

KM6312 Group Project (35 marks)

This is a group project. The size of each group is 4-5 students (excluding auditing students). Please enroll yourself into a group via **NTUlearn->Content->Group Project Self-enrollment**. After your group is formed, discuss with your group members and come up with a group name (start with "Team"). Please post your group name using the "Blogs" under your group. Then start working together on your project proposal and the fantastic project!

This project will account for **35%** of the overall grade, consisting of **project proposal** submission (5%), **project report** submission (15%), and **group presentation** (15%). Please refer to the class schedule for the due date of the project proposal, final report, and group presentation. Note that the team members will not receive the same grade, we will take **individual contributions** to the group project into consideration.

This is a data mining project where you collect your own sample dataset or use an existing dataset, and using data mining techniques and tools, build an interesting model that mines and analyzes knowledge/information from the dataset. Generally, the project scope is entirely up to you, but I suggest that you build a useful and interesting model (or application). Then, write a project report explaining your methodology and presenting the results.

You may conduct investigative analyses of your dataset using two or more of the following data mining approaches:

- Exploratory data analysis (e.g. data visualization)
- Regression (e.g. score prediction)
- Classification (e.g. sentiment classification)
- Clustering (e.g. document clustering)

Project Proposal Submission:

No specific format requirements, write clearly about your project planning including: introduction, motivation, data, methodology, experiments, expected outcome, etc.

Only require the soft copy. Please submit your project proposal via NTUlearn->Assignments->Group Project Proposal Submission by the due date. Include the names of team members in the message.

Project Report Submission:

Must include a bibliography listing all references (including URLs, if any) cited.

Length: 8-10 pages.

Formatting: The use of 10-point Times font is mandatory. The formatting should be referred to [ICML style](#). The word template could be found [here](#). The latex template is provided in the [overleaf](#).

Only require the soft copy. The soft copy should be submitted through **Turnitin** under NTUlearn->Assignments->Group Project Final Report Submission.

Create your [GitHub project page](#) that contains the **dataset and code** used by the project. And put your GitHub repo link in your report.

Group Presentation:

Each team will be asked to conduct a 15/20-minute presentation on their project work. This will be followed by a 5-minute question-and-answer session to allow for clarification by students and the lecturer. Schedules for the presentation will be announced later.

Every team member is required to present.

Note that reports and required files/documents submitted after the due date will be marked down by **10% per day**.

IDEAS INSPIRATION

You might look at recent deep learning publications from top-tier machine learning conferences and labs, as well as other resources below.

- [NeurIPS](#): Neural Information Processing Systems
- [ICML](#): International Conference on Machine Learning
- [Kaggle challenges](#): An online machine learning competition website (**recommended**)

How to get Data?

- Get data from social networks, such as Twitter or other websites
- [Kaggle Datasets](#)
- [UCI](#)