

Guía de trabajo Módulo 3 – Aprendizaje No Supervisado

1. Reto

Cada equipo debe seleccionar uno de los datasets relacionados en la tabla 1 y plantear un problema o caso de estudio que pueda ser resuelto mediante la aplicación de un algoritmo de aprendizaje no supervisado. Estos son los algoritmos que veremos en el módulo 3:

- K-Means
- Hierarchical Clustering
- Gaussian Mixture Clustering
- DBScan clustering

Dataset	Enlace para descargar y más información
Dataset de su elección	Pueden consultar en kaggle o bases de datos públicas
Customer Personality Analysis	https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis
Ship Performance Clustering Dataset	https://www.kaggle.com/datasets/jeleeladekun/lefijabi/ship-performance-clustering-dataset/data

Adicionalmente, se debe aplicar el algoritmo de reducción de la dimensionalidad **PCA** como parte de las exploraciones realizadas para mejorar el rendimiento de los algoritmos de clustering seleccionados.

El trabajo que se entrega a final del módulo debe contener:

1. El notebook(s) generado para la solución del problema planteado, organizado y documentado con los análisis realizados y las principales conclusiones. Si en los notebooks no se detalla el proceso de análisis y desarrollo, la nota de la entrega se verá afectada.
2. Un informe con el diseño de la solución propuesta, los análisis realizados, las conclusiones y recomendaciones finales (El informe debe tener máximo 5 páginas).

Se propone que los siguientes elementos sean analizados en el trabajo:

- Planteamiento del problema
- Limpieza y transformación de los datos (eliminación de datos nulos, cambio de variables categóricas a numéricas, limpieza de campos, etc)
- Análisis exploratorio de los datos
- Comparación y selección de técnicas para la aplicación de 3 de los algoritmos de clustering vistos en clase
- Afinamiento de hiperparámetros
- Incluya al final del notebook un análisis comparativo de los 4 métodos, la complejidad para establecer los hiperparámetros y si bajo su criterio logra un buen resultado.
- Conclusiones y recomendaciones



Facultad de Ingeniería
Departamento de Ingeniería Industrial
Analítica para la toma de decisiones
Aprendizaje no supervisado

Nota: Dependiendo de la solución que el equipo proponga, que es libre, puede que se omitan pasos o se incluyan adicionales, en este caso, se debe analizar por qué no se incluyó el paso o por qué se adicionaron.

Fecha de entrega: 28 de octubre

Criterios de evaluación del trabajo:

1. Pertinencia y creatividad de la solución entregada 15%
2. Estructura, documentación y justificación del código 30%
3. Análisis de resultados 35%.
4. Redacción adecuada del informe (incluye ortografía) 20%

Sustentación:

- Cada grupo debe sustentar sus entregas. Si el equipo no puede sustentarlo la nota será cero independientemente de la solución entregada.
- La presentación es de máximo 15 minutos.

Criterios de evaluación de la sustentación:

- Claridad y orden en la presentación (presentar lento y escoger bien los contenidos a mostrar, dado que el tiempo es limitado).
- Compresión adecuada del problema y del algoritmo aplicado.
- No tener errores conceptuales (mal uso o mala interpretación de conceptos o herramientas).