

PwC Challenge

**Reto: Detección de fraudes
en tarjetas de crédito**

Grupo 4

Noemi Luengo
Jerika Castellero
Jorge Morales
Salomea Stepien





Índice

Detección y Predicción del Fraude 03

Detalles del Reto y EDA 04

Solución Propuesta 05

Retos encontrados + Próximos pasos 06



Detección y Predicción del Fraude



Pérdidas de decena de miles de millones de dólares

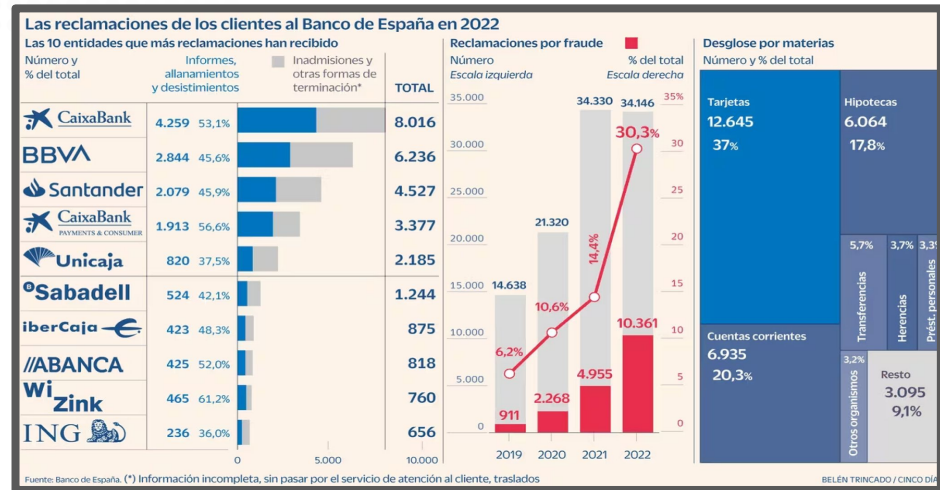


Necesidad de solución: Machine Learning (detección más precisa, proactiva y reducción de revisión manual)



Fraud Prevention Resources

Reducción de Pérdidas
Evitar daño de Marca
Clientes Experiencia Segura

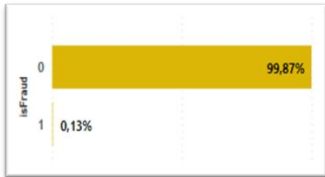


<https://cincodias.elpais.com/companias/2023-10-06/las-reclamaciones-por-fraude-de-los-clientes-ante-el-banco-de-espana-se-duplicaron-en-2022.html>



Análisis exploratorio de los datos

Para analizar el caso se realizó una exploración del dataset que nos permitiera obtener pistas sobre los datos, interpretarlos y modelarlos más adelante



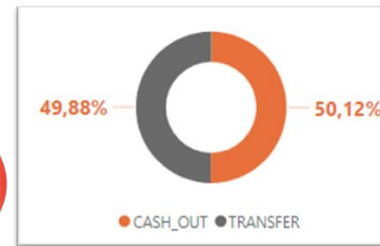
Bajo porcentaje de fraudes en los datos.

type	Promedio	Min.	Máx.	Desv std	Coef_var
TRANSFER	\$910.647	\$3	\$92.445.517	\$1.879.572	206 %
CASH_OUT	\$176.274	\$0	\$10.000.000	\$175.330	99 %
CASH_IN	\$168.920	\$0	\$1.915.268	\$126.508	75 %
DEBIT	\$5.484	\$1	\$569.078	\$13.318	243 %
PAYMENT	\$13.058	\$0	\$238.638	\$12.556	96 %

Alta dispersión de los datos - outliers

type	nameDest	nameOrig	amount	isfraud
CASH_OUT	C958479953	C1861878353	\$63,80	1
TRANSFER	C1368130863	C1293504491	\$63,80	1
CASH_OUT	C2102058838	C1584512618	\$119,00	1
TRANSFER	C1480876722	C1085557472	\$119,00	1
CASH_OUT	C1518370196	C773613907	\$119,65	1
TRANSFER	C543477940	C1497766467	\$119,65	1
CASH_OUT	C517676411	C790340353	\$151,00	1
TRANSFER	C315826176	C1172437299	\$151,00	1

Montos iguales marcados como fraudes



Fraude solo se da en **Cashout** y **Transfer**

Interrogantes

- ¿Por qué los fraudes solo se dan en **cashout** y **transfer**?
- ¿Que función tiene la variable **IsFlaggedFraud**?
- ¿Porque hay transacciones en cero marcadas como fraude?
- ¿Qué patrón existe en las transacciones con igual monto y diferente **type** marcadas como fraude?

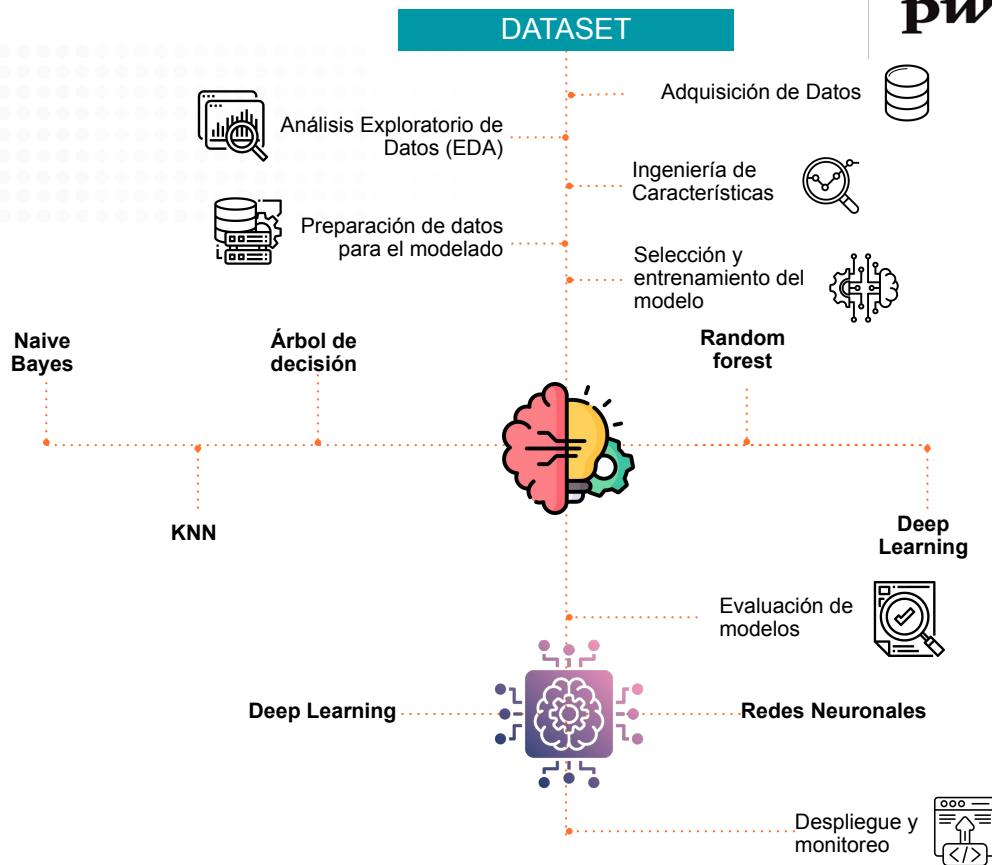
Solución Propuesta

- Undersampling para equilibrar el dataset.

NV	KNN	AD	RF	DL
93.68%	90.88%	77.98%	98.66%	93.20%

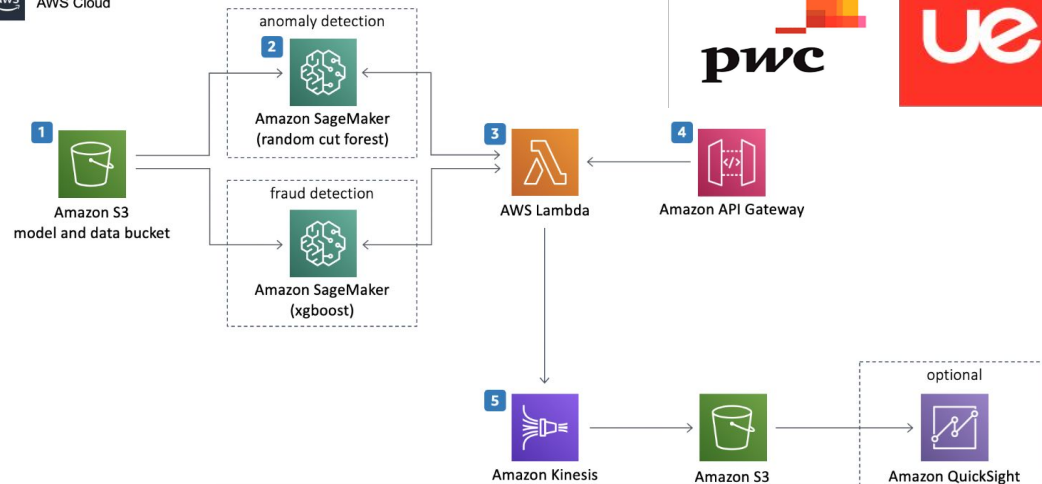
- One-Hot Encoding para optimizar con Deep Learning y Redes Neuronales

DL	NN
90.48%	95.26%



Retos + Próximos Pasos

AWS Cloud

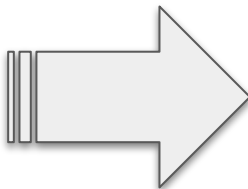


Falta de información en los datos proporcionados

Limpieza y Balanceo de Datos

Falta de Capacidad Computacional

Aprendizaje limitado para modelo ML



Alimentación y Optimización de la base de Datos

Cloud Computing y Optimización del Modelo ML

Puesta en Producción, Despliegue y Monitoreo