

CSCI 6350/7350 Data Mining

Semester Project (300 pts)

For the semester project, you may work by yourself, or in a group of two students. Here is a tentative schedule for completing the project:

- (50 pts) Wednesday, March 14th, project proposal due. The proposal writeup should include description of:
 - What data mining problem will you work on?
 - What data will you use or collect? Description of the features in the data as well as the data size.
 - Preliminary ideas about how you will approach this problem.
 - Past work related to your work, include
 - Past work addressing the same problem, and/or used the same data, and/or applied the similar approaches to your preliminary ideas.
 - Implementation effort estimation:
 - How much effort on programming, learning other software, collecting and pre-processing (cleaning) data, learning new material, etc.
 - An estimated timeline to complete the project

I will try and talk to each team after receiving your proposal to help refining the project idea. You are encouraged to discuss your project ideas with me as you are developing the proposal.
- (50 pts) Wednesday, March 21st, in class 5-minute presentation about your proposed project
- (100 pts) Monday, April 9th, project progress report due. The report should include updates on
 - Changes (if any) you have made about your project proposal;
 - The status of data collection/selection/preliminary usage;
 - Progress made in your project:
 - What has been achieved
 - Preliminary results
 - Any difficulties encountered, and your solutions to them, etc.
 - What remains to be done

The form of the progress report (writeup or presentation) will be announced later.
- (100 pts) April 23rd and 25th, project due.
 - In class project presentations;
 - Electronic submission of your project. Your project package includes:
 - Final project report. This will be an extension of the progress report with final results and conclusions.
 - Data and code used in your project