

CS 8331 - Data Mining

Guidelines for the Main Project

For the final project you have a choice between a data mining project, implementing a data mining method or writing a review paper. The topic of the final project should be coordinated with your tutorial's topic.

Tasks

1. I will set up individual meetings in **Week 2** to discuss the topic. Please come to the meeting already with an idea about the project. Note no regular class meetings are held that that week.
2. A mid-term progress report (1 page) for the project. Bring a printout to class (see Syllabus for deadline).
3. The final project report is due via email (see Syllabus for deadline).
4. You need to give a short presentation of the highlights of your project report during the last two weeks of class.

Project types

- **Implementation:** Identify an algorithm which you would like to implement. Implement the algorithm (preferable with an interface to a data mining environment (R, Python, WEKA, MOA, etc.). Test/compare your implementation with a very small data set. Write a short report (10 pages + tables, figures, etc.)
- **Review paper:** Identify a sub-area of the topic you choose for your tutorial. Write an extensive review paper where you compare results from different researchers. In this paper it is very important to critically evaluate the achievements in the field and results presented in the literature. The paper should be 30 pages + an extensive bibliography (30+ relevant references).

Guidelines for the Project Paper

Your project paper needs to be structured as a paper ready to be submitted to a conference or journal. Use the ACM template:

<http://www.acm.org/sigs/publications/proceedings-templates>

You may use the Word template or the Latex template. Submit your paper in PDF to me.

Your paper will be evaluated using the review form that we have used in class. So make sure that:

- Your paper has an abstract that explains what your project is about and highlights some key results.
- Your paper clearly states the background and motivation for your work.
- The paper is technically sound.
- The paper contains enough information (including equations, definitions, pseudo-code, etc) to make any results reproducible.
- Discuss the results sufficiently such that the reader understands what they mean for possible applications.
- The bibliography is adequate and presented in a consistent format.

Grading

Your work will be graded using the following categories:

- Abstract (10%)
- Difficulty of topic (10%)
- Clear definition of problem area (10%)
- Technical detail (40%)
- Results (10%)
- References (10%)
- Presentation style and support (10%)

Distance students

Distance students need to prepare a video presentation (using e.g. Zoom).