



# **PRACTICO 5- ANÁLISIS DE DATOS**

**ESTUDIANTES: GERMÁN GOICOECHEA, MARTÍN LASALA, AGUSTÍN  
PANIZZA, SANTIAGO ROURE Y MAURO VALDEZ**

**DOCENTE: JUAN CARLOS PELLEGRINI**

## Parte 1: Aprendizaje Supervisado (Estrellas)

### Preprocesamiento de datos:

Dedicamos uno de los archivos para esta tarea la cual tuvo los siguientes objetivos.

Los datos de estrellas se cargaron y verificaron para asegurar que no hubiera valores faltantes.

Variables categóricas como tipo de estrella, categoría y color se convirtieron en dummies para el modelo.

Se dividieron en conjuntos de entrenamiento, validación y prueba, y se normalizaron con StandardScaler.

### Modelos:

Para este practico llegamos a crear tres modelos distintos, los cuales fueron:

**Árbol de Decisión:** Entrenado y visualizado mediante gráficos de árboles de decisión.

**Random Forest** y **KNN:** Los dos fueron probados con el conjunto de validación.

Cada modelo se evaluó con métricas de precisión, recall, F1 y exactitud.

### Optimización de Hiperparámetros:

Se modificaron parámetros en los modelos y se reentrenaron para mejorar las métricas.

### Selección del Modelo:

El mejor modelo se probó en el conjunto de testeo y se compararon sus métricas con las del conjunto de validación para confirmar su consistencia.

## Parte 2: Aprendizaje No Supervisado

### Preprocesamiento de datos:

Los datos de clientes fueron normalizados y se realizaron gráficos de dispersión, comparando la edad, el ingreso anual y el puntaje de gasto.

### Clustering:

Se aplicó un algoritmo de clustering, donde se evaluó la cantidad y la medida de distancia óptimas para formar los grupos.

Finalmente, se generaron gráficos coloreados por clusters para observar patrones en los clientes.

### **Conclusiones de la parte 5:**

#### **1. Annual Income vs Spending Score:**

Se observan diferentes grupos de clientes que presentan distintos patrones de ingreso y gasto. Algunos clusters con ingresos altos tienen puntajes de gasto elevados, mientras que otros con ingresos similares muestran puntajes de gasto bajos. Esto indica que el nivel de ingreso no es el único factor que determina el comportamiento de gasto de los clientes, sino que otros aspectos, posiblemente relacionados con sus preferencias o estilo de vida, influyen en su disposición a gastar.

#### **2. Age vs Spending Score:**

En este gráfico, se evidencia que los clientes jóvenes tienden a estar en un cluster con puntajes de gasto más altos. Este patrón sugiere que los clientes más jóvenes pueden ser un segmento clave para el shopping en términos de gasto. Los clusters también revelan una segmentación en función de la edad y el gasto, lo que permite identificar diferentes perfiles de clientes que podrían beneficiarse de campañas de marketing personalizadas.

#### **3. Age vs Annual Income:**

En el análisis de edad contra ingreso anual, se observan varios clusters de clientes que, aunque tienen ingresos similares, pertenecen a diferentes grupos de edad. Esto muestra que los clientes con ingresos similares pueden tener comportamientos y necesidades distintos según su edad, lo que sugiere oportunidades para una segmentación adicional basada en este criterio.