

# MSBD 5018 Group Project Guideline

## Spring 2023

### Final grades compositions:

The final grade is determined by the project and the presentation. This shared doc shows the detailed grading policy for the project. The details for the presentation of the project will be released later.

Activities	Max Point Values
Project	100% (40 pts)
- Proposal of the Project	- 15% (6 pts)
- Results (improvement, novelty, etc)	- 15% (6 pts)
- Final Report	- 60% (24 pts)
- Reproducibility (codes and data)	- 10% (4 pts)
Presentation	10 pts

### Details:

#### 1. Project grouping (due 23:59 21<sup>st</sup> March.)

The project is supposed to be a group project with a maximum of 3 students in a group. Please indicate the contribution of each student in the group. **If you choose to do it solo, we will take your workload into consideration.**

To form a group, you could register the group or look for teammates at this shared doc: [MSBD5018 Grouping Info.docx](#).

#### 2. Proposal of the selected project (15%, 1-page proposal, due 23:59 2<sup>nd</sup> April.)

You can select a research project to work on. The related topics include:

- **Project 1:** SemEval 2021 Task 5: Toxic Spans Detection.  
<https://aclanthology.org/2021.semeval-1.6.pdf>
- **Project 2:** Temporal Commonsense Reasoning: MC-TACO dataset.  
<https://arxiv.org/abs/1909.03065>
- **Project 3:** DROP: reading comprehension. Task Description Paper.  
<https://aclanthology.org/N19-1246.pdf>
- **Project 4:** SemEval-2021 Task 4: Reading Comprehension of Abstract Meaning. Task description: <https://competitions.codalab.org/competitions/26153>
- **Project 5:** You can choose any other NLP tasks you prefer.

You are supposed to write a proposal about the topic you choose for the course project within 1 page in **PDF Format**, including the task introduction, why this task is important, and your proposed methodologies to solve the task.

#### 3. Results (15%, due 23:59 30<sup>th</sup> April.)

You should submit your main results (evaluation scores, e.g., accuracy of your model) on canvas before the presentation.

#### **4. Report (60%, due 23:59 7<sup>th</sup> May.)**

The report should be within 8 pages using ACL templates (<https://github.com/acl-org/acl-style-files>). It should include a clear introduction to the project problem, related works, methodologies you propose/try to investigate, main results, discussions, and conclusions.

Grading rubrics:

- Content clearness. It is better that you introduce the problem and algorithms clearly in the report.
- Description of your methods. You should justify your proposed methodologies clearly. Clear figures or flow-chart illustrations may help.
- Experimental results. It is good to have significantly better results than baseline scores. However, do not worry if you want to try some risky ideas and cannot beat the baseline. Just try to analyze the reasons clearly. It is also good to have some ablation studies about why the model works/does not work.
- Discussions and analysis. Some case studies or visualization may help you understand your model. Also, some intuitive error analysis can make the paper clearer. It is also interesting to have unique insights into your research problem.

#### **5. Codes (10%, due 23:59 7<sup>th</sup> May.)**

To avoid plagiarism, we need reproducible codes and data (the data can be an url on cloud platform. HKUST onedrive may be a choice).

#### **6. Presentation (To be determined. Temporarily scheduled on 3<sup>rd</sup> May)**

You need to prepare a 5-min (the concrete time limit will be announced later) presentation to illustrate your topic of the project, methodologies, results, and insights.