Dataset: [Estimates of lake nitrogen, phosphorus, and chlorophyll-a concentrations to characterize harmful algal bloom risk across the United States](https://catalog.data.gov/dataset/estimates-of-lake-nitrogen-phosphorus-and-chlorophyll-a-concentrations-to-characterize-har)

What do we do? Predict the concentration of chlorophyll-a (which are associated with harmful algal bloom, hypoxia, and eutrophication effects in US lakes) and which can help to evaluate what water bodies are in danger of those harmful conditions.

ML 1- Yilmaz

2024-2025

Name of the group members (max 2 people): Jacob Dipasupil, Petr Kisselev

If you are working alone, please state it clearly.

If you are working in groups, only one group member should submit the proposal.

Must haves:

* Link to dataset
* Information about your data
  + Meaning of attributes
  + Dimension
  + Number of instances
  + How many missing values
  + Is it uniform? Skewed?
  + What are the class distributions?
* What are you classifying/predicting?
* How will this be useful?
* Plans for preprocessing

Don’t just insert your answers in the above, make it look like you're writing the portions of the final report. See the final report sample I posted at Schoology.

As soon as you are ready with your proposal, please see me, so that I can approve it.

Also, please add your project to the spreadsheet I will provide under Q1 Project Folder.