

2025 Introduction to Artificial Intelligence

Machine Learning Project

- **Schedule**
 - 4/28 Competition starts
 - 5/11 23:59 Competition ends
 - 5/14 Report due (eeclass)

Outline

- Introduction
- Kaggle
- Report

Introduction

Data

Data contents

- # examples: 156,076 (training set)
- # attributes: 96
- # class (label 0/1): 2

Data

Data provided

- Training data (train.csv)
 - ▣ Data with labels, for model training
- Testing data (test.csv)
 - ▣ Data without labels, you need to predict the label of each data point
 - ▣ **Each data point has an id number**
 - ▣ The testing data is divided into two parts:
 - ▣ One for the public leaderboard
 - ▣ And the other for the private leaderboard
- Sample submission
 - ▣ The prediction you submit on Kaggle should follow the same format as the sample_submission.csv

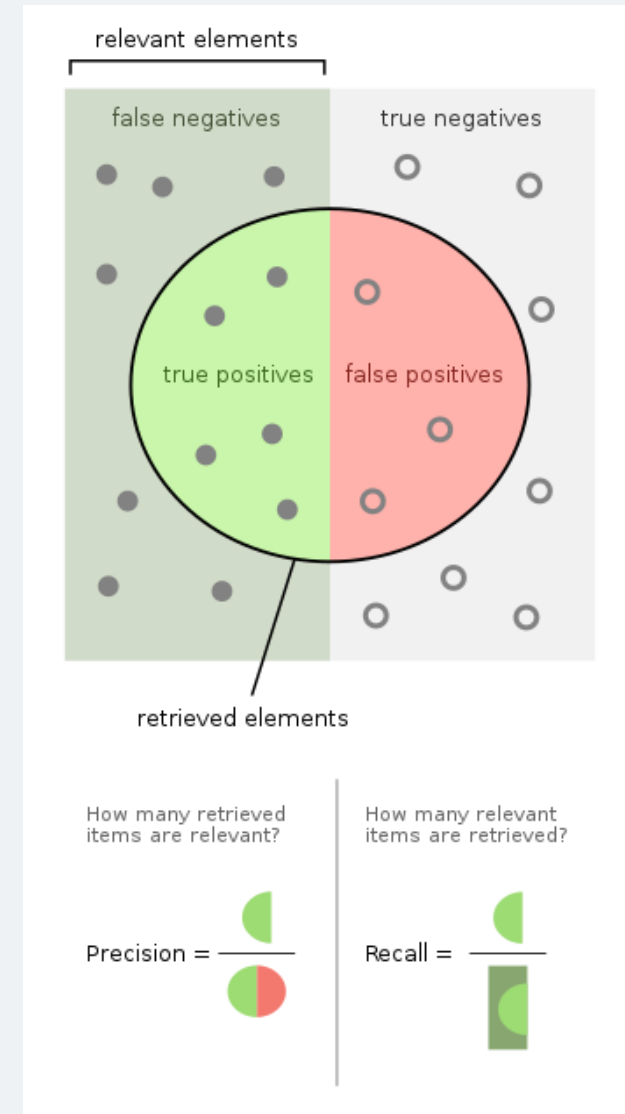
Performance indexes

F-score

- $Precision\ score = \frac{tp}{tp+fp}$
- $Recall\ score = \frac{tp}{tp+fn}$
- $F - score = 2 \cdot \frac{precision \cdot recall}{precision + recall}$

Macro F-score

- The metric used for Kaggle leaderboard
 - ▣ Can refer to the [code](#)





Kaggle

Join the Kaggle contest

Join through the Kaggle link

- The [Kaggle](#) link
 - ▣ Please contact TA 李松憶 if you do not have access to the contest

Join the Kaggle contest

Accept the rules

- Click "I understand and Accept" to start competing on the "Rules" page

Rules

Read the following instructions carefully before entering this competition!

- You should finish your homework on your own.
- You should not modify your prediction files manually.
- Do not share codes or prediction files with any living creatures.
- Do NOT use any approaches to submit your results more than 5 times a day.
- Do NOT search or use additional data or pre-trained models.

Your grade on this hw will get deducted if you violate any of the above rules.

Be careful and wish you good luck!

For more detailed regulations of this course, please visit the course page.

If you have any questions, we encourage you to post the question on the discussion page.

By clicking on "I Understand and Accept", you agree to be bound by the competition rules.

I Understand and Accept

Team name(Important!!)

Change your name to student ID

Overview Data Discussion Leaderboard Rules **Team** Submissions Settings

Your Team

Everyone that competes in a Competition does so as a team - even if you're competing by yourself. [Learn more.](#)

General

TEAM NAME

113062999|

This name will appear on your team's leaderboard position.

Team name(Important!!)

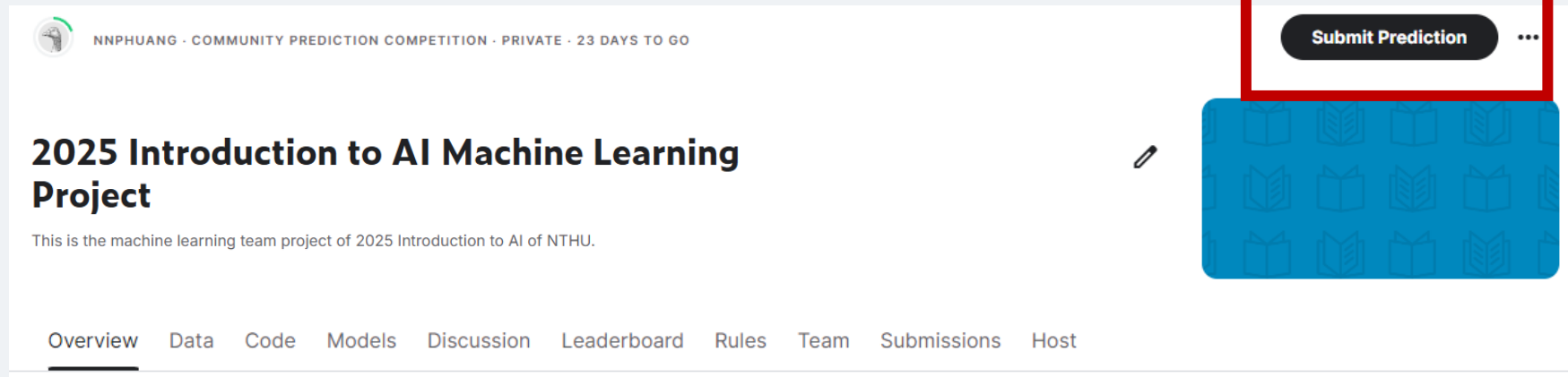
Anyone who doesn't follow these regulations will be revoked from the contest!

You need to do these before you make any submissions

Kaggle submission

Public leaderboard submission

- You can submit up to 5 times per day
- Your submission file must follow the format of sample_submission.csv



Kaggle submission

Private leaderboard submission

- Select 2 submissions as the private leaderboard submission
- The private leaderboard ranking will be released after the competition ends



Submissions

Select up to 2 submissions that will count towards your final leaderboard score. If less than 2 are selected, Kaggle will automatically select from your best scoring submissions. [Learn More](#)

☒ Auto-selection candidates [?](#)

All Successful Selected Errors

Recent ▾

Submission and Description	Public Score ?	Select
 sample_submission.csv Complete · 24s ago · example 2	0.50046	<input checked="" type="checkbox"/>
 sample_submission.csv Complete · 38s ago · example 1	0.50046	<input checked="" type="checkbox"/>

Report

Report

Things you should write in your report

- Methodology and data analysis
 - ▣ Clearly describe the methods and techniques you used
 - ▣ Include data processing steps, model choices, parameter settings, etc
- Describe how you improve the results
 - ▣ Explain how you optimized or enhanced your results
 - ▣ Examples: feature engineering, hyperparameter tuning, model comparison, etc

Format

- Save as a pdf file

Grading

Evaluation

- Report 70%
 - ▣ 30%: Awarded only if your model's performance exceeds baseline
- Kaggle rank 30%
 - ▣ Calculated based on ranking in public (15%) and private (15%) leaderboards
- No need to be stressed out about your grade 😊
 - ▣ If you aren't satisfied with your performance on the private leaderboard, you can still make up by writing an informative report

Contact us if you have problems...

Kaggle discussion page

- We strongly encourage you to post your question here
- You can also search the competition forums in case a similar question has been answered

Email

- Questions regarding Kaggle
 - ▣ 李松憶 lisong@gapp.nthu.edu.tw

