CS 177 Data Warehousing and Data Mining (Spring 2019)

Final Project Proposal Guidelines

Each team leader submits the proposal on Canvas. As for the length, I would expect a proposal <u>as least with one full page</u>.

Suggested Outline for Your Proposal

Paper Title: Make the title eye-catching…

Project Type: A or B or C?

Team Members: Name and Id of each team member.

Project Description/Abstract: provide a description of the problem you plan to address

Motivation: briefly explain why this problem matters

Background: provide a brief summary of what is known in the literature about this problem.

Data sets: provide a link to the data set (or sets) that you will use and briefly describe it (them).

System/Algorithmic design (optional): describe what algorithms you plan to use in your project.

- If you plan to improve any existing design in a paper, state how you think your design would work, or at least your current ideas.
- If you plan to use any model/algorithms in scikit-learn API or in an academic paper, provide brief descriptions and references.

Evaluation plan (optional): you define how you will evaluate your results.

- For prediction problems, such as classification or regression, this is straightforward, since we may use learning/test data split. For classification metrics, use precision, recall, F-measure or ROC AUC. For regression metrics, use RMSE or R2.

- For techniques such as clustering, you might apply the algorithm on a data set where the class labels are known, but you remove them and see if the algorithm can recover them. As for clustering metrics, use precision, recall and F-measure.

Task Division: how you would divide the task over all team members.