

Final Project Information

It's time to begin thinking about your final project! You can choose to do essentially anything in the realm of data mining as long as I approve it in advance. For example, you can build on material we have done in class, or you can use it as an opportunity to learn about subject matter we don't have time to cover.

I strongly encourage you to work with a partner for this project. If you would like a partner, but have not been able to find one, please email me and I can help facilitate. You may also want to use Piazza as a forum for finding people to work with or to look for others with similar interests for the project.

Your Task: Learn about and implement a new algorithm

Pick an algorithm that extends something we have discussed in class, or that relates to another area of data mining that we haven't covered. You can get ideas from the textbook, another book, or even a research paper. For this project, I'll expect that you'll turn in the following items:

- You working code and a detailed README.
- A small sample dataset that the code can be run on.
- A short paper (≤ 5 pages including figures) that describes the approach you learned about and what types of problems it may be useful for solving. You should also include some analysis of how this approach works. For example, this could include plots that shows how the runtime changes with more data, or how accuracy is affected by a particular parameter, etc. You should also include some analysis of what you learned and tried along the way including anything you found surprising or particularly interesting. Finally, your paper should also include any citations to materials you used when learning about the approach. This "short paper" is essentially a more fleshed out version of the type of analysis you have been doing on many of the homework assignments.

Deliverables

There are three main deliverables for your project:

1. An initial project proposal due by 10pm on **Monday May 8**.
2. A project presentation to occur in class on either: **Friday May 26, Monday May 29** or **Wednesday May 31**.
3. The completed final project is due by 10pm on **Wednesday May 31**.

The remainder of this document is dedicated to describing the different pieces of of the final project and what you will be expected to hand in for each.

Project Proposal

Your project proposal is due by 10pm on Monday May 8. Please come by office hours or setup an appointment if you'd like to discuss your project before submitting your proposal. Submit to Moodle a proposal that includes the following pieces of information:

- A description of what you plan to do. You should provide enough details to convince me that what you plan to do is reasonable given the time that you have. As a guideline, I expect this to be ~ 1 full page in length.
- The names of the team members.
- At least three references you plan to use for researching the project (textbooks, research papers, etc.).
- Any strong preferences you have for the presentation dates. I can't guarantee that I'll be able to give you what you request.

Project Presentation

Each group will give a short (probably 10 minutes - depending on the number of groups) presentation to the class on your project on either Friday May 26, Monday May 29 or Wednesday May 31. This is a chance for you share with others what you have learned, but also to practice presenting and discussing technical material (not an easy task!). Because the presentations will take place over several class periods, you may end up presenting before your project is fully completed. But you should certainly have enough done to make a coherent and interesting presentation.

Final Project

You will need to submit your final project by 10pm on Wednesday May 31. If you miss this deadline, I am forbidden by college policy to extend it. You must contact your class dean if a situation arises for which you need an extension. You must submit to Moodle the following:

Useful Resources

Here are some sources of “nicely curated” datasets that may be useful.

- [UCI Machine Learning Repository](#)
- [UCI KDD Archive](#)
- [KDnuggets](#)

Here are some textbooks or other sources of ideas for algorithms.

- A [bunch of ideas](#) from the class that our textbook comes from.
- A bunch of [online data mining textbooks](#).