



1ACC0241- Data Mining & Data Analysis

Trabajo Final

2025-02

1. Objetivo

El presente documento define el trabajo final y la rúbrica que permite evaluar el logro del curso 1ACC0241 – Data Mining & Data Analysis.

2. Logro del curso

Competencia General: Pensamiento

innovador Nivel de logro: 2

Capacidad para generar propuestas sostenibles y creativas de solución a un problema, que implica la mejora o creación de un producto, servicio o proceso, impactando positivamente en un determinado contexto.

Competencia Específica: ABET 6 – Técnicas en computación

Nivel de logro: 1

La capacidad para comprender y brindar soporte para el uso, entrega y gestión de sistemas de información dentro de un entorno de sistemas de información

3. Enunciado

El trabajo final consiste en aplicar algoritmos de aprendizaje no supervisado haciendo uso de los modelos de Kmeans, DBSCAN, HDBSCAN, OPTICS y Gaussian mixtures, proponiendo un aporte de conocimiento, acorde con las competencias generales y específicas del curso.

La principal referencia bibliográfica se encuentra en la página web de scikit-learn: <https://scikit-learn.org/stable/modules/clustering.html>

4. Estructura del Informe

El informe debe contener la siguiente estructura:

- **Descripción del caso de uso.** Redactar la descripción y fundamentación del problema y/o caso de uso del cual se propone encontrar algún conocimiento (citar fuentes).
- **Descripción del conjunto de datos obtenidos de kaggle u otra fuente de datos.** Redactar las características y origen de los datos recolectados motivo de análisis y para su posterior analítica.
- **Análisis exploratorio de los datos (EDA).** Se debe incluir la descripción de las tareas de inspección, preprocesamiento, análisis univariado, bivariado, visualización de los datos o cualquier análisis que crea genere valor.
- **Modelización.** Comprende la aplicación de los algoritmos de aprendizaje no supervisado **Kmeans, DBSCAN, HDBSCAN, OPTICS** y Gaussian mixtures.
- **Resultados.** Comunicar los principales resultados obtenidos y su comparación mediante el uso del ancho de silueta, uso de métricas y tablas comparativas que crea conveniente.
- **Conclusiones.** En un párrafo redactar las conclusiones del trabajo, especificando las técnicas utilizadas y los resultados obtenidos (positivos o no).
- **Recomendaciones.** Redactar las sugerencias para el negocio con la información analizada y los segmentos encontrados.
- **Referencias bibliográficas**

5. Acerca del grupo de trabajo

El trabajo se deberá desarrollar en grupo de máximo 3 estudiantes.

6. Lenguaje de programación

Las tareas de análisis y analítica deben estar desarrolladas en el lenguaje Python.

7. Exposición

- La exposición es individual (todos los integrantes del grupo exponen) y es parte de la evaluación del trabajo final que se realizará durante la sesión 15, la cual durará máximo 20 minutos.

8. Entrega en el aula virtual

Crear un archivo con el nombre:

TF_GRUPO_X.docx

TF_GRUPO_X.ipynb

- Subir los archivos al Aula Virtual en Fecha: **05/12/2025 antes de la 09:50h**
- Puntaje asignado: 0 a 20 puntos