

# RiesKo

Anticipa el riesgo, protege el futuro

Proyecto - Bootcamp IA

Presentado por

Catalina Zapata

Manuel Guzmán

# Equipo



**Catalina Zapata**  
Ing. Industrial



**Manuel Guzmán**  
Ing. Mecánico

# Contenido



1 Introducción



2 Análisis exploratorio



3 Experimentos



4 Análisis resultados

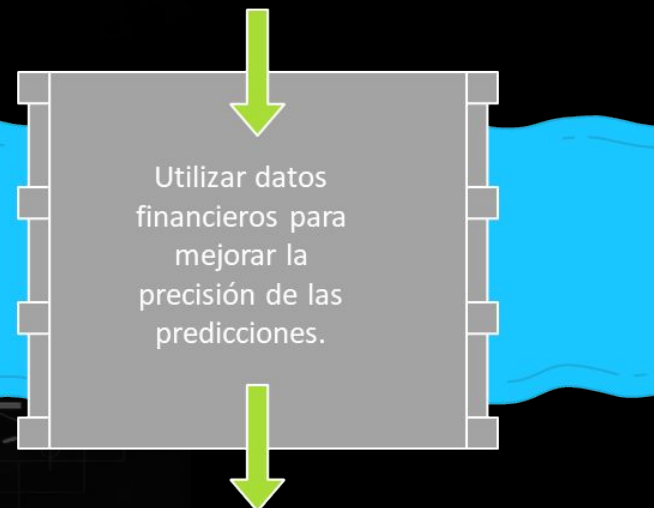


5 Conclusiones



# Introducción

Los inversionistas  
enfrentan altos  
riesgos  
financieros.



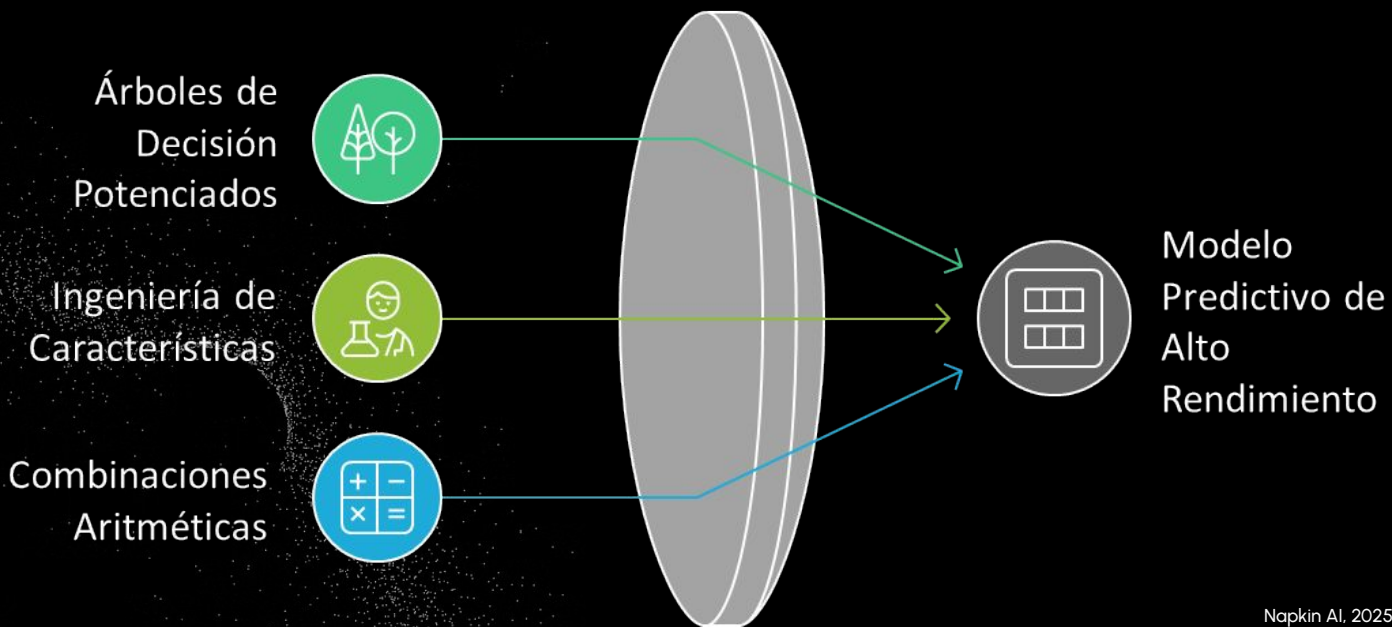
Utilizar datos  
financieros para  
mejorar la  
precisión de las  
predicciones.

Decisiones  
informadas  
reducen los  
riesgos  
financieros.

Napkin AI, 2025

# Introducción

## Construyendo un Modelo Predictivo Avanzado



Napkin AI, 2025

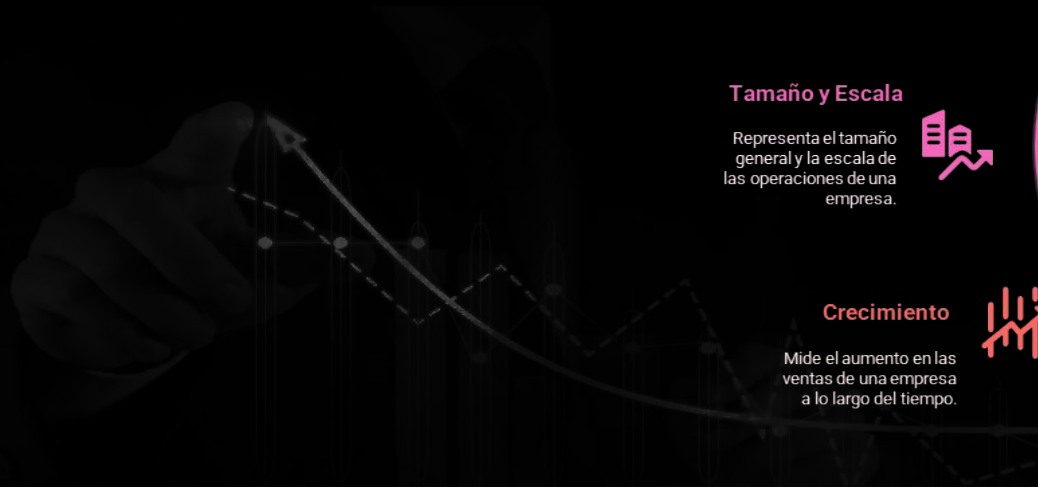
# Objetivo

Predecir la quiebra de empresas, mediante algoritmos de Machine Learning, ayudando a tomar mejores decisiones y reducir riesgos.





# Variables econométricas



## Tamaño y Escala

Representa el tamaño general y la escala de las operaciones de una empresa.



## Crecimiento

Mide el aumento en las ventas de una empresa a lo largo del tiempo.



## Apalancamiento

Muestra el grado en que una empresa utiliza deuda.



## Rentabilidad

Mide la capacidad de una empresa para generar ganancias.

## Liquidez

Evalúa la capacidad de una empresa para cumplir con sus obligaciones a corto plazo.



## Eficiencia

Determina qué tan bien una empresa utiliza sus activos.



# ¿Por qué tener cuidado al aplicarlas directamente?



Implementación consistente de NIF



Empresas formales más grandes



Tasas de interés estables



Variaciones en la implementación de NIF



Empresas informales más pequeñas



Inflación volátil



Polonia



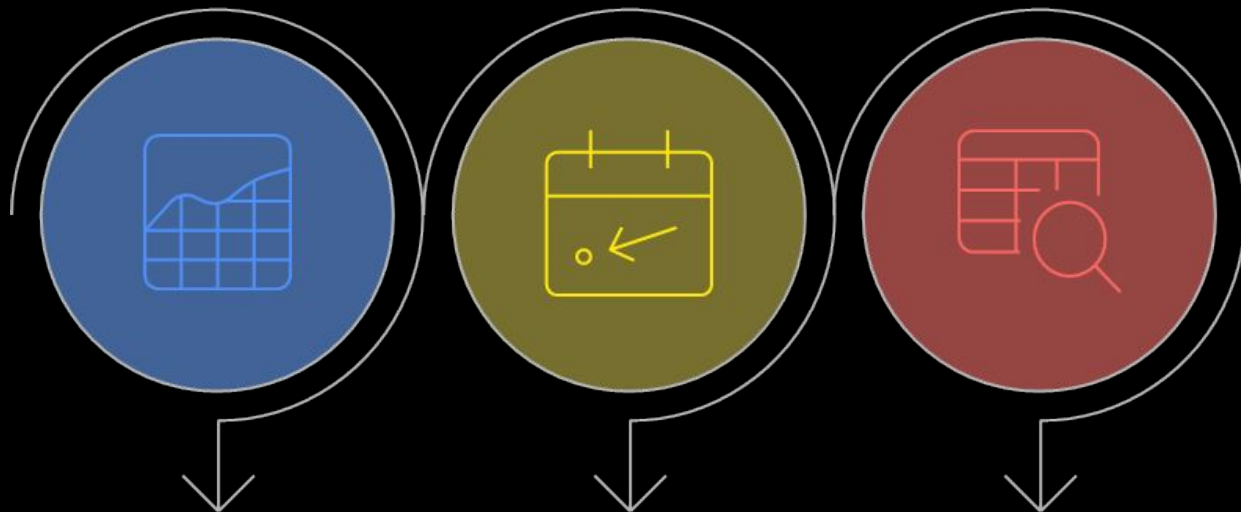
Colombia

Comparación de factores de riesgo de quiebra en Polonia y Colombia





# Data set



## Uso de datos

Evaluación de  
modelos  
predictivos para  
datos tabulares.

## Año

Datos limitados al  
año 2010

## Anomalía

Características  
creadas a través  
de la ingeniería de  
características.

# Análisis exploratorio

Información de los datos

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5910 entries, 0 to 5909
Data columns (total 65 columns):
#   Column
---  ---
0   Utilidad neta sobre activos totales
1   Pasivos totales sobre activos totales
2   Capital de trabajo sobre activos totales
3   Activos corrientes sobre pasivos a corto plazo
4   Días de liquidez
5   Utilidades retenidas sobre activos totales
6   EBIT sobre activos totales
7   Valor contable del patrimonio sobre pasivos totales
8   Ventas sobre activos totales
9   Patrimonio sobre activos totales
10  Utilidad extendida sobre activos totales
11  Utilidad bruta sobre pasivos a corto plazo
12  Utilidad bruta más depreciación sobre ventas
13  Utilidad bruta más intereses sobre activos totales
14  Días de pasivos
15  Utilidad bruta más depreciación sobre pasivos totales
16  Activos totales sobre pasivos totales
17  Utilidad bruta sobre activos totales
18  Utilidad bruta sobre ventas
```

Non-Null Count	Dtype
5907 non-null	float64
5907 non-null	float64
5907 non-null	float64
5889 non-null	float64
5899 non-null	float64
5907 non-null	float64
5907 non-null	float64
5892 non-null	float64
5909 non-null	float64
5907 non-null	float64
5907 non-null	float64

```
# ver los datos faltantes de cada variable
dataset.isnull().sum()
```

	0
Utilidad neta sobre activos totales	3
Pasivos totales sobre activos totales	3
Capital de trabajo sobre activos totales	3
Activos corrientes sobre pasivos a corto plazo	21
Días de liquidez	11
...	...
Rotación de cuentas por cobrar	15
Días de pasivos a corto plazo	0
Ventas sobre pasivos a corto plazo	21
Ventas sobre activos fijos	107
Quiebra de la empresa (Sí/No)	10

65 rows x 1 columns

# Análisis exploratorio

Codificación de la variable categórica

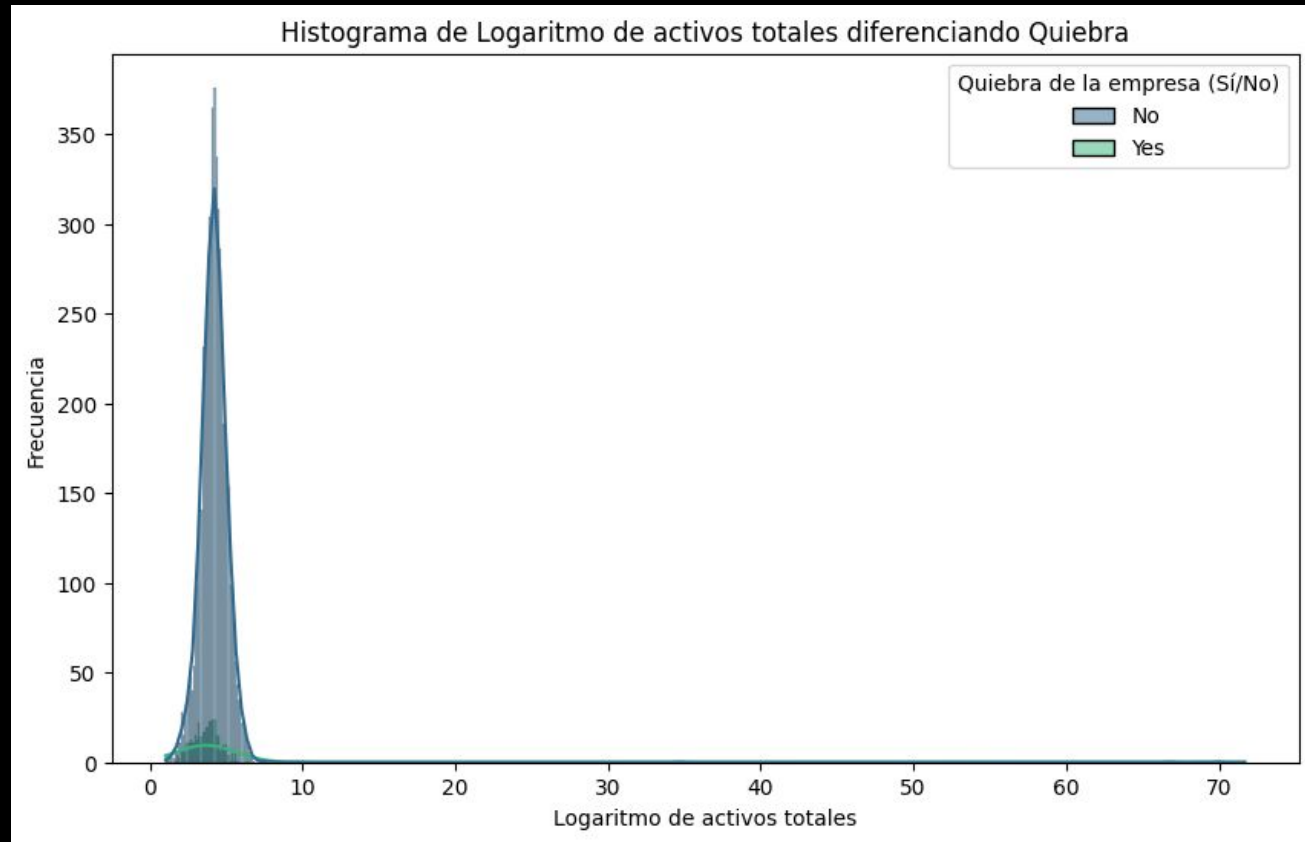
Ventas sobre activos totales	Patrimonio sobre activos totales	...	Razón de liquidez ajustada	Costos totales sobre ventas totales	Pasivos a largo plazo sobre patrimonio	Rotación de inventario	Rotación de cuentas por cobrar	Días de pasivos a corto plazo	Ventas sobre pasivos a corto plazo	Ventas sobre activos fijos	Quiebra de la empresa (Sí/No)	Quiebra
1.1324	39.883	...	21.455	88.3110	1.4171	80.9820	2.7538	13.478	27.0820	1.5454	No	0
1.0573	3.363	...	21.247	94.5820	1.0190	86.3110	3.1653	91.824	3.9750	8.6122	No	0
1.6743	36.696	...	19.682	99.5850	0.0000	5.0694	8.2175	138.010	2.6448	3.6015	No	0
1.2601	35.268	...	5.006	98.1730	16.6680	2.0090	15.8810	165.930	2.1998	4.3415	No	0
2.8297	80.558	...	25.097	91.0390	14.0040	18.1960	11.0470	23.623	15.4510	8.1001	No	0
1.0892	85.664	...	76.759	91.8130	0.0000	4.7479	5.1122	54.823	6.6578	1.5686	No	0
61.0890	91.612	...	15.598	77.7350	0.0000	4.5950	5.2391	50.119	7.2826	6.6321	No	0

# Análisis exploratorio

Histograma de:

Logaritmo de  
activos totales

Tamaño/estructura

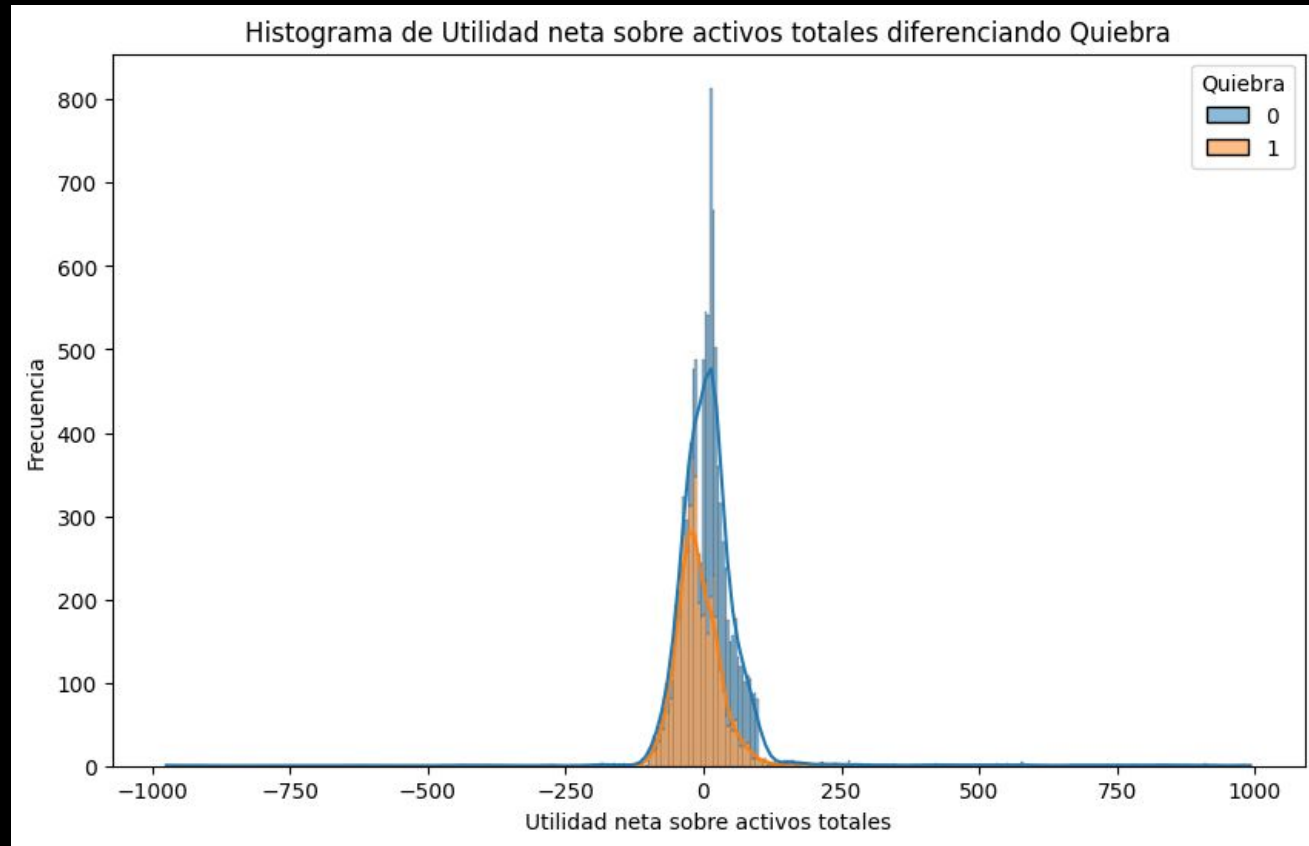


# Análisis exploratorio

Histograma de:

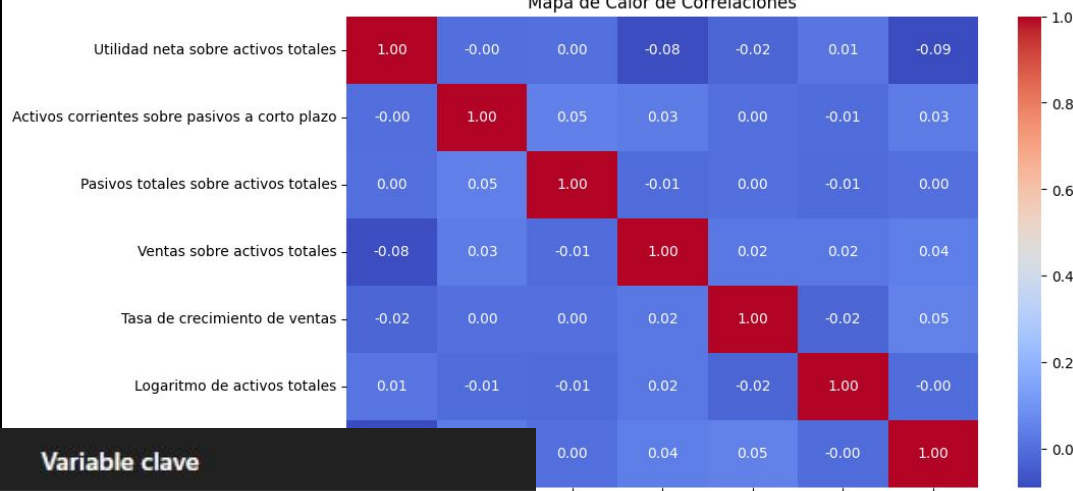
Utilidad neta /  
activos totales

Rentabilidad



# Análisis exploratorio

Selección de las variables más representativas



Dimensión	Variable clave
Rentabilidad	Utilidad neta / activos totales
Liquidez	Activos corrientes / pasivos a corto plazo
Apalancamiento	Pasivos totales / activos totales
Eficiencia operativa	Ventas / activos totales
Crecimiento	Tasa de crecimiento de ventas
Tamaño / estructura	Logaritmo de activos totales



Matriz de  
correlación de  
todas las  
variables



# Análisis exploratorio

Imputación de valores faltantes usando la media

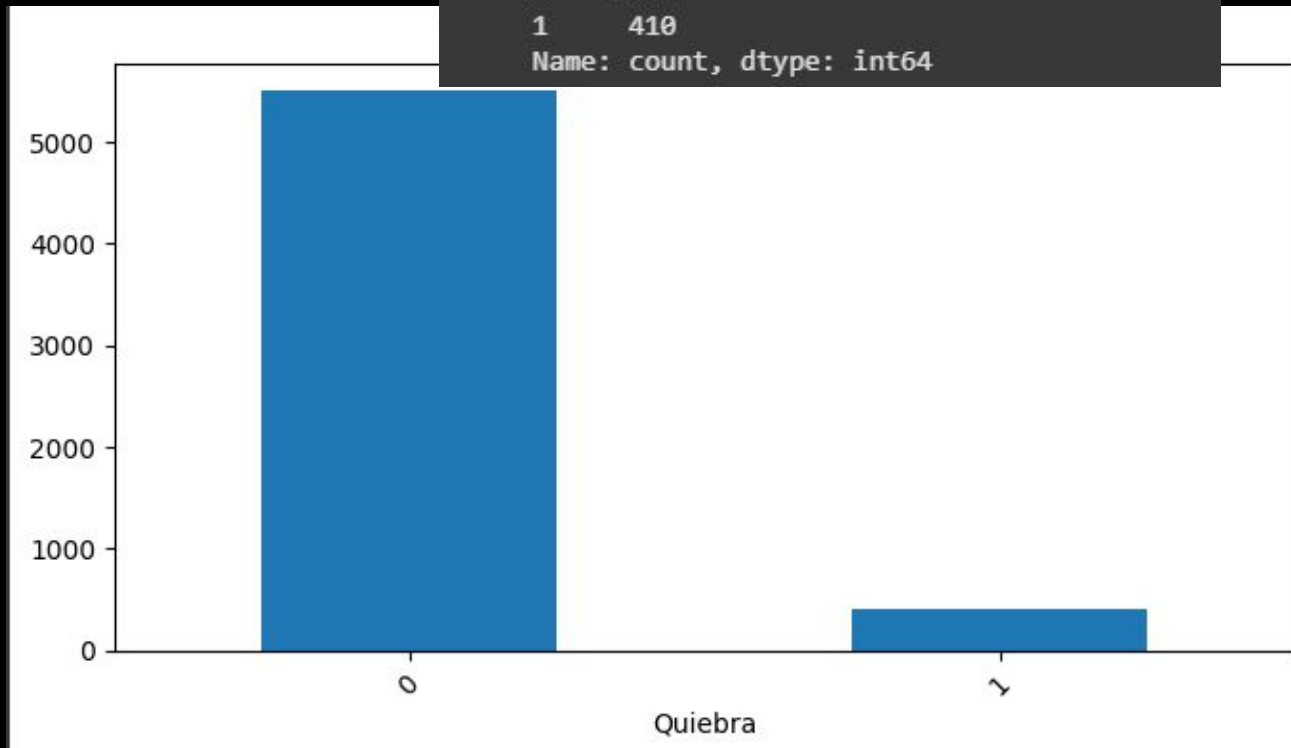
Data columns (total 66 columns):

#	Column	Non-Null Count	Dtype
0	Utilidad neta sobre activos totales	5910 non-null	float64
1	Pasivos totales sobre activos totales	5910 non-null	float64
2	Capital de trabajo sobre activos totales	5910 non-null	float64
3	Activos corrientes sobre pasivos a corto plazo	5910 non-null	float64
4	Días de liquidez	5910 non-null	float64
5	Utilidades retenidas sobre activos totales	5910 non-null	float64
6	EBIT sobre activos totales	5910 non-null	float64
7	Valor contable del patrimonio sobre pasivos totales	5910 non-null	float64
8	Ventas sobre activos totales	5910 non-null	float64
9	Patrimonio sobre activos totales	5910 non-null	float64
10	Utilidad extendida sobre activos totales	5910 non-null	float64
11	Utilidad bruta sobre pasivos a corto plazo	5910 non-null	float64
12	Utilidad bruta más depreciación sobre ventas	5910 non-null	float64
13	Utilidad bruta más intereses sobre activos totales	5910 non-null	float64
14	Días de pasivos	5910 non-null	float64
15	Utilidad bruta más depreciación sobre pasivos totales	5910 non-null	float64



# Análisis exploratorio

Desbalanceo  
de datos

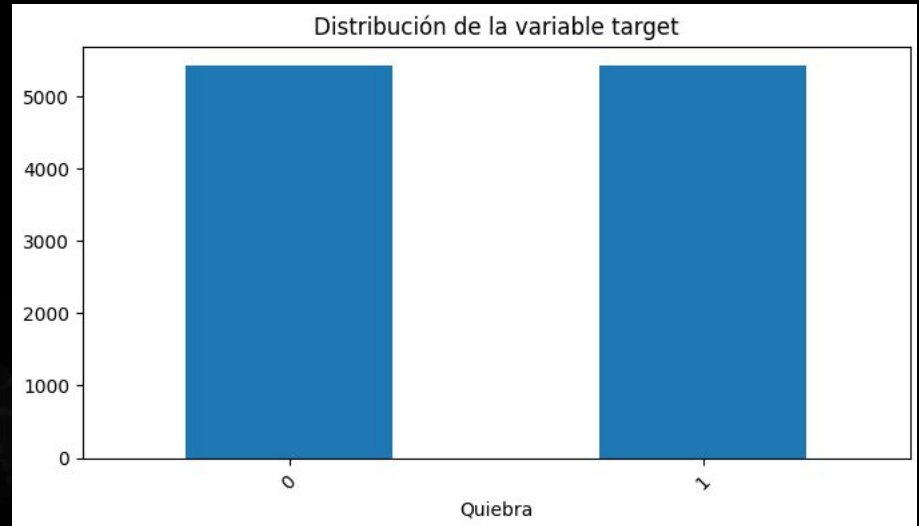


# Análisis exploratorio

Balanceo de datos:

SMOTE y Tomek

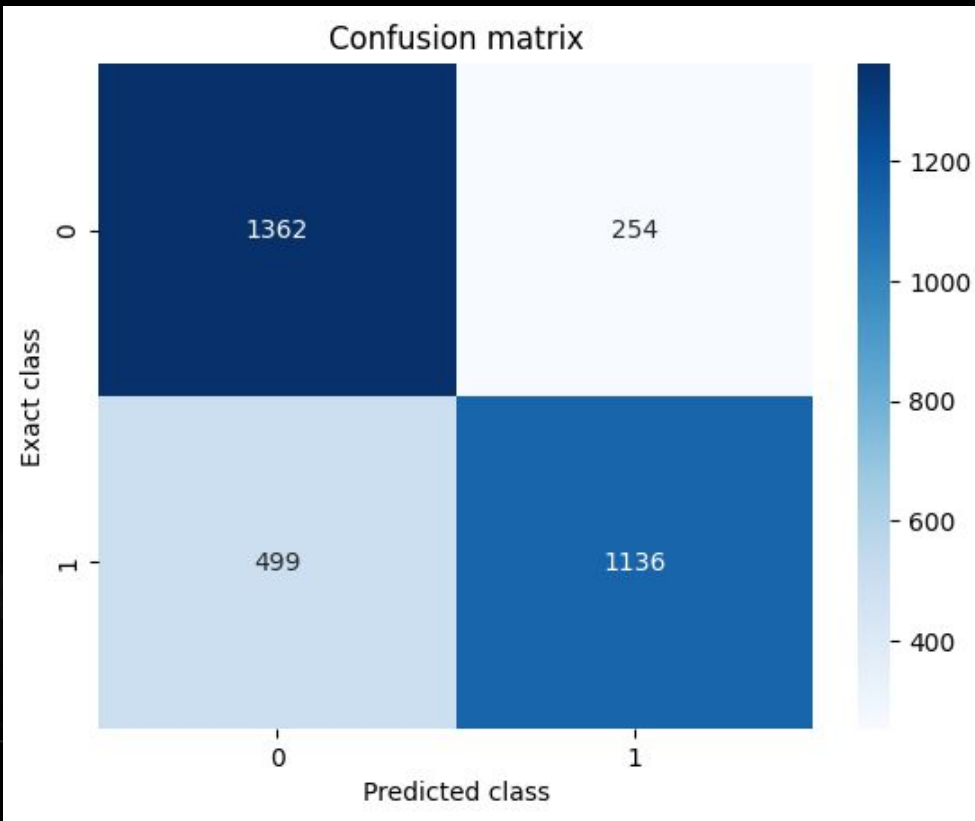
```
⇒ Número de muestras después de SMOTE + Tomek: 10836  
Distribución después de SMOTE + Tomek:  
Quiebra  
0      5418  
1      5418  
Name: count, dtype: int64
```



# Logistic regression

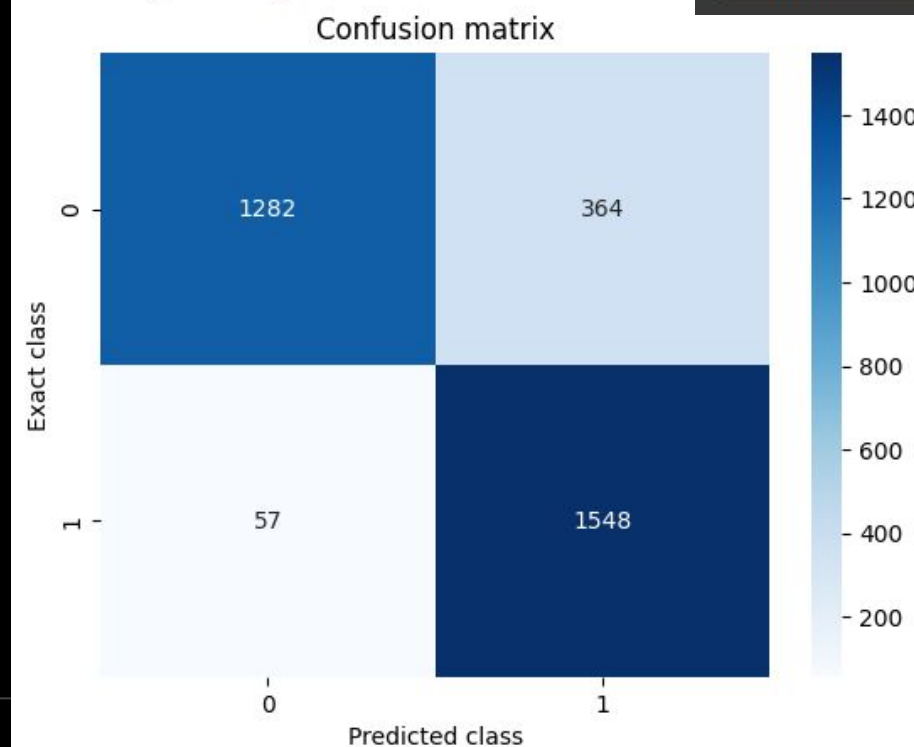
```
LogisticRegression  
LogisticRegression(max_iter=1000)
```

- Accuracy: 0.7665
- Precision (PPV): 0.8070
- Recall (Sensibilidad, TPR): 0.6928
- Specificity (TNR): 0.8384
- F1-Score: 0.7456



# KNeighborsClassifier

The accuracy of KNeighborsClassifier is 0.8705

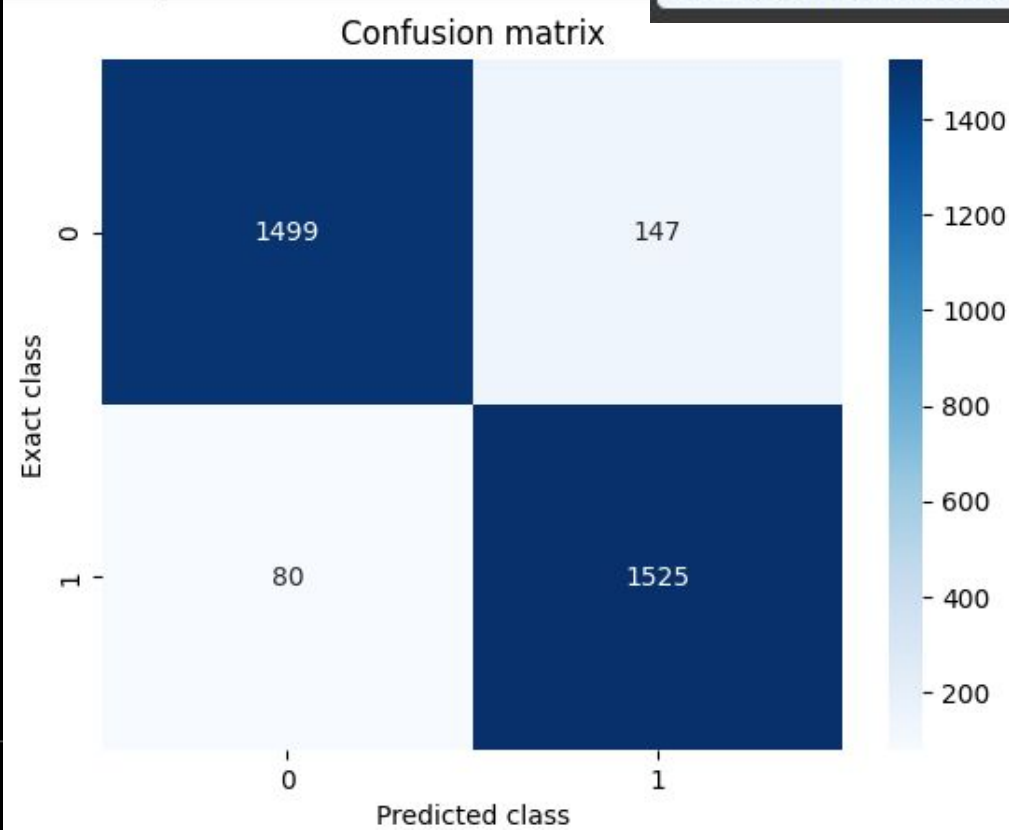


▼ KNeighborsClassifier ⓘ ?

KNeighborsClassifier()

# Árboles de decisiones

The accuracy of Decision Tree is 0.9302



DecisionTreeClassifier ⓘ ⓘ

```
DecisionTreeClassifier()
```

# Random Forest

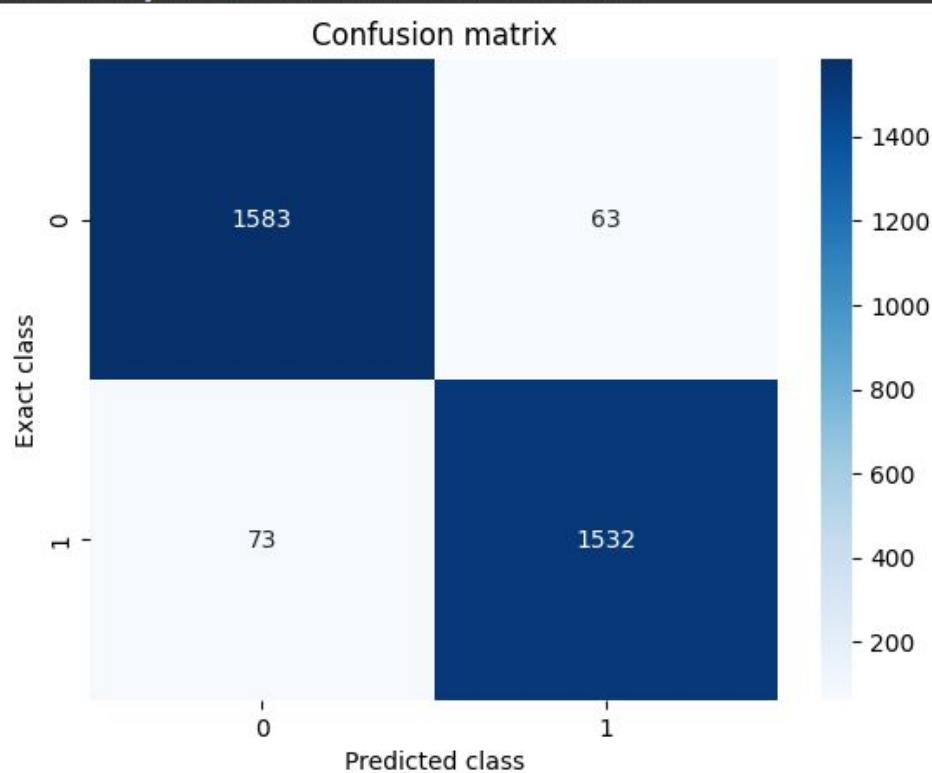
## Métricas de Clasificación para Random Forest:

- ◆ Accuracy: 0.9662
- ◆ Precision (PPV): 0.9655
- ◆ Recall (Sensibilidad, TPR): 0.9679
- ◆ Specificity (TNR): 0.9644
- ◆ F1-Score: 0.9667

RandomForestClassifier

```
RandomForestClassifier(n_estimators=10)
```

The accuracy of RandomForestClassifier is 0.9582



# Conclusiones

- El modelo Random Forest permite tener una precisión del 96% luego de imputar los datos faltantes usando la media y realizando el balanceo de los datos.
- Predecir la quiebra de empresas en Colombia permite proteger el empleo, prevenir crisis económicas, mejorar la toma de decisiones financieras, fortalecer la economía nacional y promover una gestión empresarial más responsable y sostenible.
- El modelo presentado no está limitado al caso de empresas polacas, por lo que podría adaptarse a otros contextos empresariales y económicos.

# Trabajo a futuro

- Se debe tener en cuenta la estandarización de las variables con diferentes unidades.
- Se debe comprobar la normalidad de las variables.
- Uso de modelos como XGBoost para datos no balanceados.



# Referencias

Zieba, M., Tomczak, S., Tomczak, J. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction, Expert Systems With Applications, 58 (2016).. DOI: <http://dx.doi.org/10.1016/j.eswa.2016.04.001>

OPEN ML, <https://www.openml.org/search?type=data&sort=date&status=any&id=46950>, 2025

Napkin.ai, 2025.

Notebooklm.google.co, 2025

# Gracias

Github link: <https://github.com/maguzmansan/Proyecto-TalentoTech/blob/main/README.md>