Southern New Hampshire University

DAT 520 Module One Overview

The first module introduces you to the R software needed for assignments and the final project. The problem set gives you practice with matrix multiplication and exponentiation.

The readings concentrate on describing decision analysis and showing the horizontal breadth of the discipline at large. With these three basics completed in the first module, the mystery of decision analysis should start to dissolve:

Decision Analysis . . .

- Has many moving parts, yet it is not that complex.
- Has multiple approaches toward the same end.
- Is characterized as nonlinear thinking for answering research questions.
- Is cross-discipline and applicable to any type of situation requiring deduction of an optimal choice.
- Has its roots in direct calculations of probability and use of proportions, unlike traditional inferential statistics.
- Clarifies the landscape of making decisions between discrete choices.

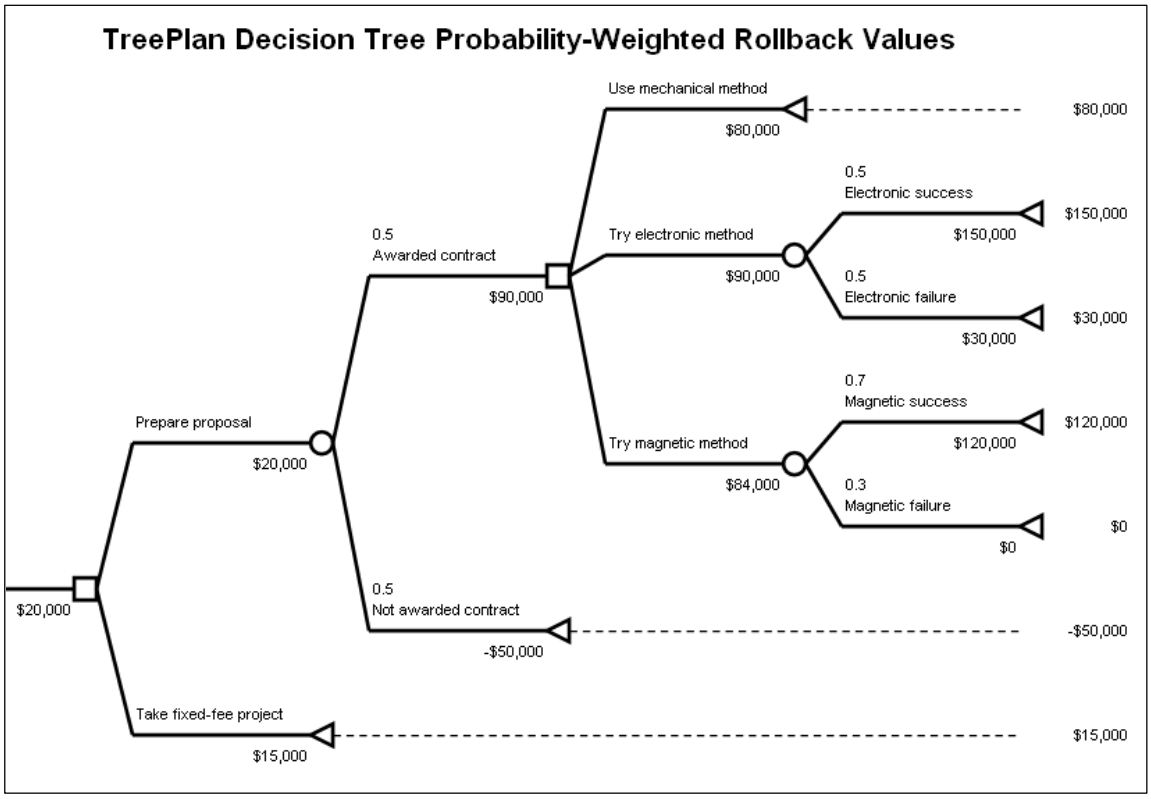
To give you a brief tour of this course, we will spend the first few modules thinking about the theory and math behind decision making. The middle modules will be spent learning how to model decisions. The final modules will be spent learning how to optimize our models and discuss their results intelligently.

The final project for this course starts early and runs as a common thread through each module, since it is broken up into milestones. For this project, you will select your own data set, create your own decision analysis research question, model it, optimize it, interpret it, and report your results.

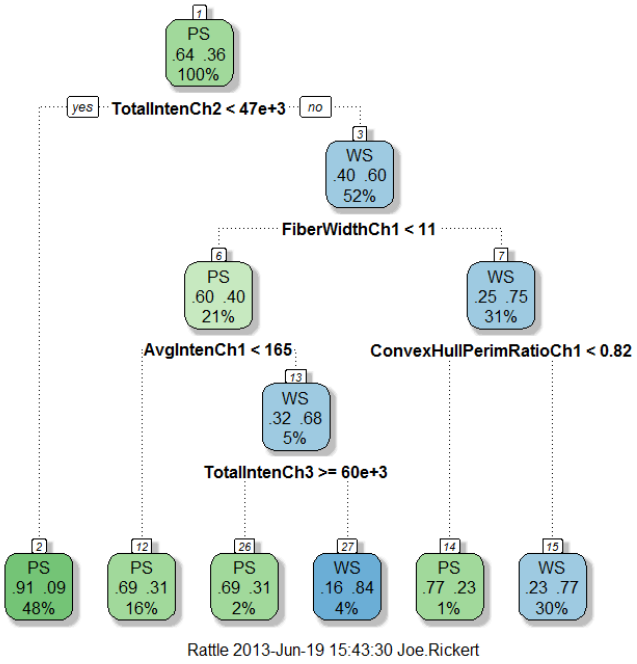
This course will give you the tools for modeling decisions, optimizing them and interpreting them by their mathematical basis and principles of decision science.

Decision Modeling

You will learn two different but related styles of modeling: top-down and bottom-up. **Top-down models** use human intuition to map the discrete elements of the model and determine their starting values. Then, the tree model is rolled back computationally to determine the optimal path. An example of a top-down model follows:



Bottom-up modeling starts with a data set that has been prepared with discrete values in categorical variables, and then a recursive partition algorithm calculates the optimal path from the raw data, followed by a plotting program that visually displays the best tree that was constructed. An example of a bottom-up model follows:



Matrix Math

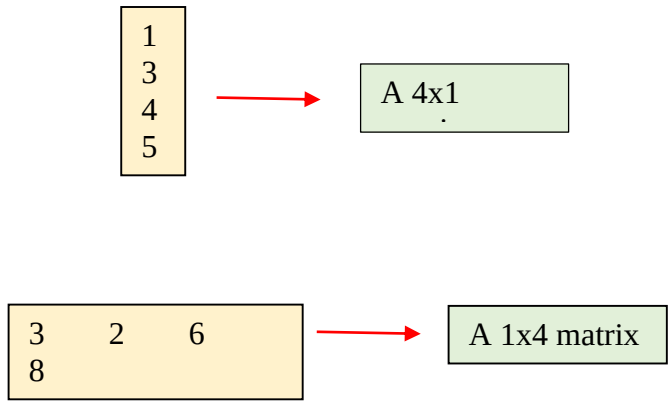
In this module we will initially focus on matrix math in R. This math is the basis of the algorithms used in computer decision tree construction. The matrix is the way the computer can define the states of the data. A matrix is a row and column design. An example of different structures is as follows:

-1	0	11
12	5	2

→ A 2x3 matrix

1	
	2
4	9
2	7
8	3

→ A 4x2 matrix



Understanding how to read a matrix is the first step. We can then perform math using the matrix. There are a variety of reasons to use a matrix and further perform math on it. For the purpose of decision trees, matrix math is mainly for scaling or applying vectors. The scalar and vector reference means that we are taking a state of variables and applying some shift that will move them in the graphical plane. Maybe in stock analysis we need to scale the returns by their trading volume weights, or in product sales we need to scale prices based on regions. In the models that recommend selections, such as with Netflix, matrices are used to overlay your movie selection to generate a “you might like” section.

In our problem set this week, we will be performing different operations on matrices in R. For the purpose of this course, we are using the problem set as a basic review. It is not expected that you become a mathematician to perform data analytics. Most tools like SAS, SPSS, and R already have functions built to handle these algorithms as you perform decision tree models. This exercise, as well as the problems found in the next three modules, will focus on performing the math behind the algorithms. For every problem set, there will be a discussion board. This is a great space for you to ask questions, share ideas, and seek support as you work through the problem sets.

This week, in preparation for our work on probability, you will calculate matrices in R. This is also a way to get familiar with how R works: its needs for data input and how R Studio windows work and give you feedback. Understanding R from day one will help you tremendously with the rest of the course. The more you learn up front about it, the smoother the rest of the term will go.