



Ahsanullah University of Science & Technology

Department of Computer Science & Engineering

Course No : CSE4142

Course Title : Data Warehousing and Mining Lab

Assignment No : 03

Date of Submission : 23.06.2024

Submitted To : Mr. Saha Reno & Mr. Raiyan Jahangir

Submitted By-

Name: Nabila Rahman

ID: 20200204065

Section: B1

(i) Create a Custom Dataset Which Will Have 8 Attributes: 4 Numeric, 3 Nominal & 1 Class (3 Class Values)

=>

```
@relation vehicle

@attribute engine_size numeric
@attribute horsepower numeric
@attribute weight numeric
@attribute fuel_efficiency numeric
@attribute color {red,green,blue}
@attribute body_type {sedan,SUV,truck}
@attribute transmission_type {manual,automatic,CVT}
@attribute class {economy,standard,luxury}
```

(ii) Create 30 Instances of That Dataset.

=>

```
@data
1.6, 120, 2800, 35.5, red, sedan, manual, economy
2.0, 150, 3000, 30.0, green, SUV, automatic, standard
3.5, 250, 4000, 25.0, blue, truck, CVT, luxury
1.8, 140, 2900, 32.5, red, sedan, manual, economy
2.2, 160, 3200, 28.0, green, SUV, automatic, standard
4.0, 300, 4200, 22.0, blue, truck, CVT, luxury
1.5, 110, 2700, 37.0, red, sedan, manual, economy
2.5, 180, 3300, 26.5, green, SUV, automatic, standard
3.8, 280, 4100, 23.5, blue, truck, CVT, luxury
1.7, 130, 2850, 34.0, red, sedan, manual, economy
2.1, 155, 3100, 29.5, green, SUV, automatic, standard
4.2, 320, 4300, 21.0, blue, truck, CVT, luxury
1.9, 145, 2950, 31.5, red, sedan, manual, economy
2.3, 170, 3250, 27.5, green, SUV, automatic, standard
3.6, 260, 4050, 24.5, blue, truck, CVT, luxury
1.4, 100, 2600, 38.5, red, sedan, manual, economy
2.4, 175, 3350, 26.0, green, SUV, automatic, standard
4.1, 310, 4250, 22.5, blue, truck, CVT, luxury
1.8, 135, 2750, 33.0, red, sedan, manual, economy
2.6, 190, 3400, 25.0, green, SUV, automatic, standard
3.7, 270, 4150, 23.0, blue, truck, CVT, luxury
1.7, 125, 2800, 36.0, red, sedan, manual, economy
2.2, 165, 3150, 28.5, green, SUV, automatic, standard
4.3, 330, 4350, 20.5, blue, truck, CVT, luxury
1.6, 120, 2900, 35.0, red, sedan, manual, economy
2.0, 150, 3100, 30.5, green, SUV, automatic, standard
3.9, 290, 4200, 24.0, blue, truck, CVT, luxury
1.5, 110, 2750, 37.5, red, sedan, manual, economy
2.5, 185, 3300, 27.0, green, SUV, automatic, standard
4.0, 300, 4400, 21.5, blue, truck, CVT, luxury
```

(iii) Construct a Classification Model using J48 Decision Tree Algorithm, Use 10-Fold Cross Validation.

=>

Classifier

Choose **J48 -C 0.25 -M 2**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) class

Result list (right-click for options)

03:37:47 - trees.J48

Classifier output

```
1 - body_type = truck: luxury (3.0)
fuel_efficiency > 30: economy (11.0/1.0)

Number of Leaves :      6

Size of the tree :      9

Time taken to build model: 0 seconds

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

Correctly Classified Instances      17           56.6667 %
Incorrectly Classified Instances    13           43.3333 %
Kappa statistic                    0.3134
Mean absolute error                 0.3336
Root mean squared error            0.5185
Relative absolute error             76.0801 %
Root relative squared error        110.3744 %
Total Number of Instances          30

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.769    0.412    0.588     0.769    0.667      0.357    0.661    0.590    economy
                0.250    0.182    0.333     0.250    0.286      0.075    0.565    0.305    standard
                0.556    0.095    0.714     0.556    0.625      0.499    0.619    0.472    luxury
Weighted Avg.   0.567    0.255    0.558     0.567    0.553      0.325    0.623    0.479

=== Confusion Matrix ===

  a  b  c  <-- classified as
10  2  1 | a = economy
 5  2  1 | b = standard
 2  2  5 | c = luxury
```

(iv) Using Filter Method, Find Out Those Attributes for Which the J48 Model Performs the Best (Least Number of "Inaccurately Classified Instances"). All Other Attributes Should Be Removed.

=>

Attribute Evaluator

Choose **ClassifierSubsetEval** -B weka.classifiers.trees.J48 -T -H "Click to set hold out or test instances" -E DEFAULT -- -C 0.25 -M 2

Search Method

Choose **BestFirst** -D 1 -N 5

Attribute Selection Mode

☒ Use full training set

☐ Cross-validation Folds 10 Seed 1

No class

Start Stop

Result list (right-click for options)

03:39:00 - BestFirst + ClassifierSubsetEval

Attribute selection output

Relation: vehicle
Instances: 30
Attributes: 8
engine_size
horsepower
weight
fuel_efficiency
color
body_type
transmission_type
class

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 32
Merit of best subset found: 0.867

Attribute Subset Evaluator (supervised, Class (nominal): 8 class):
Classifier Subset Evaluator
Learning scheme: weka.classifiers.trees.J48
Scheme options: -C 0.25 -M 2
Hold out/test set: Training data
Subset evaluation: classification error

Selected attributes: 1,3 : 2
engine_size
weight

Choose **Remove** -R 2,4-7 Apply Stop

Current relation
Relation: vehicle-weka.filters.unsupervised.attribute.Remove-R2,4-7
Instances: 30

Attributes: 3 Sum of weights: 30

Attributes

All None Invert Pattern

No.	Name
1	engine_size
2	weight
3	class

Selected attribute

Name: class
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	economy	13	13
2	standard	8	8
3	luxury	9	9

Class: class (Nom) Visualize All

Remove

(v) Also, Find Out The Percentage of Incorrect Classification for Each Folds using Weka Experiment Environment.

=>

The screenshot shows the 'Weka Experiment Environment' window with the 'Setup' tab selected. The 'Experiment Configuration Mode' is set to 'Simple'. There are buttons for 'Open...', 'Save...', and 'New'. The 'Results Destination' section has a dropdown for 'ARFF file' and a 'Filename:' field with a 'Browse...' button. The 'Experiment Type' section has a dropdown for 'Cross-validation', a 'Number of folds:' field set to '10', and radio buttons for 'Classification' (selected) and 'Regression'. The 'Iteration Control' section has a 'Number of repetitions:' field set to '10' and radio buttons for 'Data sets first' and 'Algorithms first' (selected). The 'Datasets' section has buttons for 'Add new...', 'Edit selected...', and 'Delete selected', a checkbox for 'Use relative paths', and a text field containing the path 'C:\Users\ASUS\Desktop\CSE 4.1\CSE4142 (DWM lab)\ass3\original_training(removed).arff'. The 'Algorithms' section has buttons for 'Add new...', 'Edit selected...', and 'Delete selected', and a list box containing 'J48 -C 0.25 -M 2'.

The screenshot shows the 'Weka Experiment Environment' window with the 'Run' tab selected. There is a 'Start' button. Below it is a 'Log' section with a text area containing the following text: '18:06:26: Started', '18:06:26: Finished', and '18:06:26: There were 0 errors'.

Setup

Run

Analyse

Source

Got 100 results

Actions

Perform test

Save output

Open Explorer...

Configure test

Test output

Testing with

Paired T-Tester (corrected)

Select rows and cols

Rows

Cols

Swap

Comparison field

Percent_correct

Significance

0.05

Sorting (asc.) by

<default>

Test base

Select

Displayed Columns

Select

Show std. deviations

☐

Output Format

Select

Test output

Available resultsets

(1) 33.333333

(2) 0

(3) 66.666667

(4) 100

Result list

18:08:09 - Available resultsets

18:08:52 - Available resultsets

18:09:09 - Available resultsets