

Final Project

Proposals: Tuesday, November 1 by 11:59:59pm
Presentations: Tuesday, November 29 - Thursday, December 1
Final Reports: Sunday, December 13 by 11:59:59pm

Out October 25, 2016

1 OVERVIEW

The final project is meant to provide you and your team with the opportunity to explore a specific data science problem in-depth, using the tools and techniques you've acquired and honed in this class.

The core constituents of the final project are:

- A well-defined data science question (classifying objects in images? categorizing text documents? identifying advertisers in web crawler data? detecting anomalies in network traffic? constructing product recommendations?).
- A public dataset.
- A proposed methodology for analyzing the dataset to answer the question.

The question you choose to answer should be along the lines of the project goals you've had this semester. **It should be challenging but solvable.** You are welcome to use any publicly available dataset (or any dataset you're willing to make at least partially public), and either improve upon existing solutions or design your own approach to address a completely novel question.

There are three primary deliverables in the final project (the deadlines for each are listed at the top of this document).

1. **Project Proposal:** Each team will submit a 2-page (maximum) proposal that address the three core constituents above. It should also include the names of your teammates and as much detail as you have regarding the methods and tools you'll use to answer the question.
2. **Project Presentation:** Each team will be delivering a 25-minute presentation (20 minute talk, 5 minutes for questions) on their work. These talks should summarize your question and the dataset, and spend a good amount of time on the analysis performed and results obtained.
3. **Project Write-up:** Each team will submit an 8-page (maximum) write-up in NIPS format ([template files available here](#)) detailing out their work. Grammar and punctuation count, so make sure you have someone proofread your document before final submission.

2 GUIDELINES

Base your work on sound theory. This will likely involve some kind of scholarly citation, whether you're building off existing code or not. You need to be able to justify why your proposed solution will work / work better than existing solutions, and the supporting theory will be the core of your justification.

All frameworks are available. You can use Spark, DL4J, Flink, Hadoop, Storm, Caffe, TensorFlow, mxnet, mlpack, nnpack, Paddle, Xgboost, DSSTNE, Thunder, or any other open source framework you'd like.

Use GitHub. Just like with all the projects you've done so far, please make all the code available on GitHub. You can create a private repository for development, but please keep this under the **eds-uga** organization; that way, I can see your progress and provide feedback.

Software Engineering. Employ good software engineering techniques. This includes, but is not limited to

- Use GitHub milestones and tickets to organize your development efforts.
- Develop wiki pages for your project's documentation.
- Split the work amongst your teammates, and use good commit messages to clearly describe what you're working on.
- Branching and merging in git is a huge plus with dividing and conquering the workload.

- Include `README` and `CONTRIBUTORS` to clearly spell out what your project is, what it does, and who made it.
- Use TravisCI to continually build and test your master branch every time a commit is made (they've given us free organizational integration; just ask me!).
- Use good coding practices and program design.
- Include a `LICENSE` file.

3 GRADING

Per the course syllabus / website, the entirety of the final project is worth 25% of your grade. That 25% is split among the components of the project as follows:

- **Proposal:** 15%. This is the 2-page document, pending approval from me.
- **Presentation:** 35%. This is the 25-minute presentation, and covers the quality of the talk (easy to understand, good flow of ideas and discussion, well-organized structure, clean slides).
- **Write-up:** 50%. This is the 8-page paper, but also covers your GitHub project repository (code, commits, milestones, issues, documentation, etc). In terms of the paper, this covers both the structure and organization (sections, headers, grammar, spelling, syntax) as well as the quality of the project itself (question, experimental design, framework, justifications, discussion, conclusion).