

## Reto Filtro

En esta ocasión el reto filtro consiste en crear un algoritmo de clasificación. Los datos para el reto filtro están separados en tres archivos: `datosEntrenamiento.csv`, `datosPrueba.csv` y `datosValidacion.csv`.

Los archivos **`datosEntrenamiento.csv`** y **`datosPrueba.csv`** se utilizan para generar el algoritmo y contienen cada uno 37 columnas, de las cuales 36 representan atributos y la columna 37 contiene un número que representa la clase a la que pertenece el registro. Hay 6 posibles clases indicadas por los números 1, 2, 3, 4, 5 y 6. Los archivos **`datosEntrenamiento.csv`** y **`datosPrueba.csv`** contienen 4504 y 1287 registros respectivamente.

El archivo **`datosValidacion.csv`** contiene 644 registros con 36 columnas cada uno; la columna 37 correspondiente a la clase **no está presente** en este archivo. El algoritmo debe predecir la clase correspondiente para cada registro en este archivo.

Para completar su inscripción al DCI, deberán entregar un documento describiendo la solución desarrollada indicando la efectividad de su algoritmo, la precisión para describir los datos y un archivo de texto que contenga únicamente el número entero correspondiente a la clase que predice su modelo para cada uno de los registros en el archivo **`datosValidacion.csv`**. Este archivo de texto debe ser de 644 líneas un número en cada línea con el nombre **`retoFiltro.txt`**.

La evaluación del reto se hará a partir de la precisión con la que su algoritmo pueda determinar a que clase pertenece cada uno de los objetos de la colección que se encuentra en el archivo **`datosValidacion.csv`**, tomando en cuenta la metodología empleada para generar el modelo.

Así, el objetivo del reto es crear un algoritmo de clasificación basado en el conjunto de entrenamiento, que clasifique los datos en las clases correspondientes. Una vez obtenido el modelo de clasificación, aplicarlo al conjunto de validación proporcionado y enviar los resultados obtenidos (las clases predichas por el algoritmo para cada elemento del conjunto de validación).

Recuerda que tienen 5 días naturales para entregar el reto filtro a partir de tu registro al evento. El equipo deberá de entregar un reporte donde explique el proceso de la creación del algoritmo y su implementación, así como un archivo de texto por separado con el nombre **retoFiltro.txt** que debe contener en el primer renglón el nombre del equipo y en cada renglón siguiente la clasificación de los datos contenidos en **datosValidacion.csv** obtenida por el algoritmo: por ejemplo si el algoritmo predice la clase 1 para el renglón 1 de **datosValidacion.csv** y predice la clase 6 para el renglón 76, el archivo de texto debe tener la forma

Nombre Equipo <- renglón 0

1 <- renglón 1

... <- renglón 2

...

6 <- renglón 76

...<- renglón 77

...

La entrega de los archivos se realizará a través de la página del DCI para su evaluación y asignación de categoría. El Comité de Evaluación determinará la postulación final y el equipo será notificado al correo electrónico indicado en el registro cinco días hábiles después de recibir las respuestas.

Comité Organizador

Data Challenge Industrial DCI 3.0

Xalapa, Ver., 27 de agosto de 2018