Project
Stage 2. Proposal

Due date: Wednesday, October 21, in class

Project Proposal

The second stage of the course project is the proposal. Each group shall, in consultation with all its members and the instructor, select the topic of the course project, and write a project proposal — a document specifying the problem and the proposed solution.

Deliverables and submission instructions

Each group has to submit two deliverables: the project proposal document and the proposal presentation.

Project proposal. The project proposal shall include the following information:

1. Name of the project (title);
2. Names of all people in the group (authors);
3. Abstract: a one-paragraph description of the proposed project.
4. Introduction: a slight extension of the abstract, describing the proposed project, the needs it addresses, the technologies to be involved and the structure of the rest of the proposal.
5. Problem statement: brief, but precise (formal) description of the problem.

6. Proposed solution: brief description of the proposed solution. If you propose to build software - describe what it will address, what its archetecture would be. If you propose to develop algorithms, describe their purpose. Outline any testing procedures to be performed.

7. Data/Datasets: If your project will work on specific data, provide a brief description of the data/dataset(s) you are planning to use.

8. Bibliography: all external information must be properly cited and attributed in your proposal. Please provide full citations for any sources used.

The proposal shall be submitted to your group's wiki page. You may choose to use LaTeX or MS Word to write the proposal, but, please, submit your proposal in PDF format.

There is no strict upper or lower limit on the size of the proposal. It should be succinct: you will have a chance to expand it throughout the course. At the same time, all necessary information should be presented. My estimate is that proposals will range from 4 to 10 pages in standard MS Word format.

Proposal presenation. On the due date, each group will get 10-15 minutes to present their proposed work to the entire class. The presentation should be accompanied by some displayed materials (e.g., powerpoint slides, although each group is welcome to decide what to present). These should be submitted to the instructor prior to classtime, to ensure their placement on the instructor's laptop.

Each group is welcome to choose the presentation format - either use a designated presenter, or do it as a team, etc.

Following the completion of all group presentations, we will hold a general discussion of the proposed projects.

Projects Types

As stated in the Stage 1 handout, your project may take any of the five forms listed below (I have added one more project type to the original list of four).

- New method/algorithm/technique. You propose to create/implement/test a novel approach (method/technique/algorithm) for solving a known data mining/web mining/KDD problem. To your knowledge you are the first group suggesting this specific method/algoithm/techinque to solve this particular problem.

- Improvements to existing algorithms. You propose to improve the implementation of an existing KDD/data mining/web mining techique/method/algorithm with the purpose of achieving improved performance (as measured by method accuracy measures and/or running time). To your knowledge, your particular improvement has not been considered.

- Empirical evaluation. You propose to implement and empirically evaluate a number of methods for solving a problem from the KDD/data mining/web mining domain. You plan to implement all or some of the methods you are studying, and to your knowledge, the empirical evaluation you have in mind has not been done before.

- Solving a real problem. You propose to use existing KDD/data mining/web mining techniques to solve a **real** problem in a specific application domain. To qualify, the problem must be **real** (i.e., exist outside of the course content and have an interested party/client/customer), and either unsolved, or unsolved using the methodology you want to consider.

- Building a system. You propose to create a (working prototype of a) software product designed to use available KDD/data mining/web mining techniques to solve a specific problem, or to provide some interesting (in your opinion) functionality to its users. The key difference between this and solving a real problem type of project is that you may desire to build a system that does not necessarily solve an existing applied problem (for example you want to replicate certain functionality and improve visualization).

Your project, of course, can combine any of the five purposes together.

Consultations

I strongly recommend that before each group commits the proposal to paper, it meets me for a few minutes outside of class time (either during office hours, or be scheduling an appointment) to discuss the group's idea for the project.

Unless your initial idea is out of the scope of the course I expect that I will "approve" of your initial plans. However, I would like to know about them as soon as possible. I do not need to meet with the entire group: you may choose a team representative to talk to me (this may help schedule a meeting among other things).

Due Dates

The **presentations** will happen on Wednesday, October 21.

The **presentation slides/materials** should be available on the wiki (or emailed to me) by 1:00pm on Wednsday, October 21.

The **proposals** should be available for viewing on the wiki by morning (9:00am), Tuesday, October 20. You can post drafts and replace them later with the final version, but the draft must be reasonably advanced (have significant text in all sections and not differ significantly from the final version).

I will make every effort to bring my comments back to each team by Monday, October 26.

Head Start

The official commencement of Stage 3 of the project is Monday, October 26. This gives you just over one month to complete the project. However, if (and this is another reason why early consultations with me are useful) you get an early OK from me to go ahead with your project, your team may elect to start preliminary work on the project early. This may give you (realistically) as much as extra 10-14 days to complete the tasks for the project.