Machine Learning | CSCI 0451 S25   🔍

Project Blog Post

2025-05-12

Wow! You completed an awesome project with your group. It's time for you to tell the world (and especially me) about it on your blog.

Your project blog post is the authoritative description of what you achieved and what you learned. It should describe *all* the achievement, effort, and learning that you want us to factor in to your final grade in the course. Your blog post should have eight sections, which are described below.

You can write the first seven sections of your blog post as a group and all post them on your separate blogs. If you take this path, you should just make sure to note it. The final section should be completed individually.

1 Abstract

Your abstract is a one-paragraph summary of the problem you addressed, the approach(es) that you used to address it, and the big-picture results that you obtained.

At the end of your abstract, **please include a link to the GitHub repository** that houses your project code and other deliverables.

2 Introduction

You may be able to recycle some content from your project proposal for your introduction.

Your introduction should describe the big-picture problem that you aimed to address in your project. What's the problem to be solved, and why is it important? Who has tried solving this problem already, and what did they do? I would expect most introductions to reference no fewer than 2 scholarly studies that attempted similar tasks, although 5 is probably a better target.

When citing scholarly studies in a blog post, please use Quarto citations. For a quick overview, see the appendix on references in Quarto.

3 Values Statement

Your values statement should be a few paragraphs that address the following set of questions:

- Who are the potential **users** of your project? Who, *other than your users*, might still be **affected** by your project?
- **Who benefits** from technology that solves the problem you address?
- **Who could be harmed** from technology that solves the problem you well address?
- What is **your personal** reason for working on this problem?
- Based on your reflection, **would the world be a more equitable, just, joyful, peaceful, or sustainable place** based on the technology that you implemented?

4 Materials and Methods

Your Materials and Methods section should describe:

Your Data

Include some discussion of where it came from, who collected it (include a citation), how it was collected, and what each row represents (a person, an environmental event, a body of text, etc) Please also include a discussion of potential *limitations* in the data: who or what is represented, and who or what *isn't*?

In structuring your description of the data, I encourage you to address many of the questions outlined in (**gebru2021datasheets?**), although it is not necessary for you to write a complete data sheet for your data set.

Your Approach

This is the primary section where you should describe *what you did*. Carefully describe:

- What features of your data you used as predictors for your models, and what features (if any) you used as targets.
- Whether you subset your data in any way, and for what reasons.
- What model(s) you used trained on your data, and how you chose them.
- How you trained your models, and on what hardware.
- How you evaluated your models (loss, accuracy, etc), and the size of your test set.
- If you performed an audit for bias, how you approached this and what metrics you used.

5 Results

This is the section in which you describe the main findings or achievements of your model. You can report things like accuracies on train/test data, loss scores, comparisons to previous models, etc. To compare a small set of numbers, tables are fine, but more complex phenomena should be illustrated with figures. Both figures and tables should include appropriate captions, axis labels, legends, and another professional annotations. It's fine for your figures to either be constructed manually or as computational outputs (e.g. from Pandas).

Please remember: **your results do not speak for themselves**. While figures and tables are highly effective forms of communication, your prose is necessary to tell your story.

In general, I expect that your results section may **include some code**, but that most of the code you wrote will be housed in your GitHub repository and not part of the blog post.

If the amount of code you wrote fits reasonably in its entirety into a single blog post, it may be that your project was insufficiently ambitious in its scope.

6 Concluding Discussion

Your conclusion is the right time to assess:

- In what ways did our project work?
- Did we meet the goals that we set at the beginning of the project?
- How do our results compare to the results of others who have also studied similar problems?
- If we had more time, data, or computational resources, what might we do differently in order to improve further?

7 Group Contributions Statement

When writing your group contributions statement, please keep in mind that everyone's contributions are visible in the commit history of your GitHub repository.

In your group contributions statement, please include a short paragraph for each group member describing how they contributed to the project:

- Who worked on which parts of the source code?
- Who performed or visualized which experiments?
- Who led the writing of which parts of the blog post?
- Etc.

8 Personal Reflection

This is the only section that you are required to write individually and not with your project group.

At the very end of your blog post, in a few paragraphs, respond to the following questions:

- What did you *learn* from the process of researching, implementing, and communicating about your project?
- How do you feel about what you achieved? Did meet your initial goals? Did you exceed them or fall short? In what ways?
- In what ways will you carry the experience of working on this project into your next courses, career stages, or personal life?

Appendix: References in Quarto

To manage references in Quarto, you need to create a `.bib` file (you can call it `refs.bib`). This file should live in the same directory as your blog post. Your `.bib` file is essentially a database of document information. Here's an example of a a `refs.bib` file:

```
@article{hardt2021patterns,
  title  = {Patterns, predictions, and actions: A story about machine learning},
  author = {Hardt, Moritz and Recht, Benjamin},
  journal= {arXiv preprint arXiv:2102.05242},
  year   = {2021}
}

@article{kearns1994toward,
  title     = {Toward Efficient Agnostic Learning},
  author    = {Kearns, Michael J and Schapire, Robert E and Sellie, Linda M},
  journal   = {Machine Learning},
  volume    = {17},
  pages     = {115--141},
  year      = {1994},
  publisher = {Citeseer}
}

@misc{narayanan2022limits,
  author       = {Narayanan, Arvind},
  howpublished = {Speech},
  title        = {The limits of the quantitative approach to discrimination},
  year         = {2022}
}
```

The simplest way to get entries for your references is to look them up on Google Scholar.

1. Search for the document you want.
2. Click the "Cite" link underneath and choose "Bibtex" from the options at the bottom.
3. Copy and paste the contents of the new page to your `refs.bib` file.

Once you've assembled your references, add the following line to your document metadata (the stuff in the top cell of your Jupyter notebook)

```
bibliography: refs.bib
```

Once you've followed these steps, you're ready to cite! You can reference your documents using the @ symbol and their bibliographic key, which is the first entry for each document in the `refs.bib` file. For example, typing

`@hardt2021patterns`

results in the reference

(**hardt2021patterns?**)

as well as an entry in the "References" section at the end of your blog post.

For more on how to handle citations in Quarto, check the Quarto documentation.

© Phil Chodrow, 2025