

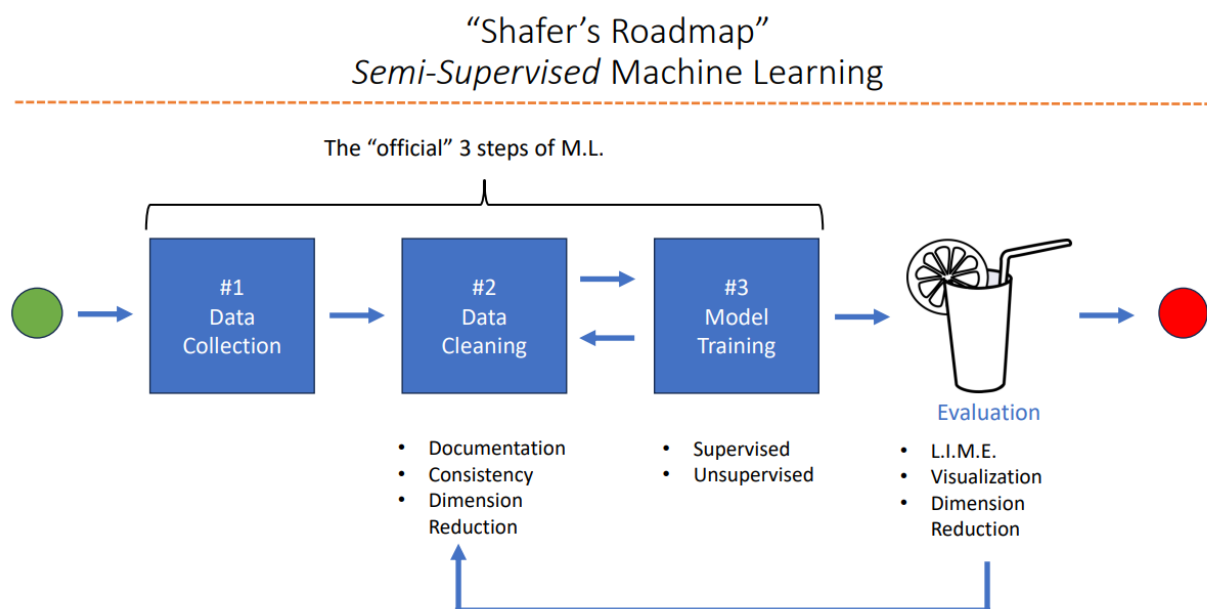
In Class Activity

Semi-Supervised Learning Roadmap

In this activity we will use a combination of unsupervised and supervised models to evaluate survey data.

The survey data is real. It was collected over three semesters here at Temple. The survey was constructed to investigate student attitudes towards “flipped classroom” instruction of a STEM topic, post COVID.

Here is version one of the semi-supervised M.L. roadmap.



We want to use appropriate models and techniques to determine answers to the following questions:

- What groupings of student responses naturally exist in the data?
- What characteristics best describe the students who do not want flipped classroom instruction?
- Is student receptiveness to flipped classroom instruction increasing or decreasing?

CONTINUED

Instructions

1. Students should visit <http://tinyurl.com/shaferaicourse> and download the files found in the **roadmap1** folder.
2. Start by examining flipped-classroom-survey.xlsx and flipped-classroom-survey.csv in Excel.
3. Our first priority will be data cleaning. As we edit flipped-classroom-survey.csv we will document our changes in the Notes tab of the xls file. Keep the spreadsheet in the xls file in its original state, in case you need to refer back to it.
4. After the data has been prepared, identify the groups of student responses naturally found in the data.

Discuss: What model should we use? Is this an example of an unsupervised model? Or a supervised model?

5. Can we, for the purposes of visualization, reduce the number of dimensions to two?

Discuss: Wouldn't it be great if I had a way to *reduce the number of dimensions* I was investigating?

6. Now revisit our data and choose a new model to identify the characteristics of students who are most likely to dislike flipped classroom instruction.
7. Now revisit our data and choose a new model to assess if student acceptance of flipped classroom instruction is trending up or down.

Note: For our purposes we will use a regression script to assess this, but this is not really the best approach. Regression should be used for continuous data. Our survey responses are categorical. It would be more proper to use a non-parametric test (like a Chi-Square test) to answer this question.

Your dear old instructor is just trying to use scripts that have already been seen previously in this class, and not turn this class into a Statistics course.

8. To document your participation in this activity, your instructor will ask you to post one of the graphics you generated, along with some of your comments regarding it. Post these to the corresponding discussion post on Canvas.

Be sure to post the graphic that your instructor requires.