

Day 4: Continued Practice of Machine Learning Activity

Lesson Plan: Summer Institutes in Computational Social Science 2024

Prepared by SICSS-Rutgers

Summary

A structured group exercise designed to practice building a classification or prediction model related to social science goals.

Learning Objectives

- Gain practice implementing machine learning processes.
- Describe strengths and weaknesses with machine learning predictions.
- Generate more potential ideas for group projects in the second week.

Activity

1. Choose an application for which [labeled data are provided](#). Examples:
 - Detecting the presence of political content in Facebook posts: “sml_practice.csv”
 - Detecting incivility in tweets: “incivility_labs.csv”
2. How might these classifications be useful for downstream social science research or practical policy applications?
 - Discuss at least one use-case of for these classifications
3. Load the data into R.
 - Locate the labeled outcome to operationalize your classification goal
 - How many positives vs. negatives do you have?
 - Discuss which performance metrics are most important in your setting given the intended use case you identified in #2. (Do you care more about false positives/false negatives?)
 - Recall, Precision, Accuracy, F1
4. Model building
 - Try to apply the supervised machine learning code from yesterday’s tutorial with at least one type of algorithm (e.g., svm) with this new dataset. Note that some of the variable names may be different for the ids, text, and outcomes. You may need to change these in the code. Together, discuss each step of the code and its purpose as you execute the code.
5. Model iteration
 - Try a new sml algorithm and/or follow some of the “improve the model” recommendations and the after-class exercise steps from yesterday’s tutorial to try to improve the performance of your model
6. Reflect on this process
 - What model/recipe/specification did you like best? Summarize its performance and how you tested it
 - What are strengths and weaknesses of the supervised machine learning approach?
 - Would you feel confident enough in your model to apply it to unlabeled data?
 - Are there any ethical considerations of using supervised machine learning in social science research or policy?