

Project Requirements for ENGO645 (2024 Winter)

1. Overview

The project is a vehicle for the student to take an active part in exploring the subject area, as appropriate for the data mining and spatial databases course. Thus, a project must be chosen that will clearly extend your knowledge and understanding of the data mining and spatial databases area relevant to the course. Moreover, the primary criterion for evaluating your project will be **what you have learned and discovered from the project**.

There are three aspects to a project:

- Problem definition, literature and technology background.
- Algorithm applied and designed
- System implementation and experimentation

The problem definition, literature and technology background should explore and describe the problem or area of interest and provide some investigation of background work in this area. It should examine the research literature and available technologies in the selected area. Because of the time limits, this stage is generally expected to be very limited, e.g. google around, read a couple of papers, study from online tutorial and books, think hard for awhile and write your project proposal.

The algorithm should try to map out an algorithmic solution that is implementable with the available time and computer resources. With the intended "prototyping" nature of the projects, the design, implementation and evaluation may be performed iteratively.

In the implementation and experimentation phase, the algorithm design and implementation need to demonstrate the algorithm feasibility, effectiveness and efficiency.

In general, the course project must be relevant to the issues and techniques and directions covered in the course. But you are allowed to use more advanced data mining, AI and machine learning methods. Original thoughts, solutions, insights are strongly encouraged and will be rewarded.

All work submitted should be done in a professional manner. Documents should be type-written and use good technical English. Software should be well-written, meeting reasonable standards of programming.

2. Data requirement

Real datasets are required for the project. The following are some open data websites, but feel free to search and use datasets from other websites. Please remember to review the terms of use for each dataset to ensure compliance with any licensing or usage restrictions.

- UCI Data Repository: A collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms. (<https://archive.ics.uci.edu/datasets>)
- Data.gov: The U.S. government's open data portal, providing access to thousands of datasets. (<https://www.data.gov/>)
- European Data Portal: A platform that provides access to open data published by European countries. (<https://www.europeandataportal.eu/>)

- UN Data: A gateway to the statistical databases of the United Nations. (<http://data.un.org/>)
- World Bank Open Data: Access to a comprehensive set of international development data. (<https://data.worldbank.org/>)
- Google Dataset Search: A tool that enables the discovery of datasets stored across the web. (<https://datasetsearch.research.google.com/>)
- Kaggle Datasets: Kaggle is a platform for predictive modeling and analytics competitions. It also hosts various datasets for exploration. (<https://www.kaggle.com/datasets>)
- GitHub - Awesome Public Datasets: A curated list of high-quality datasets on GitHub. (<https://github.com/awesomedata/awesome-public-datasets>)
- Open Data Network: A platform that provides access to open datasets from various sources. (<https://www.opendatanetwork.com/>)

3. Team

The team size varies depending on the program you are enrolled in. For Thesis-based MSc and PhD students, the recommended team size is 2 to 4 members. For MEng students, the suggested team size is 4 to 6 members. It is essential for each team member to actively contribute to the project and articulate their respective roles in both the final project presentation video and report.

4. Important Dates:

The project development must be completed by the end of the term, and the following milestones must be adhered to. Each team is required to submit the relevant materials and documents to D2L by the specified dates:

Feb 12: Submission of project team details, topic, and a brief description (0.5 page)

Mar 11: Deadline for the project proposal, including the problem statement, objectives, possible solutions, and current project progress (1~2 pages)

March 27, Apr 3,7: Project progress presentations to be conducted in class

Apr 17: Final submission of the project presentation video and project report