

Tarea 1

Machine Learning: Exploratory Data Analysis (EDA)

Entrega: 2 de febrero 2023

Instrucciones: La tarea será desarrollada en equipos de a lo más 4 estudiantes, y constará de 2 entregables: una presentación y un notebook. El plagio está estrictamente prohibido y será penado con la anulación de la tarea.

Fecha de exposición: La presentación tendrá lugar el jueves 2 de febrero a las 7:00 pm. El orden de los equipos será aleatorio, y cada presentación tendrá una duración de 10 minutos estrictos, considerando tiempo para posibles preguntas.

Fecha de entrega: La presentación y el notebook tienen como fecha de entrega el jueves 2 de febrero a las 11:59 pm. Por cada hora de retraso se descontarán 10 puntos sobre la calificación de la tarea. La tarea será enviada al correo `orlando.uc@correo.uady.mx` por un representante del equipo, y el formato del nombre será `Tarea1_EDA_Nombre1_Apellido1_ ... _Nombre4_Apellido_4.xyz`.

Ejercicio 1

A partir del dataset `youtube_dataset.csv` que contiene datos acerca de videos y canales de Youtube, realiza un EDA:

1. Identifica cuáles de las variables que contiene el dataset son categóricas y cuáles son continuas. Describe qué variables te parecen interesantes y por qué.
2. Explora si en el dataset encuentras datos faltantes, caracteres especiales y realiza el preprocesamiento que consideres necesario. Estandariza tus datos.
3. Realiza un análisis descriptivo y comenta a qué conclusiones llegaste. ¿Encuentras variables altamente correlacionadas? Si es así, o existen variables que no aportan información a tu análisis, considera eliminar variables para evitar la *maldición de la dimensionalidad*.
4. Analiza la variable `videoPublished` y comenta si la fecha o la hora de la publicación tiene relevancia para la variable `likes/views`.

5. Crea una variable categórica a partir de la variable `likes/views`. Identifica cuáles de las variables están altamente correlacionadas con la nueva variable. Supón que eres un youtuber y te interesa aumentar el ratio `likes/views` para ganar más dinero, a partir de todo el análisis previo, diseña una estrategia. ¿Intentarías aumentar el número de suscriptores? ¿Te concentrarías en aumentar el ratio `likes/suscriber`? ¿Consideras relevante la fecha de publicación de tu próximo video? Justifica.
6. Genera una presentación a partir de todo tu análisis previo y de la estrategia que diseñaste, justificando de la manera más simple posible qué es lo que propones. Apóyate de gráficas y tablas según consideres necesario.