FRA Project - Guided A

Finance and Risk Analytics

Isha Shukla

26 July 2024

Contents

Plots

Part A

Context

In the realm of modern finance, businesses encounter the perpetual challenge of managing debt obligations effectively to maintain a favorable 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.

Objective

A renowned credit rating organization wants to develop a Financial Health Assessment Tool. With the help of the tool, it endeavors to empower businesses and investors with a robust mechanism for evaluating the financial well-being and creditworthiness of companies. By harnessing machine learning techniques, the organization aims to analyze historical financial statements and extract pertinent insights to facilitate informed decision-making via the tool. Specifically, the organization foresees facilitating the following with the help of the tool:

1. Debt Management Analysis: Identify patterns and trends in debt management practices to assess the ability of businesses to fulfill financial obligations promptly and efficiently, and identify potential cases of default.
2. 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.

As a part of the data science team in the organization, you have been provided 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 default on its debt repayments in the next two quarters. The predictive model will help the organization anticipate potential challenges with the financial performance of the companies and enable proactive risk mitigation strategies.

Data Dictionary

The data consists of financial metrics from the balance sheets of different companies. The detailed data dictionary is available in the data dictionary file (*FRA_DataDictionary.xlsx*).

Exploratory Data Analysis

I. Shape of the data

```
(2058, 58)
```

II. Data Information

- There are 53 features of float data type, 4 features of integer data type and 1 feature if object data types.

```
 #   Column                                                Non-Null Count  Dtype
---  ------                                                --------------  -----
 0   Co_Code                                               2058 non-null   int64
 1   Co_Name                                               2058 non-null   object
 2   Operating_Expense_Rate                                2058 non-null   float64
 3   Research_and_development_expense_rate                 2058 non-null   float64
 4   Cash_flow_rate                                        2058 non-null   float64
 5   Interest_bearing_debt_interest_rate                   2058 non-null   float64
 6   Tax_rate_A                                            2058 non-null   float64
 7   Cash_Flow_Per_Share                                   1891 non-null   float64
 8   Per_Share_Net_profit_before_tax_Yuan_                 2058 non-null   float64
 9   Realized_Sales_Gross_Profit_Growth_Rate               2058 non-null   float64
 10  Operating_Profit_Growth_Rate                          2058 non-null   float64
 11  Continuous_Net_Profit_Growth_Rate                     2058 non-null   float64
 12  Total_Asset_Growth_Rate                               2058 non-null   float64
 13  Net_Value_Growth_Rate                                 2058 non-null   float64
 14  Total_Asset_Return_Growth_Rate_Ratio                  2058 non-null   float64
 15  Cash_Reinvestment_perc                                2058 non-null   float64
 16  Current_Ratio                                         2058 non-null   float64
 17  Quick_Ratio                                           2058 non-null   float64
 18  Interest_Expense_Ratio                                2058 non-null   float64
 19  Total_debt_to_Total_net_worth                         2037 non-null   float64
 20  Long_term_fund_suitability_ratio_A                    2058 non-null   float64
 21  Net_profit_before_tax_to_Paid_in_capital              2058 non-null   float64
 22  Total_Asset_Turnover                                  2058 non-null   float64
 23  Accounts_Receivable_Turnover                          2058 non-null   float64
 24  Average_Collection_Days                               2058 non-null   float64
 25  Inventory_Turnover_Rate_times                         2058 non-null   float64
 26  Fixed_Assets_Turnover_Frequency                       2058 non-null   float64
 27  Net_Worth_Turnover_Rate_times                         2058 non-null   float64
 28  Operating_profit_per_person                           2058 non-null   float64
 29  Allocation_rate_per_person                            2058 non-null   float64
 30  Quick_Assets_to_Total_Assets                          2058 non-null   float64
 31  Cash_to_Total_Assets                                  1962 non-null   float64
 32  Quick_Assets_to_Current_Liability                     2058 non-null   float64
 33  Cash_to_Current_Liability                             2058 non-null   float64
 34  Operating_Funds_to_Liability                          2058 non-null   float64
 35  Inventory_to_Working_Capital                          2058 non-null   float64
 36  Inventory_to_Current_Liability                        2058 non-null   float64
 37  Long_term_Liability_to_Current_Assets                 2058 non-null   float64
 38  Retained_Earnings_to_Total_Assets                     2058 non-null   float64
 39  Total_income_to_Total_expense                         2058 non-null   float64
 40  Total_expense_to_Assets                               2058 non-null   float64
 41  Current_Asset_Turnover_Rate                           2058 non-null   float64
 42  Quick_Asset_Turnover_Rate                             2058 non-null   float64
 43  Cash_Turnover_Rate                                    2058 non-null   float64
 44  Fixed_Assets_to_Assets                                2058 non-null   float64
 45  Cash_Flow_to_Total_Assets                             2058 non-null   float64
 46  Cash_Flow_to_Liability                                2058 non-null   float64
 47  CFO_to_Assets                                         2058 non-null   float64
 48  Cash_Flow_to_Equity                                   2058 non-null   float64
 49  Current_Liability_to_Current_Assets                   2044 non-null   float64
 50  Liability_Assets_Flag                                 2058 non-null   int64
 51  Total_assets_to_GNP_price                             2058 non-null   float64
 52  No_credit_Interval                                    2058 non-null   float64
 53  Degree_of_Financial_Leverage_DFL                      2058 non-null   float64
 54  Interest_Coverage_Ratio_Interest_expense_to_EBIT      2058 non-null   float64
 55  Net_Income_Flag                                       2058 non-null   int64
 56  Equity_to_Liability                                   2058 non-null   float64
 57  Default                                               2058 non-null   int64
dtypes: float64(53), int64(4), object(1)
memory usage: 932.7+ KB
```

Data Info

- There were no duplicate values.

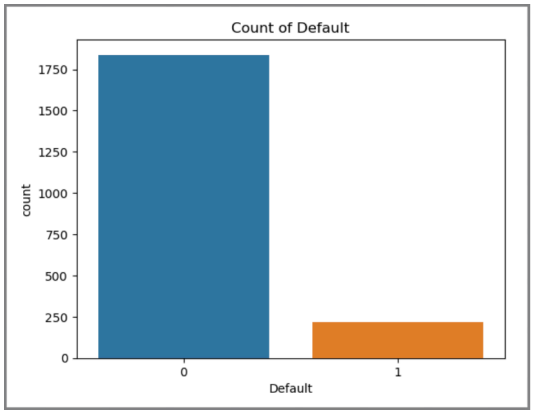
- Unique entries in the dataset.

```
Co_Code                                                  2058
Co_Name                                                  2058
Operating_Expense_Rate                                   1495
Research_and_development_expense_rate                     629
Cash_flow_rate                                           1888
Interest_bearing_debt_interest_rate                       813
Tax_rate_A                                                985
Cash_Flow_Per_Share                                       900
Per_Share_Net_profit_before_tax_Yuan_                     876
Realized_Sales_Gross_Profit_Growth_Rate                  1939
Operating_Profit_Growth_Rate                             2015
Continuous_Net_Profit_Growth_Rate                        2014
Total_Asset_Growth_Rate                                   922
Net_Value_Growth_Rate                                    1757
Total_Asset_Return_Growth_Rate_Ratio                     1428
Cash_Reinvestment_perc                                   1690
Current_Ratio                                            1972
Quick_Ratio                                              1970
Interest_Expense_Ratio                                   1716
Total_debt_to_Total_net_worth                            1949
Long_term_fund_suitability_ratio_A                       2014
Net_profit_before_tax_to_Paid_in_capital                 1798
Total_Asset_Turnover                                      283
Accounts_Receivable_Turnover                             1109
Average_Collection_Days                                  1935
Inventory_Turnover_Rate_times                            1151
Fixed_Assets_Turnover_Frequency                          1079
Net_Worth_Turnover_Rate_times                             529
Operating_profit_per_person                              1484
Allocation_rate_per_person                               2051
Quick_Assets_to_Total_Assets                             2058
Cash_to_Total_Assets                                     1962
Quick_Assets_to_Current_Liability                        2058
Cash_to_Current_Liability                                2056
Operating_Funds_to_Liability                             2058
Inventory_to_Working_Capital                             1931
Inventory_to_Current_Liability                           1932
Long_term_Liability_to_Current_Assets                    1398
Retained_Earnings_to_Total_Assets                        2058
Total_income_to_Total_expense                            2056
Total_expense_to_Assets                                  2058
Current_Asset_Turnover_Rate                              1973
Quick_Asset_Turnover_Rate                                1743
Cash_Turnover_Rate                                       1440
Fixed_Assets_to_Assets                                   2054
Cash_Flow_to_Total_Assets                                2058
Cash_Flow_to_Liability                                   2058
CFO_to_Assets                                            2058
Cash_Flow_to_Equity                                      2058
Current_Liability_to_Current_Assets                      2044
Liability_Assets_Flag                                       2
Total_assets_to_GNP_price                                2058
No_credit_Interval                                       2057
Degree_of_Financial_Leverage_DFL                         1940
Interest_Coverage_Ratio_Interest_expense_to_EBIT         1945
Net_Income_Flag                                             1
Equity_to_Liability                                      2058
Default                                                     2
dtype: int64
```

Unique entries count in each column.

III. Univariate analysis - Count of Default

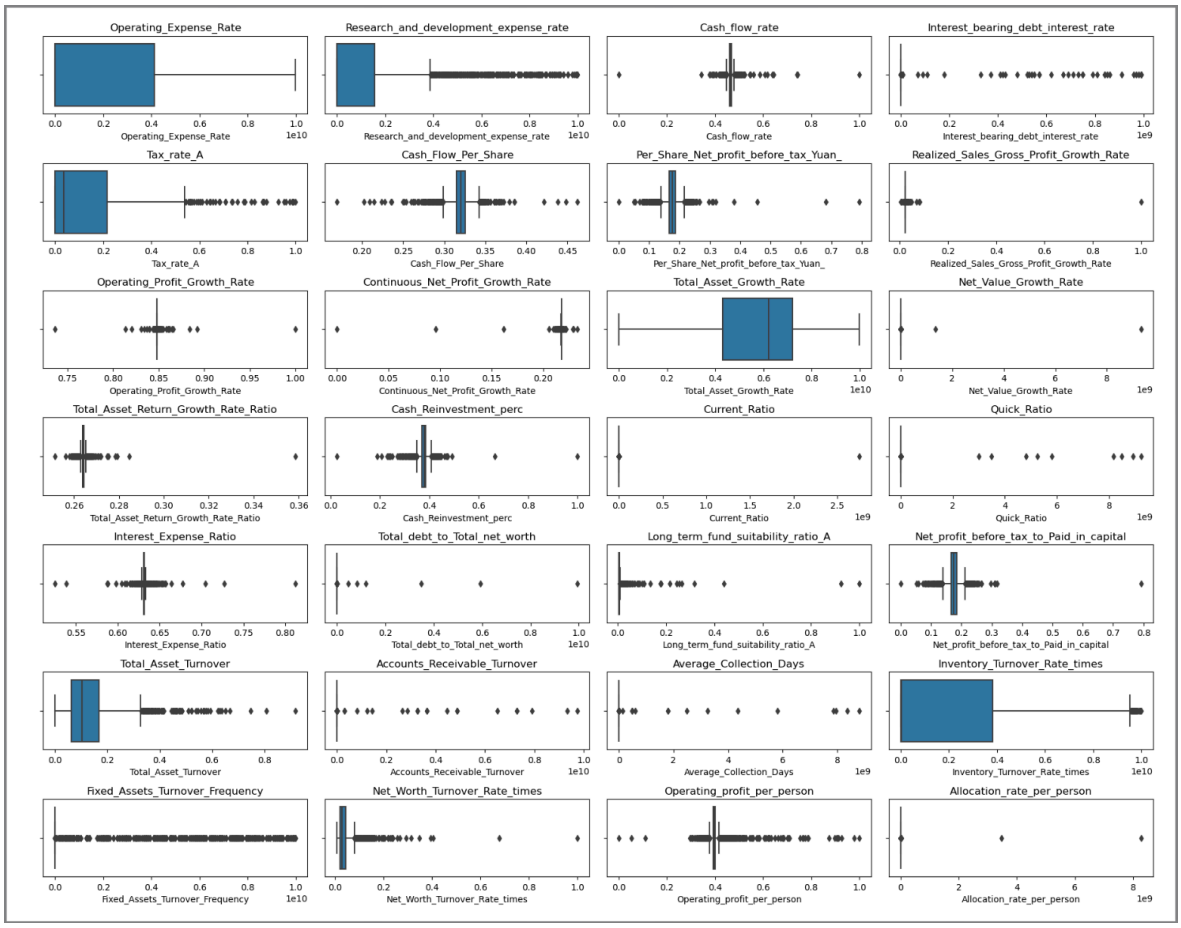
The number of instances where default did not occur (category '0') is significantly higher than the number of instances where default did occur (category '1'). This indicates that non-default cases are more prevalent than default cases within this dataset.

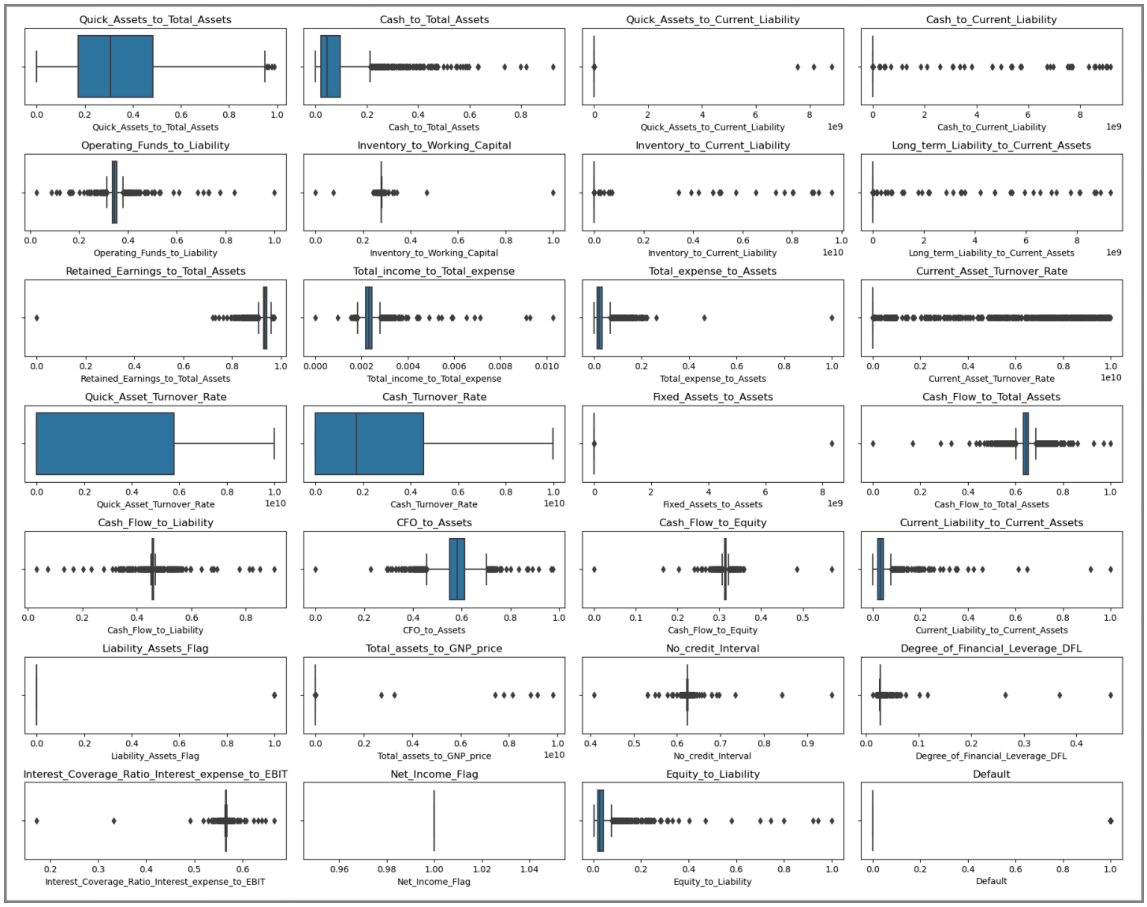


- **Percentage of Default**

```
Percentage of defaulters 10.69 %
```

- **Boxplot**

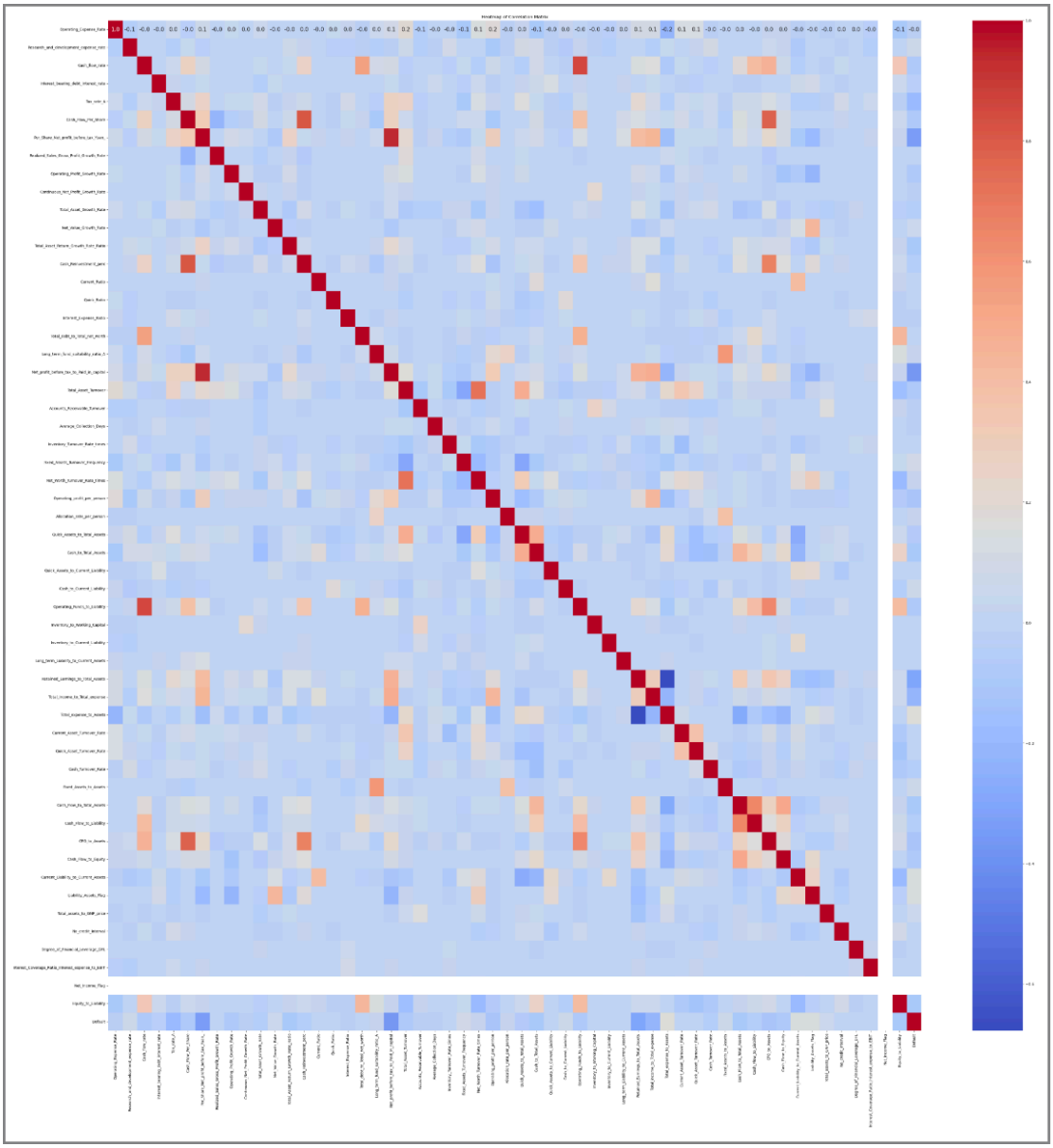




Boxplot for all the columns

Summary:

- The operating expense rate is around 50%, whereas the R&D expense rate is significantly lower.
- The total asset growth rate corresponds to the height of the blue bar, though exact values are not specified.
- The net value growth rate shows variation across different points on the x-axis.
- The net income flag appears constant (value of 1) across both default groups.
- The equity-to-liability ratio shows more variability and is lower for defaulting companies.



Heatmap

IV. Bivariate Analysis

- Various metrics show moderate correlations indicating either a positive or negative linear relationship.

Interpretation of Selected Metrics

Operating Expense Rate :
- Negatively correlated with Research and Development Expense Rate (-0.056053): Suggests that as operating expenses increase, research and development expenses tend to decrease slightly.
- Positively correlated with Per Share Net Profit Before Tax (0.071273): Indicates that as operating expenses increase, per share net profit before tax also increases slightly.

Cash Flow Rate :
- Positively correlated with Cash Flow Per Share (0.277982): Indicates a strong positive relationship between cash flow rate and cash flow per share.
- Negatively correlated with Total Expense to Assets (-0.069296): Suggests that as cash flow rate increases, total expenses relative to assets decrease slightly.

Interest Bearing Debt Interest Rate :
- Positively correlated with Current Liability to Current Assets (0.064705): Indicates a slight positive relationship between interest-bearing debt interest rate and current liability to current assets ratio.
- Negatively correlated with Cash to Total Assets (-0.107487): Suggests that higher interest-bearing debt interest rates are associated with lower cash relative to total assets.

Tax Rate A :
- Positively correlated with Cash Flow Per Share (0.110684): Indicates that higher tax rates are associated with higher cash flow per share.
- Negatively correlated with Total Expense to Assets (-0.037508): Suggests that higher tax rates are associated with slightly lower total expenses relative to assets.

Cash Flow Per Share :
- Positively correlated with Per Share Net Profit Before Tax (0.324343): Indicates a strong positive relationship between cash flow per share and per share net profit before tax.
- Negatively correlated with Realized Sales Gross Profit Growth Rate (-0.222914): Suggests that higher cash flow per share is associated with lower realized sales gross profit growth rate.

V. Data Preprocessing

1. Drop Columns with Few Unique Values:
- Remove columns like **Net_Income_Flag** and **Liability_Assets_Flag** due to their limited unique values.

2. Outliers Check:
- Examine and address outliers in the dataset.

Number of outliers in each column:

	Column	No. of outliers
0	Operating_Expense_Rate	0
1	Research_and_development_expense_rate	264
2	Cash_flow_rate	206
3	Interest_bearing_debt_interest_rate	94
4	Tax_rate_A	42
5	Cash_Flow_Per_Share	146
6	Per_Share_Net_profit_before_tax_Yuan_	186
7	Realized_Sales_Gross_Profit_Growth_Rate	283
8	Operating_Profit_Growth_Rate	317
9	Continuous_Net_Profit_Growth_Rate	340
10	Total_Asset_Growth_Rate	0
11	Net_Value_Growth_Rate	304
12	Total_Asset_Return_Growth_Rate_Ratio	226
13	Cash_Reinvestment_perc	220
14	Current_Ratio	193
15	Quick_Ratio	190
16	Interest_Expense_Ratio	328
17	Total_debt_to_Total_net_worth	105
18	Long_term_fund_suitability_ratio_A	234
19	Net_profit_before_tax_to_Paid_in_capital	173
20	Total_Asset_Turnover	101
21	Accounts_Receivable_Turnover	281
22	Average_Collection_Days	77
23	Inventory_Turnover_Rate_times	29
24	Fixed_Assets_Turnover_Frequency	501
25	Net_Worth_Turnover_Rate_times	165
26	Operating_profit_per_person	357
27	Allocation_rate_per_person	200
28	Quick_Assets_to_Total_Assets	4
29	Cash_to_Total_Assets	163
30	Quick_Assets_to_Current_Liability	185
31	Cash_to_Current_Liability	253
32	Operating_Funds_to_Liability	219
33	Inventory_to_Working_Capital	247
34	Inventory_to_Current_Liability	129
35	Long_term_Liability_to_Current_Assets	213
36	Retained_Earnings_to_Total_Assets	208
37	Total_income_to_Total_expense	136
38	Total_expense_to_Assets	168
39	Current_Asset_Turnover_Rate	464
40	Quick_Asset_Turnover_Rate	0
41	Cash_Turnover_Rate	0
42	Fixed_Assets_to_Assets	10
43	Cash_Flow_to_Total_Assets	317
44	Cash_Flow_to_Liability	407
45	CFO_to_Assets	110
46	Cash_Flow_to_Equity	306
47	Current_Liability_to_Current_Assets	121
48	Total_assets_to_GNP_price	235
49	No_credit_Interval	396
50	Degree_of_Financial_Leverage_DFL	438
51	Interest_Coverage_Ratio_Interest_expense_to_EBIT	376
52	Equity_to_Liability	190
53	Default	220

Outliers in each column

3. Data Preparation for Modeling:

- Separate the target variable (`default` column) from the rest of the data.

4. Split Data

- Divide the data into training and testing sets.

5. Missing Values Detection and Treatment:

- Identify and handle missing values in the dataset.
- Missing value in **train** dataset
  - Cash Flow Per Share - 126
  - Total debt to Total net worth - 18
  - Cash to Total Assets - 71
  - Current Liability to Current Assets - 11
- Missing value in **test** dataset
  - Cash Flow Per Share - 41
  - Total debt to Total net worth - 3
  - Cash to Total Assets - 25

Operating_Expense_Rate	0
Research_and_development_expense_rate	0
Cash_flow_rate	0
Interest_bearing_debt_interest_rate	0
Tax_rate_A	0
Cash_Flow_Per_Share	126
Per_Share_Net_profit_before_tax_Yuan_	0
Realized_Sales_Gross_Profit_Growth_Rate	0
Operating_Profit_Growth_Rate	0
Continuous_Net_Profit_Growth_Rate	0
Total_Asset_Growth_Rate	0
Net_Value_Growth_Rate	0
Total_Asset_Return_Growth_Rate_Ratio	0
Cash_Reinvestment_perc	0
Current_Ratio	0
Quick_Ratio	0
Interest_Expense_Ratio	0
Total_debt_to_Total_net_worth	18
Long_term_fund_suitability_ratio_A	0
Net_profit_before_tax_to_Paid_in_capital	0
Total_Asset_Turnover	0
Accounts_Receivable_Turnover	0
Average_Collection_Days	0
Inventory_Turnover_Rate_times	0
Fixed_Assets_Turnover_Frequency	0
Net_Worth_Turnover_Rate_times	0
Operating_profit_per_person	0
Allocation_rate_per_person	0
Quick_Assets_to_Total_Assets	0
Cash_to_Total_Assets	71
Quick_Assets_to_Current_Liability	0
Cash_to_Current_Liability	0
Operating_Funds_to_Liability	0
Inventory_to_Working_Capital	0
Inventory_to_Current_Liability	0
Long_term_Liability_to_Current_Assets	0
Retained_Earnings_to_Total_Assets	0
Total_income_to_Total_expense	0
Total_expense_to_Assets	0
Current_Asset_Turnover_Rate	0
Quick_Asset_Turnover_Rate	0
Cash_Turnover_Rate	0
Fixed_Assets_to_Assets	0
Cash_Flow_to_Total_Assets	0
Cash_Flow_to_Liability	0
CFO_to_Assets	0
Cash_Flow_to_Equity	0
Current_Liability_to_Current_Assets	11
Total_assets_to_GNP_price	0
No_credit_Interval	0
Degree_of_Financial_Leverage_DFL	0
Interest_Coverage_Ratio_Interest_expense_to_EBIT	0
Equity_to_Liability	0

dtype: int64

Missing value in train dataset

Operating_Expense_Rate	0
Research_and_development_expense_rate	0
Cash_flow_rate	0
Interest_bearing_debt_interest_rate	0
Tax_rate_A	0
Cash_Flow_Per_Share	41
Per_Share_Net_profit_before_tax_Yuan_	0
Realized_Sales_Gross_Profit_Growth_Rate	0
Operating_Profit_Growth_Rate	0
Continuous_Net_Profit_Growth_Rate	0
Total_Asset_Growth_Rate	0
Net_Value_Growth_Rate	0
Total_Asset_Return_Growth_Rate_Ratio	0
Cash_Reinvestment_perc	0
Current_Ratio	0
Quick_Ratio	0
Interest_Expense_Ratio	0
Total_debt_to_Total_net_worth	3
Long_term_fund_suitability_ratio_A	0
Net_profit_before_tax_to_Paid_in_capital	0
Total_Asset_Turnover	0
Accounts_Receivable_Turnover	0
Average_Collection_Days	0
Inventory_Turnover_Rate_times	0
Fixed_Assets_Turnover_Frequency	0
Net_Worth_Turnover_Rate_times	0
Operating_profit_per_person	0
Allocation_rate_per_person	0
Quick_Assets_to_Total_Assets	0
Cash_to_Total_Assets	25
Quick_Assets_to_Current_Liability	0
Cash_to_Current_Liability	0
Operating_Funds_to_Liability	0
Inventory_to_Working_Capital	0
Inventory_to_Current_Liability	0
Long_term_Liability_to_Current_Assets	0
Retained_Earnings_to_Total_Assets	0
Total_income_to_Total_expense	0
Total_expense_to_Assets	0
Current_Asset_Turnover_Rate	0
Quick_Asset_Turnover_Rate	0
Cash_Turnover_Rate	0
Fixed_Assets_to_Assets	0
Cash_Flow_to_Total_Assets	0
Cash_Flow_to_Liability	0
CFO_to_Assets	0
Cash_Flow_to_Equity	0
Current_Liability_to_Current_Assets	3
Total_assets_to_GNP_price	0
No_credit_Interval	0
Degree_of_Financial_Leverage_DFL	0
Interest_Coverage_Ratio_Interest_expense_to_EBIT	0
Equity_to_Liability	0

dtype: int64

Missing value in test dataset

- Use KNN Imputer to replace missing values.

6. Scaling the Data:

- Apply `StandardScaler()` to standardize the dataset.

	Operating_Expense_Rate	Research_and_development_expense_rate	Cash_flow_rate	Interest_bearing_debt_interest_rate	Tax_rate_A	Cash_Flow_Pe
0	-0.633296	-0.396806	-0.132455	-0.128462	-0.754347	0
1	-0.633296	-0.561672	-0.934352	-0.128462	-0.754347	-1
2	-0.633296	0.361946	-0.290335	-0.128462	0.061964	-0.
3	-0.633296	-0.561672	-0.179548	-0.128462	-0.754347	-0
4	-0.633296	-0.561672	-0.123892	-0.128462	-0.754347	-0

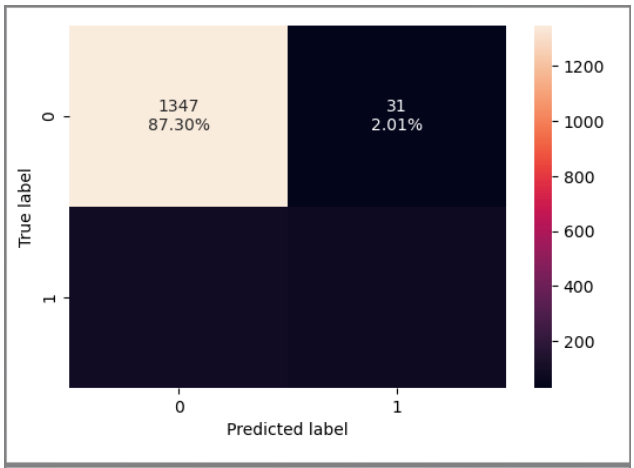
Scaled dataset for train

Total_assets_to_GNP_price	No_credit_Interval	Degree_of_Financial_Leverage_DFL	Interest_Coverage_Ratio_Interest_expense_to_EBIT	Equity_to_Liability
15.761384	-0.006331	-0.074521	-0.147943	-0.347587
-0.071478	0.033771	-0.095747	-1.144762	0.097076
-0.071478	-0.418625	-0.063956	0.098718	-0.289754
-0.071478	-0.001527	-0.074285	-0.141347	-0.381526
-0.071478	0.045728	-0.072275	-0.087539	-0.330921

Scaled dataset for test

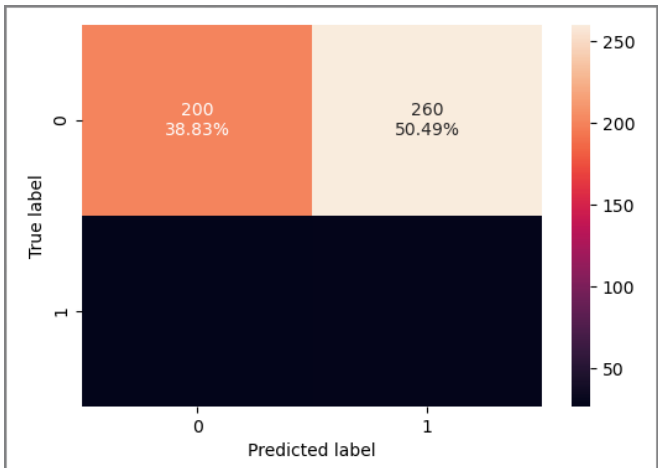
VI. Model Building

1. Logistic Regression



	Accuracy	Recall	Precision	F1
0	0.922229	0.460606	0.71028	0.558824

Training Performance



	Accuracy	Recall	Precision	F1
0	0.442718	0.509091	0.097222	0.163265

Testing Performance

Summary

```
Warning: Maximum number of iterations has been exceeded.
         Current function value: 0.193946
         Iterations: 35
                           Logit Regression Results
==============================================================================
Dep. Variable:                Default   No. Observations:                 1543
Model:                          Logit   Df Residuals:                     1489
Method:                           MLE   Df Model:                           53
Date:                Sat, 27 Jul 2024   Pseudo R-squ.:                 0.4297
Time:                        02:36:14   Log-Likelihood:                -299.26
converged:                      False   LL-Null:                       -524.71
Covariance Type:            nonrobust   LLR p-value:                1.764e-64
=============================================================================================================
                                                coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------------------------------
const                                        -7.4685   2410.787     -0.003      0.998   -4732.524    4717.587
Operating_Expense_Rate                        0.2077      0.121      1.713      0.087      -0.030       0.445
Research_and_development_expense_rate         0.3556      0.104      3.433      0.001       0.153       0.559
Cash_flow_rate                               -0.1837      1.016     -0.181      0.857      -2.175       1.808
Interest_bearing_debt_interest_rate           0.1755      0.151      1.163      0.245      -0.120       0.471
Tax_rate_A                                   -0.2580      0.174     -1.481      0.139      -0.599       0.083
Cash_Flow_Per_Share                          -0.3533      0.281     -1.260      0.208      -0.903       0.196
Per_Share_Net_profit_before_tax_Yuan_         0.2518      1.276      0.197      0.844      -2.249       2.752
Realized_Sales_Gross_Profit_Growth_Rate       0.1012      0.118      0.859      0.390      -0.130       0.332
Operating_Profit_Growth_Rate                 -0.1546      0.267     -0.579      0.563      -0.678       0.369
Continuous_Net_Profit_Growth_Rate             0.1736      0.132      1.317      0.188      -0.085       0.432
Total_Asset_Growth_Rate                      -0.0640      0.131     -0.487      0.626      -0.321       0.193
Net_Value_Growth_Rate                         0.5177   3097.843      0.000      1.000   -6071.142    6072.178
Total_Asset_Return_Growth_Rate_Ratio         -0.3299      0.361     -0.915      0.360      -1.037       0.377
Cash_Reinvestment_perc                        0.1700      0.346      0.491      0.624      -0.509       0.849
Current_Ratio                                -1.6114      0.925     -1.742      0.081      -3.424       0.201
Quick_Ratio                                  -2.7355   2.57e+04     -0.000      1.000   -5.05e+04    5.05e+04
Interest_Expense_Ratio                        0.0197      0.065      0.303      0.762      -0.107       0.147
Total_debt_to_Total_net_worth                 1.9035      0.623      3.058      0.002       0.683       3.124
Long_term_fund_suitability_ratio_A            0.1675      0.223      0.751      0.452      -0.269       0.604
Net_profit_before_tax_to_Paid_in_capital     -1.0834      1.179     -0.919      0.358      -3.394       1.227
Total_Asset_Turnover                         -0.2122      0.319     -0.666      0.506      -0.837       0.413
Accounts_Receivable_Turnover                 -1.0019      0.642     -1.560      0.119      -2.261       0.257
Average_Collection_Days                     -15.1938   2.49e+04     -0.001      1.000   -4.89e+04    4.88e+04
Inventory_Turnover_Rate_times                -0.0490      0.117     -0.420      0.675      -0.278       0.180
Fixed_Assets_Turnover_Frequency               0.1775      0.106      1.678      0.093      -0.030       0.385
Net_Worth_Turnover_Rate_times                -0.2559      0.211     -1.212      0.225      -0.670       0.158
Operating_profit_per_person                   0.0505      0.195      0.259      0.796      -0.331       0.432
Allocation_rate_per_person                  -80.4893    153.634     -0.524      0.600    -381.606     220.628
Quick_Assets_to_Total_Assets                  0.1935      0.189      1.024      0.306      -0.177       0.564
Cash_to_Total_Assets                         -0.3059      0.222     -1.380      0.168      -0.740       0.129
Quick_Assets_to_Current_Liability            -0.5860   1.49e+04  -3.92e-05      1.000   -2.93e+04    2.93e+04
Cash_to_Current_Liability                     0.0684      0.076      0.905      0.365      -0.080       0.217
Operating_Funds_to_Liability                  1.2409      0.783      1.584      0.113      -0.294       2.776
Inventory_to_Working_Capital                 -0.1714      0.158     -1.088      0.276      -0.480       0.137
Inventory_to_Current_Liability                0.1022      0.117      0.870      0.384      -0.128       0.332
Long_term_Liability_to_Current_Assets        -0.0208      0.107     -0.195      0.846      -0.230       0.188
Retained_Earnings_to_Total_Assets            -0.2111      0.207     -1.019      0.308      -0.617       0.195
Total_income_to_Total_expense                -1.4219      0.437     -3.252      0.001      -2.279      -0.565
Total_expense_to_Assets                       0.0849      0.253      0.335      0.738      -0.412       0.582
Current_Asset_Turnover_Rate                  -0.0962      0.129     -0.746      0.456      -0.349       0.157
Quick_Asset_Turnover_Rate                     0.0640      0.128      0.499      0.618      -0.188       0.316
Cash_Turnover_Rate                           -0.4286      0.130     -3.307      0.001      -0.683      -0.175
Fixed_Assets_to_Assets                       31.5359    195.727      0.161      0.872    -352.082     415.154
Cash_Flow_to_Total_Assets                     0.9901      0.270      3.668      0.000       0.461       1.519
Cash_Flow_to_Liability                       -2.7554      0.607     -4.542      0.000      -3.945      -1.566
CFO_to_Assets                                -0.3143      0.467     -0.673      0.501      -1.230       0.602
Cash_Flow_to_Equity                          -0.0344      0.085     -0.404      0.686      -0.201       0.132
Current_Liability_to_Current_Assets          -0.0863      0.121     -0.714      0.476      -0.323       0.151
Total_assets_to_GNP_price                    -0.0290      0.076     -0.384      0.701      -0.177       0.119
No_credit_Interval                            0.1051      0.079      1.326      0.185      -0.050       0.260
Degree_of_Financial_Leverage_DFL              0.0729      0.056      1.303      0.193      -0.037       0.183
Interest_Coverage_Ratio_Interest_expense_to_EBIT  0.0677  0.087      0.778      0.437      -0.103       0.238
Equity_to_Liability                          -3.0217      0.709     -4.260      0.000      -4.412      -1.632
=============================================================================================================
```

Logit Regression Results

Observation

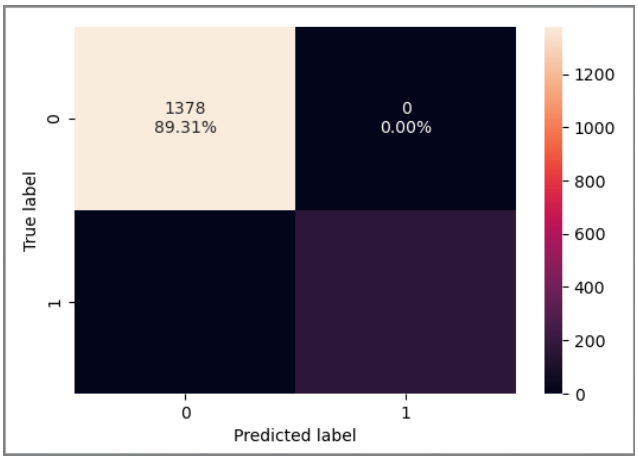
Logistic Regression Training Performance:

- **Accuracy**: 0.922229
- **Recall**: 0.460606
- **Precision**: 0.71028
- **F1 Score**: 0.558824

Logistic Regression Testing Performance:

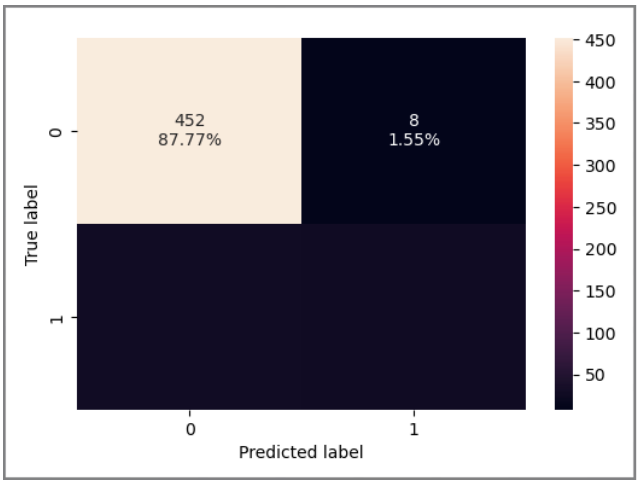
- **Accuracy**: 0.442718
- **Recall**: 0.509091
- **Precision**: 0.097222
- **F1 Score**: 0.163265

2. Random Forest



	Accuracy	Recall	Precision	F1
0	1.0	1.0	1.0	1.0

Training Performance



	Accuracy	Recall	Precision	F1
0	0.930097	0.490909	0.771429	0.6

Testing Performance

VII.Model Performance Improvement

Variance Inflation Factor (VIF) is a measure of multicollinearity in a set of multiple regression variables. A high VIF indicates that the associated independent variable is highly collinear with the other variables in the model. Here are the steps to calculate VIF using statsmodels and pandas:

1. Fit the OLS model: We fit an Ordinary Least Squares (OLS) model to each independent variable against all the other independent variables.
2. Calculate VIF: The VIF for each variable is calculated using the formula:

VIF =1/{1 - R^2}

where R^2 is the coefficient of determination of the regression of that variable against all the other variables.

The output high_vif_columns contains a list of variables that have a Variance Inflation Factor (VIF) greater than or equal to 5. This indicates that these variables are highly collinear with other independent variables in the dataset.

Dropping the columns which have VIF > 5:

1. **Cash_flow_rate**: This variable measures the rate of cash flow relative to other financial metrics.
2. **Per_Share_Net_profit_before_tax_Yuan_**: This variable represents the net profit before tax per share.
3. **Cash_Reinvestment_perc**: This variable represents the percentage of cash reinvested back into the business.
4. **Net_profit_before_tax_to_Paid_in_capital**: This variable measures the ratio of net profit before tax to the paid-in capital.
5. **Total_Asset_Turnover**: This variable measures the efficiency of a company's use of its assets to generate sales.
6. **Operating_Funds_to_Liability**: This variable measures the ratio of operating funds to liabilities.
7. **CFO_to_Assets**: This variable represents the ratio of cash flow from operations (CFO) to total assets.

Model Performance Improvement - Logistic Regression

Shape of the data after dropping columns which had vif > 5:

```
(1543, 46)
```

Shape of scaled train dataset

```
(515, 46)
```

Shape of scaled test dataset

```
         Current function value: 0.201099
         Iterations: 100
         Function evaluations: 101
         Gradient evaluations: 101
                           Logit Regression Results
==============================================================================
Dep. Variable:                Default   No. Observations:                 1543
Model:                          Logit   Df Residuals:                     1496
Method:                          MLE   Df Model:                           46
Date:               Sat, 27 Jul 2024   Pseudo R-squ.:                 0.4086
Time:                       02:36:16   Log-Likelihood:                -310.30
converged:                     False   LL-Null:                       -524.71
Covariance Type:           nonrobust   LLR p-value:                 1.462e-63
=============================================================================================================
                                                    coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------------------------------
const                                           -4.4843      0.730     -6.144      0.000      -5.915      -3.054
Operating_Expense_Rate                           0.1938      0.117      1.658      0.097      -0.035       0.423
Research_and_development_expense_rate            0.3735      0.099      3.759      0.000       0.179       0.568
Interest_bearing_debt_interest_rate              0.1971      0.152      1.301      0.193      -0.100       0.494
Tax_rate_A                                      -0.3721      0.180     -2.066      0.039      -0.725      -0.019
Cash_Flow_Per_Share                             -0.1821      0.139     -1.312      0.189      -0.454       0.090
Realized_Sales_Gross_Profit_Growth_Rate          0.1174      0.116      1.012      0.312      -0.110       0.345
Operating_Profit_Growth_Rate                    -0.2278      0.300     -0.759      0.448      -0.816       0.360
Continuous_Net_Profit_Growth_Rate                0.1505      0.123      1.227      0.220      -0.090       0.391
Total_Asset_Growth_Rate                         -0.0667      0.126     -0.531      0.595      -0.313       0.179
Net_Value_Growth_Rate                            0.1937      3.545      0.055      0.956      -6.754       7.142
Total_Asset_Return_Growth_Rate_Ratio            -0.7704      0.373     -2.063      0.039      -1.502      -0.039
Current_Ratio                                   -1.9304      0.643     -3.000      0.003      -3.192      -0.669
Quick_Ratio                                     -0.7513      7.542     -0.100      0.921     -15.533      14.030
Interest_Expense_Ratio                           0.0258      0.065      0.396      0.692      -0.102       0.154
Total_debt_to_Total_net_worth                    2.8590      0.569      5.021      0.000       1.743       3.975
Long_term_fund_suitability_ratio_A              -0.2377      0.253     -0.938      0.348      -0.734       0.259
Accounts_Receivable_Turnover                    -1.0112      0.619     -1.634      0.102      -2.224       0.202
Average_Collection_Days                         -0.3428      1.827     -0.188      0.851      -3.923       3.237
Inventory_Turnover_Rate_times                   -0.0581      0.114     -0.511      0.609      -0.281       0.165
Fixed_Assets_Turnover_Frequency                  0.1469      0.104      1.417      0.157      -0.056       0.350
Net_Worth_Turnover_Rate_times                   -0.1894      0.129     -1.472      0.141      -0.442       0.063
Operating_profit_per_person                      0.0322      0.187      0.172      0.864      -0.335       0.400
Allocation_rate_per_person                      -0.0413      1.387     -0.030      0.976      -2.759       2.677
Quick_Assets_to_Total_Assets                     0.0429      0.161      0.266      0.790      -0.273       0.359
Cash_to_Total_Assets                            -0.3529      0.212     -1.666      0.096      -0.768       0.062
Quick_Assets_to_Current_Liability               -0.1175      1.661     -0.071      0.944      -3.372       3.137
Cash_to_Current_Liability                        0.0739      0.075      0.992      0.321      -0.072       0.220
Inventory_to_Working_Capital                    -0.1518      0.143     -1.058      0.290      -0.433       0.129
Inventory_to_Current_Liability                   0.0899      0.124      0.724      0.469      -0.153       0.333
Long_term_Liability_to_Current_Assets           -0.0475      0.108     -0.439      0.661      -0.259       0.165
Retained_Earnings_to_Total_Assets               -0.2175      0.179     -1.215      0.224      -0.568       0.133
Total_income_to_Total_expense                   -2.0469      0.354     -5.783      0.000      -2.741      -1.353
Total_expense_to_Assets                          0.1727      0.206      0.837      0.403      -0.232       0.577
Current_Asset_Turnover_Rate                     -0.1299      0.120     -1.086      0.277      -0.364       0.104
Quick_Asset_Turnover_Rate                        0.0295      0.120      0.247      0.805      -0.205       0.264
Cash_Turnover_Rate                              -0.3696      0.123     -3.015      0.003      -0.610      -0.129
Fixed_Assets_to_Assets                           0.5126     17.419      0.029      0.977     -33.628      34.653
Cash_Flow_to_Total_Assets                        0.9891      0.232      4.269      0.000       0.535       1.443
Cash_Flow_to_Liability                          -2.2925      0.452     -5.077      0.000      -3.177      -1.408
Cash_Flow_to_Equity                              0.0059      0.073      0.082      0.935      -0.136       0.148
Current_Liability_to_Current_Assets             -0.0785      0.115     -0.685      0.494      -0.303       0.146
Total_assets_to_GNP_price                       -0.0347      0.075     -0.463      0.643      -0.182       0.112
No_credit_Interval                               0.1180      0.080      1.476      0.140      -0.039       0.275
Degree_of_Financial_Leverage_DFL                 0.0772      0.055      1.403      0.161      -0.031       0.185
Interest_Coverage_Ratio_Interest_expense_to_EBIT 0.0597      0.082      0.727      0.467      -0.101       0.221
Equity_to_Liability                             -2.9001      0.554     -5.236      0.000      -3.986      -1.814
=============================================================================================================
```

Logistic Regression Result

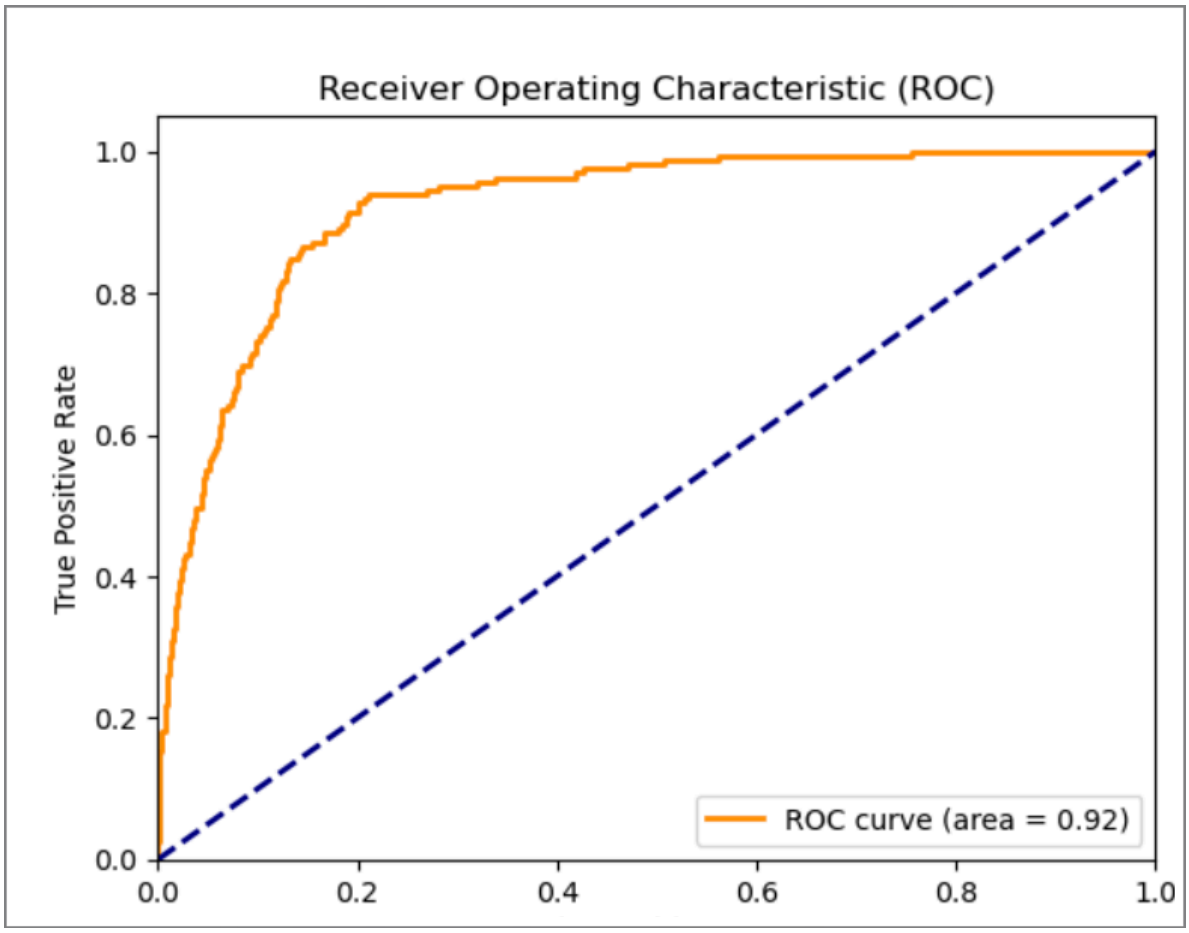
0.084

Optimal threshold value

Model Summary:

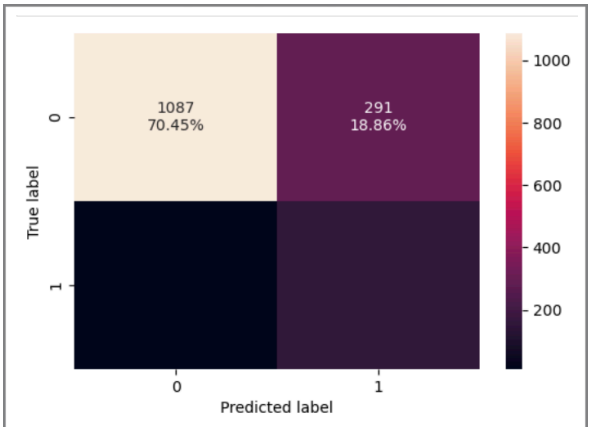
- **Convergence**: The model did not converge, indicating potential issues with model specification or data.
- **Log-Likelihood**: -310.30, indicating the fit of the model compared to a null model.
- **Pseudo R-squared**: 0.4086, meaning approximately 40.86% of the variance in the dependent variable is explained by the independent variables.

2. **Significant Variables**:
   - Research_and_development_expense_rate (p-value < 0.01): Positively associated with Default.
   - Total_debt_to_Total_net_worth (p-value < 0.01): Positively associated with Default.
   - Total_income_to_Total_expense (p-value < 0.01): Negatively associated with Default.
   - Cash_Flow_to_Total_Assets (p-value < 0.01): Positively associated with Default.
   - Cash_Flow_to_Liability (p-value < 0.01): Negatively associated with Default.
   - Equity_to_Liability (p-value < 0.01): Negatively associated with Default.
3. **Variables with High VIF**:
   - Cash_flow_rate: Potential multicollinearity issue.
   - Per_Share_Net_profit_before_tax_Yuan_: Potential multicollinearity issue.
   - Cash_Reinvestment_perc: Potential multicollinearity issue.
   - Net_profit_before_tax_to_Paid_in_capital: Potential multicollinearity issue.
   - Total_Asset_Turnover: Potential multicollinearity issue.
   - Operating_Funds_to_Liability: Potential multicollinearity issue.
   - CFO_to_Assets: Potential multicollinearity issue.
4. **Other Observations**:
   - Operating_Expense_Rate: Close to significance with p-value = 0.097.
   - Tax_rate_A: Significant (p-value < 0.05) and negatively associated with Default.
   - Current_Ratio: Significant (p-value < 0.01) and negatively associated with Default.



ROC Curve

An ROC (Receiver Operating Characteristic) curve with an area under the curve (AUC) of 0.92 indicates that the model has excellent discriminatory power, meaning it can effectively differentiate between the positive class (Default) and the negative class (Non-Default). This high AUC-ROC value signifies that the model is very good at correctly ranking positive instances higher than negative ones across various threshold settings. Essentially, there is a 92% chance that the model will assign a higher probability to a randomly chosen positive instance than to a randomly chosen negative instance, demonstrating strong overall model performance and reliability for decision-making purposes.



	Accuracy	Recall	Precision	F1
0	0.804277	0.933333	0.346067	0.504918

Logistic Regression Performance - Training Set

```
Confusion Matrix:
[[1087  291]
 [  11  154]]
True Negatives (TN): 1087
False Positives (FP): 291
False Negatives (FN): 11
True Positives (TP): 154
True Negatives (TN): 1087 (70.45%)
False Positives (FP): 291 (18.86%)
False Negatives (FN): 11 (0.71%)
True Positives (TP): 154 (9.98%)
```

Confusion Matrix

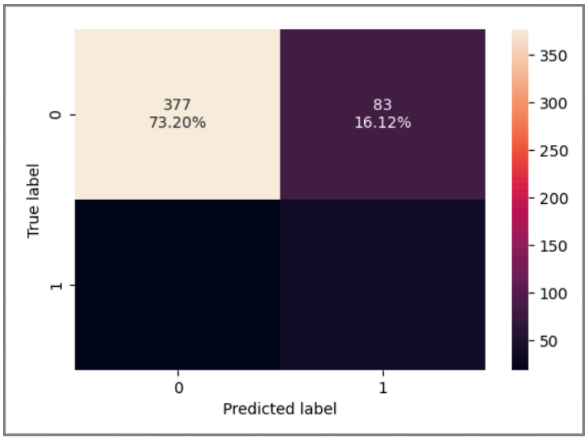
1. **True Negatives (TN)**: 1087 (70.45%)
   - **Interpretation**: The model correctly predicted 70.45% of the actual negative cases. This high percentage indicates that the model performs well in predicting negative cases.
2. **False Positives (FP)**: 291 (18.86%)
   - **Interpretation**: The model incorrectly predicted 18.86% of the actual negative cases as positive. This suggests that there is a moderate amount of misclassification in the negative class.
3. **False Negatives (FN)**: 11 (0.71%)
   - **Interpretation**: The model missed 0.71% of the actual positive cases, predicting them as negative. This is a very low percentage, indicating that the model is good at identifying positive cases.
4. **True Positives (TP)**: 154 (9.98%)
   - **Interpretation**: The model correctly predicted 9.98% of the actual positive cases. This percentage is lower compared to TN, reflecting that the number of positive cases is much smaller than the number of negative cases.

	Accuracy	Recall	Precision	F1
0	0.804277	0.933333	0.346067	0.504918

Model Performance

1. **High Recall (93.33%)**:
   - The model is very effective at identifying positive cases, correctly detecting 93.33% of actual positives. This indicates low false negative rates and good performance in recognising positive instances.
2. **Low Precision (34.60%)**:
   - The model has a low precision, meaning that only 34.60% of the cases predicted as positive are actually positive. This suggests a high rate of false positives, where many negative cases are incorrectly classified as positive.
3. **Moderate Accuracy (80.43%)**:
   - With an accuracy of 80.43%, the model correctly classifies 80.43% of all samples. However, accuracy alone may be misleading in the case of imbalanced datasets, where the model might be biased towards the majority class.
4. **Balanced F1 Score (50.50%)**:
   - The F1 score of 50.50% reflects the balance between precision and recall. While the recall is high, the low precision pulls the F1 score down, indicating that the model's performance is compromised by a significant number of false positives.

Logistic Regression Performance - Test Set




```
Confusion Matrix:
[[377  83]
 [ 19  36]]
True Negatives (TN): 377
False Positives (FP): 83
False Negatives (FN): 19
True Positives (TP): 36
True Negatives (TN): 377 (73.20%)
False Positives (FP): 83 (16.12%)
False Negatives (FN): 19 (3.69%)
True Positives (TP): 36 (6.99%)
```

	Accuracy	Recall	Precision	F1
0	0.801942	0.654545	0.302521	0.413793

Test Set

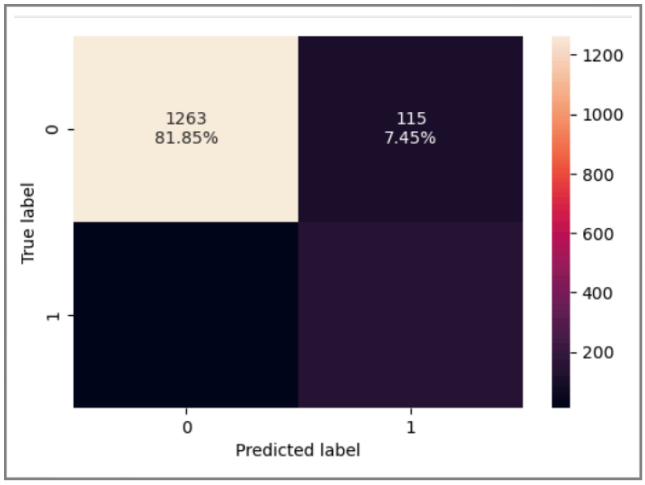
1. **High Accuracy**:
    - The model has a good accuracy of 80.19%, showing that it performs well overall in terms of correct classifications. However, accuracy alone might be misleading if the dataset is imbalanced.
2. **Moderate Recall**:
    - A recall of 65.45% indicates the model is relatively effective at identifying positive cases but still misses a significant portion of them.
3. **Low Precision**:
    - With a precision of 30.25%, the model often incorrectly predicts negatives as positives, resulting in many false positives. This suggests that the model may need adjustments to improve its specificity.
4. **Balanced F1 Score**:
    - The F1 score of 41.38% shows a trade-off between precision and recall. While the recall is higher, the low precision affects the overall performance. Improving precision would help in increasing the F1 score.

Model Performance Improvement - Random Forest

```
Parameters used in the Random Forest Classifier:
bootstrap: True
ccp_alpha: 0.0
class_weight: balanced
criterion: gini
max_depth: 5
max_features: sqrt
max_leaf_nodes: None
max_samples: None
min_impurity_decrease: 0.0
min_samples_leaf: 7
min_samples_split: 2
min_weight_fraction_leaf: 0.0
n_estimators: 200
n_jobs: None
oob_score: False
random_state: 42
verbose: 0
warm_start: False
```

Parameters used in the Random Forest Classifier

Random Forest Performance - Training Set



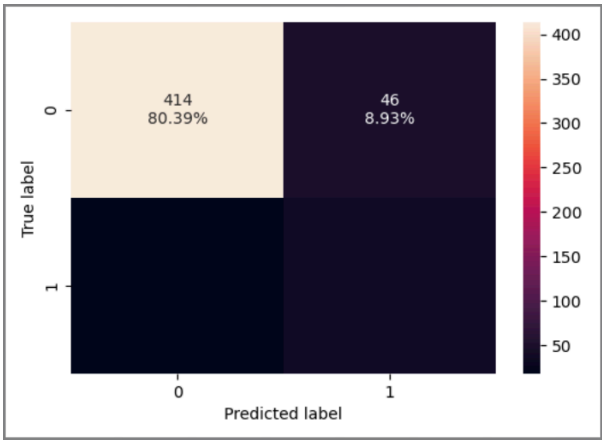

```
Confusion Matrix:
[[1263  115]
 [  13  152]]
True Negatives (TN): 1263
False Positives (FP): 115
False Negatives (FN): 13
True Positives (TP): 152
True Negatives (TN): 1263 (81.85%)
False Positives (FP): 115 (7.45%)
False Negatives (FN): 13 (0.84%)
True Positives (TP): 152 (9.85%)
```

	Accuracy	Recall	Precision	F1
0	0.917045	0.921212	0.569288	0.703704

Training set Random Forest

The Random Forest model demonstrates strong performance on the training set with an accuracy of 91.70%, indicating that it correctly classified a significant majority of the samples. It achieves an excellent recall of 92.12%, showing its effectiveness in identifying positive cases and minimizing false negatives. However, the model's precision is moderate at 56.93%, which means that about 43% of its positive predictions are actually false positives. This suggests that while the model is adept at capturing most positive cases, it also misclassifies some negatives as positives. The F1 score of 70.37% reflects a balanced performance, combining both precision and recall. Overall, the Random Forest model performs well in training, but its effectiveness should be further assessed on a test set to confirm its generalization and avoid overfitting.

Random Forest Performance - Testing Set




```
Confusion Matrix:
[[414  46]
 [ 18  37]]
True Negatives (TN): 414
False Positives (FP): 46
False Negatives (FN): 18
True Positives (TP): 37
True Negatives (TN): 414 (80.39%)
False Positives (FP): 46 (8.93%)
False Negatives (FN): 18 (3.50%)
True Positives (TP): 37 (7.18%)
```

	Accuracy	Recall	Precision	F1
0	0.875728	0.672727	0.445783	0.536232

Random Forest Performance - Test Set

The Random Forest model's performance on the testing set shows an accuracy of 87.57%, indicating strong overall classification ability. The recall is 67.27%, demonstrating the model's effectiveness in identifying positive cases, though there is room for improvement. Precision stands at 44.58%, suggesting that a notable proportion of the model's positive predictions are false positives, which points to a need for enhanced specificity. The F1 score of 53.62% reflects a balanced performance between precision and recall, indicating that while the model performs well, there is potential for optimization in terms of reducing false positives and improving overall prediction reliability.

VIII.Model Comparison

Training performance comparison:

	Logistic Regression	Tuned Logistic Regression	Random Forest	Tuned Random Forest
Accuracy	0.922229	0.804277	1.0	0.917045
Recall	0.460606	0.933333	1.0	0.921212
Precision	0.710280	0.346067	1.0	0.569288
F1	0.558824	0.504918	1.0	0.703704

Training performance comparison

Testing performance comparison:

	Logistic Regression	Tuned Logistic Regression	Random Forest	Tuned Random Forest
Accuracy	0.442718	0.801942	0.930097	0.875728
Recall	0.509091	0.654545	0.490909	0.672727
Precision	0.097222	0.302521	0.771429	0.445783
F1	0.163265	0.413793	0.600000	0.536232

Testing performance comparison

When comparing the models, the Random Forest with tuning exhibits the highest accuracy of 91.70% on the training set and 87.57% on the testing set, showing robust performance across both datasets. It also demonstrates a high recall of 92.12% on the training set and 67.27% on the testing set, indicating strong ability in detecting positive cases, although recall decreases on the test set. Precision for the tuned Random Forest is moderate, at 56.93% in training and 44.58% in testing, reflecting a balance between false positives and true positives. The F1 score of 70.37% on the training set and 53.62% on the testing set highlights its overall effectiveness in handling both precision and recall.

In contrast, Logistic Regression performs with lower accuracy on the testing set (44.27%) and the tuned version shows a significant improvement (80.19%), but still lags behind the Random Forest. The recall for Logistic Regression is high on the training set (93.33%) but falls to 65.45% on the testing set, suggesting a drop in its ability to detect positives in unseen data. Precision is notably low for Logistic Regression (30.25% on the testing set), leading to a low F1 score of 41.38%, reflecting challenges in balancing precision and recall effectively.

Overall, the Random Forest with tuning emerges as the superior model due to its high accuracy, balanced precision, and recall, making it the preferred choice for both training and testing performance.

Final Model Selection

Based on the performance metrics for both training and testing sets, the **Tuned Random Forest** is the final model selection. It excels with a high accuracy of 91.70% on the training set and 87.57% on the testing set, demonstrating robust performance across both datasets. It also achieves a high recall of 92.12% on the training set and 67.27% on the testing set, effectively identifying positive cases while maintaining a balance with a moderate precision of 56.93% in training and 44.58% in testing. The F1 score of 70.37% on the training set and 53.62% on the testing set further confirms its balanced approach in handling precision and recall.

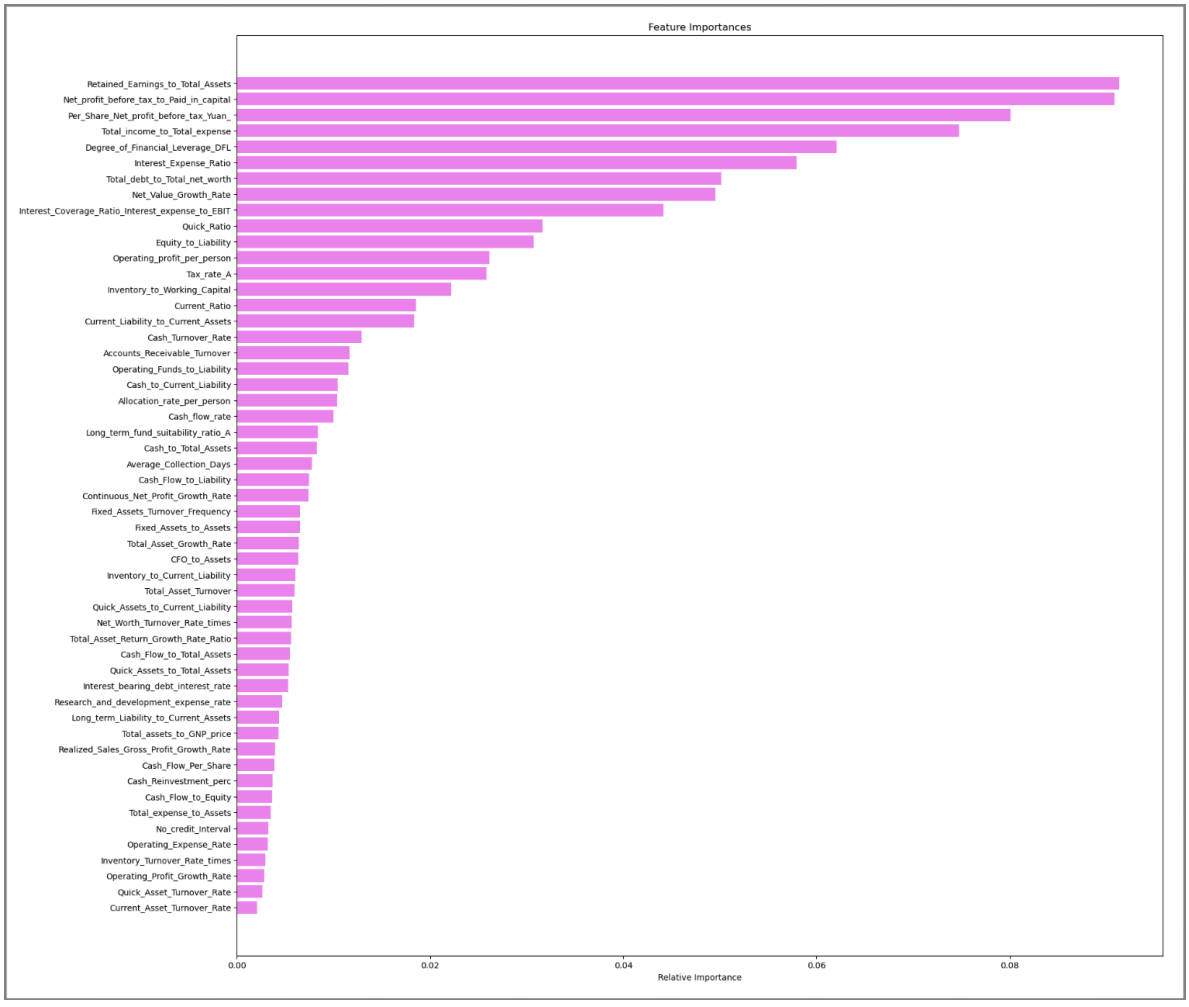
In comparison, the Logistic Regression models, including the tuned version, show lower performance in terms of accuracy, precision, and F1 score, especially on the testing set. Therefore, the Tuned Random Forest model is chosen for its superior overall performance and balance between detecting positive cases and minimizing false positives.

Feature Importance

Feature importance helps identify which features contribute the most to the predictions of a model. In the context of Random Forest, feature importance is typically derived from the average of the decrease in impurity (e.g., Gini impurity or entropy) brought by each feature across all trees in the forest.

To Determine Feature Importance in Random Forest:

1. *Train the Model*
2. *Extract Feature Importance*
3. *Interpret Feature Importance*
    - *Higher values indicate that the feature is more important in making predictions.*



Feature Importances

The feature importance analysis from the Random Forest model reveals significant insights into the predictors influencing the model's decisions. The top features include "Retained_Earnings_to_Total_Assets" with the highest importance score of 0.091, followed closely by "Net_profit_before_tax_to_Paid_in_capital" and "Per_Share_Net_profit_before_tax_Yuan_" with scores of 0.090 and 0.080, respectively. These features have the most substantial impact on the model's

predictions, reflecting their critical role in explaining the variance in the target variable.

In contrast, features such as "Current_Asset_Turnover_Rate" and "Operating_Profit_Growth_Rate" show the lowest importance scores, indicating their minimal contribution to the model's predictive power. Overall, the analysis highlights that financial ratios and profitability measures are more influential in the model compared to other less impactful metrics. This understanding can guide feature selection and refinement efforts, focusing on the most impactful features to enhance model performance and interpretability.

	Feature	Importance
36	Retained_Earnings_to_Total_Assets	0.091311
19	Net_profit_before_tax_to_Paid_in_capital	0.090854
6	Per_Share_Net_profit_before_tax_Yuan_	0.080075
37	Total_income_to_Total_expense	0.074773
50	Degree_of_Financial_Leverage_DFL	0.062086
16	Interest_Expense_Ratio	0.057904
17	Total_debt_to_Total_net_worth	0.050149
11	Net_Value_Growth_Rate	0.049492
51	Interest_Coverage_Ratio_Interest_expense_to_EBIT	0.044152
15	Quick_Ratio	0.031621
52	Equity_to_Liability	0.030703
26	Operating_profit_per_person	0.026123
4	Tax_rate_A	0.025837
33	Inventory_to_Working_Capital	0.022185
14	Current_Ratio	0.018547
47	Current_Liability_to_Current_Assets	0.018371
41	Cash_Turnover_Rate	0.012875
21	Accounts_Receivable_Turnover	0.011690
32	Operating_Funds_to_Liability	0.011520
31	Cash_to_Current_Liability	0.010455
27	Allocation_rate_per_person	0.010377
2	Cash_flow_rate	0.010003
18	Long_term_fund_suitability_ratio_A	0.008368
29	Cash_to_Total_Assets	0.008283
22	Average_Collection_Days	0.007794
44	Cash_Flow_to_Liability	0.007452
9	Continuous_Net_Profit_Growth_Rate	0.007385
24	Fixed_Assets_Turnover_Frequency	0.006548
42	Fixed_Assets_to_Assets	0.006531
10	Total_Asset_Growth_Rate	0.006375
45	CFO_to_Assets	0.006318
34	Inventory_to_Current_Liability	0.006019
20	Total_Asset_Turnover	0.005944
30	Quick_Assets_to_Current_Liability	0.005741
25	Net_Worth_Turnover_Rate_times	0.005637
12	Total_Asset_Return_Growth_Rate_Ratio	0.005570
43	Cash_Flow_to_Total_Assets	0.005451
28	Quick_Assets_to_Total_Assets	0.005329
3	Interest_bearing_debt_interest_rate	0.005272
1	Research_and_development_expense_rate	0.004657
35	Long_term_Liability_to_Current_Assets	0.004383
48	Total_assets_to_GNP_price	0.004297
7	Realized_Sales_Gross_Profit_Growth_Rate	0.003946
5	Cash_Flow_Per_Share	0.003877
13	Cash_Reinvestment_perc	0.003664
46	Cash_Flow_to_Equity	0.003639
38	Total_expense_to_Assets	0.003516
49	No_credit_Interval	0.003247
0	Operating_Expense_Rate	0.003214
23	Inventory_Turnover_Rate_times	0.002953
8	Operating_Profit_Growth_Rate	0.002793
40	Quick_Asset_Turnover_Rate	0.002622
39	Current_Asset_Turnover_Rate	0.002075

Feature name and their importances

IX. Actionable Insights & Recommendations

Actionable Insights:

1. **High-Impact Features**:
   - **Key Features**: “Retained_Earnings_to_Total_Assets”, “Net_profit_before_tax_to_Paid_in_capital”, and “Per_Share_Net_profit_before_tax_Yuan_” are the most influential features in the model. These features should be closely examined and utilized to understand their impact on the target variable.
2. **Low-Impact Features**:
   - **Minimal Influence**: Features such as “Current_Asset_Turnover_Rate” and “Operating_Profit_Growth_Rate” have low importance scores. These features contribute less to the model’s predictions and might be candidates for exclusion or further investigation to understand why they are less impactful.
3. **Performance Discrepancies**:
   - **Training vs. Testing**: The Random Forest model shows strong performance on the training set but a slight drop in testing accuracy. This indicates that while the model is well-calibrated, there might be room for improvement in generalizing to new data.

The analysis indicates that the company faces a heightened risk of default over the next two quarters, primarily due to a low equity-to-liability ratio and the presence of significant financial outliers. While factors such as constant net income, moderate asset growth, and operating and R&D expense rates also contribute to the risk context, the equity-to-liability ratio is the most critical indicator of potential default risk.

To effectively mitigate the risk of default, the company should consider the following refined strategies:

Enhance Equity Position:

- Strengthen Capital Base: Pursue new equity financing options or reinvest retained earnings to improve the equity-to-liability ratio. Converting existing debt

to equity can also help reduce financial leverage and strengthen the company's balance sheet.

Optimize Debt Management:

- Restructure Debt: Engage in proactive negotiations with creditors to restructure existing debt. Explore options such as extending repayment periods, reducing interest rates, or converting debt to equity to alleviate immediate financial pressures and improve cash flow.

Implement Rigorous Cost Control:

- Streamline Expenses: Conduct a comprehensive review of operating expenses to identify cost-saving opportunities. Focus on optimizing essential expenditures and eliminating inefficiencies to enhance overall cost management.

Drive Revenue Growth:

- Expand Market Reach: Develop and execute strategies to increase sales through targeted marketing campaigns, diversification of product offerings, or entering new markets. Expanding revenue streams can provide a more stable financial foundation.

Strengthen Liquidity Management:

- Optimize Cash Flow: Improve liquidity by enhancing cash flow management practices, deferring non-essential expenditures, and optimizing working capital. Ensure the company maintains sufficient liquidity to meet short-term obligations.

Invest in Strategic Innovation:

- Foster Growth: Invest in innovation to drive long-term growth and competitiveness. Explore cost-effective innovation strategies, such as partnerships, joint ventures, or accessing grants, to support research and development efforts.

Establish Robust Risk Monitoring:

- Continuous Assessment: Implement a comprehensive risk monitoring system to regularly assess financial health and identify early warning signs of potential issues. Continuously review key financial ratios and metrics to proactively address risks and ensure timely interventions.

By adopting these strategies, the company can strengthen its financial stability, improve its ability to meet obligations, and position itself for sustained growth and resilience against default risk.