

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) Thursday, March 27th, project proposal due. The proposal write-up 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.
 - Data mining methods to be used
 - 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

(25 pts) Each person/team presents a 5-minute presentation about your project **proposal**.
- (50 pts) Thursday, April 10th, project progress report due. The report should include discussions on
 - Changes (if any) you have made about your project proposal:
 - changes made to the methodology proposed
 - 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

(25 pts) Each person/team presents a 5-minute presentation about the **progress** of your project.
- (150 pts) Thursday, May 8th, final project due.
 - (50 pts) project presentations;
 - (100 pts) 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