

INFORME DE RESULTADOS DEL ANALISIS DEL DATASET IRIS.

Iris, es uno de los datasets más conocidos en la ciencia de datos y la estadística, el cual sirve para analizar su conjunto de datos y realizar una serie de estadísticas descriptivas que ofrecen información relevante sobre las características físicas de las plantas de iris, dentro de lo que se encuentra:

Datos Característicos.

Dentro del conjunto de datos Iris se encuentran cuatro características (columnas de entrada) que describen las dimensiones de los sépalos y pétalos de la planta: sepal length (longitud del sépalo), sepal width (ancho del sépalo), petal length (longitud del pétalo), y petal width (ancho del pétalo). Estas características se encuentran en el objeto `X` según el código en python, mientras que el objetivo (`y`) representa la clasificación de cada flor en una de las tres especies de iris: Iris-virginica, Iris-setosa e Iris-versicolor.

Por medio del código elaborado y la información suministrada del dataset, se logra apreciar que los primeros cinco registros muestran cómo varían estas características. Un ejemplo de esto es en el primer registro que tiene una longitud de sépalo de 5.1 cm y un ancho de 3.5 cm, mientras que las dimensiones de sus pétalos son 1.4 cm de largo y 0.2 cm de ancho. Las especies de estas muestras iniciales pertenecen a Iris-setosa, una de las tres clases del conjunto de datos.

Análisis de las Estadísticas Descriptivas.

Centrándose en tres medidas clave: la media, la mediana y la desviación estándar de las características de las flores de iris, se logra concluir las siguientes estadísticas descriptivas:

- **Media:** La media de cada variable proporciona una referencia general de los valores centrales. De acuerdo con los resultados obtenidos, la longitud promedio del sépalo es de 5.84 cm, mientras que el ancho promedio es de 3.05 cm. En cuanto a los pétalos, la longitud media es de 3.76 cm y el ancho medio es de 1.20 cm. Estos datos sugieren que, en promedio, los sépalos son más grandes que los pétalos, lo cual es una característica frecuente en muchas especies florales.
- **Mediana** La mediana, al ser el valor central en un conjunto de datos ordenados, es una herramienta útil para comprender cómo se distribuyen las características y para minimizar el impacto de posibles valores atípicos. En este caso, las medianas son ligeramente inferiores a las medias, con una longitud del sépalo de 5.8 cm y un ancho de 3.0 cm. Las medianas de

los pétalos son 4.35 cm y 1.30 cm, respectivamente. Esto sugiere que los datos están distribuidos de manera relativamente simétrica, aunque con un leve sesgo hacia la derecha.

- **Desviación Estándar:** La desviación estándar cuantifica cuánto se alejan los datos de la media. Para las dimensiones de los sépalos, la desviación estándar es de 0.83 cm para la longitud y 0.43 cm para el ancho. En cambio, las características de los pétalos muestran una mayor variabilidad, con una desviación estándar de 1.76 cm para la longitud y 0.76 cm para el ancho. Esta mayor dispersión en las medidas de los pétalos podría indicar una mayor diversidad entre las especies en cuanto al tamaño de los pétalos.

Interpretación del Conjunto de Datos.

El análisis del conjunto de datos Iris ofrece una comprensión clara de las diferencias físicas entre las diversas especies de iris. Las medidas de tendencia central, como la media y la mediana, permiten tener una visión precisa del tamaño promedio de los sépalos y pétalos, mientras que la desviación estándar proporciona información sobre la variabilidad entre las especies. La menor variabilidad observada en las dimensiones de los sépalos en comparación con los pétalos podría estar relacionada con las diferentes funciones biológicas que cumplen estas estructuras dentro de la flor.

En cuanto a la clasificación, el hecho de que una de las clases de iris sea separable de las otras dos de manera lineal es un aspecto crucial en el análisis. Esto indica que las características de los sépalos y los pétalos varían lo suficiente entre las especies, lo que facilita una clasificación precisa en la mayoría de los casos, lo cual convierte a este conjunto de datos en un recurso fundamental para la investigación y desarrollo de técnicas de clasificación.