Juan Carlos Estebes González – A01204421
Salomón Olivera Abud – A011715003
Artificial Intelligence
Decision Trees
Lab Report

For this lab, we use the same tests that are available in Alphagrader.

Example 1 - Play ball?

Weka tool solution:

```
Time taken to build model: 0.01 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances          14              100      %
Incorrectly Classified Instances         0                0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0       %
Root relative squared error              0       %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
              1.000    0.000    1.000      1.000   1.000      1.000  1.000     1.000     yes
              1.000    0.000    1.000      1.000   1.000      1.000  1.000     1.000     no
Weighted Avg. 1.000    0.000    1.000      1.000   1.000      1.000  1.000     1.000

=== Confusion Matrix ===

 a b   <-- classified as
 9 0 | a = yes
 0 5 | b = no
```

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     weather.symbolic
Instances:    14
Attributes:   5
              outlook
              temperature
              humidity
              windy
              play
Test mode:    evaluate on training data

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

J48 pruned tree
------------------

outlook = sunny
|   humidity = high: no (3.0)
|   humidity = normal: yes (2.0)
outlook = overcast: yes (4.0)
outlook = rainy
|   windy = TRUE: no (2.0)
|   windy = FALSE: yes (3.0)

Number of Leaves  :     5

Size of the tree :      8
```
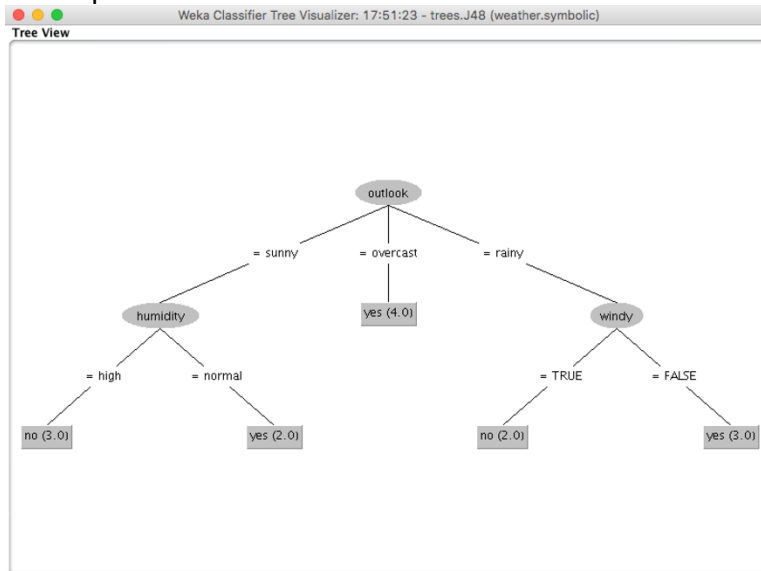
Juan Carlos Estebes González – A01204421
Salomón Olivera Abud – A011715003
Artificial Intelligence
Decision Trees
Lab Report

Our solution:

```
[MacBook-Pro-de-Juan:Decision-Trees estebes10$ python3 id3.py tests/play_ball.arff
outlook: sunny
  humidity: high
    ANSWER: no
  humidity: normal
    ANSWER: yes
outlook: overcast
  ANSWER: yes
outlook: rainy
  windy: TRUE
    ANSWER: no
  windy: FALSE
    ANSWER: yes
MacBook-Pro-de-Juan:Decision-Trees estebes10$
```

Example 2 - Contact lenses

Weka tool solution:

Juan Carlos Estebes González – A01204421
Salomón Olivera Abud – A011715003
Artificial Intelligence
Decision Trees
Lab Report

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     contact-lenses
Instances:    24
Attributes:   5
              age
              spectacle-prescrip
              astigmatism
              tear-prod-rate
              contact-lenses
Test mode:    evaluate on training data

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

J48 pruned tree
------------------

tear-prod-rate = reduced: none (12.0)
tear-prod-rate = normal
|   astigmatism = no: soft (6.0/1.0)
|   astigmatism = yes
|   |   spectacle-prescrip = myope: hard (3.0)
|   |   spectacle-prescrip = hypermetrope: none (3.0/1.0)

Number of Leaves  :     4

Size of the tree :      7
```

```
=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances          22              91.6667 %
Incorrectly Classified Instances         2               8.3333 %
Kappa statistic                          0.8447
Mean absolute error                      0.0833
Root mean squared error                  0.2041
Relative absolute error                 22.6257 %
Root relative squared error             48.1223 %
Total Number of Instances               24

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
                 1.000    0.053    0.833      1.000   0.909      0.889  0.974     0.833     soft
                 0.750    0.000    1.000      0.750   0.857      0.845  0.988     0.917     hard
                 0.933    0.111    0.933      0.933   0.933      0.822  0.967     0.972     none
Weighted Avg.    0.917    0.080    0.924      0.917   0.916      0.840  0.972     0.934

=== Confusion Matrix ===

  a  b  c   <-- classified as
  5  0  0 |  a = soft
  0  3  1 |  b = hard
  1  0 14 |  c = none
```
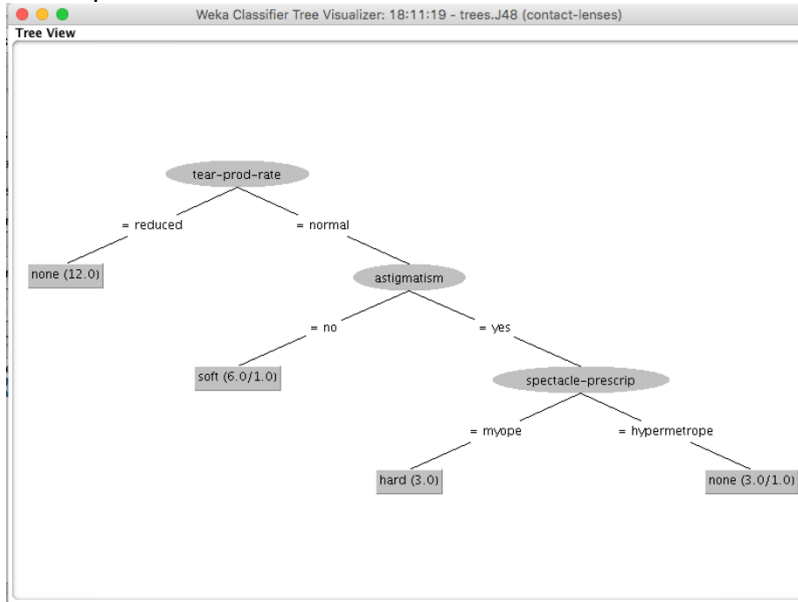
Juan Carlos Estebes González – A01204421
Salomón Olivera Abud – A011715003
Artificial Intelligence
Decision Trees
Lab Report

Out solution:

```
[MacBook-Pro-de-Juan:Decision-Trees estebes10$ python3 id3.py tests/contact_lenses.arff
tear-prod-rate: reduced
  ANSWER: none
tear-prod-rate: normal
  astigmatism: no
    age: young
      ANSWER: soft
    age: pre-presbyopic
      ANSWER: soft
    age: presbyopic
      spectacle-prescrip: myope
        ANSWER: none
      spectacle-prescrip: hypermetrope
        ANSWER: soft
  astigmatism: yes
    spectacle-prescrip: myope
      ANSWER: hard
    spectacle-prescrip: hypermetrope
      age: young
        ANSWER: hard
      age: pre-presbyopic
        ANSWER: none
      age: presbyopic
        ANSWER: none
MacBook-Pro-de-Juan:Decision-Trees estebes10$
```

Example 3 - OR function

Juan Carlos Estebes González – A01204421
Salomón Olivera Abud – A011715003
Artificial Intelligence
Decision Trees
Lab Report
Weka tool solution:

```
=== Summary ===

Correctly Classified Instances        3               75      %
Incorrectly Classified Instances      1               25      %
Kappa statistic                       0
Mean absolute error                   0.375
Root mean squared error               0.433
Relative absolute error               90      %
Root relative squared error           98.1981 %
Total Number of Instances             4

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
              1.000    1.000    0.750      1.000   0.857      ?      0.500     0.750     TRUE
              0.000    0.000    ?          0.000   ?          ?      0.500     0.250     FALSE
Weighted Avg. 0.750    0.750    ?          0.750   ?          ?      0.500     0.625

=== Confusion Matrix ===

 a b   <-- classified as
 3 0 | a = TRUE
 1 0 | b = FALSE
```

```
=== Run information ===

Scheme:         weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:       or
Instances:      4
Attributes:     3
                A
                B
                AorB
Test mode:      evaluate on training data

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

J48 pruned tree
------------------
: TRUE (4.0/1.0)

Number of Leaves  :     1

Size of the tree :      1


Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds
```


Our solution:

Juan Carlos Estebes González – A01204421
Salomón Olivera Abud – A011715003
Artificial Intelligence
Decision Trees
Lab Report

```
[MacBook-Pro-de-Juan:Decision-Trees estebes10$ python3 id3.py tests/or_function.arff
 A: TRUE
   ANSWER: TRUE
 A: FALSE
   B: TRUE
     ANSWER: TRUE
   B: FALSE
     ANSWER: FALSE
MacBook-Pro-de-Juan:Decision-Trees estebes10$ ▏
```

As conclusion doing this lab, both implementations provides a solution for decision trees problems, however, implement our own program give us the chance to understand how Decision Trees algorithms work and if we have to modify the algorithm to adapt it to others problems, we know how it is implemented and we can make necessary changes without complications, however, WEKA tool provides a good visual representation of the tree created and it is easier for users to understand how decisions are represented, due to this, our solution requires a user interface more suitable for users. On the other hand, WEKA also gives values for each leaf node and not just the result.

As we said before, to compare results between our implementation and Weka tool, we use the same tests available in Alphagrader platform, however in the algorithm that we use, we implement create a node with its corresponding information, a method to calculate the entropy for each node and a function to calculate the gain per node and for the total gain, then we just create a simple visual representation. By the way, there was little difference in the last test, because Weka just shows the result node of the solution and not the whole tree.

Based on we learned about Decision Trees, we could use them in problems to make predictions, accept or reject a proposal given some information and goals, increase or not the production capacity for a company, probability of success of an advertising campaign, minimize costs, etc. Finally, we can say that decision trees are useful for problems where alternatives or actions are well defined, if uncertainties can be quantified and the goals are clear.