

ASSIGNMENT ON OUTLIER

Apply the outlier detection to your chosen dataset and remove the outliers with extreme values

OUTLIER:

The screenshot shows the Orange3 software interface. On the left, a widget palette lists various data processing tools under 'unsupervised' and 'attribute' categories. The 'Outlier Detection' widget is highlighted with a yellow circle. In the center, a data preview window shows a table with two columns: 'No.' and 'Label'. The table contains two rows: '1 SPAM' and '2 HAM'. Below the table, a bar chart shows the distribution of the 'class' variable, with a blue bar for '1' and a red bar for '2'. The 'Class: class (Nom)' label is visible. At the bottom, a configuration dialog for the 'Outlier Detection' widget is open. It contains several settings: 'attributeIndices' is set to 'first-last', 'debug' is 'False', 'detectionPerAttribute' is 'False', 'extremeValuesAsOutliers' is 'False', 'extremeValuesFactor' is '6.0', 'outlierFactor' is '3.0', and 'outputOffsetMultiplier' is 'False'. The dialog has buttons for 'Open...', 'Save...', 'OK', and 'Cancel'.

unsupervised
attribute
Add
AddCluster
AddExpression
AddID
AddNoise
AddValues
Center
ChangeDateFormat
ClassAssigner
ClusterMembership
Copy
Discretize
FirstOrder
InterquartileRange
KernelFilter
MakeIndicator

No.	Label
1	SPAM
2	HAM

Class: class (Nom)

1378 29

attributeIndices first-last
debug False
detectionPerAttribute False
extremeValuesAsOutliers False
extremeValuesFactor 6.0
outlierFactor 3.0
outputOffsetMultiplier False

Open... Save... OK Cancel

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

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

Filter
Choose: InterquartileRange -R first-last -O 3.0 -E 6.0 Apply

Current relation
Relation: spam-detection-training-weka.filters.unsupervised.attribute....
Instances: 4326 Attributes: 23

Attributes
All None Invert Pattern

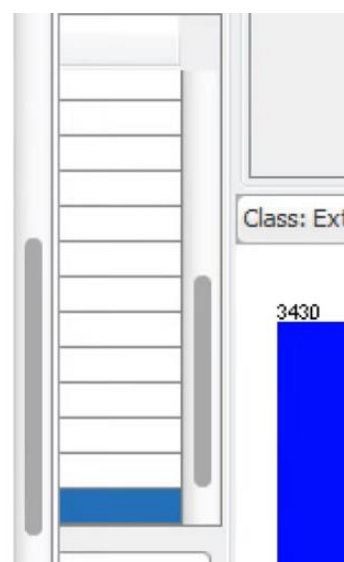
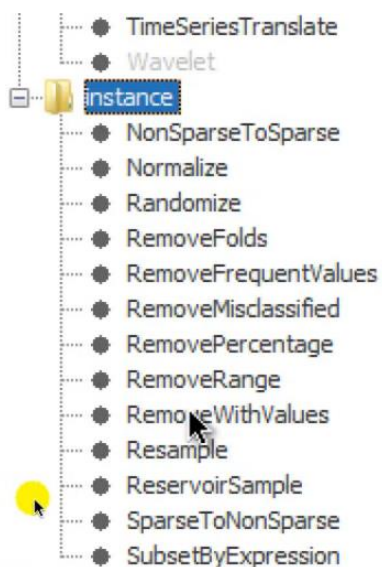
No.	Name
11	<input type="checkbox"/> document_length
12	<input type="checkbox"/> word_length
13	<input type="checkbox"/> word_length_WSW
14	<input type="checkbox"/> tf_idf
15	<input type="checkbox"/> tf_idf_WSW
16	<input type="checkbox"/> alpha_numeric
17	<input type="checkbox"/> alpha_numeric_WSW
18	<input type="checkbox"/> verb
19	<input type="checkbox"/> verb_WSW
20	<input type="checkbox"/> stop_word
21	<input type="checkbox"/> class
22	<input checked="" type="checkbox"/> Outlier
23	<input type="checkbox"/> ExtremeValue

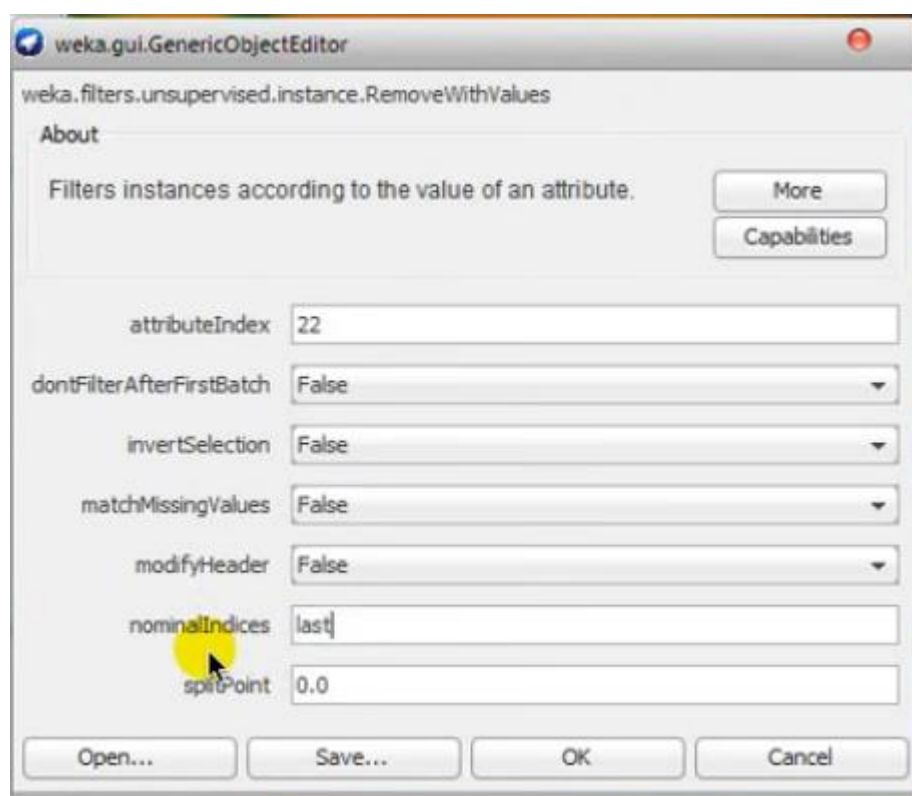
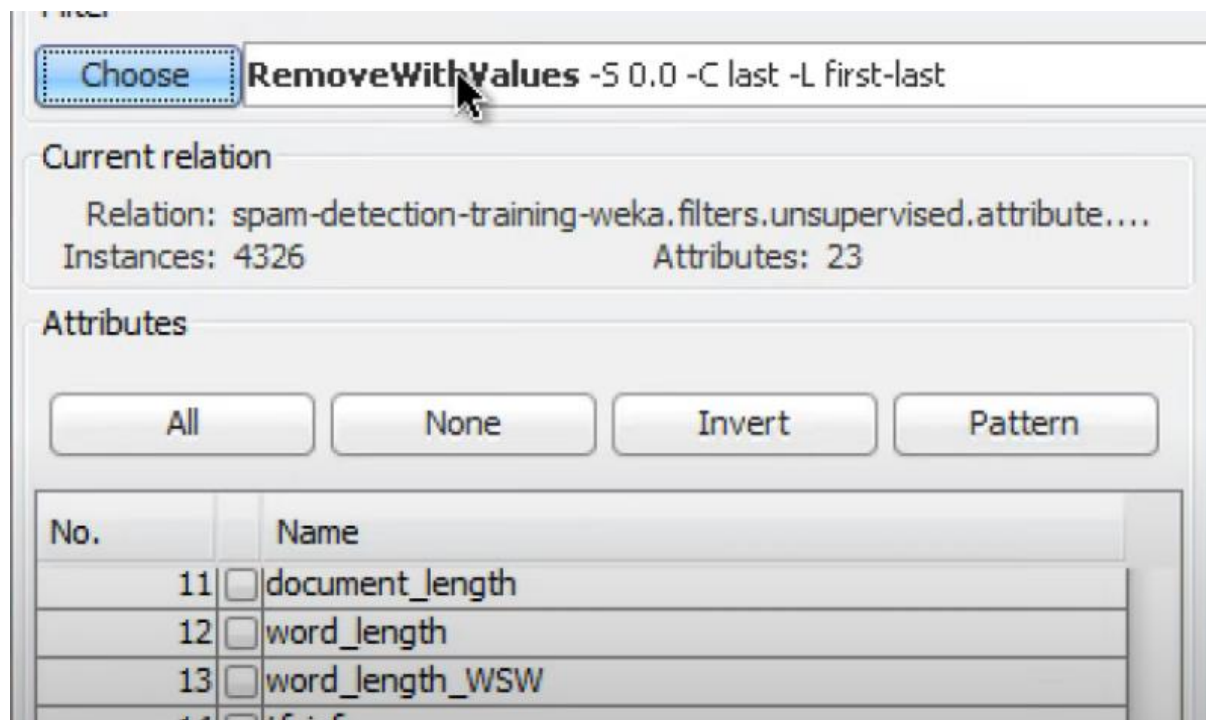
Remove

Selected attribute
Name: Outlier
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

No.	Label	Count
1	no	3552
2	yes	774

Class: ExtremeValue (Nom) Visualize All





Preprocess **Classify** Cluster Associate Select attributes Visualize

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

Filter
Choose **RemoveWithValues -S 0.0 -C 22 -L last** Apply

Current relation
Relation: spam-detection-training-weka.filters.unsupervised.attribute....
Instances: 3552 Attributes: 23

Attributes
All None Invert Pattern

No.	Name
11	<input type="checkbox"/> document_length
12	<input type="checkbox"/> word_length
13	<input type="checkbox"/> word_length_WSW
14	<input type="checkbox"/> tf_isf
15	<input type="checkbox"/> tf_isf_WSW
16	<input type="checkbox"/> alpha_numeric
17	<input type="checkbox"/> alpha_numeric_WSW
18	<input type="checkbox"/> verb
19	<input type="checkbox"/> verb_WSW
20	<input type="checkbox"/> stop_word
21	<input type="checkbox"/> class
22	<input checked="" type="checkbox"/> Outlier
23	<input type="checkbox"/> ExtremeValue

Remove

Selected attribute
Name: Outlier
Missing: 0 (0%) Distinct: 1 Type: Nominal
Unique: 0 (0%)

No.	Label	Count
1	no	3552
2	yes	0

Class: Outlier (Nom) Visualize All

```

17.52,15.07,38.46,23.4,0.06,0.07,66,57,346,346,17,1,1,1567.9,820.25,54,54,18,16,290,HAM,no,no
18.7,19.98,32.79,17.09,0.05,0.05,37,35,105,105,7,1,2,362,223.04,7,7,4,2,96,HAM,no,no
27.72,22.73,45.5,26.37,0.04,0.04,28,24,96,96,3,1,2,208.74,85.69,21,21,2,1,76,HAM,no,no
22.9,22.71,36.23,31.17,0.04,0.04,16,16,38,38,2,1,2,45.05,38.12,12,12,0,0,14,HAM,no,no
14.12,12.66,36.87,28.42,0.07,0.08,5,5,34,34,2,1,1,32.58,23.57,8,8,0,0,18,HAM,no,no
50.24,47.53,71.35,68.12,0.02,0.02,19,18,76,76,1,1,1,0,0,56,56,0,0,7,SPAM,no,no
18.34,20.01,35.57,18.35,0.05,0.05,61,55,201,201,12,1,2,903.15,439.32,6,6,11,7,180,HAM,no,no
45.79,29.18,69.83,50.35,0.02,0.03,17,16,145,145,2,1,1,338.95,120.61,41,41,3,1,94,SPAM,no,no
10.73,12.35,30.66,17.87,0.09,0.08,34,30,170,170,19,1,2,953.42,499.8,5,5,6,4,150,SPAM,no,no
30.55,20.41,52.46,32.25,0.03,0.05,20,18,129,129,3,1,1,368.04,293.33,11,11,3,3,101,SPAM,no,no
25.93,21.18,45.03,29.81,0.04,0.05,30,28,125,125,4,1,1,228.74,134.47,15,15,1,0,97,HAM,no,no
15.36,17.75,30.5,20.39,0.07,0.06,72,68,195,195,20,1,2,1251.15,763.5,18,18,18,14,128,SPAM,no,no
13.9,14.23,28.9,19.94,0.07,0.07,3,2,9,1,1,2,0,0,0,0,0,0,9,HAM,no,no

```

<https://www.youtube.com/watch?v=WripO7CmUoQ>