Finance and Risk Analytics

Individual Assignment –

Credit Risk

By

Yash Saxena

Contents

Problem Statement

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

Missing Value Treatment

There are **13** columns with missing values. These columns have missing value ranging from 1 to 103.

```
In [912]: len(null_columns)

Out[912]: 13
```

```
In [913]: null_columns

Out[913]: ['Book_Value_Adj_Unit_Curr',
           'Curr_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',
           'Inventory_Vel_Days']
```

Count of Missing Values –

```
Book_Value_Adj_Unit_Curr               4
Curr_Ratio_Latest                      1
Fixed_Assets_Ratio_Latest              1
Inventory_Ratio_Latest                 1
Debtors_Ratio_Latest                   1
Total_Asset_Turnover_Ratio_Latest      1
Interest_Cover_Ratio_Latest            1
PBIDTM_perc_Latest                     1
PBITM_perc_Latest                      1
PBDTM_perc_Latest                      1
CPM_perc_Latest                        1
APATM_perc_Latest                      1
Inventory_Vel_Days                   103
dtype: int64
```

Treatment –

Step #1: For 88% of companies Book Value (Adj.) (Unit Curr) is equal to Book Value (Unit Curr).

Therefore, null values in Book Value (Adj.) (Unit Curr) can be replaced with their Book Value (Unit Curr)

Step #2: Null values for columns like Curr_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 have null values corresponding to just one row i.e., pertaining to the data of company named "**G M Breweries**". As most of the columns for this company have values missing, **hence** this row will be dropped.

By performing the above two steps most of the columns with missing values, but Inventory_Vel_Days are to be taken careof.

For Inventory_Vel_Days, null values are replaced with their average.

Outlier Treatment

There are a lot of outliers in the data. **64** variables have outliers.

	Number of Outliers
ROG_Rev_exp_in_forex_perc	1615
ROG_Rev_earn_in_forex_perc	1317
Cash_Flow_From_Fin	1005
PAT	958
Adjusted_PAT	954
PBT	940
APATM_perc_Latest	933
Cash_Flow_From_Inv	877
ROG_Gross_Block_perc	830
CP	819
PBDT	814
Cash_Flow_From_Opr	802
ROG_Net_Worth_perc	747
Rev_earn_in_forex	738
Interest_Cover_Ratio_Latest	725
PBIT	720
CPM_perc_Latest	720
PBITM_perc_Latest	717
PBDTM_perc_Latest	695
Capital_exp_in_forex	694
Rev_exp_in_forex	693
PBIDT	675
ROG_Cost_of_Prod_perc	675
ROG_Gross_Sales_perc	671
ROG_Net_Sales_perc	667
Networth	650
Market_Capitalisation	639
ROG_CP_perc	637
ROG_PBDT_perc	628
Net_Working_Capital	625
ROG_PBIT_perc	616
ROG_PBIDT_perc	611
ROG_PBT_perc	611
Selling_Cost	608
CEPS_annualised_Unit_Curr	605
Other_Income	602
ROG_PAT_perc	598
Capital_Employed	596
PBIDTM_perc_Latest	595
Total_Debt	583
Curr_Liab_and_Prov	581
Curr_Assets	577
Total_Assets_to_Liab	574
ROG_Capital_Employed_perc	572
Curr_Ratio_Latest	565
Cost_of_Prod	562
Value_Of_Output	559

Value_Of_Output	559
Net_Sales	556
Gross_Sales	553
Gross_Block	540
ROG_Market_Capitalisation_perc	497
Fixed_Assets_Ratio_Latest	495
Book_Value_Unit_Curr	487
Book_Value_Adj_Unit_Curr	486
ROG_Total_Assets_perc	483
Value_of_Output_to_Gross_Block	481
Equity_Paid_Up	448
Debtors_Vel_Days	398
Creditors_Vel_Days	391
Inventory_Ratio_Latest	375
Debtors_Ratio_Latest	371
Inventory_Vel_Days	278
Total_Asset_Turnover_Ratio_Latest	201
Value_of_Output_to_Total_Assets	149

- **All the outliers are capped.**

Observation -

'Inventory Velocity (Days): Average number of days the company needs to turn its inventory into sales

```
count    3482.000000
mean       79.660827
std       137.864247
min      -199.000000
25%         0.000000
50%        35.000000
75%        96.000000
max       996.000000
Name: Inventory_Vel_Days, dtype: float64
```

Since this field cannot be negative, the negative value is replaced with 0.

Target Variable Transformation

Create a default variable that should take the value of 1 when Net worth next year is negative & 0 when Net worth next year is positive.

```
df["Networth_Next_Year"]=np.where(df["Networth_Next_Year"]>=0,0,1)
```

```
df["Networth_Next_Year"]=np.where(df["Networth_Next_Year"]>=0,0,1)
```

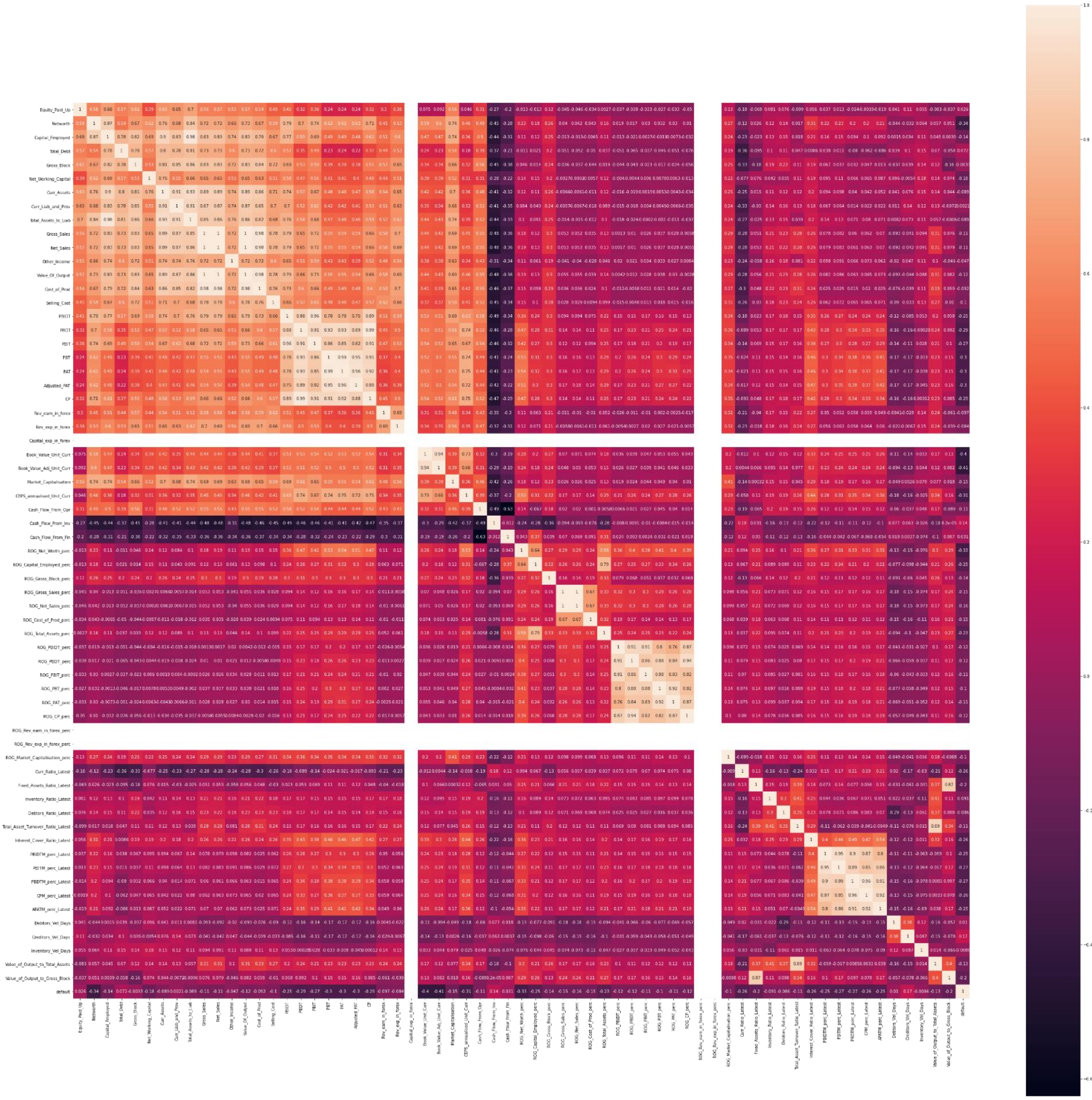
```
df["Networth_Next_Year"].value_counts()
```

```
0    3199
1     387
Name: Networth_Next_Year, dtype: int64
```

```
df["default"]=df["Networth_Next_Year"]
```

Correlation

As per the heatmap below, most of the fields do not have correlation between them. But there are fields which have moderate to high correlation between them.



Variance Inflation factor for all the variables

	variables	VIF
10	Net_Sales	1706.985568
9	Gross_Sales	977.400778
12	Value_Of_Output	691.500675
35	ROG_Gross_Sales_perc	542.389954
36	ROG_Net_Sales_perc	541.877682
16	PBDT	117.763864
8	Total_Assets_to_Liab	105.585213
21	CP	104.281389
19	PAT	76.905151
18	PBT	75.454532
2	Capital_Employed	73.231778
13	Cost_of_Prod	50.988206
6	Curr_Assets	33.213692
57	CPM_perc_Latest	32.970630
15	PBIDT	32.645791
54	PBIDTM_perc_Latest	32.636688
55	PBITM_perc_Latest	30.263099
17	PBIT	29.343091
56	PBDTM_perc_Latest	27.663456
40	ROG_PBDT_perc	25.118176
7	Curr_Liab_and_Prov	23.404073
58	APATM_perc_Latest	19.414271
42	ROG_PBT_perc	19.393456
44	ROG_CP_perc	18.353908
25	Book_Value_Unit_Curr	17.564033
20	Adjusted_PAT	17.276777
39	ROG_PBIDT_perc	15.127911
43	ROG_PAT_perc	14.744726
26	Book_Value_Adj_Unit_Curr	13.929218
62	Value_of_Output_to_Total_Assets	12.791951
1	Networth	12.771591
41	ROG_PBIT_perc	12.461225
4	Gross_Block	12.367991
52	Total_Asset_Turnover_Ratio_Latest	11.211383
63	Value_of_Output_to_Gross_Block	8.881699
49	Fixed_Assets_Ratio_Latest	8.667840
3	Total_Debt	7.336446
28	CEPS_annualised_Unit_Curr	6.703459
5	Net_Working_Capital	5.339842
0	Equity_Paid_Up	4.611477
14	Selling_Cost	4.609168
27	Market_Capitalisation	4.579631
11	Other_Income	4.506014
33	ROG_Capital_Employed_perc	4.022409
23	Rev_exp_in_forex	3.865463
29	Cash_Flow_From_Opr	3.761472
38	ROG_Total_Assets_perc	3.395773
22	Rev_earn_in_forex	3.135023
32	ROG_Net_Worth_perc	2.955247
31	Cash_Flow_From_Fin	2.916610
53	Interest_Cover_Ratio_Latest	2.551925
51	Debtors_Ratio_Latest	2.538926
59	Debtors_Vel_Days	2.485328
30	Cash_Flow_From_Inv	2.480988
48	Curr_Ratio_Latest	2.409424
50	Inventory_Ratio_Latest	2.383155
60	Creditors_Vel_Days	2.324994
37	ROG_Cost_of_Prod_perc	2.038591
61	Inventory_Vel_Days	1.965503
47	ROG_Market_Capitalisation_perc	1.672925
34	ROG_Gross_Block_perc	1.583099
24	Capital_exp_in_forex	NaN

45	ROG_Rev_earn_in_forex_perc	NaN
46	ROG_Rev_exp_in_forex_perc	NaN

- Net_sales had the highest correlation of all the factors

The following **40 variables** were removed one-by-one from the above model, based on their high VIF values –

Net_Sales	PBIDTM_perc_Latest
ROG_Gross_Sales_perc	ROG_PBDT_perc
Gross_Sales	PBITM_perc_Latest
PBDT	Capital_Employed
Capital_exp_in_forex	PBT
ROG_Rev_earn_in_forex_perc	ROG_PBT_perc
ROG_Rev_exp_in_forex_perc	ROG_CP_perc
Total_Assets_to_Liab	PBDTM_perc_Latest
PAT	Book_Value_Unit_Curr
Value_Of_Output	CP
Cost_of_Prod	ROG_PBIDT_perc
CPM_perc_Latest	Adjusted_PAT
PBIDT	ROG_PAT_perc
Curr_Assets	Book_Value_Adj_Unit_Curr
PBIT	APATM_perc_Latest
Value_of_Output_to_Total_Assets	Curr_Liab_and_Prov
ROG_PBIT_perc	Value_of_Output_to_Gross_Block
Networth	Fixed_Assets_Ratio_Latest
Total_Asset_Turnover_Ratio_Latest	Total_Debt
Gross_Block	CEPS_annualised_Unit_Curr]

Retained the following **24** variables – With VIF <5

Equity_Paid_Up	ROG_Gross_Block_perc
Net_Working_Capital	ROG_Net_Sales_perc
Other_Income	ROG_Cost_of_Prod_perc
Selling_Cost	ROG_Total_Assets_perc
Rev_earn_in_forex	ROG_Market_Capitalisation_perc
Rev_exp_in_forex	Curr_Ratio_Latest
Market_Capitalisation	Inventory_Ratio_Latest
Cash_Flow_From_Opr	Debtors_Ratio_Latest
Cash_Flow_From_Inv	Interest_Cover_Ratio_Latest
Cash_Flow_From_Fin	Debtors_Vel_Days
ROG_Net_Worth_perc	Creditors_Vel_Days
ROG_Capital_Employed_perc	Inventory_Vel_Days

Train Test Split

Split the data into Train and Test dataset in a ratio of 67:33 and use random_state =42.

```
X_train,X_test,Y_train,Y_test = train_test_split(X11,Y,test_size = 0.33, random_state = 42,stratify=Y)
```

Logistic Regression Model

On data with **24** variables selected so far –

Model 1 – Statsmodels logistic regression

```
                           Logit Regression Results
==============================================================================
Dep. Variable:                default   No. Observations:                 2401
Model:                          Logit   Df Residuals:                     2376
Method:                           MLE   Df Model:                           24
Date:                Sun, 03 Apr 2022   Pseudo R-squ.:                  0.3954
Time:                        19:24:03   Log-Likelihood:                -496.56
converged:                       True   LL-Null:                       -821.24
Covariance Type:            nonrobust   LLR p-value:                 1.077e-121
====================================================================================================
                                       coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------------
Intercept                           -0.7085      0.184     -3.859      0.000      -1.068      -0.349
Equity_Paid_Up                       0.0161      0.008      2.093      0.036       0.001       0.031
Net_Working_Capital                 -0.0074      0.002     -3.677      0.000      -0.011      -0.003
Other_Income                         0.0445      0.039      1.151      0.250      -0.031       0.120
Selling_Cost                         0.0262      0.045      0.586      0.558      -0.061       0.114
Rev_earn_in_forex                   -0.0271      0.019     -1.396      0.163      -0.065       0.011
Rev_exp_in_forex                     0.0472      0.020      2.324      0.020       0.007       0.087
Market_Capitalisation               -0.0099      0.002     -5.204      0.000      -0.014      -0.006
Cash_Flow_From_Opr                  -0.0172      0.013     -1.304      0.192      -0.043       0.009
```

```
Cash_Flow_From_Opr                  -0.0172      0.013     -1.304      0.192      -0.043       0.009
Cash_Flow_From_Inv                  -0.0010      0.026     -0.039      0.969      -0.052       0.050
Cash_Flow_From_Fin                  -0.0125      0.023     -0.549      0.583      -0.057       0.032
ROG_Net_Worth_perc                  -0.0465      0.009     -5.324      0.000      -0.064      -0.029
ROG_Capital_Employed_perc            0.0026      0.008      0.316      0.752      -0.013       0.019
ROG_Gross_Block_perc                -0.0344      0.015     -2.316      0.021      -0.063      -0.005
ROG_Net_Sales_perc                  -0.0003      0.003     -0.102      0.919      -0.007       0.006
ROG_Cost_of_Prod_perc               -0.0079      0.003     -2.463      0.014      -0.014      -0.002
ROG_Total_Assets_perc               -0.0120      0.008     -1.493      0.136      -0.028       0.004
ROG_Market_Capitalisation_perc       0.0005      0.002      0.251      0.802      -0.004       0.005
Curr_Ratio_Latest                   -0.6355      0.076     -8.344      0.000      -0.785      -0.486
Inventory_Ratio_Latest              -0.0162      0.014     -1.153      0.249      -0.044       0.011
Debtors_Ratio_Latest                -0.0231      0.015     -1.510      0.131      -0.053       0.007
Interest_Cover_Ratio_Latest         -0.2043      0.033     -6.229      0.000      -0.269      -0.140
Debtors_Vel_Days                    -0.0033      0.001     -3.141      0.002      -0.005      -0.001
Creditors_Vel_Days                   0.0034      0.001      2.966      0.003       0.001       0.006
Inventory_Vel_Days                  -0.0016      0.001     -1.290      0.197      -0.004       0.001
====================================================================================================
```

As per the above model – The insignificant variables are:

Other_Income
Selling_Cost
Rev_earn_in_forex
Cash_Flow_From_Opr
Cash_Flow_From_Inv
Cash_Flow_From_Fin
ROG_Capital_Employed_perc
ROG_Net_Sales_perc
ROG_Total_Assets_perc
ROG_Market_Capitalisation_perc

Inventory_Ratio_Latest
Debtors_Ratio_Latest
Inventory_Vel_Days

Each of these insignificant variables are removed one-by-one and **14 models** were created. The variables were removed based on their p-values if they were greater than 0.05, they were removed.

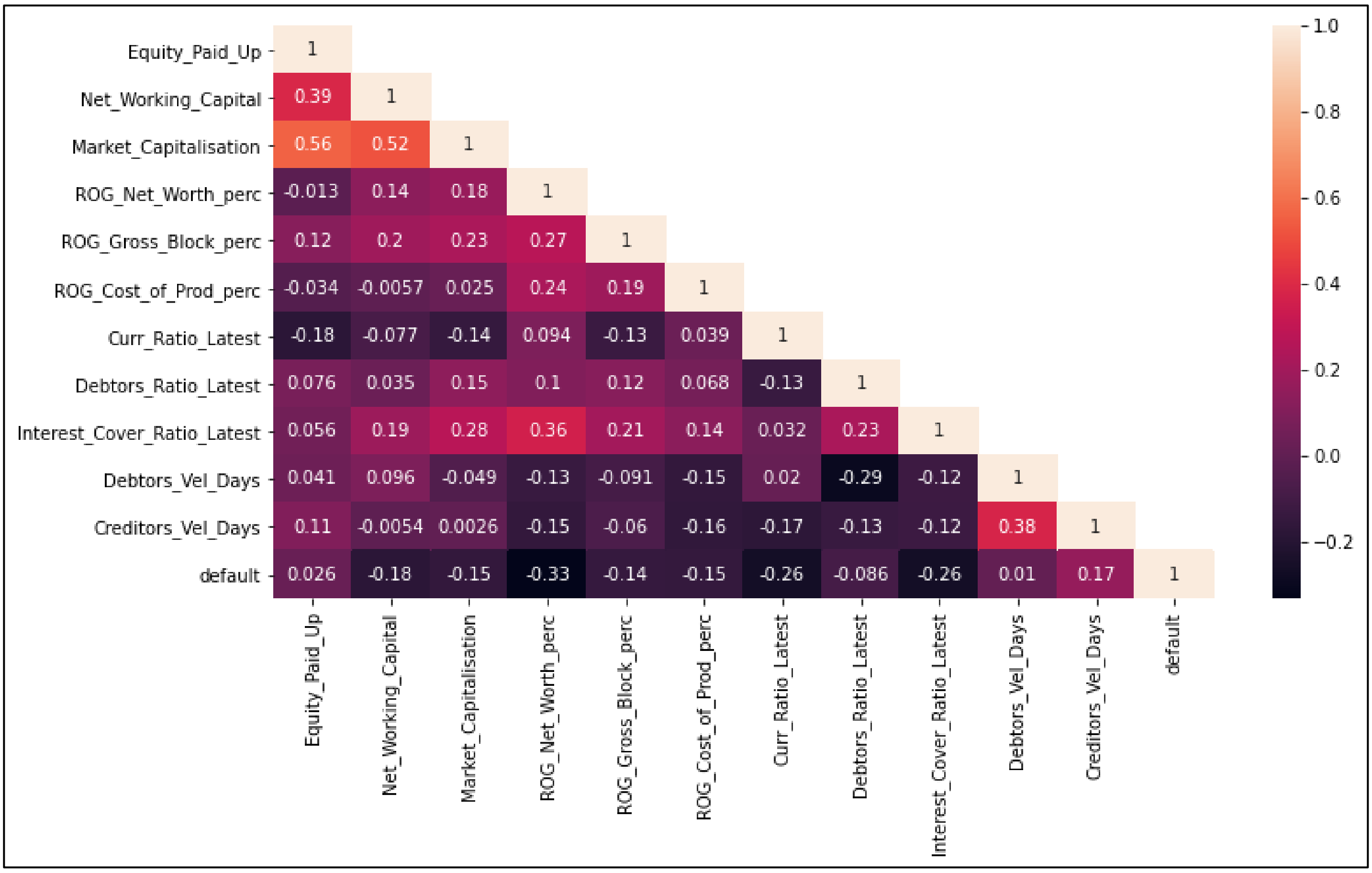
Eventually, these came out to be significant variables.

Equity_Paid_Up	ROG_Cost_of_Prod_perc
Net_Working_Capital	Curr_Ratio_Latest
Market_Capitalisation	Debtors_Ratio_Latest
ROG_Net_Worth_perc	Interest_Cover_Ratio_Latest
ROG_Gross_Block_perc	Debtors_Vel_Days
	Creditors_Vel_Days

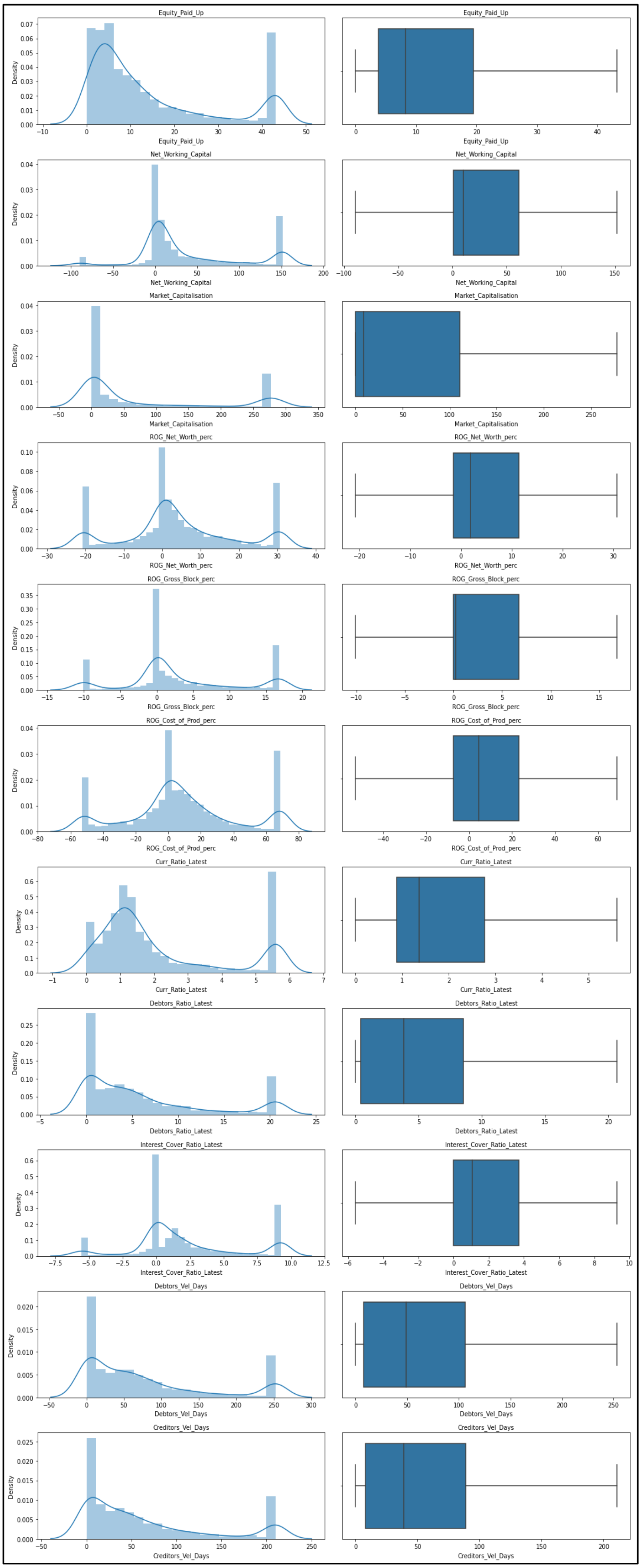
Exploratory Data Analysis

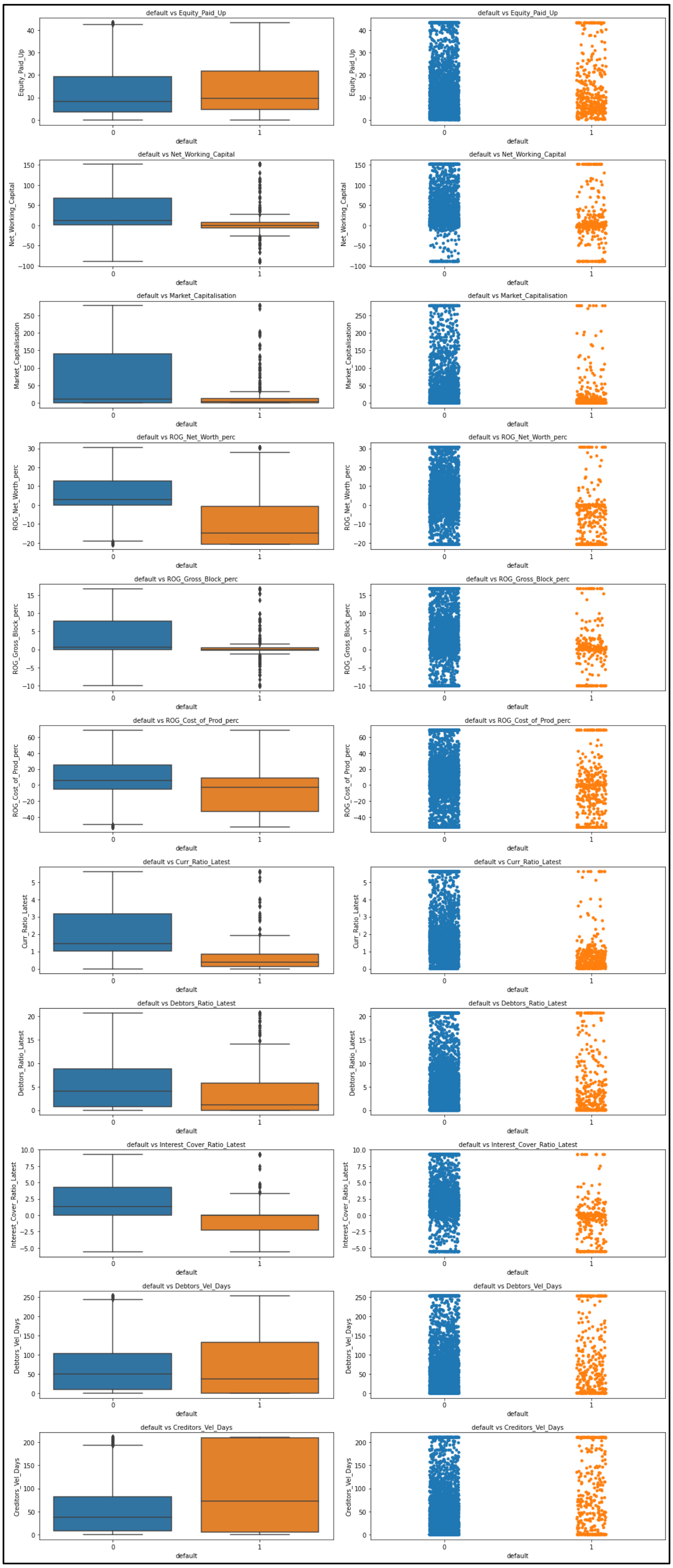
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3585 entries, 0 to 3585
Data columns (total 12 columns):
 #   Column                        Non-Null Count  Dtype  
---  ------                        --------------  -----  
 0   Equity_Paid_Up                3585 non-null   float64
 1   Net_Working_Capital           3585 non-null   float64
 2   Market_Capitalisation         3585 non-null   float64
 3   ROG_Net_Worth_perc            3585 non-null   float64
 4   ROG_Gross_Block_perc          3585 non-null   float64
 5   ROG_Cost_of_Prod_perc         3585 non-null   float64
 6   Curr_Ratio_Latest             3585 non-null   float64
 7   Debtors_Ratio_Latest          3585 non-null   float64
 8   Interest_Cover_Ratio_Latest   3585 non-null   float64
 9   Debtors_Vel_Days              3585 non-null   float64
 10  Creditors_Vel_Days            3585 non-null   float64
 11  default                       3585 non-null   int32  
dtypes: float64(11), int32(1)
memory usage: 350.1 KB
```

	count	mean	std	min	25%	50%	75%	max
Equity_Paid_Up	3585.0	13.996070	14.004971	0.000	3.75	8.29	19.52	43.175
Net_Working_Capital	3585.0	36.813095	59.350655	-89.395	0.95	10.15	61.18	151.525
Market_Capitalisation	3585.0	72.136657	106.952525	0.000	0.00	8.34	110.97	277.425
ROG_Net_Worth_perc	3585.0	4.122555	14.309276	-20.780	-1.49	1.83	11.37	30.660
ROG_Gross_Block_perc	3585.0	2.948232	7.652718	-10.080	0.00	0.25	6.72	16.800
ROG_Cost_of_Prod_perc	3585.0	7.890653	33.112475	-52.820	-7.25	4.39	23.13	68.700
Curr_Ratio_Latest	3585.0	2.084286	1.806562	0.000	0.88	1.36	2.77	5.605
Debtors_Ratio_Latest	3585.0	5.990901	6.626860	0.000	0.42	3.82	8.52	20.670
Interest_Cover_Ratio_Latest	3585.0	2.078743	3.912734	-5.565	0.00	1.08	3.71	9.275
Debtors_Vel_Days	3585.0	75.307671	81.863712	0.000	8.00	49.00	106.00	253.000
Creditors_Vel_Days	3585.0	62.456485	68.147526	0.000	8.00	39.00	89.00	210.500
default	3585.0	0.107950	0.310360	0.000	0.00	0.00	0.00	1.000



The selected 11 variables are not highly correlated with each-other


Equity_Paid_Up
Density
Equity_Paid_Up
Net_Working_Capital
Density
Net_Working_Capital
Market_Capitalisation
Density
Market_Capitalisation
ROG_Net_Worth_perc
Density
ROG_Net_Worth_perc
ROG_Gross_Block_perc
Density
ROG_Gross_Block_perc
ROG_Cost_of_Prod_perc
Density
ROG_Cost_of_Prod_perc
Curr_Ratio_Latest
Density
Curr_Ratio_Latest
Debtors_Ratio_Latest
Density
Debtors_Ratio_Latest
Interest_Cover_Ratio_Latest
Density
Interest_Cover_Ratio_Latest
Debtors_Vel_Days
Density
Debtors_Vel_Days
Creditors_Vel_Days
Density
Creditors_Vel_Days


default vs Equity_Paid_Up
Equity_Paid_Up
default
default vs Net_Working_Capital
Net_Working_Capital
default
default vs Market_Capitalisation
Market_Capitalisation
default
default vs ROG_Net_Worth_perc
ROG_Net_Worth_perc
default
default vs ROG_Gross_Block_perc
ROG_Gross_Block_perc
default
default vs ROG_Cost_of_Prod_perc
ROG_Cost_of_Prod_perc
default
default vs Curr_Ratio_Latest
Curr_Ratio_Latest
default
default vs Debtors_Ratio_Latest
Debtors_Ratio_Latest
default
default vs Interest_Cover_Ratio_Latest
Interest_Cover_Ratio_Latest
default
default vs Debtors_Vel_Days
Debtors_Vel_Days
default
default vs Creditors_Vel_Days
Creditors_Vel_Days
default

Insights:

- About 50% of defaulters have less than 0 net working capital
- About 75% of defaulters have negative rate of growth for net worth as compared to just 25% of non-defaulters having negative rate of growth for net worth
- Average number of days required for receiving the payments for non-defaulters is 75 and average number of days required for receiving the payments for non-defaulters is 77. But Average number of days non-defaulters takes to pay suppliers is 58, on the other hand average number of days defaulters take to pay supplier

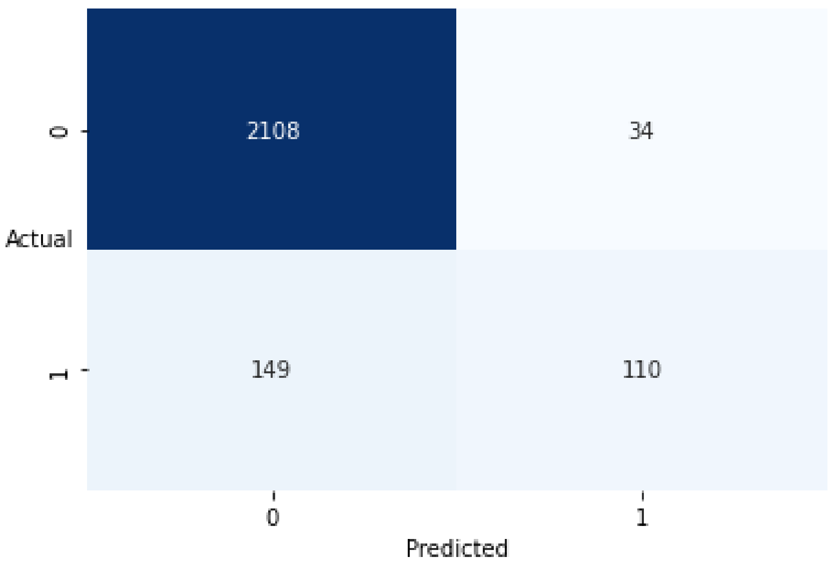
Final Logistic Regression Model

Model Final

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-0.7742	0.179	-4.335	0.000	-1.124	-0.424
Equity_Paid_Up	0.0216	0.007	3.038	0.002	0.008	0.036
Net_Working_Capital	-0.0063	0.002	-3.468	0.001	-0.010	-0.003
Market_Capitalisation	-0.0088	0.002	-5.345	0.000	-0.012	-0.006
ROG_Net_Worth_perc	-0.0515	0.007	-7.180	0.000	-0.066	-0.037
ROG_Gross_Block_perc	-0.0385	0.014	-2.766	0.006	-0.066	-0.011
ROG_Cost_of_Prod_perc	-0.0082	0.003	-3.192	0.001	-0.013	-0.003
Curr_Ratio_Latest	-0.6456	0.077	-8.435	0.000	-0.796	-0.496
Debtors_Ratio_Latest	-0.0287	0.014	-2.092	0.036	-0.055	-0.002
Interest_Cover_Ratio_Latest	-0.1950	0.032	-6.175	0.000	-0.257	-0.133
Debtors_Vel_Days	-0.0038	0.001	-3.689	0.000	-0.006	-0.002
Creditors_Vel_Days	0.0034	0.001	3.044	0.002	0.001	0.006

Model Validation

<u>Confusion Matrix Train Data (cut off = 0.5) –</u>

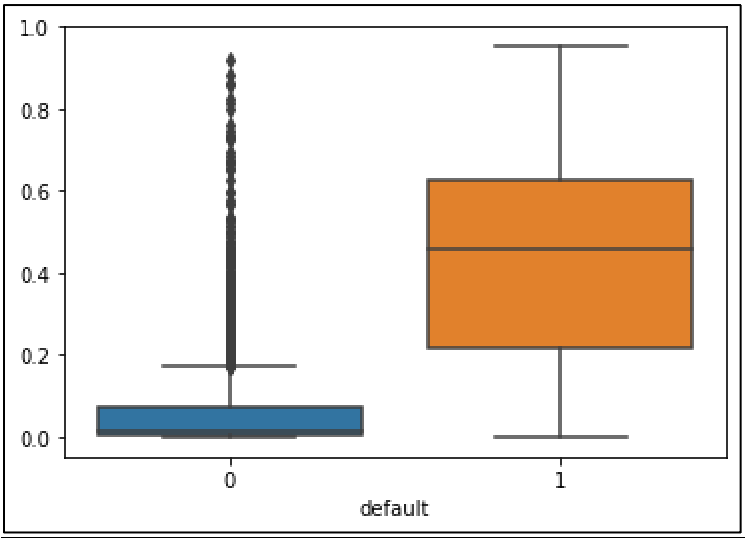


False Positive = 34
False Negative =
149

Classification Report – Train Data

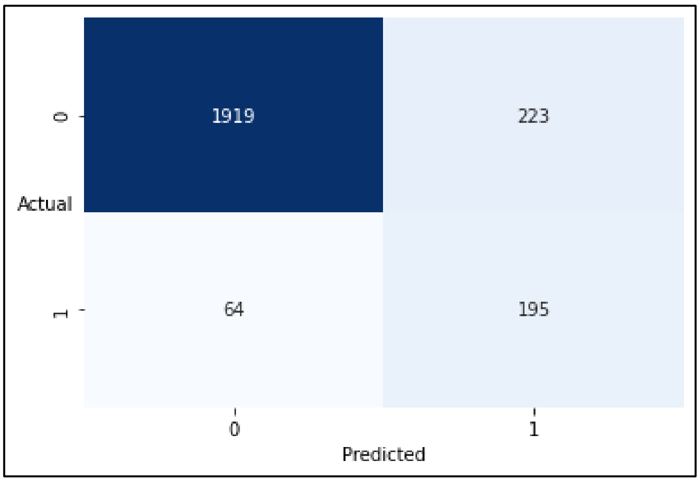
	precision	recall	f1-score	support
0	0.93	0.98	0.96	2142
1	0.76	0.42	0.55	259
accuracy			0.92	2401
macro avg	0.85	0.70	0.75	2401
weighted avg	0.92	0.92	0.91	2401

Probability distribution for different classes:



Probability cut-off was chosen as 0.21

After cut-off was changed to 0.21. The updated training confusion matrix and classification reports are as follows:

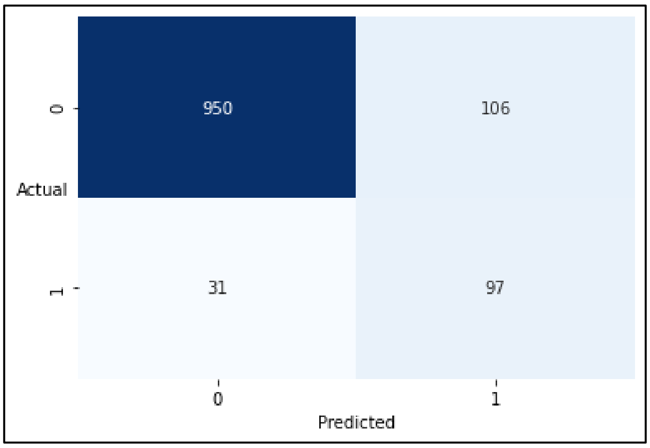


False negatives have reduced from **149** to **64.** Consequently, recall for class 1 has improved.

	precision	recall	f1-score	support
0	0.97	0.90	0.93	2142
1	0.47	0.75	0.58	259
accuracy			0.88	2401
macro avg	0.72	0.82	0.75	2401
weighted avg	0.91	0.88	0.89	2401

Test set classification report

Predictions on the test set



Test set confusion matrix

False positives are just 31.

	precision	recall	f1-score	support
0	0.97	0.90	0.93	1056
1	0.48	0.76	0.59	128
accuracy			0.88	1184
macro avg	0.72	0.83	0.76	1184
weighted avg	0.92	0.88	0.90	1184

Test set classification report

Model Interpretation

A model capable to identify companies which can default in future has been created. The class value 1 signifies that the company will default and 0 signifies that the company will not default. Since we want to identify the companies which won't default in the future, we will be focusing on the metric recall i.e., will give importance to the false negatives.

We have obtained an optimum threshold of 21% which means that any company having probability of defaulting more than 21% will be categorized as a defaulter. In other words, we are categorizing only those companies as non-defaulters which have probability of defaulting less than 21.6%.

Using this threshold value, the model thus created has an accuracy of 88%. Its precision, recall and f1-score for class 1 is 0.48, 0.76, 0.59 respectively, whereas that of class 0 is 0.97, 0.90, 0.93 respectively.

The model is neither overfit nor underfit and can be used for real data.

Business Insights

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 can handle its financial obligations, can grow quickly, and is able to manage the growth scale. We can use this model to differentiate defaulters from non-defaulters. Any investor can look up to these companies to invest his money.