



MATRÍCULA: 1797222

INTRODUCCIÓN

La clasificación de datos es una tarea fundamental en el campo del aprendizaje automático y la inteligencia artificial. Consiste en asignar etiquetas o categorías a instancias de datos en función de sus características o atributos. Es un tipo de problema supervisado en el que se utiliza un conjunto de datos etiquetados previamente para entrenar un modelo que pueda realizar predicciones sobre nuevos datos no etiquetados.

El objetivo de la clasificación de datos es construir un modelo capaz de aprender patrones y relaciones en los datos de entrenamiento para poder generalizar y realizar predicciones precisas sobre nuevos datos. Esto implica identificar y extraer características relevantes de los datos, seleccionar un algoritmo de clasificación adecuado y entrenar el modelo utilizando ejemplos etiquetados.

Para esta práctica, se utilizó la herramienta de Kaggle para obtener los datos, en este caso se usó los datos de “Popular Video Oranges vs. Grapefruit”, con el compilador PyCharm ejecutamos el código.

El conjunto de datos contiene las siguientes columnas: name, diameter, weight, red, green, blue. Estas columnas representan diferentes atributos de las frutas, como el nombre, el diámetro, el peso y los componentes de color (rojo, verde y azul).

A continuación, realiza una exploración de los datos mediante la generación de gráficos. Utiliza la función `scatter()` de `matplotlib` para crear un gráfico de dispersión que muestra la relación entre el diámetro y el peso de las frutas. Cada punto en el gráfico representa una fruta, y se utiliza un color diferente para distinguir entre naranjas y pomelos.



