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

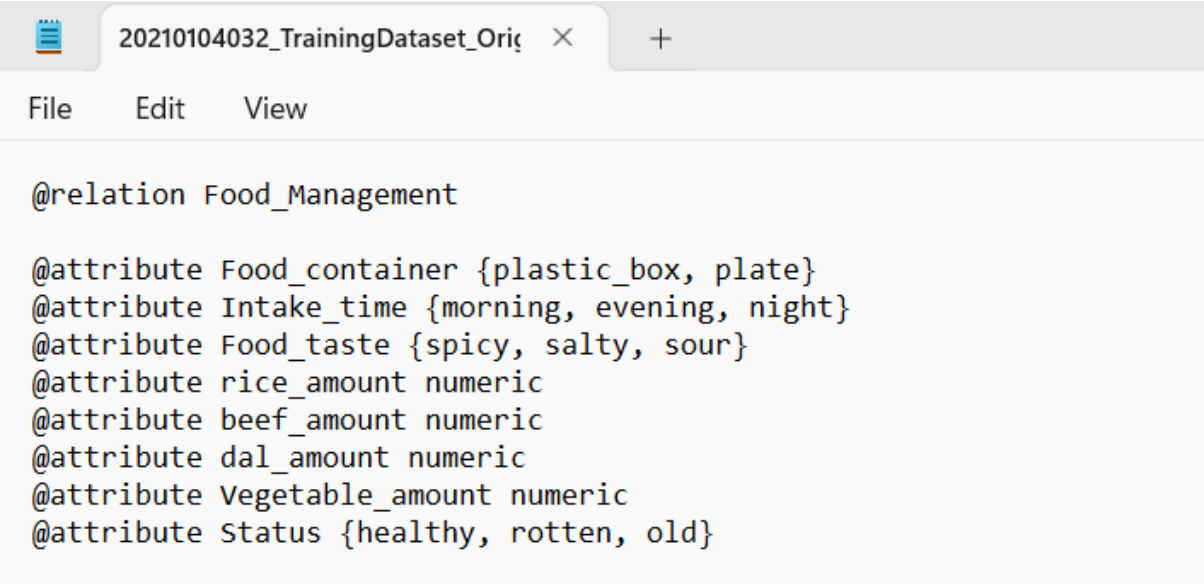
Assignment No: 02

Date of Submission: 13.02.2025

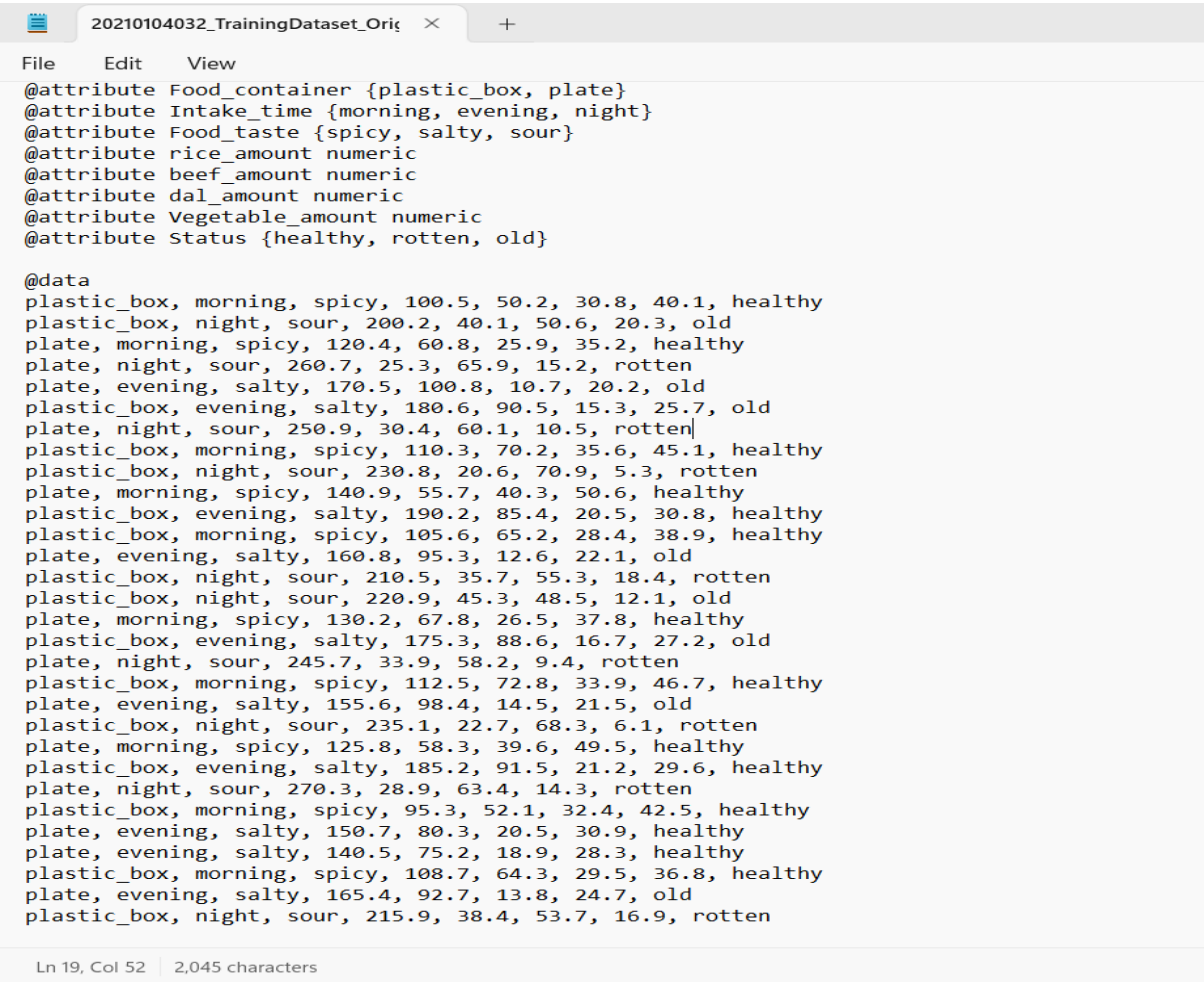
Submitted to:
Mr. Saha Reno
Assistant Professor
Department of CSE, AUST

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.



(ii) I have taken 30 instances.



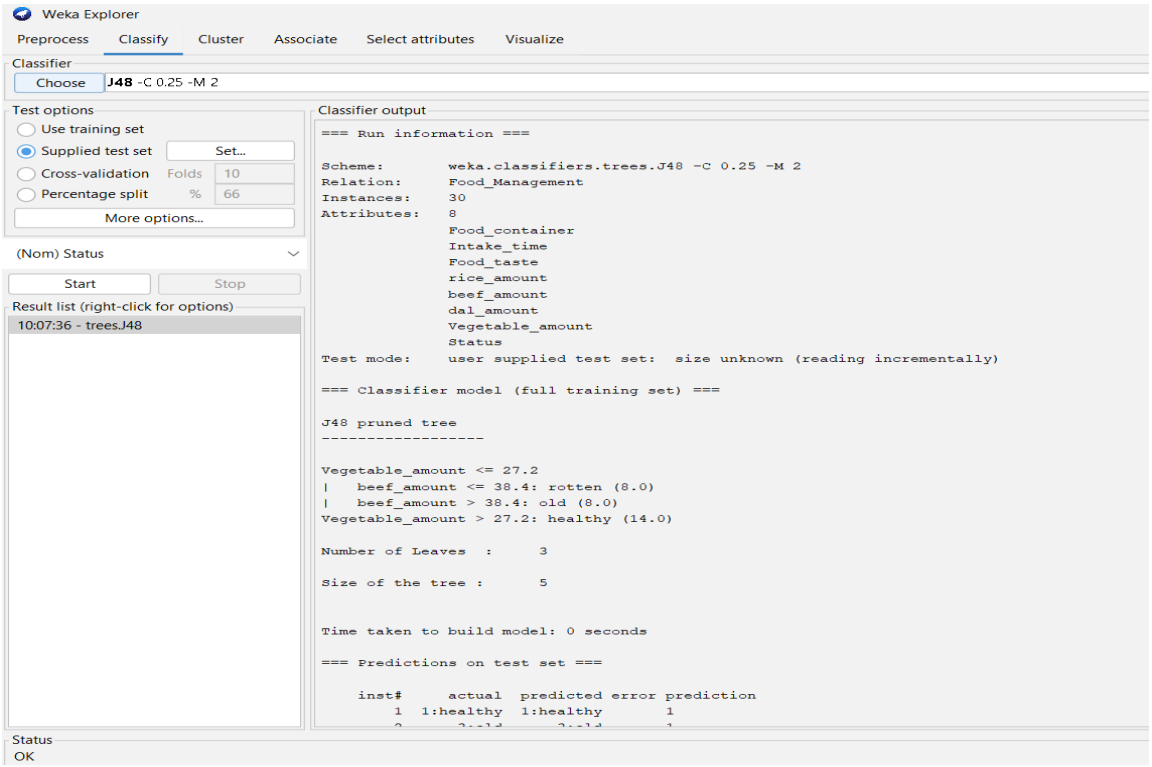
(iii) Next I have classified using J48 Decision Tree algorithm. For that I have prepared a test dataset:

20210104032_TrainingDataset_Original.arff | 20210104032_TestDataset_Original

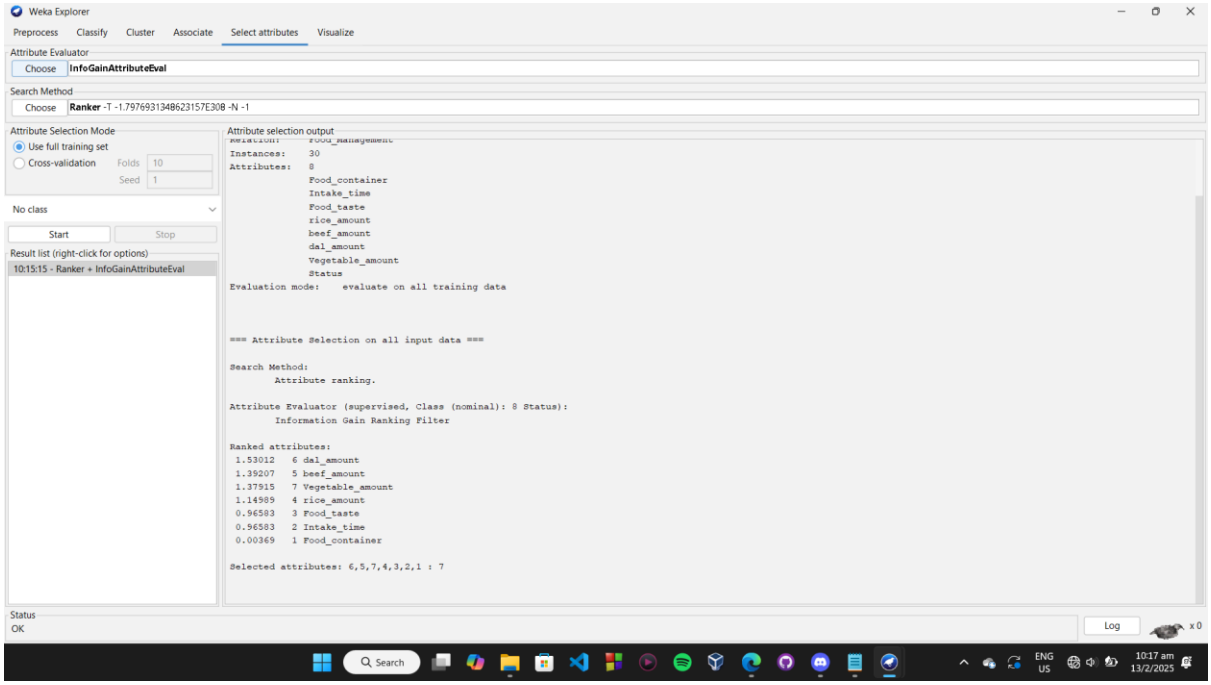
File Edit View

```
@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:



(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 *stratifidRemoveFolds*.

(vi) All fold values was present and the fold was balanced