Business Report

FRA Coded Project Part A

PGPDSBA

Chithira Raj

Table of Contents

List of Tables

List of Figures

1. Context

In the realm of modern finance, businesses encounter the perpetual challenge of managing debt obligations effectively to maintain a favourable credit standing and foster sustainable growth. Investors keenly scrutinize companies capable of navigating financial complexities while ensuring stability and profitability. A pivotal instrument in this evaluation process is the balance sheet, which provides a comprehensive overview of a company's assets, liabilities, and shareholder equity, offering insights into its financial health and operational efficiency. In this context, leveraging available financial data, particularly from preceding fiscal periods, becomes imperative for informed decision-making and strategic planning.

2. Objective

A group of venture capitalists want to develop a Financial Health Assessment Tool. With the help of the tool, it endeavours to empower businesses and investors with a robust mechanism for evaluating the financial well-being and creditworthiness of companies. By harnessing machine learning techniques, they aim to analyze historical financial statements and extract pertinent insights to facilitate informed decision-making via the tool. Specifically, they foresee facilitating the following with the help of the tool:

Debt Management Analysis: Identify patterns and trends in debt management practices to assess the ability of businesses to fulfil financial obligations promptly and efficiently, and identify potential cases of default.

Credit Risk Evaluation: Evaluate credit risk exposure by analyzing liquidity ratios, debt-to-equity ratios, and other key financial indicators to ascertain the likelihood of default and inform investment decisions.

They have hired you as a data scientist and provided you with the financial metrics of different companies. The task is to analyze the data provided and develop a predictive model leveraging machine learning techniques to identify whether a given company will be tagged as a defaulter in terms of net worth next year. The predictive model will help the organization anticipate potential challenges with the financial performance of the companies and enable proactive risk mitigation strategies.

3. Data Dictionary

S.No	Variable	Description
1	Networth Next Year	Net worth of the customer in the next year
2	Total assets	Total assets of customer
3	Net worth	Net worth of the customer of the present year
4	Total income	Total income of the customer
5	Change in stock	Difference between the current value of the stock and the value of stock in the last trading day
6	Total expenses	Total expenses done by the customer
7	Profit after tax	Profit after tax deduction
8	PBDITA	Profit before depreciation, income tax, and amortization
9	PBT	Profit before tax deduction
10	Cash profit	Total Cash profit
11	PBDITA as % of total income	PBDITA / Total income
12	PBT as % of total income	PBT / Total income
13	PAT as % of total income	PAT / Total income
14	Cash profit as % of total income	Cash Profit / Total income
15	PAT as % of net worth	PAT / Net worth
16	Sales	Sales done by the customer
17	Income from financial services	Income from financial services
18	Other income	Income from other sources
19	Total capital	Total capital of the customer

20	Reserves and funds	Total reserves and funds of the customer
21	Borrowings	Total amount borrowed by the customer
22	Current liabilities & provisions	current liabilities of the customer
23	Deferred tax liability	Future income tax customer will pay because of the current transaction
24	Shareholders funds	Amount of equity in a company which belongs to shareholders
25	Cumulative retained profits	Total cumulative profit retained by customer
26	Capital employed	Current asset minus current liabilities
27	TOL/TNW	Total liabilities of the customer divided by Total net worth
28	Total term liabilities / tangible net worth	Short + long term liabilities divided by tangible net worth
29	Contingent liabilities / Net worth (%)	Contingent liabilities / Net worth
30	Contingent liabilities	Liabilities because of uncertain events
31	Net fixed assets	The purchase price of all fixed assets
32	Investments	Total invested amount
33	Current assets	Assets that are expected to be converted to cash within a year
34	Net working capital	Difference between the current liabilities and current assets
35	Quick ratio (times)	Total cash divided by current liabilities
36	Current ratio (times)	Current assets divided by current liabilities
37	Debt to equity ratio (times)	Total liabilities divided by its shareholder equity
38	Cash to current liabilities (times)	Total liquid cash divided by current liabilities
39	Cash to average cost of sales per day	Total cash divided by the average cost of the sales
40	Creditors turnover	Net credit purchase divided by average trade creditors
41	Debtors turnover	Net credit sales divided by average accounts receivable
42	Finished goods turnover	Annual sales divided by average inventory
43	WIP turnover	The cost of goods sold for a period divided by the average inventory for that period
44	Raw material turnover	Cost of goods sold is divided by the average inventory for the same period
45	Shares outstanding	Number of issued shares minus the number of shares held in the company
46	Equity face value	cost of the equity at the time of issuing
47	EPS	Net income divided by the total number of outstanding share
48	Adjusted EPS	Adjusted net earnings divided by the weighted average number of common shares outstanding on a diluted basis during the plan year
49	Total liabilities	Sum of all types of liabilities
50	PE on BSE	Company's current stock price divided by its earnings per share

Table 1: Data Dictionary

4. Data Overview

4.1. Import libraries and load the data

	Num	Networth Next Year	Total assets	Net worth	Total income	Change in stock	Total expenses	Profit after tax	PBDITA	PBT	...	Debtors turnover	Finished goods turnover	WIP turnover	Raw material turnover	Shares outstanding	Equity face value	EPS	Adjusted EPS	Total liabilities
0	1	395.30	827.60	336.50	534.10	13.50	508.70	38.90	124.40	64.60	...	5.65	3.99	3.37	14.87	8760056.00	10.00	4.44	4.44	827.60
1	2	36.20	67.70	24.30	137.90	-3.70	131.00	3.20	5.50	1.00	...	NaN	NaN	NaN	NaN	NaN	NaN	0.00	0.00	67.70
2	3	84.00	238.40	78.90	331.20	-18.10	309.20	3.90	25.80	10.50	...	2.51	17.67	8.76	8.35	NaN	NaN	0.00	0.00	238.40
3	4	2041.40	6883.50	1443.30	8448.50	212.20	8482.40	178.30	418.40	185.10	...	1.91	18.14	18.62	11.11	10000000.00	10.00	17.60	17.60	6883.50
4	5	41.80	90.90	47.00	388.60	3.40	392.70	-0.70	7.20	-0.60	...	68.00	45.87	28.67	19.93	107315.00	100.00	-6.52	-6.52	90.90

5 rows × 51 columns

Figure 1: Data Overview

4.2. Check the structure of data

Shape of the dataset: 4256 rows and 51 columns

4.3. Check the types of the data

#	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 % of total income	4177 non-null	float64
12	PBT as % of total income	4177 non-null	float64
13	PAT as % of total income	4177 non-null	float64
14	Cash profit as % of total income	4177 non-null	float64
15	PAT as % 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 & 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/TNW	4256 non-null	float64
28	Total term liabilities / tangible net worth	4256 non-null	float64
29	Contingent liabilities / Net worth (%)	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

Figure 2: Datatypes

4.4. Check for and treat (if needed) missing values

Num	0
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 % of total income	79
PBT as % of total income	79
PAT as % of total income	79
Cash profit as % of total income	79
PAT as % 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 & provisions	110
Deferred tax liability	1369
Shareholders funds	0
Cumulative retained profits	45
Capital employed	0
TOL/TNW	0
Total term liabilities / tangible net worth	0
Contingent liabilities / Net worth (%)	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

Figure 3: Missing values check

4.5. Data Duplicates

There are no duplicate rows.

4.6. Statistical Summary

	count	mean	std	min	25%	50%	75%	max
Num	4256.00	2128.50	1228.75	1.00	1064.75	2128.50	3192.25	4256.00
Networth_Next_Year	4256.00	1344.74	15936.74	-74265.60	3.98	72.10	330.82	805773.40
Total_assets	4256.00	3573.62	30074.44	0.10	91.30	315.50	1120.80	1176509.20
Net_worth	4256.00	1351.95	12961.31	0.00	31.48	104.80	389.85	613151.60
Total_income	4025.00	4688.19	53918.95	0.00	107.10	455.10	1485.00	2442828.20
Change_in_stock	3706.00	43.70	436.92	-3029.40	-1.80	1.60	18.40	14185.50
Total_expenses	4091.00	4356.30	51398.09	-0.10	96.80	426.80	1395.70	2366035.30
Profit_after_tax	4102.00	295.05	3079.90	-3908.30	0.50	9.00	53.30	119439.10
PBDITA	4102.00	605.94	5646.23	-440.70	6.93	36.90	158.70	208576.50
PBT	4102.00	410.26	4217.42	-3894.80	0.80	12.60	74.17	145292.60
Cash_profit	4102.00	408.27	4143.93	-2245.70	2.90	19.40	96.25	176911.80
PBDITA_as_perc_of_total_income	4177.00	3.18	172.26	-6400.00	4.97	9.68	16.47	100.00
PBT_as_perc_of_total_income	4177.00	-18.20	419.91	-21340.00	0.56	3.34	8.94	100.00
PAT_as_perc_of_total_income	4177.00	-20.03	423.58	-21340.00	0.35	2.37	6.42	150.00
Cash_profit_as_perc_of_total_income	4177.00	-9.02	299.96	-15020.00	2.00	5.66	10.73	100.00
PAT_as_perc_of_net_worth	4256.00	10.17	61.53	-748.72	0.00	8.04	20.20	2466.67
Sales	3951.00	4645.68	53080.90	0.10	113.35	468.60	1481.20	2384984.40
Income fincial services	3145.00	81.36	1042.76	0.00	0.50	1.90	9.80	51938.20

Other_income	2700.00	55.95	1178.42	0.00	0.40	1.50	6.20	42856.70
Total_capital	4251.00	224.56	1684.95	0.10	13.20	42.60	103.15	78273.20
Reserves_and_funds	4158.00	1210.56	12816.23	-6525.90	5.30	55.15	282.52	625137.80
Borrowings	3825.00	1176.25	8581.25	0.10	24.40	99.80	358.30	278257.30
Current_liabilities_&_provisions	4146.00	960.63	9140.54	0.10	17.50	70.30	265.92	352240.30
Deferred_tax_liability	2887.00	234.50	2106.25	0.10	3.20	13.50	51.30	72796.60
Shareholders_funds	4256.00	1376.49	13010.69	0.00	32.30	107.60	408.90	613151.60
Cumulative_retained_profits	4211.00	937.18	9853.10	-6534.30	1.10	37.40	206.20	390133.80
Capital_employed	4256.00	2433.62	20496.40	0.00	61.30	221.20	790.30	891408.90
TOL_to_TNW	4256.00	4.03	20.88	-350.48	0.60	1.42	2.83	473.00
Total_term_liabilities__to__tangible_net_worth	4256.00	1.85	15.88	-325.60	0.05	0.34	1.00	456.00
Contingent_liabilities__to__Net_worth_perc	4256.00	55.71	369.17	0.00	0.00	5.36	31.01	14704.27
Contingent_liabilities	2854.00	948.55	12056.74	0.10	6.00	37.85	195.32	559506.80
Net_fixed_assets	4124.00	1209.49	12502.40	0.00	26.20	93.85	352.82	636604.60
Investments	2541.00	721.87	6793.86	0.00	1.00	8.20	63.80	199978.60
Current_assets	4176.00	1350.36	10155.57	0.10	36.60	148.35	515.00	354815.20
Net_working_capital	4219.00	162.87	3182.03	-63839.00	-1.10	16.70	86.50	85782.80
Quick_ratio_times	4151.00	1.50	9.33	0.00	0.41	0.67	1.03	341.00
Current_ratio_times	4151.00	2.26	12.48	0.00	0.93	1.23	1.72	505.00
Debt_to_equity_ratio_times	4256.00	2.87	15.60	0.00	0.22	0.79	1.75	456.00
Cash_to_current_liabilities_times	4151.00	0.53	4.80	0.00	0.02	0.07	0.19	165.00
Cash_to_average_cost_of_sales_per_day	4156.00	145.16	2521.99	0.00	2.88	8.04	21.97	128040.76
Creditors_turnover	3865.00	16.81	75.67	0.00	3.72	6.17	11.69	2401.00
Debtors_turnover	3871.00	17.93	90.16	0.00	3.81	6.47	11.85	3135.20
Finished_goods_turnover	3382.00	84.37	562.64	-0.09	8.19	17.32	40.01	17947.60
WIP_turnover	3492.00	28.68	169.65	-0.18	5.10	9.86	20.24	5651.40
Raw_material_turnover	3828.00	17.73	343.13	-2.00	3.02	6.41	11.82	21092.00
Shares_outstanding	3446.00	23764909.56	170979041.33	-2147483647.00	1308382.50	4750000.00	10906020.00	4130400545.00
Equity_face_value	3446.00	-1094.83	34101.36	-999998.90	10.00	10.00	10.00	100000.00
EPS	4256.00	-196.22	13061.95	-843181.82	0.00	1.49	10.00	34522.53
Adjusted_EPS	4256.00	-197.53	13061.93	-843181.82	0.00	1.24	7.62	34522.53
Total_liabilities	4256.00	3573.62	30074.44	0.10	91.30	315.50	1120.80	1176509.20
PE_on_BSE	1629.00	55.46	1304.45	-1116.64	2.97	8.69	17.00	51002.74

Figure 4: Statistical Summary

4.7. Insights

Financial Stability and Performance

- Net Worth & Assets: The average net worth is 1,351.95K, but there's a high standard deviation (12,961.31K), indicating significant variation among companies. Total assets range widely, with an average of 3,573.62K, but some companies have minimal assets (0.10K).
- Liabilities & Borrowings: Borrowings have a mean of 1,176.25K, with some companies having debt as high as 278,257.30K. Debt-to-Equity Ratio averages 2.87, meaning companies rely significantly on borrowed capital.
- Cash Position & Liquidity: The quick ratio (1.50) and current ratio (2.26) suggest that most companies have enough short-term assets to cover liabilities. However, cash to current liabilities (0.53) is relatively low, meaning cash reserves might not be strong enough for immediate obligations.

Profitability & Revenue

- Revenue & Profitability: Total income varies widely, averaging 4,688.19K, but some firms have income as high as 2,442,828.20K.

- Profit after tax (PAT) has a mean of 295.05K but a high standard deviation (3,079.90K), showing some firms are significantly more profitable than others. PAT as a percentage of net worth is 10.17%, meaning firms, on average, generate decent returns.
- Margins: PBDITA as % of Total Income (3.18%) suggests operating profit margins are thin for most firms. PBT as % of Total Income (-18.20%) indicates that many companies are struggling to generate pre-tax profits.

Stock Market & Valuation

- Earnings Per Share (EPS): Average EPS is negative (-196.22), suggesting that a large portion of companies are unprofitable. Adjusted EPS is similar (-197.53), reinforcing the trend.
- PE Ratio (Price to Earnings on BSE): The mean PE ratio is 55.46, but the extremely high standard deviation (1,304.45) suggests that valuation multiples are all over the place, with some companies being significantly overvalued or undervalued.

Working Capital & Efficiency

- Inventory & Receivables Turnover: Finished goods turnover (84.37) and WIP turnover (28.68) suggest inventory is moving quickly. Debtors turnover (17.93) suggests that companies are efficient in collecting payments.
- Working Capital: Net working capital is positive (162.87K), meaning firms generally have more current assets than liabilities.

Risk & Contingent Liabilities

- Contingent Liabilities: Contingent liabilities as % of net worth is 55.71%, meaning that many companies face potential off-balance-sheet risks. Some companies have extraordinarily high contingent liabilities (up to 14,704.27% of net worth).

4.8. Column Modification

Removes index column Num and derives default column from Net worth next year column.

A company will not be tagged as a defaulter if its net worth next year is positive, or else, it'll be tagged as a defaulter.

Proportion of Default

default	proportion
0	0.79
1	0.21

Figure 5: Proportion of default

5. Exploratory Data Analysis

5.1. Univariate Analysis

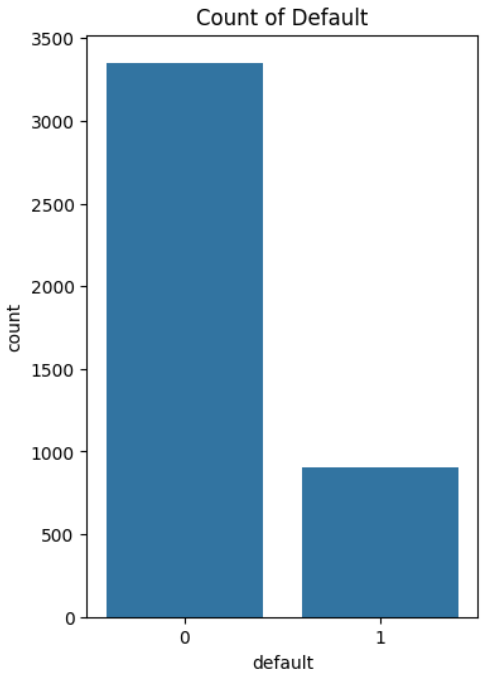

Figure 6: default

The ratio of Non-Defaulters to Defaulters is 80:20.

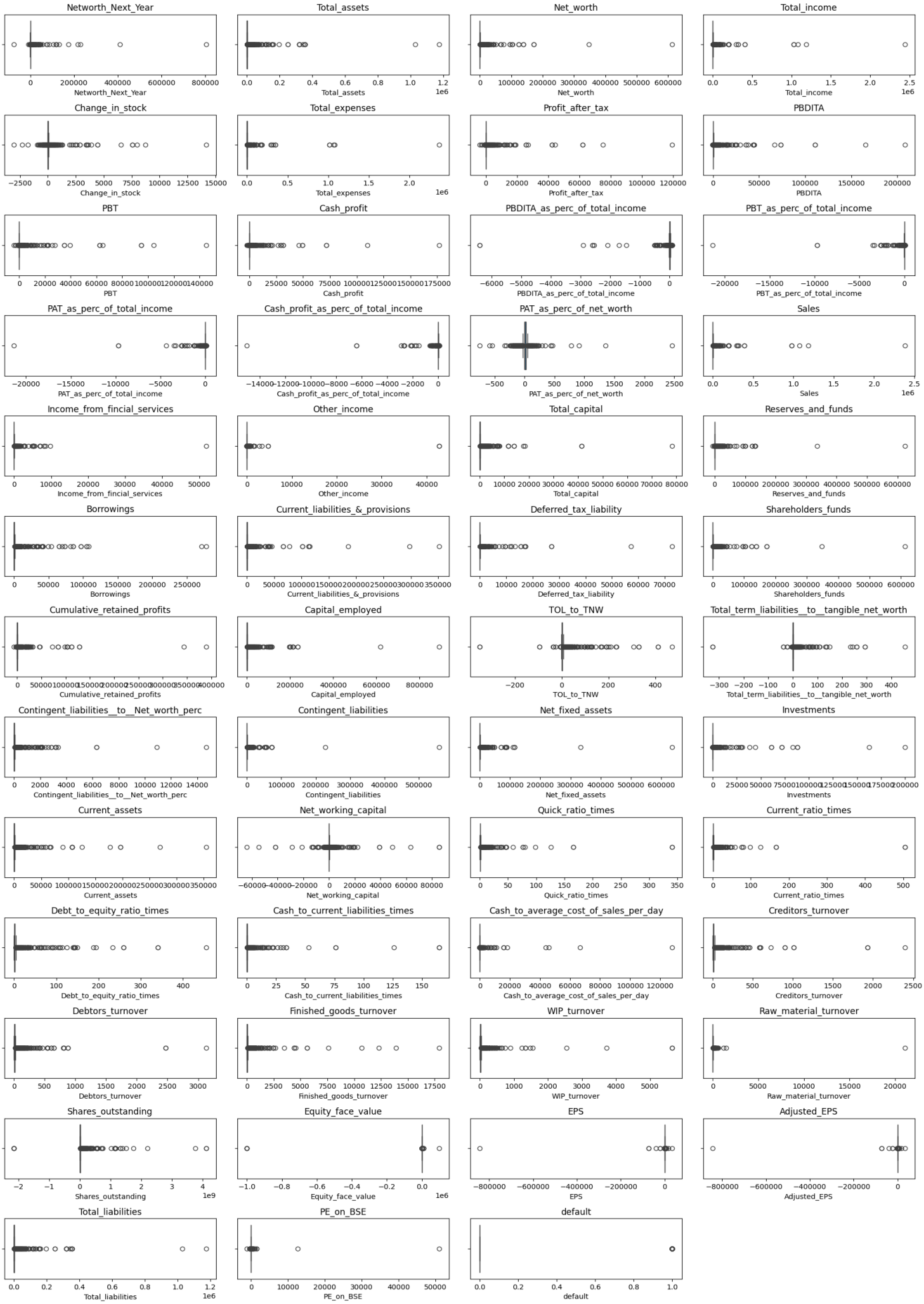

Figure 7: Numerical Boxplots

Insights

1. Presence of Significant Outliers
    - Many variables, such as Net Worth, Total Assets, Borrowings, and EPS, exhibit extreme outliers.
    - This suggests data skewness and the need for outlier treatment before modeling.
    - Recommendation: Use winsorization or log transformation to manage extreme values.
2. High Variability in Key Financial Indicators
    - PBT (Profit Before Tax), PAT (Profit After Tax), EBITDA, and Total Income show large spread, indicating high variance across companies.
    - Implication: Some companies are highly profitable, while others struggle with low or negative earnings.
    - Recommendation: Further segmentation is needed to understand what drives this disparity (e.g., industry-specific trends).
3. Liquidity & Solvency Ratios Show Extreme Variability
    - Debt-to-Equity Ratio, Quick Ratio, and Current Ratio display heavy-tailed distributions.
    - Some companies have very high leverage, which increases financial risk.
    - Recommendation: Companies with excessively high debt-to-equity should focus on debt restructuring or boosting equity.
4. Turnover Ratios Show Heavy-Tailed Distribution
    - Debtors Turnover, Finished Goods Turnover, WIP Turnover, and Raw Material Turnover show significant dispersion.
    - This indicates operational inefficiencies in some firms while others manage assets efficiently.
    - Recommendation: Improve inventory and receivables management to optimize working capital.
5. Negative & Extreme Values in Key Profitability Metrics
    - Some profitability metrics (e.g., PAT %, EBITDA %) have negative values, indicating loss-making companies.
    - Implication: Certain businesses are struggling with operational inefficiencies or financial distress.
    - Recommendation: Perform a deep dive into loss-making firms to identify underlying causes.

5.2. Bivariate Analysis

Correlation Check

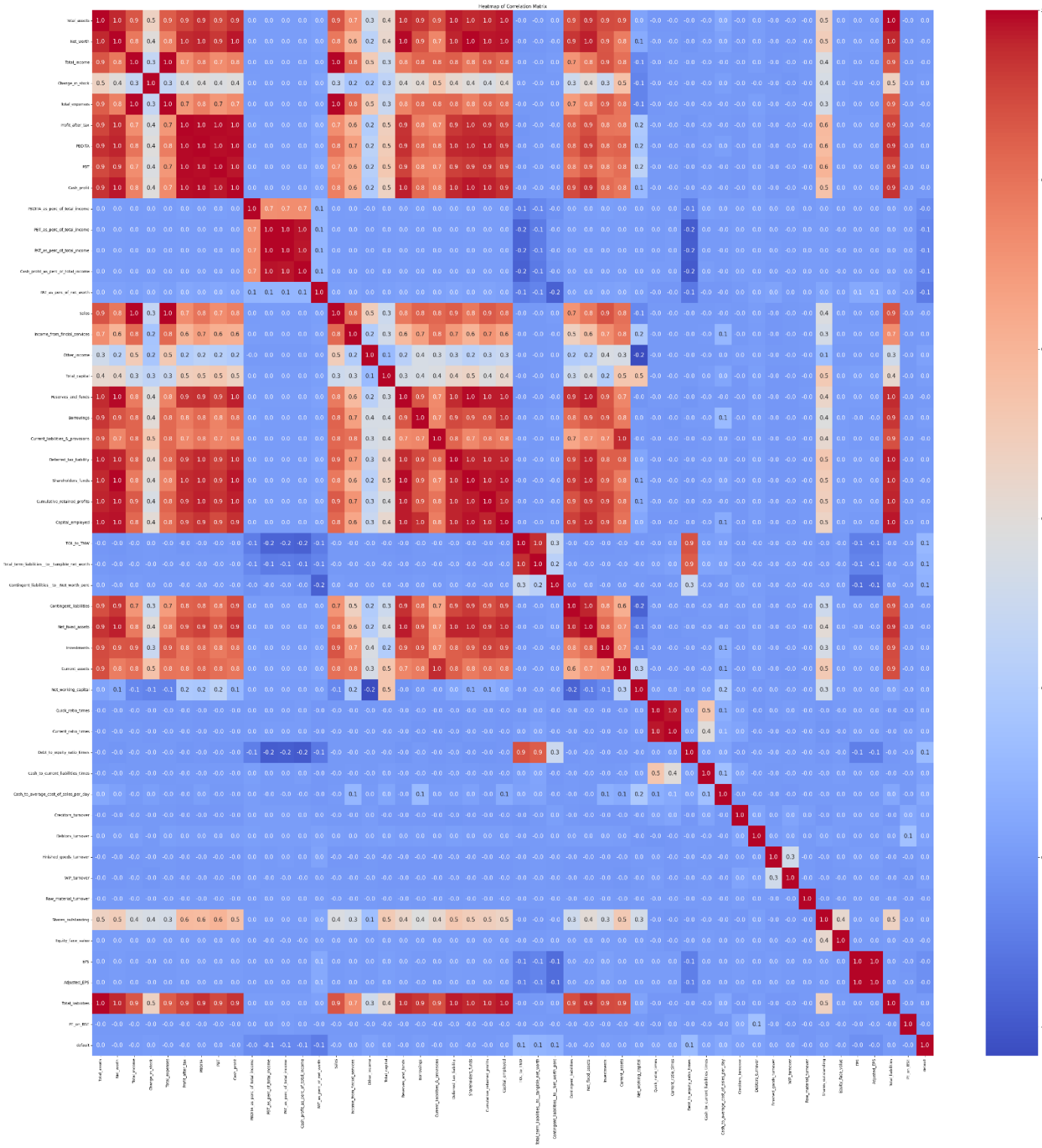

Figure 8: Heatmap

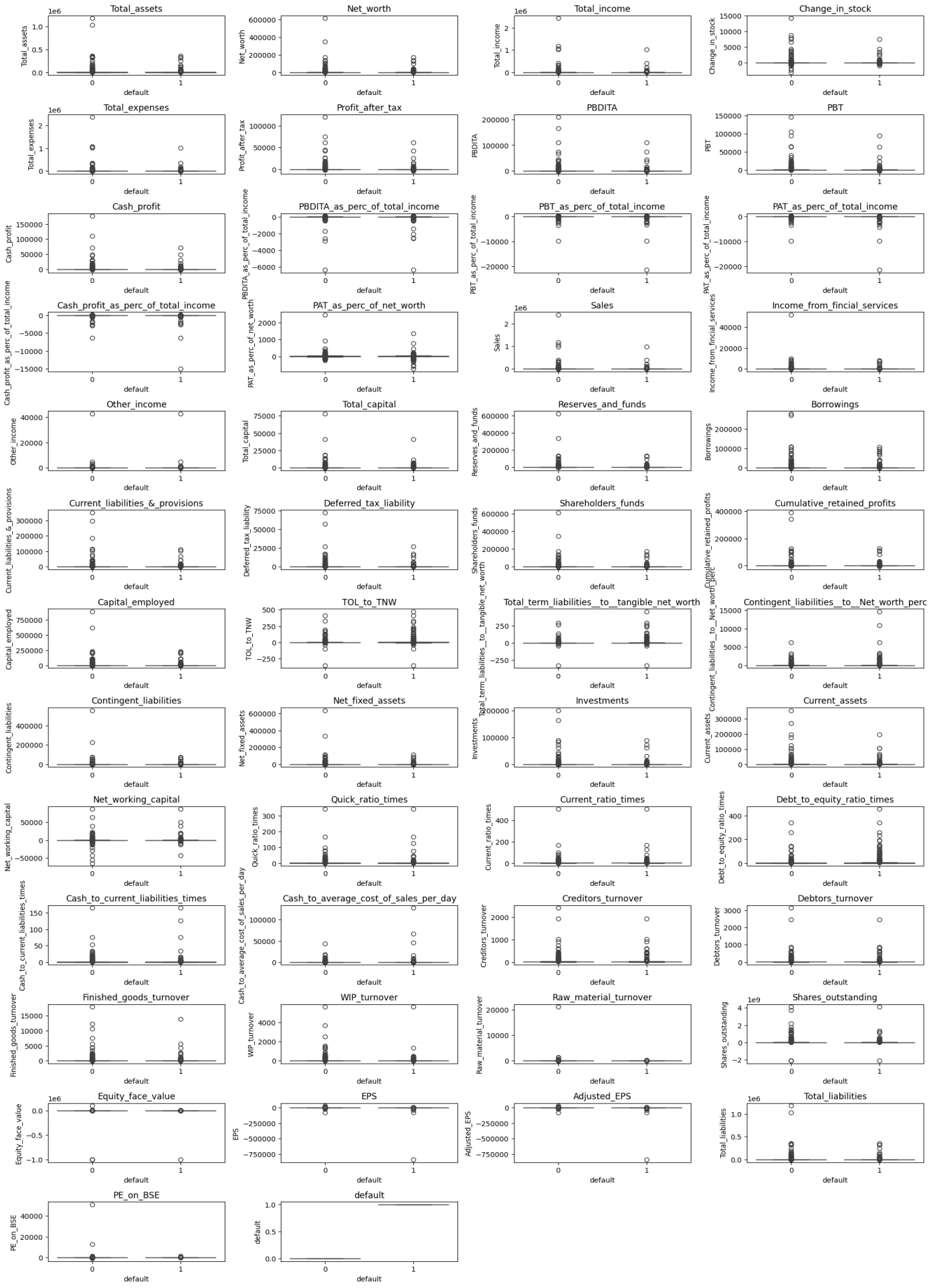

Figure 9: Boxplots Numerical columns with default column

- Higher Total Assets & Net Worth in Non-Defaulters:

  Non-defaulters generally have higher total assets and net worth, indicating financial stability.

- Debt-to-Equity Ratio is Higher in Defaulters:

  Defaulters tend to have higher debt-to-equity ratios, suggesting over-leverage and financial risk.

- Lower Quick Ratio & Current Ratio in Defaulters:

  Defaulters show lower liquidity ratios, meaning they may struggle with short-term obligations.

- Higher Contingent Liabilities for Defaulters:

  Companies that default tend to have higher contingent liabilities, indicating additional financial stress.

- Cash Flow Metrics Show Variability:

  Cash profit as a percentage of total income varies significantly, with some defaulters having extreme negative values.

Correlation

- Debt-to-Equity Ratio is Positively Correlated with Default:

  Higher debt relative to equity is associated with a higher likelihood of default, suggesting over-leveraged firms are more prone to financial distress.

- Liquidity Ratios (Current Ratio, Quick Ratio) Show Negative Correlation with Default:

  Firms with better liquidity (higher quick ratio and current ratio) are less likely to default, as they can meet short-term obligations more easily.

- Profitability Metrics (PBDITA, PAT as % of Net Worth) are Negatively Correlated with Default:

  Higher profitability is linked to a lower likelihood of default, indicating that financially strong companies are better positioned to avoid default.

- Total Liabilities Show a Moderate Positive Correlation with Default:

  Firms with higher total liabilities tend to have a higher probability of default, reinforcing the impact of financial burden on business stability.

- Contingent Liabilities and Default Show a Positive Correlation:

  Companies with significant contingent liabilities (potential obligations) are more likely to default, highlighting financial risk exposure.

- Highly Correlated Features:
  - Some variables have strong positive correlations (close to +1), suggesting they move together.
    - For example:
    - PBT as % of total income and PAT as % of total income
    - Total assets and Net worth
    - Current ratio times and Quick ratio times
  - These features might contain redundant information and could be considered for feature reduction using PCA or Variance Inflation Factor (VIF).

- Negative Correlations:

- Some variables have strong negative correlations (close to -1), indicating inverse relationships.
  - Debt-to-equity ratio times and Net worth
  - Contingent liabilities to net worth percentage and Total term liabilities to tangible net worth
- These insights suggest that as debt increases, net worth tends to decrease, which is expected in financial analysis.

- Weak Correlations:

Several features exhibit low correlations (near 0), implying little to no linear relationship. These variables may not contribute significantly to a predictive model and could be reconsidered.

- Potential Multicollinearity Issues:
  - The presence of several high correlations suggests possible multicollinearity, which could impact model performance if not addressed.
  - Techniques such as dropping redundant features, applying PCA, or using regularization (Lasso/Ridge) can help mitigate these issues.

- Defaulter vs. Non-Defaulter Insights:
  - If the target variable (Default/Non-Default) is included, analyzing its correlation with financial metrics can help identify key risk indicators.
  - Debt-related ratios and liquidity measures (such as Quick Ratio and Current Ratio) might show strong relationships with default probability.

6. Data Preprocessing

6.1. Missing Value treatment

There are missing values in the dataset.

Proportion of missing values

```
[ ]  17778/208544

0.08524819702317017
```

Figure 10: Missing value proportion

Dropping columns with more than 30% missing values.

Imputing the remaining missing values using KNNImputer with neighbours = 5

	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

Figure 11: Missing value check after imputing

6.2. Duplicate value check

There are no duplicate rows.

6.3. Outlier Detection

Total_assets	585	Total_capital	551	Current_ratio_times	397
Net_worth	595	Reserves_and_funds	643	Debt_to_equity_ratio_times	381
Total_income	508	Borrowings	532	Cash_to_current_liabilities_times	539
Change_in_stock	750	Current_liabilities_&_provisions	581	Cash_to_average_cost_of_sales_per_day	583
Total_expenses	518	Deferred_tax_liability	406	Creditors_turnover	442
Profit_after_tax	712	Shareholders_funds	588	Debtors_turnover	408
PBDITA	584	Cumulative_retained_profits	699	Finished_goods_turnover	399
PBT	704	Capital_employed	572	WIP_turnover	378
Cash_profit	627	TOL_to_TNW	414	Raw_material_turnover	296
PBDITA_as_perc_of_total_income	346	Total_term_liabilities__to__tangible_net_worth	406	Shares_outstanding	476
PBT_as_perc_of_total_income	546	Contingent_liabilities__to__Net_worth_perc	478	Equity_face_value	533
PAT_as_perc_of_total_income	610	Contingent_liabilities	393	EPS	638
Cash_profit_as_perc_of_total_income	426	Net_fixed_assets	569	Adjusted_EPS	694
PAT_as_perc_of_net_worth	427	Investments	451	Total_liabilities	585
Sales	500	Current_assets	532	PE_on_BSE	237
Income_from_fincial_services	517	Net_working_capital	806		
Other_income	389				

Figure 12: Outliers

There are outliers in the few columns. We have a few options for handling these outliers:

- Use the IQR (Interquartile Range) to determine the lower and upper bounds of the column and either replace or remove the outliers.

- However, since we lack additional information from a subject matter expert, we may decide not to treat these outliers for now.

6.4. Data Preparation for Modeling

6.4.1. Handling imbalanced data using SMOTE

```
After over sampling, predictors data volume  After over sampling, target variable volume
6704                                         6704
```

Figure 13: SMOTE

6.4.2. Train – Test Split

- Number of rows in train data = 4692
- Number of rows in test data = 2012

6.4.3. Scaling the data using StandardScaler

It is used to standardize features by removing the mean and scaling to unit variance. It transforms the data so that it has a mean of 0 and a standard deviation of 1.

7. Model building

7.1. Logistic Regression

- Feature Selection with RFECV: Used RFECV (Recursive Feature Elimination with Cross-Validation) to select the most relevant features for logistic regression, optimizing model performance.
- Cross-Validation Approach: Applied 5-fold cross-validation (cv=5) to iteratively remove less important features based on accuracy scoring.
- Selected Features: Identified the optimal subset of features using selector.support_ and extracted them from the scaled dataset.
- Model Training: Trained a LogisticRegression model using the selected features with class_weight='balanced' to handle class imbalance.
- Feature Importance Analysis: Retrieved feature rankings to understand their relative importance in the model.
- Final Dataset Preparation: Created refined training and sets using only the selected features before fitting the logistic regression model.

7.1.1. Model Performance

Train Data

Accuracy	Recall	Precision	F1
0.61	0.36	0.74	0.48

Figure 14: Model Performance

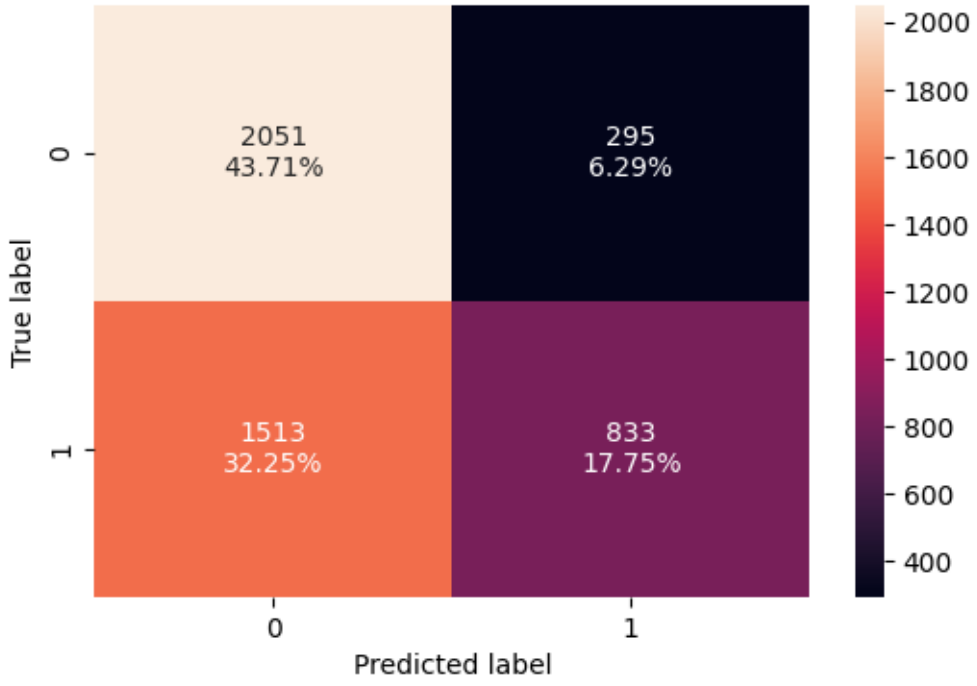

Figure 15: Confusion Matrix

Test Data

Accuracy	Recall	Precision	F1
0.63	0.38	0.75	0.51

Figure 16: Model Performance

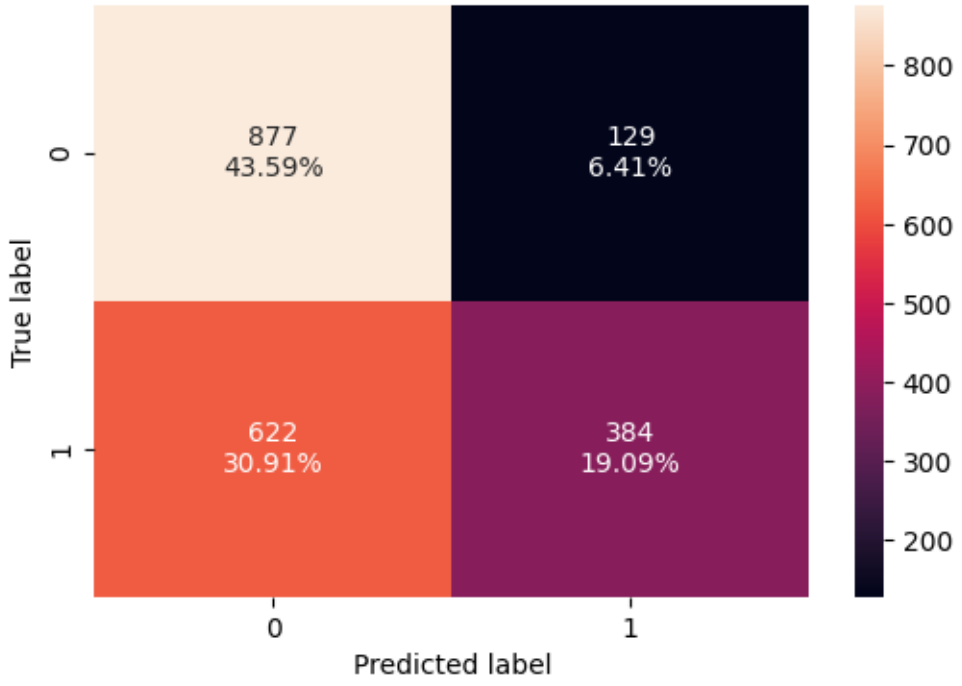

Figure 17: Confusion Matrix

7.2. Random Forest

7.2.1. Model Performance

Train Data

Accuracy	Recall	Precision	F1
0.92	0.92	0.93	0.92

Figure 18: Model Performance

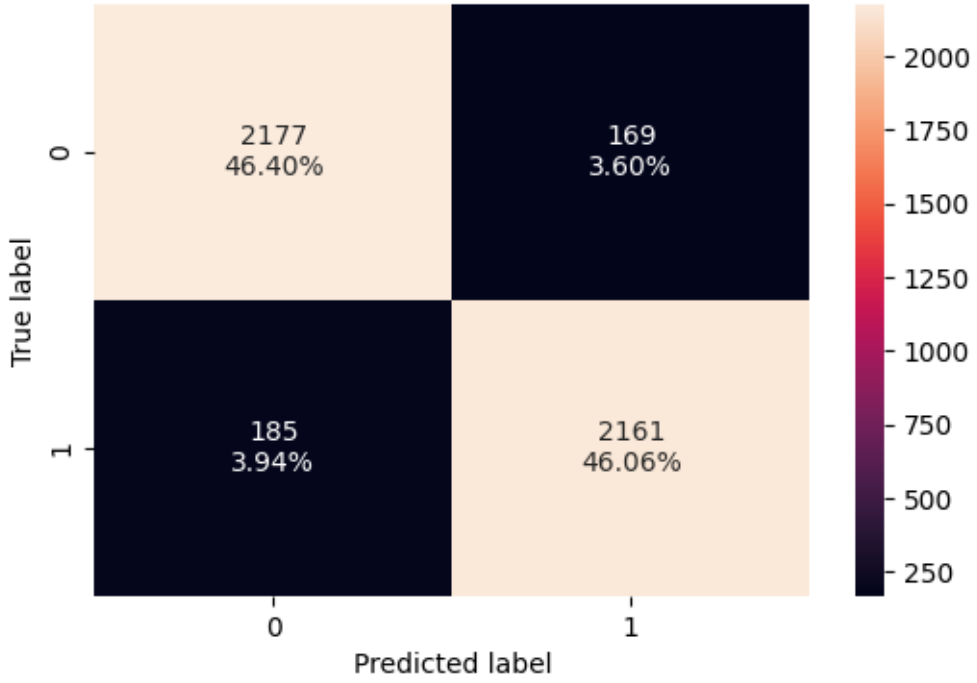

Figure 19: Confusion Matrix

Test Data

Accuracy	Recall	Precision	F1
0.69	0.69	0.69	0.69

Figure 20: Model Performance

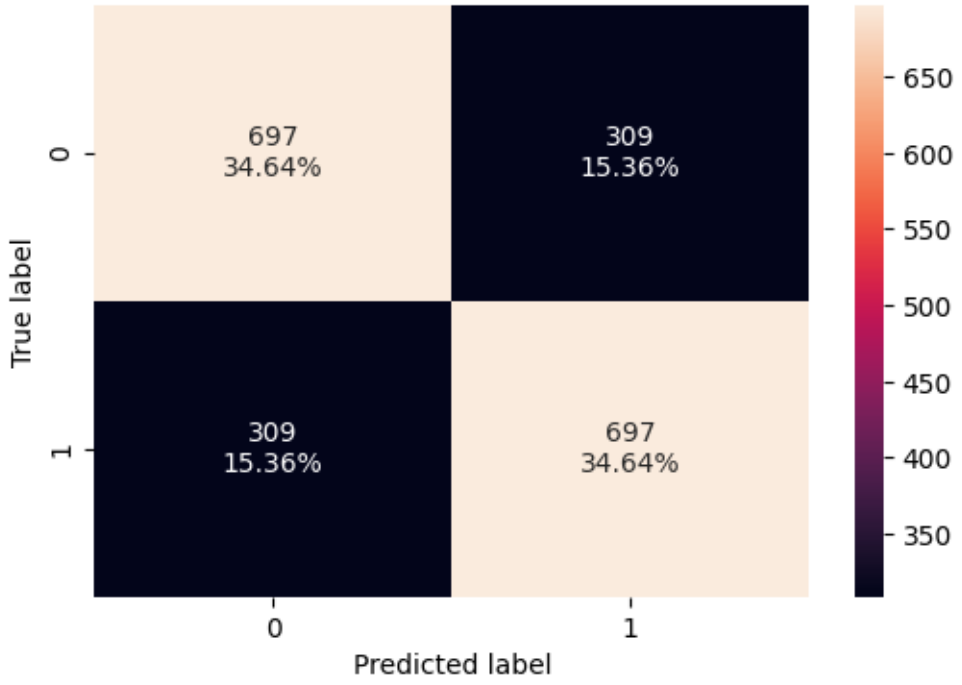

Figure 21: Confusion Matrix

8. Model Performance Improvement

- Precision (Positive Predictive Value): Important if minimizing false positives (misclassifying a non-defaulter as a defaulter) is critical.
- Recall (Sensitivity): Crucial when detecting defaults is more important (e.g., minimizing financial risk).
- F1-Score: Balances precision and recall, useful when both false positives and false negatives carry significant costs.

8.1. Logistic Regression

8.1.1. Dealing with Multicollinearity

VIF is used to measure how much the variance of an estimated regression coefficient increases when your predictors are correlated.

Here's a quick overview of what the VIF values indicate:

- VIF = 1: No correlation between the variable and other variables.
- 1 < VIF < 5: Moderate correlation; generally considered acceptable.
- VIF >= 5: High correlation; may indicate problematic multicollinearity.
- VIF > 10: Very high correlation; suggests significant multicollinearity issues.

	Variable	VIF
0	Total_assets	inf
1	Net_worth	5684.34
2	Total_income	36851.73
3	Total_expenses	34779.44
4	Profit_after_tax	1092.45
5	PBDITA	809.07
6	PBT	1220.57
7	Cash_profit	467.40
8	PBDITA_as_perc_of_total_income	2.27
9	PBT_as_perc_of_total_income	218.36
10	PAT_as_perc_of_total_income	151.51
11	Cash_profit_as_perc_of_total_income	63.51
12	PAT_as_perc_of_net_worth	1.13
13	Sales	6935.31
14	Income_from_fincial_services	6.07
15	Total_capital	29.32
16	Reserves_and_funds	735.73
17	Borrowings	2588.36
18	Current_liabilities_&_provisions	638.79
19	Shareholders_funds	10526.56
20	Cumulative_retained_profits	173.02
21	Capital_employed	15550.65
22	TOL_to_TNW	14.77
23	Total_term_liabilities__to__tangible_net_worth	12.39
24	Contingent_liabilities__to__Net_worth_perc	1.23
25	Net_fixed_assets	86.88
26	Current_assets	121.55
27	Net_working_capital	23.39
28	Debt_to_equity_ratio_times	5.01
29	Cash_to_current_liabilities_times	1.19
30	Cash_to_average_cost_of_sales_per_day	2.79
31	Creditors_turnover	1.01
32	Debtors_turnover	1.01
33	WIP_turnover	1.00
34	Raw_material_turnover	1.00
35	Shares_outstanding	7.79
36	Equity_face_value	3.13
37	EPS	2044962.44
38	Adjusted_EPS	2044960.71
39	Total_liabilities	inf

Figure 22: VIF

Drops all highly correlated variables.

Variable	VIF
PBDITA_as_perc_of_total_income	1.03
PAT_as_perc_of_net_worth	1.08
Contingent_liabilities__to__Net_worth_perc	1.07
Cash_to_current_liabilities_times	1.07
Cash_to_average_cost_of_sales_per_day	1.07
Creditors_turnover	1.00
Debtors_turnover	1.00
WIP_turnover	1.00
Raw_material_turnover	1.00
Equity_face_value	1.00

Figure 23: VIF after removing correlated variables

8.1.2. Determining optimal threshold using ROC Curve

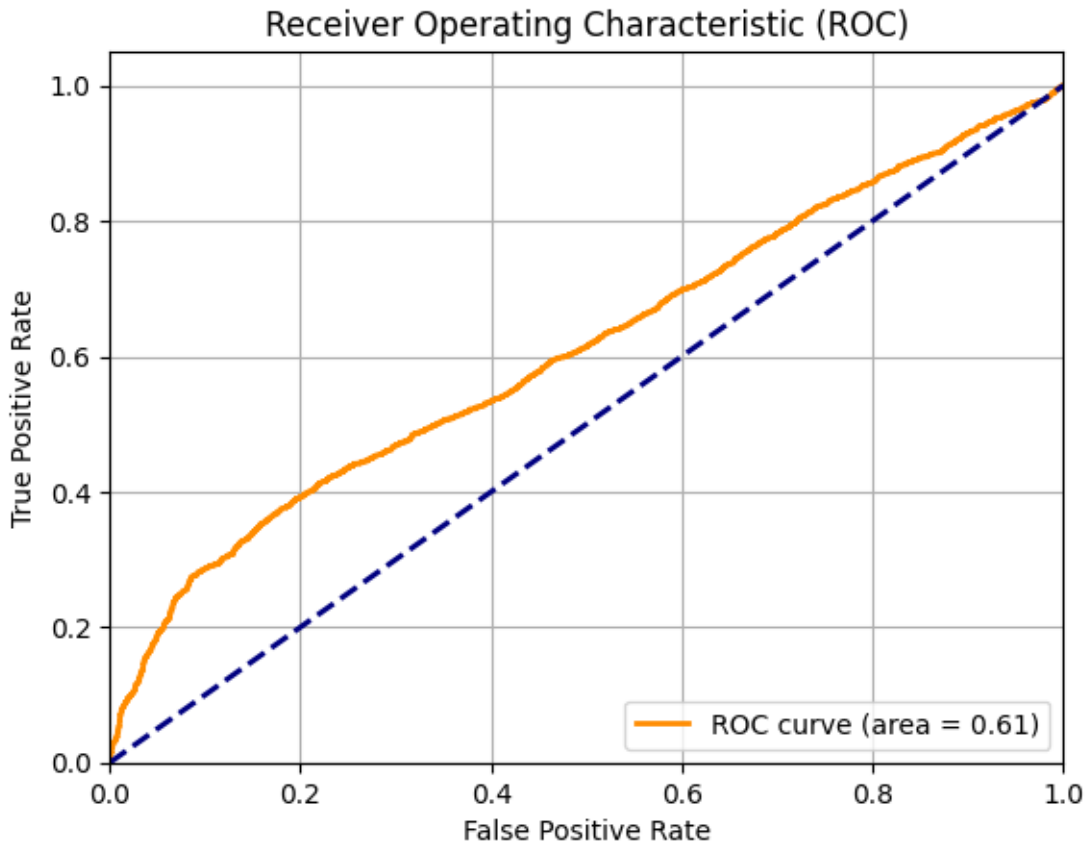

Figure 24: ROC curve

8.1.3. Tuning Logistic Regression model with significant features

After evaluating feature multicollinearity using Variance Inflation Factor (VIF) and assessing model performance with the ROC curve, the logistic regression model was retrained using only the most significant features. Removing highly correlated variables (via VIF) helped reduce redundancy, while the ROC curve analysis ensured optimal feature selection for better classification performance.

8.1.4. Tuned Model Performance

Train Data

Accuracy	Recall	Precision	F1
0.60	0.37	0.68	0.48

Figure 25: Model Performance

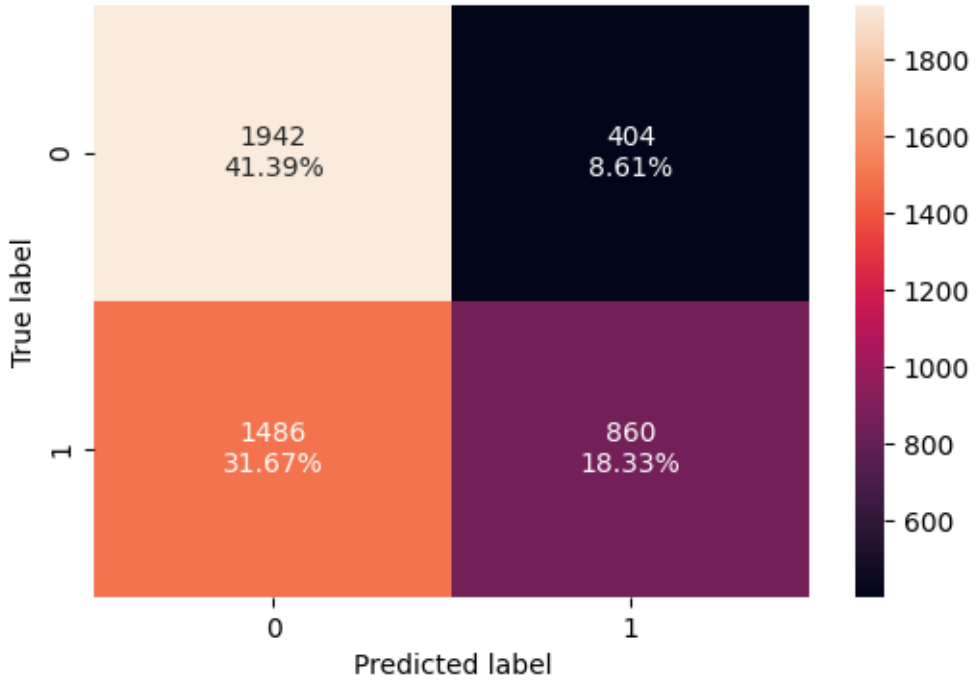

Figure 26: Confusion Matrix

Test Data

Accuracy	Recall	Precision	F1
0.59	0.40	0.64	0.49

Figure 27: Model Performance

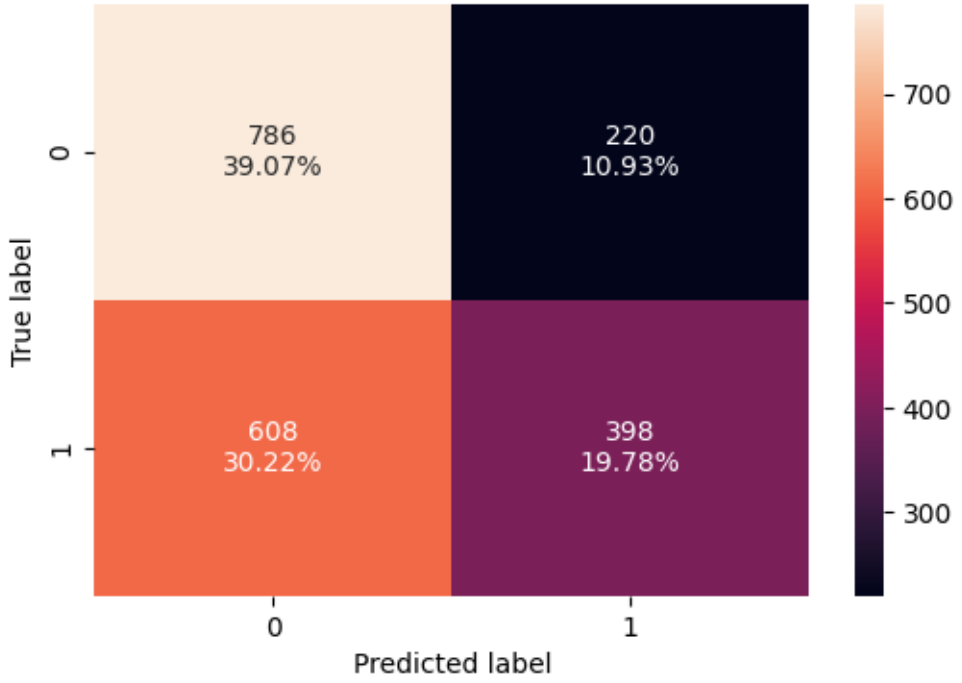

Figure 28: Confusion Matrix

8.2. Random Forest

8.2.1. Hyperparameter Tuning

```
Parameters used in the Random Forest Classifier:
bootstrap: True
ccp_alpha: 0.0
class_weight: balanced
criterion: gini
max_depth: 9
max_features: sqrt
max_leaf_nodes: None
max_samples: None
min_impurity_decrease: 0.0
min_samples_leaf: 6
min_samples_split: 2
min_weight_fraction_leaf: 0.0
monotonic_cst: None
n_estimators: 200
n_jobs: None
oob_score: False
random_state: 42
verbose: 0
warm_start: False
```

Figure 29: Best Estimators

8.2.2. Tuned Model Performance

Train Data

Accuracy	Recall	Precision	F1
0.80	0.73	0.86	0.79

Figure 30: Model Performance

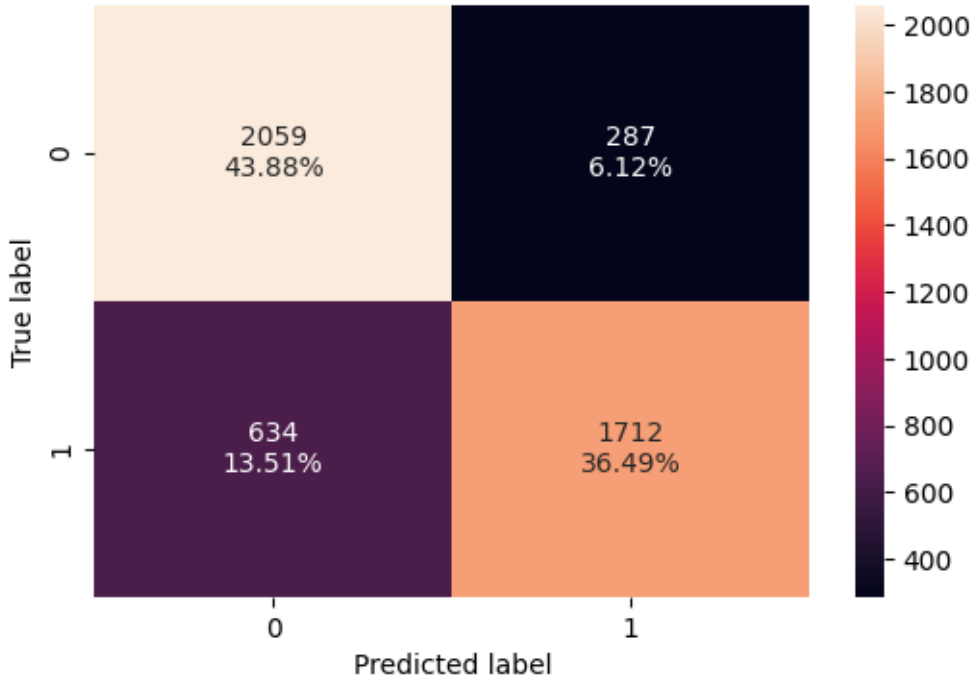

Figure 31: Confusion Matrix

Test Data

Accuracy	Recall	Precision	F1
0.69	0.62	0.71	0.67

Figure 32: Model Performance

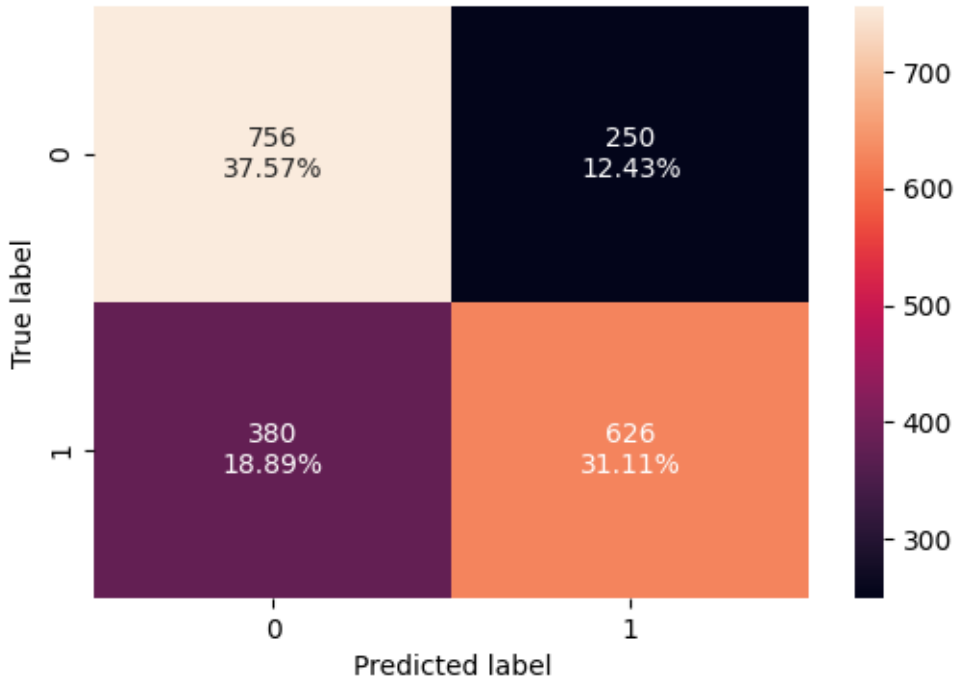

Figure 33: Confusion Matrix

9. Model Performance Comparison and Final Model Selection

Train:

	Logistic Regression Base	Logistic Regression Tuned	Random Forest Base	Random Forest Tuned
Accuracy	0.61	0.60	0.92	0.80
Recall	0.36	0.37	0.92	0.73
Precision	0.74	0.68	0.93	0.86
F1	0.48	0.48	0.92	0.79

Figure 34: Train data Model Performance Comparison

Test:

	Logistic Regression Base	Logistic Regression Tuned	Random Forest Base	Random Forest Tuned
Accuracy	0.63	0.59	0.69	0.69
Recall	0.38	0.40	0.69	0.62
Precision	0.75	0.64	0.69	0.71
F1	0.51	0.49	0.69	0.67

Figure 35: Test data Model Performance Comparison

9.1. Insights and Final Model Selection - Random Forest

Based on the results:

- If precision is the priority (reducing false positives), Logistic Regression Base is the best.
- If recall is the priority (reducing false negatives), Random Forest Base performs best but is likely overfitting.
- Random Forest Tuned balances accuracy, recall, and precision better than the base RF model.

Final Selection: Random Forest Tuned

- It has strong generalization and avoids overfitting.
- The recall (73%) is decent, ensuring we correctly identify positive cases.
- The precision (86%) is still high, avoiding false positives.

9.2. Feature Importance

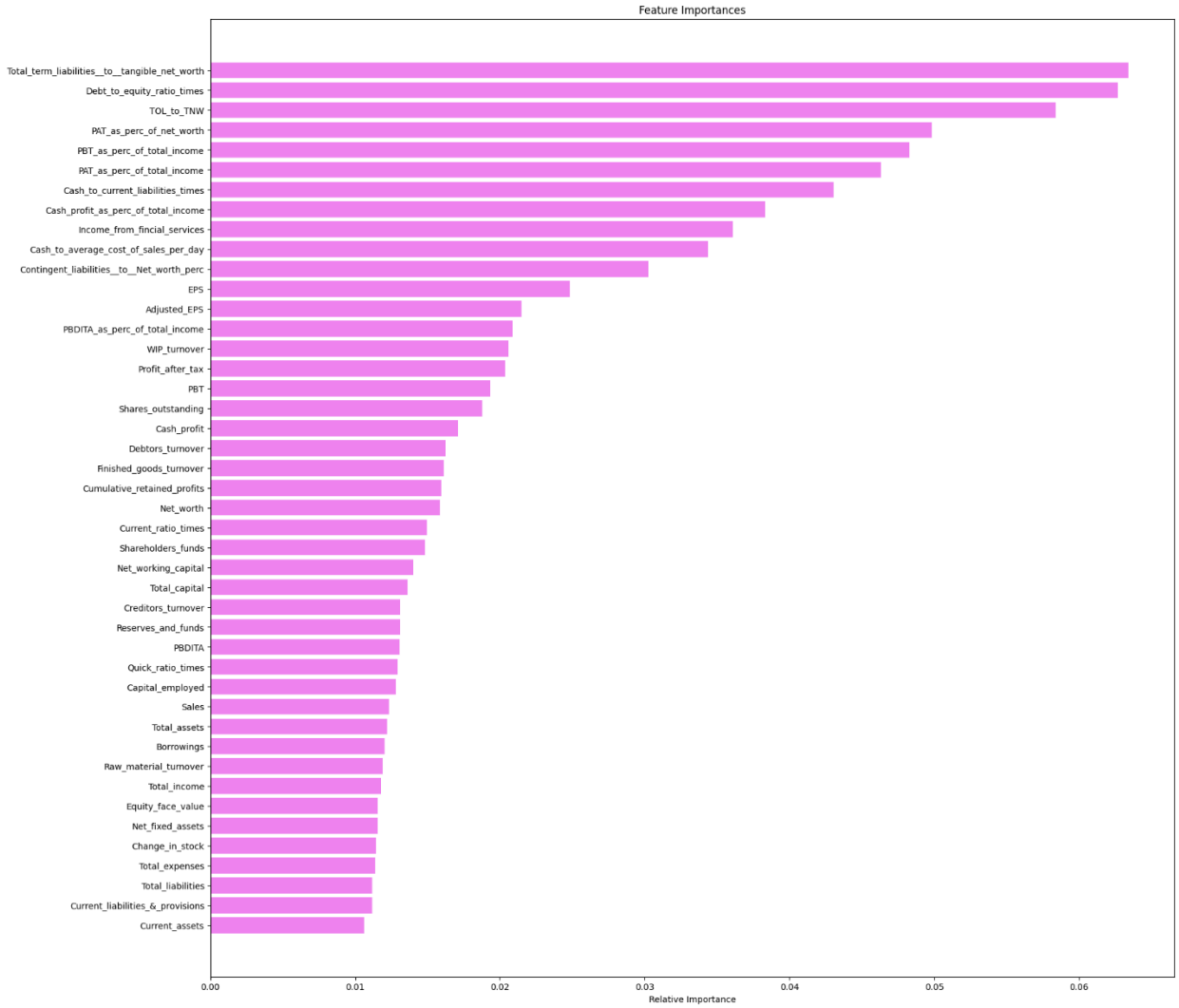

Figure 36: Feature Importance

Insights

1. Top Features Driving Predictions:
    - Total Term Liabilities to Tangible Net Worth: The most important feature, indicating that a company's leverage ratio significantly impacts predictions.
    - Debt to Equity Ratio (times): A key measure of financial risk, showing that companies with high leverage are more prone to classification changes.
    - TOL to TNW (Total Outside Liabilities to Tangible Net Worth): Another measure of debt burden that plays a crucial role in predictions.
    - PBT (Profit Before Tax) as a Percentage of Net Worth and Total Income: Indicates profitability relative to company size and revenue.
2. Cash Flow and Liquidity Metrics Matter:
    - Cash to Current Liabilities (times): Highlights the importance of liquidity management.
    - Cash Profit as a Percentage of Total Income: Indicates operational efficiency in generating cash.

- Cash to Average Cost of Sales per Day: Reflects how well a company can sustain operations using cash reserves.

3. Revenue and Profitability Play a Role:
    - Income from Financial Services: Indicates that companies with diverse revenue streams (including financial services) impact predictions.
    - Adjusted EPS (Earnings Per Share): A key profitability metric that influences classification.
    - EBITDA and PBDITA (Profit Before Depreciation, Interest, Tax, and Amortization): Important profitability measures that contribute to decision-making.
4. Turnover Ratios Have Moderate Influence:
    - WP (Work-in-Progress) Turnover: Suggests that companies with higher work-in-progress efficiency have better financial health.
    - Debtors and Finished Goods Turnover: Indicates that how fast companies convert assets into revenue is crucial.
    - Creditors Turnover: Highlights the importance of managing supplier payments efficiently.
5. Lower-Impact Features:
    - Total Assets, Borrowings, and Fixed Assets: Though relevant, they have lower importance in driving predictions.
    - Quick Ratio and Reserves & Funds: While useful, they don't have as much impact as profitability and leverage ratios.

10. Actionable Insights and Recommendations

1. Improve Financial Stability by Managing Debt & Liabilities
    - High debt ratios (Debt-to-Equity, Total Term Liabilities to Tangible Net Worth) indicate financial risk.
    - Recommendations:
    - Reduce excessive borrowing and explore equity financing. Optimize debt restructuring to lower interest costs. Maintain a healthy debt-equity balance to attract investors.
2. Focus on Profitability for Sustainable Growth
    - Key Profitability Indicators: PBT %, Adjusted EPS, EBITDA.
    - Recommendations:
    - Improve operational efficiency to maximize profit margins. Implement cost control measures and pricing strategies. Leverage automation and technology for productivity improvements.
3. Strengthen Liquidity & Cash Flow Management
    - Critical Factors: Cash to Liabilities, Cash Profit %, Cash to Cost of Sales.
    - Recommendations:

- Maintain strong cash reserves to handle economic downturns. Optimize accounts receivable management for faster cash inflows. Reduce unnecessary expenses by negotiating better supplier terms.

4. Enhance Operational Efficiency & Asset Utilization
    - Turnover Ratios (Debtors, Creditors, Finished Goods, WIP) impact financial health.
    - Recommendations:
    - Implement inventory optimization to prevent overstocking. Strengthen supply chain management to reduce costs. Improve collections strategy to enhance working capital.
5. Optimize Capital Structure for Long-Term Growth
    - Net Worth, Shareholders' Funds, Total Capital influence investor confidence.
    - Recommendations:
    - Diversify funding sources to reduce reliance on debt. Reinvest profits into high-growth areas for expansion. Improve investor relations with financial transparency.