



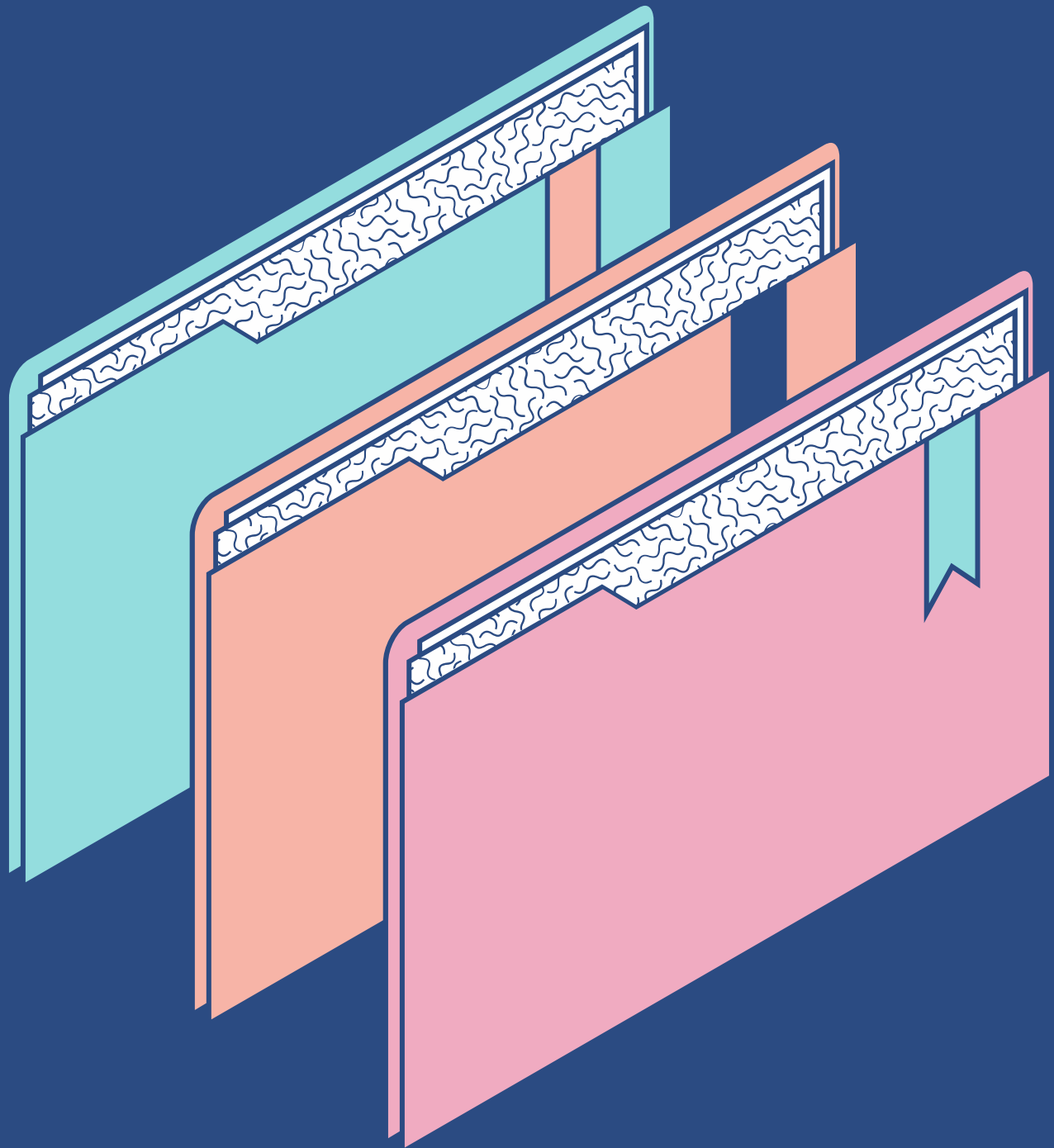
DATA SCIENCE FEM

Métodos de ensamble - XGBoost

Laura Andrea Orozco

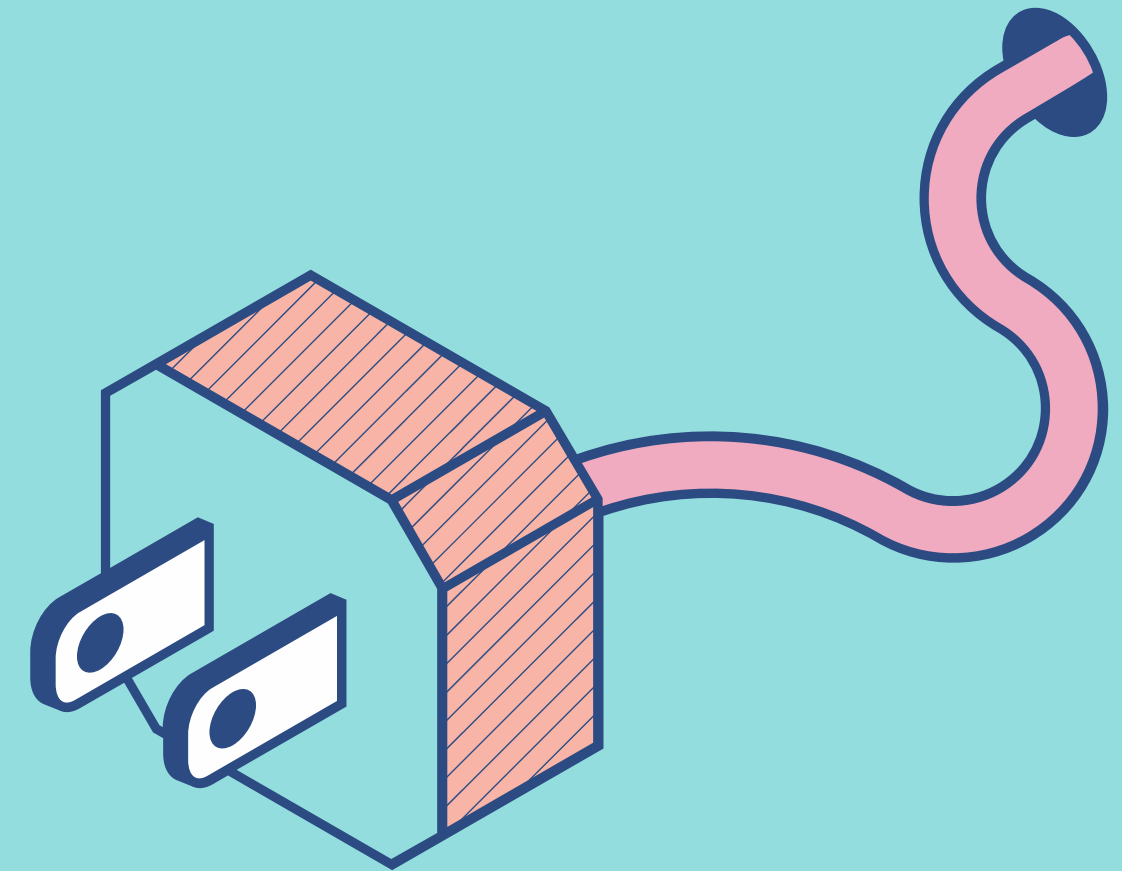
Metodos de ensamble

También conocidos como métodos combinados tienen como objetivo principal mejorar el rendimiento de los modelos de machine learning, esto se logra a través de la construcción de múltiples modelos que buscan resolver una problemática en particular, de manera que se obtienen mejores resultados que ajustando modelos individuales.

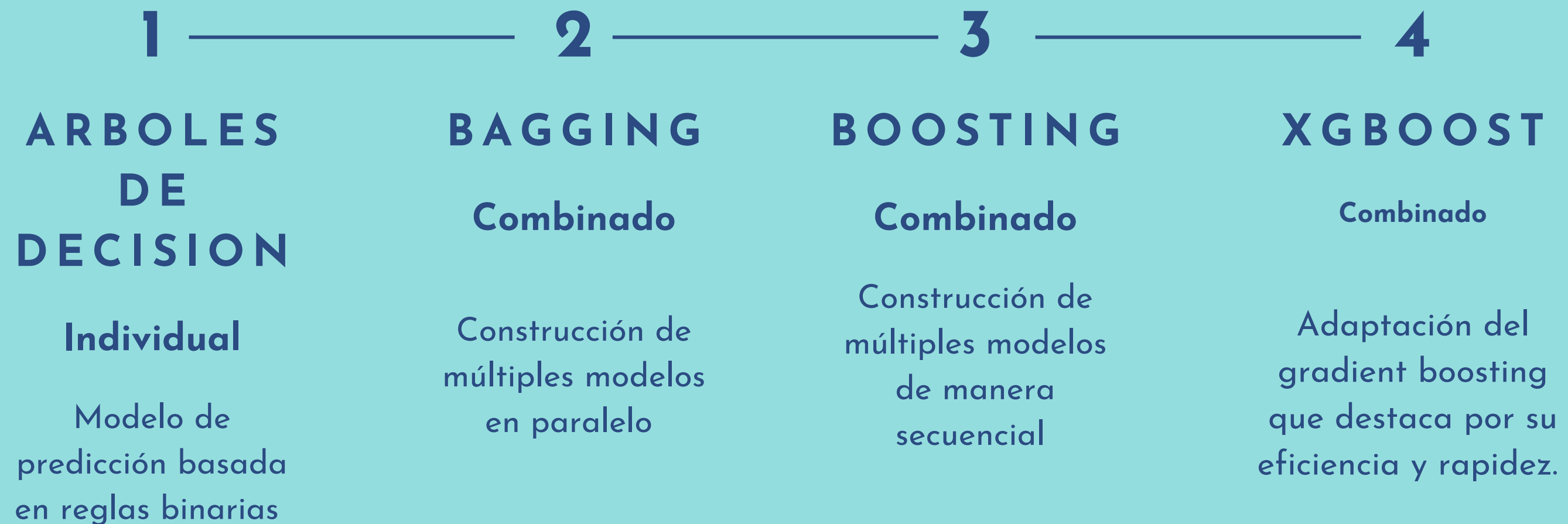


Boosting

Este concepto propone la construcción de una serie de modelos donde cada uno va aprendiendo de los errores de los modelos anteriores, el gradient descent propone la construcción de múltiples árboles de decisión de manera secuencial donde cada uno de ellos se ajusta a los errores de los árboles previamente ajustados. Y el valor final de la predicción es igual a la combinación de las predicciones de cada uno de los modelos multiplicado por una tasa de error resultante de cada modelo.



Técnicas de clasificación





Ventajas

- Puede manejar grandes bases de datos con múltiples variables
- Sus resultados son muy precisos
- Excelente velocidad de ejecución
- Pueden aplicarse en problemas de regresión y clasificación.

Desventajas

- Consume mucho recurso computacional si no se ajustan correctamente los hiperparametros.
- Solo trabaja con vectores numéricos, por lo que se deben hacer conversiones.

Consideraciones

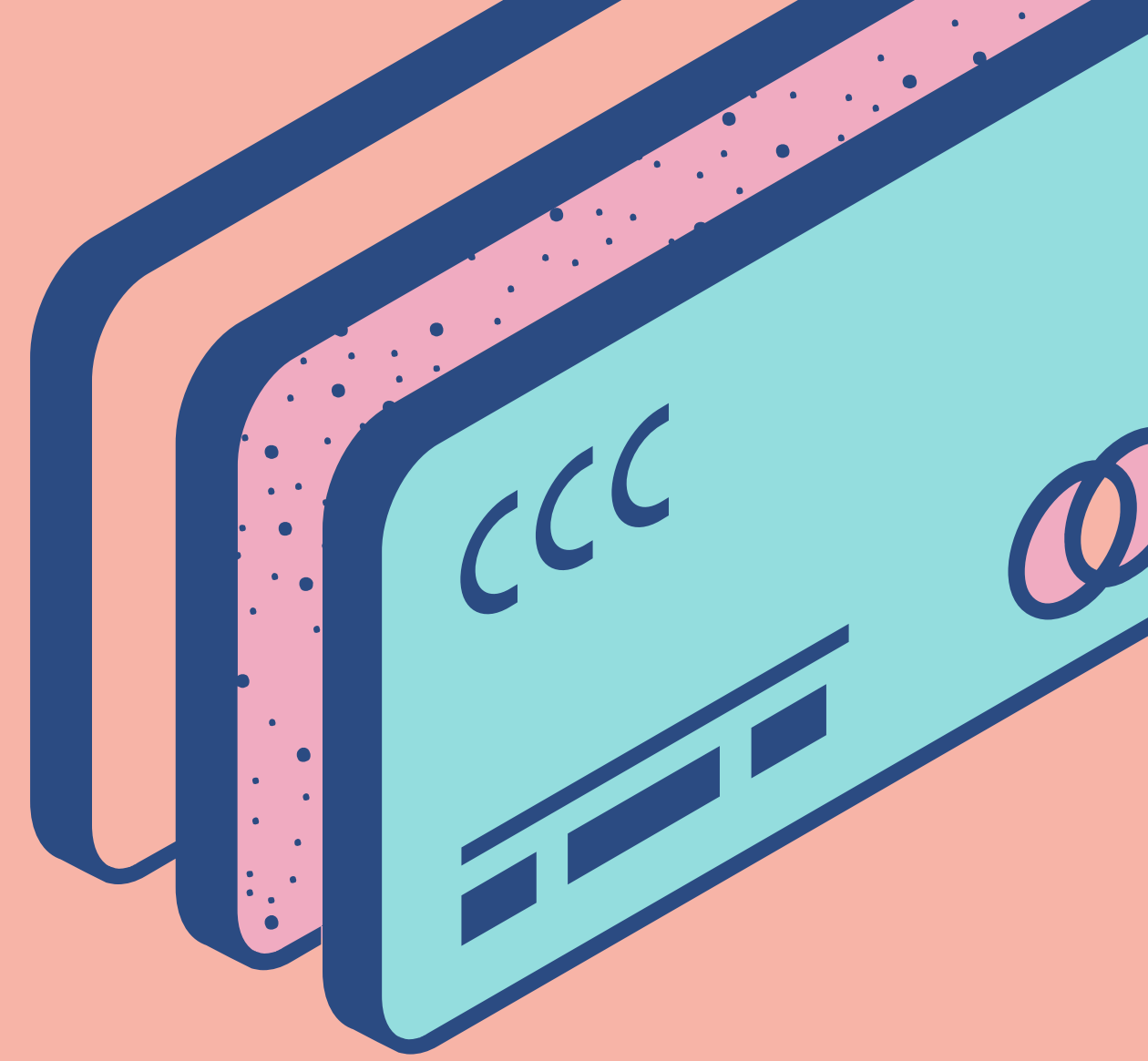
POR QUÉ Y CUÁNDO USAR XGBOOST

¿Por qué?

- Velocidad
- Rendimiento
- Supera constantemente a otros algoritmos

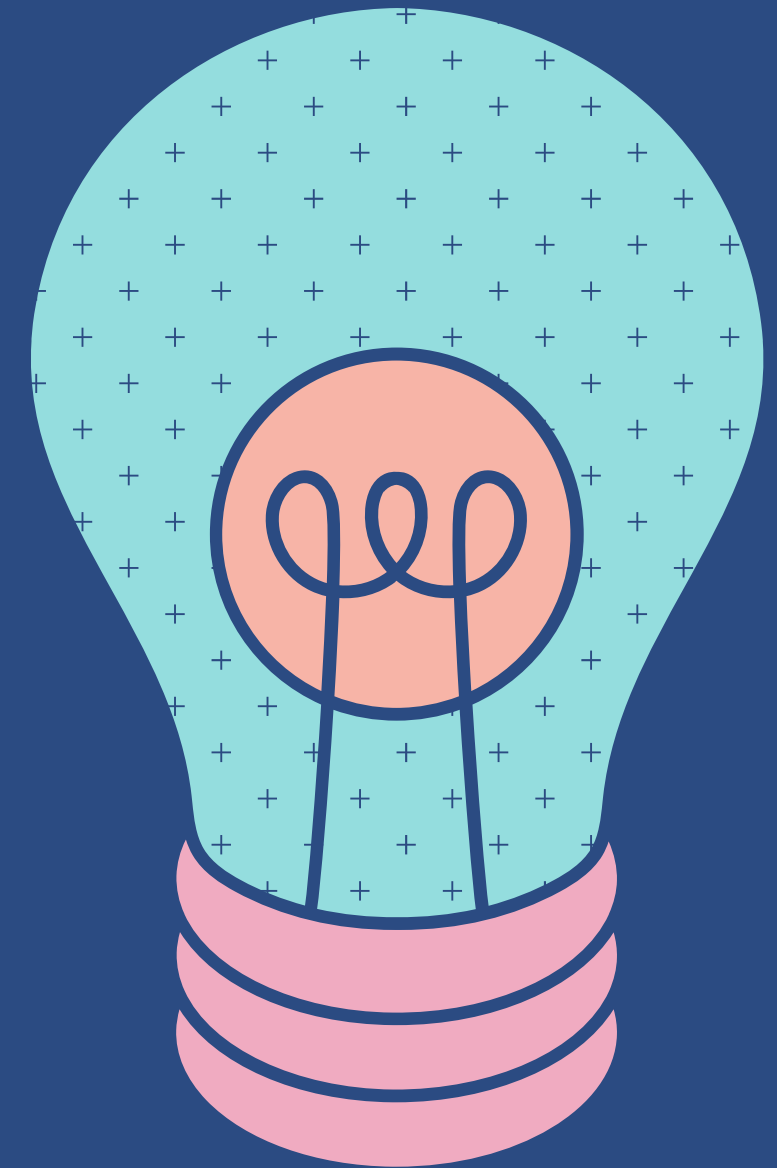
¿Cuándo?

- Se tiene una gran cantidad de ejemplos de entrenamiento
- Cuando se tiene mezcla entre variables categóricas o numéricas



Para tener presente...

- BAGGING O BOOSTING GENERAN BUENOS RESULTADOS SEGÚN EL CASO APLICADO
- LOS HIPERPARÁMETROS QUE SE USARON EN EL EJEMPLO NO SON LOS ÚNICOS QUE TIENE XGBOOST.
- LOS RESULTADOS OBTENIDOS TAMBIÉN DEPENDEN DE LA PREPARACIÓN DE LOS DATOS



**Vamos al caso
práctico**

