

## **Research Project Logistics & Rubrics**

The aim of the IS 733 project is to give you hands-on experience in data mining, and to help synthesize and apply what you are learning in the course. This course is about data mining techniques for finding hidden patterns in complicated data sets in order to solve real-world data analytic tasks, and your projects should reflect this. Projects will address either the application of data mining techniques to solve a particular task, or the development of new techniques, validated by experimental results. A typical project might involve:

- Selecting an interesting data mining task to work on, e.g., with industrial significance, or motivated from computational biology, sociology, etc. Acquiring a dataset (often by downloading an existing dataset, but could instead involve web scraping, etc.).
- Exploratory analysis and/or visualization to get to know the data (see Lesson 3).
- Data preprocessing (see Lesson 3).
- Implementing and applying one or more data mining techniques to solve the task.
- Rigorous experimental evaluation. For most projects, this will involve cross-validation or train-tests splits, validation sets, and comparisons to baseline methods (see Lesson 8).

### **Tentative Poster Presentation Schedule:**

The poster presentations have been scheduled on the last day of the class.

- Tuesday 11<sup>th</sup> May (13 minutes each presentation with 2 minutes Q&A)

### **Presentation Grading/Rubrics:**

Innovative Application (20 pts)

Selection of Appropriate Datasets (20 pts)

Appropriateness of Methodology (20 pts)

Experimental Results/Evaluations (20 pts)

Overall Presentation (20 pts)

Total = (100 points)

### **Project Report Formatting and Submission Procedure**

Papers must be 6 pages or more, including references, figures and tables (at least 10pt font, 2-column format) and should follow the IEEE conference proceedings formatting guidelines. The IEEE LaTeX and Microsoft Word templates, as well as formatting instructions, can be found at the following link: <http://www.computer.org/portal/web/cscps/formatting> The final project report should be uploaded as a soft copy (pdf) on the blackboard by **Tuesday, May 18<sup>th</sup> midnight.**