total income`	`PAT as % of total income`
3.014299	3.207403
`PBT as % of total income`	Sales
1.948787	1.077922
`Total capital`	Investments
3.495892	2.199766

MODELLING

LOGISTICS Regression models are being built to predict the Default & NON-Default status of a Company – for the CREDIT RISK Analysis.

Multiple Model performance values will be generated by varying the threshold factor – such that the banking management can take a rightful decision as to – WHICH LEVEL OF THRESHHOLD TO BE ACCEPTED FOR CREDIT RISK , SUCH AS TO MAINTAING A HEALTHY CUSTOMER BASE.

The given data set represents a 6.9% probability that a customer will DEFAULT. Thereby. Our MODEL predictions will hover around this functional limit – and to judge the optimum performance of the MODEL.

The first LOGIT Model is generated with all the TWENTY-SIX Independent variables taken into evaluation. The Significance of the variables as well as the AUC Values are noted.

After THREE variation, A MODEL WITH ELEVEN VARIABLES , OUT OF WHICH NINE THAT ARE MOST SIGNIFICANT, AND THE AUC VALUE TO BE AT ITS BEST LOW , IS SELECTED.

REFER ANNEXURE -IV for the SUMMARY of the three variants of models built.

This MODEL built of the TRAIN Data Set is now applied on to the VALIDATION DATA SET.

Performance measure are evaluated and recorded for decision making.

COEFFICIENT ANALYSIS

From the model generated upon the validation dataset, the significant variables with its coefficient values are as below –

1. Debt. To Equity Ratio(times) – Coeff. Value = 0.0327340; P-Value = 0.001648

Coeff. Values being positive, this means that MORE THE 'DEBT. TO EQUITY RATIO', MORE ARE THE CHANCES OF THE COMPANY GETTING DEFAULT.

p-value less than 0.05, indicates the variable to be significant, 3.147 SD away from the mean.

2. OTHER INCOME – Coeff. Value = 0.0615415; p-value = 0.0117

Coeff. Values being positive, this means that HIGHER THE 'OTHER INCOME', MORE ARE THE CHANCES OF THE COMPANY GETTING DEFAULT.

p-value less than 0.05,BUT CLOSER TO 0.05, indicates the variable to be less significant, 2.52 SD away from the mean.

3. NET Working Capital – Coeff. Value = -0.00932; p-value = 0.00097

Coeff. Values being negative, this means that HIGHER THE 'NET WORKING CAPITAL', LESS ARE THE CHANCES OF THE COMPANY GETTING DEFAULT.

p-value less than 0.05, indicates the variable to be highly significant, 3.299 SD before the mean.

4. CONTINGENT Liabilities – Coeff. Value = -0.00492; p-value = 0.019017

Coeff. Values being negative, this means that HIGHER THE 'CONTINGENT LIABILITIES', LESS ARE THE CHANCES OF THE COMPANY GETTING DEFAULT.

p-value less than 0.05, indicates the variable to be significant, 2.345 SD before the mean.

5. PAT as % of NET WORTH – Coeff. Value = -0.0295389; p-value = 3.12e-05

Coeff. Values being negative, this means that HIGHER THE 'PAT as % of NET WORTH', LESS ARE THE CHANCES OF THE COMPANY GETTING DEFAULT.

p-value less than 0.05, indicates the variable to be highly significant, 4.165 SD before the mean.

6. DEFERED TAX Liability – Coeff. Value = 0.014928; p-value = 2.28e-05

Coeff. Values being positive, this means that HIGHER THE 'DEFERED TAX LIABILITY', HIGHER ARE THE CHANCES OF THE COMPANY GETTING DEFAULT.

p-value less than 0.05, indicates the variable to be highly significant, 4.236 SD away from the mean.

7. SALES – Coeff. Value = -0.0007979; p-value = 0.000836

Coeff. Values being negative, this means that HIGHER THE 'SALES', LEAST POSSIBILTY OF THE COMPANY GETTING DEFAULT.

p-value less than 0.05, indicates the variable to be highly significant, 3.341 SD before the mean.

Other variables used for the model building – 'Contingent Liabilities/Net Worth'; 'Cash to average cost of Sales per day'; 'Cash profit as % of total Income'& 'PAT as % of Total Income' is found to be NOT SIGNIFICANT in defining its IMPACT of the Dependent Variable – DEFAULT / NON-DEFAULT.

MODEL PERFORMANCE MEASURES

The Model performance is measured by its-

1. Accuracy to correctly predict the DEFAULT and the NON-DEFAULT customers over the total DATA.
2. SENSITIVITY – the efficiency to predict the NON-DEFAULT Correctly over the total NON-DEFAULT customers.
3. SPECIFICITY – the efficiency to predict the DEFAULT Correctly over the total DEFAULT Customers.

The MODEL PERFORMANCE data recorded and tabulated as below , also indicates the importance of OUTLIER Treatment has over the MODEL PREDICTIONS.

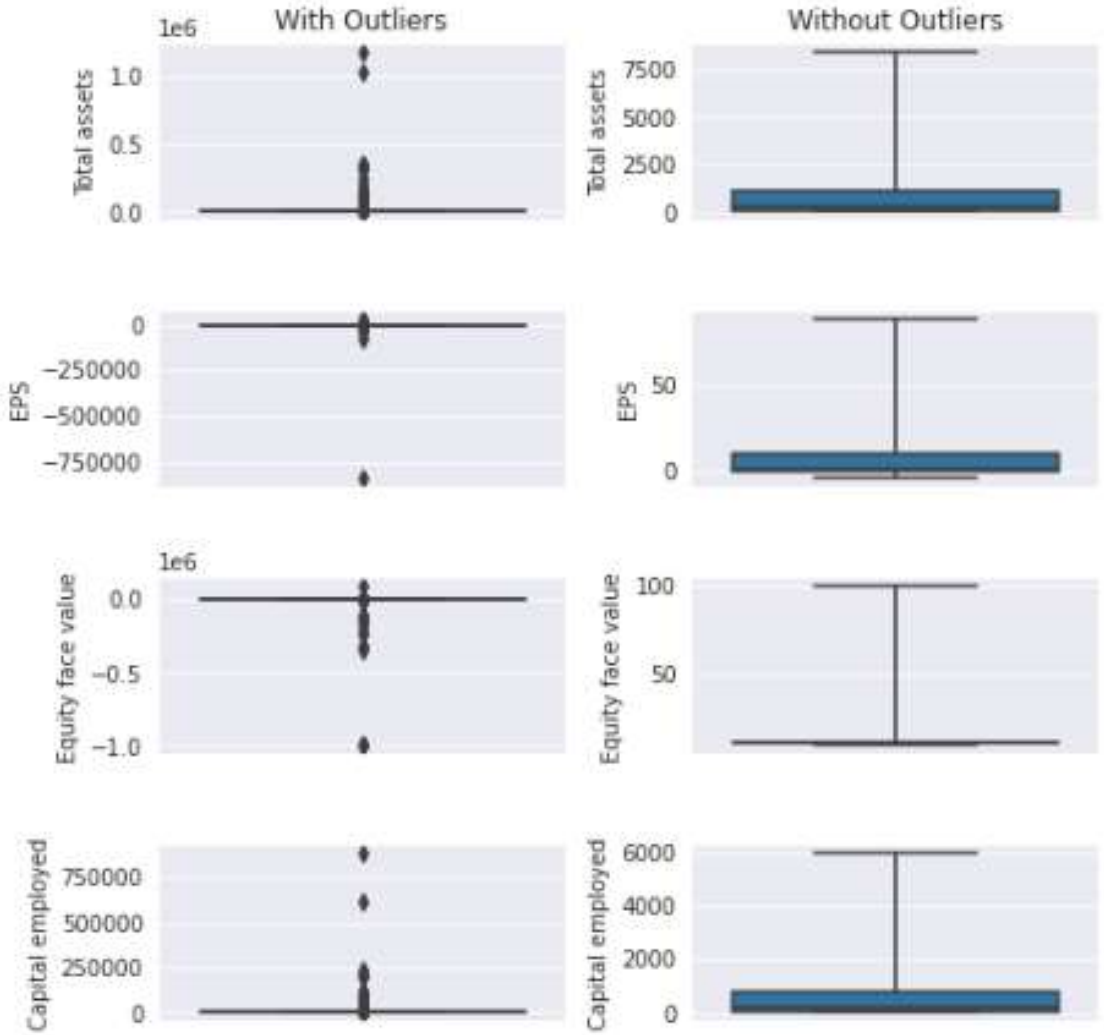
	TRAIN DATA		VALIDATION DATA	
THRESHOLD VALUE = .07	W/O Outlier Treatment	With Outlier Treatment	W/O Outlier Treatment	With Outlier Treatment
SENSITIVITY	93.45%	87.69%	93.79%	90.92%
SPECIFICITY	67.48%	73.66%	75.92%	90.74%
ACCURACY	93.13%	93.13%	92.44%	92.45%
THRESHOLD VALUE = .06				
SENSITIVITY	86.81%	84.29%	92.28%	89.56%
SPECIFICITY	76.54%	77.78%	77.78%	90.74%
ACCURACY	93.13%	93.13%	92.44%	92.45%
THRESHOLD VALUE = .058				
SENSITIVITY	82.53%	83.26%	91.53%	89.56%
SPECIFICITY	82.71%	79.01%	77.78%	90.74%
ACCURACY	93.13%	93.13%	92.44%	92.45%
THRESHOLD VALUE = .05				
SENSITIVITY	64.98%	77.93%	88.65%	88.05%
SPECIFICITY	91.35%	83.54%	81.48%	92.59%
ACCURACY	93.13%	93.13%	92.44%	92.44%

========== THE END ========

ANNEXURE – I ➔ Outlier Data in Tabulated

	1st Quart[A]	3rd Quart[B]	IQR	MIN Value	[A] - 1.5IQR	MAX. Value	[B] + 1.5IQR	Inference	Low Fence range	High Fence Range	DATA Count(Low Fence Range)	DATA Count(High Fence Range)
Capital employed	60.8	767.3	706.5	0	-998.95	891408.9	1827.05	Outlier Present	NIL	1827.05 to 891408	0	477
Test data	64.35	913.65	849.3	0.1	-1209.6	235389.5	2187.6	Outlier Present	NIL	2187.6 to 235389.5	0	94
Contingent liabilities / Net worth (%)	0	30.76	30.76	0	-46.14	14704.27	76.9	Outlier Present	NIL	76.9 to 14704.27	0	1314
Test data	0	31.49	31.49	0	-47.235	6295.24	78.725	Outlier Present	NIL	78.725 to 6295.24	0	306
EPS	0	9.6	9.6	-843181.8	-14.4	34522.5	24	Outlier Present	-14.4 to -843181.8	24 to 34522.5	68	473
Test data	0	11.46	11.46	-72750	-17.19	8784	28.65	Outlier Present	-17.19 to -72750	28.65 to 8784	9	98
Net worth	31.3	377.3	346	0	-487.7	613151.6	896.3	Outlier Present	NIL	896.3 to 613151.6	0	499
Test data	34.4	451.5	417.1	0.1	-591.25	171840	1077.15	Outlier Present	NIL	1077.15 to 171840	0	95
Total assets	91.3	1098.7	1007.4	0.1	-1419.8	1176509.2	2609.8	Outlier Present	NIL	2609.8 to 1176509	0	484
Test data	93.2	1315.3	1222.1	0.1	-1739.95	354727.3	3148.45	Outlier Present	NIL	3148.45 to 354727.3	0	97
Total liabilities	91.3	1098.7	1007.4	0.1	-1419.8	1176509.2	2609.8	Outlier Present	NIL	2609.8 to 1176509	0	484
Test data	93.2	1315.3	1222.1	0.1	-1739.95	354727.3	3148.45	Outlier Present	NIL	3148.45 to 354727.3	0	97
Total capital	13.1	100.5	87.4	0.1	-118	78273.2	231.6	Outlier Present	NIL	231.6 to 78273.2	0	446
Test data	14.1	123	108.9	0.1	-149.25	41304	286.35	Outlier Present	NIL	286.35 to 41304	0	98
Net working capital	-1	84.4	85.4	-63839	-129.1	85782.8	212.5	Outlier Present	-129.1 to -63839	212.5 to 85782.8	203	489
Test data	-1.2	100.2	101.4	-41908.3	-153.3	85782.8	252.3	Outlier Present	-153.3 to -41908.3	252.3 to 85782.8	39	94
Cumulative retained profits	1.2	202.4	201.2	-6534.3	-300.6	390133.8	504.2	Outlier Present	-300.6 to -6534.3	504.2 to 390133.8	60	516
Test data	0.8	264.2	263.4	-2582.4	-394.3	128183.1	659.3	Outlier Present	-394.3 to -2582.4	659.3 to 128183.1	14	107
Current assets	37.2	501.6	464.4	0.1	-659.4	354815.2	1198.2	Outlier Present	0	1198.2 to 354815.2	0	451
Test data	40.5	578.6	538.1	0.1	-766.65	196614.6	1385.75	Outlier Present	0	1385.75 to 196614.6	0	85
Cash to average cost of sales per day	2.88	22.29	19.41	0	-26.235	128040.76	51.405	Outlier Present	0	51.405 to 128040.8	0	487
Test data	3.355	21.715	18.36	0	-24.185	15999.17	49.255	Outlier Present	0	49.255 to 15999.17	0	105
Current liabilities & provisions	18.7	254.2	235.5	0.1	-334.55	352240.3	607.45	Outlier Present	0	607.45 to 352240.3	0	493
Test data	17.4	303.1	285.7	0.1	-411.15	112712.1	731.65	Outlier Present	0	731.65 to 112712.1	0	96
Net fixed assets	27.3	334.3	307	0	-433.2	636604.6	794.8	Outlier Present	0	794.8 to 636604.6	0	485
Test data	27.95	434.5	406.55	0.1	-581.875	115737.5	1044.325	Outlier Present	0	1044.325 to 115737.5	0	93
Cash profit	3.1	89.1	86	-2245.7	-125.9	176911.8	218.1	Outlier Present	-125.9 to -2245.7	218.1 to 176911.8	29	508
Test data	3.1	119.5	116.4	-894.6	-171.5	71681.6	294.1	Outlier Present	-171.5 to -894.6	294.1 to 71681.6	7	96
PBDITA	7.3	144.2	136.9	-440.7	-198.05	208576.5	349.55	Outlier Present	-198.1 to -440.7	349.55 to 208576.5	5	503
Test data	7.2	192.9	185.7	-393.9	-271.35	110557.1	471.45	Outlier Present	-271.35 to -393.9	471.45 to 110557.1	2	99
PBT	0.9	69.21	68.31	-3894.8	-101.565	145292.6	171.675	Outlier Present	-101.6 to -3894.8	171.68 to 145292.6	52	545
Test data	1	88.45	87.45	-993.9	-130.175	94565.2	219.625	Outlier Present	-130.2 to -993.9	219.63 to 94565.2	16	116
Profit after tax	0.6	49.2	48.6	-3908.3	-72.3	119439.1	122.1	Outlier Present	-72.3 to -3908.3	122.1 to 119439.1	63	538
Test data	0.7	67.1	66.4	-998	-98.9	62022.9	166.7	Outlier Present	-98.9 to -998	166.7 to 62022.9	10	109
Total expenses	104.1	1317.4	1213.3	-0.1	-1715.85	2366035.3	3137.35	Outlier Present	0	3137.4 to 2366035	0	452
Test data	113.4	1666.2	1552.8	0	-2215.8	1014813.1	3995.4	Outlier Present	0	3995.4 to 1014813.1	0	82
Total income	121.2	1386.1	1264.9	0	-1776.15	2442828.2	3283.45	Outlier Present	0	3283.5 to 2442828.2	0	454
Test data	122	1721.8	1599.8	0	-2277.7	1028087.7	4121.5	Outlier Present	0	4121.5 to 1028087.4	0	84
Borrowings	29.7	350.7	321	0.1	-451.8	278257.3	832.2	Outlier Present	0	832.2 TO 278257.3	0	461
Test data	32.1	449.5	417.4	0.2	-594	105175.3	1075.6	Outlier Present	0	1075.6 to 105175.3	0	94
Change in stock	-1.3	18.59	19.89	-3029.4	-31.135	14185.5	48.425	Outlier Present	-31.14 to 3029.4	48.425 to 14185.5	223	433
Test data	-3.065	18.3	21.365	-488.1	-35.1125	7540	50.3475	Outlier Present	-35.11 to -488.1	50.34 to 7540	62	96
Equity face value	10	10	0	-999999	10	100000	10	Outlier Present	not a valid data ranage for outlier evaluation			
Shares outstanding	2200000	11000000	8800000	-2147000000	-11000000	4130000000	24200000	Outlier Present	-11000000 to 2147000000	24200000 to 4130000000	4	432
Test data	2244000	12020000	9776000	128000	-12420000	4130000000	26684000	Outlier Present	0	26684000 to 4130000000	0	99
Finished goods turnover	9.97	41.61	31.64	-0.09	-37.49	17947.6	89.07	Outlier Present	0	89.07 to 17947.6	0	386
Test data	9.97	38.56	28.59	-0.09	-32.915	5614.8	81.445	Outlier Present	0	81.445 to 5614.8	0	93
Income from financial services	0.7	10.1	9.4	0	-13.4	51938.2	24.2	Outlier Present	0	24.2 to 51932.2	0	471
Test data	0.85	13.85	13	0.1	-18.65	8097.2	33.35	Outlier Present	0	33.35 to 8097.2	0	102
Deferred tax liability	6.6	51.3	44.7	0.1	-60.45	72796.6	118.35	Outlier Present	0	118.35 to 72796.6	0	423
Test data	6.75	83.11	76.36	0.1	-107.79	27077.9	197.65	Outlier Present	0	197.65 to 27077.9	0	66
Contingent liabilities	13.8	138.4	124.6	0.1	-173.1	559506.8	325.3	Outlier Present	0	325.3 to 559506.8	0	465
Test data	11.85	177.51	165.66	0.1	-236.64	72620.8	426	Outlier Present	0	426 to 72620.8	0	82
Other income	0.7	5.5	4.8	0	-6.5	42856.7	12.7	Outlier Present	0	12.7 to 42856.7	0	398
Test data	0.9	9.7	8.8	0	-12.3	42856.7	22.9	Outlier Present	0	22.9 to 42856.9	0	67
Investments	3.82	69.59	65.77	0	-94.835	199978.6	168.245	Outlier Present	0	168.245 to 199978.6	0	56
Test data	4.1	144	139.9	0	-205.75	88047.8	353.85	Outlier Present	0	353.85 to 88047.8	0	9
Adjusted EPS	0	7.5	7.5	-843181.8	-11.25	34522.5	18.75	Outlier Present	-11.25 to -843181.8	18.75 to 34522.5	81	500
Test data	0	8.35	8.35	-72750	-12.525	8784	20.875	Outlier Present	-12.525 to -72750	20.875 to 8784	13	97
Cash profit as % of total income	2.03	10.57	8.54	-15020	-10.78	100	23.38	Outlier Present	-10.78 to -15020	23.38 to 100	172	189
Test data	1.92	10.8	8.88	-6400	-11.4	100	24.12	Outlier Present	-11.4 to -6400	24.12 to 100	42	34
PAT as % of total income	0.34	6.26	5.92	-21340	-8.54	150	15.14	Outlier Present	-8.54 to 21340	15.14 to 150	270	240
Test data	0.39	6.66	6.27	-9700	-9.015	100	16.065	Outlier Present	-9.015 to -9700	16.065 to 100	60	50

ANNEXURE – II ➔ BOXPLOT Samples



ANNEXURE – III ➔ VARIABLES GROUPING

ANNEXURE - III

Variable Name	Type	NAs	OUTLIER PRESENCE	Discreption
Networth Next Year	Dependent Variable	0	NA	Net worth of the customer in next year
TOL/TNW	Leverage	0	NO	Total liabilities of the customer divided by Total net worth
Total term liabilities / tangible net worth	Leverage	0	NO	Short + long term liabilities divided by tangible net worth
Contingent liabilities / Net worth (%)	Leverage	0	YES	Contingent liabilities / Net worth
Total liabilities	Leverage	0	YES	Sum of all type of liabilities
Borrowings	Leverage	366	YES	Total amount borrowed by customer
Debt to equity ratio (times)	Liquidity	0	NO	Total liabilities divided by its shareholder equity
Cash to current liabilities (times)	Liquidity	93	NO	Total liquid cash divided by current liabilities
Current ratio (times)	Liquidity	93	NO	Current assets divided by current liabilities
Quick ratio (times)	Liquidity	93	NO	Total cash divided by current liabilities
Debtors turnover	Liquidity	328	NO	Net credit sales divided by average accounts receivable
Raw material turnover	Liquidity	361	NO	Cost of goods sold is divided by the average inventory for the same period
WIP turnover	Liquidity	640	NO	The cost of goods sold for a period divided by the average inventory for that period
Cash to average cost of sales per day	Liquidity	85	YES	Total cash divided by average cost of the sales
Current liabilities & provisions	Liquidity	96	YES	current liabilities of the customer
Finished goods turnover	Liquidity	740	YES	Annual sales divided by average inventory
Deferred tax liability	Liquidity	1140	YES	Future income tax customer will pay because of the current transaction
Contingent liabilities	Liquidity	1188	YES	Liabilities because of uncertain events
Creditors turnover	Liquidity	333		Net credit purchase divided to average trade creditors
Net working capital	Liquidity/Size	32	YES	Difference of current liabilities and current assets
PE on BSE	Market Sentiment	23	NO	Company current stock price divided by its earning per share
PAT as % of net worth	Profit	0	NO	PAT / Net worth
Reserves and funds	Profit	85	NO	Total reserves and funds of the customer
EPS	Profit	0	YES	Net income divided by total number of outstanding share
Cumulative retained profits	Profit	38	YES	Total cumulative profit retained by customer
Cash profit	Profit	131	YES	Total Cash profit
PBDITA	Profit	131	YES	Profit before depreciation, income tax and amortization
PBT	Profit	131	YES	Profit before tax deduction
Profit after tax	Profit	131	YES	Profit after tax deduction
Income from financial services	Profit	935	YES	Income from financial services
Other income	Profit	1295	YES	Income from other sources
Adjusted EPS	Profit	0	-YES	Adjusted net earning divided by the weighted average number of common share outstanding on a diluted basis during the
Cash profit as % of total income	Profit	68	-YES	Cash Profit / Total income
PAT as % of total income	Profit	68	-YES	PAT / Total income
PBDITA as % of total income	Profit	68	-YES	PBDITA / Total income
PBT as % of total income	Profit	68	-YES	PBT / Total income
Deposits (accepted by commercial banks)	Profit/ Size	3541	NO DATA	All blank values
Shareholders funds	Size	0	NO	Amount of equity in a company, which is belong to
Sales	Size	259	NO	Sales done by customer
Capital employed	Size	0	YES	Current asset minus current liabilities
Net worth	Size	0	YES	Net worth of the customer of present year
Change in stock	Change in Size	458	YES	difference between value of current stock and the value of stock in last trading day
Total assets	Size	0	YES	Total assets of customer
Total capital	Size	4	YES	Total capital of the customer
Current assets	Size	66	YES	Assets that are expected to be converted to cash within a
Net fixed assets	Size	118	YES	purchase price of all fixed assets
Total income	Size	198	YES	Total income of the customer
Equity face value	Size	692	YES	cost of the equity at the time of issuing
Shares outstanding	Size	692	YES	Number of issued shares minus the number of share held in the company
Investments	Size	1435	YES	Total invested amount
Total expenses	Size/Cost	139	YES	Total expense done by customer

ANNEXURE – IV ➔ LIST OF MODELS ON TRAIN DATA SET

MODEL 1

```
Call:
glm(formula = risk ~ `Debt to equity ratio (times)` + `Contingent liabilities / Net worth (%)` +
    `Raw material turnover` + `Cash to average cost of sales per day` +
    tdata13out$`Creditors turnover` + `Other income` + `Equity face value` +
    `Change in stock` + `Net working capital` + `Contingent liabilities` +
    `Cash profit as % of total income` + `PBT as % of total income` +
    `Total capital` + `Cash to current liabilities (times)` +
    `Debtors turnover` + `WIP turnover` + `Finished goods turnover` +
    `PAT as % of net worth` + `PBDITA as % of total income` +
    `Shares outstanding` + tdata13out$`PE on BSE` + `Deferred tax liability` +
    `Income from financial services` + `PAT as % of total income` +
    Sales + Investments, family = binomial, data = tdata13out)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.3529  -0.3029  -0.2222  -0.1266   5.6067

Coefficients:
                                        Estimate Std. Error z value Pr(>|z|)
(Intercept)                         -2.884e+00  1.755e-01 -16.434  < 2e-16 ***
`Debt to equity ratio (times)`       2.876e-02  6.188e-03   4.647 3.37e-06 ***
`Contingent liabilities / Net worth (%)`  7.369e-03  1.857e-03   3.969 7.21e-05 ***
`Raw material turnover`             -8.748e-04  3.486e-03  -0.251 0.801830
`Cash to average cost of sales per day`   3.723e-03  1.735e-03   2.146 0.031909 *
tdata13out$`Creditors turnover`     -4.368e-03  4.190e-03  -1.043 0.297162
`Other income`                       2.710e-02  1.316e-02   2.059 0.039482 *
`Equity face value`                  1.948e-03  3.342e-03   0.583 0.559981
`Change in stock`                    6.638e-04  2.757e-03   0.241 0.809727
`Net working capital`               -2.958e-03  8.519e-04  -3.472 0.000517 ***
`Contingent liabilities`            -1.655e-03  6.463e-04  -2.561 0.010436 *
`Cash profit as % of total income`  -8.324e-02  1.795e-02  -4.637 3.53e-06 ***
`PBT as % of total income`          -8.278e-05  3.620e-04  -0.229 0.819128
`Total capital`                     -1.382e-03  1.059e-03  -1.305 0.191853
`Cash to current liabilities (times)`   -1.304e-03  1.474e-02  -0.088 0.929504
`Debtors turnover`                   6.499e-04  6.767e-04   0.960 0.336851
`WIP turnover`                      -2.138e-03  2.293e-03  -0.932 0.351247
`Finished goods turnover`            1.215e-03  1.791e-03   0.679 0.497449
`PAT as % of net worth`             -9.561e-03  2.329e-03  -4.105 4.04e-05 ***
`PBDITA as % of total income`        5.013e-04  5.961e-04   0.841 0.400414
`Shares outstanding`                 1.263e-08  1.013e-08   1.246 0.212649
tdata13out$`PE on BSE`               2.921e-06  9.241e-05   0.032 0.974787
`Deferred tax liability`             3.956e-03  1.740e-03   2.273 0.023031 *
`Income from financial services`     5.882e-03  6.832e-03   0.861 0.389260
`PAT as % of total income`          -3.951e-02  1.519e-02  -2.600 0.009309 **
Sales                               -1.006e-04  5.160e-05  -1.949 0.051260 .
Investments                         -8.392e-04  1.007e-03  -0.834 0.404482
```

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1771.0 on 3540 degrees of freedom
Residual deviance: 1183.4 on 3514 degrees of freedom
AIC: 1237.4

Number of Fisher Scoring iterations: 10

MODEL 2

> summary(logit12)
Call:
glm(formula = risk ~ `Debt to equity ratio (times)` + `Contingent liabilities / Net worth (%)` +
`Cash to average cost of sales per day` + `Other income` +
`Net working capital` + `Contingent liabilities` + `Cash profit as % of total income` +
`PAT as % of net worth` + `Deferred tax liability` + `PAT as % of total income` +
Sales, family = binomial, data = tdata13out)

Deviance Residuals:
Min 1Q Median 3Q Max
-3.4526 -0.3050 -0.2283 -0.1294 5.6169

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.871e+00	1.379e-01	-20.818	< 2e-16	***
`Debt to equity ratio (times)`	3.064e-02	5.759e-03	5.320	1.04e-07	***
`Contingent liabilities / Net worth (%)`	7.215e-03	1.791e-03	4.027	5.64e-05	***
`Cash to average cost of sales per day`	4.024e-03	1.620e-03	2.484	0.013003	*
`Other income`	2.335e-02	1.225e-02	1.906	0.056627	.
`Net working capital`	-2.998e-03	8.434e-04	-3.555	0.000378	***
`Contingent liabilities`	-1.681e-03	6.092e-04	-2.760	0.005783	**
`Cash profit as % of total income`	-8.206e-02	1.796e-02	-4.569	4.90e-06	***
`PAT as % of net worth`	-9.646e-03	2.302e-03	-4.191	2.78e-05	***
`Deferred tax liability`	3.762e-03	1.549e-03	2.429	0.015138	*
`PAT as % of total income`	-3.746e-02	1.478e-02	-2.535	0.011258	*
Sales	-8.677e-05	4.685e-05	-1.852	0.064029	.

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1771.0 on 3540 degrees of freedom
Residual deviance: 1193.4 on 3529 degrees of freedom
AIC: 1217.4

Number of Fisher Scoring iterations: 9

MODEL 3

```
> summary(logit13)

Call:
glm(formula = risk ~ `Debt to equity ratio (times)` + `Contingent liabilities / Net worth (%)` +
    `Cash to average cost of sales per day` + `Net working capital` +
    `Contingent liabilities` + `Cash profit as % of total income` +
    `PAT as % of net worth` + `Deferred tax liability` + `PAT as % of total income`,
    family = binomial, data = tdata13out)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5428  -0.3050  -0.2307  -0.1318   5.6989

Coefficients:
                                        Estimate Std. Error z value Pr(>|z|)
(Intercept)                          -2.8221858  0.1344416 -20.992  < 2e-16 ***
`Debt to equity ratio (times)`        0.0307824  0.0058139   5.295 1.19e-07 ***
`Contingent liabilities / Net worth (%)`  0.0072811  0.0017885   4.071 4.68e-05 ***
`Cash to average cost of sales per day`   0.0041354  0.0016197   2.553 0.010672 *
`Net working capital`                -0.0030540  0.0008432  -3.622 0.000292 ***
`Contingent liabilities`             -0.0019736  0.0005525  -3.572 0.000354 ***
`Cash profit as % of total income`   -0.0802556  0.0180015  -4.458 8.26e-06 ***
`PAT as % of net worth`              -0.0093333  0.0022608  -4.128 3.65e-05 ***
`Deferred tax liability`              0.0035862  0.0013909   2.578 0.009928 **
`PAT as % of total income`           -0.0406094  0.0145937  -2.783 0.005391 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1771.0  on 3540  degrees of freedom
Residual deviance: 1199.4  on 3531  degrees of freedom
AIC: 1219.4

Number of Fisher Scoring iterations: 7
```