

## Curso: Análisis de datos

### Práctico 5 Aprendizaje automático

#### Parte 1: Aprendizaje supervisado

En este apartado se trabajará con un conjunto de datos sobre estrellas.

El objetivo del mismo es crear 3 modelos que clasifiquen a las estrellas según su valor de 'Spectral class' a partir del resto de variables.

Pueden elegir los 3 modelos que prefieran.

1. Preprocesar preprocesar los datos, dividirlos en conjuntos de entrenamiento, validación y testeo y normalizar los mismos.
2. Crear 3 modelos de entrenamiento, probarlos y comparar métricas obtenidas para cada uno.
3. Investigar acerca de hiperparámetros de cada modelo, cambiar dos en cada uno y probar nuevamente los modelos comparando sus métricas de evaluación.
4. Seleccionar el mejor modelo obtenido, probarlo en el conjunto de testeo y comparar las métricas obtenidas respecto al mismo modelo y sus métricas obtenidas con el conjunto de validación.

#### Parte 2: Aprendizaje no supervisado

En este apartado se trabajará con un conjunto de datos de clientes de un shopping.

El objetivo del mismo es crear un modelo que agrupe a los clientes según los patrones encontrados por el algoritmo (ya que en este caso no se cuenta con datos etiquetados).

1. Preprocesar los datos incluyendo la normalización de los mismos.
2. Realizar los siguiente gráficos de dispersión:
  - a. Anual Income (k\$) vs Spending Score (1-100)
  - b. Age vs Spending Score (1-100)
  - c. Age vs Anual Income (k\$)
3. Aplicar el modelo elegido, evaluar cuál es la cantidad de clusters adecuada a utilizar y una medida de distancia adecuada.

4. Crear nuevamente los gráficos solicitados, esta vez coloreando los clusters predichos.
5. Escriba las conclusiones que identifique a partir de la clasificación realizada.

Se debe entregar un informe **PDF** con el análisis solicitado en cada punto y el archivo donde hayan ejecutado el código para procesar los datos y realizar el análisis.