week12-final-project

March 5, 2023

File : week12_final_project.ipynb Name : Avinash Alapati Date : 03/01/2023 Course : DSC-630 - Predictive Analytics

[42]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from imblearn.over_sampling import SMOTE
from scipy import stats
from sklearn import linear_model
from sklearn.metrics import confusion_matrix
from sklearn import tree
from sklearn import ensemble
from sklearn import neural_network
import scikitplot as skplt
from sklearn.neighbors import KNeighborsClassifier
from sklearn import metrics
from sklearn.linear_model import LogisticRegression
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
```

[43]:
```
df = pd.read_csv("/users/avinash_alapati/Downloads/DataScience/DSC630/data.csv")
pd.set_option('display.max_columns', None)
```

[44]:
```
#Dropping NA values
df = df.dropna()
```

[45]:
```
#Displaying head of the DataFrame
df.head()
```

[45]:

	Bankrupt?	ROA(C) before interest and depreciation before interest	ROA(A) before interest and % after tax
0	1	0.370594	
1	1	0.464291	
2	1	0.426071	
3	1	0.399844	
4	1	0.465022	

	Net Value Per Share (B)
0	0.424389
1	0.538214
2	0.499019
3	0.451265
4	0.538432

	ROA(B) before interest and depreciation after tax
0	0.405750
1	0.516730
2	0.472295
3	0.457733
4	0.522298

	Operating Gross Margin	Realized Sales Gross Margin
0	0.601457	0.601457
1	0.610235	0.610235
2	0.601450	0.601364
3	0.583541	0.583541
4	0.598783	0.598783

	Operating Profit Rate	Pre-tax net Interest Rate
0	0.998969	0.796887
1	0.998946	0.797380
2	0.998857	0.796403
3	0.998700	0.796967
4	0.998973	0.797366

	After-tax net Interest Rate	Non-industry income and expenditure/revenue
0	0.808809	0.302646
1	0.809301	0.303556
2	0.808388	0.302035
3	0.808966	0.303350
4	0.809304	0.303475

	Continuous interest rate (after tax)	Operating Expense Rate
0	0.780985	1.256969e-04
1	0.781506	2.897851e-04
2	0.780284	2.361297e-04
3	0.781241	1.078888e-04
4	0.781550	7.890000e+09

	Research and development expense rate	Cash flow rate
0	0.0	0.458143
1	0.0	0.461867
2	25500000.0	0.458521
3	0.0	0.465705
4	0.0	0.462746

	Interest-bearing debt interest rate	Tax rate (A)
0	0.000725	0.0
1	0.000647	0.0
2	0.000790	0.0
3	0.000449	0.0
4	0.000686	0.0

	Net Value Per Share (B)	Net Value Per Share (A)
0	0.147950	0.147950
1	0.182251	0.182251
2	0.177911	0.177911
3	0.154187	0.154187
4	0.167502	0.167502

	Net Value Per Share (C)	Persistent EPS in the Last Four Seasons
0	0.147950	0.169141
1	0.182251	0.208944
2	0.193713	0.180581
3	0.154187	0.193722
4	0.167502	0.212537

	Cash Flow Per Share	Revenue Per Share (Yuan ¥)
0	0.311664	0.017560
1	0.318137	0.021144
2	0.307102	0.005944
3	0.321674	0.014368
4	0.319162	0.029690

	Operating Profit Per Share (Yuan ¥)
0	0.095921
1	0.093722
2	0.092338
3	0.077762
4	0.096898

	Per Share Net profit before tax (Yuan ¥)
0	0.138736
1	0.169918
2	0.142803
3	0.148603
4	0.168412

	Realized Sales Gross Profit Growth Rate	Operating Profit Growth Rate
0	0.022102	0.848195
1	0.022080	0.848088
2	0.022760	0.848094

3	0.022046	0.848005
4	0.022096	0.848258

	After-tax Net Profit Growth Rate	Regular Net Profit Growth Rate
0	0.688979	0.688979
1	0.689693	0.689702
2	0.689463	0.689470
3	0.689110	0.689110
4	0.689697	0.689697

	Continuous Net Profit Growth Rate	Total Asset Growth Rate
0	0.217535	4.980000e+09
1	0.217620	6.110000e+09
2	0.217601	7.280000e+09
3	0.217568	4.880000e+09
4	0.217626	5.510000e+09

	Net Value Growth Rate	Total Asset Return Growth Rate Ratio
0	0.000327	0.263100
1	0.000443	0.264516
2	0.000396	0.264184
3	0.000382	0.263371
4	0.000439	0.265218

	Cash Reinvestment %	Current Ratio	Quick Ratio
0	0.363725	0.002259	0.001208
1	0.376709	0.006016	0.004039
2	0.368913	0.011543	0.005348
3	0.384077	0.004194	0.002896
4	0.379690	0.006022	0.003727

	Interest Expense Ratio	Total debt/Total net worth	Debt ratio %
0	0.629951	0.021266	0.207576
1	0.635172	0.012502	0.171176
2	0.629631	0.021248	0.207516
3	0.630228	0.009572	0.151465
4	0.636055	0.005150	0.106509

	Net worth/Assets	Long-term fund suitability ratio (A)
0	0.792424	0.005024
1	0.828824	0.005059
2	0.792484	0.005100
3	0.848535	0.005047
4	0.893491	0.005303

	Borrowing dependency	Contingent liabilities/Net worth
0	0.390284	0.006479

```
1              0.376760                          0.005835
2              0.379093                          0.006562
3              0.379743                          0.005366
4              0.375025                          0.006624

    Operating profit/Paid-in capital   Net profit before tax/Paid-in capital  \
0                           0.095885                                0.137757
1                           0.093743                                0.168962
2                           0.092318                                0.148036
3                           0.077727                                0.147561
4                           0.096927                                0.167461

    Inventory and accounts receivable/Net value   Total Asset Turnover  \
0                                      0.398036               0.086957
1                                      0.397725               0.064468
2                                      0.406580               0.014993
3                                      0.397925               0.089955
4                                      0.400079               0.175412

    Accounts Receivable Turnover   Average Collection Days  \
0                       0.001814                 0.003487
1                       0.001286                 0.004917
2                       0.001495                 0.004227
3                       0.001966                 0.003215
4                       0.001449                 0.004367

    Inventory Turnover Rate (times)   Fixed Assets Turnover Frequency  \
0                      1.820926e-04                      1.165007e-04
1                      9.360000e+09                      7.190000e+08
2                      6.500000e+07                      2.650000e+09
3                      7.130000e+09                      9.150000e+09
4                      1.633674e-04                      2.935211e-04

    Net Worth Turnover Rate (times)   Revenue per person  \
0                          0.032903            0.034164
1                          0.025484            0.006889
2                          0.013387            0.028997
3                          0.028065            0.015463
4                          0.040161            0.058111

    Operating profit per person   Allocation rate per person  \
0                      0.392913                    0.037135
1                      0.391590                    0.012335
2                      0.381968                    0.141016
3                      0.378497                    0.021320
4                      0.394371                    0.023988
```

	Working Capital to Total Assets	Quick Assets/Total Assets	\
0	0.672775	0.166673	
1	0.751111	0.127236	
2	0.829502	0.340201	
3	0.725754	0.161575	
4	0.751822	0.260330	

	Current Assets/Total Assets	Cash/Total Assets	\
0	0.190643	0.004094	
1	0.182419	0.014948	
2	0.602806	0.000991	
3	0.225815	0.018851	
4	0.358380	0.014161	

	Quick Assets/Current Liability	Cash/Current Liability	\
0	0.001997	1.473360e-04	
1	0.004136	1.383910e-03	
2	0.006302	5.340000e+09	
3	0.002961	1.010646e-03	
4	0.004275	6.804636e-04	

	Current Liability to Assets	Operating Funds to Liability	\
0	0.147308	0.334015	
1	0.056963	0.341106	
2	0.098162	0.336731	
3	0.098715	0.348716	
4	0.110195	0.344639	

	Inventory/Working Capital	Inventory/Current Liability	\
0	0.276920	0.001036	
1	0.289642	0.005210	
2	0.277456	0.013879	
3	0.276580	0.003540	
4	0.287913	0.004869	

	Current Liabilities/Liability	Working Capital/Equity	\
0	0.676269	0.721275	
1	0.308589	0.731975	
2	0.446027	0.742729	
3	0.615848	0.729825	
4	0.975007	0.732000	

	Current Liabilities/Equity	Long-term Liability to Current Assets	\
0	0.339077	0.025592	
1	0.329740	0.023947	
2	0.334777	0.003715	
3	0.331509	0.022165	

```
4                    0.330726                                  0.000000

   Retained Earnings to Total Assets   Total income/Total expense  \
0                           0.903225                     0.002022
1                           0.931065                     0.002226
2                           0.909903                     0.002060
3                           0.906902                     0.001831
4                           0.913850                     0.002224

   Total expense/Assets   Current Asset Turnover Rate  \
0              0.064856                 7.010000e+08
1              0.025516                 1.065198e-04
2              0.021387                 1.791094e-03
3              0.024161                 8.140000e+09
4              0.026385                 6.680000e+09

   Quick Asset Turnover Rate   Working capitcal Turnover Rate  \
0               6.550000e+09                        0.593831
1               7.700000e+09                        0.593916
2               1.022676e-03                        0.594502
3               6.050000e+09                        0.593889
4               5.050000e+09                        0.593915

   Cash Turnover Rate   Cash Flow to Sales   Fixed Assets to Assets  \
0       4.580000e+08             0.671568                 0.424206
1       2.490000e+09             0.671570                 0.468828
2       7.610000e+08             0.671571                 0.276179
3       2.030000e+09             0.671519                 0.559144
4       8.240000e+08             0.671563                 0.309555

   Current Liability to Liability   Current Liability to Equity  \
0                        0.676269                     0.339077
1                        0.308589                     0.329740
2                        0.446027                     0.334777
3                        0.615848                     0.331509
4                        0.975007                     0.330726

   Equity to Long-term Liability   Cash Flow to Total Assets  \
0                       0.126549                   0.637555
1                       0.120916                   0.641100
2                       0.117922                   0.642765
3                       0.120760                   0.579039
4                       0.110933                   0.622374

   Cash Flow to Liability   CFO to Assets   Cash Flow to Equity  \
0                0.458609        0.520382              0.312905
1                0.459001        0.567101              0.314163
```

```
2                0.459254        0.538491             0.314515
3                0.448518        0.604105             0.302382
4                0.454411        0.578469             0.311567

   Current Liability to Current Assets   Liability-Assets Flag  \
0                             0.118250                       0
1                             0.047775                       0
2                             0.025346                       0
3                             0.067250                       0
4                             0.047725                       0

   Net Income to Total Assets   Total assets to GNP price  \
0                    0.716845                   0.009219
1                    0.795297                   0.008323
2                    0.774670                   0.040003
3                    0.739555                   0.003252
4                    0.795016                   0.003878

   No-credit Interval   Gross Profit to Sales  \
0            0.622879                0.601453
1            0.623652                0.610237
2            0.623841                0.601449
3            0.622929                0.583538
4            0.623521                0.598782

   Net Income to Stockholder's Equity   Liability to Equity  \
0                            0.827890              0.290202
1                            0.839969              0.283846
2                            0.836774              0.290189
3                            0.834697              0.281721
4                            0.839973              0.278514

   Degree of Financial Leverage (DFL)  \
0                            0.026601
1                            0.264577
2                            0.026555
3                            0.026697
4                            0.024752

   Interest Coverage Ratio (Interest expense to EBIT)   Net Income Flag  \
0                                           0.564050                 1
1                                           0.570175                 1
2                                           0.563706                 1
3                                           0.564663                 1
4                                           0.575617                 1

   Equity to Liability
```

```
0              0.016469
1              0.020794
2              0.016474
3              0.023982
4              0.035490
```

[46]:
```
#print the DataFrame values
for name in df:
    print(name)
```

```
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
 Continuous interest rate (after tax)
 Operating Expense Rate
 Research and development expense rate
 Cash flow rate
 Interest-bearing debt interest rate
 Tax rate (A)
 Net Value Per Share (B)
 Net Value Per Share (A)
 Net Value Per Share (C)
 Persistent EPS in the Last Four Seasons
 Cash Flow Per Share
 Revenue Per Share (Yuan ¥)
 Operating Profit Per Share (Yuan ¥)
 Per Share Net profit before tax (Yuan ¥)
 Realized Sales Gross Profit Growth Rate
 Operating Profit Growth Rate
 After-tax Net Profit Growth Rate
 Regular Net Profit Growth Rate
 Continuous Net Profit Growth Rate
 Total Asset Growth Rate
 Net Value Growth Rate
 Total Asset Return Growth Rate Ratio
 Cash Reinvestment %
 Current Ratio
 Quick Ratio
 Interest Expense Ratio
 Total debt/Total net worth
```

Debt ratio %
Net worth/Assets
Long-term fund suitability ratio (A)
Borrowing dependency
Contingent liabilities/Net worth
Operating profit/Paid-in capital
Net profit before tax/Paid-in capital
Inventory and accounts receivable/Net value
Total Asset Turnover
Accounts Receivable Turnover
Average Collection Days
Inventory Turnover Rate (times)
Fixed Assets Turnover Frequency
Net Worth Turnover Rate (times)
Revenue per person
Operating profit per person
Allocation rate per person
Working Capital to Total Assets
Quick Assets/Total Assets
Current Assets/Total Assets
Cash/Total Assets
Quick Assets/Current Liability
Cash/Current Liability
Current Liability to Assets
Operating Funds to Liability
Inventory/Working Capital
Inventory/Current Liability
Current Liabilities/Liability
Working Capital/Equity
Current Liabilities/Equity
Long-term Liability to Current Assets
Retained Earnings to Total Assets
Total income/Total expense
Total expense/Assets
Current Asset Turnover Rate
Quick Asset Turnover Rate
Working capitcal Turnover Rate
Cash Turnover Rate
Cash Flow to Sales
Fixed Assets to Assets
Current Liability to Liability
Current Liability to Equity
Equity to Long-term Liability
Cash Flow to Total Assets
Cash Flow to Liability
CFO to Assets
Cash Flow to Equity
Current Liability to Current Assets

```
Liability-Assets Flag
Net Income to Total Assets
Total assets to GNP price
No-credit Interval
Gross Profit to Sales
Net Income to Stockholder's Equity
Liability to Equity
Degree of Financial Leverage (DFL)
Interest Coverage Ratio (Interest expense to EBIT)
Net Income Flag
Equity to Liability
```

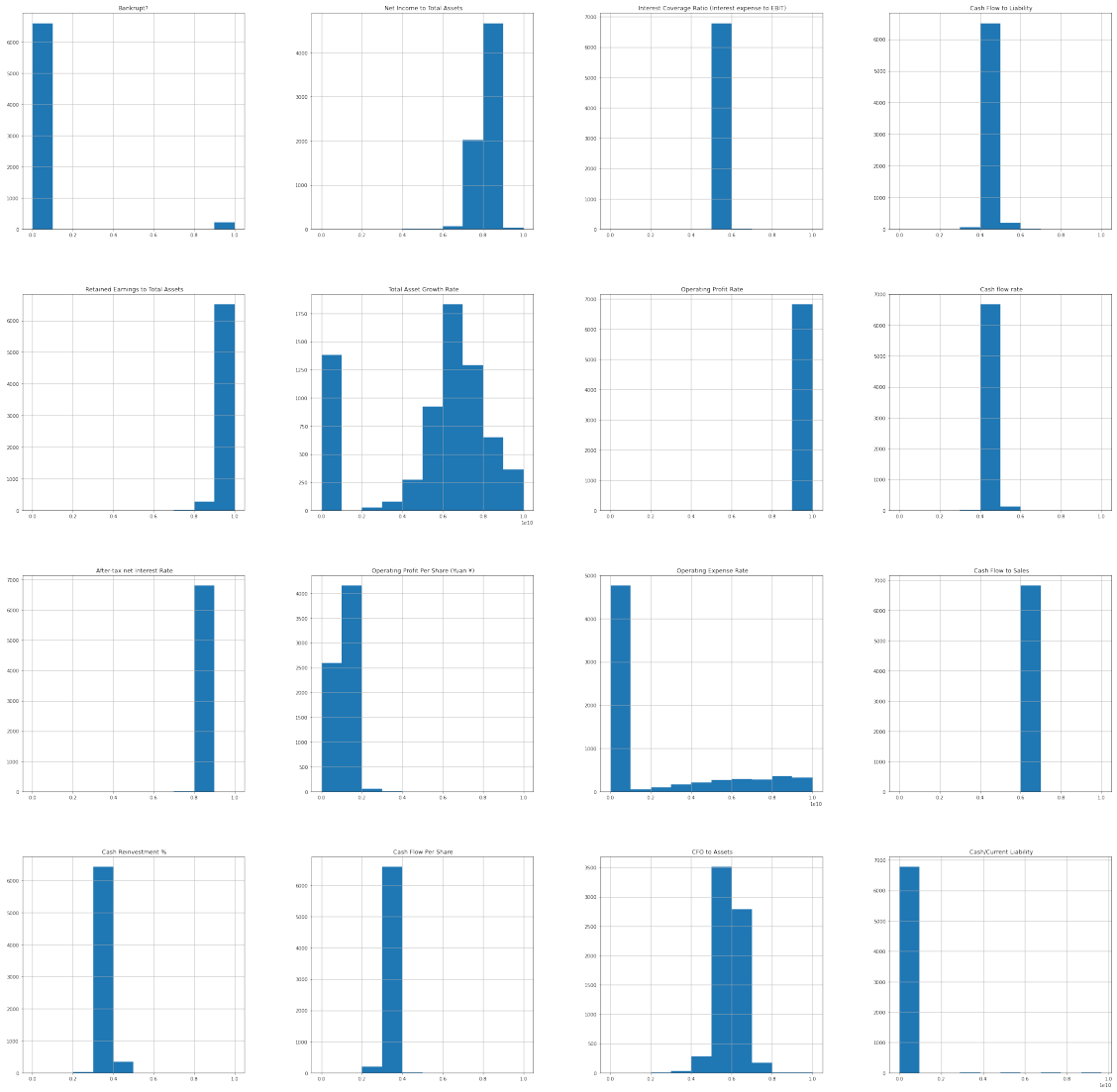
[47]:
```
#Construct a new DataFrame with only needed entries
df1 = df[["Bankrupt?", " Net Income to Total Assets", " Interest Coverage Ratio␣
 ↪(Interest expense to EBIT)",
        " Cash Flow to Liability", " Retained Earnings to Total Assets", "␣
 ↪Total Asset Growth Rate",
        " Operating Profit Rate", " Cash flow rate", " After-tax net Interest␣
 ↪Rate", " Operating Profit Per Share (Yuan ¥)",
        " Operating Expense Rate", " Cash Flow to Sales", " Cash Reinvestment␣
 ↪%", " Cash Flow Per Share",
        " CFO to Assets", " Cash/Current Liability"]]
```

[48]:
```
#Print DataFrame objects
df1.dtypes
```

[48]:
```
Bankrupt?                                                 int64
 Net Income to Total Assets                             float64
 Interest Coverage Ratio (Interest expense to EBIT)     float64
 Cash Flow to Liability                                 float64
 Retained Earnings to Total Assets                      float64
 Total Asset Growth Rate                                float64
 Operating Profit Rate                                  float64
 Cash flow rate                                         float64
 After-tax net Interest Rate                            float64
 Operating Profit Per Share (Yuan ¥)                    float64
 Operating Expense Rate                                 float64
 Cash Flow to Sales                                     float64
 Cash Reinvestment %                                    float64
 Cash Flow Per Share                                    float64
 CFO to Assets                                          float64
 Cash/Current Liability                                 float64
dtype: object
```

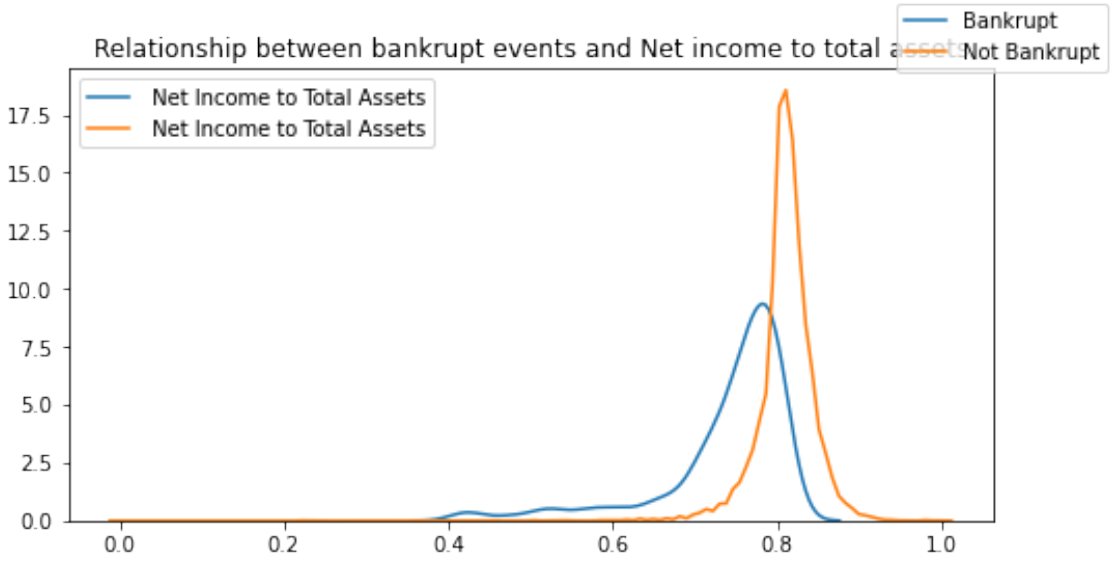
[49]:
```
#Display histogram objects
df1.hist(figsize = (40,40))
```

```
[49]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7fcbfd9b93a0>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x7fcbfd96f1c0>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x7fcbf762d940>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x7fcbf7ce0df0>],
             [<matplotlib.axes._subplots.AxesSubplot object at 0x7fcbf8a3d250>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x7fcbf8a6a5e0>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x7fcbf8a6a6d0>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x7fcbf8a97b80>],
             [<matplotlib.axes._subplots.AxesSubplot object at 0x7fcbfa20d3d0>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x7fcbfa23b820>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x7fcbfa878ca0>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x7fcbfa8b0130>],
             [<matplotlib.axes._subplots.AxesSubplot object at 0x7fcbfa8dc580>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x7fcbfa9099d0>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x7fcbfa937e20>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x7fcbfa9712b0>]],
            dtype=object)
```



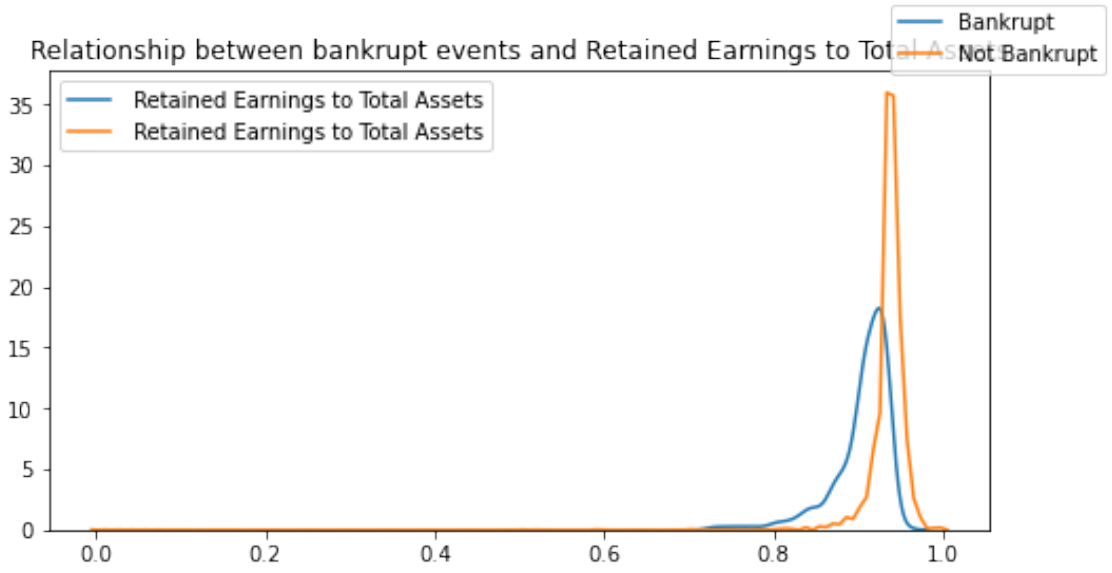
[70]:
```
#plot sns plot for Bankrupt to Net Income to Totak Assests
fig = plt.figure(figsize=(8,4))
sns.kdeplot(df1[df1['Bankrupt?']==1][' Net Income to Total Assets'])
sns.kdeplot(df1[df1['Bankrupt?']==0][' Net Income to Total Assets'])
fig.legend(labels=['Bankrupt', 'Not Bankrupt'])
plt.title('Relationship between bankrupt events and Net income to total assets')
```

[70]:
```
Text(0.5, 1.0, 'Relationship between bankrupt events and Net income to total
assets')
```



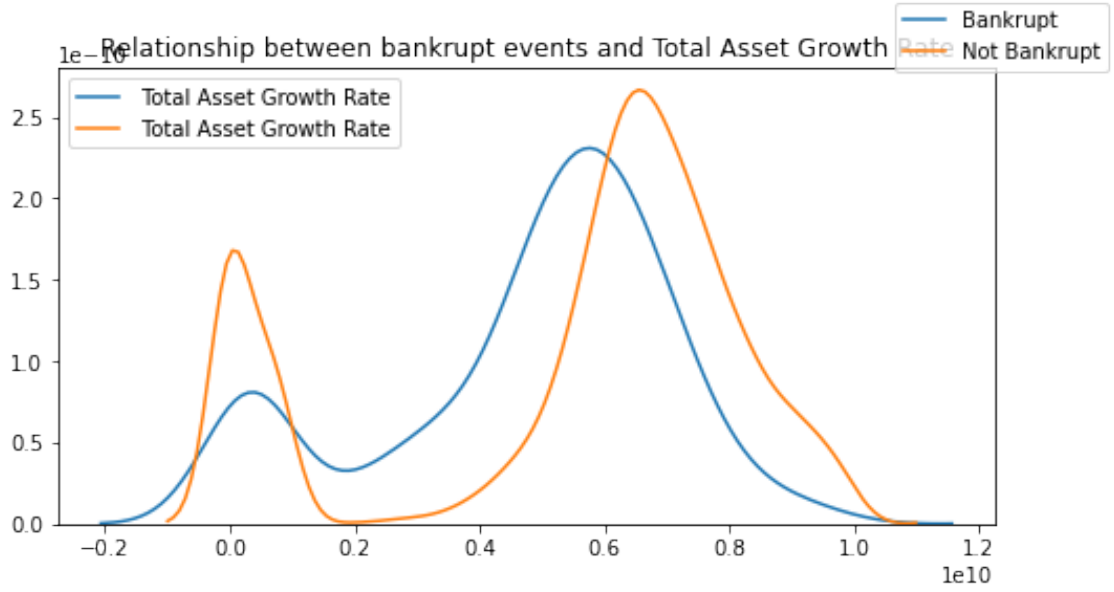
[72]:
```
#plot sns plot for Bankrupt to Retained Earnings to Total Assests
fig = plt.figure(figsize=(8,4))
sns.kdeplot(df1[df1['Bankrupt?']==1][' Retained Earnings to Total Assets'])
sns.kdeplot(df1[df1['Bankrupt?']==0][' Retained Earnings to Total Assets'])
fig.legend(labels=['Bankrupt', 'Not Bankrupt'])
plt.title('Relationship between bankrupt events and Retained Earnings to Total␣
 ↪Assets')
```

[72]:
```
Text(0.5, 1.0, 'Relationship between bankrupt events and Retained Earnings to
Total Assets')
```



[53]:
```
#plot sns plot for Bankrupt to Retained Earnings to Total Asset Growth Rate
fig = plt.figure(figsize=(8,4))
sns.kdeplot(df[df['Bankrupt?']==1][' Total Asset Growth Rate'])
sns.kdeplot(df[df['Bankrupt?']==0][' Total Asset Growth Rate'])
fig.legend(labels=['Bankrupt', 'Not Bankrupt'])
plt.title('Relationship between bankrupt events and Total Asset Growth Rate')
```

[53]:
```
Text(0.5, 1.0, 'Relationship between bankrupt events and Total Asset Growth
Rate')
```



[54]:
```
#Dropping Bankrupt column for training and testing the model
X = df1.drop(columns = "Bankrupt?")
Y = df1["Bankrupt?"]
```

[55]:
```
X_train,X_test,Y_train,Y_test = train_test_split(X,Y,random_state=1)
```

[56]:
```
X_train,Y_train = SMOTE(random_state=1).fit_resample(X_train,Y_train)
```

[57]:
```
X_train = stats.zscore(X_train)
X_test = stats.zscore(X_test)
```

[58]:
```
#LogisticRegression Fit for Train and Test Data sets
model = LogisticRegression()
model.fit(X_train,Y_train)
pred = model.predict(X_test)
cm = confusion_matrix(Y_test,pred)
```

```
print("Train set Accuracy: ", metrics.accuracy_score(Y_train, model.
  ↪predict(X_train)))
print("Test set Accuracy: ", metrics.accuracy_score(Y_test, pred))
```

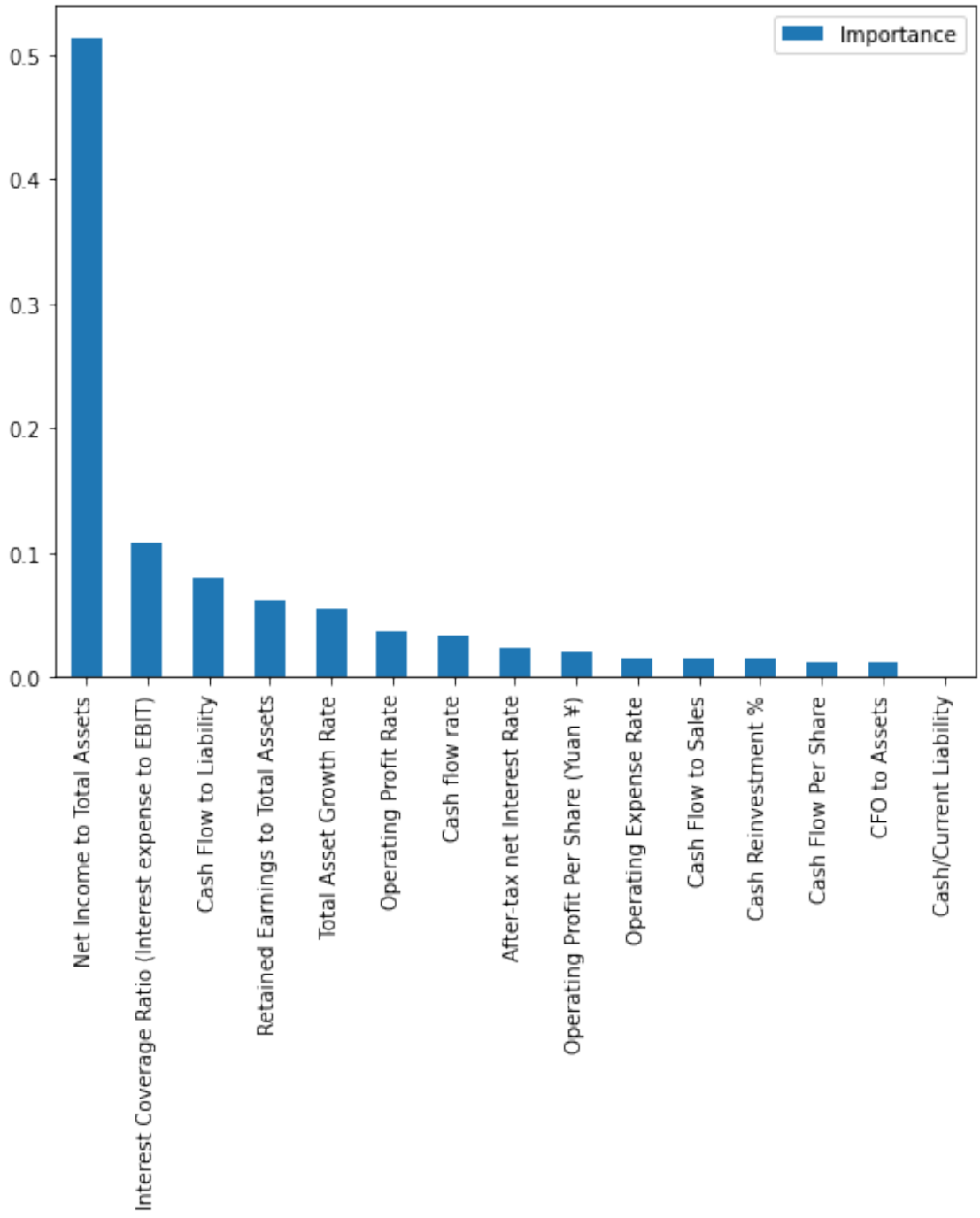
```
Train set Accuracy:  0.8418768920282543
Test set Accuracy:  0.4656891495601173
```

[59]:
```
#DecisionTree Fit for Train and Test Data sets
model = tree.DecisionTreeClassifier(random_state=1)
model.fit(X_train,Y_train)
pred = model.predict(X_test)
cm = confusion_matrix(Y_test,pred)
print("Train set Accuracy: ", metrics.accuracy_score(Y_train, model.
  ↪predict(X_train)))
print("Test set Accuracy: ", metrics.accuracy_score(Y_test, pred))
```

```
Train set Accuracy:  1.0
Test set Accuracy:  0.673900293255132
```

[60]:
```
#Analayze feature variables based on DecisionTree model analysis
feat_importances = pd.DataFrame(model.feature_importances_, index=X_train.
  ↪columns, columns=["Importance"])
feat_importances.sort_values(by='Importance', ascending=False, inplace=True)
feat_importances.plot(kind='bar', figsize=(8,6))
```

[60]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fcbfbb111c0>
```



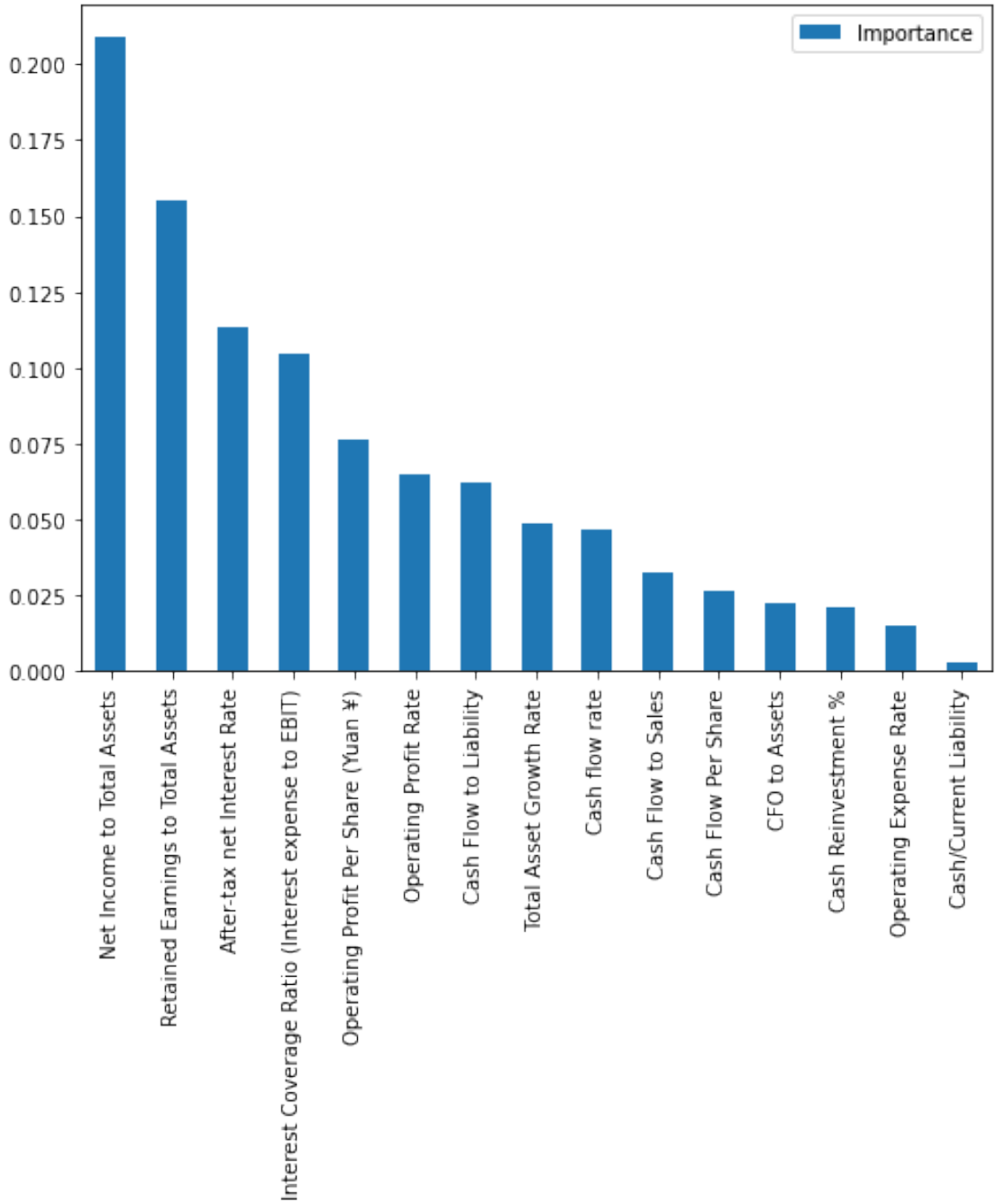
[40]:
```
#Analayze feature variables based on KNN model analysis
model = KNeighborsClassifier(n_neighbors=2)
model.fit(X_train,Y_train)
pred = model.predict(X_test)
cm = confusion_matrix(Y_test,pred)
```

```
print("Train set Accuracy: ", metrics.accuracy_score(Y_train, model.
  ↪predict(X_train)))
print("Test set Accuracy: ", metrics.accuracy_score(Y_test, pred))
```

```
Train set Accuracy:  0.998183652875883
Test set Accuracy:  0.7589442815249267
```

[63]:
```
#Analayze feature variables based on DecisionTree model analysis
feat_importances = pd.DataFrame(model.feature_importances_, index=X_train.
  ↪columns, columns=["Importance"])
feat_importances.sort_values(by='Importance', ascending=False, inplace=True)
feat_importances.plot(kind='bar', figsize=(8,6))
```

[63]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fcbfb8b7fa0>
```



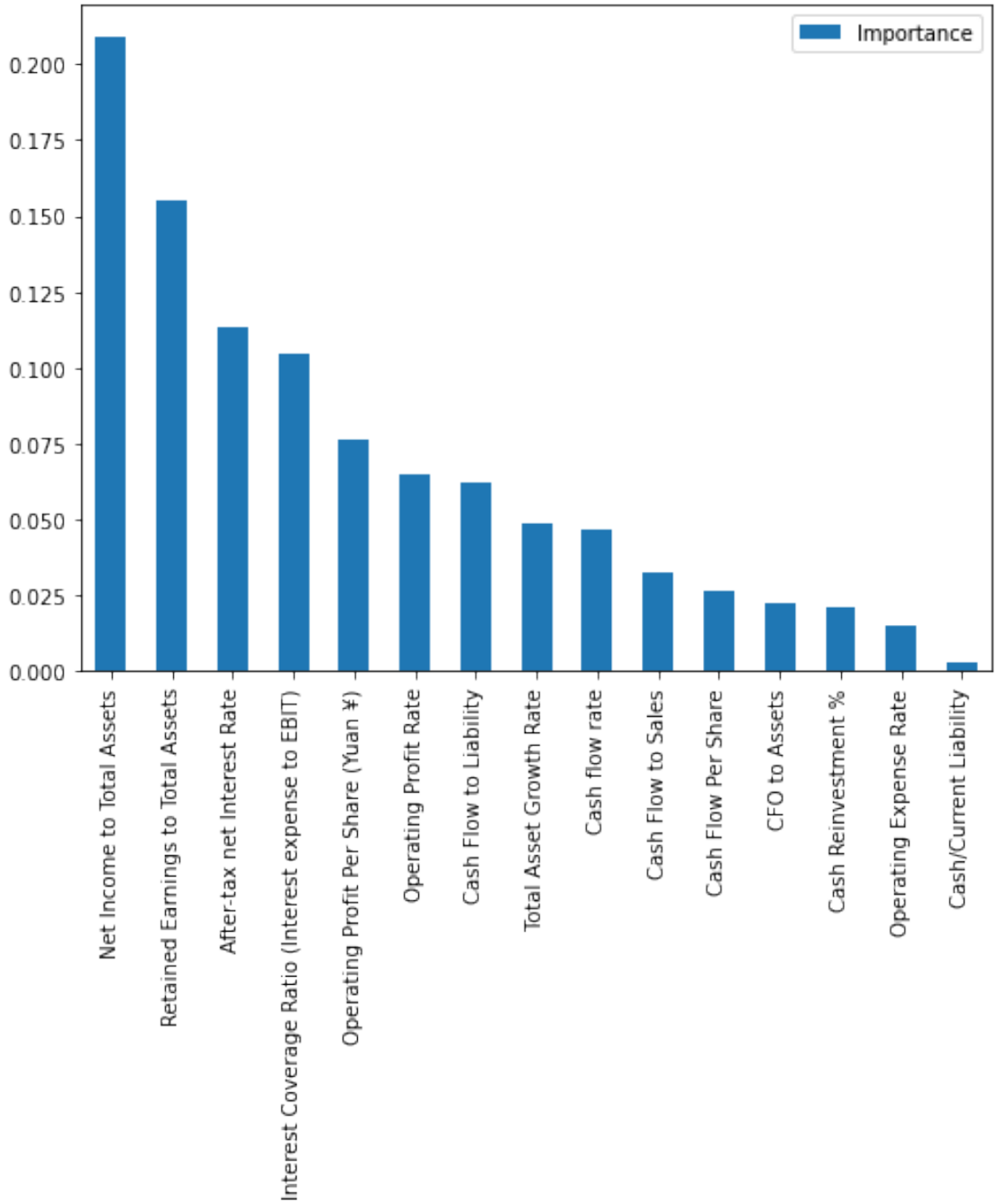
[61]:
```
#Analayze feature variables based on RandomForest Classifier
model = ensemble.RandomForestClassifier(random_state=1)
model.fit(X_train,Y_train)
pred = model.predict(X_test)
cm = confusion_matrix(Y_test,pred)
```

```
print("Train set Accuracy: ", metrics.accuracy_score(Y_train, model.
 ↪predict(X_train)))
print("Test set Accuracy: ", metrics.accuracy_score(Y_test, pred))
```

```
Train set Accuracy:  1.0
Test set Accuracy:  0.6299120234604105
```

[62]:
```
#Analayze feature variables based on RandomForest model analysis
feat_importances = pd.DataFrame(model.feature_importances_, index=X_train.
 ↪columns, columns=["Importance"])
feat_importances.sort_values(by='Importance', ascending=False, inplace=True)
feat_importances.plot(kind='bar', figsize=(8,6))
```

[62]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fcbfb817550>
```



Results observation and Analysis

From the comparison of the models, K neighbor classifier has the most accuracy with a K value of 2, test set accuracy up to 75%

Net Income to Total Assets growth will decrease bankruptcy chance

Retained earnings to Total Assests will decrease bankruptcy chance

After Tax net interest rate has direct impact on bankruptcy event.

There is no straightforward relationship between Bankrupt and Total Assests growth rate

There is no straightforward relationship between Bankrupt and Total Assests growth rate