

Final Project proposal

The final exam (30% of final grade) will be conducted in the form of a final project plus an in-class short quiz on March 3rd with project presentation. The aim of the final project is to apply what we learn in class to a real-world problem. One of the goals of this project is to add a credential to your CV of your data analysis skills.

The project topic is up to you. You can get some idea of what others do for a final project at <http://cs229.stanford.edu/projects.html>

To get started:

o Find 1-3 persons to form a group (ideally 2)

Write a proposal with the following sections and submit it to dropbox.cse.sc.edu

o Title of the project

o Team members of the project

o Sections:

Abstract: what is the problem, why it is interesting

Background: details and related work

Input data: where do you get the data, data format, example of the data entries

output data: what do you predict or do with the data? Regression problem?

Classification?

Evaluation: how you will evaluate your machine learning model performance?

Notes:

- State the problem clearly and explain why it is interesting, significant, or publishable
- A sample proposal and final project report are available in the dropbox Finalproject folder.
- Upload your proposal to <https://dropbox.cse.sc.edu>
- Your proposal will be graded and needs to be approved by the instructor. If approved, you can start working on it. Otherwise, you need to talk to your instructor or modify it based on the feedback.

Currently, due to the dominance of deep learning in application fields, a good deep learning project may make your CV shine.

To find a topic for your final project, you can

Step1: first pick a dataset from:

- (1) <https://archive-beta.ics.uci.edu/> try to pick the latest datasets
- (2) <https://www.kaggle.com/datasets?fileType=csv> kaggle competition datasets
- (3) <https://paperswithcode.com/datasets> image/text/numerical data
- (4) Work on materials property prediction problem using a different algorithm.

Check here to figure out how to get features for materials:

https://github.com/hackingmaterials/matminer_examples/tree/main/matminer_examples/machine_learning-nb

43 datasets description:

https://hackingmaterials.lbl.gov/matminer/dataset_summary.html

Step2: determine the algorithm to use, use a deep learning algorithm or a few basic scikit-learn machine learning algorithms such as decision tree, random forest, Gaussian processes.

Step3: develop the code, and run experiments, tune hyper-parameters, generate the results, and write the final report. (check the sample project_report.pdf as reference) or more reports <http://cs229.stanford.edu/proj2021spr/>

Step4: prepare the slides for final exam day presentation of your project, submit your final report, data, and code to <http://dropbox.cse.sc.edu>