SI 650 / EECS 549
Project Blog Post Guidelines

Due: Wednesday December 14, 11:59pm

Introduction

Communicating technical concepts to a general audience is an important and life-long skill. This type of communication can come in many ways, from in-person sharing, to poster sessions, to talks to a general audience (e.g., at a library), or through public blogging. For your final project, instead of presenting your work to a general audience as a talk, you will write up a blog post describing your project, how you did it, and its findings.

The final project "presentation" has the following (very practical) learning goals:

1. Learn how to communicate your results to a general audience
2. Become familiar with blogging technology and style usage
3. Learn how to explain basic programming and technical materials to peers

In summary, we want to write a public blog post that shows off your skills and expertise and provides some technical insight into how to do this kind of project. You have likely seen similar blog posts before when you yourself were looking around for ideas for this or another class. This time, you'll be the one writing such a blog post, in hopes of inspiring and teaching others.

Your Blog Post

A key part of your blog post is trying to communicate to your intended audience. Your blog post should be aimed at two related audiences: (1) an educated reader who is low-to-moderately tech savvy. Think of someone who knows how to program decently but hasn't done extensive work and (2) someone who is interested in the topic and appreciates the technical approach, but not necessarily understand it yet. To aim for this audience, try to explain your steps and motivations, as you would to a peer or junior colleague in your program, or to someone who is interested in your topic but doesn't know how to develop software.

Your blog post should include at a minimum the following structure:

1. Introduction -- where you talk about what the problem is and why it's useful or interesting to work on it. Appeal to the person's curiosity and highlight what they'll learn from your post
    a. You should being your introduction with a tl;dr (too long didn't read) which is a 1-2 sentence summary of the whole blog post for a reader

2. Data -- what is the data you used, where did you get it (URLs are great here). If you had to do any data cleaning here, describe it here. This is a great place to show an example.
    a. If you did any preprocessing or cleaning, include a few snippets of code here.
3. Methods -- here's where you explain how you solved your problem.
    a. You should definitely include at least some code here that shows exactly how you did things. This is the kind of code that others will find useful when solving or re-implementing your approach
4. Results and Discussions -- did it work? describe your results and explain what they mean to someone who might not know how to interpret the number.
    a. Figures and tables are great here.
5. What's Next -- Usually, there's a few next steps you think could be useful. Inspire the next person by suggesting things they could try.

For each of these sections, you are encouraged to include things that will help readers follow along at an intuitive level and benefit from your work. Example include:

- snippets of code that show exactly how something was done
    - However: Avoid including long blocks of code!
- links to the software library or github repo for things you used
- examples of data to help illustrate your problem
- figures and tables throughout the post (figures > tables in blog posts)
- key equations (when they make sense)

You are welcomed to include additional sections and to rename these sections as you see fit. The five points mostly relate to content we want to see there. Your main goal should be to write in a way that is engaging to the reader.

Your blog post is expected to be at least 500 words. If you're using Medium or a similar blog platform, this should translate into at least a 5 minute read.

You can borrow *some* of the text from your technical report for the blog post. However, be sure that this text is accessible to your intended audience. The audience for your technical reports is experts in the field; the audience for the blog post is the general audience described above.

Examples of Good Blog Posts

Good blog posts from SI650 F20

- Customer SatisfactionPrediction for Grocery E-Commerce Customer Reviews
- Prost! A Deep Neural Networks-Based Beer Recommender System
- Tour City Search Engine: Locate Your Favorite Destination
- How to build an Emotion-Based Song Retrieval Model in Python
- Personalized Recommendation on Sephora using Neural Collaborative Filtering

Good blog posts from SI630 W20:

- Who is the angriest Avenger? An NLP approach beyond sentiment analysis
- Deus Ex Machina: Fine Tuning GPT2 to Generate Christian Song Lyrics
- Using natural language processing techniques to classify mathematics teachers' responses to representations of practice
- Assessing the Funniness of Edited News Headlines
- Yer A Wizard, NLP: Using Neural Networks to Determine Character Depth in Harry Potter
- Multiple-Choice Question Answering Through Semantic Representation Space
- Using Tensor Networks to perform Visual Question Answering tasks
- Auto-highlighter: extractive text summarization with sequence-to-sequence model

Good general data science blog posts:

- https://medium.com/glose-team/how-to-evaluate-text-readability-with-nlp-9c04bd3f46a2
- https://medium.com/analytics-vidhya/part-of-speech-tagging-what-when-why-and-how-9d250e634df6
- https://towardsdatascience.com/fine-grained-sentiment-analysis-in-python-part-1-2697bb111ed4
- https://towardsdatascience.com/embeddings-with-word2vec-in-non-nlp-contexts-details-e879b093d34d
- https://medium.com/analytics-vidhya/using-machine-learning-to-create-music-6581f1048e24
- http://www.matthewjockers.net/2015/02/02/syuzhet/
- https://medium.com/@s.aakash3431/lipstick-on-a-pig-existing-debiasing-methods-simply-cover-up-systematic-gender-biases-in-word-65459d5388fb
- https://medium.com/@uwdata/errudite-55d5fbf3232e
- https://medium.com/analytics-vidhya/using-machine-learning-to-create-music-6581f1048e24
- good but too short on the analysis: https://towardsdatascience.com/a-machine-learning-approach-to-author-identification-of-horror-novels-from-text-snippets-3f1ef5dba634
- good but light on the methods (links to its code, which is nice): https://medium.com/data-story-teller/nlp-harry-potter-and-the-philosophers-stone-6e393d46b0ff

What to submit

On canvas, submit (1) a link to your blog post and (2) a PDF copy of the web page

Important caveat: If your project involves on-going research that can't be shared publicly yet (e.g., due to double-blind review) or involves sensitive data (e.g., hospital records), you should

not post it publicly, and can instead only send us a PDF copy of the post (and keep it offline). If you think you're in this category, please contact us before doing it, as the expectation is everything should be public.