

CS6220 - Data Mining Techniques

Final Project Information

Facilitator/Instructor: Sara Arunagiri

Email: s.arunagiri@northeastern.edu

Final Project FAQ

(This document is meant to help you plan your schedules ahead of time. Based on many other factors including student requirements, the exact dates might change as the course progresses and you still need to pay close attention to the deadlines specified in the Assignment Section)

*When is **the final** project report due?*

Friday the 12th Aug 2022.

What is a Project Proposal?

The project proposal is your plan of action, it has more specific information than the previous step in the collaborative activity where you communicated your ideas. The proposal must consist of one to three paragraphs clearly outlining -

- o Your name and your project partners' names
- o what you plan to do,
- o how you plan to approach it, and
- o what your final deliverable will be.

You will be asked to submit a proposal during the Week of 26th June.

Can I change this proposal based on information gathered later?

Yes, you may change if there's a good reason to.

What's a project report draft?

You will be asked to submit the draft round Week 10 of this course. At about this time, you would have just started to give full-attention to the project work, so we don't expect you to have any substantial content in this draft and certainly not the final results. We ask you for a draft to ensure that you get started with writing the report. You do not have to do a lot - it can be just

the section headings and any content you created for the sections. If you have any exploratory analysis - you may include it.

Since this is not expected to be a publication quality paper, you don't have to follow the IEEE or ACM or other format restrictions. You just need to make it good and readable, for example you may choose a font like "Times New Roman" and about 12 pt size. It does not need to be double column. You may choose any style for references, but stay consistent for all references in your document.

How much of the final score is assigned to this project?

45% of the final score is for the project work – 30 for the report and 15 for the presentation.

How many pages should the final report have?

Typically, 5-15 pages, there's no length restriction.

I have not done any formal research before, how do I start?

There are three important steps in this task.

Step 1. Identify an interesting/intriguing/burning question to answer and a dataset that could help you answer it.

Step 2. Conduct the research and find answer(s) to the identified research question using Data Mining techniques learnt so far in this class.

Step 3. Write a report narrating not just your methods and findings but also the story and inspiring ideas related to the research you conducted. In short, this is like story writing but in a scientific style and methodology.

Why are we spending so much time writing this report? Isn't this a data mining course?

Yes, introducing you to data mining techniques is the main goal of this class. However, in real world, in addition to performing the data analysis you need to be able to write and communicate the results, to be able to do so is of utmost importance to be a successful analyst. You may be required to communicate your results to a lay audience, stake holders, decision makers, administrators, a scientific community, or a gathering of people in a forum. Preparing the report and presentation gives you a glimpse of the process.

What's the format of the final report?

Although you are not expected to write a full-fledged research paper, this exercise helps you understand the structure of the components thereof. The project report might have the following seven main sections -

- a. Abstract - This is a brief summary not exceeding 500 - 750 words describing the problem, the solution method, and the results.

- b. Introduction – Approximately one to two pages long and contains
- (i) Statement of the problem you are trying to solve
 - (ii) Why is it important to solve this problem?
 - (iii) Background information and bit of literature survey to present what's already known about this problem.
- c. Methodology - What's your solution, how do you propose to solve the identified problem?
- d. Code - Brief explanation of the code.
- e. Results – here you just state the results as you see them. E.g., In a fictitious study you were studying a data collected over a period of 50 years of height of children aged 12 years. You can state the trends shown in the data, like in all countries the average height of 12-year-olds seem to be increasing by 3 mm every decade, except in countries C and D. In these two countries it seemed to decrease by 2 mm in the 1980s.
- f. Discussion - interpretation of the results - what do the numbers really mean, do they make sense. E.g., In the fictitious study we used to illustrate the results section, Does it make sense that in countries C and D the average height decreased. Probably, there was a famine in Country C and Country D had an epidemic that affected the health of children in the 1980s.
- g. Future Work - Is this study conclusive or does it lead to some future work?
- h. Conclusion - what conclusions can you draw.
- i. References – Research papers, articles, and Internet resources referred to in the rest of the report.
- j. Annex-A – This Annex has information about contribution of each individual in the project group.

Can we work in groups for our final project?

Yes, you are strongly encouraged to work in groups. How you split the work between you is totally up to you. But you need to -

- Recommended group size is between 3 to 5.
- mention all names in the proposal,
- make sure that all of you participate in all stages of the project and each of you is prepared to make a presentation on request,
- all of you submit a copy of all documents asked of you, or one person in the group submits it on behalf of everybody else.
- be aware that the quality of project needs to reflect the size of the group. If you have more members, try to do more.

What are some important stages of the project?

| Approximate Date/Module | Task |
|---|--|
| Sometime during the Week of 30 th May. | Each of you publish a brief description of your project idea in the discussion forum. You give feedback on at least two of your classmates' ideas. |
| Sometime during the Week of 12 th Jun. | From reading your classmates' ideas and interacting with them on the discussion forums, self-select groups of 3 or 4. |
| Sunday 3 rd July | Submit a project proposal. |
| Sunday 24 th July | Submit a draft of the project report. |
| Friday the 12 th Aug. | Submit (a) Video presentation of your project (b) Code and the final project report. In addition, upload a recorded presentation in the discussion forum for your classmates to comment; details will be provided later. If requested, be prepared for each person in the group to make the presentation in a video conferencing session. |
| Monday the 15 th Aug. | On the discussion forum, provide your feedback for at least 3 of your classmates' project presentations. |

Do I follow IEEE or ACM format in writing the report?

Since this is not expected to be a publication quality paper, you don't have to follow the IEEE or ACM or other format restrictions. You just need to make it good and readable, for example you may choose a font like "Times New Roman" and about 12 pt size. It does not need to be double column. For references, it may be convenient for you to use one of these styles for references, you may choose any but stay consistent for all references in your document.