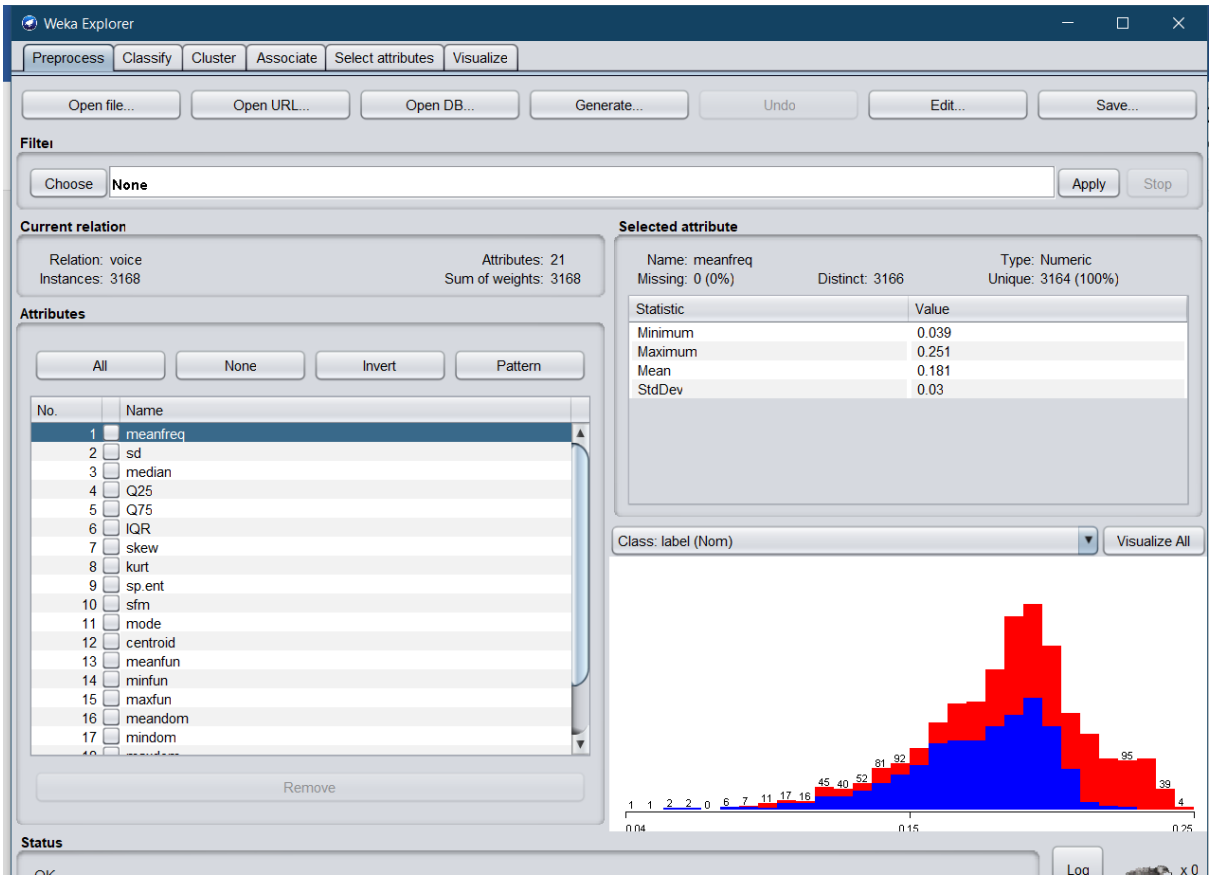
ALEXANDER HUMBERTO NINA PACAJES 5950236 LP

1. Abriendo el voice.csv seleccionado

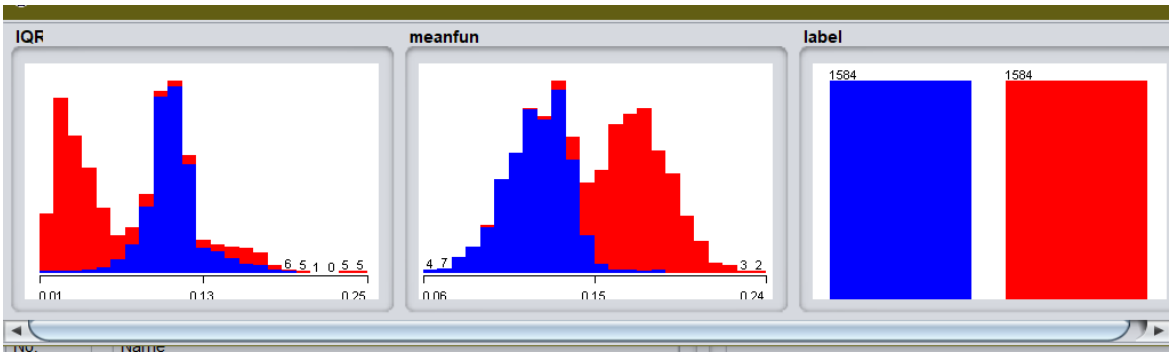


2. Algoritmos de Preprocesamiento (3 Supervisado)

Atribute Selection

Entre los subconjuntos de atributos calcula el grado de redundancia y capacidad predictiva individual, (escalada codiciosa aumentada con una función de retroceso.) backtracking

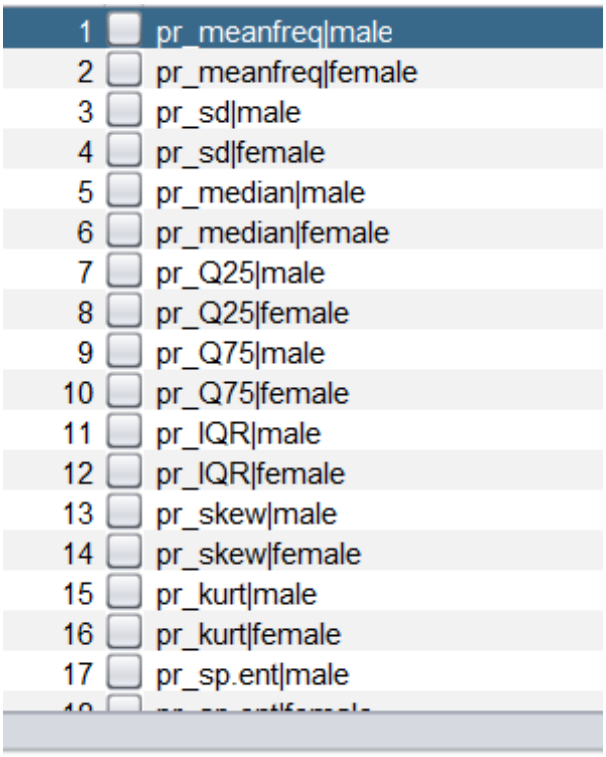
Aplicado el filtro tenemos los atributos con mas importancia (perdida de precisión) del dataset:



Siendo el IQR y meanfun los mas relevantes a la hora de revisión de algún clasificador. Comparando con los atributos eliminados

ClassConditionalValues

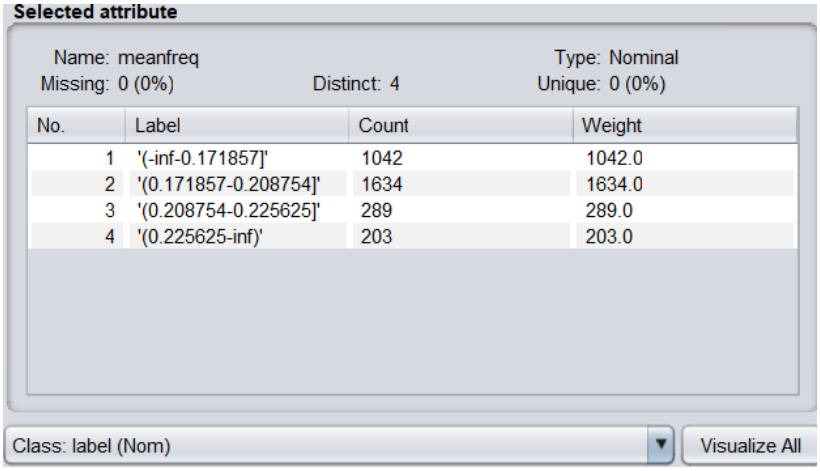
Simplemente separa los atributos por el tipo de clase es decir en este dataset se usa de output target ‘Masculino’ y ‘Femenino’

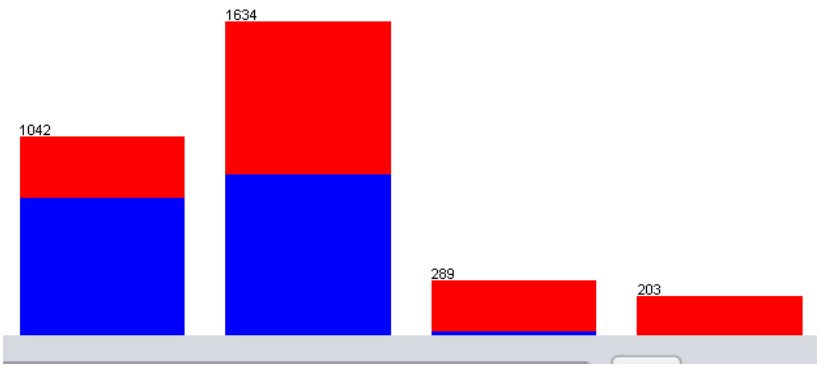


Creando valores nominales de los númericos usando el target output o la clase.

Discretize

Convertir los datos numéricos en nominales categóricos, en este dataset aplico división por intervalo del mismo tamaño como se ve la imagen



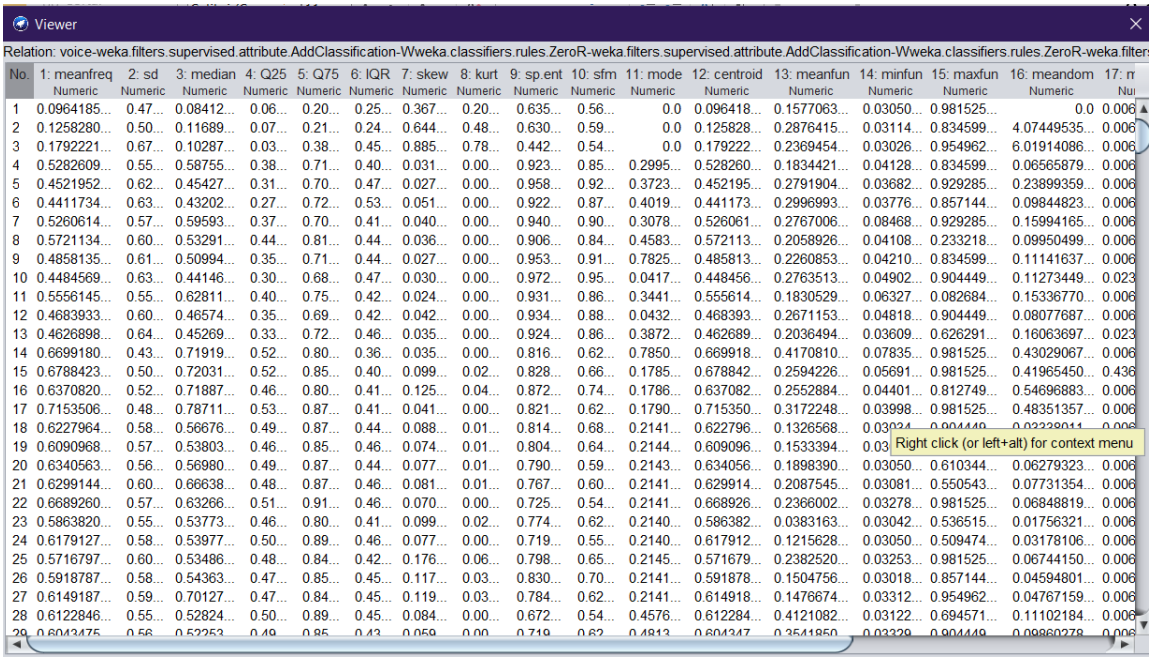


Sin supervisión (3)

Por ser Weka una herramienta de data mining, los algoritmos para preprocesamiento son más amplios

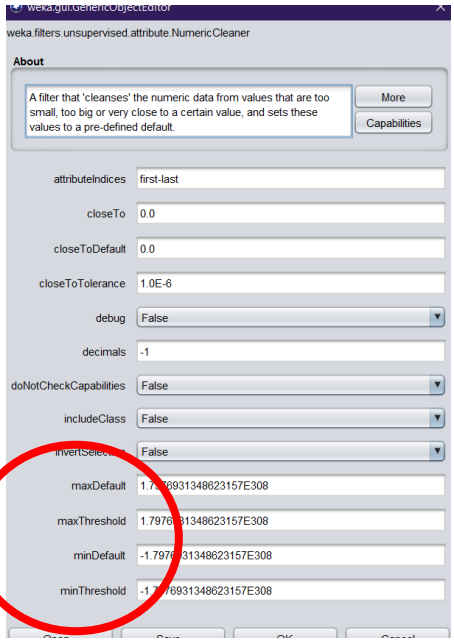
Normalize

Este es uno de los mas usados según el algoritmo de clasificación que se use, el único que conozco que actúa mejor con atributos normalizados es el kNN, simplemente le dará un rango a todas las variables de forma que no tengan mas relevancia si los valores son mayores



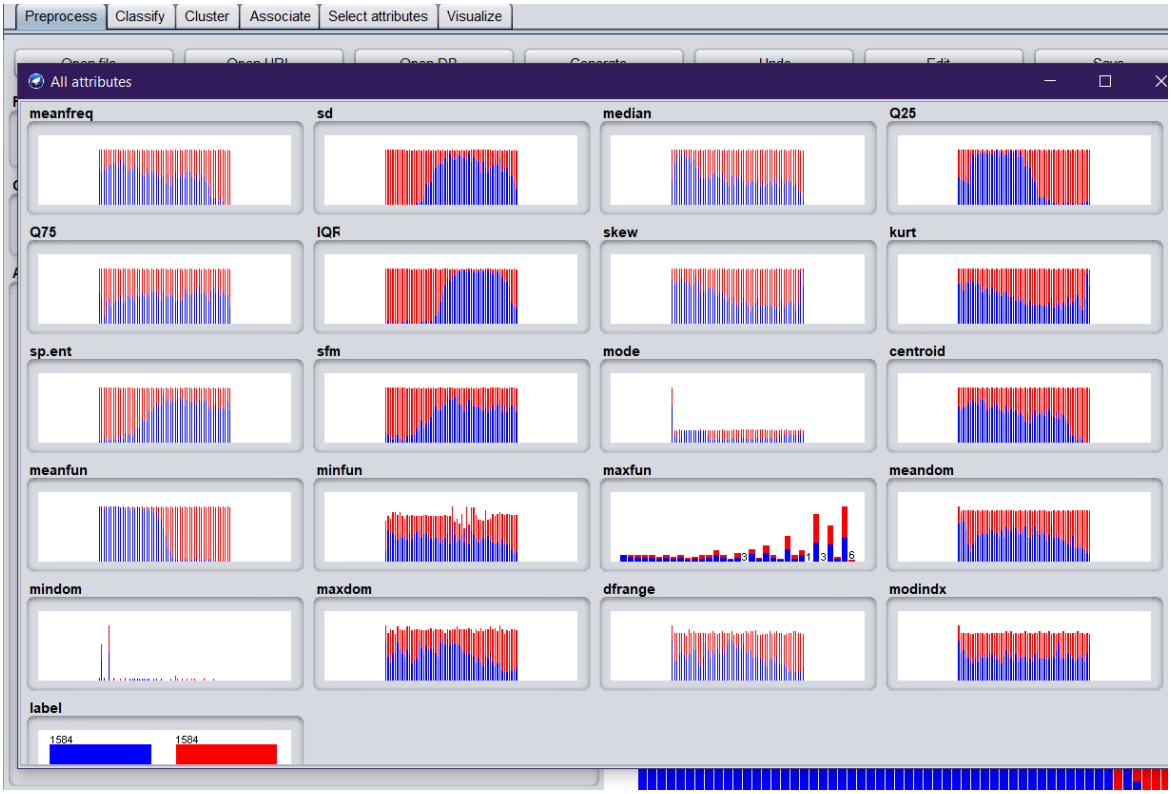
NumericCleaner

Este hará que los valores demasiado pequeños sean ignorados, eliminándolos de los examples. Como se ve en la imagen eliminará valores demasiado pequeños que no 'afectan' en gran medida la predicción



PKIDiscretize:

Esta forma de discretización es una 'mejora' del método que mostré más arriba, ahora con este se puede hacer binning de atributos con la misma frecuencia en sus examples como lo dice 'obliga a que el número de bins sea igual a la raíz cuadrada del número de valores del atributo numérico.' Su comportamiento como se espera será entonces


Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
All attributes
meanfreq
sd
median
Q25
Q75
IQR
skew
kurt
sp.ent
sfm
mode
centroid
meanfun
minfun
maxfun
meandom
mindom
maxdom
dfrange
modindx
label
1584
1584