



VIT[®]
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

DATA WAREHOUSING AND DATA MINING

NAME : MOTHISHWARAN C.
REG No : 19MID0017
COURSE CODE : CSI3010
SLOT : L39+L40
FACULTY : DR. CHELLATHAMILAN T.

OUTLIERS AND EXTREME VALUES DETECTION USING WEKA

Dataset: <https://www.kaggle.com/ramjasmaurya/upi-apps-transactions-in-2021>

No	1:UPI Banks	2: Volume (Mn) By Costumers	3: Value (Cr) by Costumers	4: Volume (Mn)	5: Value (Cr)	6: Month	7: Year
1	Airtel Pay...	9.04	2248.49	15.12	5732.09	12.0	2021.0
2	Airtel Pay...	8.41	1923.95	15.9	5554.63	11.0	2021.0
3	Airtel Pay...	7.6	1491.34	14.31	4746.79	10.0	2021.0
4	Airtel Pay...	6.91	1520.92	11.52	3759.12	9.0	2021.0
5	Airtel Pay...	6.14	1396.65	9.31	2941.88	8.0	2021.0
6	Airtel Pay...	5.79	1534.21	7.36	2342.76	7.0	2021.0
7	Airtel Pay...	5.06	1323.26	6.69	2821.94	6.0	2021.0
8	Airtel Pay...	4.06	429.12	4.06	429.12	5.0	2021.0
9	Airtel Pay...	3.53	363.66	14.7	11172.43	4.0	2021.0
10	Airtel Pay...	3.79	381.74	3.79	381.74	3.0	2021.0
11	Airtel Pay...	3.33	348.7	3.33	348.7	2.0	2021.0
12	Airtel Pay...	3.71	355.77	3.71	355.77	1.0	2021.0
13	Allahabad...	0.02	7.43	0.02	7.43	12.0	2021.0
14	Allahabad...	0.02	6.28	0.02	6.28	11.0	2021.0
15	Allahabad...	0.02	6.35	0.02	6.35	10.0	2021.0
16	Allahabad...	0.02	5.34	0.02	5.34	9.0	2021.0
17	Allahabad...	0.01	4.44	0.01	4.44	8.0	2021.0
18	Allahabad...	0.01	2.03	0.01	2.03	7.0	2021.0
19	Allahabad...	0.01	4.43	0.01	4.43	5.0	2021.0
20	Allahabad...	0.02	5.21	0.02	5.21	4.0	2021.0
21	Allahabad...	0.05	13.51	0.05	13.51	3.0	2021.0
22	Allahabad...	0.05	12.23	0.05	12.23	2.0	2021.0
23	Allahabad...	0.09	21.14	0.09	21.14	1.0	2021.0
24	Amazon Pay	76.18	6641.62	76.18	6641.62	12.0	2021.0
25	Amazon Pay	71.04	6041.96	71.04	6041.96	11.0	2021.0
26	Amazon Pay	68.82	6286.6	68.82	6286.6	10.0	2021.0
27	Amazon Pay	62.9	5267.79	62.9	5267.79	9.0	2021.0
28	Amazon Pay	60.67	5297.16	60.67	5297.16	8.0	2021.0
29	Amazon Pay	62.83	5223.57	62.83	5223.57	7.0	2021.0
30	Amazon Pay	51.24	4914.74	51.24	4914.74	6.0	2021.0

Weka Explorer

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

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

Filter: Choose **None** Apply Stop

Current relation
 Relation: UPI apps transaction data in 2021
 Instances: 654
 Attributes: 7
 Sum of weights: 654

Attributes
 All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> UPI Banks
2	<input type="checkbox"/> Volume (Mn) By Costumers
3	<input type="checkbox"/> Value (Cr) by Costumers
4	<input type="checkbox"/> Volume (Mn)
5	<input type="checkbox"/> Value (Cr)
6	<input type="checkbox"/> Month
7	<input type="checkbox"/> Year

Remove

Selected attribute
 Name: UPI Banks
 Missing: 0 (0%)
 Distinct: 69
 Type: Nominal
 Unique: 7 (1%)

No.	Label	Count	Weight
1	Airtel Payments Bank...	12	12.0
2	Allahabad Bank App	11	11.0
3	Amazon Pay	12	12.0
4	Andhra Bank App	1	1.0
5	AU Small Finance Ba...	12	12.0
6	Axis Bank Apps	12	12.0
7	Bajaj Finserv	12	12.0
8	Bank of Baroda Apps	12	12.0
9	Bank of India App	9	9.0
10	Bank of Maharashtra	12	12.0

Class: Year (Num) Visualize All

Status: OK Log x 0

Choosing Inter Quartile Range Filter:

Weka Explorer

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

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

Filter: Apply Stop

Selected attribute
 Name: UPI Banks
 Missing: 0 (0%)
 Distinct: 69
 Type: Nominal
 Unique: 7 (1%)

No.	Label	Count	Weight
1	Airtel Payments Bank...	12	12.0
2	Allahabad Bank App	11	11.0
3	Amazon Pay	12	12.0
4	Andhra Bank App	1	1.0
5	AU Small Finance Ba...	12	12.0
6	Axis Bank Apps	12	12.0
7	Bajaj Finserv	12	12.0
8	Bank of Baroda Apps	12	12.0

Filter

- MultiFilter
- RenameRelation
- supervised
- unsupervised
 - attribute
 - Add
 - AddCluster
 - AddExpression
 - AddID
 - AddNoise
 - AddUserFields
 - AddValues
 - CartesianProduct
 - Center
 - Charn
 - Class
 - Class
 - Cluster
 - Copy
 - Date
 - Discretize
 - FirstC
 - Fixed
 - InterQuartileRange
 - Kernal
 - Makel
 - MathExpression

A filter for performing the Cartesian product of a set of nominal attributes
 The weight of the new Cartesian product attribute is the sum of the weights of the combined attributes.

CAPABILITIES
 Class -- Binary class, Date class, Empty nominal class, Missing class values, No class, Nominal class, Numeric class, Relational class, String class, Unary class

Attributes -- Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Relational attributes, String attributes, Unary attributes

Additional
 Minimum number of instances: 0

Filter... Remove filter Close

Status: OK Log x 0

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

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

Filter: Choose InterquartileRange -R 2,3,4,5-

Current relation: Relation: UPI apps transaction data in 2021 Instances: 654

Attributes: All None

No.	Name
1	<input checked="" type="checkbox"/> UPI Banks
2	<input type="checkbox"/> Volume (Mn) By Costumers
3	<input type="checkbox"/> Value (Cr) by Costumers
4	<input type="checkbox"/> Volume (Mn)
5	<input type="checkbox"/> Value (Cr)
6	<input type="checkbox"/> Month
7	<input type="checkbox"/> Year

Remove

Status: OK Log x 0

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.InterquartileRange

About

A filter for detecting outliers and extreme values based on interquartile ranges.

More Capabilities

attributeIndices: 2,3,4,5

debug: False

detectionPerAttribute: False

doNotCheckCapabilities: False

extremeValuesAsOutliers: False

extremeValuesFactor: 6.0

outlierFactor: 3.0

outputOffsetMultiplier: False

Open... Save... OK Cancel

Type: Nominal Unique: 7 (1%)

Weight
12.0
11.0
12.0
1.0
12.0
12.0
12.0
12.0
9.0
12.0

Visualize All

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

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

Filter: Choose InterquartileRange -R 2,3,4,5 -O 3.0 -E 6.0

Current relation: Relation: UPI apps transaction data in 2021 Instances: 654

Attributes: All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> UPI Banks
2	<input type="checkbox"/> Volume (Mn) By Costumers
3	<input type="checkbox"/> Value (Cr) by Costumers
4	<input type="checkbox"/> Volume (Mn)
5	<input type="checkbox"/> Value (Cr)
6	<input type="checkbox"/> Month
7	<input type="checkbox"/> Year

Remove

Status: OK Log x 0

Selected attribute

Name: UPI Banks Missing: 0 (0%) Distinct: 69 Type: Nominal Unique: 7 (1%)

No.	Label	Count	Weight
1	Airtel Payments Bank...	12	12.0
2	Allahabad Bank App	11	11.0
3	Amazon Pay	12	12.0
4	Andhra Bank App	1	1.0
5	AU Small Finance Ba...	12	12.0
6	Axis Bank Apps	12	12.0
7	Bajaj Finserv	12	12.0
8	Bank of Baroda Apps	12	12.0
9	Bank of India App	9	9.0
10	Bank of Maharashtra	12	12.0

Class: Year (Num)

Visualize All

Outlier & Extreme value Attribute:

Weka Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Open file...Open URL...Open DB...Generate...UndoEdit...Save...

Filter: Choose InterquartileRange -R 2,3,4,5 -O 3.0 -E 6.0 Apply Stop

Current relation

Relation: UPI apps transaction data in 2021-weka.filters.unsup... Instances: 654 Attributes: 9 Sum of weights: 654

Attributes

AllNoneInvertPattern

No.	Name
1	<input checked="" type="checkbox"/> UPI Banks
2	<input type="checkbox"/> Volume (Mn) By Costumers
3	<input type="checkbox"/> Value (Cr) by Costumers
4	<input type="checkbox"/> Volume (Mn)
5	<input type="checkbox"/> Value (Cr)
6	<input type="checkbox"/> Month
7	<input type="checkbox"/> Year
8	<input type="checkbox"/> Outlier
9	<input type="checkbox"/> ExtremeValue

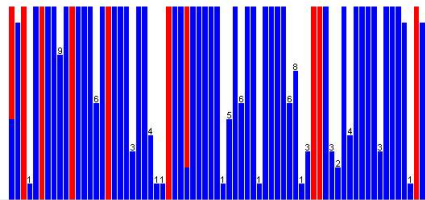
Remove

Selected attribute

Name: UPI Banks Missing: 0 (0%) Distinct: 69 Type: Nominal Unique: 7 (1%)

No.	Label	Count	Weight
1	Airtel Payments Bank...	12	12.0
2	Allahabad Bank App	11	11.0
3	Amazon Pay	12	12.0
4	Andhra Bank App	1	1.0
5	AU Small Finance Ba...	12	12.0
6	Axis Bank Apps	12	12.0
7	Bajaj Finserv	12	12.0
8	Bank of Baroda Apps	12	12.0
9	Bank of India App	9	9.0
10	Bank of Maharashtra	12	12.0

Class: ExtremeValue (Nom) Visualize All



Status

OK Log x 0

Weka Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Open file...Open URL...Open DB...Generate...UndoEdit...Save...

Filter: Choose InterquartileRange -R 2,3,4,5 -O 3.0 -E 6.0 Apply Stop

Current relation

Relation: UPI apps transaction data in 2021-weka.filters.unsup... Instances: 654 Attributes: 9 Sum of weights: 654

Attributes

AllNoneInvertPattern

No.	Name
1	<input type="checkbox"/> UPI Banks
2	<input type="checkbox"/> Volume (Mn) By Costumers
3	<input type="checkbox"/> Value (Cr) by Costumers
4	<input type="checkbox"/> Volume (Mn)
5	<input type="checkbox"/> Value (Cr)
6	<input type="checkbox"/> Month
7	<input type="checkbox"/> Year
8	<input checked="" type="checkbox"/> Outlier
9	<input type="checkbox"/> ExtremeValue

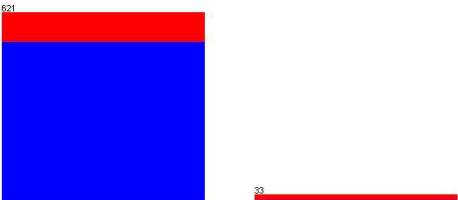
Remove

Selected attribute

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

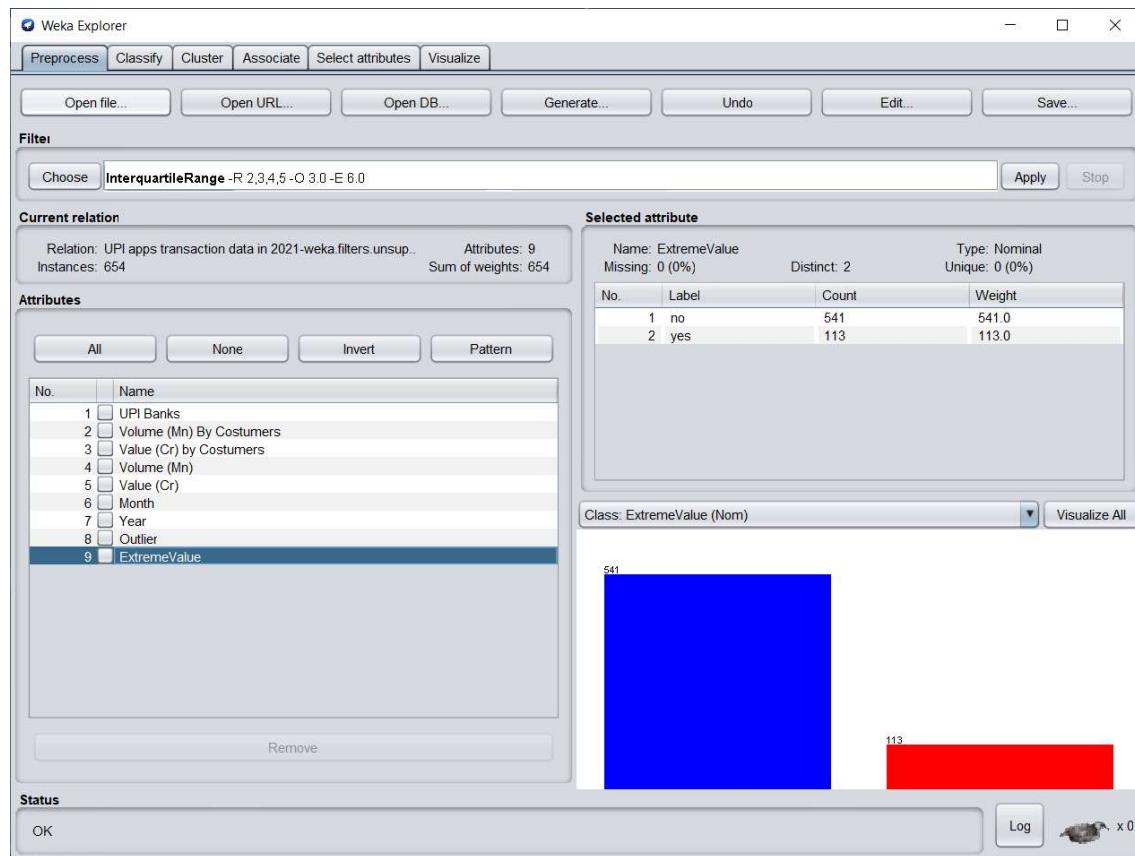
No.	Label	Count	Weight
1	no	621	621.0
2	yes	33	33.0

Class: ExtremeValue (Nom) Visualize All

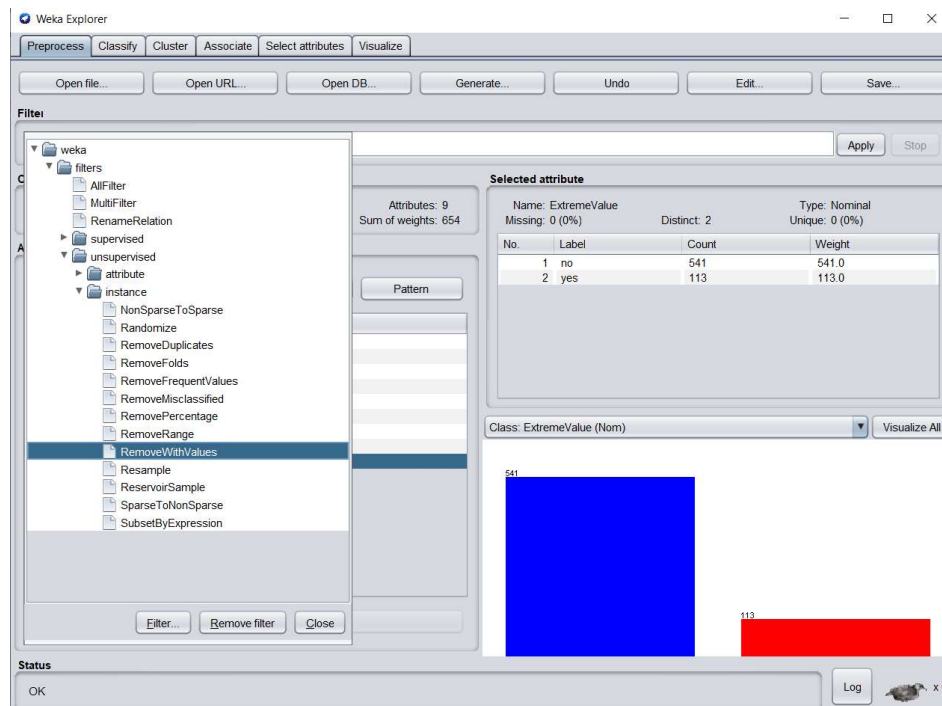


Status

OK Log x 0



Removing Outliers:



 weka.gui.GenericObjectEditor 

weka.filters.unsupervised.instance.RemoveWithValues

About

Filters instances according to the value of an attribute.

More


Capabilities

attributeIndex

8

debug

False




doNotCheckCapabilities

False




dontFilterAfterFirstBatch

False



invertSelection

False




matchMissingValues

False



modifyHeader

False



nominalIndices

last

splitPoint

0.0

Open...

Save...

OK

Cancel

Weka Explorer

PreprocessClusterAssociateSelect attributesVisualize

Open file...Open URL...Open DB...Generate...UndoEdit...Save...

FilterChooseRemoveWithValues -S 0.0 -C 8 -L lastApplyStop

Current relation

Relation: UPI apps transaction data in 2021-weka.filters.unsup...
Instances: 621Attributes: 9
Sum of weights: 621

Attributes

AllNoneInvertPattern

No.	Name
1	<input type="checkbox"/> UPI Banks
2	<input type="checkbox"/> Volume (Mn) By Costumers
3	<input type="checkbox"/> Value (Cr) by Costumers
4	<input type="checkbox"/> Volume (Mn)
5	<input type="checkbox"/> Value (Cr)
6	<input type="checkbox"/> Month
7	<input type="checkbox"/> Year
8	<input checked="" type="checkbox"/> Outlier
9	<input type="checkbox"/> ExtremeValue

Remove

Selected attribute

Name: Outlier
Missing: 0 (0%)

Distinct: 1

Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	no	621	621.0
2	yes	0	0.0

Class: ExtremeValue (Nom)


Visualize All

621

0

Status

OK

Log x 0

Removing Extreme values:

weka.gui.GenericObjectEditor ✕

weka.filters.unsupervised.instance.RemoveWithValues

About

Filters instances according to the value of an attribute. More
Capabilities

attributeIndex

debug ▼

doNotCheckCapabilities ▼

dontFilterAfterFirstBatch ▼

invertSelection ▼

matchMissingValues ▼

modifyHeader ▼

nominalIndices

splitPoint

Open... Save... OK Cancel

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 last -L last
Apply
Stop

Current relation
Relation: UPI apps transaction data in 2021-weka.filters.unsup...
Instances: 526
Attributes: 9
Sum of weights: 526

Attributes
All
None
Invert
Pattern

No.	Name
1	<input type="checkbox"/> UPI Banks
2	<input type="checkbox"/> Volume (Mn) By Costumers
3	<input type="checkbox"/> Value (Cr) by Costumers
4	<input type="checkbox"/> Volume (Mn)
5	<input type="checkbox"/> Value (Cr)
6	<input type="checkbox"/> Month
7	<input type="checkbox"/> Year
8	<input type="checkbox"/> Outlier
9	<input type="checkbox"/> ExtremeValue

Remove

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

No.	Label	Count	Weight
1	no	526	526.0
2	yes	0	0.0

Class: ExtremeValue (Nom)
Visualize All

Status
OK
Log
x 0

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 last -L last
Apply
Stop

Current relation
Relation: UPI apps transaction data in 2021-weka.filters.unsup...
Instances: 526
Attributes: 9
Sum of weights: 526

Attributes
All
None
Invert
Pattern

No.	Name
1	<input type="checkbox"/> UPI Banks
2	<input type="checkbox"/> Volume (Mn) By Costumers
3	<input type="checkbox"/> Value (Cr) by Costumers
4	<input type="checkbox"/> Volume (Mn)
5	<input type="checkbox"/> Value (Cr)
6	<input type="checkbox"/> Month
7	<input type="checkbox"/> Year
8	<input checked="" type="checkbox"/> Outlier
9	<input checked="" type="checkbox"/> ExtremeValue

Remove

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

No.	Label	Count	Weight
1	no	526	526.0
2	yes	0	0.0

Class: ExtremeValue (Nom)
Visualize All

Status
OK
Log
x 0

Final Dataset:

There were 621 instances at first. After removing outliers and extreme values there are 526 instances now.

The screenshot shows the Weka Explorer application window. The 'Preprocess' tab is active. The 'Filter' section shows 'RemoveWithValues -S 0.0 -C last -L last' applied. The 'Current relation' is 'UPI apps transaction data in 2021-weka.filters.unsup.' with 526 instances and 7 attributes. The 'Attributes' list on the left includes 'UPI Banks' (selected), 'Volume (Mn) By Costumers', 'Value (Cr) by Costumers', 'Volume (Mn)', 'Value (Cr)', 'Month', and 'Year'. The 'Selected attribute' section shows 'UPI Banks' with 59 distinct values and 7 unique values. A table below lists the top 10 banks by count. The 'Visualize All' button is visible, and a bar chart is displayed at the bottom right.

No.	Label	Count	Weight
1	Airtel Payments Bank...	4	4.0
2	Allahabad Bank App	11	11.0
3	Amazon Pay	0	0.0
4	Andhra Bank App	1	1.0
5	AU Small Finance Ba...	12	12.0
6	Axis Bank Apps	0	0.0
7	Bajaj Finserv	12	12.0
8	Bank of Baroda Apps	12	12.0
9	Bank of India App	9	9.0
10	Bank of Maharashtra	12	12.0

Class: UPI Banks (Nom)

Visualize All

Status: OK