Lab 12 : Implementing Bayesian Networks

Ramón Romero - A01700318

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

  Depending on the scenario is to decide if use a tool as WEKA or write your own code.

  Some advantages of using your own code are that you can always change it according to your needs, but on the other hand, is necessarily a correct understanding of the algorithms in order to implement them in a proper way. By using tools similars to WEKA it is quicker to generate a solution, but it is limited to the approach that this kind of tools has already implemented.

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

  I choose datasets which data was labelled and which only contains defined elements inside these labels ex: @attribute value { true, false}. My implementation was not designed to manage reserved words as Real or Integer.

- 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?

  They are very similar, but it starts to differ as the tree goes deeper, it is because of the algorithm that is implemented in WEKA, C4.5 (**J48**), is an extension of ID3 which add others factors to evaluate and generate.
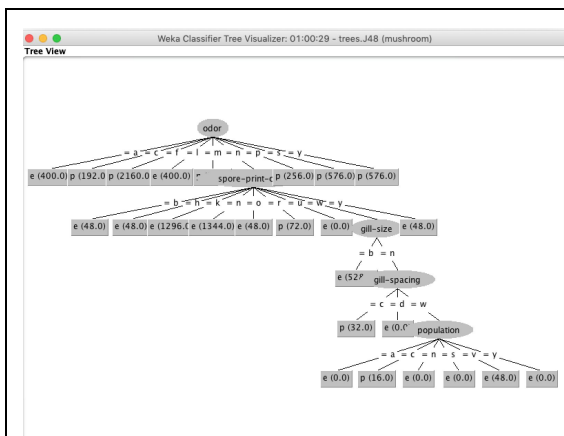
	'odor': 'a'
	  ANSWER: 'e'
	'odor': 'c'
	  ANSWER: 'p'
	'odor': 'f'
	  ANSWER: 'p'
	'odor': 'l'
	  ANSWER: 'e'
	'odor': 'm'
	  ANSWER: 'p'
	'odor': 'n'
	  'spore-print-color': 'b'
	    ANSWER: 'e'
	  'spore-print-color': 'h'
	    ANSWER: 'e'
	  'spore-print-color': 'k'
	    ANSWER: 'e'
	  'spore-print-color': 'n'

	ANSWER: 'e' 'spore-print-color': 'o' 　ANSWER: 'e' 'spore-print-color': 'r' 　ANSWER: 'p' 'spore-print-color': 'w' 　'habitat': 'd' 　'gill-size': 'b' 　　ANSWER: 'e' 　'gill-size': 'n' 　　ANSWER: 'p' 　'habitat': 'g' 　　ANSWER: 'e' 　'habitat': 'l' 　'cap-color': 'c' 　　ANSWER: 'e' 　'cap-color': 'n' 　　ANSWER: 'e' 　'cap-color': 'w' 　　ANSWER: 'p' 　'cap-color': 'y' 　　ANSWER: 'p' 　'habitat': 'p' 　　ANSWER: 'e' 　'habitat': 'w' 　　ANSWER: 'e' 　'spore-print-color': 'y' 　　ANSWER: 'e' 'odor': 'p' 　ANSWER: 'p' 'odor': 's' 　ANSWER: 'p' 'odor': 'y' 　ANSWER: 'p'

The other threes and images can be found on the directory ./results

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

I would use this technique when the amount of attributes in my data set is small, because the complexity (deepness) of the tree is completely related to them. Also, it is important to consider all techniques and variants of this technique, as could be pruning or J48, which can be useful to find a better approach to our problem.