

IDENTIFY OUTLIERS/EXTREME VALUES AND REMOVE IT

Step 1: Open Weka

Step 2: Open DataBase

The screenshot shows the Weka Explorer window. The 'Filter' tab is selected, and the filter 'InterquartileRange -R first-last-O 3.0 -E 6.0' is applied. The 'Current relation' is 'pima_diabetes' with 768 instances and 9 attributes. The 'Attributes' list shows 9 attributes: preg, plas, pres, skin, insu, mass, pedi, age, and class. The 'Selected attribute' panel shows statistics for 'preg' (Minimum: 0, Maximum: 17, Mean: 3.845, StdDev: 3.37). A histogram of the 'preg' attribute is displayed at the bottom right.

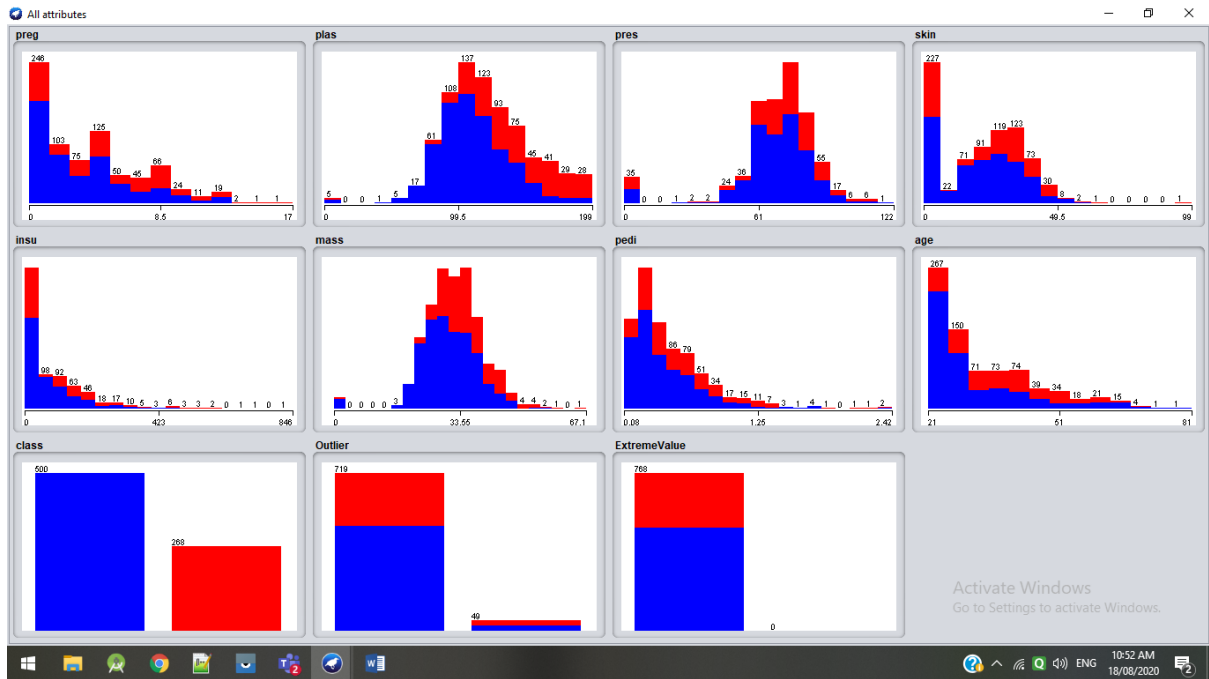
Step 3: Observe the value of no. of attributes.

Click on choose -> Filter -> Unsupervised -> Attribute -> InterquartileRange

Click on apply

Observe the value of no. of attributes

The screenshot shows the Weka Explorer window after applying the 'InterquartileRange' filter. The 'Current relation' is 'pima_diabetes-weka.filters.unsupervised.attribute.InterquartileRange-Rfirst-la...' with 768 instances and 11 attributes. The 'Attributes' list now includes 11 attributes: preg, plas, pres, skin, insu, mass, pedi, age, class, Outlier, and ExtremeValue. The 'Selected attribute' panel shows the same statistics for 'preg'. A histogram of the 'preg' attribute is displayed at the bottom right.



Step 4: To remove

Click on choose -> Filter -> Unsupervised -> Instances-> RemoveWithValues

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **RemoveWithValues** -S 0.0 -C 11 -L last Apply Stop

Current relation: Relation: pima_diabetes-weka.filters.unsupervised.attribute.InterquartileRange-Rfirst-la... Attributes: 11 Instances: 768 Sum of weights: 768

Attributes:

No.	Name
1	preg
2	plas
3	pres
4	skin
5	insu
6	mass
7	pedi
8	age
9	class
10	<input checked="" type="checkbox"/> Outlier
11	<input type="checkbox"/> ExtremeValue

Remove

Selected attribute:

No.	Label	Count	Weight
1	no	719	719.0
2	yes	49	49.0

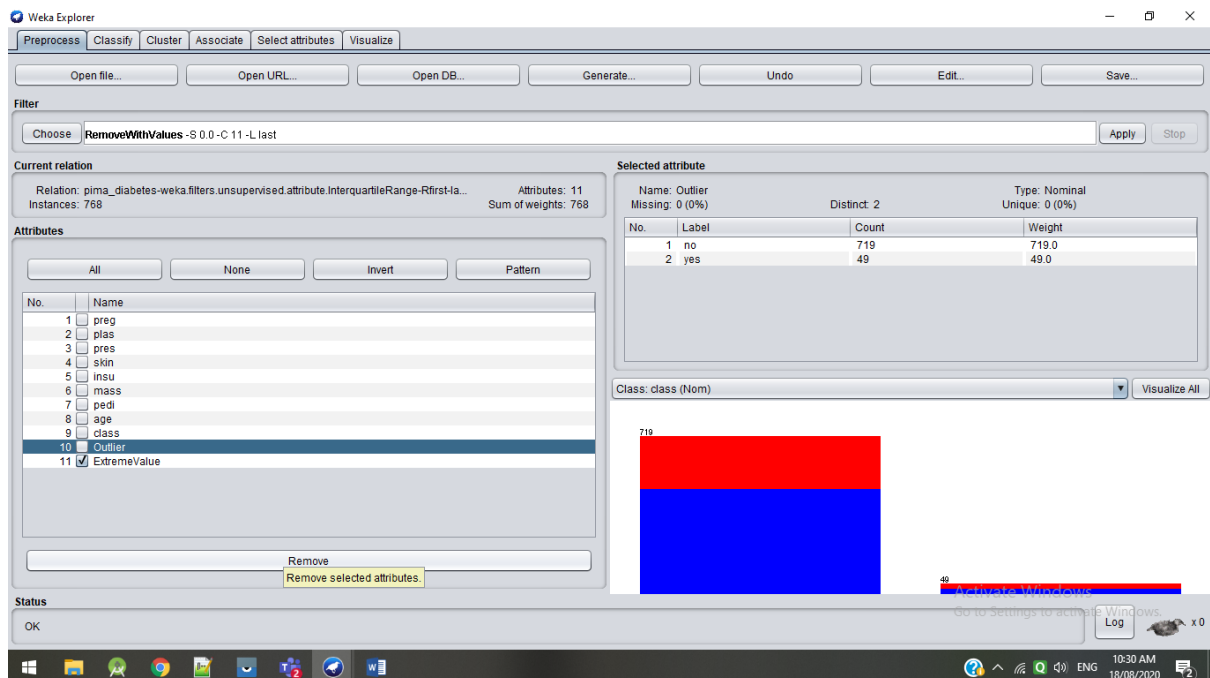
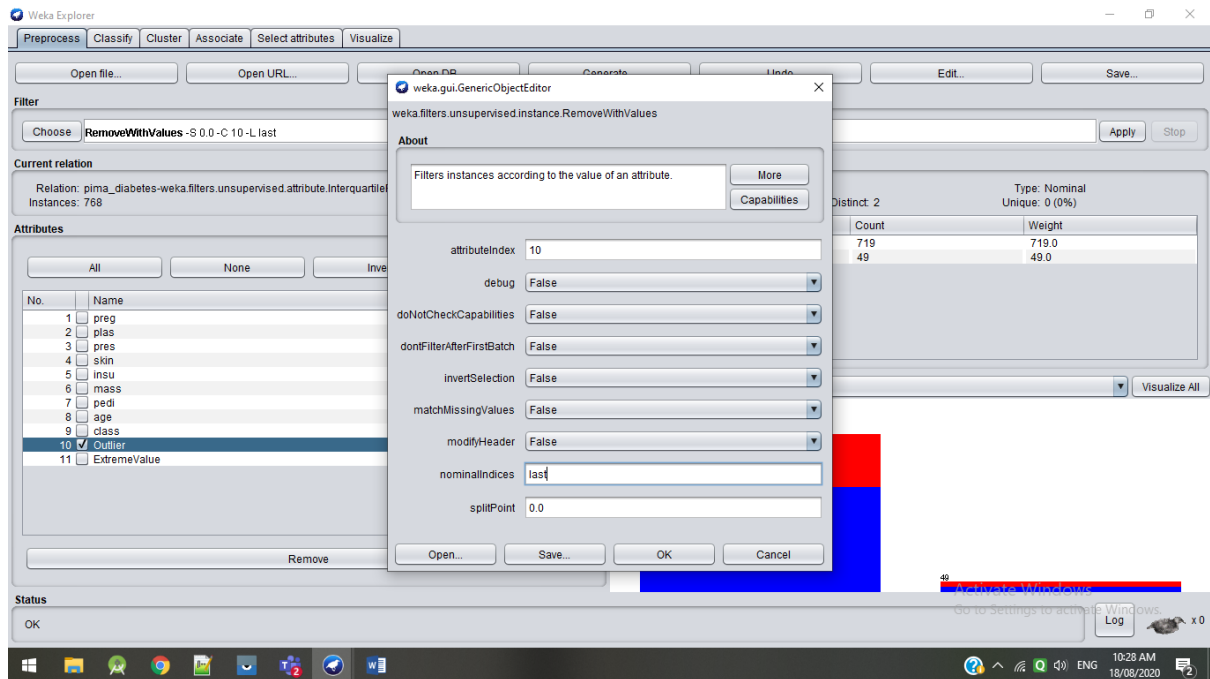
Class: class (Nom) Visualize All

Status: OK

Step 5: Select Outlier ->Click on which is followed by RemoveWithValues

Set attributeIndex: 10

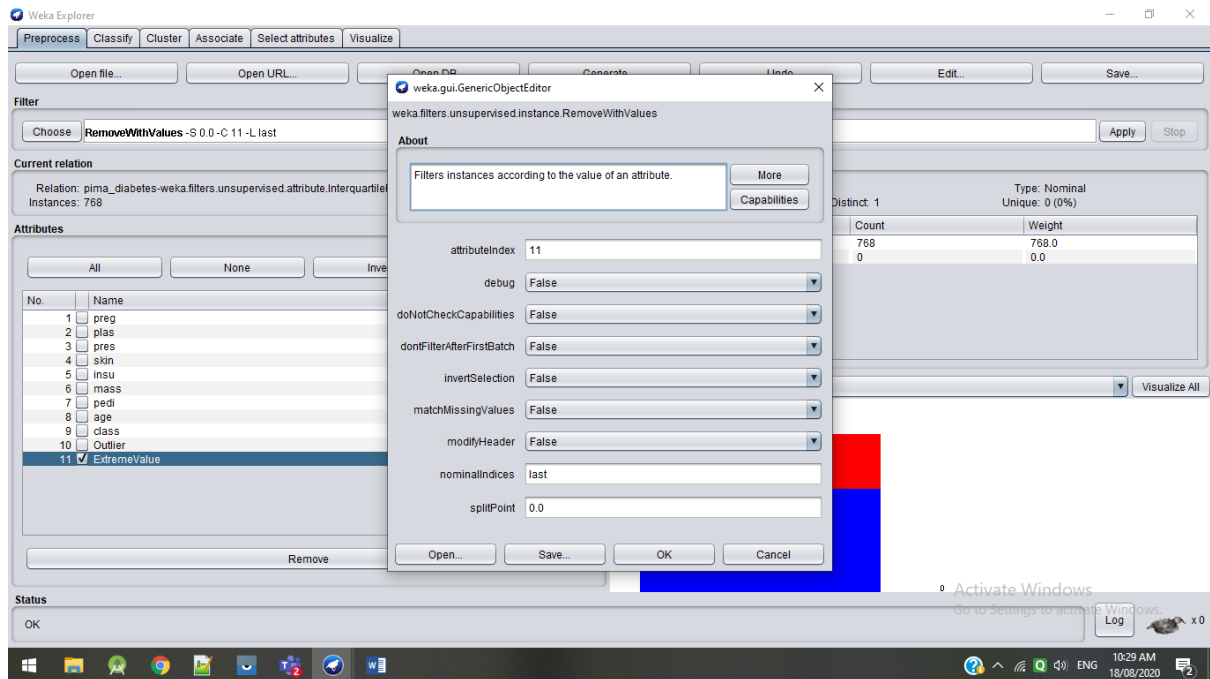
nominalIndices: last



Step 6: Select Extreme Valuesr ->Click on which is followed by RemoveWithValues

Set attributeIndex: 11

nominalIndices: last



Step 7: Click on Apply and see the result

