

Actividad 4.1 - Segmentación de clientes según datos bancarios

El objeto de esta actividad es aplicar técnicas de clustering para obtener una serie de clusters que pueden ser utilizados para futuras campañas de Marketing.

Ahora bien, a menudo en la vida real podremos encontrarnos con conjuntos de datos relacionados con la adquisición/compra de productos y servicios, para a continuación realizar un estudio de clustering con el objeto de ver/estudiar una posible relación entre perfiles/segmentos/clusters de clientes y la adquisición de los productos/servicios.

Para esta actividad se ofrecen dos datasets, **de los que has de elegir SÓLO uno de ellos según tu criterio personal.**

Inicialmente se utilizarían para problemas/casos de uso de aprendizaje supervisado (Clasificación).

Título: Bank Marketing

Url: <https://archive.ics.uci.edu/dataset/222/bank+marketing>

Descripción: Los datos están relacionados con campañas de marketing directo (llamadas telefónicas) de una institución bancaria portuguesa. El objetivo de la clasificación es predecir si el cliente suscribirá un depósito a plazo (variable y)

Título: Statlog (German Credit Data)

Url: <https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data>

Descripción: Este conjunto de datos clasifica a las personas descritas según un conjunto de atributos como riesgos crediticios buenos o malos. Viene en dos formatos (uno todo numérico).

Ejemplo de referencia a considerar, por si ves algún mecanismo/utilidad/herramienta que desees incorporar a tu solución.:

- **Título:** Clustering for Effective Marketing Strategy
Url: <https://www.kaggle.com/code/caesarmario/clustering-for-effective-marketing-strategy>

La rúbrica a utilizar para evaluar

Criterios para valorar Proyecto/Reto		
Actividad_4_1_Segmentación de clientes según datos bancarios		
Peso %	Tareas (Se evalúa entre 0 y 10)	
5	Argumenta la elección del dataset de estudio	
5	Utilizar el drive/github como origen de ficheros para la importación del dataset	
5	Importación del dataset y preparación de los datos: Normaliza, ajusta la calidad de los datos.	
10	Número de clusters: Utiliza la técnica de Silhouette Coefficient para determinar el número de clusters	
15	Entrenamiento: Utiliza alguno de los modelos vistos en clase (dendogramas , DBSCAN, DPC, K-MEANS) - Desarrolla las diversas pruebas propuestas para la selección y justifica el criterio utilizado para la selección del modelo. Nota: Recuerda y valora la necesidad de aplicar escalamiento de las características.	
10	Entrenamiento: Aplica técnicas de reducción de la dimensionalidad (PCA)	
10	Entrenamiento: Muestra gráficamente los resultados del agrupamiento en ua gráfica de dos dimensiones	
10	Exportar a Excel los resultados obtenidos, asociando a cada fila/muestra con los clusters obtenidos.	
10	Conclusiones de clusters: Obtiene conclusiones en relación a los clusters obtenidos y la compra o adquisición de un producto o servicio. Se puede apoyar en el uso de Power BI o cualquier otra herramienta que consideres adecuado.	
10	Propone soluciones creativas e innovadoras	
5	El pdf final tiene una portada., utiliza un índice, apartado de conclusiones y referencias (web). Se hace mención a referencias externas, no recogidas en el material suministrado.	
5	Comenta con claridad cada uno de los pasos realizados	
Total	100	

Formato de entrega

- Entregar un fichero en un Archivo PDF con capturas del código y resultados obtenidos, así como la url de GitHub y Google Colab donde has publicado el código.
- Nombrar el archivo siguiendo el siguiente patrón:

SNS_ACT4_1_NombreApellidos.pdf