

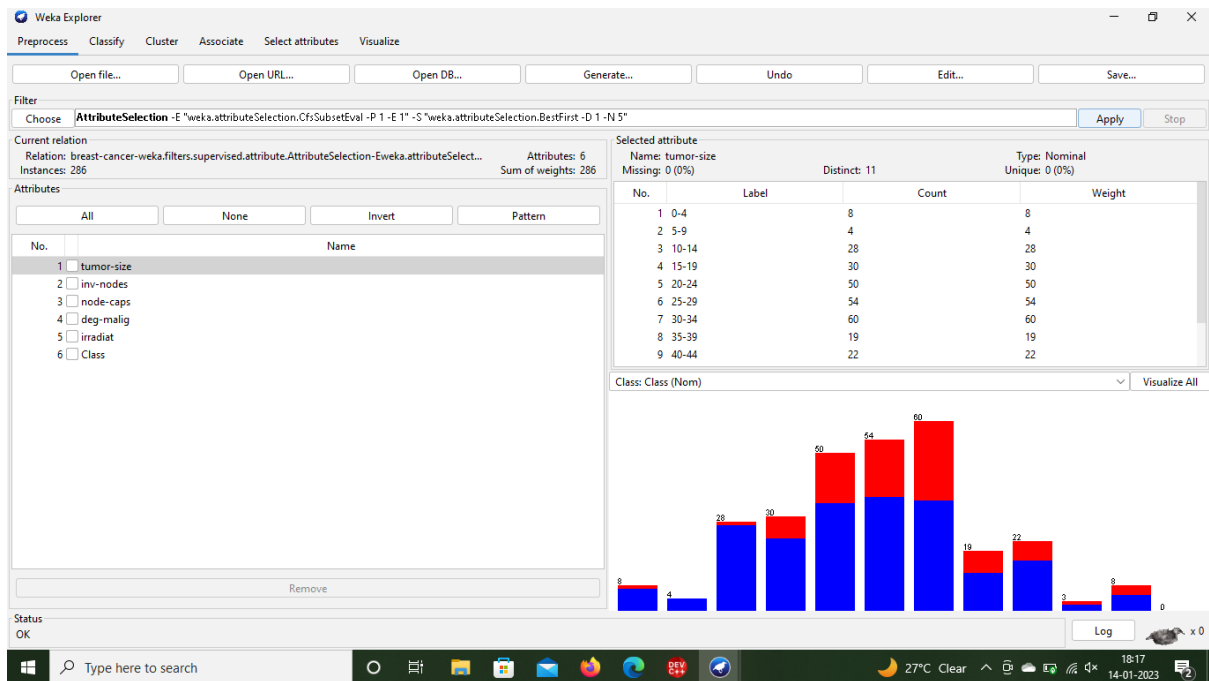
CSA1668-DATAWAREHOUSING AND DATA MINING

P.PRANAVI

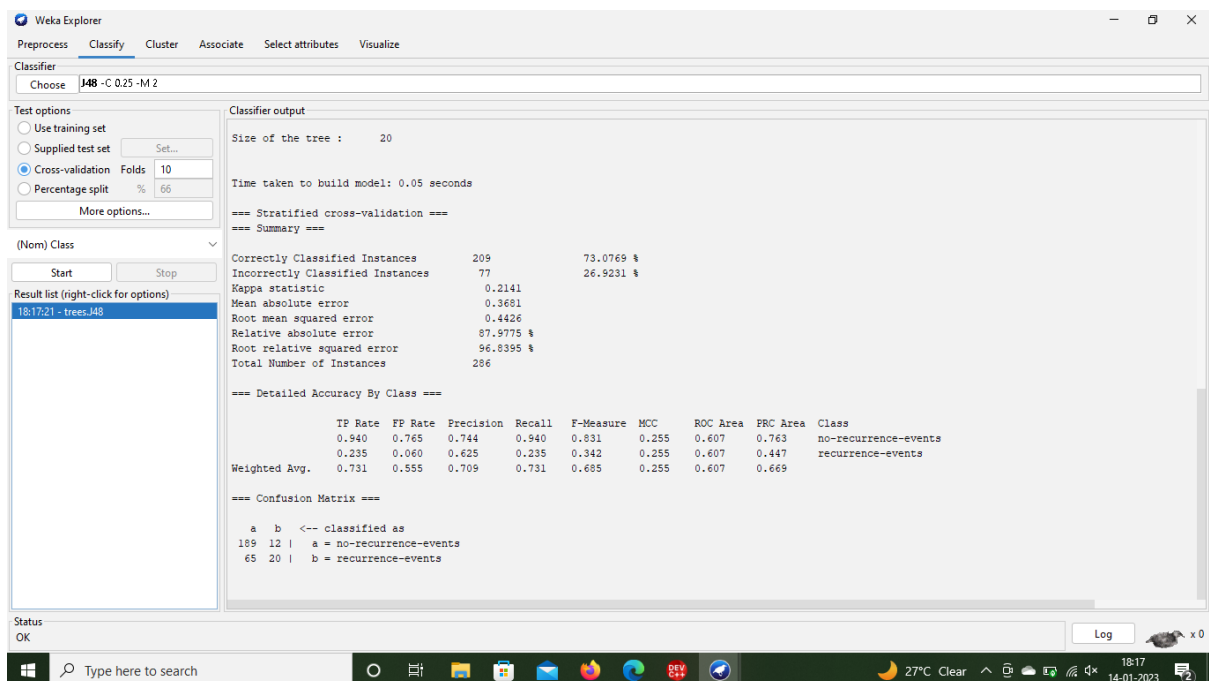
192110530

WEKA TOOL FOR BREAST CANCER

(a)DATA PREPROCESSING



J48 CLASSIFIER:



REPLACE THE MISSING VALUES:

Weka Explorer

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

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

Filter: Choose **ReplaceMissingValues** Apply Stop

Current relation: Relation: breast-cancer-weka.filters.supervised.attribute.AttributeSelection-Eweka.attributeSelect...
Instances: 286 Attributes: 6 Sum of weights: 286

Attributes: All None Invert Pattern

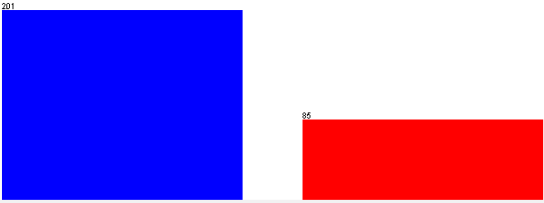
No.	Name
1	tumor-size
2	inv-nodes
3	node-caps
4	deg-malig
5	irradiat
6	Class

Remove

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

No.	Label	Count	Weight
1	no-recurrence-events	201	201
2	recurrence-events	85	85

Class: Class (Nom) Visualize All



Status: OK Log x 0

Type here to search 27°C Clear 18:18 14-01-2023

INTERQUARTILE RANGE

Weka Explorer

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

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

Filter: Choose **InterquartileRange -R 2 -O 3.0 -E 6.0** Apply Stop

Current relation: Relation: breast-cancer-weka.filters.supervised.attribute.AttributeSelection-Eweka.attributeSelect...
Instances: 286 Attributes: 8 Sum of weights: 286

Attributes: All None Invert Pattern

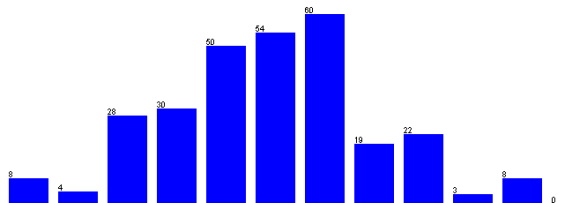
No.	Name
1	tumor-size
2	inv-nodes
3	node-caps
4	deg-malig
5	irradiat
6	Class
7	Outlier
8	ExtremeValue

Remove

Selected attribute: Name: tumor-size
Missing: 0 (0%) Distinct: 11 Type: Nominal Unique: 0 (0%)

No.	Label	Count	Weight
1	0-4	8	8
2	5-9	4	4
3	10-14	28	28
4	15-19	30	30
5	20-24	50	50
6	25-29	54	54
7	30-34	60	60
8	35-39	19	19
9	40-44	22	22

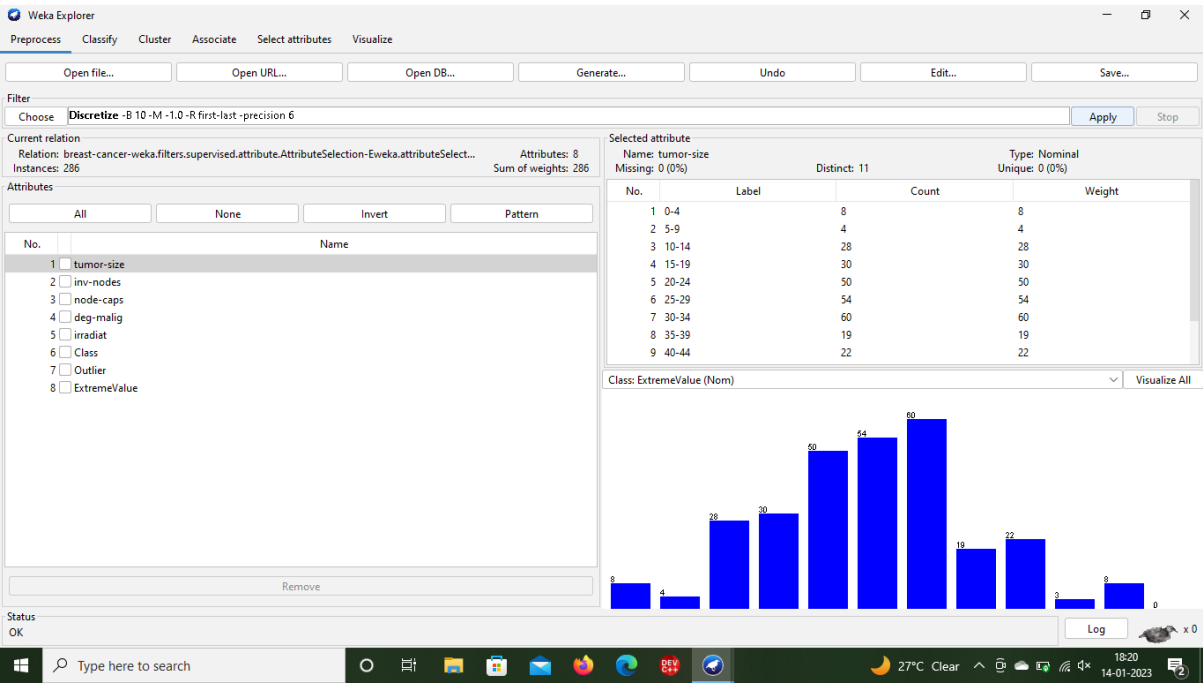
Class: ExtremeValue (Nom) Visualize All



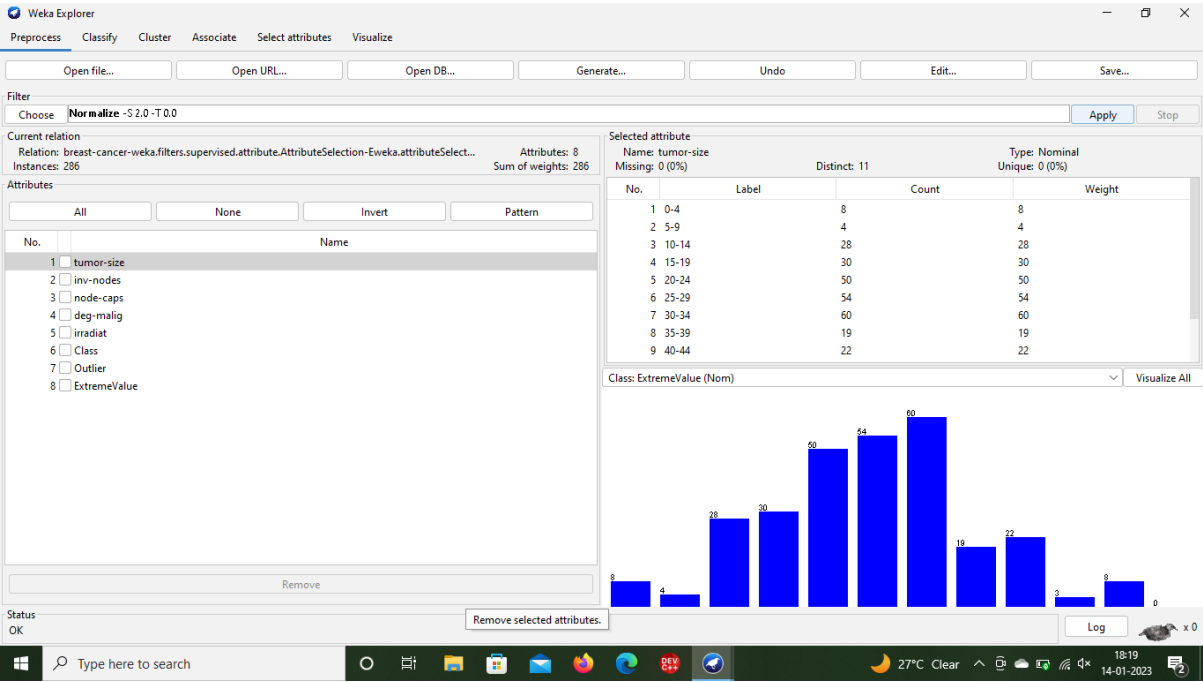
Status: OK Log x 0

Type here to search 27°C Clear 18:18 14-01-2023

DISCRETIZE



NORMALIZE



DATA CLASSIFICATION

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options:
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds: **10**
☐ Percentage split % **66**
More options...

(Nom) ExtremeValue
Start Stop

Result list (right-click for options):
18:17:21 - trees.J48
18:20:14 - trees.J48

Classifier output:

Size of the tree : 1

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Metric	Value	Percentage
Correctly Classified Instances	286	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	
Mean absolute error	0	
Root mean squared error	0	
Relative absolute error	0	%
Root relative squared error	0	%
Total Number of Instances	286	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000 ?	1.000	1.000	1.000	1.000	1.000	?	?	1.000	no
? 0.000 ?	?	?	?	?	?	?	?	?	yes
Weighted Avg.	1.000	?	1.000	1.000	1.000	?	?	1.000	

=== Confusion Matrix ===

a	b	←-- classified as
286	0	a = no
0	0	b = yes

Status: OK

OneDrive Screenshot saved
The screenshot was added to your OneDrive.

TREE CLASSIFIER

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options:
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds: **10**
☐ Percentage split % **66**
More options...

(Nom) ExtremeValue
Start Stop

Result list (right-click for options):
18:17:21 - trees.J48
18:20:14 - trees.J48

Weka Classifier Tree Visualizer: 18:17:21 - trees.J48 (breast-cancer-weka.filters.supervised.attribute.AttributeSelection-Eweka.attributeSelect...

Tree View

Inv-nodes

0-2 3-5 6-8 9-11 12-14 15-17 18-20 21-23 24-26 27-29 30-32 33-35

no-recurrence-events deg-malign ? recurrence-recurrence-events (0.0)

= 1 = 2 = 3 = 1 = 2 = 3

no-r no-recur recurrence-rec no-recu recurrence-events (10.0/1.0)

Status: OK

Log

KMEANS CLUSTERING:

The screenshot shows the Weka Explorer interface with the 'Cluster' tab selected. The 'SimpleKMeans' algorithm is chosen, and the 'Cluster mode' is set to 'Use training set'. The 'Store clusters for visualization' option is checked. The 'Clusterer output' pane displays the following information:

Cluster 0: 10-14,0-2,no,2,no,no-recurrence-events,no,no
Cluster 1: 15-19,0-2,yes,3,no,recurrence-events,no,no

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data (286.0)	Cluster# 0 (230.0)	1 (56.0)
tumor-size	30-34	20-24	30-34
inv-nodes	0-2	0-2	0-2
node-caps	no	no	yes
deg-malig	2	2	3
irradiat	no	no	no
Class	no-recurrence-events	no-recurrence-events	recurrence-events
Outlier	no	no	no
ExtremeValue	no	no	no

Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

Cluster	Count	Percentage
0	230	(80%)
1	56	(20%)

EM:

The screenshot shows the Weka Explorer interface with the 'Cluster' tab selected. The 'EM' algorithm is chosen, and the 'Cluster mode' is set to 'Use training set'. The 'Store clusters for visualization' option is checked. The 'Clusterer output' pane displays the following information:

Cluster 0: 10-14,0-2,no,2,no,no-recurrence-events,no,no
Cluster 1: 15-19,0-2,yes,3,no,recurrence-events,no,no

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data (286.0)	Cluster# 0 (230.0)	1 (56.0)
tumor-size	30-34	20-24	30-34
inv-nodes	0-2	0-2	0-2
node-caps	no	no	yes
deg-malig	2	2	3
irradiat	no	no	no
Class	no-recurrence-events	no-recurrence-events	recurrence-events
Outlier	no	no	no
ExtremeValue	no	no	no

Time taken to build model (full training data) : 0.93 seconds

=== Model and evaluation on training set ===

Clustered Instances

Cluster	Count	Percentage
0	211	(74%)
1	75	(26%)

Log likelihood: -5.35071

ASSOCIATE:

The screenshot shows the Weka Explorer application window. The 'Associate' tab is selected, and the 'Apriori' algorithm is chosen. The 'Associate output' pane displays the results of the Apriori algorithm run on the 'full training set'.

Associate output
=== Associate model (full training set) ===

Apriori
=====

Minimum support: 0.75 (214 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 5

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4
Size of set of large itemsets L(2): 5
Size of set of large itemsets L(3): 2

Best rules found:

1. ExtremeValue=no 286 ==> Outlier=no 286 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. Outlier=no 286 ==> ExtremeValue=no 286 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. node-caps=no 230 ==> Outlier=no 230 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. node-caps=no 230 ==> ExtremeValue=no 230 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. node-caps=no ExtremeValue=no 230 ==> Outlier=no 230 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. node-caps=no Outlier=no 230 ==> ExtremeValue=no 230 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. node-caps=no 230 ==> Outlier=no ExtremeValue=no 230 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. irradiat=no 218 ==> Outlier=no 218 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. irradiat=no 218 ==> ExtremeValue=no 218 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. irradiat=no ExtremeValue=no 218 ==> Outlier=no 218 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

The status bar at the bottom shows 'Status OK' and a 'Log' button. The Windows taskbar at the very bottom displays the date and time as 18:21 on 14-01-2023, along with system icons for weather (27°C Clear) and network connectivity.