

**PRACTICA 3**  
**APRENDIZAJE DE MÁQUINAS**  
**MARIA C TORRES**

**FECHA DE ENTREGA: JULIO 24 DE 2025**

**INSTRUCCIONES DE ENTREGA:** Presente un Jupyter Notebook para cada uno de los problemas. Cargue a la plataforma Moodle, los Notebook en una carpeta comprimida \*.ZIP.

**PARTE 1: SELECCIÓN DE DATOS Y ACONDICIONAMIENTO**

Seleccione UNO de los dataset adjuntos a la guía para su procesamiento. Podrá encontrar mayor información de cada conjunto de datos en Kaggle.

Dataset	Enlace mayor información
Installed capacity by state	<a href="https://www.kaggle.com/datasets/tahzeer/indias-power-generation-capacity-by-state">https://www.kaggle.com/datasets/tahzeer/indias-power-generation-capacity-by-state</a>
second_hand_car_sales	<a href="https://www.kaggle.com/code/hosseinzavari/analysing-second-hand-car-sales-data">https://www.kaggle.com/code/hosseinzavari/analysing-second-hand-car-sales-data</a>
Country-data	<a href="https://www.kaggle.com/datasets/rohan0301/unsupervised-learning-on-country-data">https://www.kaggle.com/datasets/rohan0301/unsupervised-learning-on-country-data</a>

Realice el acondicionamiento de los datos (eliminación de datos nulos, cambio de variables categóricas a numéricas) para usar los métodos de agrupamiento.

**PARTE 2: COMPARACIÓN DE METODOS DE CLUSTERING**

Para el conjunto de datos seleccionado en la parte 1, aplique los siguientes métodos de clustering:

- KMEANS
- Clustering aglomerativo
- DBSCAN
- Spectral Clustering

Para cada método seleccione los parámetros empleando el índice de la silueta. Los parámetros mínimos que debe establecer mediante este índice son:

- KMEANS: número de clusters
- Clustering aglomerativo: tipo de enlace (simple, completo, promedio o Ward) y número de clusters
- DBSCAN: eps y min\_samples
- Spectral Clustering: número de clústeres

Incluya al final del notebook un análisis comparativo de los 4 métodos, la complejidad para establecer los parámetros y si bajo su criterio logra un buen resultado.