INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY

Artificial Intelligence

Report Lab Decision Trees

Equipo:

Melannie Isabel Torres Soto	A01361808
Bernardo Gómez Romero	A01209704

Due Date:

October 29, 2018

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

A great advantage of WEKA is that you can choose the algorithm you want to use, so you aren't forced to stick with a specific algorithm like in our program. Another advantage is that you can visualize the tree in a graphical way which is very helpful to get a better insight on the data.

On the other hand, writing your own program gives you more flexibility to build a solution specifically for your problem and your needs. You can clean your data, make the necessary data transformations and set more restrictions to avoid overfiting your data, in order to build a more precise model.

Which datasets can our program handle and which ones it can't (will break it).

Our program can handle:

- discrete possible results
- discrete possible inputs
- null input values

Our program can't handle:

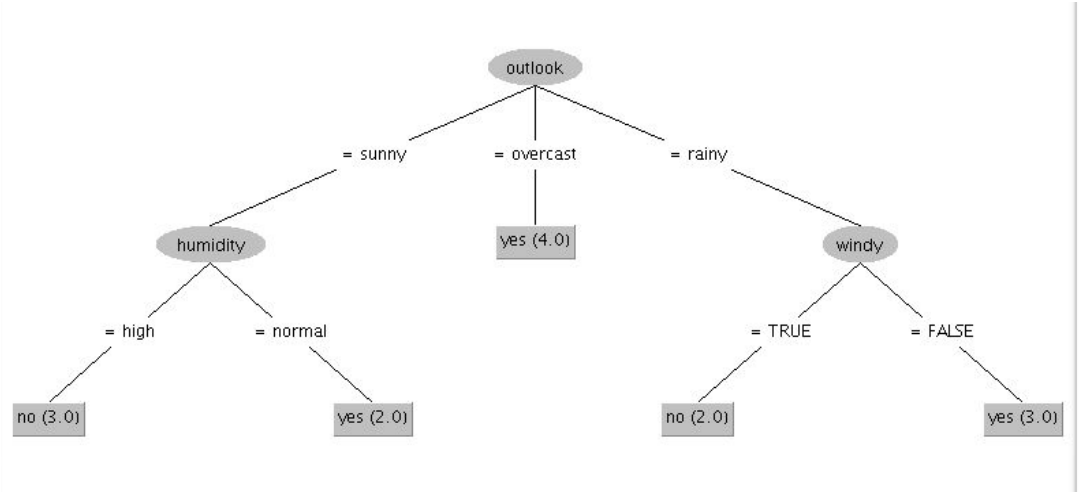
- continuos data as results
- continuos data as inputs

To summarize, our program just classifies the input, it doesn't perform calculations to make continuos estimations of an event, it can only handle discrete values and thanks to pandas it can also handle null values.

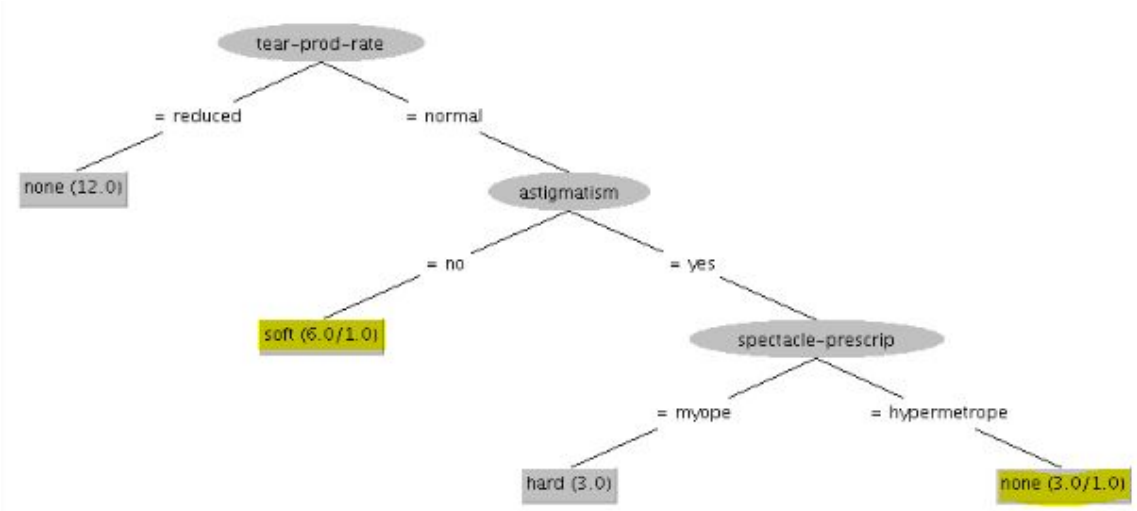
Include the graphics of the trees or part of the trees you generated in WEKA and your own program. Are they different, and if so, why?

Test 3

```
outlook: sunny
  humidity: high
    ANSWER: no
  humidity: normal
    ANSWER: yes
outlook: overcast
  ANSWER: yes
outlook: rainy
  windy: TRUE
    ANSWER: no
  windy: FALSE
    ANSWER: yes
```



Test 4




```
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
```

Test 5

```
A: TRUE
  ANSWER: TRUE
A: FALSE
  B: TRUE
    ANSWER: TRUE
  B: FALSE
    ANSWER: FALSE
```

Our tests results are different because our program and WEKA use distinct algoritms to generate the decision tree, our program uses ID3 while these Weka examples use J48. In the first test case the tree is identical and this could be because the dataset is too small to have significant changes on the model generated. In the second test we can clearly see that the outputs are different, being this diference that WEKA allows some impurity on the leaf nodes. The algorithm used in the WEKA may stop making splits at some value of information gain. This is done to avoid overfiting, while in our program, all our trees will be overfitted to the data given.

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

So far we've seen examples in medical diagnosis, weather predictions, business decision making, but to generalize we can use supervised learning algorithms and in this case, decision trees, to solve clasification problems. In these type of problems we must have previous taken data that we can use to train and test a model. For example we can build a decision tree using some training data, after that we can test its performance with an unseen cross-origin dataset and after proving its eficiency and precision we can make predictions using this model. This allows us to take better decisions in deterministic and non-deterministic contexts.