

Proyecto 2

Día de entrega: 31/3/2020

Introducción Ciencia de Datos

Lead University.

Profesor: Rodrigo Herrera.

Instrucciones del Proyecto

Se puede hacer en grupos de 4 como máximo

Proyecto: Aplicación de la materia vista hasta el momento (Limpieza de datos, Exploración de datos y utilización de algoritmos)

Descripción:

Cada grupo debe de trabajar con el objetivo de poder predecir si un chofer de la formula 1 va a poder terminar o no la carrera.

En el archivo comprimido llamado “Datos Proyecto” van a tener 2 carpetas llamadas “F1 Race Data” y “F1 External Data”.

“F1 Race Data” =>Va a tener los datos internos de las carreras, pits, corredores, grupos, etc, deben de revisar el archivo llamado “f1db_schema.txt” para poder revisar el diccionario de cada una de las tablas o archivos que estén dentro de la carpeta.

“F1 External Data” =>Son archivos de datos externos que se deben de usar para poder crear la tabla final antes del entrenamiento. Estos datos son relacionados al clima y las llantas que se utilizaron durante las carreras.

Por hacer:

- Revisar cada uno de los archivos o tablas, de todos crear una sola tabla final, que este limpia y la cual contenga solamente las columnas que consideren necesarias para hacer el entrenamiento de la clasificación.
- Con la tabla final lista, deben de realizar un proceso de entrenamiento para crear un clasificador binario en la aplicación Orange Data Mining tal como lo vimos en clase.
- Deben de probar 4 diferentes algoritmos y escoger el mejor algoritmo según su criterio y explicar porque los restantes no son eficientes.
- Deben crear una presentación en la cual expliquen el procesamiento de limpieza que realizaron para crear la tabla final y exponer los resultados obtenidos y su explicación

Entregar:

- Deben de subir al aula virtual, la tabla final limpia, archivo de Orange Data Mining y la presentación que realizaron.

TIPS:

Considerar fechas, recuerden el proceso One-Hot Encoded