



UNIVERSIDAD
AUSTRAL

**CIENCIAS
EMPRESARIALES**

Trabajo Práctico Big Data

INTEGRANTES:

Solana Brun

Delfina Linck

Pilar Solorzano

EQUIPO BIG DREAMS

Fecha de entrega: última clase

Análisis y Modelado de Datos con Orange Data Mining

Objetivo: Utilizar Orange Data Mining para llevar a cabo un proyecto completo de análisis y modelado de datos, desde la definición del problema hasta la interpretación de las predicciones.

Consigna:

1. Selección del Dataset asignado: revisá según tu equipo que dataset tenes que utilizar.

Esta base de datos de Kaggle se usa para **predecir si una empresa va a quebrar o no**, a partir de distintos indicadores financieros.

Reúne información de muchas compañías y muestra qué tan sanas estaban en términos de liquidez, endeudamiento y rentabilidad.

El **desafío** es que la mayoría de las empresas del conjunto no quebraron, así que los datos están desbalanceados. Sirve mucho para armar modelos que ayuden a bancos o inversores a anticipar riesgos, y con un análisis exploratorio se pueden ver patrones claros: las empresas con menos liquidez, más deudas o baja rentabilidad suelen estar más cerca de la quiebra.

2. Definición del Problema: deberán definir un problema específico de negocio o investigación que quieran resolver. Esta definición debe incluir una clara formulación de qué quieren predecir o clasificar y por qué es importante o

relevante.

Problema: Las instituciones financieras, como bancos y aseguradoras, enfrentan un **gran riesgo** al otorgar créditos a empresas que luego podrían quebrar. Predecir con anticipación la probabilidad de quiebra de una compañía es crucial para reducir pérdidas económicas

Queremos **predecir si una empresa quebrará o no** (variable binaria: 0 = no quiebra, 1 = quiebra), utilizando ratios financieros históricos como predictores.

3. Análisis Exploratorio de Datos:

- Realicen un análisis exploratorio inicial para familiarizarse con los datos.

Lo primero que observamos es que el dataset tiene 6819 empresas y 96 variables predictivas, entendemos que son ratios financieros.

La variable principal es **"Bankrupt"**, que nos explica si quebró o no la empresa. Podemos ver que la gran mayoría de las empresas no quiebran (aprox 97 %), mientras que solo un pequeño grupo sí lo hizo (3 %).

Luego, vimos que los datos provienen de registros financieros de empresas de Taiwán, recopilados en un período de alrededor de 10 años.

Vemos que el uso principal es la investigación y desarrollo de modelos de machine learning que busquen predecir riesgo de quiebra, un problema clásico en finanzas.

- Planteen hipótesis sobre qué variables creen que son más relevantes para predecir el resultado de interés.

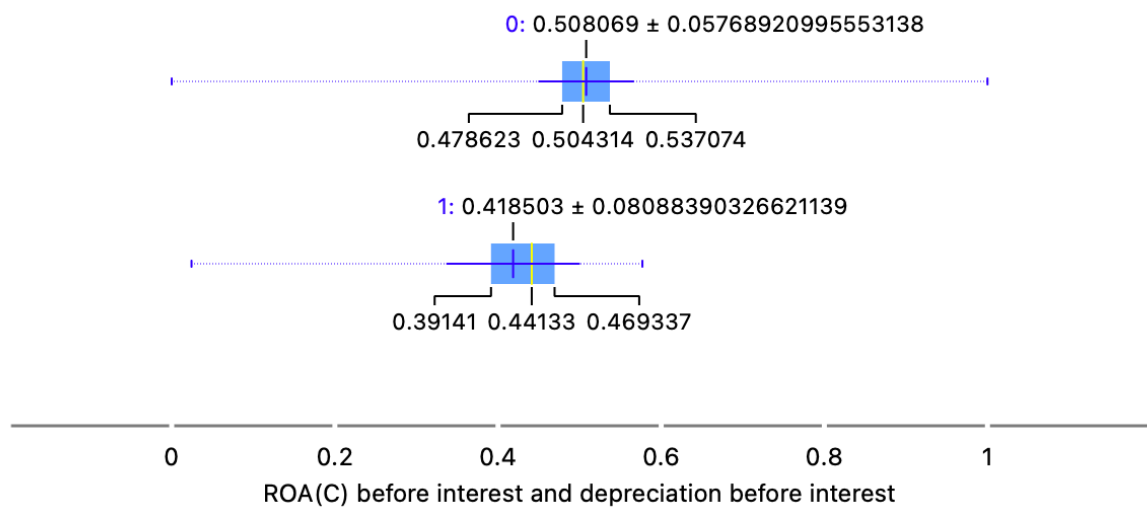
En este conjunto de datos, es probable que las variables relacionadas con la **liquidez**, como el **Current Ratio** o el **Quick Ratio**, tengan un papel importante, ya que las empresas con menor liquidez suelen estar más expuestas a la quiebra.

También resultan críticas las variables de **endeudamiento**, por ejemplo el **Debt Ratio** o el **Interest Coverage Ratio**, porque un mayor nivel de deuda puede aumentar significativamente el riesgo financiero.

Finalmente, los indicadores de **rentabilidad**, como el **ROA**, el **ROE** o el **Net Income to Total Assets**, son útiles para diferenciar a las empresas que se encuentran en una situación saludable de aquellas que ya muestran señales de problemas.

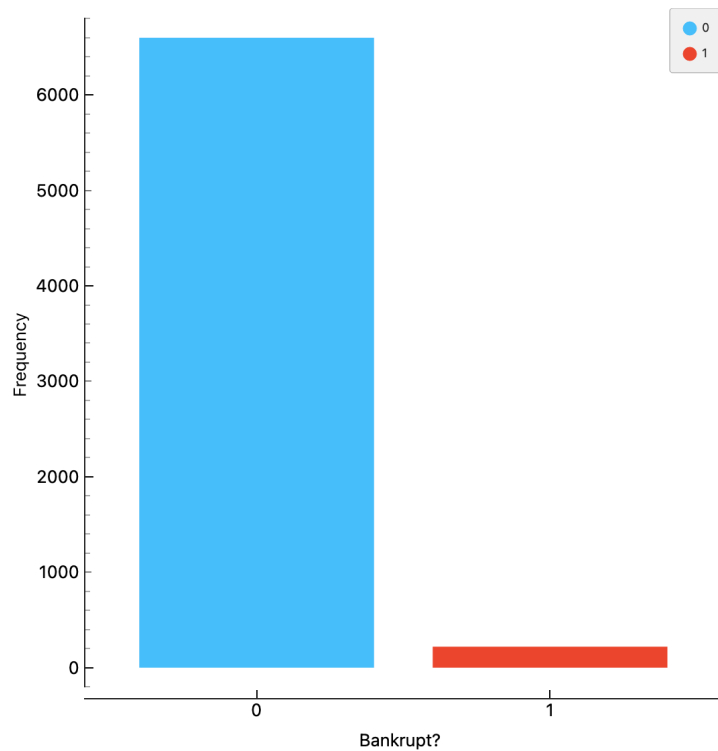
- Utilicen gráficos y estadísticas para apoyar estas hipótesis y documenten sus hallazgos.

Gráfico 1: Box plot del ROA según bankruptcy



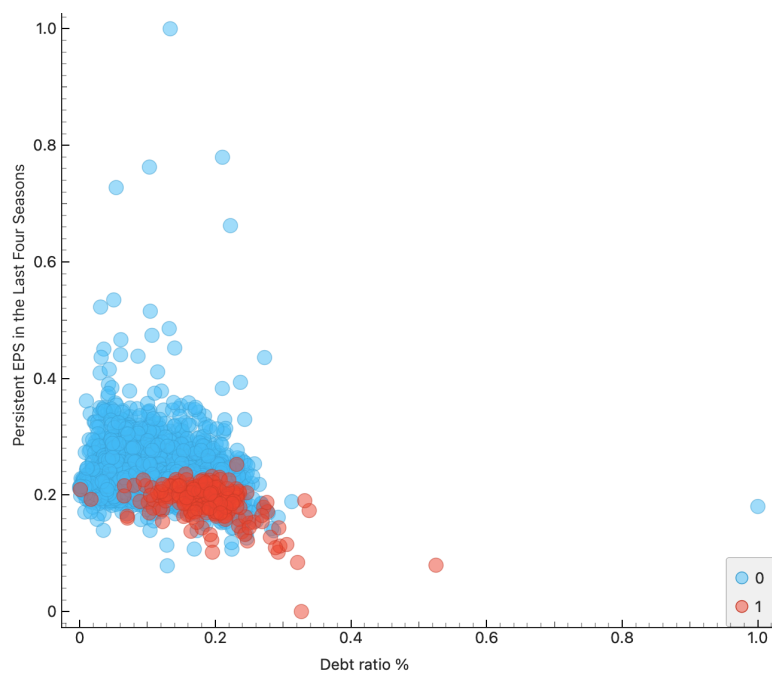
Hallazgos: El ROA es un factor que puede llegar a predecir si una empresa puede entrar en quiebra ya que hay un claro patrón que demuestra que el ROA en las empresas en quiebra es en promedio más bajo que el ROA en las empresas que no quebraron.

Gráfico 2: Gráfico de barras de cantidad de casos con bankruptcy



Hallazgos: En su mayoría, las empresas que fueron evaluadas en esta base de datos son empresas que no entraron en quiebra

Gráfico 3: Scatter plot de la persistencia del Equity por acción y el porcentaje de deuda



Hallazgos: Las empresas que no entraron en quiebra suelen tener un Equity por acción sostenido en las últimas 4 estaciones mucho más alto