

## PROYECTO DE CURSO – FASE 1

### Objetivo

Aplicar todos los conceptos y métodos aprendidos durante el curso para resolver un problema de predicción.

### Requisitos

1. Deberá seleccionar un dataset que considere adecuado para construir un modelo de **clasificación** que contenga al menos 15 variables mas la variable a predecir, es decir 16 variables. Adicionalmente el dataset debe contener al menos 500 filas para que todo el proceso que se desarrollará funciona adecuadamente. El dataset debe puede ser algo de su trabajo o un dataset público (Kaggle por ejemplo).
2. Análisis exploratorio: debe realizarlo para todo el dataset elegido, mostrando información relevante. Mostrar resultados y gráficas.
3. Selección de Variables:
  - a. Variable a predecir: **tipo categórico**
  - b. Variables predictoras: variables que considere esenciales.
4. Ingeniería de características: deberá desarrollar todo el procedimiento de ingeniería de características que considere necesario para resolver el problema de predicción:
  - a. Imputación de variables con data faltante:
    - i. Numérica
    - ii. Categórica
  - b. Codificación de variables categóricas
  - c. Transformación de variables numéricas
  - d. Tratamiento de outliers
  - e. Estandarización de variables

### Entregables – Fase 1

- Link de github incluyendo:
  - Dataset seleccionado.
  - **Notebook #1:** con Análisis Exploratorio con tablas y gráficas de sus datos.
  - **Notebook #2:** con el análisis, procedimiento y construcción del Pipeline de ingeniería de características.
  - Readme del proyecto y la información incluida en el repositorio.