



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

Tarea Extra: Ejercicio usando el procesamiento de datos y la selección de características

Problema:

1. Dado el dataset **MAS_126_SE_1072CasesOriginal** ubicado en la carpeta Dataset del D2L. Se desea:
 - Conformar un *dataset* de entrenamiento (*train*) que no contenga la primera (NAME) ni la última columna (CLASS), las cuales transmiten información de las variables y clases numéricas respectivamente.
 - Aplicar la tarea de **normalización** de los datos para el conjunto de entrenamiento (*train*).
 - Aplicar la tarea de **feature selection** usando el paradigma wrapper al nuevo conjunto de entrenamiento.
 - El modelo debe maximizar el valor de performance en el training lo más cerca posible de 0.95 (umbral heurístico de calidad).
 - Es obligatorio mostrar la trazabilidad del método durante la ejecución del programa:
 - i. Subconjunto optimizado y valor de performance obtenido
 - Es obligatorio realizar una **presentación científica (NO FILOSÓFICA)** al profesor con todos los elementos del método wrapper.
 - Cargar al D2L los códigos implementados (archivo compactado que incluye el ejecutable ej: el .JAR de java) dentro del plazo de entrega.

Nota Importante: Esta tarea constituye una segunda oportunidad para mejorar la nota del proyecto 4.