

CMP 466 – Machine Learning and Data Mining
Project Proposal

Choose a dataset of labeled samples for a classification task. This dataset will be used during the semester for classification.

Submit a report including the following items by the deadline:

- 1- Cover page that includes a tentative project title and group members information.
- 2- Introduction and problem statement: what is the problem you are trying to solve? Why do you think it is interesting and important?
- 3- A brief overview of the literature (existing works) about the problem.
- 4- Description of the dataset including the following:
 - 4.1. Size of the dataset in terms of the number of samples and the number of features.
 - 4.2. List of features such as age of a patient, salary of an employee, number of rooms in a house, etc.
 - 4.3. Number and description of the classes/labels such as spam/not spam email, malignant/benign, etc.
 - 4.4. Number of samples in each class. Example: the dataset has 520 samples belonging to class malignant, and 1000 samples belonging to class benign.
 - 4.5. Other interesting information.
- 5- References including the following:
 - Link to the dataset or to the paper(s) that generated/ first used the dataset.
 - Papers using the dataset that their work is close to your work.
 - Any interesting research papers you cited in the introduction or literature review above.

Students are highly encouraged to conduct a quick literature review to understand the problem and the previous work done on this topic before committing to an idea.

Good luck!