



Bogotá D.C., 27 de Octubre de 2022

**MAESTRIA EN INTELIGENCIA ARTIFICIAL**

**APRENDIZAJE DE MAQUINA**

Prof. Cesar Diaz. Ph.D.

**Parte 1**

Un sistema de recomendación tiene varias maneras de implementarse, uno como el que se realizo de tarea, utilizando filtros colaborativos y otra forma es utilizando filtros basado en contenidos. Para este reto, van a trabajar una base de datos no numérica, lo que tendrán que hacer son pasos, que hemos hecho en nuestros laboratorios, pero con variables no numéricas.

URL de los datos: <https://query.data.world/s/uikepcpffyo2nhig52xxeevdialfl7>

**Hint:** Ojo!! Es solo una pista, pueden tomar de este set solo las variables que a su parecer puedan utilizar y sirvan para su motor de recomendación. i.e. Title, Genre, Director, Actors, Plot.

1. Análisis descriptivo de los datos. Realice un análisis descriptivo completo del dataset utilizado en el parcial (10%)
2. Realice normalización, limpieza, estandarización o lo que necesite hacerle a los datos (10%).
3. Realice una función donde tome el titulo de una película como entrada y retorne el top 10 de películas recomendadas. (40%)

**Parte 2**

De acuerdo al set de datos recogido por SOCRATA, el del SECOP II, realice en el lenguaje de su escogencia, lo siguiente:

1. Aplique un modelo de clusterización a estos datos (10%).
2. Cree un modelo de predicción utilizando arboles de decisión, la variable target es de libre escogencia, pero explicar muy bien porque se escogió, de igual forma aplique al mismo problema Random Forest y realice la comparación de los modelos. (15%)
3. Hacer lo mismo que el punto anterior pero con otra variable target diferente. (15%)

Se debe entregar en archivo comprimido .zip (ojo!! No se puede entregar .rar) la secuencia de archivos y directorios para cada parte del parcial debe ser así: un directorio para la parte 1, donde se encontraran los respectivas fuentes y programas de la solución y otro directorio para la parte 2 correspondiente.

**Fecha de entrega:** Noviembre 24 del 2022 6:00pm