Nancy Espinosa Castillo - A01206782
Joel Alexis Bustamante Calvario - A00226215

Decision Trees

Decision trees (DT) are diagrams in tree-shaped that are used to determine a course of action or show a static probability. Each branch of the decision tree represents a possible decision, occurrence or reaction and each option may lead to the next.

As an example of Decision Tree, the team used a data set that represented different conditions of an experiment with balloons. For search a property data set, we specified the next three facts:

- A data set Multivariate, to have a data set with more of one variable.
- A data set Categorical, to have discrete attributes.
- A data set with Classification, to helps to ensure the discrete attributes.

With this features we selected the data set of Balloons [1] that contains 16 instances and 5 attributes.

The attributes of the set are the next:

- Color with the values yellow, purple
- size with the values large, small
- act with the values stretch, dip
- age with the values adult, child
- inflated with the values T, F

We made a Python program to generate the tree and we compared the result of our program with a specialized tool, Weka.

In general, the graphic representation of the tree is shown in the Figure 1.

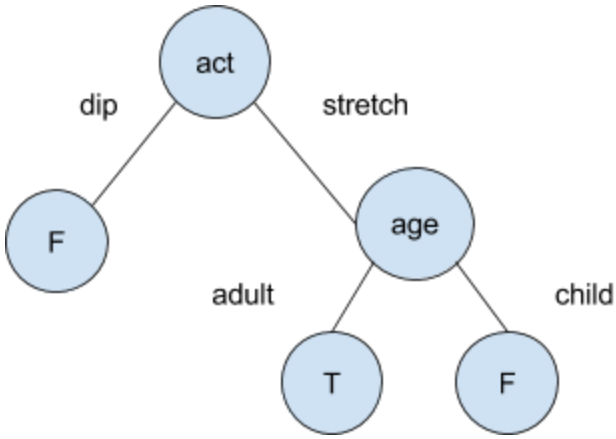

Figure 1. Balloons data set

In the Python program, the input is represented by the next code

@attribute Color {YELLOW,PURPLE}
@attribute size {LARGE,SMALL}
@attribute act {STRETCH,DIP}
@attribute age {ADULT, CHILD}
@attribute inflated {T,F}

@data
YELLOW,SMALL,STRETCH,ADULT,T
YELLOW,SMALL,STRETCH,ADULT,T
YELLOW,SMALL,STRETCH,CHILD,F
YELLOW,SMALL,DIP,ADULT,F
YELLOW,SMALL,DIP,CHILD,F
YELLOW,LARGE,STRETCH,ADULT,T
YELLOW,LARGE,STRETCH,ADULT,T
YELLOW,LARGE,STRETCH,CHILD,F
YELLOW,LARGE,DIP,ADULT,F
YELLOW,LARGE,DIP,CHILD,F
PURPLE,SMALL,STRETCH,ADULT,T
PURPLE,SMALL,STRETCH,ADULT,T
PURPLE,SMALL,STRETCH,CHILD,F
PURPLE,SMALL,DIP,ADULT,F
PURPLE,SMALL,DIP,CHILD,F
PURPLE,LARGE,STRETCH,ADULT,T
PURPLE,LARGE,STRETCH,ADULT,T
PURPLE,LARGE,STRETCH,CHILD,F
PURPLE,LARGE,DIP,ADULT,F

PURPLE,LARGE,DIP,CHILD,F

And the output was

```
act: STRETCH
  age: ADULT
    ANSWER: T
  age: CHILD
    ANSWER: F
act: DIP
  ANSWER: F
```

On the other hand, after load the data set in Weka, we obtained the representation and the results of the Figure 2.

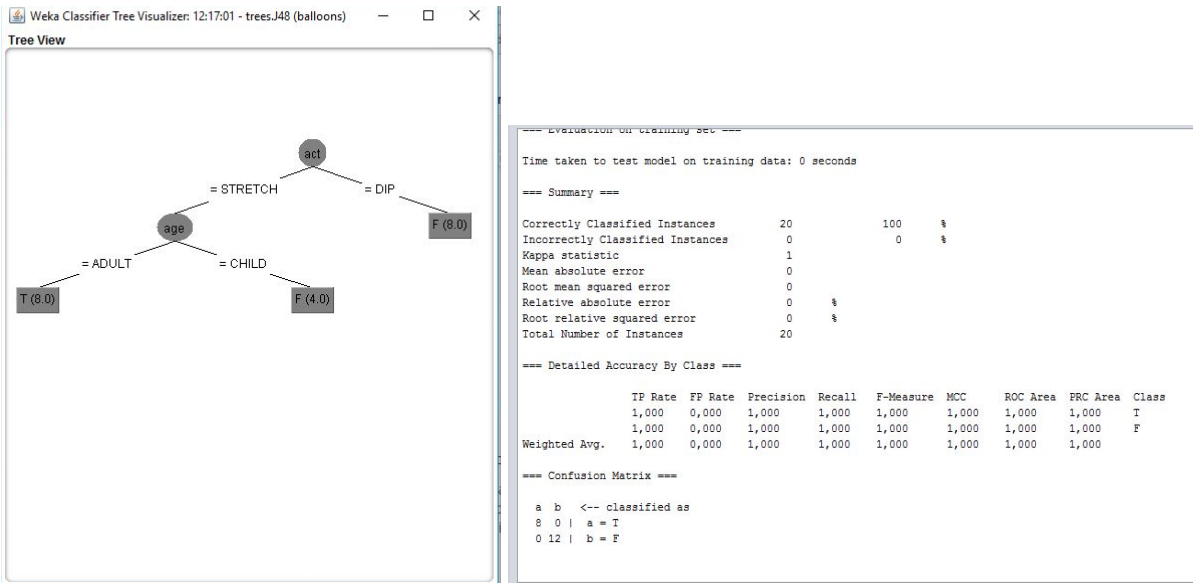

Figure 2.

Also, after working and comparing these tools between them, we can say that the advantages of the Python program over Weka are :

- It is possible to know exactly how the algorithm works.
- It is possible to create a more intuitive interface than Weka.
- The code can be adapted according to custom needs.
- The code can be optimized.
- You don't need to pass the learning curve such as Weka.

Although the disadvantages are:

- It is necessary to make the code and this takes certain amount of time.
- It could have defects..

- It does not have the large set of different data sets and algorithms like Weka.
- It could have a low level of accuracy because it has not been tested sufficiently.

In terms of differences between WEKA and the Python program according to the results, we can say that in the case of Weka, the algorithm which generates the tree is J48 that is a model based on C4.5 which in general is a better version of ID3. The reason is that ID3 model is implemented to know a fast result without counting factors like unavailable values, continuous attribute value ranges, etc. On the other hand, the Python program reads all the data set for the training and creates the tree using the ID3 algorithm.

Finally, the team considers that the fields of application of decision trees are huge, they can be useful in all the situations where a decision should be taken and exists previous information to take in consideration. For example in market positioning, crime forensics, predicting performance of personnel or machines, in the videogames are used to define the decisions an enemy is going to take in different situations, a bank to know the risk of giving a loan, etc.

References:

1. https://archive.ics.uci.edu/ml/datasets/Balloons