DSCI550 Data Science at Scale

Spring 2022 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 three students.
- Team members are responsible for dividing up the work equally and making sure that each member contributes.
- If you are having troubles, contact the instructor.

Your project will be worth **25%** of your total class grade, and will have 2 deliverables:
- Report your team formation by **Feb. 18, 2022**. (names, emails, team leader)
- Project Proposal Report (20% of project)
    - Due on **Feb. 28, 11:59 PM**
    - 2 pages
    - Submit to Blackboard in PDF file with the name, Pre-proj-"your name".pdf.
    - Late penalty: 30% deduction each day
- Project Progress Report (10% of project)
    - Due on **Mar. 28, 11:59 PM**
    - 4 pages (2 page proposal + 2 page updates & progress)
    - Submit to Blackboard in PDF file with the name, Pro-proj-"your name".pdf.
    - Late penalty: 30% deduction each day
    - No late submission allowed
    - Presentation opportunity in the class for the top projects
- Final Report (70% of project)
    - Due on **May. 4, 11:59 PM**
    - No more than 8 pages
    - Submit to Blackboard in a single zip file, Final-proj-"your name", 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.

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)
- Project plan: how you do the project with the description of system and software 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 in the proposal
- Any updates or modification in plan
- Progress of the project (should be significant)

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? Contribution?)
- Description of dataset (quantity, quality)
- Description of method used
- Experiment: analysis results
- Observation and Conclusion
- References

More details will be explained in the class.