

| | | | | |
|----|---|------|----------|---------|
| 9 | Non-industry income and expenditure/revenue | 6819 | non-null | float64 |
| 10 | Continuous interest rate (after tax) | 6819 | non-null | float64 |
| 11 | Operating Expense Rate | 6819 | non-null | float64 |
| 12 | Research and development expense rate | 6819 | non-null | float64 |
| 13 | Cash flow rate | 6819 | non-null | float64 |
| 14 | Interest-bearing debt interest rate | 6819 | non-null | float64 |
| 15 | Tax rate (A) | 6819 | non-null | float64 |
| 16 | Net Value Per Share (B) | 6819 | non-null | float64 |
| 17 | Net Value Per Share (A) | 6819 | non-null | float64 |
| 18 | Net Value Per Share (C) | 6819 | non-null | float64 |
| 19 | Persistent EPS in the Last Four Seasons | 6819 | non-null | float64 |
| 20 | Cash Flow Per Share | 6819 | non-null | float64 |
| 21 | Revenue Per Share (Yuan ¥) | 6819 | non-null | float64 |
| 22 | Operating Profit Per Share (Yuan ¥) | 6819 | non-null | float64 |
| 23 | Per Share Net profit before tax (Yuan ¥) | 6819 | non-null | float64 |
| 24 | Realized Sales Gross Profit Growth Rate | 6819 | non-null | float64 |
| 25 | Operating Profit Growth Rate | 6819 | non-null | float64 |
| 26 | After-tax Net Profit Growth Rate | 6819 | non-null | float64 |
| 27 | Regular Net Profit Growth Rate | 6819 | non-null | float64 |
| 28 | Continuous Net Profit Growth Rate | 6819 | non-null | float64 |
| 29 | Total Asset Growth Rate | 6819 | non-null | float64 |
| 30 | Net Value Growth Rate | 6819 | non-null | float64 |
| 31 | Total Asset Return Growth Rate Ratio | 6819 | non-null | float64 |
| 32 | Cash Reinvestment % | 6819 | non-null | float64 |
| 33 | Current Ratio | 6819 | non-null | float64 |
| 34 | Quick Ratio | 6819 | non-null | float64 |
| 35 | Interest Expense Ratio | 6819 | non-null | float64 |
| 36 | Total debt/Total net worth | 6819 | non-null | float64 |
| 37 | Debt ratio % | 6819 | non-null | float64 |
| 38 | Net worth/Assets | 6819 | non-null | float64 |
| 39 | Long-term fund suitability ratio (A) | 6819 | non-null | float64 |
| 40 | Borrowing dependency | 6819 | non-null | float64 |
| 41 | Contingent liabilities/Net worth | 6819 | non-null | float64 |
| 42 | Operating profit/Paid-in capital | 6819 | non-null | float64 |
| 43 | Net profit before tax/Paid-in capital | 6819 | non-null | float64 |
| 44 | Inventory and accounts receivable/Net value | 6819 | non-null | float64 |
| 45 | Total Asset Turnover | 6819 | non-null | float64 |
| 46 | Accounts Receivable Turnover | 6819 | non-null | float64 |
| 47 | Average Collection Days | 6819 | non-null | float64 |
| 48 | Inventory Turnover Rate (times) | 6819 | non-null | float64 |
| 49 | Fixed Assets Turnover Frequency | 6819 | non-null | float64 |
| 50 | Net Worth Turnover Rate (times) | 6819 | non-null | float64 |
| 51 | Revenue per person | 6819 | non-null | float64 |
| 52 | Operating profit per person | 6819 | non-null | float64 |
| 53 | Allocation rate per person | 6819 | non-null | float64 |
| 54 | Working Capital to Total Assets | 6819 | non-null | float64 |
| 55 | Quick Assets/Total Assets | 6819 | non-null | float64 |
| 56 | Current Assets/Total Assets | 6819 | non-null | float64 |
| 57 | Cash/Total Assets | 6819 | non-null | float64 |
| 58 | Quick Assets/Current Liability | 6819 | non-null | float64 |
| 59 | Cash/Current Liability | 6819 | non-null | float64 |
| 60 | Current Liability to Assets | 6819 | non-null | float64 |
| 61 | Operating Funds to Liability | 6819 | non-null | float64 |
| 62 | Inventory/Working Capital | 6819 | non-null | float64 |
| 63 | Inventory/Current Liability | 6819 | non-null | float64 |
| 64 | Current Liabilities/Liability | 6819 | non-null | float64 |
| 65 | Working Capital/Equity | 6819 | non-null | float64 |
| 66 | Current Liabilities/Equity | 6819 | non-null | float64 |
| 67 | Long-term Liability to Current Assets | 6819 | non-null | float64 |
| 68 | Retained Earnings to Total Assets | 6819 | non-null | float64 |
| 69 | Total income/Total expense | 6819 | non-null | float64 |
| 70 | Total expense/Assets | 6819 | non-null | float64 |
| 71 | Current Asset Turnover Rate | 6819 | non-null | float64 |
| 72 | Quick Asset Turnover Rate | 6819 | non-null | float64 |
| 73 | Working capitcal Turnover Rate | 6819 | non-null | float64 |
| 74 | Cash Turnover Rate | 6819 | non-null | float64 |
| 75 | Cash Flow to Sales | 6819 | non-null | float64 |
| 76 | Fixed Assets to Assets | 6819 | non-null | float64 |
| 77 | Current Liability to Liability | 6819 | non-null | float64 |
| 78 | Current Liability to Equity | 6819 | non-null | float64 |
| 79 | Equityto Long-term Liability | 6819 | non-null | float64 |
| 80 | Cash Flow to Total Assets | 6819 | non-null | float64 |

```

81 Cash Flow to Liability 6819 non-null float64
82 CFO to Assets 6819 non-null float64
83 Cash Flow to Equity 6819 non-null float64
84 Current Liability to Current Assets 6819 non-null float64
85 Liability-Assets Flag 6819 non-null int64
86 Net Income to Total Assets 6819 non-null float64
87 Total assets to GNP price 6819 non-null float64
88 No-credit Interval 6819 non-null float64
89 Gross Profit to Sales 6819 non-null float64
90 Net Income to Stockholder's Equity 6819 non-null float64
91 Liability to Equity 6819 non-null float64
92 Degree of Financial Leverage (DFL) 6819 non-null float64
93 Interest Coverage Ratio (Interest expense to EBIT) 6819 non-null float64
94 Net Income Flag 6819 non-null int64
95 Equity to Liability 6819 non-null float64
dtypes: float64(93), int64(3)
memory usage: 5.0 MB

```

In [153]:

```
data.shape
```

Out[153]:

```
(6819, 96)
```

In [154]:

```
data.head(5)
```

Out[154]:

| | Bankrupt? | ROA(C) before interest and depreciation before interest | ROA(A) before interest and % after tax | ROA(B) before interest and depreciation after tax | Operating Gross Margin | Realized Sales Gross Margin | Operating Profit Rate | Pre-tax net Interest Rate | After-tax net Interest Rate | Non-industry income and expenditure/revenue |
|---|-----------|--|--|---|------------------------------|--------------------------------------|-----------------------------|------------------------------------|--------------------------------------|---|
| 0 | 1 | 0.370594 | 0.424389 | 0.405750 | 0.601457 | 0.601457 | 0.998969 | 0.796887 | 0.808809 | 0.302646 |
| 1 | 1 | 0.464291 | 0.538214 | 0.516730 | 0.610235 | 0.610235 | 0.998946 | 0.797380 | 0.809301 | 0.303556 |
| 2 | 1 | 0.426071 | 0.499019 | 0.472295 | 0.601450 | 0.601364 | 0.998857 | 0.796403 | 0.808388 | 0.302035 |
| 3 | 1 | 0.399844 | 0.451265 | 0.457733 | 0.583541 | 0.583541 | 0.998700 | 0.796967 | 0.808966 | 0.303350 |
| 4 | 1 | 0.465022 | 0.538432 | 0.522298 | 0.598783 | 0.598783 | 0.998973 | 0.797366 | 0.809304 | 0.303475 |

5 rows x 96 columns



In [155]:

```
data.isnull().sum().any()
```

Out[155]:

```
False
```

In [156]:

```
data.tail()
```

Out[156]:

| | Bankrupt? | ROA(C) before interest and depreciation before interest | ROA(A) before interest and % after tax | ROA(B) before interest and depreciation after tax | Operating Gross Margin | Realized Sales Gross Margin | Operating Profit Rate | Pre-tax net Interest Rate | After-tax net Interest Rate | Non-indust income an expenditure/revenu |
|------|-----------|--|--|---|------------------------------|--------------------------------------|-----------------------------|------------------------------------|--------------------------------------|---|
| 6814 | 0 | 0.493687 | 0.539468 | 0.543230 | 0.604455 | 0.604462 | 0.998992 | 0.797409 | 0.809331 | 0.30351 |

| | | | | | | | | | | | | |
|------|---|---|----------|---|----------|---|----------|----------------|----------|-------------|---------------|---------------------------------|
| 6815 | 0 | ROA(C) before interest and depreciation before interest | 0.538269 | ROA(A) before interest and depreciation after tax | 0.524172 | ROA(B) before interest and depreciation after tax | 0.598308 | 0.598308 | 0.998992 | 0.797414 | 0.809327 | 0.30352 |
| 6816 | 0 | Bankrupt? | 0.472725 | 0.383714 | 0.325036 | Operating Gross Margin | 0.610213 | Realized Sales | 0.998992 | Pre-tax net | After-tax net | Non-interest income and expense |
| 6817 | 0 | depreciation before interest | 0.506264 | 0.559911 | 0.554945 | 0.607959 | 0.607850 | 0.999074 | 0.797300 | 0.809399 | 0.30346 | Revenue |
| 6818 | 0 | interest | 0.570105 | 0.570105 | 0.549548 | 0.627409 | 0.627409 | 0.998080 | 0.801987 | 0.813800 | | 0.31341 |

5 rows x 96 columns



In [157]:

```
X_train.shape
```

Out[157]:

(5114, 95)

In [158]:

```
data.describe().T
```

Out[158]:

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|--------|----------|----------|-----|----------|----------|----------|-----|
| Bankrupt? | 6819.0 | 0.032263 | 0.176710 | 0.0 | 0.000000 | 0.000000 | 0.000000 | 1.0 |
| ROA(C) before interest and depreciation before interest | 6819.0 | 0.505180 | 0.060686 | 0.0 | 0.476527 | 0.502706 | 0.535563 | 1.0 |
| ROA(A) before interest and % after tax | 6819.0 | 0.558625 | 0.065620 | 0.0 | 0.535543 | 0.559802 | 0.589157 | 1.0 |
| ROA(B) before interest and depreciation after tax | 6819.0 | 0.553589 | 0.061595 | 0.0 | 0.527277 | 0.552278 | 0.584105 | 1.0 |
| Operating Gross Margin | 6819.0 | 0.607948 | 0.016934 | 0.0 | 0.600445 | 0.605997 | 0.613914 | 1.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Liability to Equity | 6819.0 | 0.280365 | 0.014463 | 0.0 | 0.276944 | 0.278778 | 0.281449 | 1.0 |
| Degree of Financial Leverage (DFL) | 6819.0 | 0.027541 | 0.015668 | 0.0 | 0.026791 | 0.026808 | 0.026913 | 1.0 |
| Interest Coverage Ratio (Interest expense to EBIT) | 6819.0 | 0.565358 | 0.013214 | 0.0 | 0.565158 | 0.565252 | 0.565725 | 1.0 |
| Net Income Flag | 6819.0 | 1.000000 | 0.000000 | 1.0 | 1.000000 | 1.000000 | 1.000000 | 1.0 |
| Equity to Liability | 6819.0 | 0.047578 | 0.050014 | 0.0 | 0.024477 | 0.033798 | 0.052838 | 1.0 |

96 rows x 8 columns

Clean and Prepare the data

In [159]:

```
data['Bankrupt?'].value_counts()
```

Out[159]:

0 6599
1 220
Name: Bankrupt?, dtype: int64

In [160]:

```
X=data.drop(['Bankrupt?'],axis=1)
```

In [161]:

```
X=data.drop('Bankrupt?',axis=1)
X
```

Out[161]:

| | ROA(C) before interest and depreciation before interest | ROA(A) before interest and % after tax | ROA(B) before interest and depreciation after tax | Operating Gross Margin | Realized Sales Gross Margin | Operating Profit Rate | Pre-tax net Interest Rate | After-tax net Interest Rate | Non-industry income and expenditure/revenue | Continuo inter rate (af t |
|------|--|--|---|------------------------------|--------------------------------------|-----------------------------|------------------------------------|--------------------------------------|---|------------------------------------|
| 0 | 0.370594 | 0.424389 | 0.405750 | 0.601457 | 0.601457 | 0.998969 | 0.796887 | 0.808809 | 0.302646 | 0.7809 |
| 1 | 0.464291 | 0.538214 | 0.516730 | 0.610235 | 0.610235 | 0.998946 | 0.797380 | 0.809301 | 0.303556 | 0.7815 |
| 2 | 0.426071 | 0.499019 | 0.472295 | 0.601450 | 0.601364 | 0.998857 | 0.796403 | 0.808388 | 0.302035 | 0.7802 |
| 3 | 0.399844 | 0.451265 | 0.457733 | 0.583541 | 0.583541 | 0.998700 | 0.796967 | 0.808966 | 0.303350 | 0.7812 |
| 4 | 0.465022 | 0.538432 | 0.522298 | 0.598783 | 0.598783 | 0.998973 | 0.797366 | 0.809304 | 0.303475 | 0.7815 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 6814 | 0.493687 | 0.539468 | 0.543230 | 0.604455 | 0.604462 | 0.998992 | 0.797409 | 0.809331 | 0.303510 | 0.7815 |
| 6815 | 0.475162 | 0.538269 | 0.524172 | 0.598308 | 0.598308 | 0.998992 | 0.797414 | 0.809327 | 0.303520 | 0.7815 |
| 6816 | 0.472725 | 0.533744 | 0.520638 | 0.610444 | 0.610213 | 0.998984 | 0.797401 | 0.809317 | 0.303512 | 0.7815 |
| 6817 | 0.506264 | 0.559911 | 0.554045 | 0.607850 | 0.607850 | 0.999074 | 0.797500 | 0.809399 | 0.303498 | 0.7816 |
| 6818 | 0.493053 | 0.570105 | 0.549548 | 0.627409 | 0.627409 | 0.998080 | 0.801987 | 0.813800 | 0.313415 | 0.7860 |

6819 rows x 95 columns



Building Model

In [163]:

```
Y=data['Bankrupt?']
Y
```

Out[163]:

```
0      1
1      1
2      1
3      1
4      1
..
6814   0
6815   0
6816   0
6817   0
6818   0
Name: Bankrupt?, Length: 6819, dtype: int64
```

In [164]:

```
from sklearn.model_selection import train_test_split
```

In [166]:

```
X_train, X_test, y_train, y_test = train_test_split(
    X, Y, test_size=0.33, random_state=4)
```

In [167]:

```
X_train
```

Out[167]:

| | ROA(C) before interest and depreciation before interest | ROA(A) before interest and % after tax | ROA(B) before interest and depreciation after tax | Operating Gross Margin | Realized Sales Gross Margin | Operating Profit Rate | Pre-tax net Interest Rate | After-tax net Interest Rate | Non-industry income and expenditure/revenue | Continuo inter rate (af t |
|--|--|--|---|------------------------------|--------------------------------------|-----------------------------|------------------------------------|--------------------------------------|---|------------------------------------|
|--|--|--|---|------------------------------|--------------------------------------|-----------------------------|------------------------------------|--------------------------------------|---|------------------------------------|

| | | | | | | | | | | | | | | | | | | |
|------|---|----------|--|----------|------------------------|----------|-----------------------------|----------|-----------------------|----------|-----------------------|----------|-------------------------|----------|---|----------|---|--------|
| 3955 | ROA(C) before interest and depreciation | 0.442597 | ROA(C) before interest and % after tax | 0.400099 | Operating Gross Margin | 0.631891 | Realized Sales Gross Margin | 0.631891 | Operating Profit Rate | 0.998672 | Pre-tax Interest Rate | 0.797016 | After-tax Interest Rate | 0.808964 | Non-industry income and expenditure/revenue | 0.303492 | Continued operation rate (after interest) | 0.7812 |
| 3063 | | 0.52499 | | 0.582043 | | 0.624911 | | 0.63499 | | 0.999180 | | 0.797600 | | 0.809593 | | 0.303608 | | 0.7816 |
| 3573 | | 0.49266 | | 0.537124 | | 0.609241 | | 0.609241 | | 0.998981 | | 0.797401 | | 0.809318 | | 0.303519 | | 0.7816 |
| 1052 | | 0.387072 | | 0.608155 | | 0.400878 | | 0.597940 | | 0.598466 | | 0.998872 | | 0.797708 | | 0.809531 | | 0.7811 |
| 5586 | | 0.537757 | | 0.579808 | | 0.575566 | | 0.608268 | | 0.608268 | | 0.999080 | | 0.797561 | | 0.809431 | | 0.7817 |
| ... | | ... | | ... | | ... | | ... | | ... | | ... | | ... | | ... | | ... |
| 6017 | | 0.550431 | | 0.601123 | | 0.586595 | | 0.606459 | | 0.606459 | | 0.999091 | | 0.797510 | | 0.809397 | | 0.7816 |
| 709 | | 0.551650 | | 0.607992 | | 0.588147 | | 0.620260 | | 0.620260 | | 0.999120 | | 0.797623 | | 0.809484 | | 0.7817 |
| 2487 | | 0.468045 | | 0.537178 | | 0.519246 | | 0.639343 | | 0.639343 | | 0.999373 | | 0.798015 | | 0.809783 | | 0.7819 |
| 174 | | 0.520694 | | 0.539468 | | 0.568285 | | 0.605435 | | 0.605435 | | 0.998997 | | 0.797396 | | 0.809325 | | 0.7816 |
| 1146 | | 0.488276 | | 0.548626 | | 0.533540 | | 0.600607 | | 0.600607 | | 0.998996 | | 0.797418 | | 0.809330 | | 0.7815 |

4568 rows x 95 columns



Feature scaling

In [168]:

```
from sklearn.preprocessing import StandardScaler
```

In [169]:

```
ss = StandardScaler()
x_train = ss.fit_transform(X_train)
x_test = ss.transform(X_test)
```

Models and find the Best one

In [170]:

```
from sklearn.linear_model import LogisticRegression
```

In [171]:

```
lr=LogisticRegression()
```

In [172]:

```
model=lr.fit(X_train,y_train)
```

/home/anjali/.local/lib/python3.6/site-packages/sklearn/linear_model/_logistic.py:765: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
<https://scikit-learn.org/stable/modules/preprocessing.html>
Please also refer to the documentation for alternative solver options:
https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)

In [173]:

```
prediction=model.predict(X_test)
```

In [174]:

```
prediction
```

Out[174]:

```
array([0, 0, 0, ..., 0, 0, 0])
```

```
In [175]:
```

```
from sklearn.metrics import confusion_matrix
```

```
In [176]:
```

```
cm=confusion_matrix(y_test,prediction)
cm
```

```
Out[176]:
```

```
array([[2160,   16],
       [   74,    1]])
```

```
In [177]:
```

```
from sklearn.metrics import accuracy_score
```

```
In [178]:
```

```
accuracy_score(y_test,prediction)
```

```
Out[178]:
```

```
0.9600177698800533
```

Decision Tree

```
In [179]:
```

```
from sklearn.tree import DecisionTreeClassifier
```

```
In [180]:
```

```
dec=DecisionTreeClassifier()
modell=dec.fit(X_train,y_train)
```

```
In [181]:
```

```
modell
```

```
Out[181]:
```

```
DecisionTreeClassifier()
```

```
In [182]:
```

```
prediction=modell.predict(X_test)
```

```
In [183]:
```

```
prediction
```

```
Out[183]:
```

```
array([0, 0, 0, ..., 0, 0, 0])
```

```
In [184]:
```

```
cnn=confusion_matrix(y_test,prediction)
```

```
In [185]:
```

```
cnn
```

```
Out[185]:
```

```
array([[2109,   67],
       [   47,   28]])
```

In [186]:

```
accuracy_score(y_test,prediction)
```

Out[186]:

0.9493558418480675

In [187]:

```
from sklearn.metrics import classification_report
```

In [188]:

```
print(classification_report(y_test,prediction))
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.98 | 0.97 | 0.97 | 2176 |
| 1 | 0.29 | 0.37 | 0.33 | 75 |
| accuracy | | | 0.95 | 2251 |
| macro avg | 0.64 | 0.67 | 0.65 | 2251 |
| weighted avg | 0.96 | 0.95 | 0.95 | 2251 |

In []: