Project Overview
Stage 1 (Group Formation)

Due date (Stage 1): Monday, September 28

Project Overview

This is one of the two quarter-long assignments in the course (the second one is the in-class presentation).

The course project shall be done in small teams. Given the number of students in the class, 3-4 person teams are expected. 2- and 5- person teams will be considered on individual basis. Team formation is discussed below.

Each team will submit a proposal outlining the project it wants to do and upon instructor's approval, will perform the activities necessary to complete the project.

While more specific instructions will be given to you at the beginning of Stage 2 of the project, here is some preliminary information about the nature of the course project.

Project topic. You are expected to propose a project on a topic related to the course. Any subarea of data mining/KDD is considered fair game. For example, proposals concentrating on information retrieval (especially from the Web), collaborative filtering, text mining are all considered to have an appropriate topic. On the other hand, proposals to build a web-based system with little or no KDD/data mining content will be considered inappropriate.

Project type. Generally speaking, your proposed project should have a research component. In short, a research component can occur via one of the following means:

- 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.

Project flow. The project consists of three stages:

Stage 1. Team formation. On this stage you will form the teams.

Stage 2. Proposal preparation. On this stage you will determine the topic you want to work on, and will write a short proposal outlining your team's putative project.

Stage 3. Project implementation. On this stage you will conduct all the work necessary to complete the project.

The timeline is as follows:

Stage	Name	Starts	Ends
Stage 1	Team formation	September 23	September 28
Stage 2	Project Proposal	September 28	October 19
Stage 3	Project Implementation	October 19	December 7

Project deliverables. The project will have the following deliverables:

No.	Deliverable	Stage	Medium	Madatory/Optional
1.	Team Name/Roster	Stage 1	wiki	*Mandatory*
2.	Proposal	Stage 2	document/wiki	*Mandatory*
3.	Proposal presentation	Stage 2	oral/slides	*Mandatory*
4.	Project report	Stage 3	document/wiki	*Mandatory*
5.	Project presentation	Stage 3	oral/slides	*Mandatory*
6.	Datasets	Stage 3	files/wiki	*Optional*
7.	Code	Stage 3	files/wiki	*Mandatory*

Each team will prepare two written documents: the project proposal and the project report. Each team will give two presentations: proposal presentation and project presentation. Teams will submit any datasets used, generated, constructed, acquired during the performance of the project, for which public display and distribution of the data is permissible.

Use of wikis. Each team shall maintain a space on the course wiki. You should feel free to use the wiki and the code repository for the day-to-day activities of the team in project execution. Additionally, all artifacts generated by the team while working on the project need to be placed on the wiki (e.g., all reports, presentation slides, etc.).

In-class presentation overview.

Presentation length. Based on your work throughout the quarter, each of you will prepare an in-class presentation. We have alotted six class periods (possibly seven) for these presentations, at the end of the quarter (tentatively starting November 9). We can fit three presentations into a single class period, so, each presentation should be about 35 minutes in length.

Presentation style. Your in-class presentations should be styled as **lectures** or **mini-lectures**, rather than research reports. The objective of each presentation is to introduce the class to the new technical content (problems, techniques, methods, algorithms, etc). You can use any means you find useful to deliver the presentation (board, slides, handouts, etc.).

Scheduling of presentations. The schedule of presentations will be worked out by me in consultation with everyone. My goal is to schedule presentations on related topics and issues close to each other. Some of you **will have to** present on November 9, while some will present much later (December 2).

Stage 1: Team Formation

Your first task it to break into teams. Please use the following guidelines when forming the teams:

- I expect teams to be 3-4 people in size. I will consider allowing teams of 2 and 5 people to stand, but you have to have an explanation of why you want a team of this size. For larger teams, I will expect project proposals outlining proportionally more work to be done.
- There are six people on the class roster who have taken CSC 466 course last Spring. It would be great if each group had at least one student with CSC 466 experience in it. (This is not a hard requirement,

however, CSC 466 alums have done some intensive programming in the area and may be good assets to any team).

- When forming a team, please make certain your schedules allow you a number of opportunities to meet as a team outside of lecture hours.

Deliverables

By September 28 class time, please do the following:

- Wiki Space. Create a wiki page for your group. Put on the page the list of all students in the team with their CalPoly IDs.

- Joint CV. To understand better the assets your team has in terms of CS experience, create and post on the wiki (either as a document attachment, or as a wiki page) a simple joint team CV. This CV essentially combines the experience of your team into a single list. List all skills (programming languages, environments etc.) with the combined number of years between you, as well as your combined professional experience (e.g., if three people had summer internships at Google, and one person - at Microsoft, write "three Google internships, one Microsoft internship"). You do not need to associate names with specific skills/experiences.