

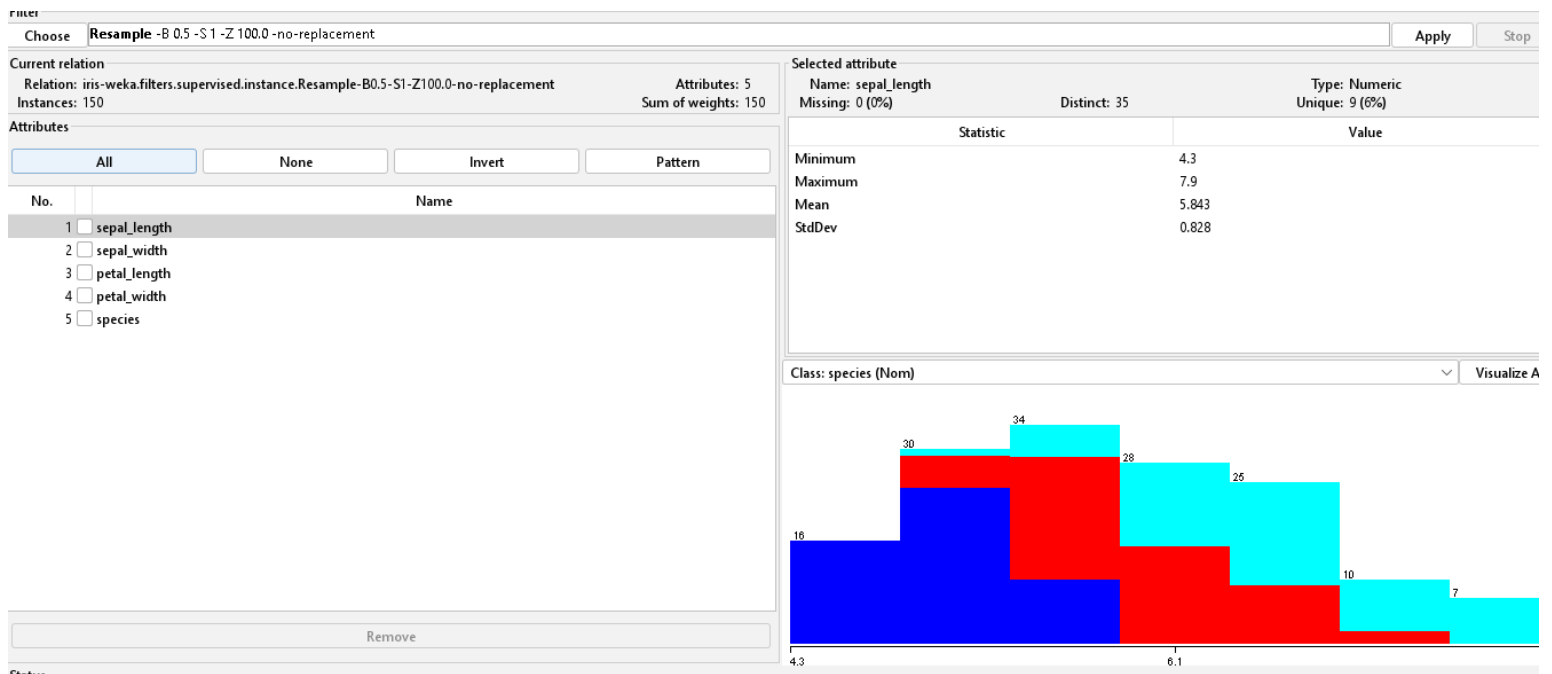
CHOQUE VEGA ALEX FRANCO

SUPERVISADOS:

Resample:

Descripción: Resample es un método que modifica la distribución de clases en el dataset. Puede usarse para equilibrar clases desbalanceadas mediante la sobre o sub-muestreo.

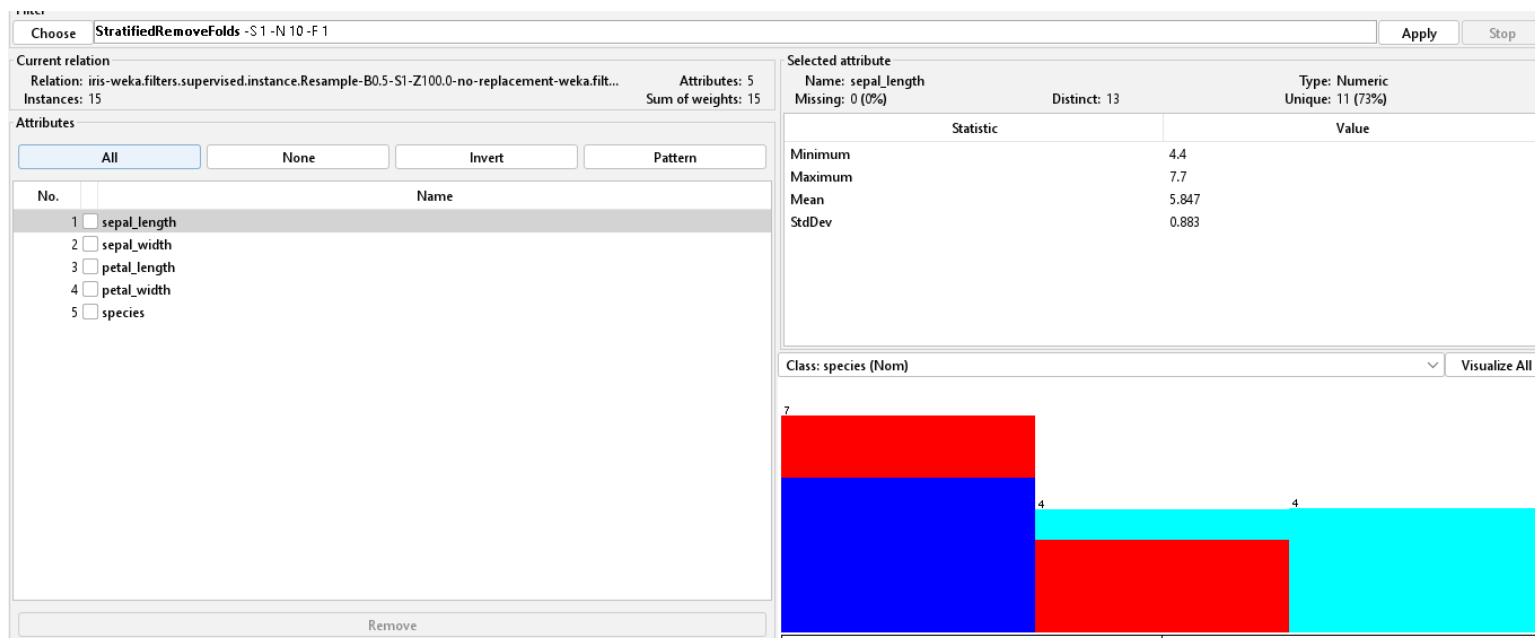
Razón de uso: Si tu dataset tiene clases desbalanceadas, es decir, algunas clases tienen significativamente más instancias que otras, esto puede afectar negativamente la precisión del modelo. Resample permite crear un dataset más balanceado, lo que puede mejorar el rendimiento de los algoritmos de clasificación al proporcionar una mejor representación de las clases minoritarias.



Stratified Remove Folds:

Descripción: Este método elimina un conjunto de pliegues (folds) del dataset, manteniendo la distribución de clases en cada pliegue.

Razón de uso: Se utiliza en el contexto de validación cruzada para evaluar el rendimiento de un modelo de manera más robusta. Al mantener la distribución de clases durante la eliminación de pliegues, se asegura que cada subconjunto utilizado en el entrenamiento y prueba del modelo sea representativo del dataset completo.



NO SUPERVISADO

Normalize:

Descripción: Normalize ajusta los valores de las características numéricas a una escala común, típicamente [0, 1] o [-1, 1].

Razón de uso: La normalización es crucial cuando las características tienen diferentes escalas y magnitudes. Los algoritmos de aprendizaje automático como k-NN y SVM pueden ser sensibles a la escala de los datos. Normalizar los datos puede mejorar la eficiencia y precisión de estos algoritmos.



Principal Component Analysis (PCA):

Descripción: PCA es una técnica de reducción de dimensionalidad que transforma las características originales en un nuevo conjunto de características no correlacionadas llamadas componentes principales.

Razón de uso: PCA se utiliza para reducir la cantidad de características en el dataset mientras se retiene la mayor cantidad posible de variabilidad en los datos. Esto no solo puede reducir el tiempo de entrenamiento del modelo, sino también mejorar su rendimiento al eliminar el ruido y la redundancia en los datos.

