

Ahsanullah University of Science and Technology



Department of Computer Science and Engineering

Program: Bachelor of Science in Computer Science and Engineering

Course No: CSE 4142

Course Title: Data Warehousing and Mining Lab

Assignment No: 02

Date of Submission: 13.02.2025

Submitted to:

Mr. Saha Reno

Assistant Professor

Department of CSE, AUST

Submitted by:

Name: Afia Fahmida

Student ID: 20210104032

Group: 02

- (i) I have created a customized dataset with 4 Numeric, 3 Nominal and 1 Class value attributes of Food management deciding if the food is healthy, old or rotten.

```
20210104032_TrainingDataset_Orig x +
File Edit View

@relation Food_Management

@attribute Food_container {plastic_box, plate}
@attribute Intake_time {morning, evening, night}
@attribute Food_taste {spicy, salty, sour}
@attribute rice_amount numeric
@attribute beef_amount numeric
@attribute dal_amount numeric
@attribute Vegetable_amount numeric
@attribute Status {healthy, rotten, old}
```

- (ii) I have taken 30 instances.

```
20210104032_TrainingDataset_Orig x +
File Edit View

@attribute Food_container {plastic_box, plate}
@attribute Intake_time {morning, evening, night}
@attribute Food_taste {spicy, salty, sour}
@attribute rice_amount numeric
@attribute beef_amount numeric
@attribute dal_amount numeric
@attribute Vegetable_amount numeric
@attribute Status {healthy, rotten, old}

@data
plastic_box, morning, spicy, 100.5, 50.2, 30.8, 40.1, healthy
plastic_box, night, sour, 200.2, 40.1, 50.6, 20.3, old
plate, morning, spicy, 120.4, 60.8, 25.9, 35.2, healthy
plate, night, sour, 260.7, 25.3, 65.9, 15.2, rotten
plate, evening, salty, 170.5, 100.8, 10.7, 20.2, old
plastic_box, evening, salty, 180.6, 90.5, 15.3, 25.7, old
plate, night, sour, 250.9, 30.4, 60.1, 10.5, rotten
plastic_box, morning, spicy, 110.3, 70.2, 35.6, 45.1, healthy
plastic_box, night, sour, 230.8, 20.6, 70.9, 5.3, rotten
plate, morning, spicy, 140.9, 55.7, 40.3, 50.6, healthy
plastic_box, evening, salty, 190.2, 85.4, 20.5, 30.8, healthy
plastic_box, morning, spicy, 105.6, 65.2, 28.4, 38.9, healthy
plate, evening, salty, 160.8, 95.3, 12.6, 22.1, old
plastic_box, night, sour, 210.5, 35.7, 55.3, 18.4, rotten
plastic_box, night, sour, 220.9, 45.3, 48.5, 12.1, old
plate, morning, spicy, 130.2, 67.8, 26.5, 37.8, healthy
plastic_box, evening, salty, 175.3, 88.6, 16.7, 27.2, old
plate, night, sour, 245.7, 33.9, 58.2, 9.4, rotten
plastic_box, morning, spicy, 112.5, 72.8, 33.9, 46.7, healthy
plate, evening, salty, 155.6, 98.4, 14.5, 21.5, old
plastic_box, night, sour, 235.1, 22.7, 68.3, 6.1, rotten
plate, morning, spicy, 125.8, 58.3, 39.6, 49.5, healthy
plastic_box, evening, salty, 185.2, 91.5, 21.2, 29.6, healthy
plate, night, sour, 270.3, 28.9, 63.4, 14.3, rotten
plastic_box, morning, spicy, 95.3, 52.1, 32.4, 42.5, healthy
plate, evening, salty, 150.7, 80.3, 20.5, 30.9, healthy
plate, evening, salty, 140.5, 75.2, 18.9, 28.3, healthy
plastic_box, morning, spicy, 108.7, 64.3, 29.5, 36.8, healthy
plate, evening, salty, 165.4, 92.7, 13.8, 24.7, old
plastic_box, night, sour, 215.9, 38.4, 53.7, 16.9, rotten

Ln 19, Col 52 | 2,045 characters
```

- (iii) Next I have classified using J48 Decision Tree algorithm. For that I have prepared a test dataset:

```
20210104032_TrainingDataset_Original.arff 20210104032_TestDataset_Original × +
File Edit View

@relation Food_Management

@attribute Food_container {plastic_box, plate}
@attribute Intake_time {morning, evening, night}
@attribute Food_taste {spicy, salty, sour}
@attribute rice_amount numeric
@attribute beef_amount numeric
@attribute dal_amount numeric
@attribute Vegetable_amount numeric
@attribute Status {healthy, rotten, old}

@data
plastic_box, morning, spicy, 101.5, 150.2, 20.8, 40.15, healthy
plastic_box, night, sour, 204.2, 90.1, 54.6, 23.3, old
plate, morning, spicy, 130.45, 67.8, 65.9, 15.2, healthy
plate, night, sour, 260.7, 25.31, 75.9, 75.2, rotten
plate, evening, salty, 190.5, 100.8, 10.7, 20.2, old
plastic_box, evening, salty, 280.6, 100.5, 15.3, 25.7, old
plate, night, sour, 250.9, 30.4, 60.1, 101.52, rotten
plastic_box, morning, spicy, 110.3, 71.2, 75.6, 45.1, healthy
plastic_box, night, sour, 230.87, 20.6, 72.9, 5.3, rotten
plate, morning, spicy, 143.9, 57.7, 40.3, 60.6, healthy
```

Next I have selected the classifier J48, selected plain text format and set cross fold validation to 10:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

☐ Use training set

☒ Supplied test set Set...

☐ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) Status

Start Stop

Result list (right-click for options)

10:07:36 - trees.J48

Classifier output

```
=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    Food_Management
Instances:    30
Attributes:   8
  Food_container
  Intake_time
  Food_taste
  rice_amount
  beef_amount
  dal_amount
  Vegetable_amount
  Status

Test mode:    user supplied test set: size unknown (reading incrementally)

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

J48 pruned tree
-----
Vegetable_amount <= 27.2
|  beef_amount <= 38.4: rotten (8.0)
|  beef_amount > 38.4: old (8.0)
Vegetable_amount > 27.2: healthy (14.0)

Number of Leaves :    3
Size of the tree :    5

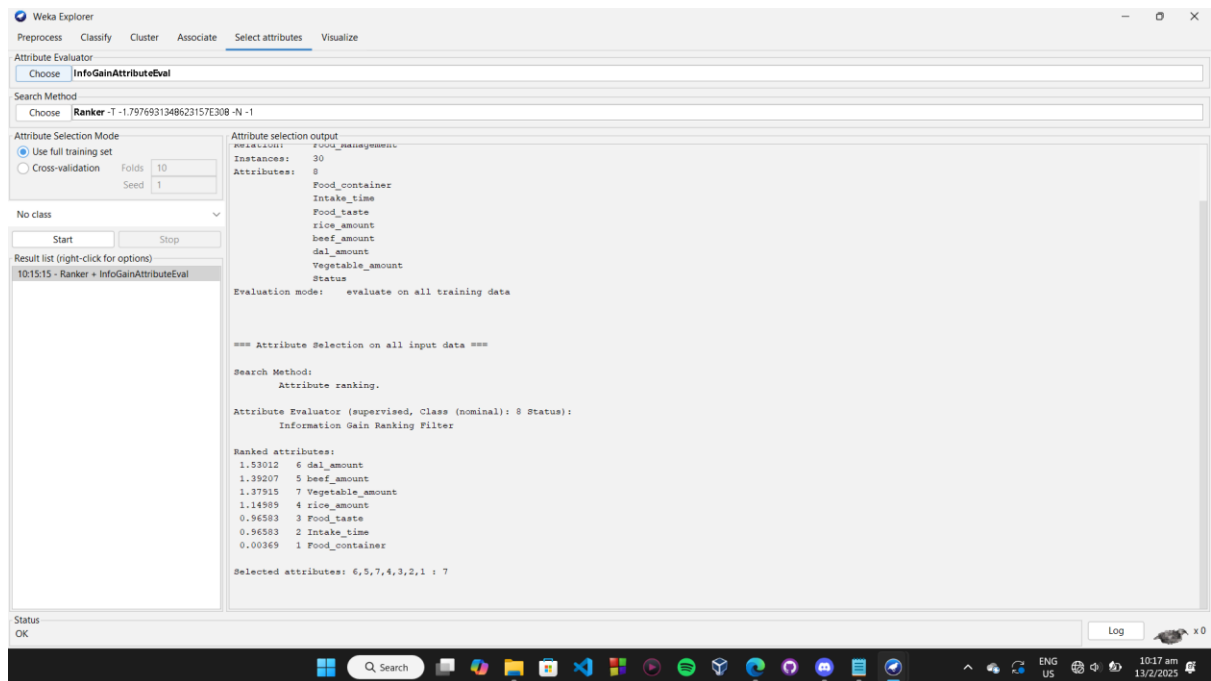
Time taken to build model: 0 seconds

=== Predictions on test set ===

inst#    actual    predicted error prediction
1 1:healthy 1:healthy 1
2 1:healthy 1:healthy 1
3 1:healthy 1:healthy 1
4 1:healthy 1:healthy 1
5 1:healthy 1:healthy 1
6 1:healthy 1:healthy 1
7 1:healthy 1:healthy 1
8 1:healthy 1:healthy 1
9 1:healthy 1:healthy 1
10 1:healthy 1:healthy 1
11 1:healthy 1:healthy 1
12 1:healthy 1:healthy 1
13 1:healthy 1:healthy 1
14 1:healthy 1:healthy 1
15 1:healthy 1:healthy 1
16 1:healthy 1:healthy 1
17 1:healthy 1:healthy 1
18 1:healthy 1:healthy 1
19 1:healthy 1:healthy 1
20 1:healthy 1:healthy 1
21 1:healthy 1:healthy 1
22 1:healthy 1:healthy 1
23 1:healthy 1:healthy 1
24 1:healthy 1:healthy 1
25 1:healthy 1:healthy 1
26 1:healthy 1:healthy 1
27 1:healthy 1:healthy 1
28 1:healthy 1:healthy 1
29 1:healthy 1:healthy 1
30 1:healthy 1:healthy 1
```

Status OK

- (iv) Using filter method I have found out the best classified attributes by this algorithm. For that I have chosen *InfoGainAttributeEval* as attribute evaluator and *Ranker* as search method. The results showd that dal_amount attribute has the best performance.



- (v) I have extracted 1 fold of the best performing attribute using *stratifiedRemoveFolds*.
- (vi) All fold values was present and the fold was balanced

