

Final Project

TA email: ntueemlta2022@gmail.com

RELEASE DATE: 10/21/2022

PROJECT DUE DATE: 12/30/2022

Introduction

The final project will take **20%** of the total grade and contain three different tasks. In this project, students will group up and choose one of the three tasks as their topic to explore. Each task is challenging, exciting and inspiring, and we believe that this project can absolutely help students gain more knowledge and practical experience of machine learning. Can't wait to see what's going on? Let's get started!

General Rules of Each Task and Bonus Opportunities

Before talking about the grouping rule and the 3 tasks, here are some general rules:

1. About plagiarism, cheating and lying:

You should specify all the references and collaboration clearly. If any form of plagiarism, cheating or lying, such as copying others' code/articles, is detected and confirmed, you will get **ZERO for the ENTIRE final project!**

2. About the proposal:

Each group should hand in their own proposal for TAs to audit (refer to the tasks description). If TAs think the content of your proposal is not appropriate, we will ask you to modify it and resubmit it before the final deadline of proposal (see the schedule). The proposal of every task can be handed in at most **3** times, and please send an email to TAs' address to inform us that you've made resubmission. The grading of proposal will take the following things into consideration:

- If you don't submit the first edition of your proposal before the first submission deadline, you will get **1 points deducted**.
- If your proposal is still unsatisfactory after the final deadline, you will get **0.5 points deducted**, and you may get further deduction for other parts, reports for example.

To prevent unsatisfactory proposal, you can discuss with TA first through emails or TA hours. TAs are willing to give you some suggestions.

3. About oral presentation and bonus:

TAs will announce the groups that should give an oral presentation one week before the deadline of the entire final project. The selected groups should record your

presentation, and submit the video. However, if you want to voluntarily give presentation, you can let TAs know and follow the rule of submission of the presentation given above. We provide **2 bonus points** for all the volunteers.

4. **About free-rider prevention and bonus for hardworking:**

To prevent free riding, the allowed minimum contribution of each member in percentage is set as: 25% for a 3-member group, and 20% for a 4-member group. TAs will judge your contribution totally based the workload distribution table in your group's report. If you fail to meet the minimum contribution, you will **get ZERO for the entire final project**. On the contrary, we honor students with strong diligence. Therefore, we provide 2 extra points for those who are in a 3-member group and have contribution larger than 45%, and those who are in a 4-member group and have contribution larger than 35%. If you are confused, the following examples may help:

Example 1: If A, B and C form a 3-member team having contribution 20%, 30% and 50% respectively. Then A will get 0 for the project, and C will get the bonus.

Example 2: If A, B, C, and D form a 4-member team with contribution 10%, 15%, 35% and 40% respectively. Then A and B will get 0 for the project, and C and D can get bonus.

5. If there are changes on any regulations (including the rule or the deadline), TAs will make an announcement on NTU Cool and the FB forum to inform all of you.
6. Prof. Wu and the TAs reserve the right of final interpretation of the final project.

Rules of Group Forming and Task Choosing

1. **Group forming:**

Typically, each group should have at most 3 students in it. However, if a group have an undergraduate member, the maximum allowable capacity of that group can be extended to 4. In general, you can form your group on your own, and sign up before 10/26/2022. If you have any problem about forming groups, you can contact TAs for help. So, don't worry about this. After 10/26, TAs will randomly assign groups to all the students who don't belong to any existing group, and announce the group list in the course of 10/28.

2. **Task Choosing:**

We limit **the maximum number of groups that can have the same task to be 11** in order to ensure the diversity of the final project. Therefore, TAs will announce a Google form for you to choose your topic. The deadline of the form is tentatively scheduled on 10/30/2022. Then we will assign the topics according to your response on the Google form, and announce the final assigned tasks of all the group on 11/01.

Task 1 – Real-world ML Competition

Students who choose this task as their topic should attend the real-world AI competition assigned by the TAs, and follow all the rules set up by the TAs and the host of the competition. The official website of the competition can be found here:

<https://tbrain.trendmicro.com.tw/Competitions/Details/24?fbclid=IwAR3Uwo0OGweCJNZgmKIKVvlxOCCeYd wHJIfiRNISVrEyGnsg1AMS2f-Pkzk>

The goal of this competition is to predict the probability of raising SAR (Suspicious Activity Report) for each case next month, and you can refer to the website given above for more details and information. As aforementioned, you should follow the rules established by the host. However, here are some additional regulations that you should follow:

- Though the team size allowed by the host is much larger than that allowed by TAs, your team should be no larger than the maximum size given in preceding section.
- Your team name should in this form: <Student ID>_<Anything you want>, where the student ID refers to the ID of your team leader (selected by yourselves!).

The grading policy is as below:

1. **Proposal (3%):**

Students should hand in a proposal in written form to demonstrate their plan to conquer the competitions. The so-called “plan” may include but not limited to data processing methods, ML algorithms and model designs.

2. **Simple baseline (1%):**

Students will get full points of this part if they pass the simple baseline assigned by TAs, which is

3. **Medium baseline (1%):**

Students will get full points of this part if they pass the medium baseline assigned by TAs.

4. **Strong baseline (2%):**

Students will get full points of this part if they pass the strong baseline assigned by TAs.

5. **Boss baseline (2%):**

Students will get full points of this part if they pass the boss baseline assigned by TAs.

6. **Ranking (5%):**

The overall ranking of the students will be calculated to the score by the following

formula:

$$\text{score of ranking} = \left(1 - \frac{\text{Public Ranking} - 1}{\text{Total Team number} - 1}\right) \times 5$$

7. **Source code** (3%):

The source code (with some explanation) that can reproduce your results should be submitted.

8. **Report** (3%):

A report briefly explaining what you have done in the competition and all of your source codes are required. In addition, please include table of workload distribution of your group in your report as well.

Pay attention! The baselines will be announced on 12/2. Since we wish to assign a set of baselines that are not only challenging, but also suitable and reasonable for you, we will set the baselines based on the overall performance of you. We believe that the baselines established by this principle will be more affordable for you than the hard-coded one. Don't worry too much about this!

Task 2 – Paper Survey

In this task, students will gain experience of surveying papers after completing all the requirements in this task. Specifically, each group should select at least **9** papers in the field of machine learning to read, and summary the research status in the chosen domain.

The detailed grading policy is here:

1. **Proposal** (3%):

Each group should hand in their proposals in written form. In the proposal, you should list the papers you want to survey. The general principle of selecting papers is to choose the papers from well-known conferences in the recent **3** years that are highly related to each other or a common domain.

2. **Reports** (16%):

Each group should submit a report in written form. The grading of the report will base on the following components:

- Abstract
- Introduction and motivation (ex: The reason you choose these papers.)
- Related works (ex: Are there any papers on the same/similar topics? What makes them different from the paper chosen by you?)
- Background
- Discussion and your own opinion (ex: Try to compare or integrate the results

from different papers and make your own comments on them.)

- Conclusion
- References

In addition, please include table of workload distribution of your group in your report as well. The evaluation of contribution will be mainly based on the table.

3. **Required number of papers surveyed (1%):**

You can get the full point of this part if you meet the requirement of surveying 9 or more papers.

Task 3 – Paper Review and Implementation

Paper review and reproduction is quite important for researcher. In task 3, you will experience this! Please choose at least 1 paper related to ML to review. In addition, you should try your best to make code implementation of that paper.

The detailed grading policy is:

1. **Proposal (3%):**

Each group should hand in their proposals in written form. In the proposal, you have to provide the paper you would like to review and implement and reasons of make this choice. The general principle of selecting papers is to choose the paper from well-known conferences in the recent **3** years that are highly related to each other or a common domain.

2. **Reports (14%):**

Each group should submit a report in written form. The grading of the report should contain following components:

- Abstract
- Introduction and motivation (ex: The reason you choose this paper.)
- Problems addressed by this paper. (ex: What did the author want to deal with?)
- Solutions by the original paper and the comparison with existing approaches.
- Description of your implementation (ex: The functionality and performance of your implementation.)
- Discussion (ex: Have you encountered any troubles? How did you solve them?)
- References

In addition, please include table of workload distribution of your group in your report as well. The evaluation of contribution will be mainly based on the table.

3. **Source Code with explanation (3%):**

You should submit your source code and a brief readme to let TAs understand your code easily and execute it in correct manner.

Tentative Schedule

10/21	Final project announced
10/26	Deadline for signing up groups
10/28	Announcement of group list
10/30	Deadline of the task-choosing form
11/01	Announcement of the assigned topics
11/04	
11/11	First submission deadline of proposals
11/18	Final submission deadline of proposals
11/25	
12/02	Announcement of baselines of Task 1
12/09	
12/16	
12/23	Final exam
12/30	Final Project & presentation video due (Happy New Year!)

Enjoy Your Project!

If you have any questions/problems about the final project, please send an email to TAs' address (the title should begin with [Final Project]) or drop by the TA hour. Good luck!