

PWC

CHALLENGE

WWW.PWC.COM



CONTENIDO

01

CONTEXTO INICIAL

02

¿QUÉ PROBLEMAS ENCONTRAMOS?

03

SOLUCIÓN (MODELO)
¿POR QUÉ ELEGIMOS EL MODELO?

04

VALIDEZ DEL MODELO

05

CONCLUSIÓN



Documentación del proyecto
GitHub

CONTEXTO INICIAL

¿QUÉ PROBLEMAS ENCONTRAMOS?



5 Variables que entregan valor para identificar Fraude ¿tarjeta de crédito?



Mayores transacciones realizadas a través de Transfer y Cash out.



Cash out 50,1%(4116) y Transfer 49,9% (4097) presentan el 100% de los casos de Fraude.



0.13% registros clasificados como Fraude de 6362620 registros: **Descompensación.**



6 millones de registros: **Problemas computacionales.**



Falta de **contexto y origen** de la información recogida en la BBDD.

- *La variable `IsFlaggedFraud` (representación casi nula)*
- *Gran desviación (eliminación de datos outliers)*
- *A mayor amount, mayor probabilidad de fraude*



○ ○ ○ ○ **SOLUCIÓN**

¿QUÉ SOLUCIÓN OFRECEMOS?

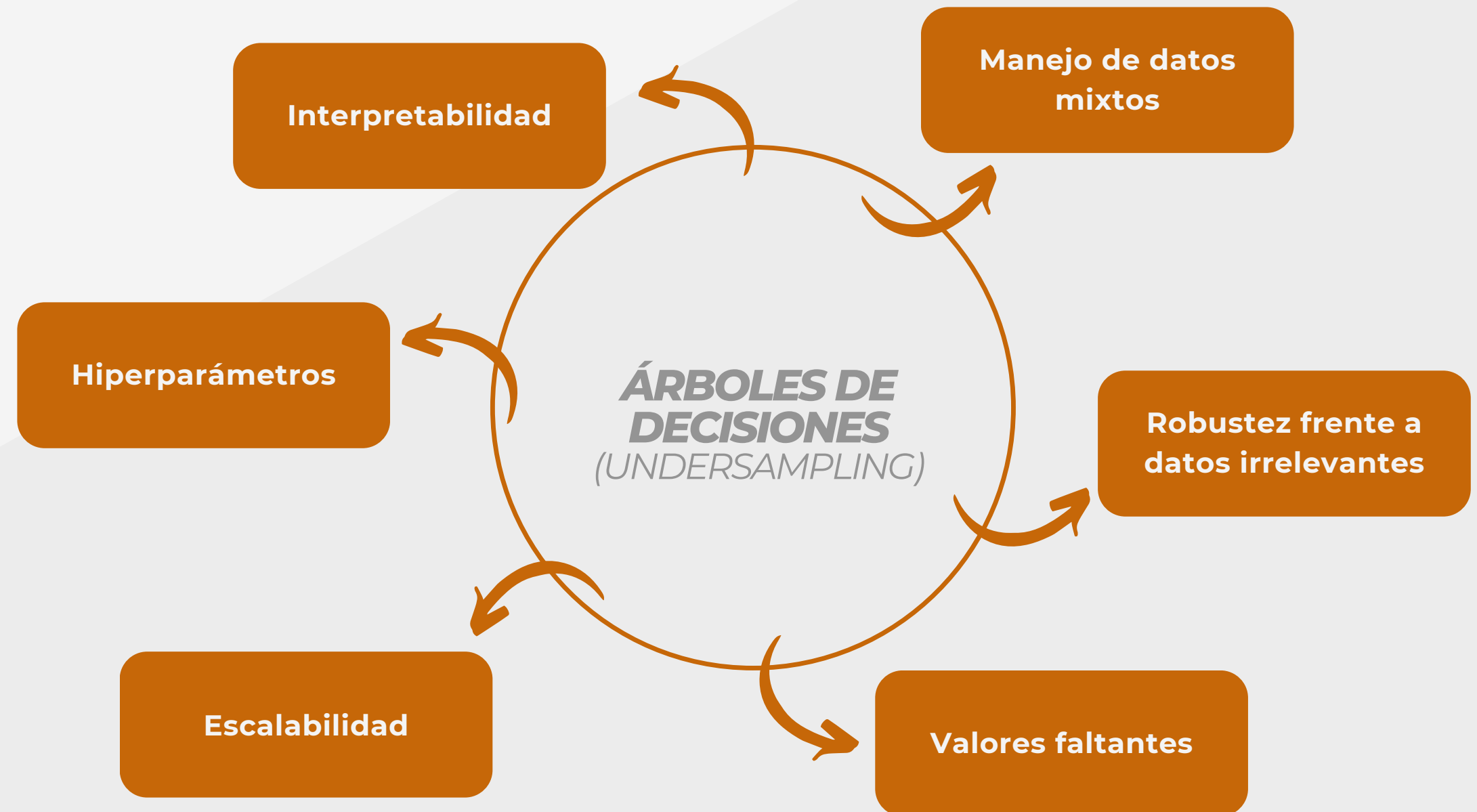
Elección de modelo de clasificación:

- Random Forest
- Decision Tree
- Decision Tree Balanced

Random Undersampling:

- 8213 Fraudes
- 8213 No Fraudes

Modelo de aprendizaje supervisado



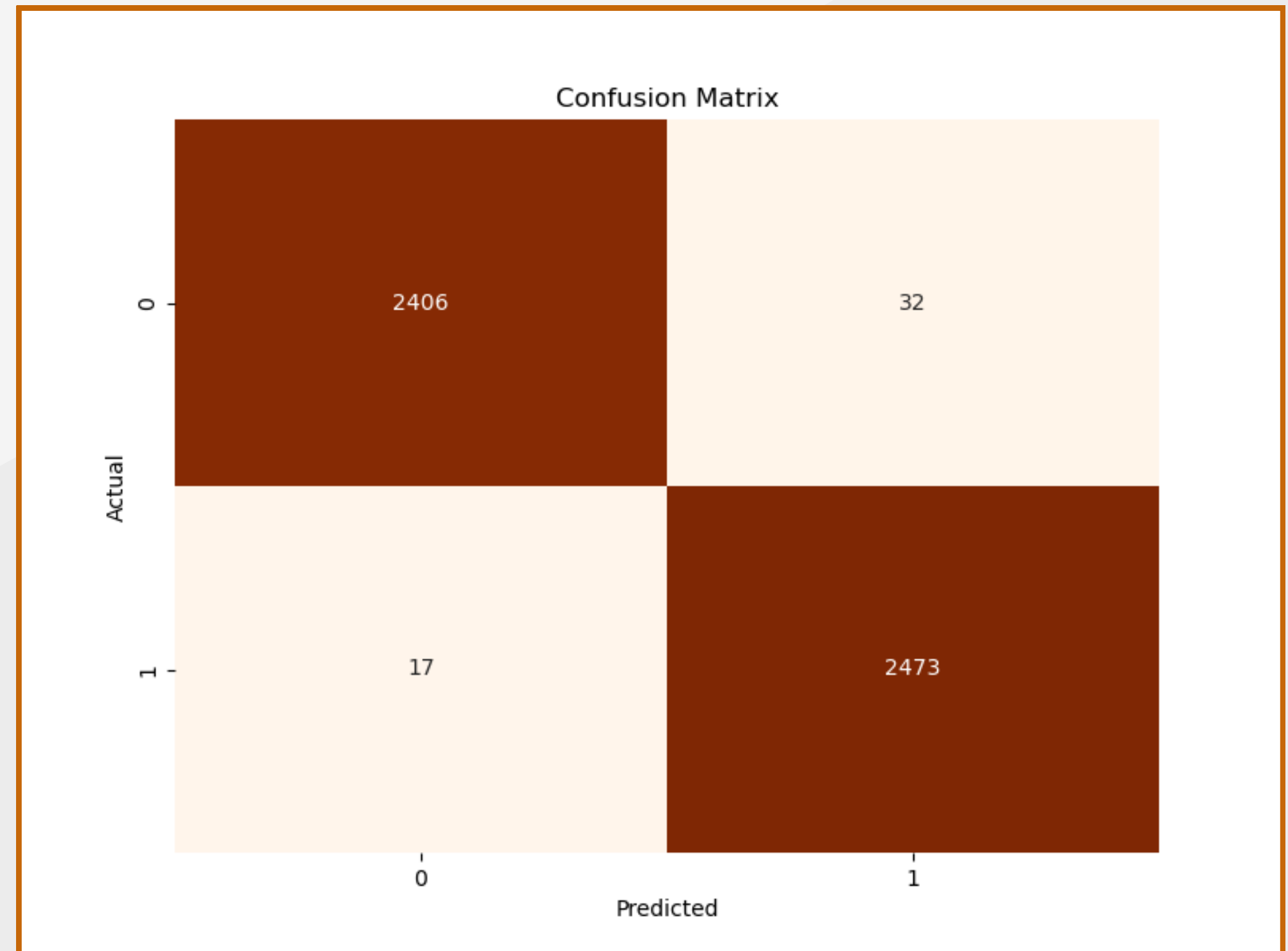
VALIDACIÓN DEL MODELO

¿POR QUÉ ESTA

SOLUCIÓN ES BUENA?

Otros parámetros de validación del modelo:

- AUC: 96.17%
- Sensibilidad: 99.43%
- Precisión: 96.20%





CÓDIGO Y CONCLUSIÓN



- Estudio del dato
- Distribución de variables
- Correlaciones

ETL + EDA

INGENIERÍA DE
CARACTERÍSTICAS

- Selección de variables
- Creación nuevas características
- Balanceo de datos (Undersampling)

MODELING

- Selección Modelo

VALIDATION

- Evaluación Modelo

APPLICATION

NUESTRA SOLUCIÓN

EL FRAUDE NOS TEME

MUCHAS GRACIAS POR SU TIEMPO

