



CURSO: CMP 0575 - TÓPICOS 2 (DATA MINING)
COLEGIO: POLITÉCNICO
Semestre: Primer Semestre 2018/2019 NRC: 1068

Tarea 7: Ejercicio usando agrupaciones (clustering)

Problema:

1. Dado el conjunto de datos “**Dataset(Clustering).csv**” proporcionado en el D2L, se desea aplicar un algoritmo de agrupamiento que permita realizar asignaciones de los datos a los posibles **clusters** formados. Para la realización de la tarea se exige:
 - Usar **uno** de los tres algoritmos vistos en clase de acuerdo a la distribución por estudiante presentada en el punto 2 de este documento: **k-means**, **BRF** y **CURE**.
 - Es obligatorio mostrar la trazabilidad del método durante la ejecución del programa:
 - i. El valor del **centroide** de los **clusters** óptimos de solución.
 - ii. El proceso de selección de **k** (cantidad de **clusters**), o sea como converge la distancia promedio de los centroides a medida que aumenta el número de **clusters** (optimizar **k** en un rango de 1..10). Un plot parecido al que se presentó en la clase de **K vs D** (distancia).
 - iii. La cantidad de instancias asignadas a cada uno de los **cluster** creados.
 - iv. Es obligatorio realizar una **presentación científica (NO FILOSÓFICA)** al profesor con todos los elementos requeridos.
 - v. La programación del algoritmo utilizado debe ser en el lenguaje **JAVA** o **C++**
 - Cargar al D2L los códigos implementados (fichero compactado que incluye el ejecutable ej: el .JAR de java) y la presentación dentro del plazo de entrega.
2. Distribución de estudiantes por algoritmos:
 - K-means (Pedro S., Christian M., Paolo A., Santiago R.)
 - BRF (Christian S., Alejandro B., Raymond M., Diego M., Javier C.)
 - CURE (Jonathan O., Esteban F., Andres E., Samik R.)