Instructions for Project Proposal for Distributed Data Mining

The purpose of the project proposal is for you to give the instructor an idea of what problem you want to study as part of the class project.

Your writeup should be atmost ONE page with normal fonts and margins. If participating in a group project, please provide ONE report per group. The proposal could contain the following items:

1. **Project Participants:** Clearly identify the project participant(s) and state what you expect would be each person's contribution if working in a group.

2. **Summary of proposed research:** About half a page should be used to succinctly summarize the problem you are trying to address; Define conceptually / mathematically the problem and its relevance. Elaborate on the *distributed* component of your problem.

3. **Novelty of your technique:** Describe in a few sentences why what you are doing in your project is novel. You could try and distinguish this work from very closely related work in literature.

4. **Algorithm Description:** Mention the **algorithm(s)** you are planning to use. Is an implementation of your algorithm available already (if so, where?).

5. **Data sets and Software:** Explain what data sets and software you are going to use. Provide pointers (websites, wikipages, references etc.) as relevant.

6. **Evaluation:** Present a brief summary of the proposed empirical evaluation if relevant. If doing a theoretical project, indicate ways by which you can prove lemmas and theorems on convergence etc.

7. **Analysis:** What is the intuition for proofs of convergence of your distributed algorithm? Do you have an idea on how to come up with message and time complexity analysis? What kind of protocols are you going to use (if any)? Are you going to study the effect of graph structure on your algorithm?

8. **Proposed Future Work:** If your project has an interesting sub-problem that needs exploration in future, beyond the scope of the course project, describe this clearly.

9. **Bibliography:** Provide *complete* references for papers, software, sites from where data sets are downloaded, network packages used and any other material that will be made use of in this project. If using latex, this should be done using Bibtex.