

Primer desafío de ciencia de datos - ML UDD

Se cuenta con un archivo con datos de entrada a modo de ejemplo para clasificar si el usuario que visita un sitio web usa como sistema operativo Windows, Macintosh o Linux. La información de entrada son 4 características que se extraen de una web que utiliza Google Analytics. Las características son:

- Duración de la visita en Segundos
- Cantidad de Páginas Vistas durante la Sesión
- Cantidad de Acciones del usuario (click, scroll, uso de checkbox, sliders, etc)
- Suma del Valor de las acciones (cada acción lleva asociada una valoración de importancia)

Como la salida es discreta, asignaremos los siguientes valores a las etiquetas:

- 0 Windows
- 1 Macintosh
- 2 Linux

La muestra es pequeña: son 170 registros para poder comprender los métodos que hemos abordado hasta ahora. Para ello, se solicita:

1. Generar análisis de ciencia de datos en Python, usando Jupyter Notebooks y Scikit Learn como biblioteca de métodos.
2. Utilizar métodos de Machine Learning referidos en clases, y en el material disponible del curso, para predecir el sistema operativo que usa el usuario en base a sus métricas.

Una vez generado el modelo con el archivo **usuarios_win_mac_lin_train.csv**. Predecir para el archivo **data validation without class.csv**. En los próximos días les llegará una invitación al desafío de Binnario School, para que suban sus resultados y vean cómo está la efectividad de su modelo. Tendrán un máximo de 3 intentos para “evaluar” sus resultados. Por ello, es importante que efectúen en el tiempo inicial asignado un buen análisis fundamental (estadístico y/o visual) y experimentación de métodos en ambiente de train/test.

Respecto a la métrica a evaluar, el desafío medirá el grado de satisfacción respecto al F1-Score (Revisar lectura complementaria disponible en Canvas o directamente en el siguiente enlace: [Métricas de evaluación de clasificación](#)).

Por otro lado, se solicita que una vez finalizado el desafío, haga entrega de su código notebook (ojalá documentado en el paso a paso) vía Canvas, en el link autorizado para ello.

Nota: Puede ser desarrollado de manera individual o en equipos de hasta 4 personas.