

▼ Laboratorio 2: Algoritmos de aprendizaje de máquinas

- Eric Mendoza 15002
 - Marlon Fuentes 15240
-

▼ Preparación de entorno de ejecución

▼ Carga de librerías

```
import os

import pandas as pd
import numpy as np

import seaborn as sns
color = sns.color_palette()
sns.set_style('darkgrid')

import matplotlib.pyplot as plt
%matplotlib inline

from scipy import stats
import copy

from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split, cross_val_score

from sklearn import neighbors
from sklearn import preprocessing
from sklearn.model_selection import StratifiedShuffleSplit
from sklearn.preprocessing import StandardScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.cluster import KMeans
```

▼ Carga de data

Cargar train.csv desde github.

```
url = 'https://raw.githubusercontent.com/fuentesmarlon/CLUSTERING-PCA-APRIORI/master/train.csv'
```

▼ 1. Preprocesamiento de Data

Kaggle provee dos conjuntos de datos para realizar las pruebas, por lo cual no es necesario realizar la división e muestreo aleatorio. Sin embargo, por razones educativas, se procede a realizar la división del dataset *train.csv*.

