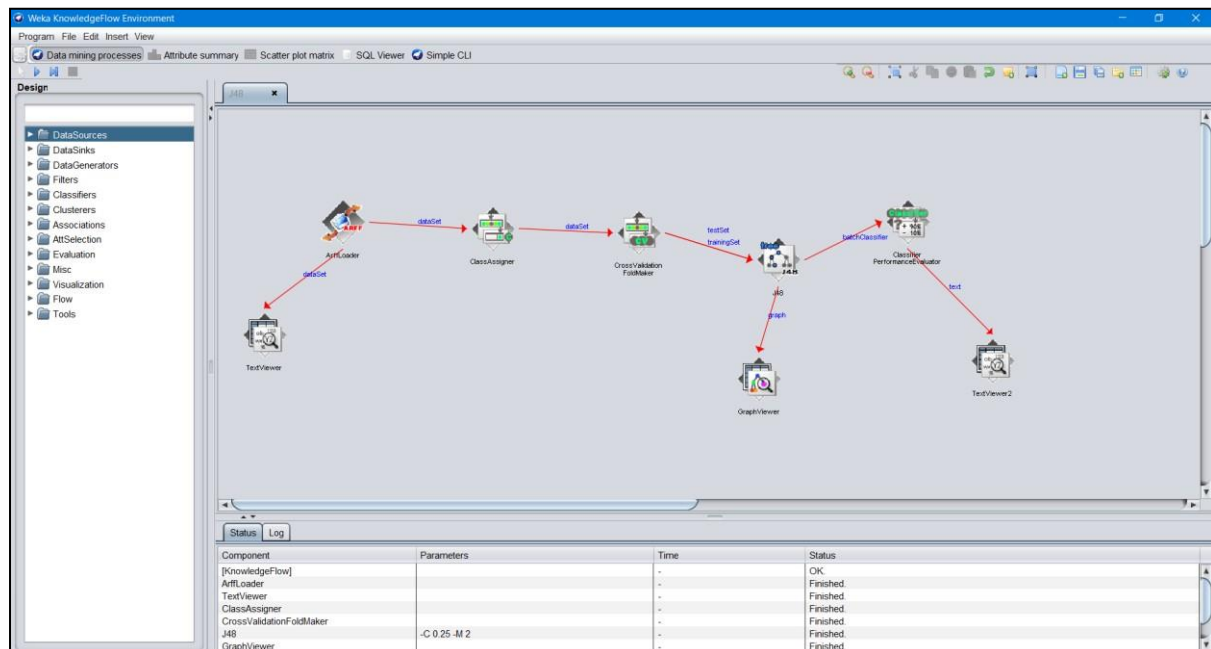


## Experiment-7B: J48

**Experiment Title:** Perform data Pre-processing task and Demonstrate performing Classification, Clustering, Association algorithm on data sets using data mining tool (WEKA,R tool, XL Miner, etc.)

- **Knowledge Flow Environment:**

### J48 Classification Model design:

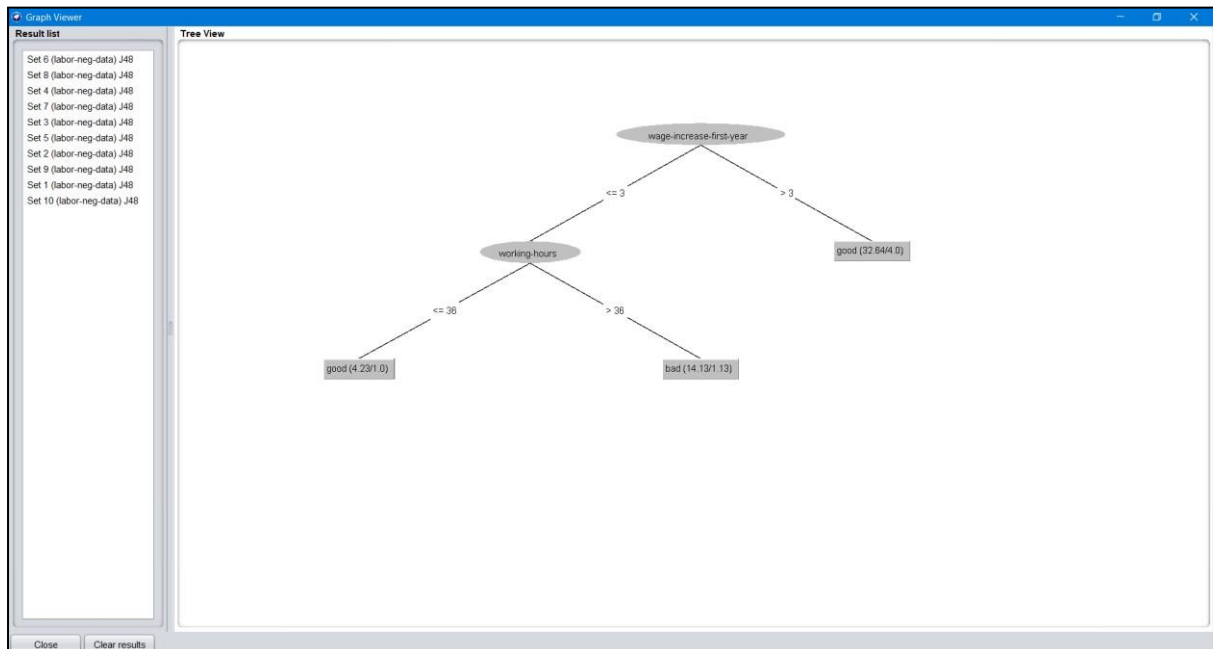


## DataSet Passed to the ArffLoader:

```
#relation labor-neg-data
@attribute duration numeric
@attribute wage-increase-first-year numeric
@attribute wage-increase-second-year numeric
@attribute wage-increase-third-year numeric
@attribute cost-of-living-adjustment {none,tcf,tc}
@attribute working-hours numeric
@attribute pension {none,ret_allw,empl_contr}
@attribute standby-pay numeric
@attribute shift-differential numeric
@attribute education-allowance {yes,no}
@attribute statutory-holidays numeric
@attribute vacation {below_average,average,generous}
@attribute longterm-disability-assistance {yes,no}
@attribute contribution-to-dental-plan {none,half,full}
@attribute bereavement-assistance {yes,no}
@attribute contribution-to-health-plan {none,half,full}
@attribute class {bad,good}

#data
1,5,7,7,7,40,7,2,7,11,average,7,7,yes,7,good
2,4,5,5,8,7,7,35,ret_allw,7,7,yes,11,below_average,7,full,7,good
7,7,7,7,7,38,empl_contr,7,5,7,11,generous,yes,half,yes,half,good
3,3,7,4,5,tc,7,7,7,2,yes,7,7,7,7,good
3,4,5,4,5,7,40,7,7,7,12,average,7,half,yes,half,good
2,2,2,5,7,7,35,7,7,6,yes,12,average,7,7,7,7,good
3,4,5,5,tc,7,empl_contr,7,7,7,12,generous,yes,none,yes,half,good
3,6,9,4,8,2,3,7,40,7,7,7,12,below_average,7,7,7,7,good
2,3,7,7,7,38,7,12,25,yes,11,below_average,yes,half,yes,7,good
1,5,7,7,7,none,40,empl_contr,7,4,7,11,generous,yes,full,7,7,good
3,3,5,4,4,6,none,36,7,7,7,13,generous,7,7,yes,full,good
2,6,4,6,4,7,7,38,7,7,4,7,15,7,7,full,7,7,good
2,3,5,4,7,none,40,7,7,2,no,10,below_average,no,half,7,half,bad
3,3,5,4,5,1,tc,37,7,7,4,7,13,generous,7,full,yes,full,good
1,3,7,7,none,36,7,7,10,no,11,generous,7,7,7,7,good
2,4,5,4,7,none,37,empl_contr,7,7,7,11,average,7,full,yes,7,good
1,2,8,7,7,7,35,7,7,2,7,12,below_average,7,7,7,7,good
1,2,1,7,7,tc,40,ret_allw,2,3,no,9,below_average,yes,half,7,none,bad
1,2,7,7,none,38,none,7,7,yes,11,average,no,none,no,none,bad
2,4,5,7,tc,35,7,13,5,7,15,generous,7,7,7,7,good
2,4,3,4,4,7,7,38,7,7,4,7,12,generous,7,full,7,full,good
3,3,3,3,3,3,40,none,3,3,3,11,below_average,3,3,3,3,bad
```

## Result Generated in GraphViewer:



## Result on TextViewer:

The screenshot shows the Weka Text Viewer window. The left pane displays the 'Result list' with the entry '16:48:34.171 - J48'. The right pane shows the evaluation results for the J48 classifier on the 'labor-neg-data' relation.

```
=== Evaluation result ===  
  
Scheme: J48  
Options: -C 0.25 -M 2  
Relation: labor-neg-data  
  
=== Summary ===  
  
Correctly Classified Instances      42      73.6842 %  
Incorrectly Classified Instances    15      26.3158 %  
Kappa statistic                    0.4415  
Mean absolute error                 0.3192  
Root mean squared error            0.4669  
Relative absolute error            69.7715 %  
Root relative squared error        97.7888 %  
Total Number of Instances         57  
  
=== Detailed Accuracy By Class ===  
  
      TP Rate  PP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class  
-----  
0.700    0.243    0.609    0.700    0.651    0.444    0.695    0.559    bad  
0.757    0.300    0.824    0.757    0.789    0.444    0.695    0.738    good  
Weighted Avg.    0.737    0.280    0.748    0.737    0.740    0.444    0.695    0.675  
  
=== Confusion Matrix ===  
  
a  b  <-- classified as  
14  6 | a = bad  
 9 28 | b = good
```

## • Explorer:

## Passing Dataset:

The screenshot shows the Weka Explorer window. The 'Preprocess' tab is selected. The 'Current relation' is 'iris' with 150 instances and 5 attributes. The 'Attributes' list on the left shows 'sepalwidth', 'sepalwidth', 'petalwidth', 'petalwidth', and 'class' selected. The 'Selected attribute' pane on the right shows the 'class' attribute with 3 distinct values: 'Iris-setosa' (50 instances), 'Iris-versicolor' (50 instances), and 'Iris-virginica' (50 instances). The 'Visualize All' button is visible at the bottom right.

## Results After J48 Classification Execution:

The screenshot shows the Weka Explorer interface with the J48 classifier selected. The 'Test options' section on the left indicates 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section on the right displays the following information:

```
=== Run information ===
Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     iris
Instances:    150
Attributes:   5
  sepallength
  sepalwidth
  petallength
  petalwidth
  class

Test mode:    10-fold cross-validation

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

J48 pruned tree

petalwidth <= 0.6: Iris-setosa (50.0)
petalwidth > 0.6
| petalwidth <= 1.7
| | petallength <= 4.9: Iris-versicolor (46.0/1.0)
| | petallength > 4.9
| | | petalwidth <= 1.5: Iris-virginica (3.0)
| | | petalwidth > 1.5: Iris-versicolor (3.0/1.0)
| petalwidth > 1.7: Iris-virginica (46.0/1.0)

Number of Leaves :    5
Size of the tree :    5

Time taken to build model: 0 seconds
```

The 'Result list' on the left shows a single entry: '10.55.12 - trees.J48'.

The screenshot shows the Weka Explorer interface with the J48 classifier selected. The 'Test options' section on the left indicates 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section on the right displays the following information:

```
Size of the tree :    5

Time taken to build model: 0 seconds

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

Correctly Classified Instances      144          96 %
Incorrectly Classified Instances     6           4 %
Kappa statistic                    0.94
Mean absolute error                 0.035
Root mean squared error             0.1586
Relative absolute error              7.8705 %
Root relative squared error         33.6353 %
Total Number of Instances          150

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
    0.980    0.000    1.000    0.980    0.990    0.985    0.990    0.987    Iris-setosa
    0.940    0.030    0.940    0.940    0.940    0.910    0.952    0.880    Iris-versicolor
    0.960    0.030    0.941    0.960    0.950    0.925    0.961    0.905    Iris-virginica
Weighted Avg.    0.940    0.020    0.960    0.960    0.960    0.940    0.968    0.924

=== Confusion Matrix ===

a b c <-- classified as
45 1 0 | a = Iris-setosa
 0 47 3 | b = Iris-versicolor
 0 2 48 | c = Iris-virginica
```

The 'Result list' on the left shows a single entry: '10.55.12 - trees.J48'.