

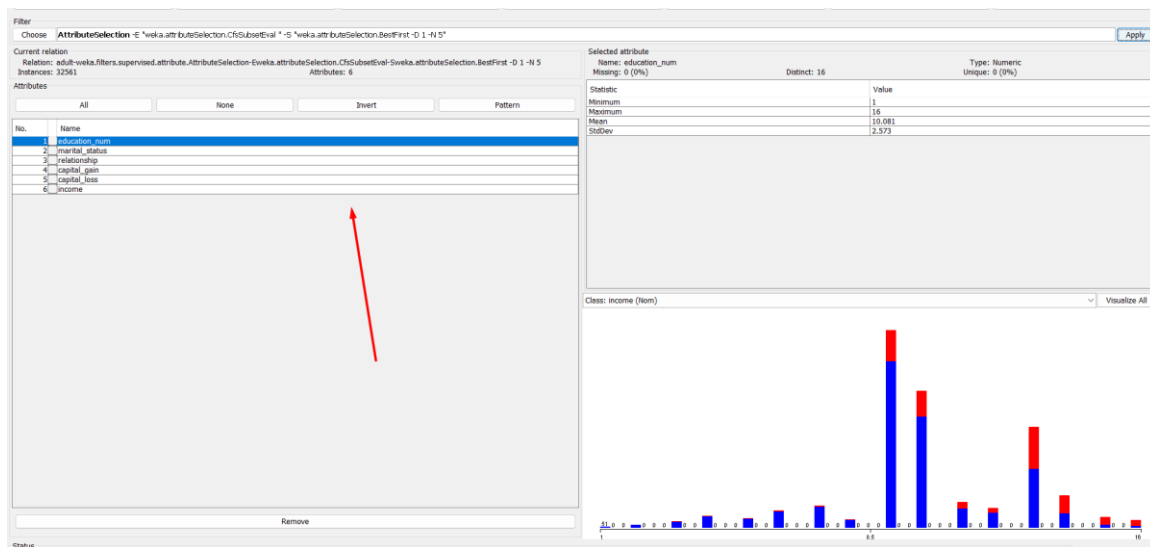
2. Del dataset del punto 1 realice en WEKA, cuatro algoritmos de preprocesamiento (2 supervisados y 2 no supervisados). Explique por cada uno la razón de su uso.

PREPROCESAMIENTOS SUPERVISADO

AttributeSelection (Selección de Atributos):

Esta técnica selecciona un subconjunto de atributos relevantes. Puedes utilizar un evaluador como CfsSubsetEval junto con un buscador como BestFirst para seleccionar los atributos más significativos.

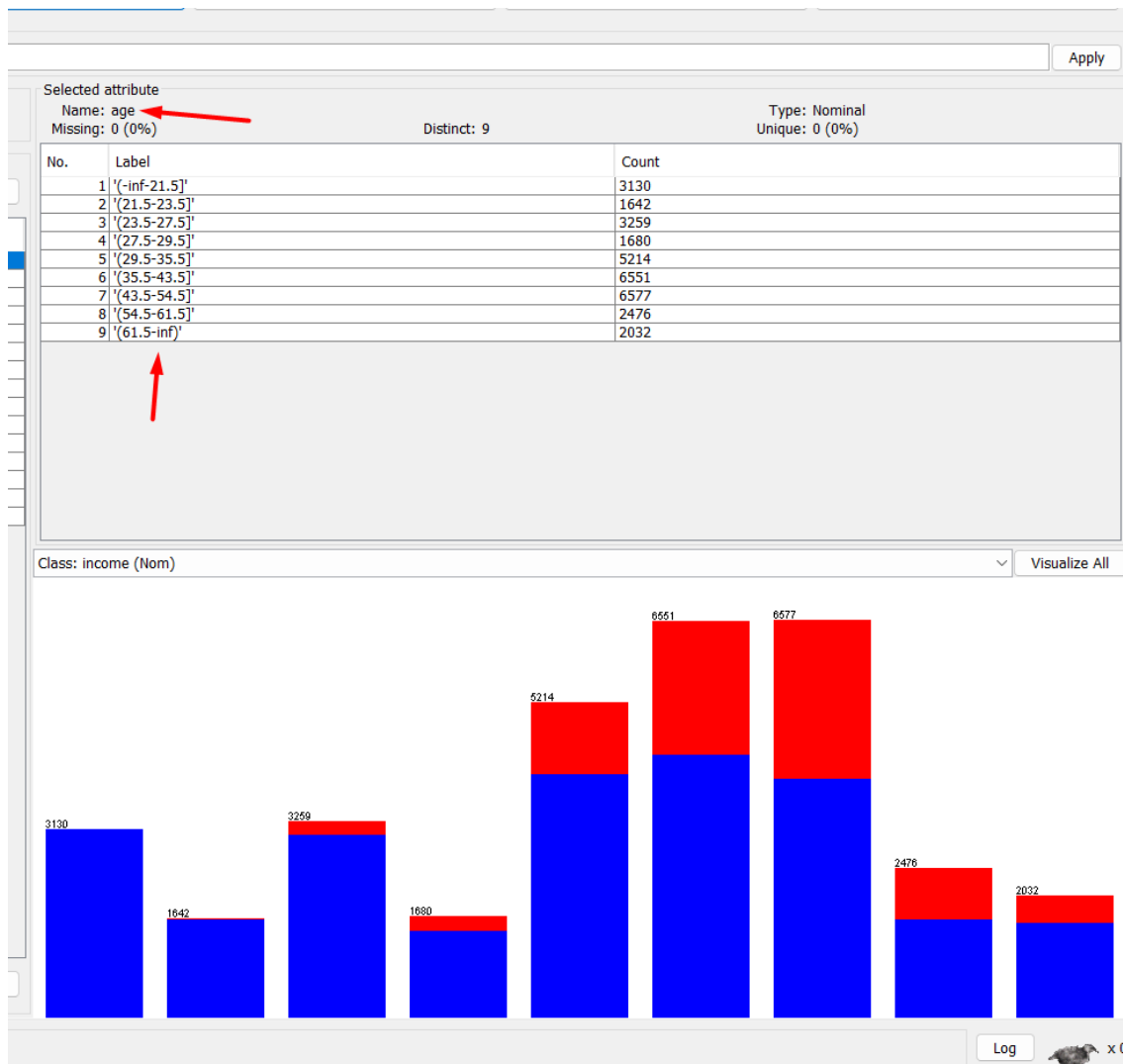
Adaptación al Dataset Adult: Este método ayuda a reducir la dimensionalidad y eliminar atributos irrelevantes, mejorando así la eficiencia y precisión del modelo de machine learning.



Discretize (Discretización):

Esta técnica convierte atributos numéricos en nominales mediante la creación de intervalos o "bins". Puedes configurar el número de intervalos y decidir si deseas que los intervalos tengan igual frecuencia.

Adaptación al Dataset Adult: Este método es útil para atributos numéricos continuos en el dataset, como la edad y los ingresos, facilitando el uso de algoritmos que funcionan mejor con atributos discretos. Además, ayuda a simplificar los datos y puede mejorar la interpretabilidad del modelo.



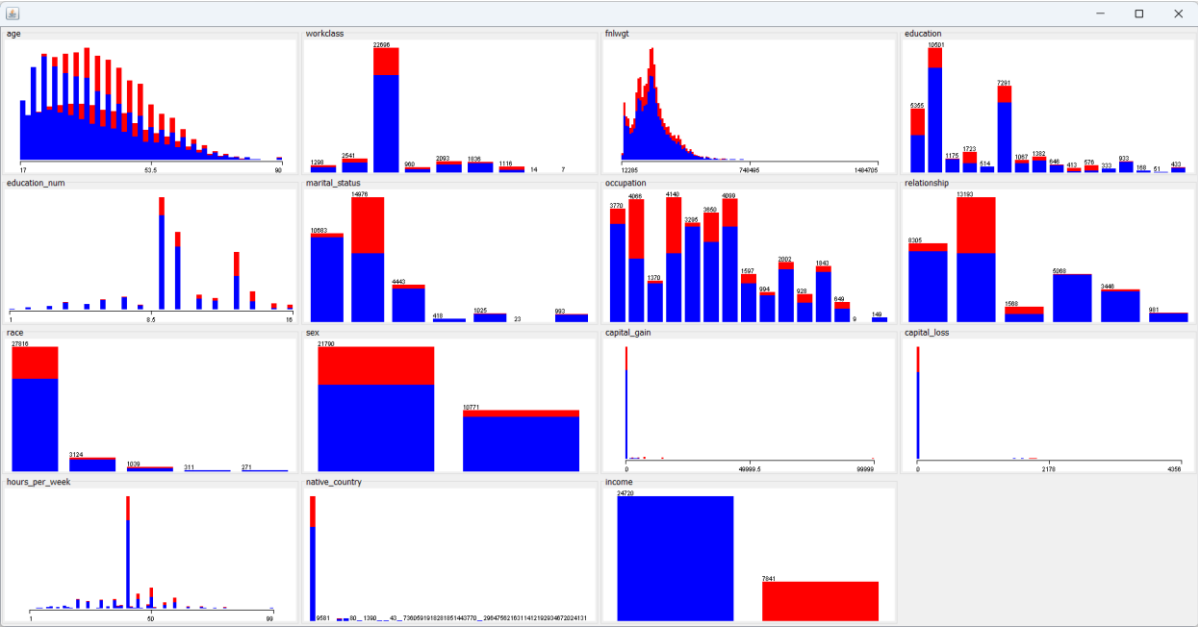
NO SUPERVIZADOS

ReplaceMissingValues

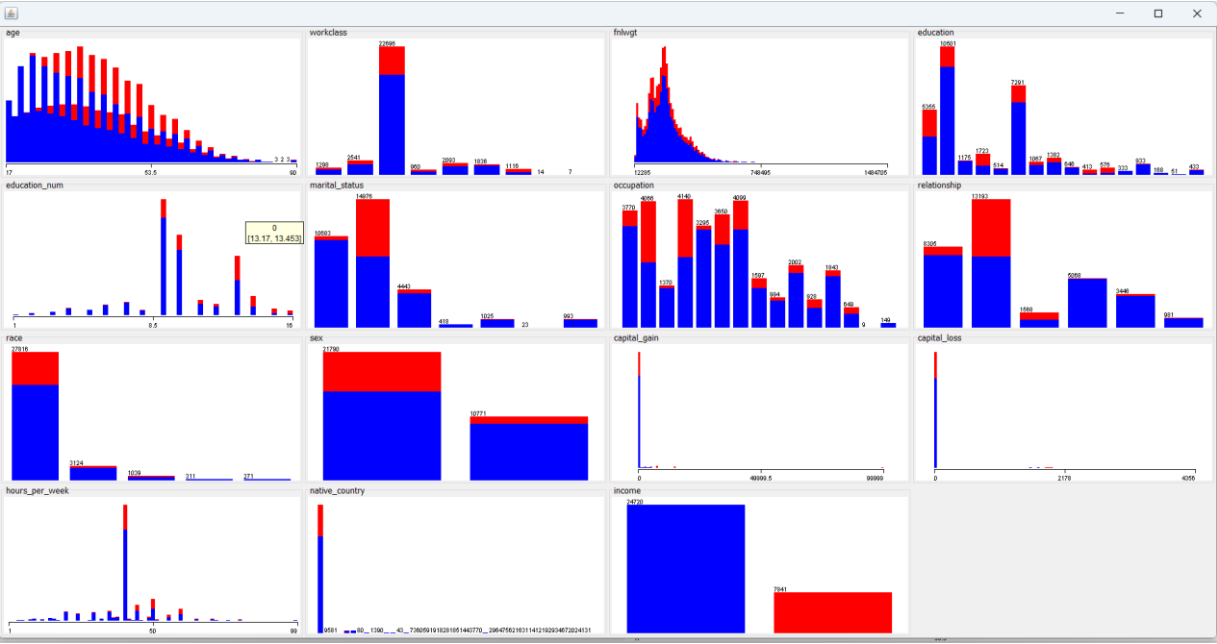
ReplaceMissingValues es un filtro en Weka que reemplaza los valores faltantes en los atributos de un dataset con valores adecuados. Para atributos numéricos, generalmente utiliza la media del atributo, y para atributos nominales, utiliza la moda (el valor más frecuente).

Adaptación al Dataset Adult, este método es particularmente útil para manejar datos incompletos en el dataset Adult, donde pueden haber registros con valores faltantes en atributos como "workclass", "occupation" o "native-country". Al reemplazar estos valores faltantes, se asegura que el dataset esté completo y listo para ser utilizado por los algoritmos de machine learning, evitando errores y mejorando la calidad del modelo.

ANTES



DESPUES



NumericToNominal (Numérico a Nominal)

NumericToNominal es un filtro en Weka que convierte atributos numéricos en atributos nominales. Este filtro es útil cuando se desea tratar atributos numéricos como categóricos para utilizar algoritmos que funcionan mejor con datos nominales o para simplificar la interpretación de los resultados.

En el dataset Adult, podríamos querer convertir ciertos atributos numéricos en nominales. Por ejemplo, convertir la edad en rangos de edad (20-29, 30-39, etc.) o convertir el número de horas trabajadas por semana en categorías (bajo, medio, alto). Esto puede facilitar el análisis y el modelado de datos.

