

캐글 리더보드 만들기

2022-2 인공지능 수업

1. 새로운 Competition 추가

The image shows the Kaggle homepage. On the left sidebar, the 'Create' button is highlighted with a red box and a circled '1'. Below it, the 'New Competition' option is also highlighted with a red box and a circled '2'. The main content area features a 'How to start: Choose a focus for today' section with three cards: 'Learn to compete on Kaggle', 'Take a short course', and 'Browse inspiring data and code'. Below this is a 'Newsfeed' section with a post by Jiwei Liu about competition solutions. On the right, there's a user profile for Juyoung Hong and a 'Novice' progress bar with a checklist of tasks.

① Create

② New Competition

How to start: Choose a focus for today
Help us make relevant suggestions for you

Learn to compete on Kaggle
Improve and test your skills
Get started →

Take a short course
Our courses are the fastest way to learn data science
Get started →

Browse inspiring data and code
Improve your data science projects
Get started →

Newsfeed

Jiwei Liu • Follow
created this topic a day ago ...

Recordings and Source codes of winners' solutions of this competition last year.
in the Open Problems - Multimodal Single-Cell Integration forum

As you might already know, this competition was hosted in 2021 NeurIPS as well. Winners presented their solutions in the workshop remotely. You can find the recording here:
<https://drive.google.com/file/d/1aQss-KyfYlZdrBQCh5joiXMITwpG5gdf/view?usp=sharing>

Source code of winning solutions can be found here: <https://github.com/openproblems-bio/neur...> See More

Takoi upvoted this topic 31

Juyoung Hong
Joined a year ago

Novice

- ✓ Run 1 notebook or script
- ✓ Make 1 competition submission
- ✓ Make 1 comment
- Give 1 upvote

Your Datasets +

mnist_gt

캐글 메인 >> Create >> New Competition

2. Competition 정보 설정



New to Creating a Competition?

Learn how to set up your own competition with our setup guide.

[Read Setup Guide](#)

배추 가격 예측

2022학년도 2학기 인공지능 텀프로젝트

◀ 제목, 구글 시트에 작성한 주제 작성

◀ 소제목, 양식 통일 !!

Competition Setup

Competition URL <http://www.kaggle.com/competitions/2022-ai-w3p1>

Choose a short URL. All links to your competition pages will start with this URL.

◀ URL 양식 통일 !! 2022-ai-tp-학번

Participant Access

Access Type ☐ Open to Everyone

Competition details are public on Kaggle. Anyone can join.

☒ Invitation Only

◀ 권한 설정 반드시 ★ Invitation Only ★

Competition details are public on Kaggle. Participation is limited to those with an invitation link. The link is available after you create the competition.



I confirm that I am creating a competition in accordance with Kaggle site terms, and will not offer a cash prize unless otherwise approved by the Kaggle Team

Create Competition

공란 모두 작성 후 하단 Create Competition 클릭

3. Overview 작성

Community Prediction Competition

선형분류를 이용한 MNIST 손글씨 분류
2022학년도 2학기 인공지능 실습문제 (5주차)

Hos

Overview

Data

Code

Discussion

Leaderboard

Rules

Team

My Submissions

...

Overview

Edit 눌러서 수정▶

Edit

Description


Evaluation

+ Add Page

교과목 정보

• 세종대학교 지능기전공학부 (최유경교수)

선형분류를 이용한 MNIST 손글씨 분류



본 데이터셋은 인공지능의 대표적인 손글씨 이미지 데이터셋인 MNIST입니다.
MNIST를 구성하는 숫자는 0~9까지 총 10개의 클래스로 구성됩니다. (상단 이미지 참고)

제출 시 주의사항

배추가격 예측문제

2022학년도 2학기 인공지능 텀프로젝트

문제를 풀기 위해 알아야할 배경지식 등 자유롭게 작성해주세요

▲ Description 예시
(참고하세요)

Description & Evaluation 작성

개설한 텀프로젝트에 대한 설명을 자세히 작성할 것

4. Data (Description)

Community Prediction Competition

선형분류를 이용한 MNIST 손글씨 분류
2022학년도 2학기 인공지능 실습문제 (5주차)

Host

Overview

Data

Code

Discussion

Leaderboard

Rules

Team

My Submissions

...

✓ Save successful.

Data Description

Edit

제공되는 파일 설명

- train.csv : 60,000개의 데이터가 있고, 각 행은 label과 784개의 픽셀값으로 구성 (60,000 x 785)
- test.csv : 10,000개의 데이터가 있고, 각 행은 784개의 픽셀값으로 구성 (10,000 x 784)
- sample_submit.csv : submission 파일 예시

Column name 과 정보

- label : 손글씨 숫자값 (0~9)
- 1×1 ~ 28×28 : 컬럼명이 해당하는 위치의 픽셀값

Data

This competition doesn't have any data attached to it yet.

Upload first version

Edit 눌러서 수정

Data Description 작성
(데이터 Column에 대한 설명을 자세히 작성할 것 !!)

4. Data (Upload)

Community Prediction Competition

선형분류를 이용한 MNIST 손글씨 분류

2022학년도 2학기 인공지능 실습문제 (5주차)

Host

Overview

Data

Code

Discussion

Leaderboard

Rules

Team

My Submissions

...

✓ Save successful.

Data Description

Edit

제공되는 파일 설명

- train.csv : 60,000개의 데이터가 있고, 각 행은 label과 784개의 픽셀값으로 구성 (60,000 x 785)
- test.csv : 10,000개의 데이터가 있고, 각 행은 784개의 픽셀값으로 구성 (10,000 x 784)
- sample_submit.csv : submission 파일 예시

Column name 과 정보

- label : 손글씨 숫자값 (0~9)
- 1×1 ~ 28×28 : 컬럼명이 해당하는 위치의 픽셀값

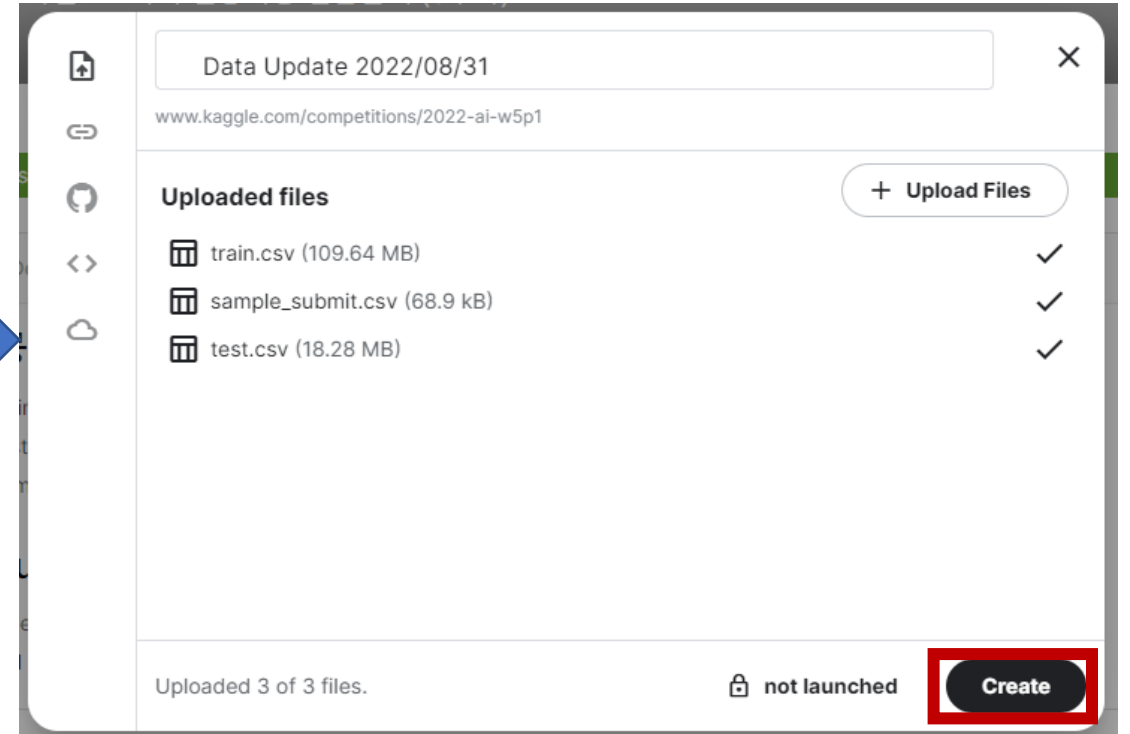
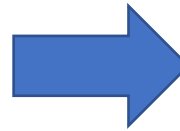
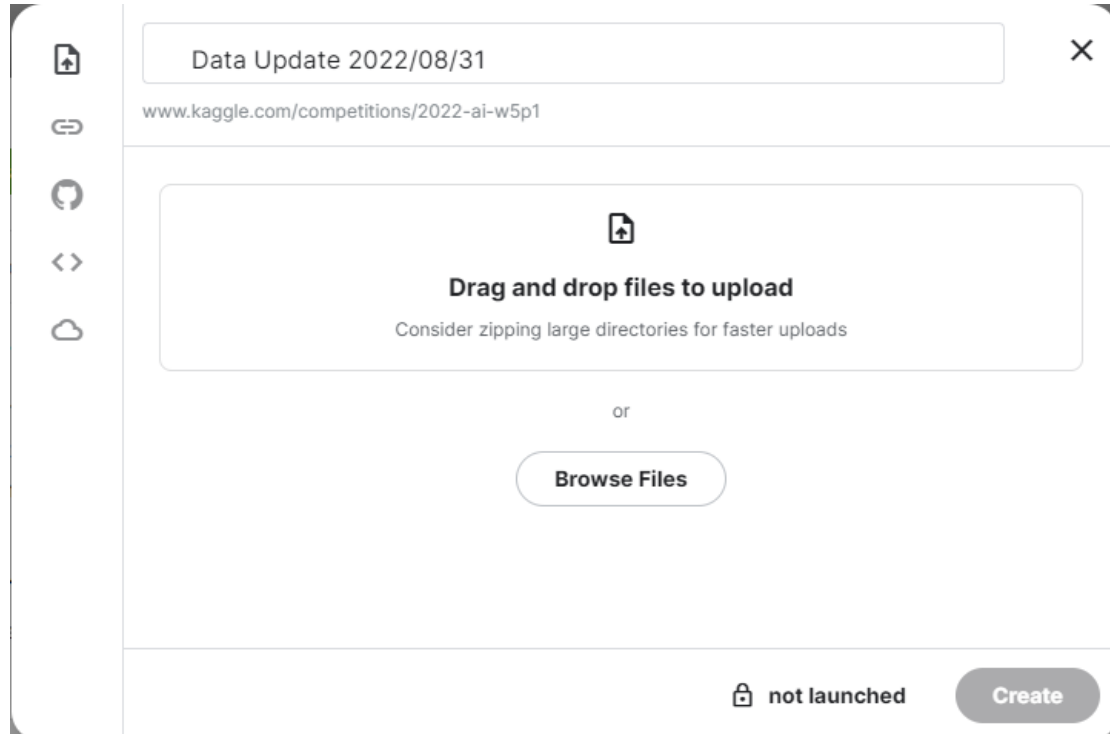
Data

This competition doesn't have any data attached to it yet.

Upload first version

Upload first version 클릭
(캐글 리더보드 개설을 위한 데이터셋을 업로드하는 과정입니다)

4. Data (Upload)

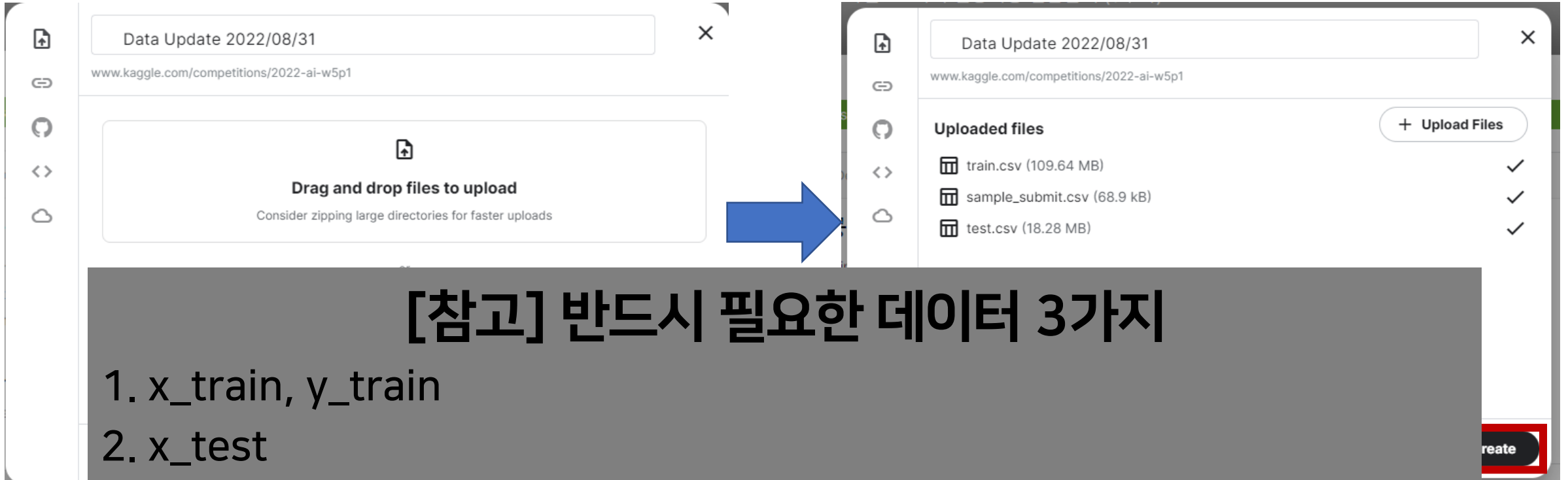


[제공할 데이터 첨부하기]

- x_train.csv
- y_train.csv
- x_test.csv
- sample_submit.csv

업로드 후 Create 버튼

4. Data (Upload)



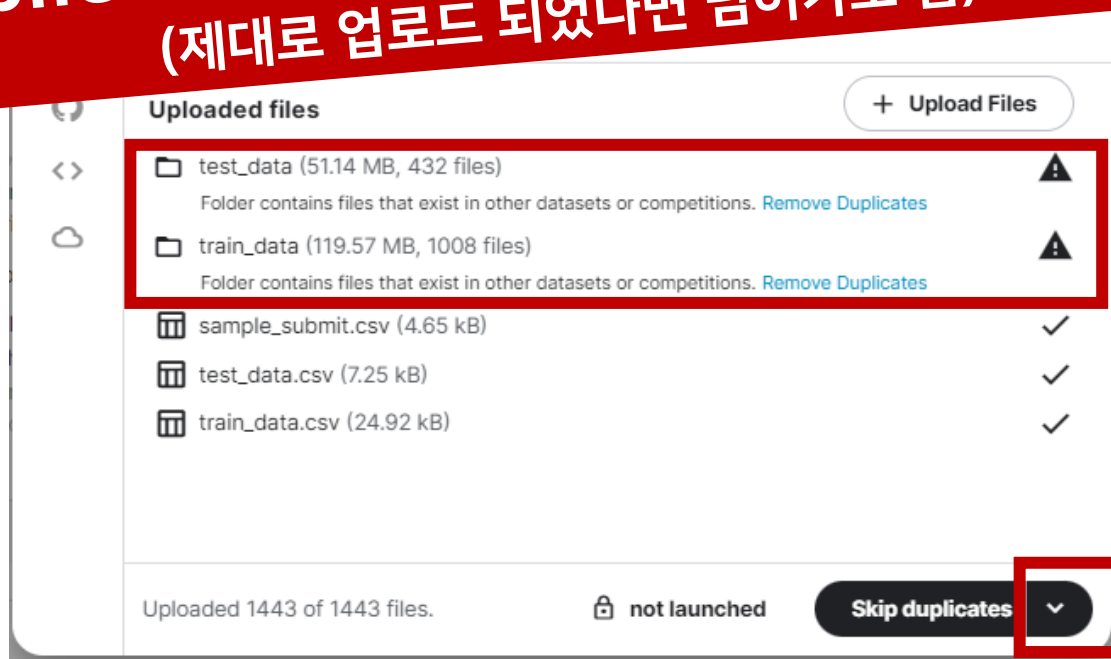
[참고] 반드시 필요한 데이터 3가지

1. x_train, y_train
2. x_test
3. sample_submit

- Sample은 특히!! 반드시 필요합니다!!
- y_test에서 정답값을 지우고 올리면 됩니다.
- 이때, 정답값인 y_test를 올리지 않도록 주의해주세요!!
- 모르겠다면 기존 실습문제에서 제공하는 데이터셋을 참고해주세요~

4. Data (Upload)

만일 Duplicate 경고로 인해 업로드에 실패할 경우
(제대로 업로드 되었다면 넘어가도 됨)



이미 캐글에 올라와있는 데이터라는 Duplicate 경고가 뜨는 경우,
전체 파일을 하나로 압축한 뒤 **zip 파일 하나만 업로드** 하세요

5. Evaluation

Community Prediction Competition

선형분류를 이용한 MNIST 손글씨 분류
2022학년도 2학기 인공지능 실습문제 (5주차)

Host

Overview

Data

Code

Discussion

Leaderboard

Rules

Team

My Submissions

...

Host Controls

Settings

Images

Manage Hosts

Evaluation

Sandbox Submissions

All Submissions

Launch Checklist

User submissions are evaluated by comparing their **Submission CSV** to the ground truth **Solution CSV** with respect to a chosen **Scoring Metric**, for example the Root-mean-squared Error (RMSE) metric.

Each **Scoring Metric** will expect the Solution and Submission CSVs to have certain columns, defined in a header row. For example, RMSE expects an **Id** column identifying the row and an **Expected** column specifying the true value which competitors are attempting to predict. Your Solution CSV's header row does not need to use the exact column names, you will be able to map your column names onto the Metric's expected columns.

To ensure that competitors cannot game the system, the rows in the solution file are sampled into **Public** and **Private** rows. Public scores are shown on the public leaderboard, while only the private scores are used to determine the competition winner. The sampling can be done manually by including a **Usage** column with values of **Public**, **Private**, or **Ignored**. Otherwise, rows will automatically be sampled into these buckets.

Scoring Metric ⓘ

Choose a scoring metric to use...

Solution File ⓘ

리더보드 점수 측정을 위한 y_test 값 넣어주기

5. Evaluation

Images

Manage Hosts

Evaluation

Sandbox Submissions

All Submissions

Launch Checklist

Scoring Metric ⓘ

CategorizationAccuracy Categorization Accuracy

Solution File ⓘ

Download Raw

✕ Remove

Solution Mapping ⓘ

Solution Column	Expected Column
id	Id (String) ▼
label	Category (Stri... ▼

Public Leaderboard Percentage 100

Set Solution Mapping

Solution Sampling

Your submission should have 10000 rows, each with 2 columns.

Type	Rows	Percentage
Public Leaderboard	0	0%
Private Leaderboard	0	0%
Ignored	10000	100%
Total	10000	100.0%

◀ 평가 메트릭 선택

◀ Y_test.csv 업로드

◀ Y_test.csv 에 대한 Column 정보 매핑

◀ Public Leaderboard Percentage 100으로 변경

설정 완료 후 Set solution mapping 클릭

6. Check Sample Submit

Host

Overview

Data

Code

Discussion

Leaderboard

Rules

Team

My Submissions

...

Settings

Images

Manage Hosts

Evaluation

Sandbox Submissions

All Submissions

Launch Checklist

Solution Mapping

id

label

Id (String)

Category (Stri...

Public Leaderboard Percentage100

Set Solution Mapping

Solution Sampling

Your submission should have 10000 rows, each with 2 columns.

Type	Rows	Percentage
Public Leaderboard	10000	100%
Private Leaderboard	0	0%
Total	10000	100.0%

Sample Submission

sample_submit.csv (68.9 kB)

CSV Parsed Remove

Sample Mapping

Submission Colu...	Expected Column
id	Id (String)
label	Category (Stri...

Set Submission Mapping

◀ sample.csv 업로드

◀ sample.csv 컬럼 정보 매핑

매핑 후 set submission mapping 클릭

7. Baseline 설정

이제부터는 정의한 데이터셋을 가지고 **적절한 베이스라인**을 설정해봅시다.
(실습 문제 풀듯 직접 데이터셋을 가지고 문제를 풀어보면 됩니다)

베이스라인을 설계하는 데에는 **수업시간에 배운 방법론만** 사용할 수 있습니다.

베이스라인이 원복가능하도록 아래 **랜덤 시드 고정 코드**를 추가해주세요.
(만일 다시 돌렸을 때 **성능 원복이 불가능한 경우 0점 처리**됩니다.)

```
import random
import torch

seed = 1
random.seed(seed)
np.random.seed(seed)
torch.manual_seed(seed)
if torch.cuda.is_available():
    torch.cuda.manual_seed(seed)
    torch.cuda.manual_seed_all(seed)
torch.backends.cudnn.benchmark = False
torch.backends.cudnn.deterministic = True
```

7. Baseline 설정

Community Prediction Competition

선형분류를 이용한 MNIST 손글씨 분류
2022학년도 2학기 인공지능 실습문제 (5주차)

Host Overview Data Code Discussion Leaderboard Rules Team My Submissions ...

Host Controls

Settings
Images
Manage Hosts
Evaluation
Sandbox Submissions

✓ baseline_1.csv (68.9 kB)

Name	Submitted	Wait time	Execution time	Score
baseline_1.csv	a minute ago	0 seconds	0 seconds	0.89220
Complete				

Name		Date	Public	Private	Benchmark
baseline_1.csv		a minute ago	0.89220	0.89220	<input checked="" type="checkbox"/>
sample_submit.csv		3 minutes ago	0.00000	0.00000	<input type="checkbox"/>
sample_submit.csv		4 minutes ago	0.00000	0.00000	<input type="checkbox"/>
sample_submit.csv		6 minutes ago	Failed	Failed	
sample_submit.csv		11 minutes ago	Failed	Failed	

< Sandbox Submission >

제작자는 Sandbox Submission에 본인의 csv 파일을 제출할 수 있습니다.

여기에 제출할 경우 제작자의 점수는 리더보드에 기록되지 않습니다.
또한 리더보드 제작자는 csv파일을 횟수 제한 없이 마음껏 제출 할 수 있습니다.

이제 인공지능 수업시간에 배운 다양한 모델을 사용하여 csv 파일을 제출해본 뒤, 점수를 보고 적절한 베이스라인을 설정해봅시다.

7. Baseline 설정


Community Prediction Competition

선형분류를 이용한 MNIST 손글씨 분류
2022학년도 2학기 인공지능 실습문제 (5주차)

[Host](#) [Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [My Submissions](#) [...](#)















Host Controls

Settings
Images
Manage Hosts
Evaluation
Sandbox Submissions
All Submissions
Launch Checklist



✓ baseline_1.csv (68.9 kB)

Name	Submitted	Wait time	Execution time	Score
baseline_1.csv	a minute ago	0 seconds	0 seconds	0.89220
Complete				

Name		Date	Public	Private	Benchmark	
baseline_1.csv			a minute ago	0.89220	0.89220	<input checked="" type="checkbox"/>
sample_submit.csv			3 minutes ago	0.00000	0.00000	<input type="checkbox"/>
sample_submit.csv			4 minutes ago	0.00000	0.00000	<input type="checkbox"/>
sample_submit.csv			6 minutes ago	Failed 	Failed 	
sample_submit.csv			11 minutes ago	Failed 	Failed 	

베이스라인으로 설정하고
싶은 스코어에 체크

8. 리더보드 관련 설정

Host Overview Data Code Discussion Leaderboard Rules Team My Submissions ...

Settings

Images

Manage Hosts

Evaluation

Sandbox Submissions

All Submissions

Launch Checklist

Title 선형분류를 이용한 MNIST 손글씨 분류

Competition URL <http://www.kaggle.com/competitions/2022-ai-w5p1>
Choose a short URL. All links to your competition pages will start with this URL.

Brief Description 2022학년도 2학기 인공지능 실습문제 (5주차)

Organization Search Organizations...

Deadline 08/30/2022 11:59 PM
Competitions typically last two to four months. Date and time are in UTC.

Private Leaderboard Release Date
You may specify a release date up to 7 days after the deadline. Updating the deadline will also update this. Leave unset to release at the deadline.

Notebook Publishing ☒ Enable Publishing

Participant Access

Access Type ☐ Open to Everyone
Competition details are public on Kaggle. Anyone can join.

☒ Invitation Only
Competition details are public on Kaggle. Participation is limited to those with an invitation link.

URL for Sharing <https://www.kaggle.com/t/af7769902f0342fab5561fd8c090730>

Participant Submissions

Maximum Daily Submissions 20
Participants will need to wait until the next UTC day after submitting the maximum number of daily submissions.

Scored Private Submissions 20
The number of submissions eligible for the final private leaderboard. Users can hand-select the eligible submissions, or will otherwise default to the best public scoring submissions.

Save Changes

◀ 데드라인 변경

* 11/27/2022 03:00 PM 으로 설정해주세요

◀ 리더보드는 반드시 이 링크로 공유해주세요

◀ 최대 제출 횟수 20으로 변경

9. 리더보드 호스트 추가

Community Prediction Competition

선형분류를 이용한 MNIST 손글씨 분류

2022학년도 2학기 인공지능 실습문제 (5주차)

[Host](#) [Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [My Submissions](#) ...

Host Controls

- Settings
- Images
- Manage Hosts**
- Evaluation
- Sandbox Submissions
- All Submissions
- Launch Checklist

Hosts

Juyoung Hong

Search Users [Search Users](#)

Search Results

Lim Guen Taek

[+ Add](#)



Hosts

Juyoung Hong

Lim Guen Taek [✕ Remove](#)

y2sman [✕ Remove](#)

이승현 [✕ Remove](#)


Yukyung Choi [✕ Remove](#)

Search Users [Search Users](#)

Search Results ...

[호스트 추가할 이름]
yukyungchoi, juyounghong, y2sman,
leeseunghyeon99, Lim Guen Taek

10. Rule 변경


 Community Prediction Competition

선형분류를 이용한 MNIST 손글씨 분류

2022학년도 2학기 인공지능 실습문제 (5주차)

[Host](#) [Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [My Submissions](#) [...](#)

✓ Save successful.

 You have accepted the rules for this competition. Good luck!

Rules [Edit](#)

This is a page where you can include rules that participants must accept before joining. You may wish to include rules like:

- Don't cheat!
- Apply yourself!
- Have fun!

Edit 누르고 저장
(내용 수정 안해도 되지만
반드시 바꿔야 리더보드가 Launch 됨)

11. CheckList 확인 후 Launch

[Host](#) [Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [My Submissions](#) [...](#)

i This competition is not yet live; only competition hosts can currently view it.

Host Controls

Settings

Images

Manage Hosts

Evaluation

Sandbox Submissions

All Submissions

Launch Checklist

☒ You have uploaded a solution

☒ You have mapped your solution columns to the metric's expected columns.

☒ You have mapped your submission columns to the metric's expected columns.

☒ You have made a successful sandbox submission



☒ You have uploaded at least one data file

☒ The deadline is in the future

☒ You have published all required pages

☒ You have no remaining template content

https://www.kaggle.com/t/af7769902f0342fabb5561fd8c090730



This is a URL that can be shared around and re-used. Anyone who visits this link will be able to participate in the competition.

Your competition passes our automated checks.

You may launch the competition when you are ready.

Set Future Launch Date

Launch Competition Now

[주의사항]

다른 사람들의 리더보드를 공격할 수 있는 기간은 11월 15일 00:00 부터입니다.

만일 리더보드 공개 전에 공격 시도가 있을 경우,
리더보드 제작자와 공격자 모두 0점 처리 됩니다.

리더보드 제작자는 자신의 리더보드가 다른 학생들에게 공개되지 않도록 주의해주세요.

또한 코드 공유 적발 시 0점 처리됩니다.

리더보드 제작자는 자신의 베이스라인을 설명할 수 있어야 합니다.

또한 베이스라인 설계 시, 시간이 너무 학습 시간이 오래걸리는 사전 학습 모델 사용은 지양바랍니다.

(Tip! 베이스라인이 너무 높게 설정하여 그 누구에게도 공격당하지 못할 경우, 디펜스의 기회가 없어질 수 있습니다~)

수고하셨습니다 😊