

# Ejercicio Práctico: Gradient Boosting con datos de Kaggle

En este ejercicio pondrás en práctica el uso de modelos de Gradient Boosting en Python utilizando un dataset real de Kaggle. El dataset recomendado es **House Prices - Advanced Regression Techniques**, disponible en Kaggle. Archivo principal: *train.csv*. Variable objetivo: *SalePrice* (precio de venta de las casas).

## Requisitos del ejercicio:

- 1 Importa las librerías necesarias para trabajar con datos, visualización y modelos de machine learning.
- 2 Carga el dataset de Kaggle (*train.csv*) y realiza una exploración inicial: número de registros, columnas, tipos de datos y valores nulos.
- 3 Preprocesa los datos: maneja valores nulos y selecciona un subconjunto de variables numéricas (ej. *OverallQual*, *GrLivArea*, *GarageCars*, *YearBuilt*). Define la variable objetivo *SalePrice*.
- 4 Divide los datos en conjuntos de entrenamiento y prueba utilizando *train\_test\_split*.
- 5 Crea y entrena un modelo de Gradient Boosting Regressor sobre los datos de entrenamiento.
- 6 Evalúa el modelo en el conjunto de prueba utilizando métricas de regresión: MAE, RMSE,  $R^2$ .
- 7 Visualiza las predicciones con un gráfico de dispersión (valores reales vs predichos) e interpreta los resultados.
- 8 Visualiza la importancia de las variables utilizadas en el modelo.
- 9 Mejora el modelo aplicando ajuste de hiperparámetros (ej. con *GridSearchCV*) sobre *n\_estimators*, *learning\_rate* y *max\_depth*.
- 10 (Opcional avanzado) Incluye variables categóricas mediante One-Hot Encoding y compara el desempeño con otros modelos como *RandomForestRegressor* o *LinearRegression*.

*Entrega tu implementación en un archivo Jupyter Notebook (.ipynb) o Python (.py). Incluye comentarios en el código que expliquen cada paso realizado.*