Project

CS 484-002: Spring 2023

1. General information

There will be one semester-long team project, which accounts for **30%** of final grade. Students are required to form teams of **3-4 students** and carry out a project related to the topic of '**Data Mining for Social Good**' or '**Socially Responsible Data Mining**'. Students are responsible for determining a specific problem related to this topic; finding appropriate datasets for conducting experiments; designing new data mining algorithms for tackling the target problem (or conduct thorough analysis to study the problem by data mining techniques); and present the problem, method, and results by written reports, in-class presentations, and videos.

2. Basic Requirements (grading criteria)

(1) the problem proposed is **reasonable, important, and is related to social good or social responsibility**;
(2) the datasets used are reasonably **large** and well **processed**;
(3) the applied data mining techniques or the designed new algorithms are **non-trivial**;
(4) conducting **comprehensive experiments** to support the conclusions;
(5) if new algorithms are proposed, empirical comparison with **baseline** methods must be included;
(6) clear and error-free presentations (documents, videos, in-class presentations).

3. Example topics

Example problems include (but not limited to): (1) study how to evaluate and alleviate movie recommendation unfairness in a movie recommender system; and (2) study how to evaluate and enhance fairness among genders and races for income level classification; (3) data mining for health; (4) study and address unfairness/bias/discrimination in a specific data mining task/system/application; (5) data mining for social issues; etc.

Some useful resources for datasets:
https://www.kaggle.com/datasets
https://github.com/awesomedata/awesome-public-datasets
https://archive.ics.uci.edu/ml/index.php

Attention: One type of project is not preferred – running models by libraries on a Kaggle dataset to solve an existing classification or regression problem.

4. Milestones

(1) Project proposal document (2%):

a) Format: 1- or 2-page document.

b) Important content: name of your team (make a cool team name!); team member information (name, NetID, G#, email); project motivation and background; project goal; dataset information; schedule.
c) Due: 02/19 11:59 pm EST

(2) Midterm Presentation (5%):

a) Format: 5-min pre-recorded video.
b) Important content: project motivation and background; project goal; data; method; and preliminary results.
c) Due: 03/26 11:59 pm EST

(3) Final Presentation (5%):

a) Format: 10-min in-class presentation or a two-hour poster session.
b) Important content: project motivation and background; core contributions; technique details; empirical result details.
c) Date: TBA

(4) Final Report (9%):

a) Format: document (minimum of 5 pages excluding references).
b) Important content: project motivation and background; core contributions; technique details; empirical result details; specifying contributions of each team member.
c) Due: 05/07 11:59 pm EST

(5) Code and data (4%):

a) Format: a link to a github repo with all code and data of the project.
b) Important content: a README file showing how to run the code, and sufficient annotations in code.
c) Due: 05/07 11:59 pm EST

*** For written documents, students are encouraged to use LaTex for typesetting and the NeurIPS LaTex template (https://nips.cc/) is highly recommended ***