

Group Project

Objective

This group project is to encourage students to study an area of big data applications in depth. Projects can be done in groups of one, two, or three.

The project is quite open ended. You can use the techniques taught in this class or any big data tools and methods you know.

In addition to your coding work, you need to present to the entire class about your findings. The total points of the this project is 15. You will receive 10 points if your code runs and generate the same results as you presented. The other 5 points will be assigned by the the entire class based on your presentation.

You will only grade other team's project. For all the grades received for a project, the highest score and lowest score will be removed before calculate the final average.

Each presentation must be done within 20 minutes. Prepare 15-16 minutes talk and leave 4-5 minutes for Q&A.

Every one should show up for every presentation. If I do not receive your grading for a project, I will subtract 0.5 points from your project score.

Bonus Points

After all presentations, we will have a vote for the best projects in each of the following catalog. You CANNOT vote for your own project. For the top project of each catalog, all team members of that project will receive 5 bonus points.

1. The project with most innovative ideas

This is to encourage students to come up with fresh ideas. The idea of this project is new and impressive. No other person has done similar analysis on this data set.

2. The project with the most advanced technology

This is to encourage students to explore more on big data. Big data is a fast involving area. New methods, approaches and tools are introduced every day. Finding

good approaches and make them work with the data set involves a lot of energy and effect. The enthusiasm and effort will be recognized here.

3. The most interesting project

The project or the findings from this project are very interesting. The author explains very well about how they approached the problem, how they get the results, and what are the fascinating findings they found in the data.

Important Dates

11/19/2018	The project proposal due.
11/04/2018-12/10/2018	Group presentations
12/10/2018	All codes, slides, reports must be submitted to Canvas.

Submission

You need to submit the following documents, project proposal, source code, slides, report of results, and project instructions paper (to compile and run the code), according to their deadlines. The deadlines are posted on Canvas.

Project Proposal

The project proposal should have the following sections.

1. A list of students in this project.
2. The data set(s) you will use.
3. What you plan to find from the data set(s).
4. The techniques/tools that you plan to use to analyze the data.
5. The tools/methods that you plan to use to present your result.

Project Presentation Slides

Each group should submit the slides they used in their presentation. The slides should cover the following topics.

1. The title of your project and the names of the team members.
-

2. Brief description about the data set you used.
3. From the data set, what you planned to find.
4. Explain in detail how you analyzed the data (the tool set, the algorithm, etc). During the analysis, what difficulties you encountered, and how you solved the problems.
5. If there are related works and literatures, compare your approach with theirs. Explain why you use this approach to achieve your goal.
6. Each student should present the portion that he/she worked on.
7. Explain the results you found. Present the results with graphics.

Project Source Code

Submit all your source code, which are compilable and runnable.

Report of the results

List of your findings from the data sets. Provide brief explanations about the results.

Instructions

In this writeup, provide step by step instructions about how to run your source code. It should include the following.

1. Where is your data set. It can be submitted together with your code, or you can provide the link to download the data set).
2. How to compile and run your code.
3. In general, how long the program can finish and generate the results.
4. Where the result is stored.
5. How to read the result out if it is not human readable.
6. If you are using tools other than the ones we installed in our lab VM. Please provide instructions about how to install them.
7. If you have packages/libraries that are needed to run your code. Provide instructions about how to install them as well.

In one word, if a grader follows your steps in this document, he should be able to compile, run your code and get the same results as you presented in your slides.
