

DAT 520 Final Project Notes

This document is intended to assist you with the final project. This course introduces a decision analysis using a decision tree. There are guidelines that can assist in completing a project of this kind.

Milestone One: Choose a Data Set and Formulate Decision Analysis Research Question

In Milestone One, you are to provide an abstract describing your decision question and high level approach. Consider this abstract as the information that your peers will read and determine if they want to read more. It needs to include a statement of your decision question, what your premise is, and how the data will be used to support the analysis. The information that follows is additional notes for Milestone One:

In terms of data, the data will lend itself to a decision-making scenario with discrete outcomes. As additional information on how to select your data, first think about the course materials and what you have learned so far about the types of situations that lend themselves to decisions under uncertainty. Then, use the table of student projects as a seed to see what other people have worked on in the past.

Selecting your data set is step one, but it also overlaps step two, which is formulating a research question. Decision analysis research questions are best stated as a discrete set of choices to be analyzed. A good decision analysis research question also possesses the following hallmarks:

- It has clarity of purpose by being framed as a discrete set of choices to be analyzed.
- It is concise.
- It is appropriate and can be answered with decision analysis techniques.
- Its parts are relevant to each other.
- It has not been answered before, or the variation from existing works is great enough to make it a novel line of inquiry.
- It is open-ended enough that it may lead to new research questions.
- But it is closed enough that direct answers are possible, even if they may not be found by this round of analysis.

Here are two references that will guide you in writing a good decision analysis research question: [Writing Research Questions](#), [What Makes a Good Research Question?](#) The discussion topic is also available to ask the instructor and classmates about your research question. The guiding light, however, should be what interests you most and makes you want to investigate.

Milestone Two: Develop Decision Analysis Model

During the last milestone, you may have been thinking about different ways to create a model that explains your thoughts. By this point in the class, you have been exposed to both top-down and bottom-up modeling styles. To complete this milestone, you may have to experiment with different modeling styles. The main objective is to draft your model, explain what you did, and explain why it is the best model for your research question. Are you leaving out any variables that could strengthen your model?

Figure out the style of decision analysis modeling that you might use toward exploring your research question. At your discretion, your analysis may include some observational data analysis methods that you learned in experiences outside of this class, but the bulk of your methods need to have been taught in this class. Be careful not to let other methods overpower what you are doing here with decision

analysis. More than 75% of your project needs to be decision analysis modeling as covered in DAT 520. For example, it will not be acceptable to revert to a regression analysis to complete your final project.

That said, how do you go about creating a model? First, you need to have a viable decision analysis research question. In other words, you need a research question that analyzes a discrete set of choices. Second, you need to have at least one viable data set: it needs to contain the variables and covariates of interest. Or, if you have multiple data sets, they need to be combined. You have been learning skills in R that will assist you in this data prep phase. If R is still uncomfortable, you can always use Microsoft Excel. The data set that you end up with needs to be cleaned of errors, and ready to go for use in Rattle, or in R to create probabilities. Third, and this is the most critical, you need to decide whether you are going to engage in a bottom-up or top-down modeling style.

To decide if you are going to use a top-down model, you need to be able to create proportions from your data set that represent the decision nodes and chance nodes that fall on the path between outcomes and choices. Or, if you are going to use a bottom-up model, you need to decide how many groups should be represented in the outcome of interest. Is that outcome represented as a continuous value in the data set? If it is, then it will need to be converted into a categorical variable. Consult your references for how to do this efficiently. Also keep in mind Rattle's setting for the number of buckets. Getting the variables ready, either as proportions for top-down modeling or as categorical variables for bottom-up modeling, is the most basic aspect of the data prep that you need to do.

To draw the model, follow the guidance for either Rattle or R. You should have done enough examples as well as the assignment to know where to start. A good decision analysis model will have more than just choice and outcome, or choice one chance node and outcome. It will be multifaceted. It will consider a number of inflection points that occur in making that decision, somewhere between three and 20. Most decision trees have somewhere between three and seven levels, so you should aim for approximately that level of complexity.

This milestone is important because it should be making you curious about your research project. What is the best way to investigate this research question? Are there alternative ways to draw the same model? What happens if you include this variable? What happens if you exclude this one but include this other one? What happens if you tried both the top-down and bottom-up model? Does your research question support either approach? Does your model match up with what your research question is asking?

The next part to consider is whether the graphical representation of the model is clear or not. Are the parts labeled clearly? Are the values present? Is there anything missing? Is the optimal path obvious? Do you know what the errors are? Is the model interesting? There are a number of different considerations, but the main outcome of this milestone is to experiment to discover the best way to answer your research question.

Milestone Three: Revise and Evaluate Decision Analysis Model

In this milestone, you will perform an evaluation of your decision model and revise your decision model as needed. Evaluation examples are if you are performing a bottom-up style recursive partitioning analysis, you should report on the error rate and variable selection. You might also consider alternative variable categorizations to improve your model. If you are performing a top-down decision tree modeling exercise, what are the threshold values that cause the tree to flip? You should perform

sensitivity analysis on the critical variables in your tree and report what those sensitivity analyses are telling you. For either style of modeling, what makes your tree stronger? What breaks the model?

- If you are performing a bottom-up style recursive partitioning analysis, you should report on the error rate and variable selection, and what you did to improve them. You might also consider alternative variable categorizations to improve your model. You might consider creating different versions of the same variable with slightly different categories and invoking them selectively in Rattle. You might consider making multiple models that represent different groups of variables to explain an answer to the research question slightly differently each way. You should also report shifts in the error rate and what that means when you do different things.
- If you are performing a top-down decision tree model, where are the threshold values that cause the tree to flip? Are there any? You have learned about sensitivity analysis at this point in class, so you should be able to identify the critical values for key variables in your tree and report what the sensitivity analyses are telling you. What happens when you include certain decision nodes in your tree but exclude others? Can you draw alternative trees that still answer the research question? What happens to the proportions and the outcomes? What method are you going to use to deduce the optimal path?

Generally, for any of these decision trees, what makes your tree stronger? What breaks the model? What kinds of variables do you wish you had but do not have data for? What is the best criticism of the tree that you drew? What are its limitations? What are its strengths? You do not need to answer all of these questions exhaustively, but can use them as launching points for your writing.

Final Submission: Decision Analysis Model and Report

For your final project, you will submit your decision analysis model and report, compiling all the components used to develop the model and produce the report, as well as a leading abstract, table of contents, and in a format that addresses all of the critical elements in the instructions. Each of the prior milestones has information needed to produce the final report. Note, the final report is not just a concatenation of the milestones. It instead is a report that is created within a defined eight page limit that will include sections that detail the limitations and justification for your analysis. You should also take the time to address any ethical or legal issues that connect with your results or decisions being analyzed. Lastly, you should address the agility of your analysis and how it might be applied to future uses.