

EXPERIMENT 3

Aim: To create an .arff file.

Theory:

Attribute Relation File Format (ARFF), this has two parts:

1. The header section defines the relation (data set) name, attribute name and the type.
2. The data section lists the data instances.

An ARFF file requires the declaration of the relation, attribute and data.

- @relation: This is the first line in any ARFF file, written in the header section, followed by the relation/data set name. The relation name must be a string and if it contains spaces, then it should be enclosed between quotes.
- @attribute: These are declared with their names and the type or range in the header section.

Weka supports the following data types for attributes:

- Numeric
- <nominal-specification>
- String
- date
- @data – Defined in the Data section followed by the list of all data segments

Procedure:

To create an .arff file the steps involved are:

1. Open Notepad or any text editor.
2. Make a relation Employee.
3. Give attributes to it like name, age, gender
4. Give values to the attributes.
5. After everything is done save the file as “filename.arff”, Click OK.
6. Open WEKA and load the .arff file created.
7. Go to classify and click on cross-validation radio button
8. The value of Folds should be equal to your instances & Press Start.
9. Perform same action in all tabs.

Code:

```
@relation Employee
@attribute name{sahra,vivek,sumit,tanya,riya,atul,stark}
@attribute age real
@attribute gender{M,F}
@data
sahra,25,F
vivek,25,M
sumit,30,M
tanya,25,F
riya,26,F
atul,29,M
stark,31,M
```

Classifier Output:

== Run information ==

```

Scheme:      weka.classifiers.rules.ZeroR
Relation:    Employee
Instances:   7
Attributes:  3
              name
              age
              gender
Test mode:   7-fold cross-validation

```

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

ZeroR predicts class value: M

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	4	57.1429 %
Incorrectly Classified Instances	3	42.8571 %
Kappa statistic	0	
Mean absolute error	0.5536	
Root mean squared error	0.557	
Relative absolute error	100	%
Root relative squared error	100	%
Total Number of Instances	7	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	1.000	0.571	1.000	0.727	?	0.000	0.571	M
	0.000	0.000	?	0.000	?	?	0.000	0.429	F
Weighted Avg.	0.571	0.571	?	0.571	?	?	0.000	0.510	

=== Confusion Matrix ===

```

a b  <-- classified as
4 0 | a = M
3 0 | b = F

```

Clustered Output:

== Run information ==

```

Scheme:      weka.clusterers.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
Relation:    Employee
Instances:   7
Attributes:  3
              name
              age
              gender
Test mode:   evaluate on training data

```

=== Clustering model (full training set) ===

EM

==

Number of clusters selected by cross validation: 1

Number of iterations performed: 2

```

Cluster
Attribute 0
           (1)

```

=====

name

```

sahra      2
vivek      2
sumit      2
tanya      2
riya       2
atul       2
stark      2
[total]    14

```

age

```

mean      27.2857
std. dev. 2.4328

```

gender

```

M          5
F          4
[total]    9

```

Time taken to build model (full training data) : 0.01 seconds

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

Clustered Instances

0 7 (100%)

Log likelihood: -4.9373

16/09/2020

Mohd. Aftab Alam

02013302717

Visualization:

