



Mahavir Education Trust's
**SHAH & ANCHOR KUTCHHI ENGINEERING
COLLEGE**
Chembur, Mumbai - 400 088
UG Program in Cyber Security

Experiment Number: 6					
Date of Performance:					
Date of Submission:					
Program Execution/ formation/ correction/ ethical practices (07)	Documentation (02)	Timely Submission (03)	Viva Answer to sample questions (03)	Experiment Total (15)	Sign



Mahavir Education Trust's
**SHAH & ANCHOR KUTCHHI ENGINEERING
COLLEGE**
Chembur, Mumbai - 400 088
UG Program in Cyber Security
Experiment 6

Aim: Perform data pre-processing task & demonstrate classification, clustering, association algorithm on datasets using data mining tool WEKA.

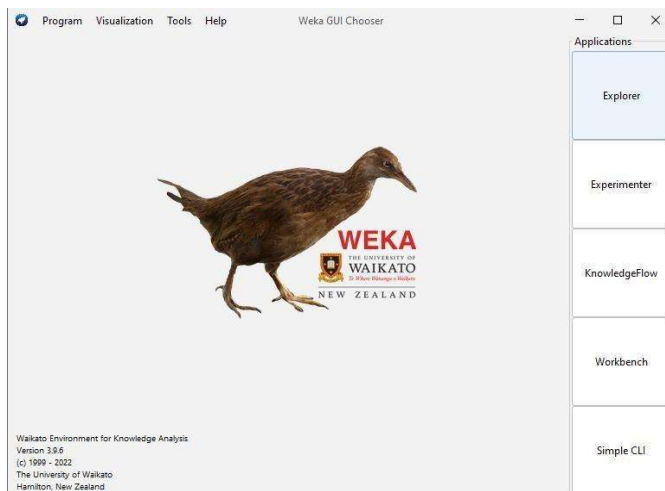
Lab outcomes: CSL 503.2: Implement data mining algorithms like classification. CSL 503.3: Implement clustering algorithms on a given set of data sample.

CSL 503.4: Implement Association rule mining and web mining algorithms.

Problem Statement: To perform data pre-processing task using data mining tool WEKA.

Theory:

WEKA - an open-source software provides tools for data pre-processing, implementation of several Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real-world data mining problems.

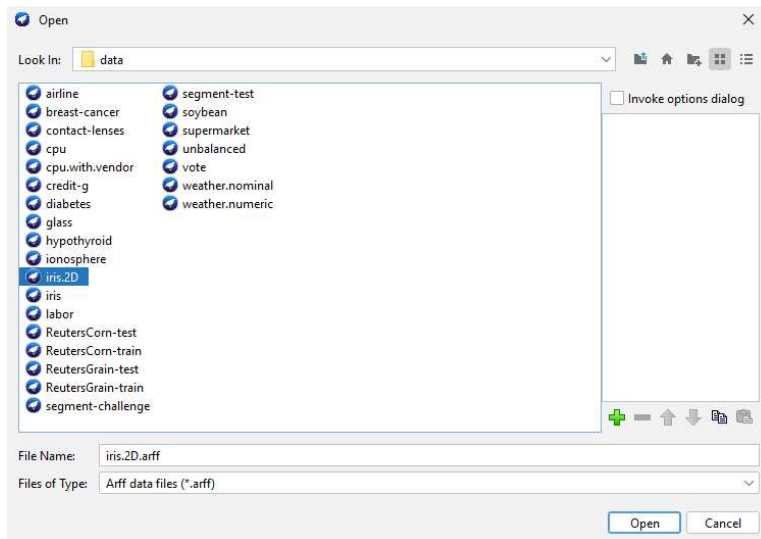
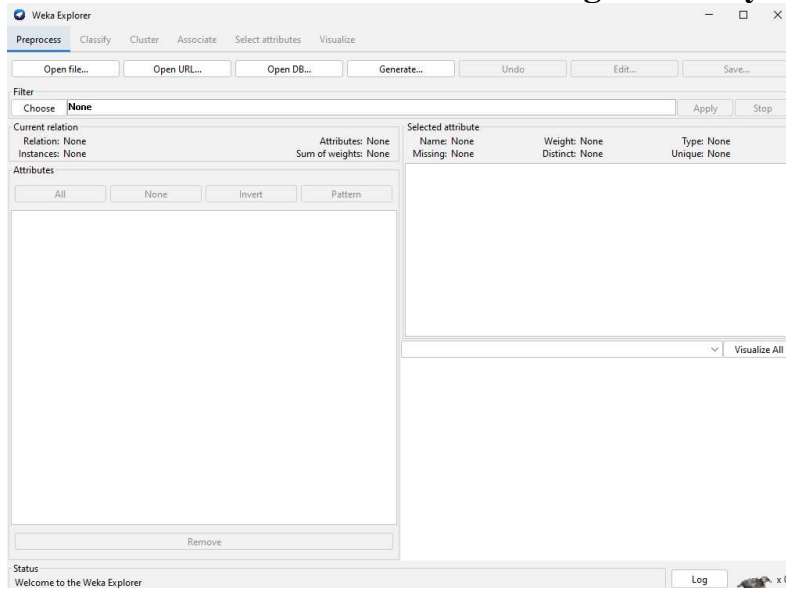


Pre-process Tab

Initially as you open the explorer, only the pre-process tab is enabled. The first step in machine learning is to pre-process the data. Thus, in the Pre-process option, you will select the data file, process it and make it fit for applying the various machine learning algorithms.

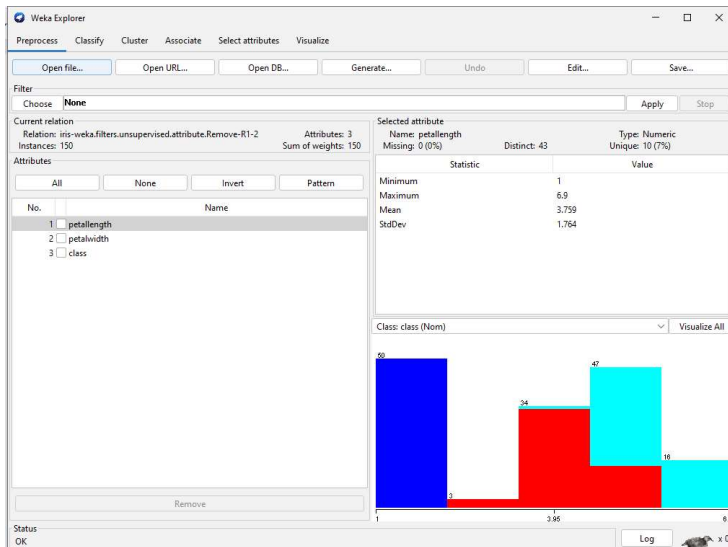


Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING
COLLEGE
Chembur, Mumbai - 400 088
UG Program in Cyber Security



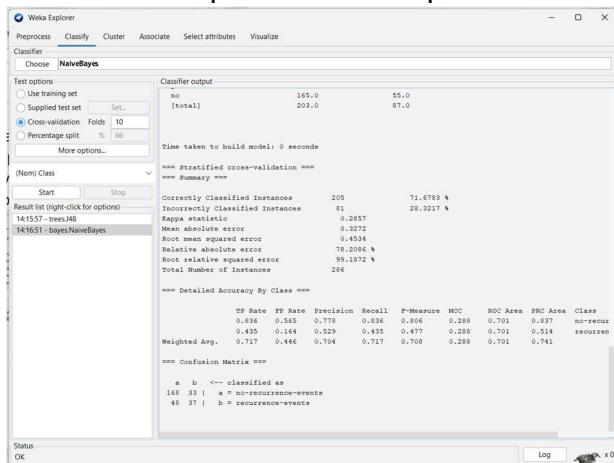


Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING
COLLEGE
Chembur, Mumbai - 400 088
UG Program in Cyber Security



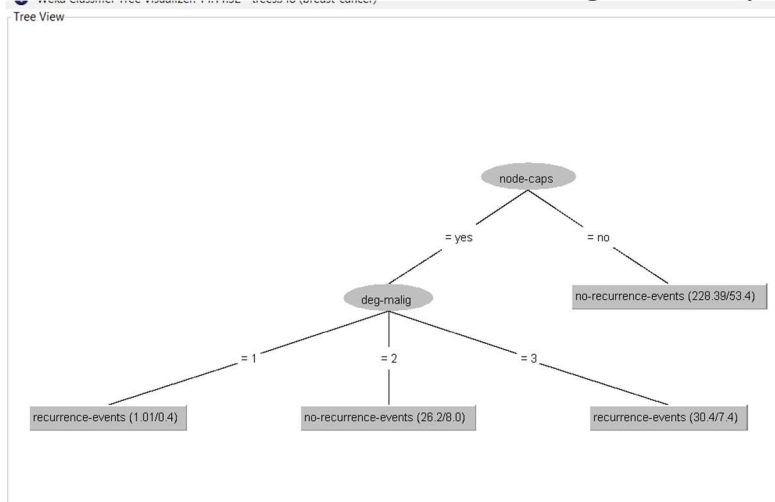
Classify Tab

The Classify tab provides you several machine learning algorithms for the classification of your data. To list a few, you may apply algorithms such as Linear Regression, Logistic Regression, Support Vector Machines, Decision Trees, Random Tree, Random Forest, and so on. The list is very exhaustive and provides both supervised and unsupervised machine learning algorithms.





Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING
COLLEGE
Chembur, Mumbai - 400 088
UG Program in Cyber Security



Cluster Tab

Under the Cluster tab, there are several clustering algorithms provided - such as Simple KMeans, Filtered Clusterer, Hierarchical Clusterer, and so on.

Weka Explorer

Preprocess Classify **Cluster** Associate Select attributes Visualize

Clusterer: Choose **EM** -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100

Cluster mode

- ☒ Use training set
- ☐ Supplied test set
- ☐ Percentage split
- ☐ Classes to clusters evaluation
- ☒ Store clusters for visualization

Ignore attributes: []

Start Stop

Result list (right-click for options)

14:16:01 - EM

Cluster output

	right	breast-quad	irradiat
[total]	48.1997	104.9785	107.9785
left_up	55.182	42.386	14.4847
left_low	33.6182	34.7548	18.5795
right_up	69.6463	39.6158	90.4938
right_low		45.9394	98.7956
central		10.4952	104.9785
[total]		13.9286	117.3751
no-recurrence-events		4.7102	86.87
recurrence-events		15.7882	89.3144
[total]		10.7714	18.1085
Log likelihood: -9.36546		9.9641	28.0608

Time taken to build model (full training data) : 0.63 seconds

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

Clustered Instances

Class	Count	Percentage
0	104	36%
1	117	41%
2	65	23%

Status: OK

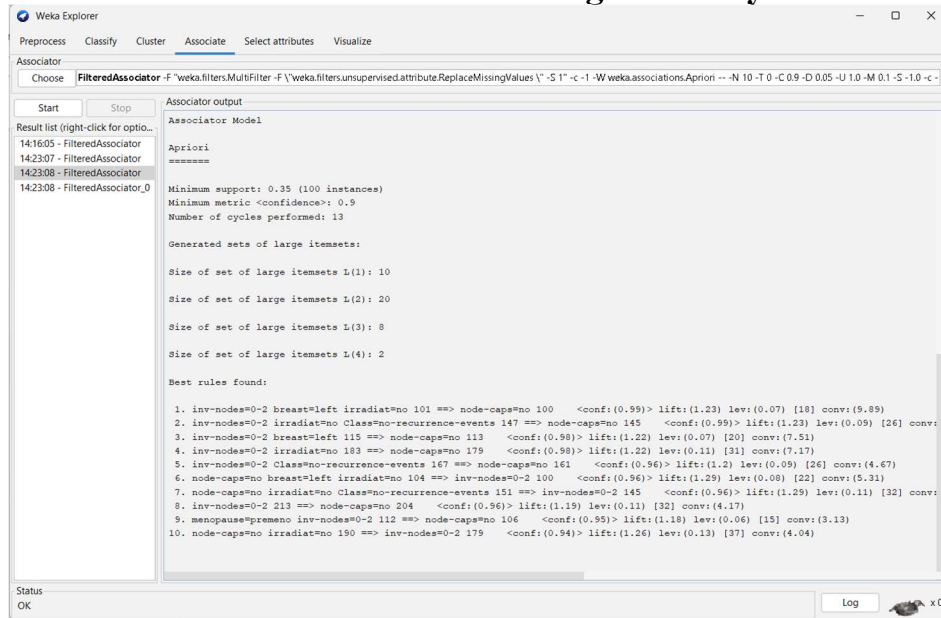
Log

Associate Tab

Under the Associate tab, you would find Apriori, Filtered Associator and FP-Growth.

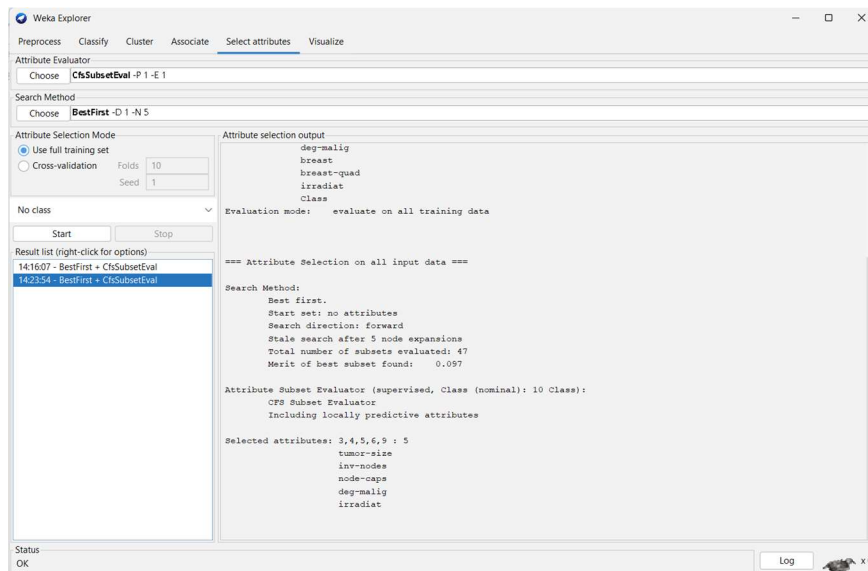


Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING
COLLEGE
Chembur, Mumbai - 400 088
UG Program in Cyber Security



Select Attributes Tab

Select Attributes allows you feature selections based on several algorithms such as Classifier Subset Eval, Principal Components, etc.

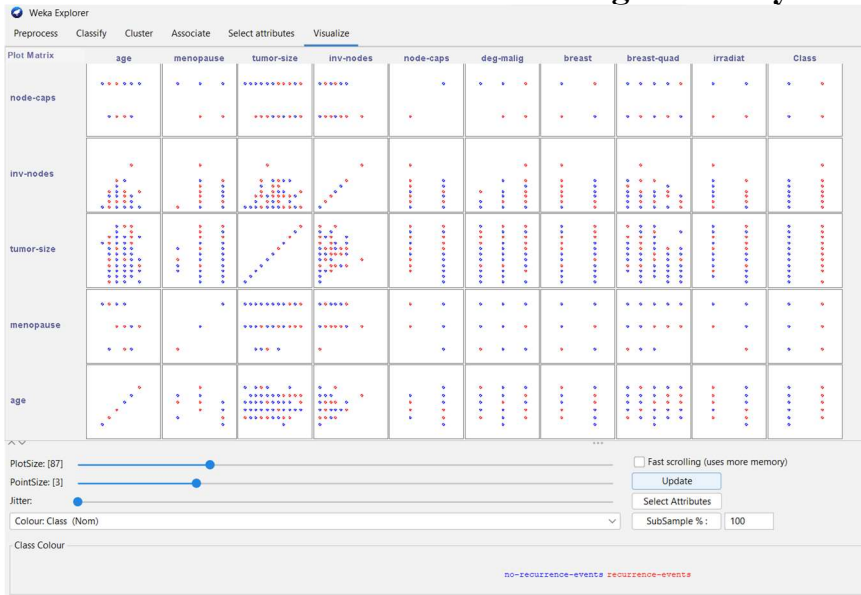


Visualize Tab

Lastly, the Visualize option allows you to visualize your processed data for analysis.



Mahavir Education Trust's
**SHAH & ANCHOR KUTCHHI ENGINEERING
COLLEGE**
Chembur, Mumbai - 400 088
UG Program in Cyber Security



Conclusion: Here we implemented data pre-processing & demonstrate different algorithm on datasets using data mining tool WEKA