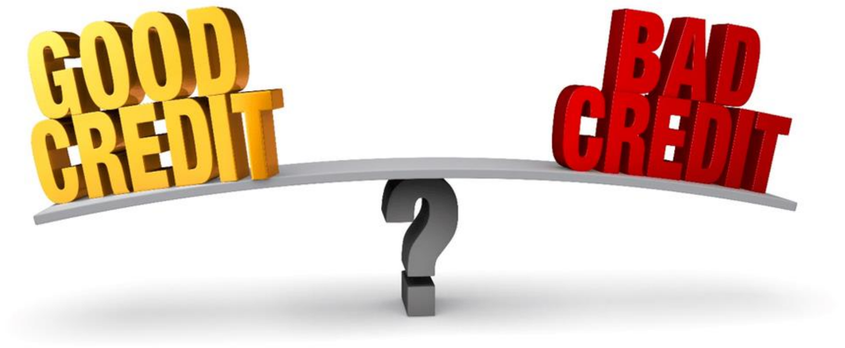
TAREA II

COMPANY BANKRUPTCY PREDICTION

RICHARD DOUGLAS G.

```
[6] data.dtypes #  tenemos 96 caracteristicas distintas (columnas)  y 6819 observaciones (filas)
                   # todas las caracteristicas presentes son del tipo númerico.  3 del tipo Int  y 93 del tipo Float

    Bankrupt?                                                  int64
     ROA(C) before interest and depreciation before interest    float64
     ROA(A) before interest and % after tax                     float64
     ROA(B) before interest and depreciation after tax          float64
     Operating Gross Margin                                     float64
                                                                 ...
     Liability to Equity                                        float64
     Degree of Financial Leverage (DFL)                         float64
     Interest Coverage Ratio (Interest expense to EBIT)         float64
     Net Income Flag                                            int64
     Equity to Liability                                        float64
    Length: 96, dtype: object
```

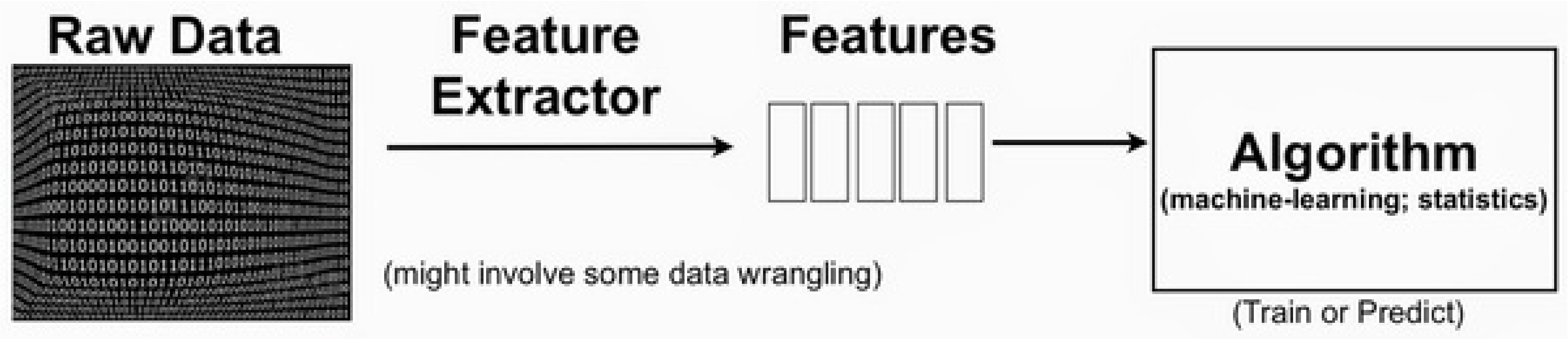


DATASET

- **Bankrupt?**: Aparece como una caracteristica binaria, en donde si indica cero 0, nos dice que la empresa es competente, mientras que si el valor es 1, nos indica que es probable que la empresa caiga en banca rota.

```python
# Importacion de bibliotecas

import numpy as np
import pandas as pd
from sklearn.datasets import make_blobs
import matplotlib.pyplot as plt
from matplotlib import style
import seaborn as sns
from sklearn.mixture import GaussianMixture #GMM
from sklearn.pipeline import make_pipeline
from sklearn.preprocessing import StandardScaler
import warnings
warnings.filterwarnings('ignore')
import argparse
from sklearn.preprocessing import MinMaxScaler
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
from kneed import KneeLocator
import plotly.graph_objects as go
from plotly.subplots import make_subplots
sns.set()
```



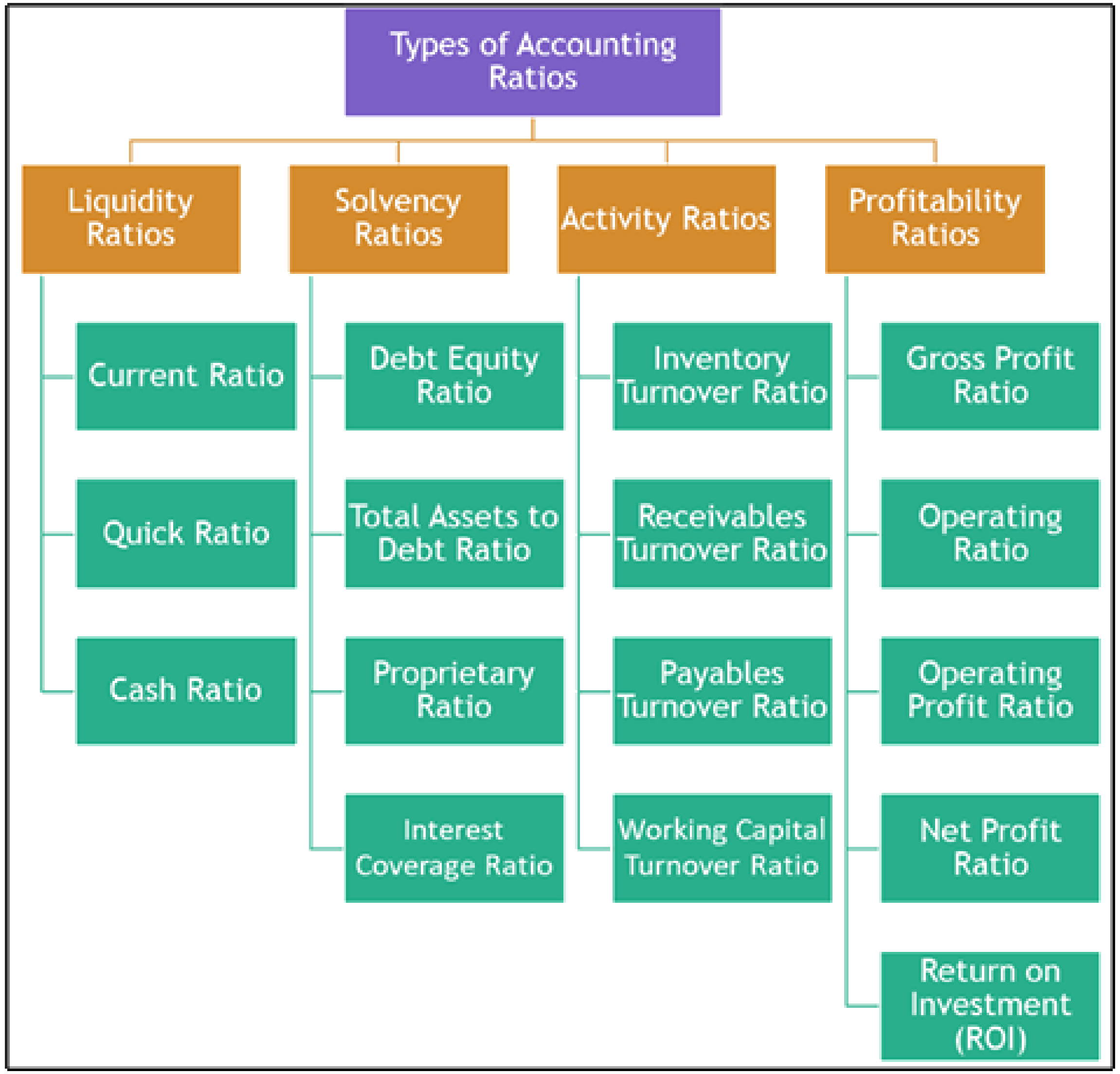
SE LE APLICA TECNICAS AL DATASET PARA LA COMPRENSION Y USO DE LA INFORMACION PARA GENERAR VALOR AL CLIENTE

EDA- COMPANY BANKRUPTCY PREDICTION

CONOCIMIENTO Y APLICACIÓN SOBRE LAS CARACTERISTICAS

EDA- COMPANY BANKRUPTCY PREDICTION

CONOCIMIENTO Y APLICACIÓN SOBRE LAS CARACTERISTICAS

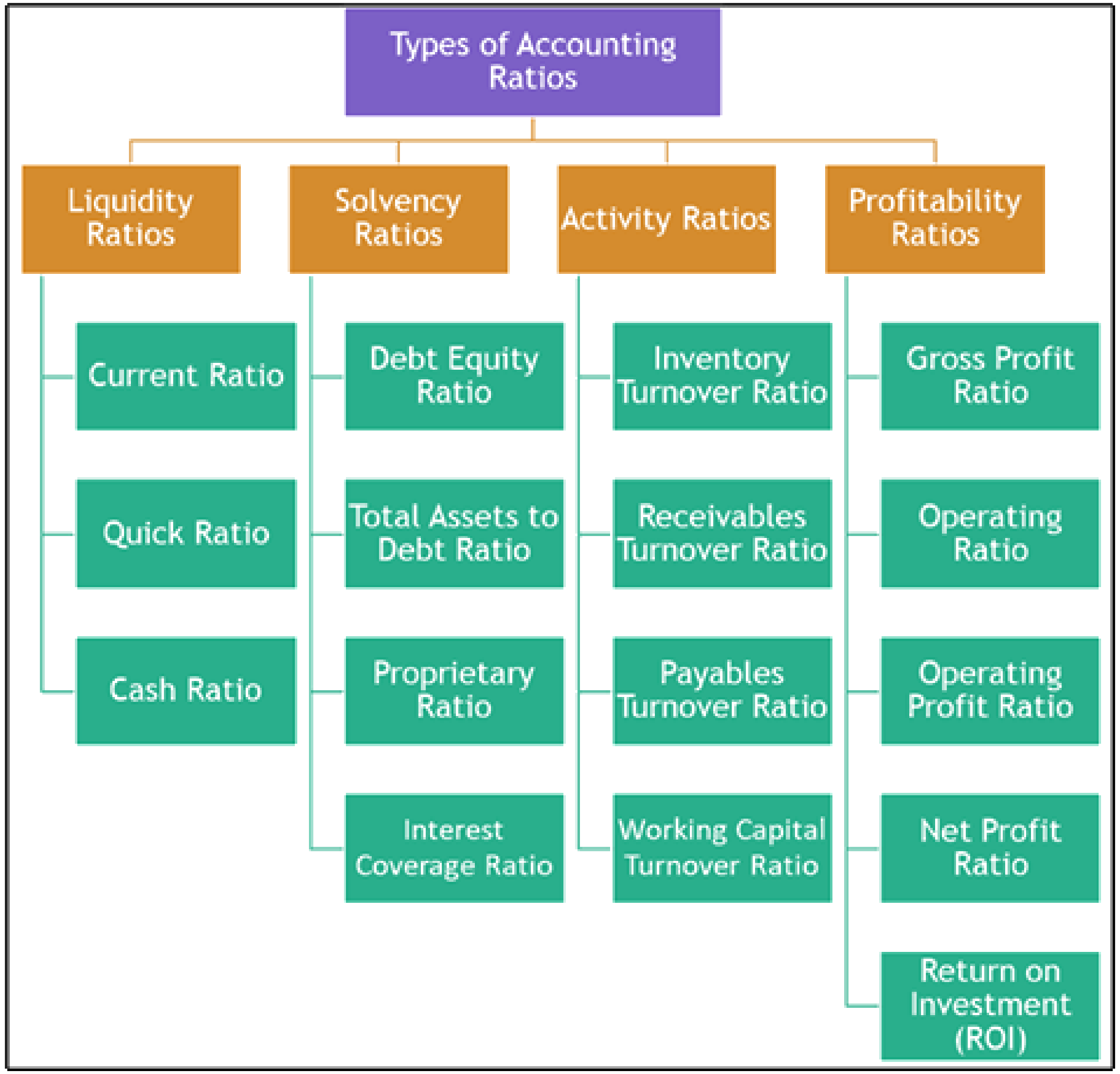


Los tres elementos determinantes de todo análisis financiero son:

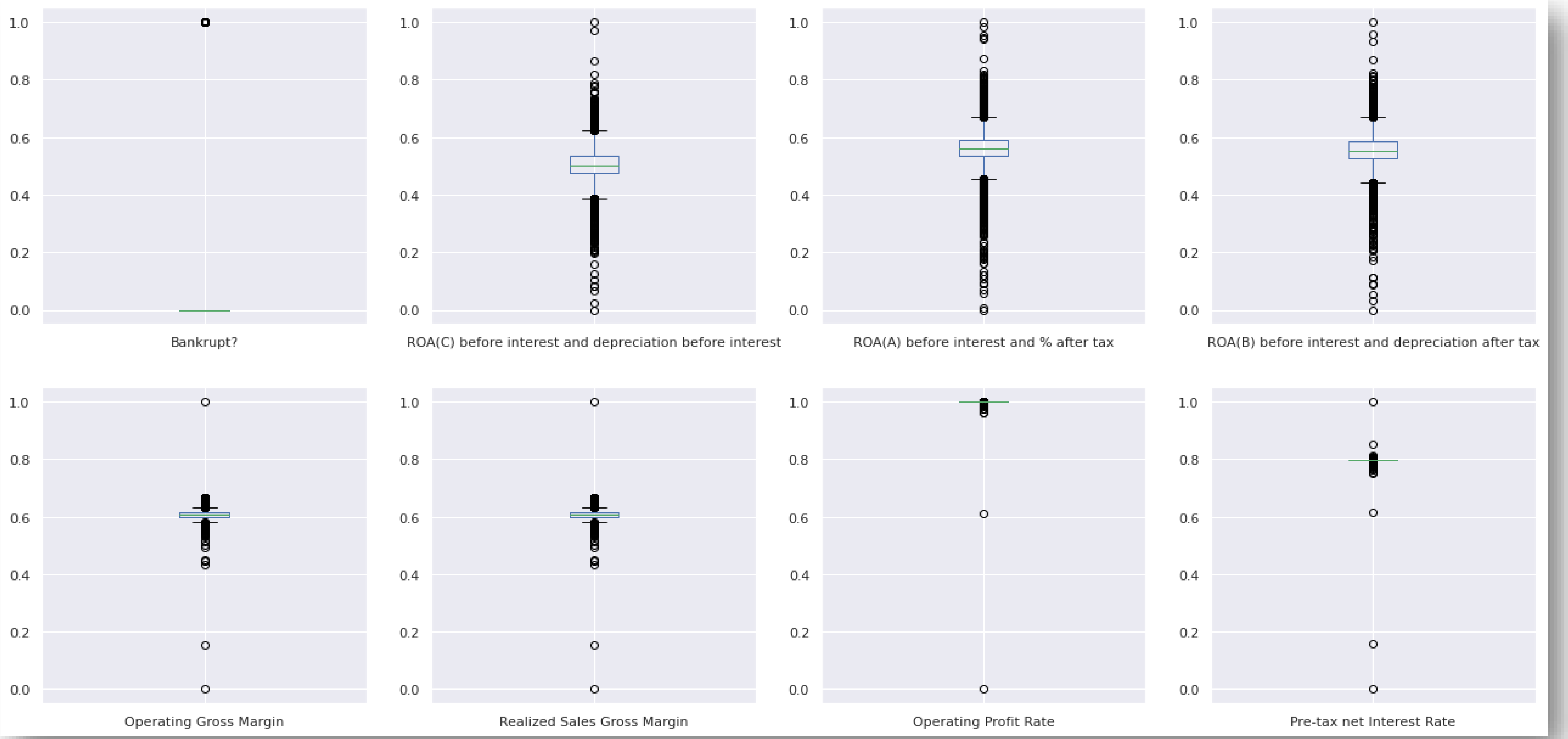
•**Liquidez**: capacidad para hacer frente y cumplir con sus obligaciones financieras a corto plazo.

•**Solvencia**: como responder a los compromisos de largo plazo (endeudamiento).

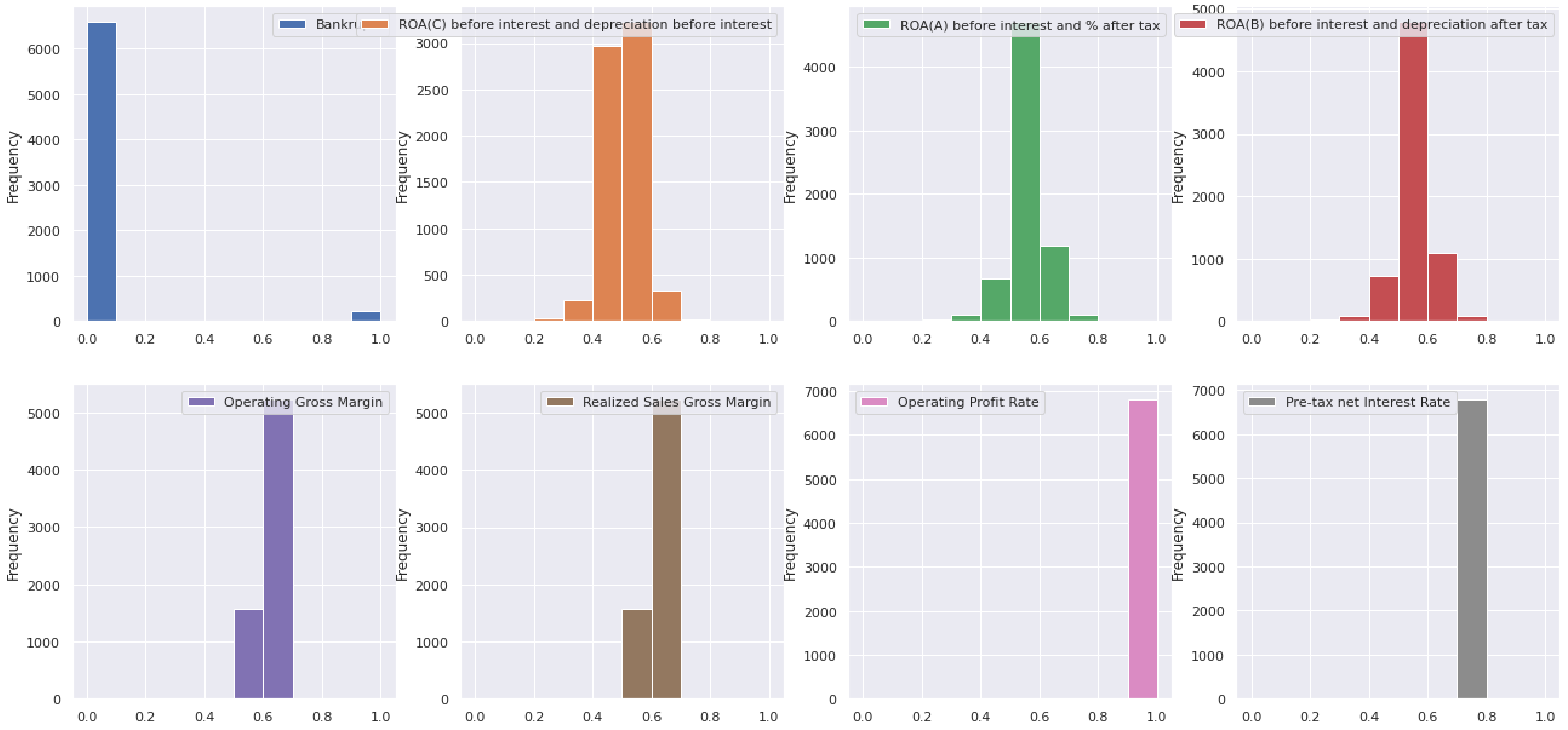
•**Rentabilidad**: capacidad de generar ingresos/beneficios, se puede medir el nivel de eficiencia con el que los recursos son utilizados en la empresa.



SE REALIZAN LOS BOXPLOTS PARA OBSERVAR LA EXISTENCIA DE VALORES ATÍPICOS



SE REALIZAN LOS HISTOGRAMAS PARA OBSERVAR LAS DISTRIBUCIONES



Correlaciones

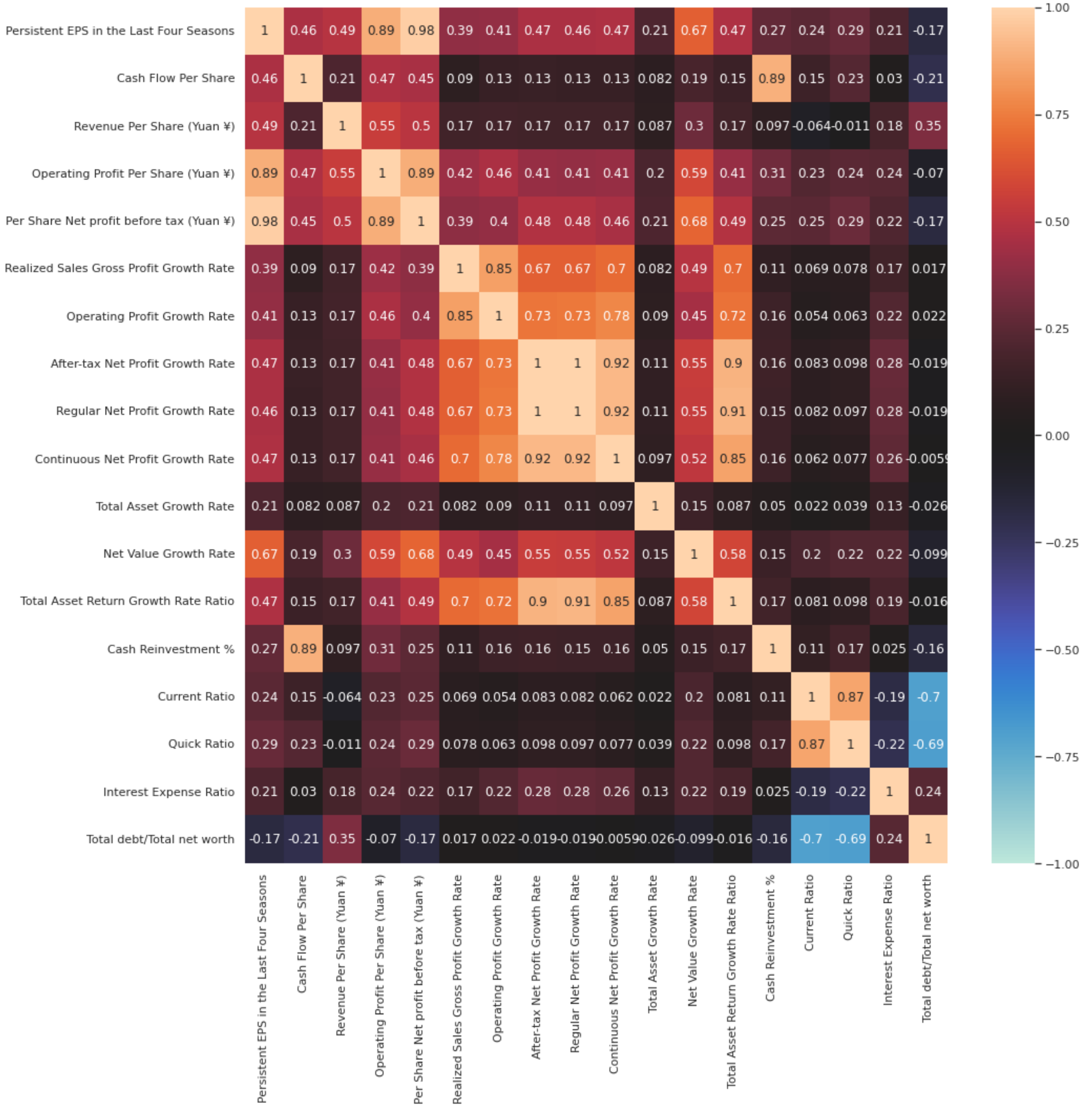
```python
cor_matrix = datos.corr().abs()   # esta versión permite colorear aquellas correlaciones que nos llaman la atención tanto positivas como negativas
cor_matrix.style.background_gradient(sns.light_palette('red', as_cmap=True))   # código tomado de la web en que aplican este método, es muy útil ayuda cuando hay muchas variables
```

	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	Continuo Net Pro Growt Rate
Bankrupt?	1.000000	0.260807	0.282941	0.273051	0.100043	0.099445	0.000230	0.008517	0.008857	0.016593	0.008395	0.006083	0.024232	0.072356	0.023063	0.109706	0.165399	0.165465	0.164784	0.219560	0.077516	0.004692	0.142051	0.201395	0.000458	0.015168	0.037783	0.036820	0.009401
ROA(C) before interest and depreciation before interest	0.260807	1.000000	0.940124	0.986849	0.334719	0.332755	0.035725	0.053419	0.049222	0.020501	0.051328	0.066869	0.106461	0.323482	0.048882	0.250761	0.505580	0.505407	0.505281	0.775006	0.379839	0.015932	0.687201	0.750564	0.000591	0.036511	0.115083	0.115040	0.025234
ROA(A) before interest and % after tax	0.282941	0.940124	1.000000	0.955741	0.326969	0.324956	0.032053	0.053518	0.049474	0.029649	0.049909	0.075727	0.084334	0.288440	0.050362	0.225897	0.531799	0.531790	0.531821	0.764828	0.326239	0.011829	0.654253	0.752578	0.003277	0.042208	0.125384	0.125872	0.024887
ROA(B) before interest and depreciation after tax	0.273051	0.986849	0.955741	1.000000	0.333749	0.331755	0.035212	0.053726	0.049952	0.022366	0.052261	0.065602	0.102147	0.323040	0.045839	0.197344	0.502052	0.502000	0.501907	0.764597	0.366216	0.014359	0.659834	0.722940	0.002142	0.036144	0.117130	0.117042	0.024414
Operating Gross Margin	0.100043	0.334719	0.326969	0.333749	1.000000	0.999518	0.005745	0.032493	0.027175	0.051438	0.029430	0.206353	0.016976	0.341188	0.017198	0.067970	0.144661	0.145031	0.145057	0.256722	0.163192	0.117045	0.267944	0.247789	0.014172	0.022867	0.054639	0.053430	0.009121
Realized Sales Gross Margin	0.099445	0.332755	0.324956	0.331755	0.999518	1.000000	0.005610	0.032232	0.026851	0.051242	0.029166	0.206439	0.017391	0.341433	0.017121	0.067708	0.142887	0.143262	0.143288	0.254753	0.163163	0.117196	0.267021	0.246004	0.014188	0.022778	0.054470	0.053259	0.009117
Operating Profit Rate	0.000230	0.035725	0.032053	0.035212	0.005745	0.005610	1.000000	0.916448	0.862191	0.592006	0.915544	0.013246	0.016387	0.023051	0.002784	0.019936	0.019257	0.019218	0.019240	0.020420	0.014244	0.044460	0.022397	0.020219	0.000831	0.004952	0.011328	0.011227	0.001318
Pre-tax net Interest Rate	0.008517	0.053419	0.053518	0.053726	0.032493	0.032232	0.916448	1.000000	0.986379	0.220045	0.993617	0.014247	0.016836	0.024950	0.004031	0.023003	0.033034	0.033015	0.033035	0.033726	0.017617	0.004931	0.026314	0.034046	0.001246	0.003909	0.035150	0.034914	0.003013
After-tax net Interest Rate	0.008857	0.049222	0.049474	0.049952	0.027175	0.026851	0.862191	0.986379	1.000000	0.115211	0.984452	0.013982	0.016521	0.022813	0.003824	0.021164	0.031369	0.031347	0.031367	0.030768	0.016140	0.005594	0.024137	0.030621	0.001226	0.002962	0.031223	0.030964	0.002565
Non-industry income and expenditure/revenue	0.016593	0.020501	0.029649	0.022366	0.051438	0.051242	0.592006	0.220045	0.115211	1.000000	0.230698	0.003597	0.006041	0.005943	0.001332	0.002270	0.019588	0.019644	0.019632	0.018148	0.000758	0.118316	0.001601	0.019279	0.000484	0.004200	0.043179	0.042951	0.002855
Continuous interest rate (after tax)	0.008395	0.051328	0.049909	0.052261	0.029430	0.029166	0.915544	0.993617	0.984452	0.230698	1.000000	0.013168	0.015728	0.027730	0.003654	0.020407	0.030839	0.030835	0.030840	0.032051	0.016343	0.051607	0.024516	0.030487	0.001207	0.002643	0.016584	0.016415	0.001842
Operating Expense Rate	0.006083	0.066869	0.075727	0.065602	0.206353	0.206439	0.013246	0.014247	0.013982	0.003597	0.013168	1.000000	0.060386	0.020147	0.006011	0.060683	0.090519	0.091263	0.091197	0.080969	0.007253	0.015838	0.071799	0.081428	0.008170	0.013374	0.007176	0.009511	0.006644
Research and development expense rate	0.024232	0.106461	0.084334	0.102147	0.016976	0.017391	0.016387	0.016836	0.016521	0.006041	0.015728	0.060386	1.000000	0.030918	0.000656	0.019201	0.088822	0.087500	0.087063	0.076486	0.052162	0.019291	0.068738	0.066085	0.011151	0.012166	0.019958	0.020703	0.007842
Cash flow rate	0.072356	0.323482	0.288440	0.323040	0.341188	0.341433	0.023051	0.024950	0.022813	0.005943	0.027730	0.020147	0.030918	1.000000	0.011986	0.049835	0.158471	0.158520	0.158255	0.197705	0.353883	0.201679	0.191974	0.177008	0.017070	0.003731	0.019071	0.018300	0.003902

MODELO DE MACHINE LEARNING
COMPANY BANKRUPTCY PREDICTION

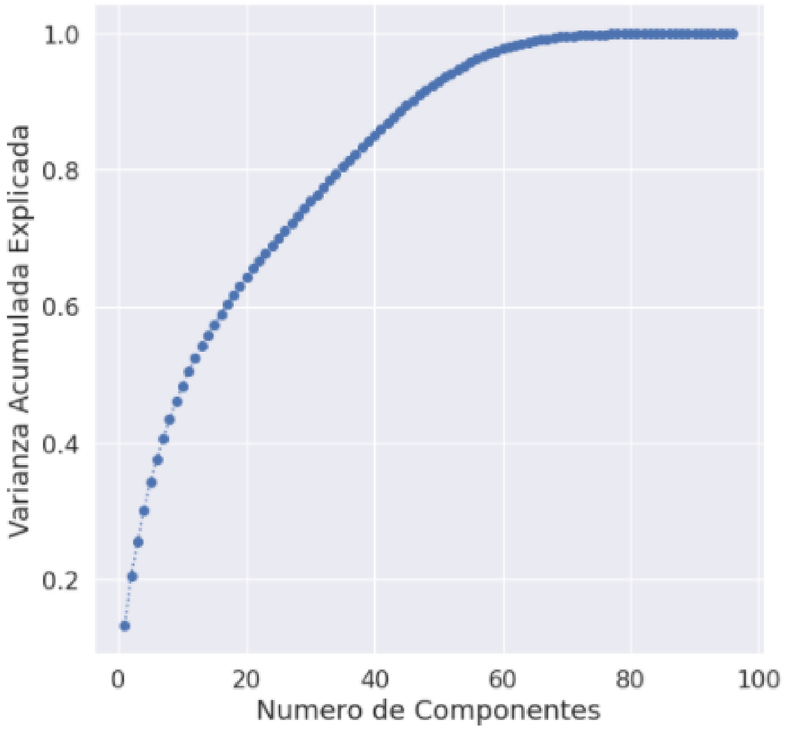
MATRIZ DE CORRELACIONES

- Se hace una gráfica con variables seleccionadas para facilitar a visualización.

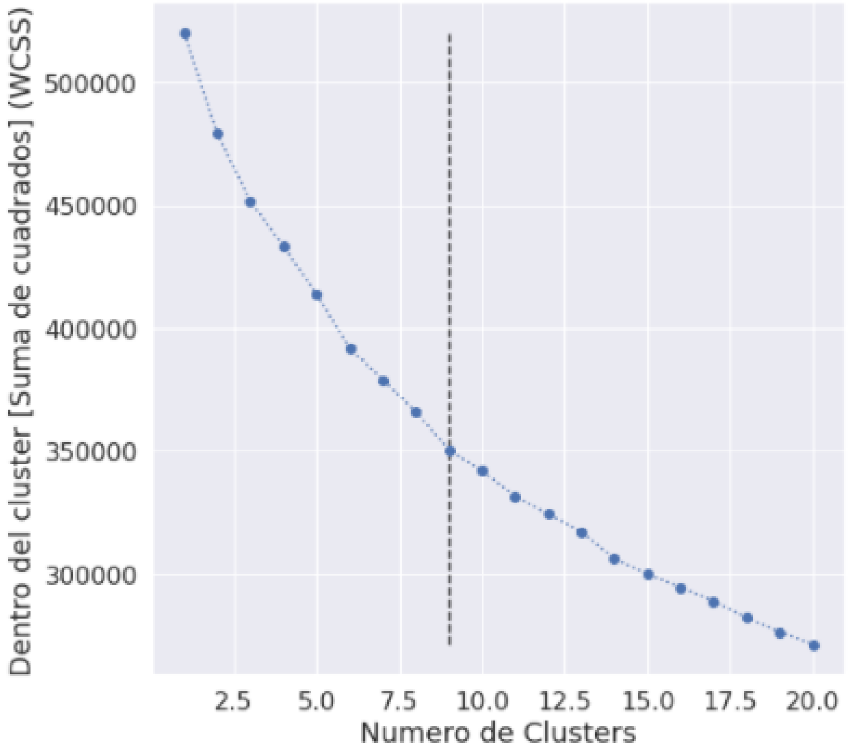


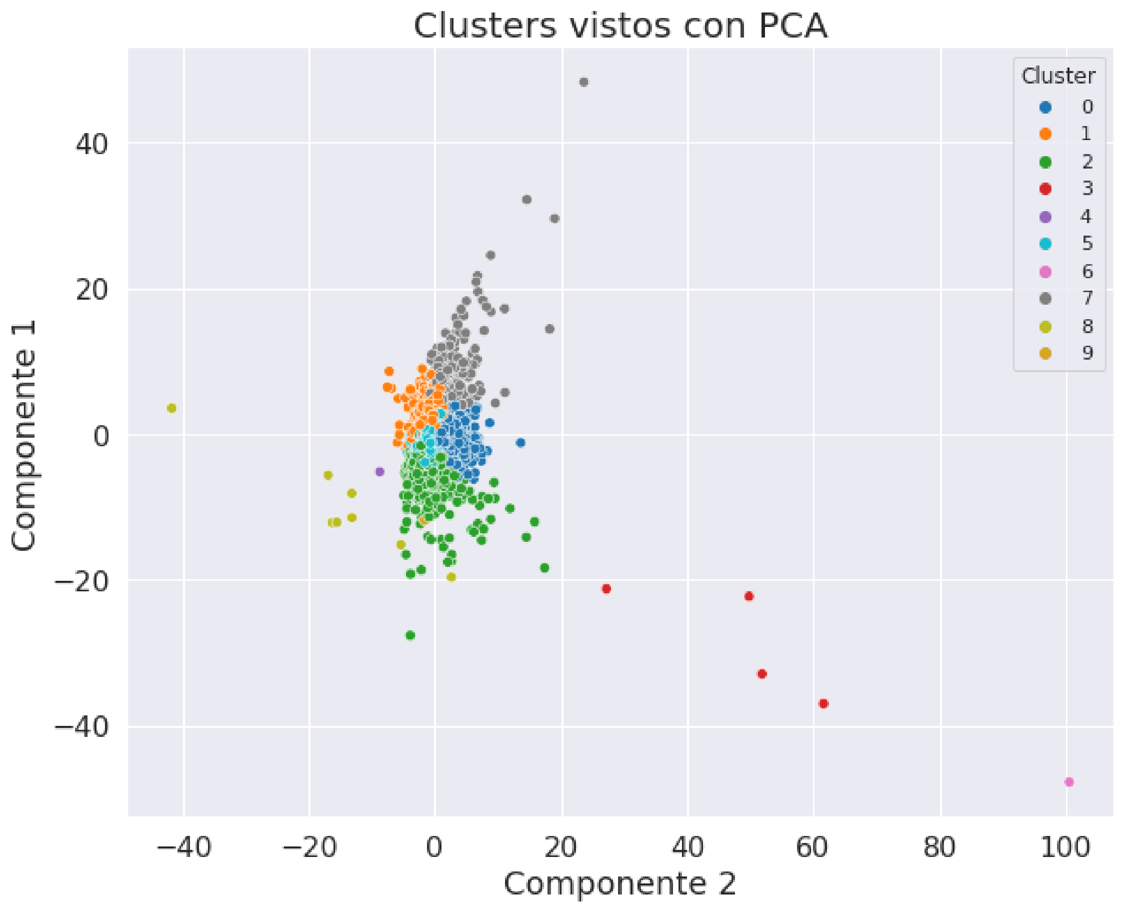
MODELO DE MACHINE LEARNING
COMPANY BANKRUPTCY PREDICTION

```
# Ploteando grafico de Componentes principales
fig = plt.figure(figsize=(8,8))
plt.plot(range(1, len(X.columns)+1), evr.cumsum(), marker='o', linestyle=':
plt.xlabel('Numero de Componentes', fontsize=18)
plt.ylabel('Varianza Acumulada Explicada',fontsize=18)
plt.xticks(fontsize=16)
plt.yticks(fontsize=16)
plt.show()
```

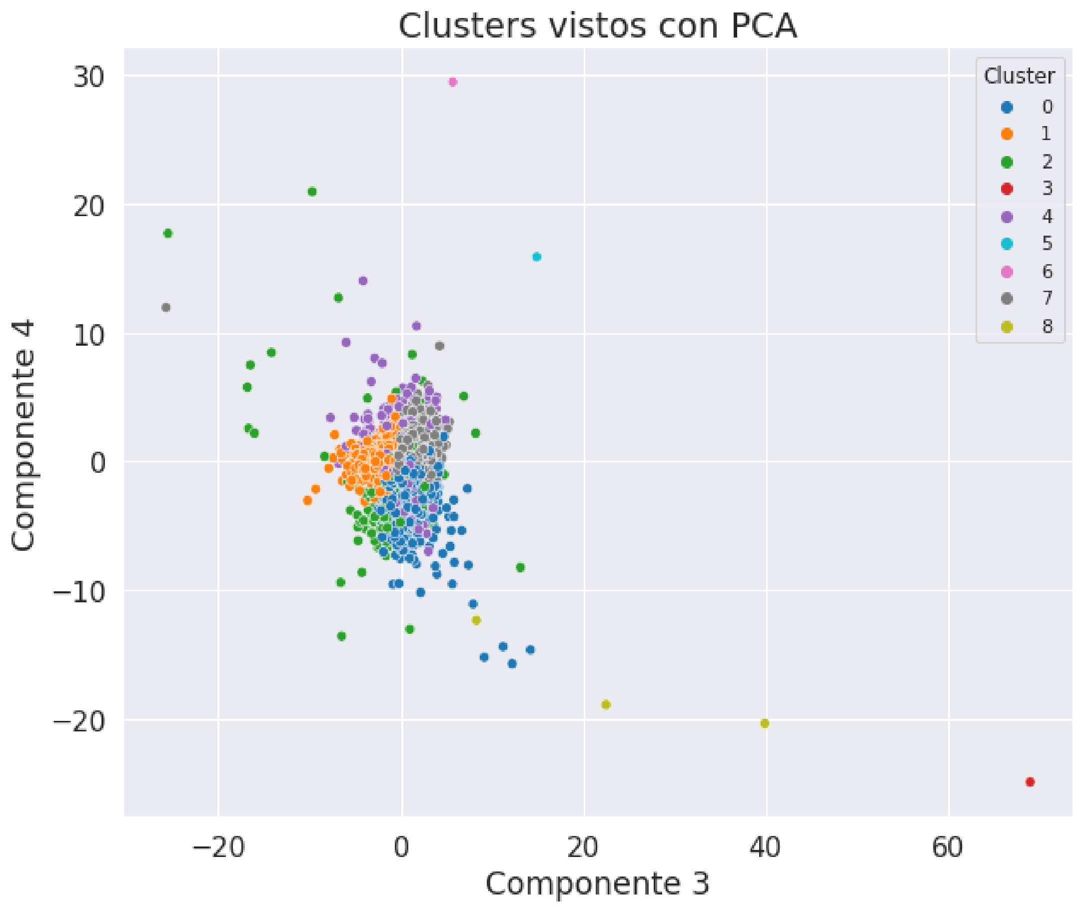



```
# Ploteando grafico
fig = plt.figure(figsize=(8,8))
plt.plot(range(1, 21), wcss, marker='o', linestyle=':')
plt.vlines(KneeLocator([i for i in range(1, max_clusters)], wcss, curve='convex',
                       direction='decreasing').knee, ymin=min(wcss), ymax=max(wcss), linestyles='dashed')
plt.xlabel('Numero de Clusters', fontsize=18)
plt.ylabel('Dentro del cluster [Suma de cuadrados] (WCSS)', fontsize=18)
plt.xticks(fontsize=16)
plt.yticks(fontsize=16)
plt.show()
```



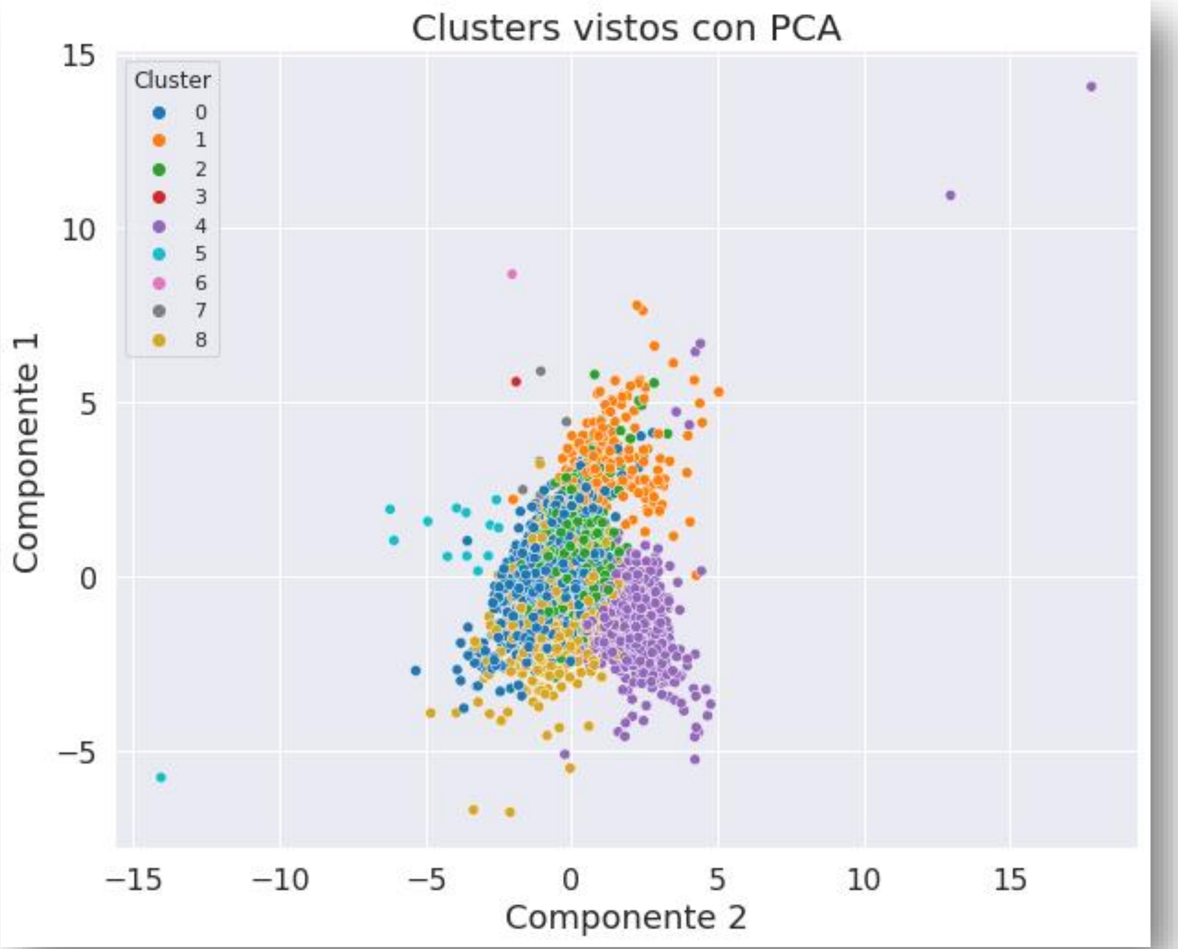


Clusters vistos con PCA
Cluster
0
1
2
3
4
5
6
7
8
9
Componente 1
Componente 2
40
20
0
−20
−40
−40
−20
0
20
40
60
80
100

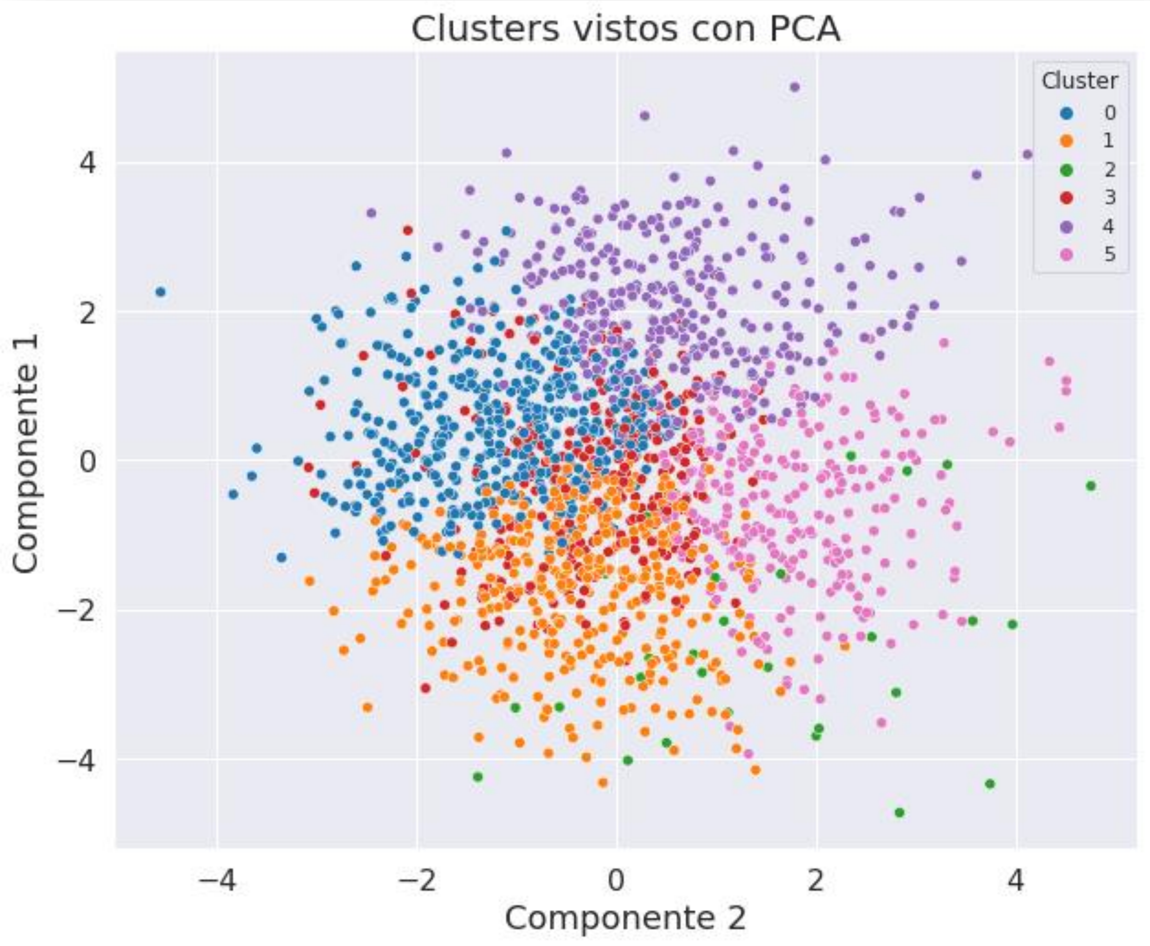


Clusters vistos con PCA
Cluster
0
1
2
3
4
5
6
7
8
Componente 4
Componente 3
30
20
10
0
−10
−20
−20
0
20
40
60

Sin Tratamiento de Atípicos



Con Tratamiento de Atípicos



Bankrupt?	ROA(C) before interest and depreciation before interest	Operating Profit Rate	Operating Profit Growth Rate	Total Asset Growth Rate	Quick Ratio	Total Asset Turnover	Accounts Receivable Turnover	Inventory Turnover Rate (times)	Current Asset Turnover Rate	Cash Turnover Rate	Cash Flow to Liability	Gross Profit to Sales	Net Income Flag

```python
#Se debe seleccionar el mejor modelo, ya que ahora se tienen 6 modelos y estimaciones de precisión para cada uno,
#por ello se necesita comparar los modelos entre sí y seleccionar el más preciso.
import warnings
warnings.filterwarnings('ignore')

resultados = []
names = []


# Si se necesita tanto el índice o nombre, así como el elemento, se usa for indice, elemento en lista
for name, model in models:
    kfold = StratifiedKFold(n_splits=10, random_state=1, shuffle=True) # Declaracion de la validación cruzada, las características
    cv_resultados = cross_val_score(model, X_train, Y_train, cv=kfold, scoring='accuracy') # genera la precisión de la validación cruzada y la guarda en la variable cv_resultados en lista
    resultados.append(cv_resultados)  # genera la precisión de la validación cruzada y la guarda en la variable cv_resultados en matrices, esto para hacer el boxplot.
    names.append(name) # names en matrices
    print('%s: %f (%f)' % (name, cv_resultados.mean(), cv_resultados.std()))
```

```
LR: 0.967920 (0.000899)
LDA: 0.965904 (0.003670)
KNN: 0.966821 (0.002366)
CART: 0.945922 (0.008950)
NB: 0.035380 (0.002721)
SVC: 0.968103 (0.000874)
```

```
LR: 0.967920 (0.000899)
LDA: 0.965904 (0.003670)
KNN: 0.966821 (0.002366)
CART: 0.948123 (0.013037)
NB: 0.035380 (0.002721)
SVC: 0.968103 (0.000874)
```

```python
# vamos a elegir el SVC
```

```python
# Haciendo predicciones y evaluación del dataset

model = SVC()
model.fit(X_train, Y_train)
prediccion = model.predict(X_test)
```

Prueba de Predicción de la Variable Dependiente

```
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(random_state=10)
classifier.fit(X_train, Y_train)
```

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                   intercept_scaling=1, l1_ratio=None, max_iter=100,
                   multi_class='auto', n_jobs=None, penalty='l2',
                   random_state=10, solver='lbfgs', tol=0.0001, verbose=0,
                   warm_start=False)
```

```
y_pred = classifier.predict(X_test)
```

```
cm = confusion_matrix(Y_test, y_pred)    # vemos una identificación de 1318  identificados correctamente, mientras que se identificaron 46 falss negativos
print(cm)
```

```
[[1318    0]
 [  46    0]]
```

```
empresa1 =  classifier.predict([[0.37,0.95,0.85,5.667,0.003,0.002,0.0126,1.0667,5.45,3.85,0.385,0]])     # predicción
print(empresa1)
```

0 Competente

1 No competente, puede caer bancarrota

```
[0.]
```

MODELO SUPERVISADO


```python
[1220] mc =pd.DataFrame(confusion_matrix(Y_test, prediccion, labels=[0,1]),
                            index = [0,1],
                            columns = [0,1])

       # Evaluando Predicciones
       print("ROC:", accuracy_score(Y_test, prediccion),sep='\n')
       print("")
       print("Matriz de Confusión:", mc,sep='\n')
```

```
ROC:
0.9633431085043989

Matriz de Confusión:
      0   1
0  1307  11
1    39   7
```

Por medio del uso de modelos de Machine Learning
Se pueden evaluar las variables que influyen en la determinación o no de la posibilidad que una empresa presente bancarrota.

Estas posibilidades que se aplican a la determinación de las posibilidades financieras de una empresa, pueden aplicarse a individuos en procesos de clasificación o aprobación de créditos.

MODELO DE MACHINE LEARNING
COMPANY BANKRUPTCY PREDICTION

REFERENCIAS

Para referencia de información (https://isslab.csie.ncu.edu.tw/download/publications/1.pdf) Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study

Definiciones de ratios o razones financieras

Ratios de Rentabilidad : Return on Assets (ROA), Return on Equity (ROE), Return on Investment (ROI), Return on Invested Capital (ROIC), EBITDA Margin, Net Profit Margin, Operating Margin.

(https://www.investopedia.com/ask/answers/031215/what-formula-calculating-return-assets-roa.asp)

Ratios de Liquidez: Current Ratio, Quick Ratio, Cash Ratio, Operating Cash Flow Ratio, Receivables Turnover Ratio, Inventory Turnover

https://www.investopedia.com/terms/q/quickratio.asp

Ratios de Solvencia: Debt-To-Equity Ratio, Total-Debt-to-Total-Assets Ratio, Interest Coverage Ratio, Shareholder Equity Ratio

https://www.investopedia.com/terms/i/interestcoverageratio.asp

Clasificaciones de los distintos ratios financieros

https://corporatefinanceinstitute.com/resources/knowledge/finance/financial-ratios/