

Create an Indian credit risk(default) model, using the data provided in the spreadsheet.

Dependent variable - We need to create a default variable which should take the value of 1 when net worth next year is negative & 0 when net worth next year is positive.

Validation Dataset - We need to build the model on train dataset and check the model performance measures on validation dataset.

TABLE OF CONTENTS:

1	Networth_Next_Year
2	Total_assets
3	Net_worth
4	Total_income
5	Change_in_stock
6	Total_expenses
7	Profit_after_tax
8	PBDITA
9	PBT
10	Cash_profit
11	PBDITA_as_perc_of_total_income
12	PBT_as_perc_of_total_income
13	PAT_as_perc_of_total_income
14	Cash_profit_as_perc_of_total_income
15	PAT_as_perc_of_net_worth
16	Sales
17	Income_from_fincial_services
18	Other_income
19	Total_capital
20	Reserves_and_funds
21	Borrowings
22	Current_liabilities_and_provisions
23	Deferred_tax_liability
24	Shareholders_funds
25	Cumulative_retained_profits
26	Capital_employed
27	TOL_to_TNW
28	Total_term_liabilities__to__tangible_net_worth
29	Contingent_liabilities__to__Net_worth_perc
30	Contingent_liabilities
31	Net_fixed_assets
32	Investments
33	Current_assets
34	Net_working_capital
35	Quick_ratio_times
36	Current_ratio_times
37	Debt_to_equity_ratio_times
38	Cash_to_current_liabilities_times
39	Cash_to_average_cost_of_sales_per_day
40	Creditors_turnover
41	Debtors_turnover
42	Finished_goods_turnover
43	WIP_turnover
44	Raw_material_turnover
45	Shares_outstanding
46	Equity_face_value
47	EPS
48	Adjusted_EPS
49	Total_liabilities
50	PE_on_BSE

TABLE :

Networth Next Year	Total assets	Net worth	Total income	Change in stock	Total expenses	Profit after tax	PBDITA	PBT	Cash profit	PBDITA as % of total income	PBT as % of total income	PAT as % of total income	Cash profit as % of total income	PAT as % of net worth	Sales	Income from fincial services
395.3	827.6	336.5	534.1	13.5	508.7	38.9	124.4	64.6	95.2	23.29	12.10	7.28	17.82	12.27	533.5	0.6
36.2	67.7	24.3	137.9	-3.7	131.0	3.2	5.5	1.0	3.8	3.99	0.73	2.32	2.76	0.00	135.5	NaN
84.0	238.4	78.9	331.2	-18.1	309.2	3.9	25.8	10.5	9.4	7.79	3.17	1.18	2.84	5.07	330.6	0.6
2041.4	6883.5	1443.3	8448.5	212.2	8482.4	178.3	418.4	185.1	178.0	4.95	2.19	2.11	2.11	13.17	8444.2	2.0
41.8	90.9	47.0	388.6	3.4	392.7	-0.7	7.2	-0.6	3.9	1.85	-0.15	-0.18	1.00	-1.48	387.6	0.2

SHAPE:

The number of rows (observations) is 4256

The number of columns (variables) is 51

INFO:

#	Column	Non-Null Count	Dtype
0	Num	4256 non-null	int64
1	Networth_Next_Year	4256 non-null	float64
2	Total_assets	4256 non-null	float64
3	Net_worth	4256 non-null	float64
4	Total_income	4025 non-null	float64
5	Change_in_stock	3706 non-null	float64
6	Total_expenses	4091 non-null	float64
7	Profit_after_tax	4102 non-null	float64
8	PBDITA	4102 non-null	float64
9	PBT	4102 non-null	float64
10	Cash_profit	4102 non-null	float64
11	PBDITA_as_perc_of_total_income	4177 non-null	float64
12	PBT_as_perc_of_total_income	4177 non-null	float64
13	PAT_as_perc_of_total_income	4177 non-null	float64
14	Cash_profit_as_perc_of_total_income	4177 non-null	float64
15	PAT_as_perc_of_net_worth	4256 non-null	float64
16	Sales	3951 non-null	float64
17	Income_from_fincial_services	3145 non-null	float64
18	Other_income	2700 non-null	float64
19	Total_capital	4251 non-null	float64
20	Reserves_and_funds	4158 non-null	float64
21	Borrowings	3825 non-null	float64
22	Current_liabilities_and_provisions	4146 non-null	float64
23	Deferred_tax_liability	2887 non-null	float64
24	Shareholders_funds	4256 non-null	float64
25	Cumulative_retained_profits	4211 non-null	float64
26	Capital_employed	4256 non-null	float64
27	TOL_to_TNW	4256 non-null	float64
28	Total_term_liabilities_to_tangible_net_worth	4256 non-null	float64
29	Contingent_liabilities_to_Net_worth_perc	4256 non-null	float64
30	Contingent_liabilities	2854 non-null	float64
31	Net_fixed_assets	4124 non-null	float64
32	Investments	2541 non-null	float64
33	Current_assets	4176 non-null	float64
34	Net_working_capital	4219 non-null	float64
35	Quick_ratio_times	4151 non-null	float64
36	Current_ratio_times	4151 non-null	float64
37	Debt_to_equity_ratio_times	4256 non-null	float64

38	Cash_to_current_liabilities_times	4151	non-null	float64
39	Cash_to_average_cost_of_sales_per_day	4156	non-null	float64
40	Creditors_turnover	3865	non-null	float64
41	Debtors_turnover	3871	non-null	float64
42	Finished_goods_turnover	3382	non-null	float64
43	WIP_turnover	3492	non-null	float64
44	Raw_material_turnover	3828	non-null	float64
45	Shares_outstanding	3446	non-null	float64
46	Equity_face_value	3446	non-null	float64
47	EPS	4256	non-null	float64
48	Adjusted_EPS	4256	non-null	float64
49	Total_liabilities	4256	non-null	float64
50	PE_on_BSE	1629	non-null	float64

DUPLICATES:

0

CREATING DEFAULT VARIABLE USING NETWORTH NEXT YEAR COLUMN:

	default	Networth_Next_Year
0	0	395.3
1	0	36.2
2	0	84.0
3	0	2041.4
4	0	41.8

VALUE COUNTS OF VARIABLE DEFAULT :

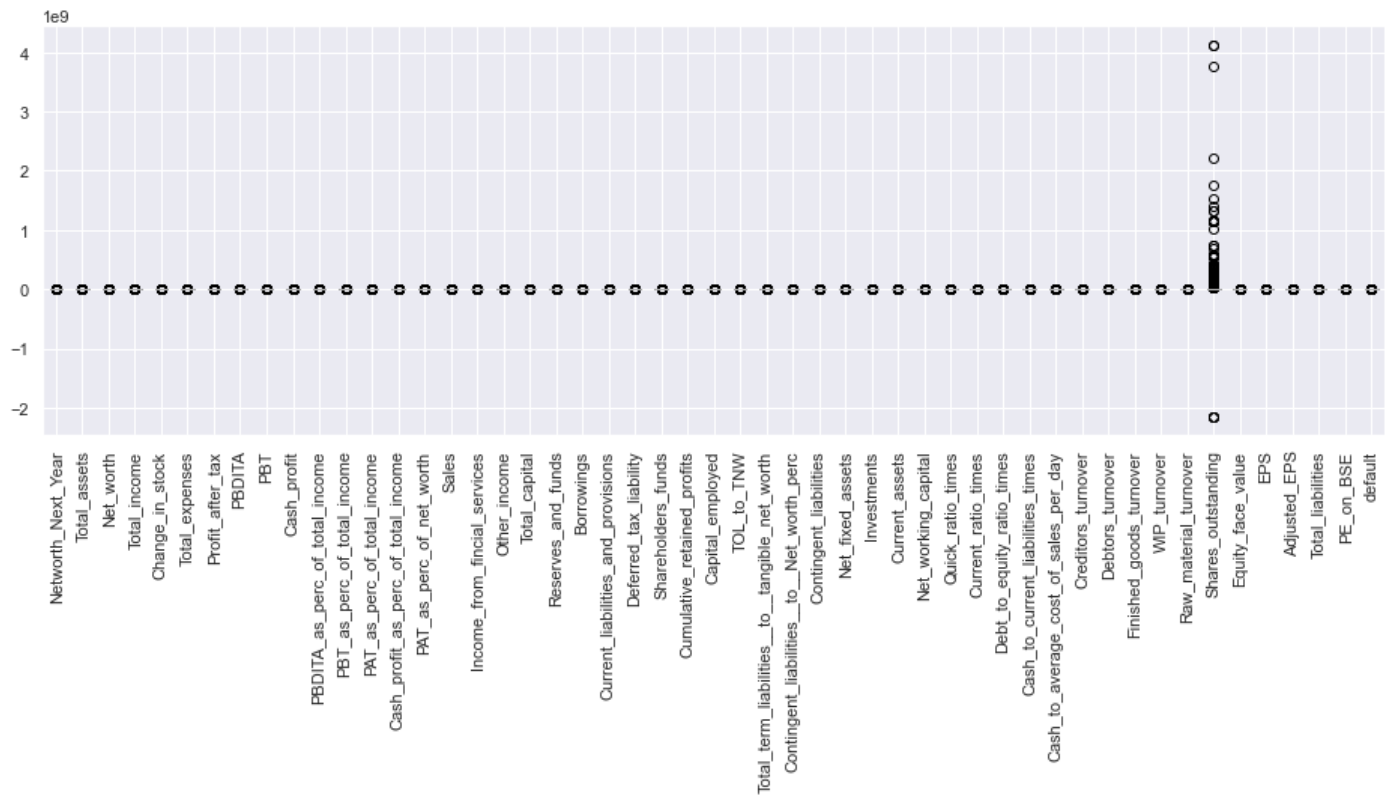
0	3352
1	904

NULL VALUES:

Networth_Next_Year	0
Total_assets	0
Net_worth	0
Total_income	231
Change_in_stock	550
Total_expenses	165
Profit_after_tax	154
PBDITA	154
PBT	154
Cash_profit	154
PBDITA_as_perc_of_total_income	79
PBT_as_perc_of_total_income	79
PAT_as_perc_of_total_income	79
Cash_profit_as_perc_of_total_income	79
PAT_as_perc_of_net_worth	0
Sales	305
Income_from_fincial_services	1111
Other_income	1556
Total_capital	5

Reserves_and_funds	98
Borrowings	431
Current_liabilities_and_provisions	110
Deferred_tax_liability	1369
Shareholders_funds	0
Cumulative_retained_profits	45
Capital_employed	0
TOL_to_TNW	0
Total_term_liabilities_to_tangible_net_worth	0
Contingent_liabilities_to_Net_worth_perc	0
Contingent_liabilities	1402
Net_fixed_assets	132
Investments	1715
Current_assets	80
Net_working_capital	37
Quick_ratio_times	105
Current_ratio_times	105
Debt_to_equity_ratio_times	0
Cash_to_current_liabilities_times	105
Cash_to_average_cost_of_sales_per_day	100
Creditors_turnover	391
Debtors_turnover	385
Finished_goods_turnover	874
WIP_turnover	764
Raw_material_turnover	428
Shares_outstanding	810
Equity_face_value	810
EPS	0
Adjusted_EPS	0
Total_liabilities	0
PE_on_BSE	2627
default	0

OUTLIERS



DESCRIPTION:

	count	mean	std	min	25%	50%	75%	max
Networth_Next_Year	425 6.0	1.344741 e+03	1.593674 e+04	- 7.426560 e+04	3.975	72.100	3.308250 e+02	8.057734 e+05
Total_assets	425 6.0	3.573617 e+03	3.007444 e+04	1.000000 e-01	91.300	315.500	1.120800 e+03	1.176509 e+06
Net_worth	425 6.0	1.351950 e+03	1.296131 e+04	0.000000 e+00	31.475	104.800	3.898500 e+02	6.131516 e+05
Total_income	402 5.0	4.688190 e+03	5.391895 e+04	0.000000 e+00	107.100	455.100	1.485000 e+03	2.442828 e+06
Change_in_stock	370 6.0	4.370248 e+01	4.369150 e+02	- 3.029400 e+03	-1.800	1.600	1.840000 e+01	1.418550 e+04
Total_expenses	409 1.0	4.356301 e+03	5.139809 e+04	- 1.000000 e-01	96.800	426.800	1.395700 e+03	2.366035 e+06
Profit_after_tax	410 2.0	2.950506 e+02	3.079902 e+03	- 3.908300 e+03	0.500	9.000	5.330000 e+01	1.194391 e+05
PBDITA	410 2.0	6.059406 e+02	5.646231 e+03	- 4.407000 e+02	6.925	36.900	1.587000 e+02	2.085765 e+05
PBT	410 2.0	4.102590 e+02	4.217415 e+03	- 3.894800 e+03	0.800	12.600	7.417500 e+01	1.452926 e+05
Cash_profit	410 2.0	4.082675 e+02	4.143926 e+03	- 2.245700 e+03	2.900	19.400	9.625000 e+01	1.769118 e+05
PBDITA_as_perc_of_total_income	417 7.0	3.179892 e+00	1.722566 e+02	- 6.400000 e+03	4.970	9.680	1.647000 e+01	1.000000 e+02
PBT_as_perc_of_total_income	417 7.0	- 1.819683 e+01	4.199111 e+02	- 2.134000 e+04	0.560	3.340	8.940000 e+00	1.000000 e+02

	count	mean	std	min	25%	50%	75%	max
PAT_as_perc_of_total_income	4177.0	-2.003367e+01	4.235762e+02	-2.134000e+04	0.350	2.370	6.420000e+00	1.500000e+02
Cash_profit_as_perc_of_total_income	4177.0	-9.021278e+00	2.999574e+02	-1.502000e+04	2.000	5.660	1.073000e+01	1.000000e+02
PAT_as_perc_of_net_worth	4256.0	1.016786e+01	6.153240e+01	-7.487200e+02	0.000	8.040	2.020250e+01	2.466670e+03
Sales	3951.0	4.645685e+03	5.308090e+04	1.000000e-01	113.350	468.600	1.481200e+03	2.384984e+06
Income_from_fincial_services	3145.0	8.136006e+01	1.042759e+03	0.000000e+00	0.500	1.900	9.800000e+00	5.193820e+04
Other_income	2700.0	5.595289e+01	1.178415e+03	0.000000e+00	0.400	1.500	6.200000e+00	4.285670e+04
Total_capital	4251.0	2.245577e+02	1.684951e+03	1.000000e-01	13.200	42.600	1.031500e+02	7.827320e+04
Reserves_and_funds	4158.0	1.210562e+03	1.281623e+04	-6.525900e+03	5.300	55.150	2.825250e+02	6.251378e+05
Borrowings	3825.0	1.176248e+03	8.581249e+03	1.000000e-01	24.400	99.800	3.583000e+02	2.782573e+05
Current_liabilities_and_provisions	4146.0	9.606314e+02	9.140536e+03	1.000000e-01	17.500	70.300	2.659250e+02	3.522403e+05
Deferred_tax_liability	2887.0	2.344951e+02	2.106253e+03	1.000000e-01	3.200	13.500	5.130000e+01	7.279660e+04
Shareholders_funds	4256.0	1.376487e+03	1.301069e+04	0.000000e+00	32.300	107.600	4.089000e+02	6.131516e+05
Cumulative_retained_profits	4211.0	9.371820e+02	9.853096e+03	-6.534300e+03	1.100	37.400	2.062000e+02	3.901338e+05
Capital_employed	4256.0	2.433618e+03	2.049640e+04	0.000000e+00	61.300	221.200	7.903000e+02	8.914089e+05

	count	mean	std	min	25%	50%	75%	max
TOL_to_TNW	425 6.0	4.025343 e+00	2.087909 e+01	- 3.504800 e+02	0.600	1.420	2.830000 e+00	4.730000 e+02
Total_term_liabilities_to_tangible_net_worth	425 6.0	1.854288 e+00	1.587506 e+01	- 3.256000 e+02	0.050	0.345	1.000000 e+00	4.560000 e+02
Contingent_liabilities_to_Net_worth_perc	425 6.0	5.570750 e+01	3.691657 e+02	0.000000 e+00	0.000	5.360	3.101250 e+01	1.470427 e+04
Contingent_liabilities	285 4.0	9.485522 e+02	1.205674 e+04	1.000000 e-01	6.000	37.850	1.953250 e+02	5.595068 e+05
Net_fixed_assets	412 4.0	1.209487 e+03	1.250240 e+04	0.000000 e+00	26.200	93.850	3.528250 e+02	6.366046 e+05
Investments	254 1.0	7.218659 e+02	6.793860 e+03	0.000000 e+00	1.000	8.200	6.380000 e+01	1.999786 e+05
Current_assets	417 6.0	1.350360 e+03	1.015557 e+04	1.000000 e-01	36.600	148.350	5.150000 e+02	3.548152 e+05
Net_working_capital	421 9.0	1.628742 e+02	3.182030 e+03	- 6.383900 e+04	-1.100	16.700	8.650000 e+01	8.578280 e+04
Quick_ratio_times	415 1.0	1.497355 e+00	9.327519 e+00	0.000000 e+00	0.410	0.670	1.030000 e+00	3.410000 e+02
Current_ratio_times	415 1.0	2.257398 e+00	1.247829 e+01	0.000000 e+00	0.930	1.230	1.720000 e+00	5.050000 e+02
Debt_to_equity_ratio_times	425 6.0	2.871563 e+00	1.559997 e+01	0.000000 e+00	0.220	0.790	1.750000 e+00	4.560000 e+02
Cash_to_current_liabilities_times	415 1.0	5.284197 e-01	4.796342 e+00	0.000000 e+00	0.020	0.070	1.900000 e-01	1.650000 e+02
Cash_to_average_cost_of_sales_per_day	415 6.0	1.451579 e+02	2.521992 e+03	0.000000 e+00	2.880	8.040	2.197000 e+01	1.280408 e+05
Creditors_turnover	386 5.0	1.681226 e+01	7.567492 e+01	0.000000 e+00	3.720	6.170	1.169000 e+01	2.401000 e+03
Debtors_turnover	387 1.0	1.792903 e+01	9.016443 e+01	0.000000 e+00	3.810	6.470	1.185000 e+01	3.135200 e+03

	count	mean	std	min	25%	50%	75%	max
Finished_goods_turnover	338 2.0	8.436999 e+01	5.626374 e+02	9.000000 e-02	8.190	17.320	4.001250 e+01	1.794760 e+04
WIP_turnover	349 2.0	2.868451 e+01	1.696509 e+02	1.800000 e-01	5.100	9.860	2.024000 e+01	5.651400 e+03
Raw_material_turnover	382 8.0	1.773393 e+01	3.431259 e+02	2.000000 e+00	3.020	6.410	1.182250 e+01	2.109200 e+04
Shares_outstanding	344 6.0	2.376491 e+07	1.709790 e+08	2.147484 e+09	1308382 .500	4750000 .000	1.090602 e+07	4.130401 e+09
Equity_face_value	344 6.0	1.094829 e+03	3.410136 e+04	9.999989 e+05	10.000	10.000	1.000000 e+01	1.000000 e+05
EPS	425 6.0	1.962175 e+02	1.306195 e+04	8.431818 e+05	0.000	1.490	1.000000 e+01	3.452253 e+04
Adjusted_EPS	425 6.0	1.975276 e+02	1.306193 e+04	8.431818 e+05	0.000	1.240	7.615000 e+00	3.452253 e+04
Total_liabilities	425 6.0	3.573617 e+03	3.007444 e+04	1.000000 e-01	91.300	315.500	1.120800 e+03	1.176509 e+06
PE_on_BSE	162 9.0	5.546229 e+01	1.304445 e+03	1.116640 e+03	2.970	8.690	1.700000 e+01	5.100274 e+04
default	425 6.0	2.124060 e-01	4.090587 e-01	0.000000 e+00	0.000	0.000	0.000000 e+00	1.000000 e+00

Removing COLUMNS WHICH ARE MISSING MORE THEN 30% OF DATA:

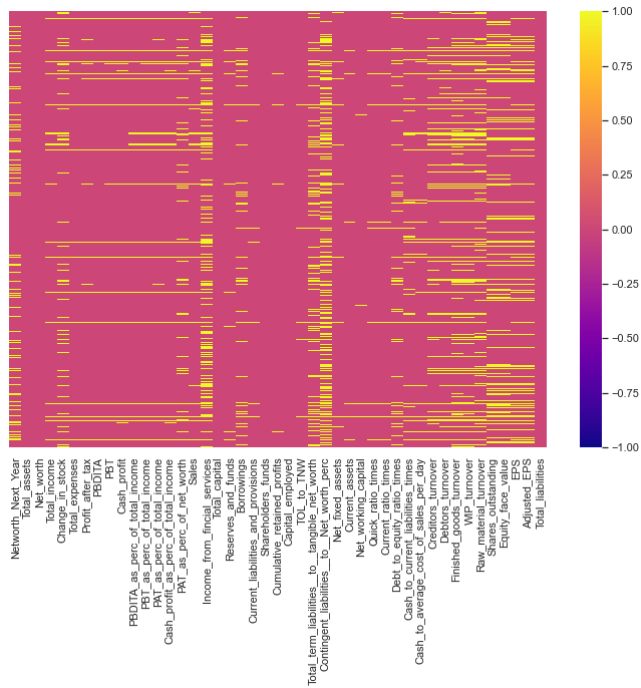
PE_on_BSE	0.617246
Investments	0.402961
Other_income	0.365602
Contingent_liabilities	0.329417
Deferred_tax_liability	0.321664
Income_from_fincial_services	0.261043
Finished_goods_turnover	0.205357
Shares_outstanding	0.190320
Equity_face_value	0.190320
WIP_turnover	0.179511

Change_in_stock	0.129229
Borrowings	0.101269
Raw_material_turnover	0.100564
Creditors_turnover	0.091870
Debtors_turnover	0.090461
Sales	0.071664
Total_income	0.054276
Total_expenses	0.038769
Cash_profit	0.036184
Profit_after_tax	0.036184
PBT	0.036184
PBDITA	0.036184
Net_fixed_assets	0.031015
Current_liabilities_and_provisions	0.025846
Cash_to_current_liabilities_times	0.024671
Current_ratio_times	0.024671
Quick_ratio_times	0.024671
Cash_to_average_cost_of_sales_per_day	0.023496
Reserves_and_funds	0.023026
Current_assets	0.018797
PAT_as_perc_of_total_income	0.018562
PBT_as_perc_of_total_income	0.018562
PBDITA_as_perc_of_total_income	0.018562
Cash_profit_as_perc_of_total_income	0.018562
Cumulative_retained_profits	0.010573
Net_working_capital	0.008694
Total_capital	0.001175
EPS	0.000000
Total_liabilities	0.000000
Adjusted_EPS	0.000000
Networth_Next_Year	0.000000
Debt_to_equity_ratio_times	0.000000
Contingent_liabilities_to_Net_worth_perc	0.000000
Total_term_liabilities_to_tangible_net_worth	0.000000
TOL_to_TNW	0.000000
Total_assets	0.000000
Shareholders_funds	0.000000
PAT_as_perc_of_net_worth	0.000000
Net_worth	0.000000
Capital_employed	0.000000

COLUMNS DROPPED:

["PE_on_BSE","Investments","Other_income","Contingent_liabilities","Deferred_tax_liability"]

HEATMAP OF NULL VALUES:



Here, we can see lot's of rows have null values in them as we can see by the yellow dots in the figure.

SCALED DATA:

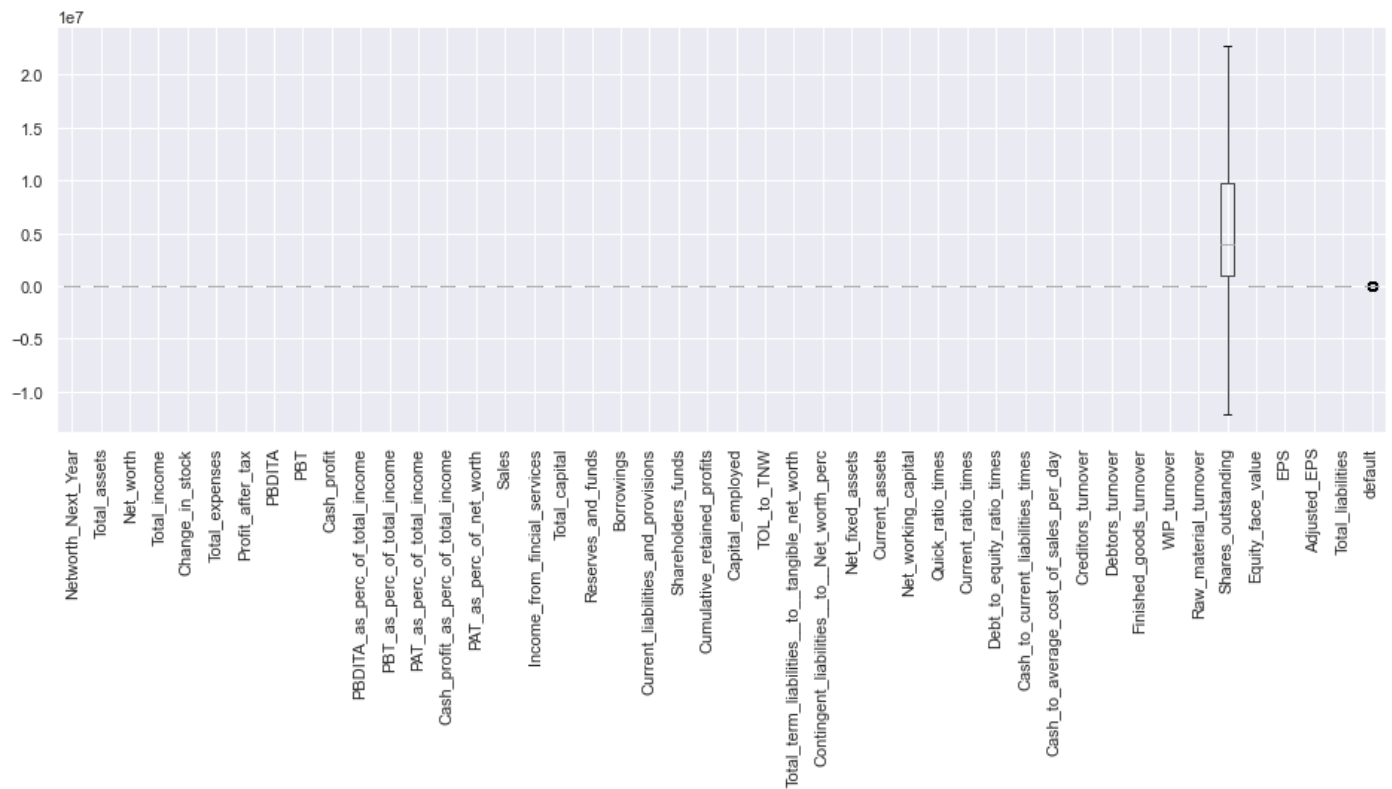
Network_Next_Year	Total_assets	Net_worth	Total_income	Change_in_stock	Total_expenses	Profit_after_tax	PBDITA	PBT	Cash_profit
-0.069211	-0.091318	-0.078369	-0.077077	-0.070705	-0.074926	-0.084231	-0.086259	-0.083062	-0.076597
-0.089911	-0.116588	-0.102456	-0.084424	-0.109617	-0.082271	-0.095714	-0.107167	-0.098002	-0.098472
-0.087155	-0.110912	-0.098244	-0.080840	-0.142194	-0.078806	-0.095489	-0.103597	-0.095770	-0.097132
0.025675	0.110069	0.007023	0.069688	0.378820	0.080135	-0.039393	-0.034560	-0.054757	-0.056781
-0.089588	-0.115817	-0.100705	-0.079775	-0.093554	-0.077182	-0.096968	-0.106868	-0.098378	-0.098448

IMPUTED VALUES BY USING KNNImputer:

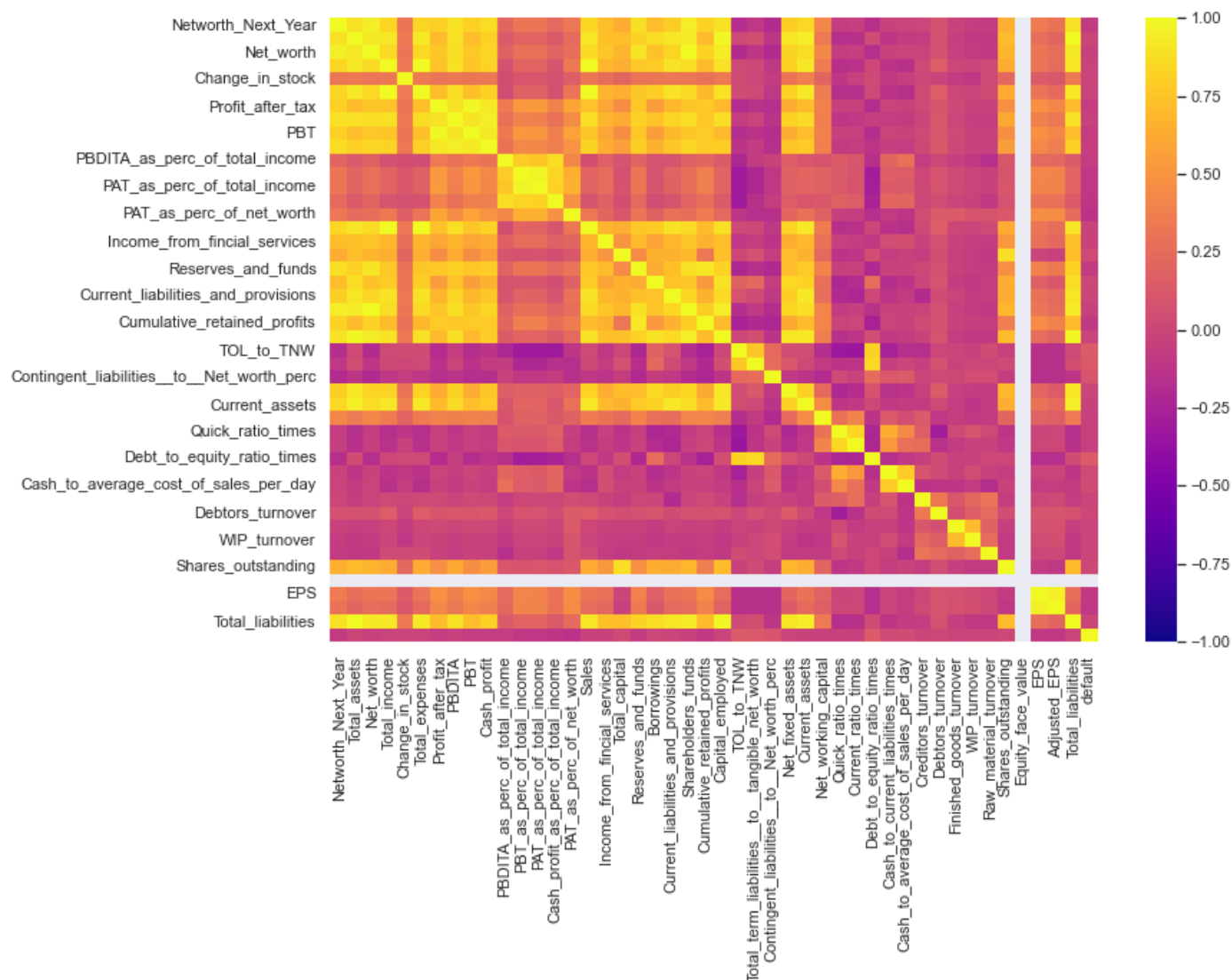
Network_Next_Year	0
Total_assets	0
Net_worth	0
Total_income	0
Change_in_stock	0
Total_expenses	0
Profit_after_tax	0
PBDITA	0
PBT	0
Cash_profit	0
PBDITA_as_perc_of_total_income	0
PBT_as_perc_of_total_income	0
PAT_as_perc_of_total_income	0
Cash_profit_as_perc_of_total_income	0
PAT_as_perc_of_net_worth	0
Sales	0
Income_from_fincial_services	0
Total_capital	0
Reserves_and_funds	0
Borrowings	0
Current_liabilities_and_provisions	0
Shareholders_funds	0
Cumulative_retained_profits	0
Capital_employed	0
TOL_to_TNW	0

Total_term_liabilities_to_tangible_net_worth	0
Contingent_liabilities_to_Net_worth_perc	0
Net_fixed_assets	0
Current_assets	0
Net_working_capital	0
Quick_ratio_times	0
Current_ratio_times	0
Debt_to_equity_ratio_times	0
Cash_to_current_liabilities_times	0
Cash_to_average_cost_of_sales_per_day	0
Creditors_turnover	0
Debtors_turnover	0
Finished_goods_turnover	0
WIP_turnover	0
Raw_material_turnover	0
Shares_outstanding	0
Equity_face_value	0
EPS	0
Adjusted_EPS	0
Total_liabilities	0
default	0

TREATED OUTLIERS WITH THE HELP OF IQR METHOD:



HEATMAP TO CHECK COLLINEARITY IN OUR DATA:



Here, we can see that lots of variables have collinearity in them so we would be using VIF method to find those significant variables which are of use to us

Using STATSMODEL TO CHOOSE THE RIGHT VARIABLES:

```
Optimization terminated successfully.
Current function value: 0.491731
Iterations 16
```

Logit Regression Results								
Dep. Variable:		default	No. Observations:		4256			
Model:		Logit	Df Residuals:		4214			
Method:		MLE	Df Model:		41			
Date:		Fri, 20 Jan 2023	Pseudo R-squ.:		0.04911			
Time:		18:25:01	Log-Likelihood:		-2092.8			
converged:		True	LL-Null:		-2200.9			
Covariance Type:		nonrobust	LLR p-value:		1.174e-25			
				coef	std err	z	P> z	
							[0.025	0.975]

Intercept	-1.5818	0.201	-7.853	0.000	-1.977	-1.187
Total_assets	0.0003	nan	nan	nan	nan	nan
Net_worth	0.0003	0.001	0.371	0.711	-0.001	0.002
Total_income	0.0001	0.001	0.209	0.834	-0.001	0.001
Change_in_stock	-0.0002	0.003	-0.096	0.923	-0.005	0.005
Total_expenses	0.0008	0.001	1.422	0.155	-0.000	0.002
Profit_after_tax	0.0036	0.005	0.674	0.500	-0.007	0.014
PBDITA	-0.0005	0.001	-0.349	0.727	-0.003	0.002
PBT	0.0018	0.004	0.452	0.651	-0.006	0.010
Cash_profit	-0.0012	0.002	-0.616	0.538	-0.005	0.003
PBDITA_as_perc_of_total_income	0.0098	0.009	1.082	0.279	-0.008	0.027
PBT_as_perc_of_total_income	-0.0076	0.023	-0.328	0.743	-0.053	0.038
PAT_as_perc_of_total_income	0.0252	0.032	0.801	0.423	-0.037	0.087
Cash_profit_as_perc_of_total_income	-0.0269	0.013	-1.995	0.046	-0.053	-0.000
PAT_as_perc_of_net_worth	-0.0160	0.003	-4.680	0.000	-0.023	-0.009
Sales	-0.0008	0.001	-1.096	0.273	-0.002	0.001
Total_capital	0.0004	0.001	0.433	0.665	-0.001	0.002
Reserves_and_funds	-0.0009	0.000	-1.857	0.063	-0.002	4.84e-05
Borrowings	-0.0006	0.000	-1.414	0.157	-0.001	0.000
Current_liabilities_and_provisions	-0.0002	0.001	-0.227	0.821	-0.002	0.001
Shareholders_funds	0.0007	0.001	0.823	0.411	-0.001	0.002
Cumulative_retained_profits	-0.0003	0.001	-0.495	0.621	-0.001	0.001

Capital_employed	-0.0005	0.000	-1.020	0.308	-0.001	0.000
TOL_to_TNW	0.0598	0.023	2.594	0.009	0.015	0.105
Total_term_liabilities_to_tangible_net_worth	-0.0235	0.091	-0.258	0.796	-0.202	0.155
Contingent_liabilities_to_Net_worth_perc	0.0015	0.001	2.078	0.038	8.51e-05	0.003
Net_fixed_assets	0.0002	0.000	0.474	0.635	-0.001	0.001
Current_assets	-0.0004	0.000	-0.885	0.376	-0.001	0.000
Net_working_capital	0.0005	0.001	0.839	0.401	-0.001	0.002
Quick_ratio_times	-0.0499	0.158	-0.316	0.752	-0.359	0.259
Current_ratio_times	-0.1026	0.111	-0.922	0.356	-0.321	0.115
Debt_to_equity_ratio_times	0.1750	nan	nan	nan	nan	nan
Cash_to_current_liabilities_times	0.5108	0.466	1.095	0.273	-0.403	1.425
Cash_to_average_cost_of_sales_per_day	0.0014	0.003	0.474	0.635	-0.004	0.007
Creditors_turnover	0.0069	0.005	1.487	0.137	-0.002	0.016
Debtors_turnover	0.0033	0.009	0.363	0.717	-0.015	0.021
Finished_goods_turnover	0.0001	0.002	0.070	0.944	-0.004	0.004
WIP_turnover	-0.0012	0.004	-0.278	0.781	-0.009	0.007
Raw_material_turnover	-0.0033	0.006	-0.546	0.585	-0.015	0.009
EPS	0.0271	0.010	2.582	0.010	0.007	0.048
Adjusted_EPS	-0.0379	0.012	-3.047	0.002	-0.062	-0.014
Total_liabilities	-3.606e-05	nan	nan	nan	nan	nan

PERFORMINF VIF :

	variables	VIF
44	Total_liabilities	inf
1	Total_assets	inf
3	Total_income	284.013129
15	Sales	266.758949
5	Total_expenses	194.518168
21	Shareholders_funds	123.102097
2	Net_worth	117.923232
23	Capital_employed	83.532451
8	PBT	54.425917
6	Profit_after_tax	51.723265
7	PBDITA	30.536290
11	PBT_as_perc_of_total_income	25.589023
12	PAT_as_perc_of_total_income	24.275834
41	Equity_face_value	23.910991
9	Cash_profit	23.053122
28	Current_assets	19.219593
20	Current_liabilities_and_provisions	15.502869
0	Networth_Next_Year	14.034693
18	Reserves_and_funds	13.119989
42	EPS	12.518374
43	Adjusted_EPS	11.747289

	variables	VIF
22	Cumulative_retained_profits	8.843055
27	Net_fixed_assets	8.542348
19	Borrowings	8.295477
13	Cash_profit_as_perc_of_total_income	7.650875
17	Total_capital	6.264607
10	PBDITA_as_perc_of_total_income	6.146431
30	Quick_ratio_times	5.870699
32	Debt_to_equity_ratio_times	5.794109
40	Shares_outstanding	5.363271
24	TOL_to_TNW	4.774305
31	Current_ratio_times	4.758383
33	Cash_to_current_liabilities_times	4.500523
25	Total_term_liabilities__to__tangible_net_worth	4.465485
34	Cash_to_average_cost_of_sales_per_day	3.516572
14	PAT_as_perc_of_net_worth	2.727997
16	Income_from_fincial_services	2.582098
38	WIP_turnover	2.329891
37	Finished_goods_turnover	2.087666
29	Net_working_capital	2.071037
36	Debtors_turnover	1.806258
35	Creditors_turnover	1.651879

	variables	VIF
39	Raw_material_turnover	1.331097
4	Change_in_stock	1.297725
26	Contingent_liabilities__to__Net_worth_perc	1.207277

ELIMINATING SOME VARIABLES THROUGH VIF WHICH HAVE VALUE OF < 5:

	variables	VIF
9	Quick_ratio_times	4.641107
4	Borrowings	4.503532
12	Debtors_turnover	3.601846
11	Creditors_turnover	3.586279
3	Income_from_fincial_services	3.456780
5	Cumulative_retained_profits	3.434983
15	Shares_outstanding	3.425728
14	Raw_material_turnover	3.231177
1	Cash_profit_as_perc_of_total_income	2.605220
10	Cash_to_average_cost_of_sales_per_day	2.600104
6	Total_term_liabilities__to__tangible_net_worth	2.528051
2	PAT_as_perc_of_net_worth	2.431624
13	Finished_goods_turnover	2.171654
16	Adjusted_EPS	2.028770
7	Contingent_liabilities__to__Net_worth_perc	2.005133
8	Net_working_capital	1.823069

	variables	VIF
0	Change_in_stock	1.331058

DIVIDED THE DATA INTO 70:30 Ratio

Now Performing LOGISTIC REGRESSION through logit function to see which variables are of no use to us and whose value is <0.05 will be eliminated one by one, so we get 7 variables that are:

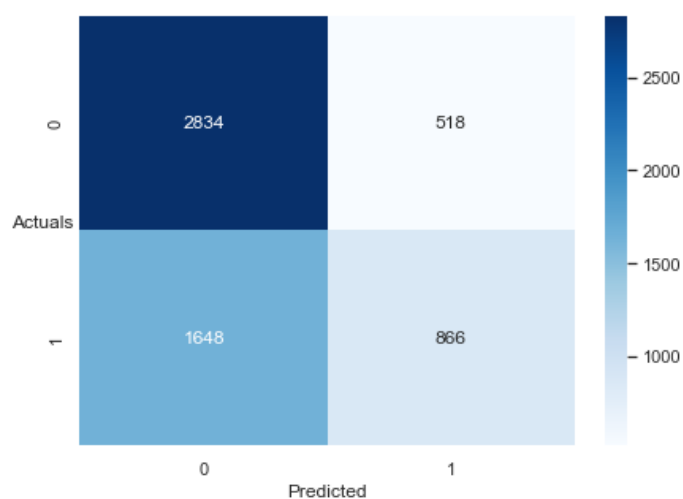
Logit Regression Results							
Dep. Variable:	default	No. Observations:	5866				
Model:	Logit	Df Residuals:	5858				
Method:	MLE	Df Model:	7				
Date:	Fri, 20 Jan 2023	Pseudo R-squ.:	0.04472				
Time:	20:48:44	Log-Likelihood:	-3826.8				
converged:	True	LL-Null:	-4005.9				
Covariance Type:	nonrobust	LLR p-value:	2.037e-73				
		coef	std err	z	P> z	[0.025	0.975]
	Intercept	-0.5820	0.063	-9.171	0.000	-0.706	-0.458
	Cash_profit_as_perc_of_total_income	-0.0185	0.004	-4.642	0.000	-0.026	-0.011
	PAT_as_perc_of_net_worth	-0.0093	0.002	-4.713	0.000	-0.013	-0.005
	Total_term_liabilities_to_tangible_net_worth	0.2823	0.037	7.722	0.000	0.211	0.354
	Contingent_liabilities_to_Net_worth_perc	0.0025	0.001	4.720	0.000	0.001	0.003
	Cash_to_average_cost_of_sales_per_day	0.0036	0.001	2.837	0.005	0.001	0.006
	Shares_outstanding	1.387e-08	3.55e-09	3.904	0.000	6.91e-09	2.08e-08
	Adjusted_EPS	-0.0087	0.003	-2.667	0.008	-0.015	-0.002

Performed prediction :

TRAIN:

	precision	recall	f1-score	support
0.0	0.63	0.85	0.72	3352
1.0	0.63	0.34	0.44	2514

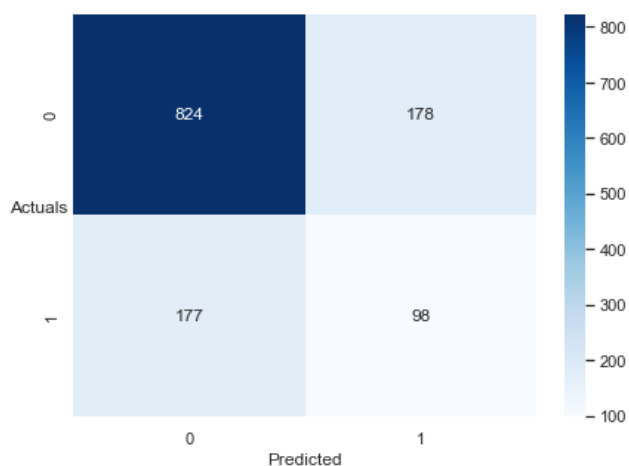
accuracy			0.63	5866
macro avg	0.63	0.59	0.58	5866
weighted avg	0.63	0.63	0.60	5866



TEST:

	precision	recall	f1-score	support
0.0	0.82	0.82	0.82	1002
1.0	0.36	0.36	0.36	275

accuracy			0.72	1277
macro avg	0.59	0.59	0.59	1277
weighted avg	0.72	0.72	0.72	1277



Random Forest Classifier

```
GridSearchCV(estimator=RandomForestClassifier(),
              param_grid={'max_depth': [3, 5, 7],
                           'min_samples_leaf': [5, 10, 15],
                           'min_samples_split': [15, 30, 45],
                           'n_estimators': [25, 50]})
```

☐ estimator: RandomForestClassifier

RandomForestClassifier()

☐ RandomForestClassifier

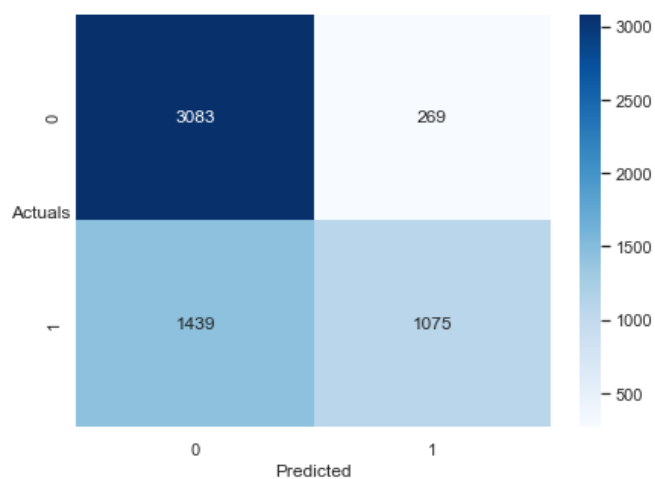
RandomForestClassifier()

Best Parameters:

```
{'max_depth': 7,  
'min_samples_leaf': 15,  
'min_samples_split': 15,  
'n_estimators': 25}
```

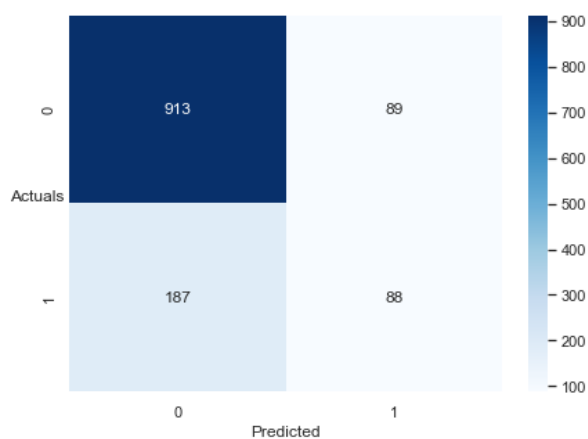
TRAIN:

	precision	recall	f1-score	support
0.0	0.68	0.92	0.78	3352
1.0	0.80	0.43	0.56	2514
accuracy			0.71	5866
macro avg	0.74	0.67	0.67	5866
weighted avg	0.73	0.71	0.69	5866



TEST:

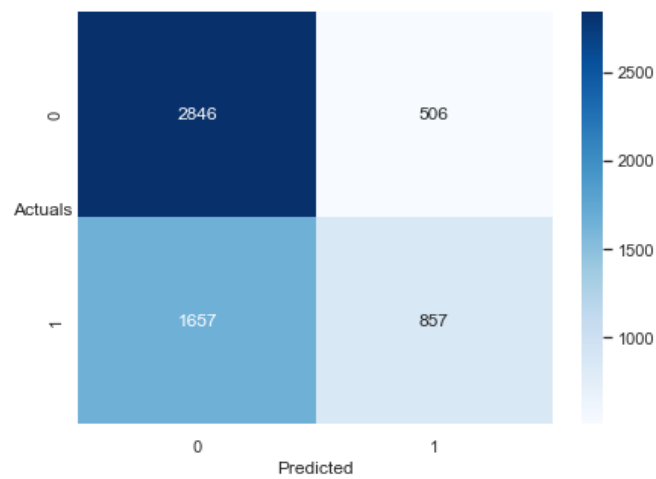
	precision	recall	f1-score	support
0.0	0.83	0.91	0.87	1002
1.0	0.50	0.32	0.39	275
accuracy			0.78	1277
macro avg	0.66	0.62	0.63	1277
weighted avg	0.76	0.78	0.77	1277



Linear Discriminant Analysis

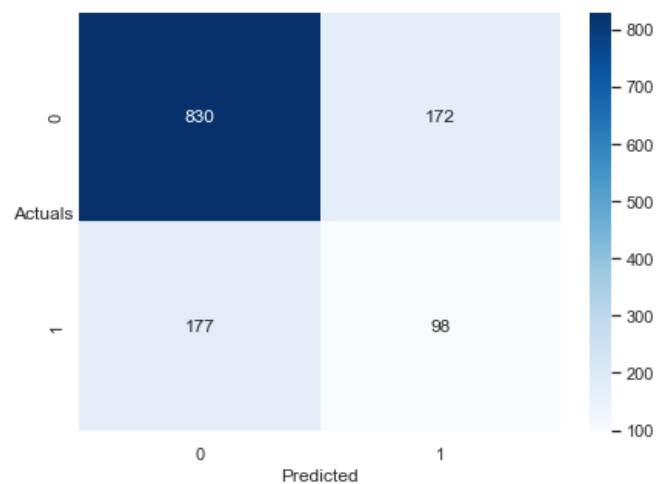
TRAIN:

	precision	recall	f1-score	support
0.0	0.63	0.85	0.72	3352
1.0	0.63	0.34	0.44	2514
accuracy			0.63	5866
macro avg	0.63	0.59	0.58	5866
weighted avg	0.63	0.63	0.60	5866



TEST:

	precision	recall	f1-score	support
0.0	0.82	0.83	0.83	1002
1.0	0.36	0.36	0.36	275
accuracy			0.73	1277
macro avg	0.59	0.59	0.59	1277
weighted avg	0.72	0.73	0.73	1277

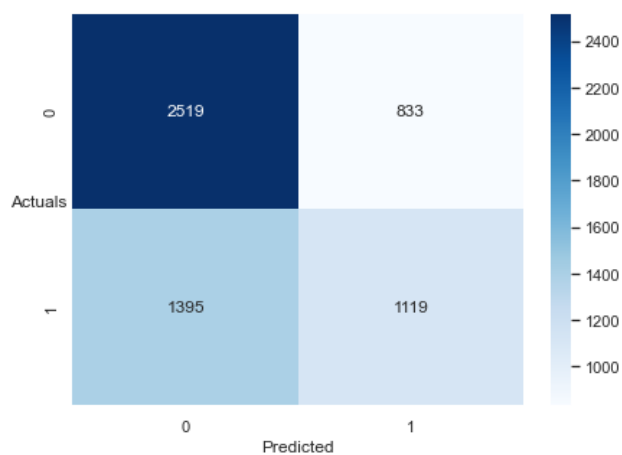


Choosing the optimal threshold

OPTIMUM THRESHOLD: 0.45243591581928677

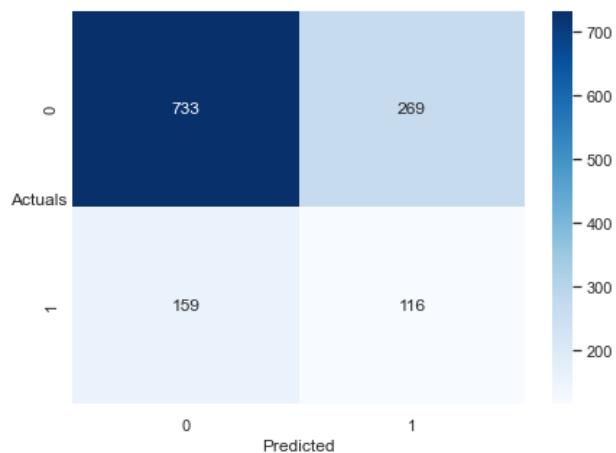
TRAIN:

	precision	recall	f1-score	support
0	0.64	0.75	0.69	3352
1	0.57	0.45	0.50	2514
accuracy			0.62	5866
macro avg	0.61	0.60	0.60	5866
weighted avg	0.61	0.62	0.61	5866



TEST:

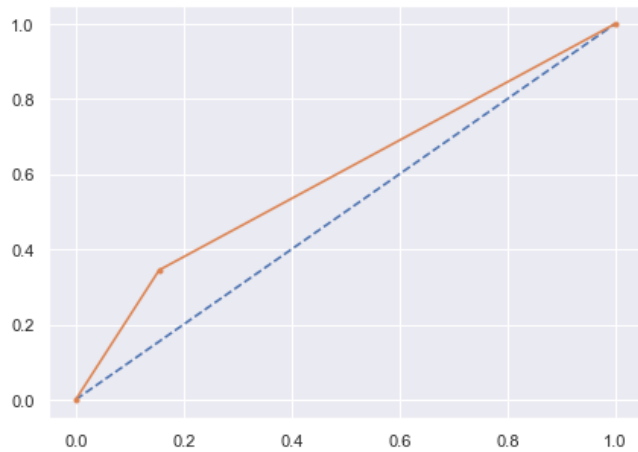
	precision	recall	f1-score	support
0	0.82	0.73	0.77	1002
1	0.30	0.42	0.35	275
accuracy			0.66	1277
macro avg	0.56	0.58	0.56	1277
weighted avg	0.71	0.66	0.68	1277



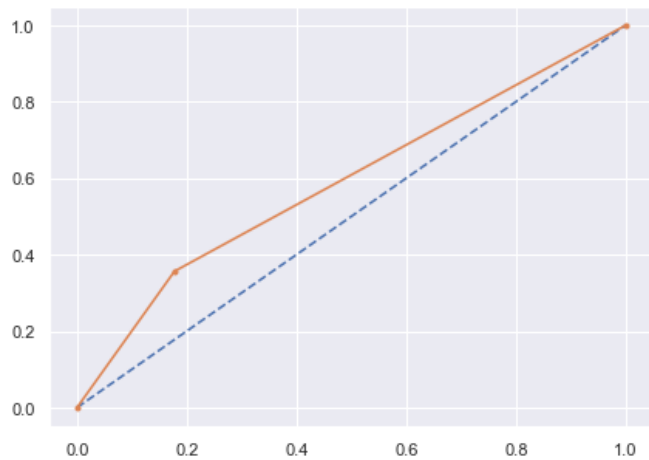
Compare the performances of Logistics, Radom Forest and LDA models (include ROC Curve)

Logistic Regression - AUC and ROC Curve for training data

TRAIN AUC: 0.595

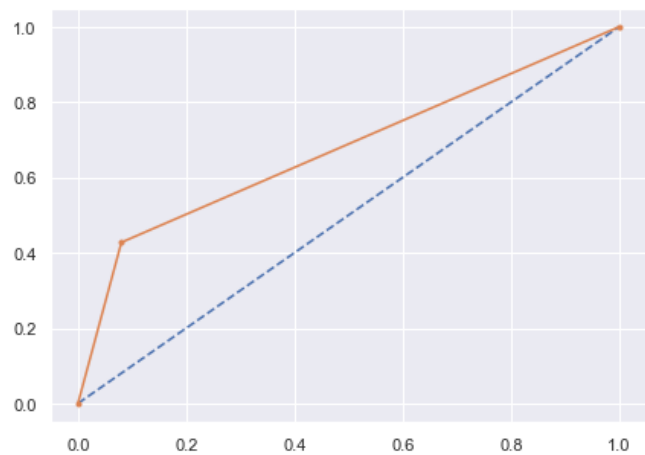


TEST AUC: 0.589

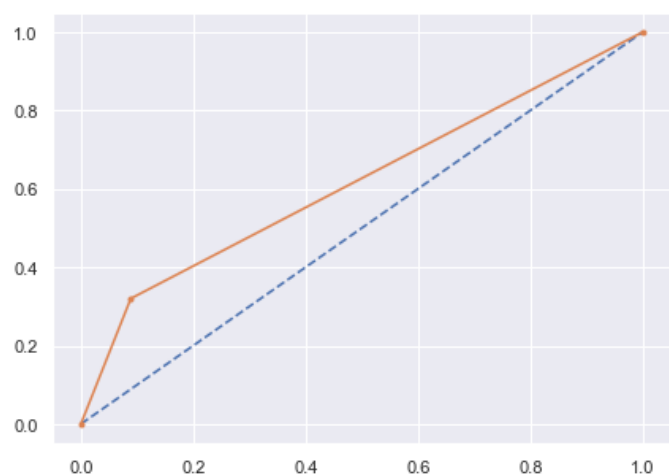


Random Forest - AUC and ROC Curve for training data

TRAIN AUC: 0.674

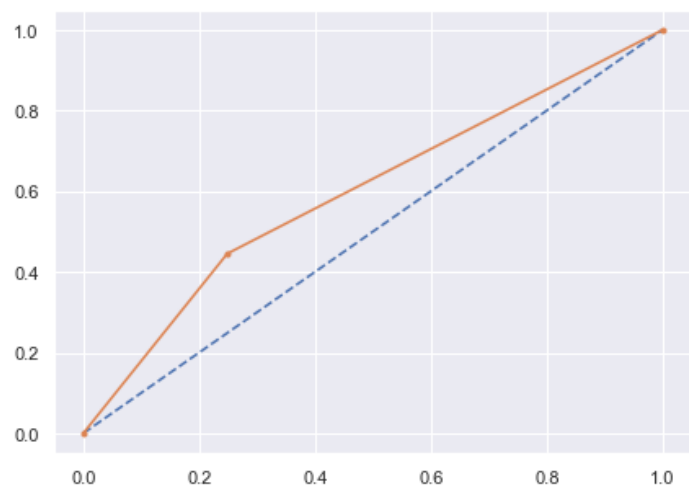


TEST AUC: 0.616

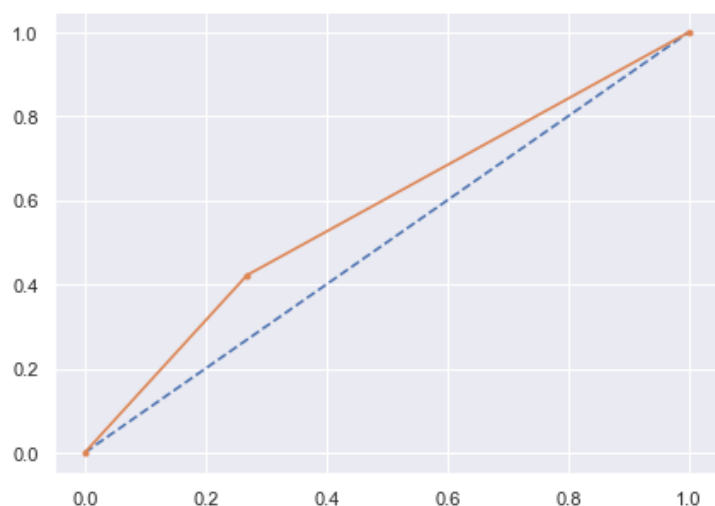


Linear Discriminant Analysis - AUC and ROC Curve for training data

TRAIN AUC: 0.595



TEST AUC: 0.592



Combined Result

	LR Train	LR Test	RF Train	RF Test	LDA Train	LDA Test
accuracy	0.63	0.72	0.71	0.78	0.62	0.66
precision	0.63	0.36	0.80	0.50	0.57	0.30
recall	0.34	0.36	0.43	0.32	0.45	0.42
f1score	0.44	0.36	0.56	0.39	0.50	0.35
auc_roc_score	0.60	0.59	0.67	0.61	0.60	0.59

Here, we can see that our recall score is not so good for the prediction so I would suggest to get more data and also that type of data which have significant columns in it but if we have to go with a model we could go with LDA one because of the highest recall score because in this recall score is the most important parameter not the accuracy If I have to suggest something I would suggest that we should pay attention to our customers who are defaulting and should check their

background and all properly and for the people who are paying properly we could give some schemes to them so that they start using more of our credit card for the defaulting one we should take strict measure and probably make them sign some papers regarding the time in which they will pay and if they delay we can increase their interest and if they fail to do so we can take some legal action on them.