

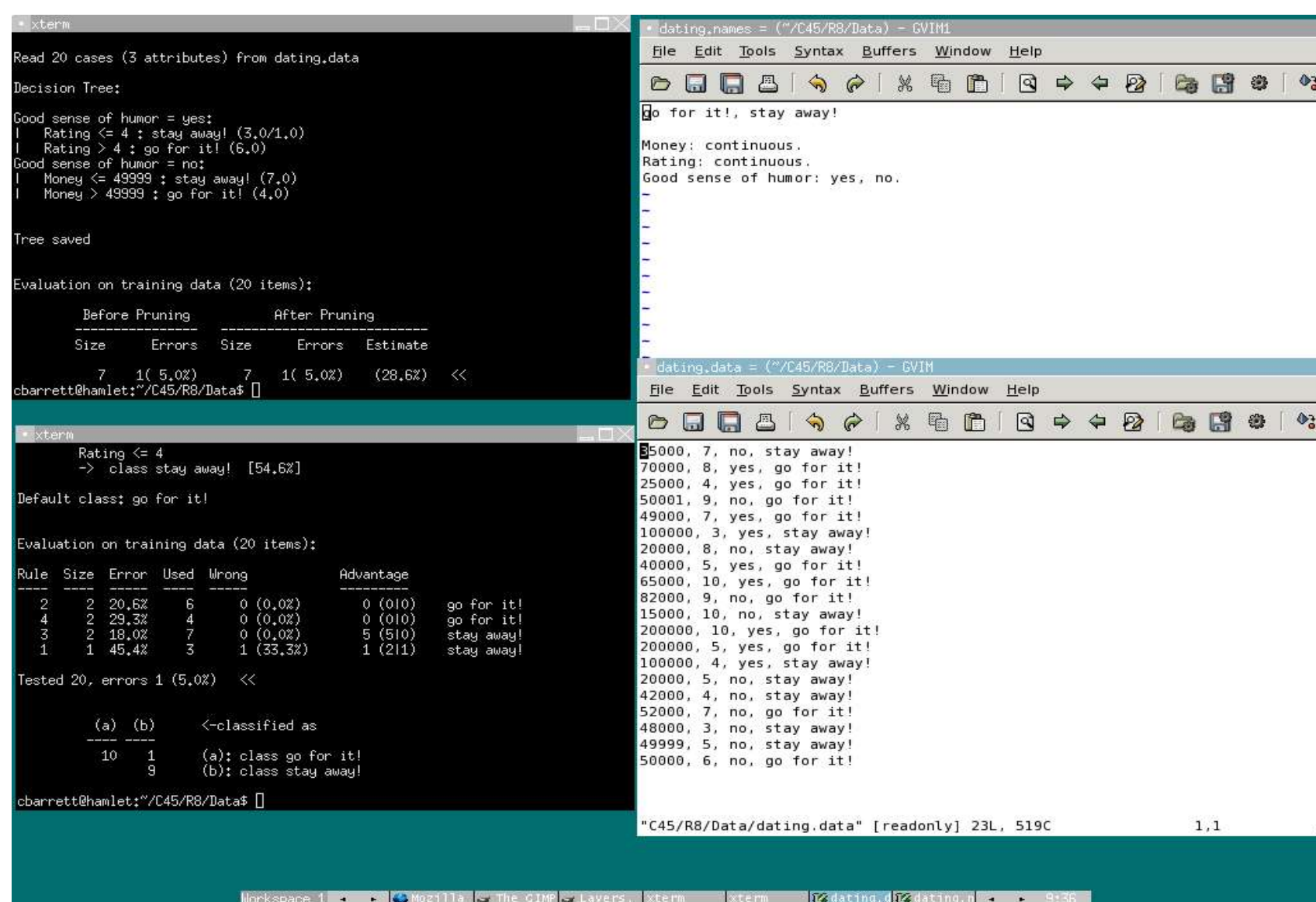
# The Implementation of a Decision Tree for Career Guidance

Casey Barrett 2004-2005  
Computer Systems Lab

## Background

A decision tree is a graphical representation of the decision analysis process. This type of tool consists of some sort of input, whether it is a situation or an object. This input is then sent through a set of parameters, or “rules”, and eventually the tree gives a Boolean output. There are many different types of parameters that can be used. Or in other words, many different types of parameter cases can be used. These cases can include numerical data, simple yes/no answers, or word answers, such as hair color (“black”, “brown”, or “blonde”). Each parameter will have a specified set of cases that correspond to the parameter.

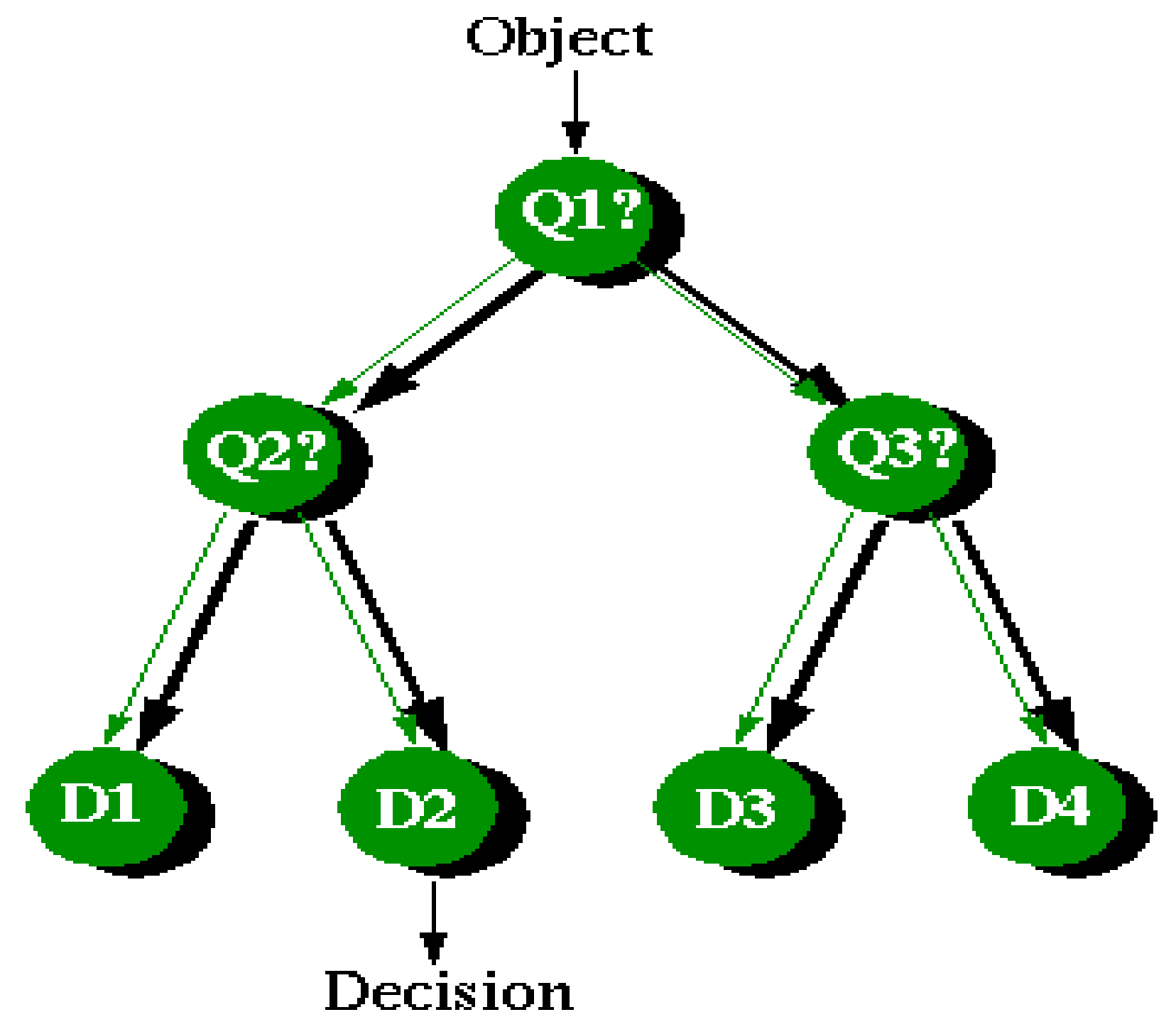
As many high school students prepare for college and in some cases, life in the real world, many are directionless and confused. They do not know what to study in college. College studies are important because generally, a student's career will be dictated by their studies. I propose a group of career guiding decision trees to help students find their interests and hopefully steer them into focusing on these interests.



```
Read 20 cases (3 attributes) from dating.data
Decision Tree:
Good sense of humor = yes:
1. Rating < 4.1: stay away! (3,0/1,0)
1. Rating > 4.1: go for it! (6,0)
Good sense of humor = no:
1. Money < 45999: stay away! (7,0)
1. Money > 45999: go for it! (4,0)
Tree saved
Evaluation on training data (20 items):
Before Pruning      After Pruning
Size  Errors  Size  Errors  Estimate
7  1 (5.0%)  7  1 (5.0%)  (20.62)
cbarett@wallet:~/C45/RR/Data$

Rule Size Error Used Wrong Advantage
2 2 20.62 0 0 (0.0%) go for it!
4 2 29.32 4 0 (0.0%) go for it!
3 2 31.02 7 0 (0.0%) stay away!
1 1 45.42 3 1 (33.3%) 1 (2/1) stay away!
Tested 20, errors 1 (5.0%)

(a) (b) <-classified on
10 1 (a): class go for it!
9 0 (b): class stay away!
```



## Abstract

This research project will be an investigation into the design and implementation of various decision trees for career guidance. A decision tree takes into account some sort of situation outlined by a group of parameters and outputs a Boolean decision to the situation. This project will take into account many aspects associated with decision trees including database building, searching and sorting, and algorithms for accessing data.

My project utilizes numerous decision trees in an effort to serve as a tool for career guidance for young adults. A user will fill out a form of specified fields that will then be analyzed by the group of decision trees until a field of study/occupation is given to the user as the outcome. This group of decision trees will be built through database building techniques.

## General Tree

For career guidance, I separated the decision trees that I would need into three different and distinct trees. The first one is a tree designed to help a user decide whether they should focus on either of the broad intellectual categories of the sciences or the liberal arts. This decision will be decided based a series of fields the user fills out in a separate C++ program. Some of these fields include the user's grades in their current English and Math class, if the user is in a science club, the number of computers that the user owns, and the number of plays the user has participated in during the last year. These starting questions are somewhat broad because this is the first preset tree that the user will be compared to. All in all, there are eight separate parameters that make up the fields for the decision tree.

## Science Tree

The other decision tree that needed to be made was the Science-specific tree. The process for setting this tree up was almost identical to that of the previously described Liberal Arts tree. The same three files were needed for the Science as in the other two cases. These files include the .txt file, the .names file, and the .data file. There are six questions that will be asked to the user in the .txt file. These questions are "How many calculus or post-calculus courses have you taken?", "Do you enjoy solving problems?", "Are you a hippie?", "Have you ever taken apart an electrical device?", "Do you enjoy living things?", and the ever-popular "Do you like explosions?". (Appendix 3.1) The answer to the first question will be a non-negative integer while the all of the other questions are basic yes/no answers. Although these questions seem a bit more fanciful, they do serve a purpose in career guidance.

## Language Arts Tree

The first of the other two decision trees created was the Liberal Arts tree. Just like the General Tree, three separate files were created that could be used by the User Input Program as well as the C4.5 program. The first one, the la-questions.txt file, has the list of questions that the user will be asked that relate to the Liberal Arts field. These questions include "Would you rather read a fictional novel or a historical account of an important event?", "Do you enjoy leading people", "How many times a month do you go to church?", "Would you rather lead children or citizens?". (Appendix 2.1) These questions will be prompted to the user who will then type in his or her responses. These questions were just a sample of the complexity that most career guidance programs will go through. After researching fields within the Liberal Arts, I decided upon these questions in order to make an easy to follow decision tree with little ambiguity.