

Using Data Mining Tools WEKA

Writing a program for making decision trees has the advantage of knowing the algorithm and what it is doing, also the output can be modified to serve specific purposes and link it to other programs without needing external components. The disadvantages are the quality, efficiency and fiability of the results depend on the implementation, if the algorithm if highly efficient then the results may contain errors, or if it is extremely accurate it can result in low efficiency. The advantages of using a pre-created suite are that it has a correct balance from the previous attributes and its trust depends on the quantity of their users. Also it provides multiple tools to understand, process, manipulate and visualize datasets. The disadvantages are that linking it to an existing solution can be difficult, also it may have more functions that what is needed thus creating more complexity in the process of understanding and using the application.

The datasets were selected using the examples from the test cases, also due to the limitations of the WEKA suite we chose a dataset with few columns. After a quick search in the dataset collection from the UCI Machine Learning Repository we found the balloon dataset, that had few columns, and the values were similar to the ones from the test cases.

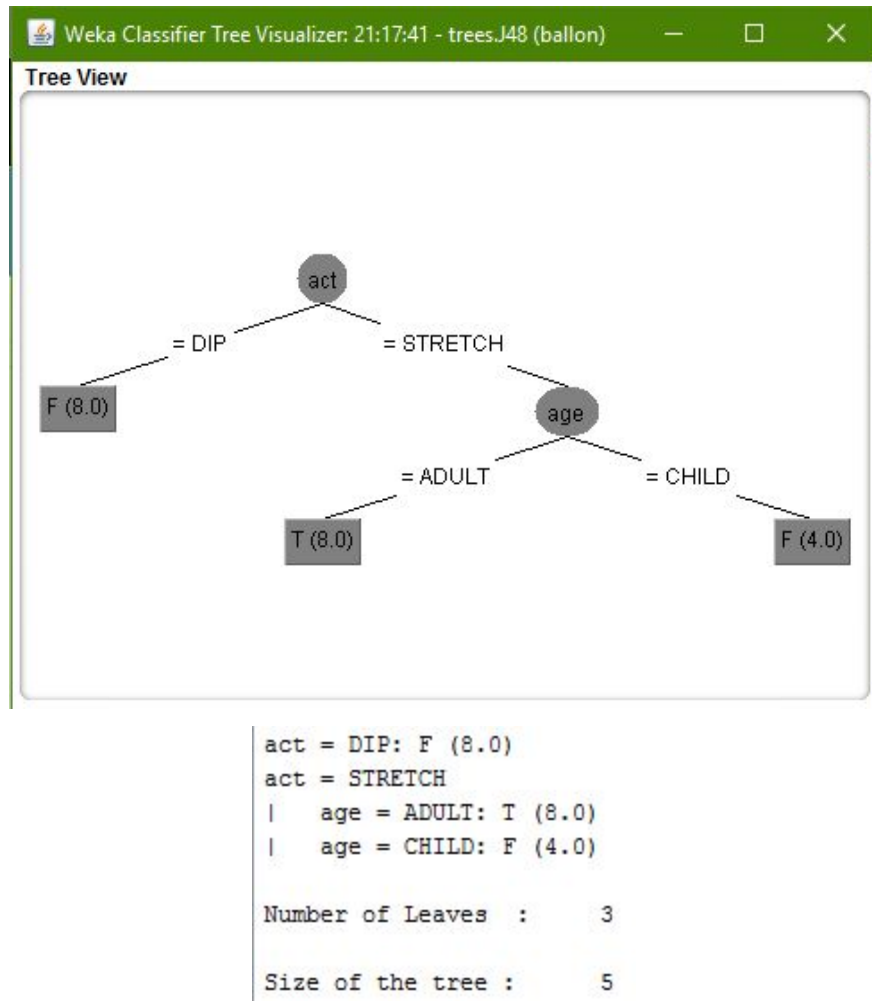
Comparing the implementation of a suite with the python script made for the lab we obtained different trees, with the same information because algorithms process the information with different approaches. Both implementations are able to read the information in a specific format (.aff files) and create text output with the tree.

Decision trees help to know the probable outcome of a defined behaviour, for example it can be used to predict the eye color of a baby using the information of its relatives. They could be used to predict the behaviour or a person after analyzing his habits, for example processing information about his food preferences and current budget to predict where he is going to eat to estimate how much time will be spent and an estimated expense. Also considering the tendency of smart cities, the same analysis could be made for restaurants to predict the quantity of clients that will order a certain dish. Another use for the decision trees is in the grammar of a compiler or interpreter, it show us the behavior of the grammar. The decision trees are helpful to easy interpret and explain it to executives or developers.

Results from each implementation processing the data set from:

<https://archive.ics.uci.edu/ml/datasets/Balloons>

Weka tree output:



Script output:

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

References:

Tutorial Exercises for the Weka Explorer. (2011). Retrieved from:

https://moodle.umons.ac.be/pluginfile.php/43703/mod_resource/content/2/WekaTutorial.pdf

Pazzani, M. (s.f.). Balloon Data Set. Retrieved from:

<https://archive.ics.uci.edu/ml/datasets/Balloons>