

UNIVERSIDAD NACIONAL DE COLOMBIA - SEDE MEDELLÍN
FACULTAD DE MINAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN Y DE LA
DECISIÓN

- **Inteligencia Artificial, Código: 3007855**
- **Introducción a la Inteligencia Artificial, Código: 3010476**

Semestre: **02/2021**

Prof. Demetrio Arturo Ovalle Carranza, Ph.D. (e-mail: dovalle@unal.edu.co)

Monitores:

Ana María Osorio Mondragón (e-mail: anmosorियो@unal.edu.co)
Juan Pablo Carvajal García (e-mail: jpcarvajalga@unal.edu.co)

Eq. 9

ENUNCIADO DEL MICRO-PROYECTO2 (15%) - CLUSTERING

Fecha: Jueves 18 de noviembre

Fecha de entrega: Lunes 29 de noviembre, entrega digital; sustentación martes 30 de noviembre

Estudiantes:

Angel Maya, Emmanuel
Orozco López, Luisa Fernanda
Zapata Querubín, Emanuel

OBJETIVO – Analizar el Índice de Suicidios Hombres - Mujeres

DATASET: 2 archivos cvs – a) Suicide rate, female (per 100,000 people); b) Suicide rate, male (per 100,000 people)

Usando un algoritmo similar al presentado en clase y en la práctica, realizar un programa en Python que determine los clústeres en los datos del archivo .csv a partir de las tres variables que según su análisis mejor segmentan los datos.

Se sugiere mezclar los datos de forma aleatoria antes de realizar cualquier otro tipo de preprocesamiento.

Indicaciones de entrega: sólo debe entregarse el archivo descargado de Google Colaboratory con el siguiente formato de nombre: "equipo#_tematicaDataset.ipynb". Los trabajos serán probados con el conjunto de datos original, por lo tanto, todas las modificaciones hechas al dataset deben realizarse desde el código.

Los integrantes deben aparecer en el .ipynb para que se les pueda asignar la calificación obtenida.

1) Explique si tuvo que realizar algún tipo de pre-procesamiento en el archivo .csv.

2) Explique la(s) técnica(s) utilizada(s) para establecer el número de clústeres del conjunto de datos. (ejm. Curva de codo, Estadístico de gap, etc.). Justifique su respuesta.

3) Basado en los clústeres establecidos, determinar en cuántos conjuntos se agrupan los datos similares y qué características son más representativas para cada grupo, realizando la mayor caracterización de los datos posible.

Para este ejercicio se usarán dos datasets: a) Índice de Suicidios Hombres vs b) Índice de Suicidios Mujeres por países, se sugiere agrupar los países en 3 años diferentes, espaciados por un tiempo similar (2000, 2010, 2016).

NOTA: Recuerde explicar qué está haciendo en cada parte del código. Para esto debe entregar la libreta (.ipynb) correspondiente, debidamente documentada. La falta de esto afectará la calificación final.

4) Realizar un agrupamiento jerárquico de los datos a través de un dendograma, según se explicó en el taller, donde se visualicen los clústeres del dataset entregado. Este código debe hacer parte de la libreta (.ipynb) y ser debidamente documentado.

5) ¿Qué puede concluir del porcentaje de suicidios en mujeres en latinoamérica? Así mismo ¿Qué puede concluir del porcentaje de suicidios en hombres en latinoamérica? ¿Qué países son más propensos al suicidio en el mundo? (ejm cite un país en cada continente)