

CSCI 183

Final Project

Your class project is an opportunity for you to explore an interesting data science problem of your choice in the context of a real-world data set. Below, you will find some project ideas, but the best idea would be to combine Data Science with problems in your own research area of interest.

In this project you will want to apply what you have learned in class and report your findings.

Select a topic you would like to work on and look for interesting project ideas related to that. Feel free to design this project based on your group's interests and do not restrict yourself just to the topics discussed in class. :)

For example, [Linear Regression Project Ideas](#) could be a starting point for your project!

SAMPLE PROJECT:

Sensor network data

Using this 54-node sensor network deployment, we collected temperature, humidity, and light data, along with the voltage level of the batteries at each node. The data was collected every 30 seconds, starting around 1 am on February 28th, 2004.

<http://www-2.cs.cmu.edu/~gustrin/Research/Data/>

This is a "real" dataset, with lots of missing data, noise, and failed sensors giving outlier values, especially when battery levels are low.

Project ideas:

1. Compare regression algorithms [A survey paper]
2. Detect failed sensors

Similarly you could explore other Machine Learning algorithms for other projects.

In a nutshell

- **Proposal (Due 2/14/24).** 10% of your project grade.
- **Project report and Code (Due 3/17/24).** 90% of your project grade.

Sample Project Reports:-

<http://cs229.stanford.edu/projects2011.html>

Project proposal format:

Proposals should be three pages maximum. Include the following information:

- Project title
- Data set
- Project idea. This should be approximately two paragraphs.
- The software you will need.
- Papers to read. Include 1-3 relevant papers. You will probably want to read at least one of them before submitting your proposal.
- Teammate: will you have a teammate? If so, whom? The maximum team size is three students. Please sign up here: [Group Sign-Up](#)

Your project will be worth 10% of your final class grade, and will have two final deliverables:

1. A report of at least 5 pages, explaining the problem, approaches, results and conclusion.(NO LATE SUBMISSION ACCEPTED since I need to get your grades in)
Make sure it is not double spaced and you should not just fill up your report with images.
2. Source code of your project.

Final Report Grading Scheme

The grades for the final project report will be loosely based on the following criteria:

Background:

1. Does the final report describe the setting and reference to related research?
2. Does the final report describe the data that you are working with and how it was derived?

Design:

1. Does the final report have a concrete well defined experimental design describing the learning task?
2. Does the final report describe which features and models are used and why they were chosen?

Implementation:

1. Did you implement the required appropriate techniques?
2. Does the final report describe what was implemented and what tools were used?

Results:

1. Does the final report have quantitative results from learning or experimenting with your data?
2. Does the final report have data analysis using visualization tools? Does the report have findings for the same?
3. Does the final report have results evaluating the learning of your model? (i.e., learning curves, precision-recall, training/testing errors, ...)
4. Does the final report make effective use of graphs that are appropriately labeled and properly described in the document?

Interpretation:

1. Does the final report attempt to interpret the results?
2. Does the interpretation correctly use concepts to justify the results?

Overall Presentation: Is the overall presentation coherent and well organized?