As a course requirement, students must work on a relevant research project and produce an associated research paper or report. In a typical project, some machine learning technique (or a collection of techniques) is applied to one or more data sets and the performance analyzed. If multiple algorithms are applied, a comparison can be made. While the project need not develop a new machine learning technique or extend existing ones, a good project should at least provide some insight into the techniques used or their application in the chosen domain. Students can work in groups of at most 3.

1 Report Requirements

The associated paper should be the equivalent of at least 15 typed pages (double-spaced, 1 inch margin, 12pt font). It should include the following:

- a title, date, and list of authors;
- an abstract summarizing both the objective and results;
- a presentation of the problem and materials and data sets used;
- a detailed discussion of related work;
- a detailed discussion of experiments performed as well as any data preprocessing;
- and an analysis of the results.

A significant number (e.g. 10) of outside works should be cited. The associated report must be in the format of a conference paper. Formatting templates are typically found in a conference's call for papers. E.g.:

- `https://aaai.org/Conferences/AAAI-18/aaai18call/`
- `http://icdm2018.org/calls/call-for-papers/`
- `https://icml.cc/Conferences/2018/CallForPapers`
- `http://reasoning.eas.asu.edu/kr2018/`
- `https://www.ijcai-18.org/`

The above conferences can be used to help identify project topics. Regarding the references, though it is likely that only a few works will play a key role in the project, it is essential that you perform a literature review and identify relevant work.

If multiple authors are involved in the project, then each should be involved in the writing of the final paper. A statement of what each did should be included.

2 Sources of Data

The following are possible sources for data sets.

- `https://www.kaggle.com/datasets`
- `https://www.data.gov/`
- `https://archive.ics.uci.edu/ml/datasets.html`
- `http://lod-cloud.net/`
- `http://wiki.dbpedia.org/`

3 Grading

1. General formatting, spelling, and grammar **(10 points)**
2. Abstract and introduction **(15 points)**
3. Literature review/discussion of related work **(15 points)**
4. Presentation of the problem and the data set including handling/preprocessing of the data **(20 points)**
5. Design and discussion of experiments **(20 points)**
6. Analysis of results **(20 points)**

4 Due Dates

- Project topics should be chosen soon. Review the course website for deadlines.
- During the term, certain milestones might be specified. These might be counted as HW assignments.
- The final report and any necessary software should be submitted via eLC on or before the last day of class.