

## **Laboratorio #2 – AutoML con PyCaret**

### Descripción:

Para el dataset de House\_Prices de la competencia de kaggle, deberá implementar un Pipeline de machine learning utilizado PyCaret.

Deberá construir un notebook que contenga los siguientes elementos:

1. Carga de Datos.
2. Realice un Split de los datos en 80% training y 20% testing. Recuerde que debe usar sample de pandas dado que el train\_test\_split de sklearn no devuelve una tabla sino un tensor y PyCaret requiere dataframes como entrada.
3. EDA y Profiling de los datos.
4. Descripción de los warnings generados en el EDA.
5. Setup: en base a su experiencia deberá definir y justificar por que ha decidido realizar las operaciones de ingeniería de características que seleccione. Recuerde que debe considerar los siguientes aspectos en la ingeniería de características:
  - a. Imputación de variables numéricas.
  - b. Imputación de variables categóricas.
  - c. Codificación de variables categóricas.
  - d. Transformación de variables.
  - e. Tratamiento de Outliers.
  - f. Normalización de características.
  - g. Eliminación de características no utilizadas en el modelo.
6. Entrenamiento y selección de modelos automática. Deberá comentar cuales son los primeros tres modelos ganadores y sus métricas resultantes.
  - a. Para cada uno de estos modelos ganadores, deberá mostrar la gráfica de importancia de características y comentar los resultados.
7. Optimización de hiper-parámetros: De los primeros dos modelos deberá realizar optimización de hiper-parámetros.
8. Selección del modelo final: en función de los resultados obtenidos por la optimización de hiper-parámetros deberá seleccionar el modelo ganador.
9. Almacenamiento del pipeline en el disco duro.
10. Predicciones: Una vez almacenado el pipeline entrenado, deberá realizar las predicciones sobre el 20% de los datos almacenados para este propósito, cargando el modelo desde el path donde lo almacenó.
11. Cálculo de métricas de rendimiento: Con los datos predichos deberá calcular el R2, RMSE, MSE y MAPE. Además de proporcionar sus comentarios finales sobre los resultados.

Deberá entregar su laboratorio por medio de github.