

## EXPERIMENT 4

**Aim:** Implementation of Classification technique on ARFF files using WEKA.

**Theory:** This experiment illustrates the use of j-48 classifier in Weka. The sample data set used in this experiment is “student” data available at .arff format. This document assumes that appropriate data pre-processing has been performed. Steps involved in this experiment:

**Step 1:** We begin the experiment by loading the data (employee1.arff) into Weka.

**Step 2:** Next we select the “classify” tab and click “choose” button to select the “j48” classifier.

**Step 3:** Now we specify the various parameters. These can be specified by clicking in the text box to the right of the choose button. In this example, we accept the default values. The default version does perform some pruning but does not perform error pruning.

**Step 4:** Under the “text” options in the main panel. We select the 10-fold cross validation as our evaluation approach. Since we don’t have separate evaluation data set, this is necessary to get a reasonable idea of accuracy of generated model.

**Step 5:** We now click “start” to generate the model. The ASCII version of the tree as well as evaluation statistic will appear in the right panel when the model construction is complete.

**Step 6:** Note that the classification accuracy of model is about 78.57%. This indicates that we may find more work. (Either in preprocessing or in selecting current parameters for the classification)

**Step 7:** Now Weka also lets us view a graphical version of the classification tree. This can be done by right clicking the last result set and selecting “visualize tree” from the pop-up menu.

**Step 8:** We will use our model to classify the new instances.

**Step 9:** In the main panel under “text” options click the “supplied test set” radio button and then click the “set” button. This will pop-up a window which will allow you to open the file containing test instances.

**Code:**

```

@relation Employee
@attribute salary{<3.5,3.5-6.5,>6.5}
@attribute income{low,mid,high}
@attribute work_ex{y,n}
@attribute gender{M,F}
@attribute pentual{y,n}
@data
<3.5,high,y,M,n
<3.5,high,n,M,n
3.5-6.5,high,n, F, y
>6.5,mid,n,F,y
>6.5,low,y,F,y
>6.5,low,y,M, n
3.5-6.5,low,y,M, y
<3.5,mid,n,F,n
<3.5,low,y,F,n
>6.5,mid,y,F,y|
<3.5,mid,y,M,y
3.5-6.5,mid,n,M,y
3.5-6.5,high,y,F,y
>6.5,mid,n,M,n

```

**Classification Window:**

**Weka Explorer**

Preprocess Classify Cluster Associate

Classifier: ZeroR

Test options:

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds: 10
- ☐ Percentage split %: 66

More options...

(Nom) salary

Start Stop

Result list (right-click for options):

23:11:06 - rules.ZeroR

Status: OK

Run information:

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2  
 Relation: Employee  
 Instances: 14  
 Attributes: salary, income, work\_ex, gender, pentual

Test mode: 10-fold cross-validation  
 Classifier model (full training set) ==  
 J48 pruned tree

```

salary = <3.5: n (5.0/1.0)
salary = 3.5-6.5: y (4.0)
salary = >6.5
|   gender = M: n (2.0)
|   gender = F: y (3.0)
  
```

Number of Leaves : 4  
 Size of the tree : 6  
 Time taken to build model: 0 seconds

Stratified cross-validation Summary:

Metric	Value	Percentage
Correctly Classified Instances	11	78.5714 %
Incorrectly Classified Instances	3	21.4286 %
Kappa statistic	0.5532	
Mean absolute error	0.25	
Root mean squared error	0.4058	
Relative absolute error	49.5283 %	
Root relative squared error	79.6745 %	
Total Number of Instances	14	

Detailed Accuracy By Class:

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.875	0.333	0.778	0.875	0.824	0.559	0.854	0.919	y
	0.667	0.125	0.800	0.667	0.727	0.559	0.854	0.727	n
Weighted Avg.	0.786	0.244	0.787	0.786	0.782	0.559	0.854	0.837	

Confusion Matrix:

```

a b  <-- classified as
7 1 | a = y
2 4 | b = n
  
```

**Classification Tree:**