DSCI550 Data Science at Scale

Spring 2024 Project

(25% of total grade)

Your class project is an opportunity for you to explore an interesting problem of your choice in the context of a real-world big data application. We will provide some project ideas, but the best idea would be to combine the topics of this course with problems in your own interested area.

This project is to practice what we have studied in the class using a real-world dataset. You are supposed; 1) to survey existing real world datasets and select one in your interested area, 2) *define your own problem*, and 3) program to solve the problem using the chosen dataset.

- Projects will be done in a team of 3-4 students.
- Team members are responsible for dividing up the work equally and making sure that each member contributes. You select a team leader who will be responsible for all submissions. Please mark the team leader's name when you report your team formation.
- If you are having troubles, contact the instructor.

Your project will be worth **25%** of your total class grade, and will have the following deliverables:

- Report your team formation by **Jan. 25, 2024**. (names, emails, team leader)
- Project Proposal Report (20% of project)
    - Due on **Feb. 6, 11:59 PM**
    - 2 page report
    - Submit a PDF file with the name, Pre-proj-"team#".pdf. **Only one submission by the team leader.**
    - Late penalty: 30% deduction each day
- Project Progress Report (15% of project)
    - Due on **Mar. 8, 11:59 PM**
    - 4 page report and presentation slides
    - Submit a single zip file with the name, Prog-proj-"team#".
    - Late penalty: 30% deduction each day
- Final Report (65% of project)
    - Due on **May 1, 11:59 PM**
    - Max 8 page report + codes
    - Submit a single zip file with the name, Final-proj-"team#", which include report and codes.
    - Late penalty: 30% deduction each day

All report should be written using 12 pt font, single column, 1.5 space. **Keep Formats**!

Project Proposal Report should include the followings:

- Project title with student's names
- Project idea: what you want to do and achieve. Make the description as clear as possible.
- Description of dataset (what to collect and how, size, content, why the dataset is appropriate)
- Project plan: how you do the project with the description of system, software, and methodology you will need to use.

- Some relevant articles or papers as reference. (No more than 5)
- Teammates and work division.

Project Progress Report should include the followings:
- Content from the proposal (not exact copy but modified version)
- Any updates or modification in plan (dataset, methodology, objectives, scope, etc.)
- Progress of the project (should be significant) including intermediate results if available

Project Final Report should include the followings:
- Project title with student's names
- Problem definition
- Description of background (why is this meaningful? What to solve? Related work? Contribution?)
- Description of dataset (quantity, quality)
- Description of methods used
- Experiment: experiment setup and analysis results
- Observation and Conclusion
- References

Note: Carefully follow report formats and submission guidelines. Otherwise, there will be penalties.

More details will be explained in the class.