Lab 5: Decision trees

Write a 400-500 word report that includes the following points:

1. Explain the advantages and disadvantages of writing a program on your own vs using a pre-created suite such as WEKA.

   The advantage of using direct implementations by programming is that you are able to understand how the algorithm works and this makes you easier the process of debugging and discover what is bad in the decision tree. As it is a row implementation you can focus on the expected output and its format, how it will be used later and also in which language will it be more suitable to be programed. In tools like WEKA the output is not that flexible, this tools give extra information perhaps that can be useful to have near. This tools make more transparent of data analysis.

2. Explain what criteria you followed to choose the datasets for your tree and the WEKA tests.

   The tutorial is clear and we follow it, we tried with several datasets but then we realize that we wanted to test with a simple ones, the ones we can evaluate if it was acting in the correct way both the program and weka. Some of datasets have different formats or characters that we were not aware of. We choosed one that classifies animals by types.

   https://raw.githubusercontent.com/renatopp/arff-datasets/master/classification/zoo.arff

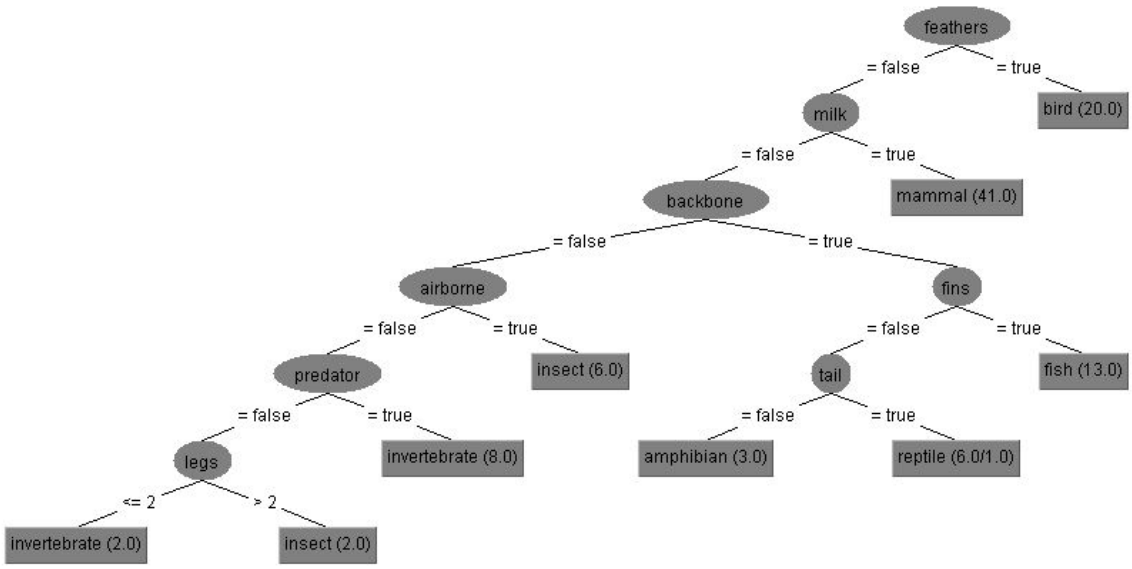
3. Include the graphics of the trees or part of the trees you generated in WEKA and your own program. Are they different, and if so, why?

   Program output:

```
animal: aardvark
    ANSWER: mammal
animal: antelope
    ANSWER: mammal
animal: bass
    ANSWER: fish
animal: bear
    ANSWER: mammal
animal: boar
    ANSWER: mammal
animal: buffalo
    ANSWER: mammal
animal: calf
    ANSWER: mammal
animal: carp
    ANSWER: fish
animal: catfish
    ANSWER: fish
animal: cavy
    ANSWER: mammal
animal: cheetah
    ANSWER: mammal
animal: chicken
    ANSWER: bird
animal: chub
    ANSWER: fish
animal: clam
                ANSWER:
invertebrate
animal: crab
                ANSWER:
invertebrate
animal: crayfish
                ANSWER:
invertebrate
animal: crow
    ANSWER: bird
animal: deer
    ANSWER: mammal
animal: dogfish
    ANSWER: fish
animal: dolphin
    ANSWER: mammal
animal: dove
    ANSWER: bird
animal: duck
    ANSWER: bird
animal: elephant
    ANSWER: mammal
animal: flamingo
    ANSWER: bird
animal: flea
    ANSWER: insect
animal: frog
    ANSWER: amphibian
animal: fruitbat
    ANSWER: mammal
animal: giraffe
    ANSWER: mammal
animal: girl
```

```
    ANSWER: mammal
animal: gnat
    ANSWER: insect
animal: goat
    ANSWER: mammal
animal: gorilla
    ANSWER: mammal
animal: gull
    ANSWER: bird
animal: haddock
    ANSWER: fish
animal: hamster
    ANSWER: mammal
animal: hare
    ANSWER: mammal
animal: hawk
    ANSWER: bird
animal: herring
    ANSWER: fish
animal: honeybee
    ANSWER: insect
animal: housefly
    ANSWER: insect
animal: kiwi
    ANSWER: bird
animal: ladybird
    ANSWER: insect
animal: lark
    ANSWER: bird
animal: leopard
    ANSWER: mammal
animal: lion
    ANSWER: mammal
animal: lobster
              ANSWER:
invertebrate
animal: lynx
    ANSWER: mammal
animal: mink
    ANSWER: mammal
animal: mole
    ANSWER: mammal
animal: mongoose
    ANSWER: mammal
animal: moth
    ANSWER: insect
animal: newt
    ANSWER: amphibian
animal: octopus
              ANSWER:
invertebrate
animal: opossum
    ANSWER: mammal
animal: oryx
    ANSWER: mammal
animal: ostrich
    ANSWER: bird
animal: parakeet
    ANSWER: bird
animal: penguin
    ANSWER: bird
animal: pheasant
    ANSWER: bird
animal: pike
    ANSWER: fish
animal: piranha
    ANSWER: fish
animal: pitviper
    ANSWER: reptile
animal: platypus
    ANSWER: mammal
animal: polecat
    ANSWER: mammal
animal: pony
    ANSWER: mammal
animal: porpoise
    ANSWER: mammal
animal: puma
    ANSWER: mammal
animal: pussycat
    ANSWER: mammal
animal: raccoon
    ANSWER: mammal
animal: reindeer
    ANSWER: mammal
animal: rhea
    ANSWER: bird
animal: scorpion
              ANSWER:
invertebrate
animal: seahorse
    ANSWER: fish
animal: seal
    ANSWER: mammal
animal: sealion
    ANSWER: mammal
animal: seasnake
    ANSWER: reptile
animal: seawasp
              ANSWER:
invertebrate
animal: skimmer
    ANSWER: bird
animal: skua
    ANSWER: bird
animal: slowworm
    ANSWER: reptile
animal: slug
              ANSWER:
invertebrate
animal: sole
    ANSWER: fish
animal: sparrow
    ANSWER: bird
animal: squirrel
    ANSWER: mammal
animal: starfish
              ANSWER:
invertebrate
animal: stingray
    ANSWER: fish
animal: swan
    ANSWER: bird
animal: termite
    ANSWER: insect
animal: toad
    ANSWER: amphibian
animal: tortoise
    ANSWER: reptile
animal: tuatara
    ANSWER: reptile
animal: tuna
    ANSWER: fish
animal: vampire
    ANSWER: mammal
animal: vole
    ANSWER: mammal
animal: vulture
    ANSWER: bird
animal: wallaby
    ANSWER: mammal
animal: wasp
    ANSWER: insect
animal: wolf
    ANSWER: mammal
animal: worm
              ANSWER:
invertebrate
animal: wren
    ANSWER: bird
```

Weka tree output:




```
feathers = false
|   milk = false
|   |   backbone = false
|   |   |   airborne = false
|   |   |   |   predator = false
|   |   |   |   |   legs <= 2: invertebrate (2.0)
|   |   |   |   |   legs > 2: insect (2.0)
|   |   |   |   predator = true: invertebrate (8.0)
|   |   |   airborne = true: insect (6.0)
|   |   backbone = true
|   |   |   fins = false
|   |   |   |   tail = false: amphibian (3.0)
|   |   |   |   tail = true: reptile (6.0/1.0)
|   |   |   fins = true: fish (13.0)
|   milk = true: mammal (41.0)
feathers = true: bird (20.0)
```

The outputs are different one from the other, at first we were thinking that our program was incorrectly build, but analysing the data we first realize that WEKA decisions were not the correct ones, then we change dataset and we realize the format or WEKA was not such appropriate so it differs from our output but it was correct.

We think that the output of our program is clearer than WEKA's output, because as our has indentation this helps to understand more the tree.

4. Based in what you have learned so far where would you use decision trees?

   Decision trees are easy to understand,to code, visualize, manipulate, and explain and more advanced classifiers, clustering, and machine learning may be more accurate for large data sets, but the advanced algorithms can't be easily visualized or manipulated.