

DS 495 Data Science Capstone

Proposal Guidelines

Due 11:59 PM on the scheduled deadline

1. Contents

Your proposal should cover the following topics. Many of them will be refined and deepened for Milestone 1.

- **Abstract:** meant to be a snapshot of the entire paper (word count < 300). It should contain a brief description of the problem statement, your project motivation, your potential contributions to the field and the expected (experiment) results.
- **Introduction:**
 - Length: $\frac{3}{4}$ of a page or an entire page.
 - A very *brief background* of the proposed topic (2-3 sentences).
 - The *existing problems* in this topic area (must be clearly stated).
 - The *existing solutions* for this topic. What are the *shortcomings* of the existing solutions (name 2-3 aspects)? This serves as a summarization of the next Section – Background.
 - What is *your solution* to solve this problem that is different from the existing ones? Most importantly, to differentiate your project from others, you must list 2-3 aspects that are uniquely designed by you.
 - What are the *expected results* from your solution?
 - A short paragraph to list the organization (section names, etc.) for the rest of the paper.
- **Background:**
 - Length: $> \frac{1}{2}$ of a page.
 - A literature review is an important part of any research. Find at least 3 most related literatures and answer the following questions for each.
 - Who has worked in this area in the past?
 - What achievements have they made?
 - What have been the major stumbling blocks (ideally, this is where you can improve)?

- **Design:**
 - Length: 1 – 1 ½ pages.
 - You need to draw some block diagrams to illustrate your overall design process. For example, if it is a software, you may want to draft the flowchart; if it is a data analytics pipeline system, you may want to draw the connectivity among components.
 - Where can you find the dataset for your project? How would you preprocess the dataset to fit in your analytical need?
 - What kind of models you will need to implement? For traditional machine learning models, what features do you need to extract and why? For deep learning models, what neural network architecture would you try and how?
 - If the dataset is too large to be handled by a single computer, or if you anticipate the model training would take very long time, would you use distributed computing systems or hardware accelerators like GPU and how?
 - In what use cases would your models be useful and how to use them?
- **Experiments:**
 - Length: > ½ of a page.
 - What experiments do you want to conduct?
 - List the possible data points you want to visualize. You can even hand draw some figures to illustrate your idea.
 - What are the metrics you need to collect to evaluate the effectiveness of your design, and what are the possible results you will get?
- **Timeline:** When do you expect to have what done? Be thoughtful and realistic. Use Gantt chart.
- **Reference:** this is a separate section by the end to place all the literatures you cited in the Section Background or others. The reference must be in the ACM reference format. E.g.:

[1] P. Li, X. Wu, Y. Ran and Y. Luo, "Designing Virtual Network Functions for 100 GbE Network Using Multicore Processors," *2017 ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS)*, Beijing, 2017, pp. 49-59. doi: 10.1109/ANCS.2017.15

2. Format

- You must have *more than* 4 pages and *less than* 8 pages of contents with the ACM double column proceedings template. See Section 3. Grading.

- For LaTeX users, if you want to use the online services such as Overleaf, you can find the format here: <https://www.overleaf.com/gallery/tagged/acm-official#.WOUOk2e1taQ>
- For Microsoft Word users, download the word template from the link on the ACM website: <https://www.acm.org/binaries/content/assets/publications/article-templates/pubform.docx>

3. Grading

Grading of the proposal will be based on the topics discussed above:

Topic	Percentage
Abstract	10%
Introduction	15%
Background	10%
Design	15%
Experiments	10%
Timeline	10%
Reference	10%
ACM Format	10%
Length	10%