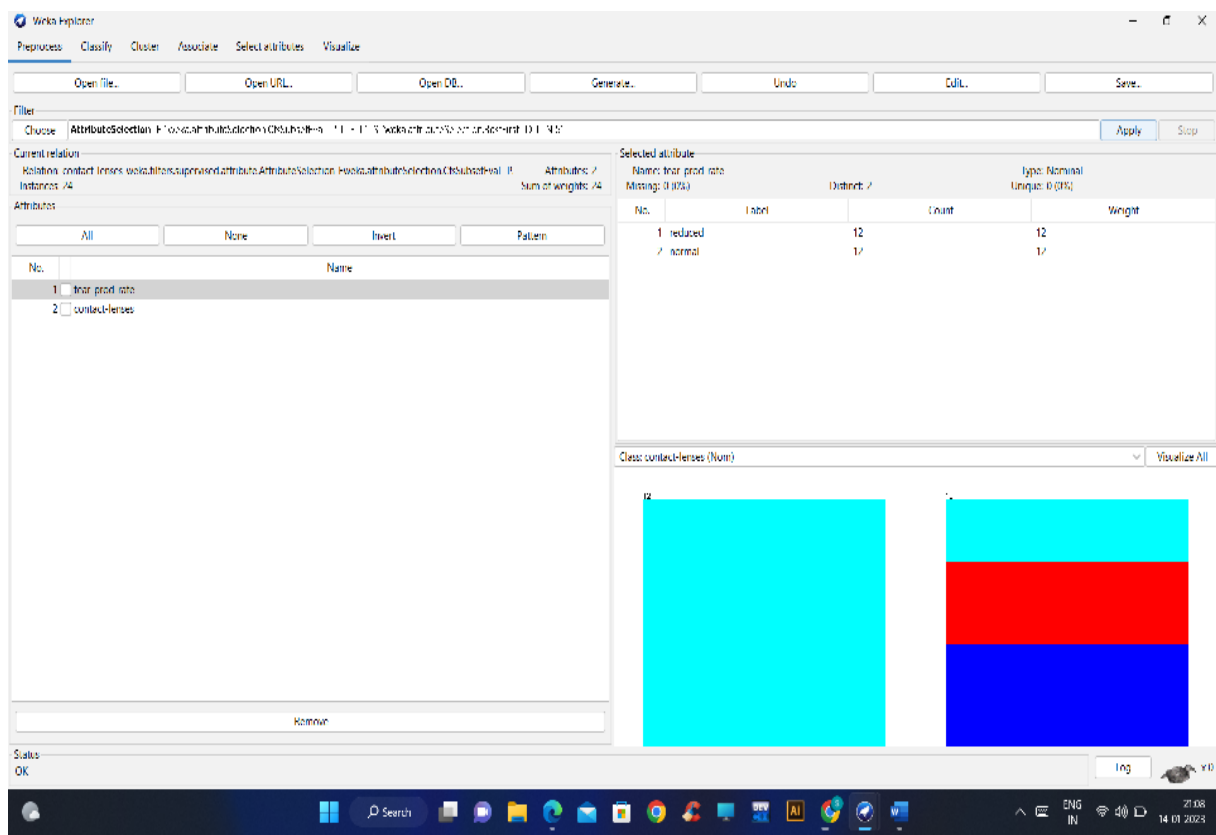


192110696

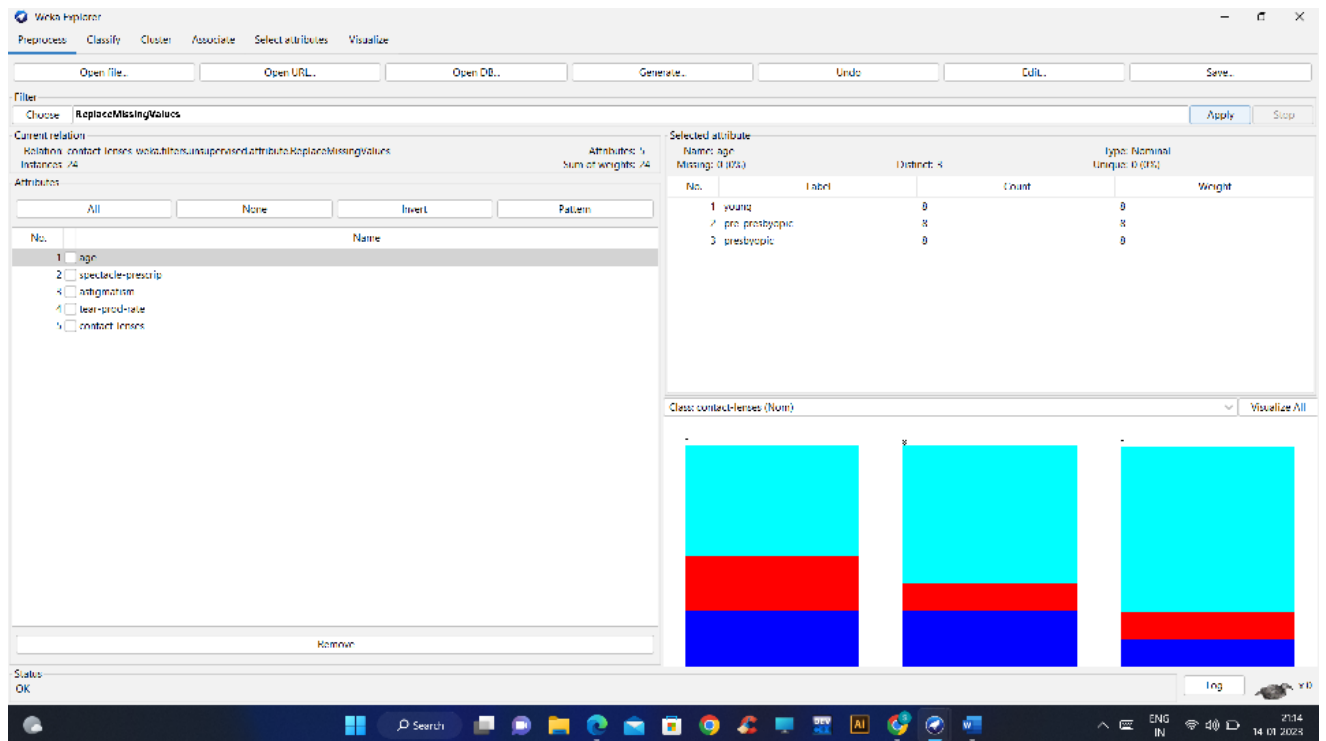
CSA-1668 Datawarehousing and Datamining for pattern analysis

Weka tool Dataset: contact-lenses

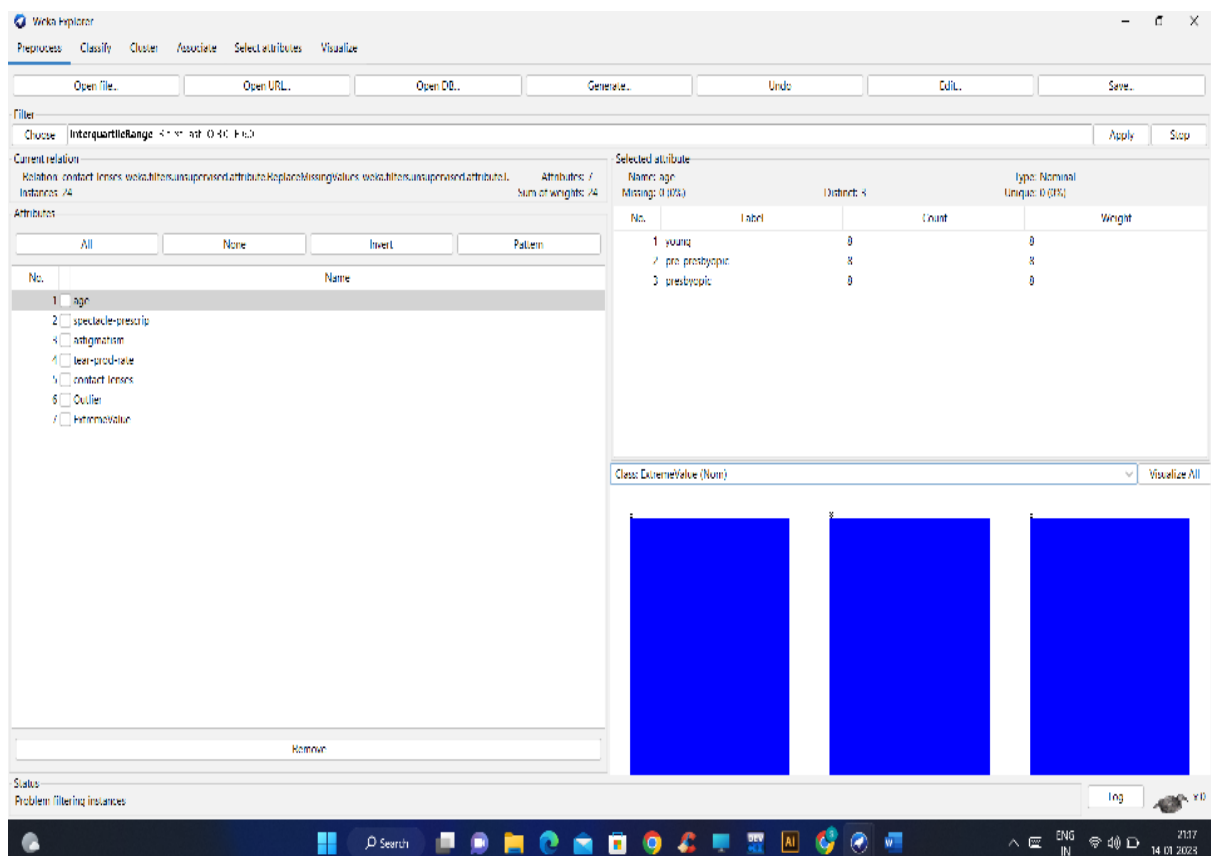
Datapreprocessing:



Remove Missing values:



Interquartile range:



Classification(Bayes):

The screenshot shows the Weka Explorer interface with the Naive Bayes classifier selected. The 'Test options' section is set to 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section displays the results for the 'contact-lenses' dataset. The output is a table with columns for 'Attribute', 'Class', 'no', and 'yes'.

Attribute	Class	no	yes
age	young	8.0	1.0
	spectacle-prescrip	8.0	1.0
	astigmatism	9.0	1.0
	tear-prod-rate	27.0	3.0
spectacle-prescrip	astigmatism	13.0	1.0
	tear-prod-rate	13.0	1.0
	contact-lenses	26.0	2.0
astigmatism	tear-prod-rate	13.0	1.0
	contact-lenses	13.0	1.0
	contact-lenses	26.0	2.0
tear-prod-rate	contact-lenses	13.0	1.0
	contact-lenses	13.0	1.0
	contact-lenses	26.0	2.0
contact-lenses	contact-lenses	13.0	1.0
	contact-lenses	13.0	1.0
	contact-lenses	26.0	2.0
contact-lenses	contact-lenses	13.0	1.0
	contact-lenses	13.0	1.0
	contact-lenses	26.0	2.0

DISCRETIZE:

The screenshot shows the Weka Explorer interface with the Discretize filter applied to the 'age' attribute. The 'Filter' section shows 'Discretize-B10-M-1.0-R-first-last-precision6'. The 'Attributes' section shows the list of attributes: age, spectacle-prescrip, astigmatism, tear-prod-rate, and contact-lenses. The 'Selected attribute' section shows the results for the 'age' attribute, including a table with 'No.', 'Label', 'Count', and 'Weight'.

No.	Label	Count	Weight
1	young	8	8
2	pre-presbyopic	8	8
3	presbyopic	8	8

The 'Class: contact-lenses (Nom)' section shows three stacked bar charts representing the distribution of the 'age' attribute across the three classes: young, pre-presbyopic, and presbyopic. The bars are colored red, blue, and green, representing the three classes.

Normalize:

The screenshot shows the Weka Explorer application with the 'Normalize' filter applied to the 'contact-lenses' dataset. The 'Attributes' list on the left includes 'age', 'spectacle-prescrip', 'astigmatism', 'tear-prod-rate', and 'contact-lenses'. The 'Selected attribute' table on the right shows the distribution of the 'age' attribute, with three distinct values: 'young', 'pre-presbyopic', and 'presbyopic', each with a count of 8. The bottom right section displays three stacked bar charts for the 'contact-lenses' class, showing the distribution of the 'age' attribute across the three classes: 'young', 'pre-presbyopic', and 'presbyopic'.

No.	Label	Count	Weight
1	young	8	8
2	pre-presbyopic	8	8
3	presbyopic	8	8

Decision tree classification:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **NaiveBayes**

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds **10**

☐ Percentage split % **66**

More options...

(Nom) contact-lenses

Start Stop

Result list (right-click for options)

21:18:26 - trees:REPtree

21:18:47 - bayes:NaiveBayes

Classifier output

==== Classifier model (full training set) ====

REPtree

====

: no (16/0) [8/0]

Size of the tree : 1

Time taken to build model: 0 seconds

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

==== Summary ====

Correctly Classified Instances	24	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	24		

==== 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	no
?	0.000	?	?	?	?	?	?	?	yes
Weighted Avg.	1.000	?	1.000	1.000	1.000	?	?	1.000	

==== Confusion Matrix ====

a b <-- classified as

24	0	a = no
0	0	b = yes

Status OK

Log

Clustering:

K-means:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Cluster

Choose **SimpleKMeans** -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode

☒ Use training set

☐ Supplied test set Set...

☐ Percentage split % **66**

☐ Classes to clusters evaluation

(Nom) contact-lenses

☒ Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

21:26:22 - SimpleKMeans

Cluster output

kMeans

====

Number of iterations: 2

Within cluster sum of squared errors: 47.0

Initial starting points (random):

Cluster 0: pre-presbyopic,myope,no,normal,soft

Cluster 1: pre-presbyopic,myope,no,reduced,none

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data	Cluster# 0	Cluster# 1
	(24.0)	(12.0)	(12.0)

====

age	young	young	young
spectacle-prescrip	myope	myope	myope
astigmatism	no	no	no
tear-prod-rate	reduced	normal	reduced
contact-lenses	none	soft	none

Time taken to build model (full training data) : 0 seconds

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

Clustered Instances

0	12 (50%)
1	12 (50%)

Status OK

Log



EM :

The screenshot shows the Weka Explorer interface with the EM (Expectation-Maximization) algorithm selected. The 'Clusterer' tab is active, and the 'EM' model is chosen. The 'Cluster output' pane displays the following data:

Attribute	0	1
age		
young	5.1102	4.8890
pre-presbyopic	5.8346	4.1654
presbyopic	6.6083	3.3917
[total]	17.5532	12.4468
spectacle-prescrip		
myope	7.8833	6.1167
hypermetrope	8.6698	5.3302
[total]	16.5532	11.4468
astigmatism		
no	7.8347	6.1653
yes	8.7185	5.2815
[total]	16.5532	11.4468
tearprod-rate		
reduced	12.7989	1.2011
normal	3.7543	10.2457
[total]	16.5532	11.4468
contact-lenses		
soft	1.2538	5.7462
hard	1.2896	4.7104
none	15.0097	1.9903
[total]	17.5532	12.4468

Time taken to build model (full training data) : 0.12 seconds

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

Clustered Instances

Cluster	Count	Percentage
0	15	63%
1	9	38%

Log likelihood: -3.82823

Association Rule:

Apriori:

The screenshot shows the Weka Explorer interface with the Apriori algorithm selected. The 'Associate' tab is active, and the 'Apriori' model is chosen. The 'Associate output' pane displays the following data:

```

Apriori
=====

Minimum support: 0.2 (5 instances)
Minimum metric (confidence): 0.4
Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11
Size of set of large itemsets L(2): 21
Size of set of large itemsets L(3): 6

Task finish found:

1. tearprod-rate=reduced => contact-lenses=none [2]  Conf: (1) > 1.00 (1.0) [1] Conf: (4.5)
2. spectacle-prescrip=hypermetrope tearprod-rate=reduced => contact-lenses=none [2]  Conf: (1) > 1.00 (1.0) [1] Conf: (2.25)
3. spectacle-prescrip=hypermetrope tearprod-rate=reduced => contact-lenses=none [2]  Conf: (1) > 1.00 (1.0) [1] Conf: (2.25)
4. astigmatism=no tearprod-rate=reduced => contact-lenses=none [2]  Conf: (1) > 1.00 (1.0) [1] Conf: (2.25)
5. astigmatism=yes tearprod-rate=reduced => contact-lenses=none [2]  Conf: (1) > 1.00 (1.0) [1] Conf: (2.25)
6. contact-lenses=none => astigmatism=no [2]  Conf: (1) > 1.00 (1.0) [1] Conf: (2.25)
7. contact-lenses=none => tearprod-rate=reduced [2]  Conf: (1) > 1.00 (1.0) [1] Conf: (2.25)
8. contact-lenses=none => spectacle-prescrip=hypermetrope [2]  Conf: (1) > 1.00 (1.0) [1] Conf: (2.25)
9. astigmatism=no contact-lenses=none => tearprod-rate=reduced [2]  Conf: (1) > 1.00 (1.0) [1] Conf: (2.25)
10. contact-lenses=none => astigmatism=yes [2]  Conf: (1) > 1.00 (1.0) [1] Conf: (2.25)
  
```