

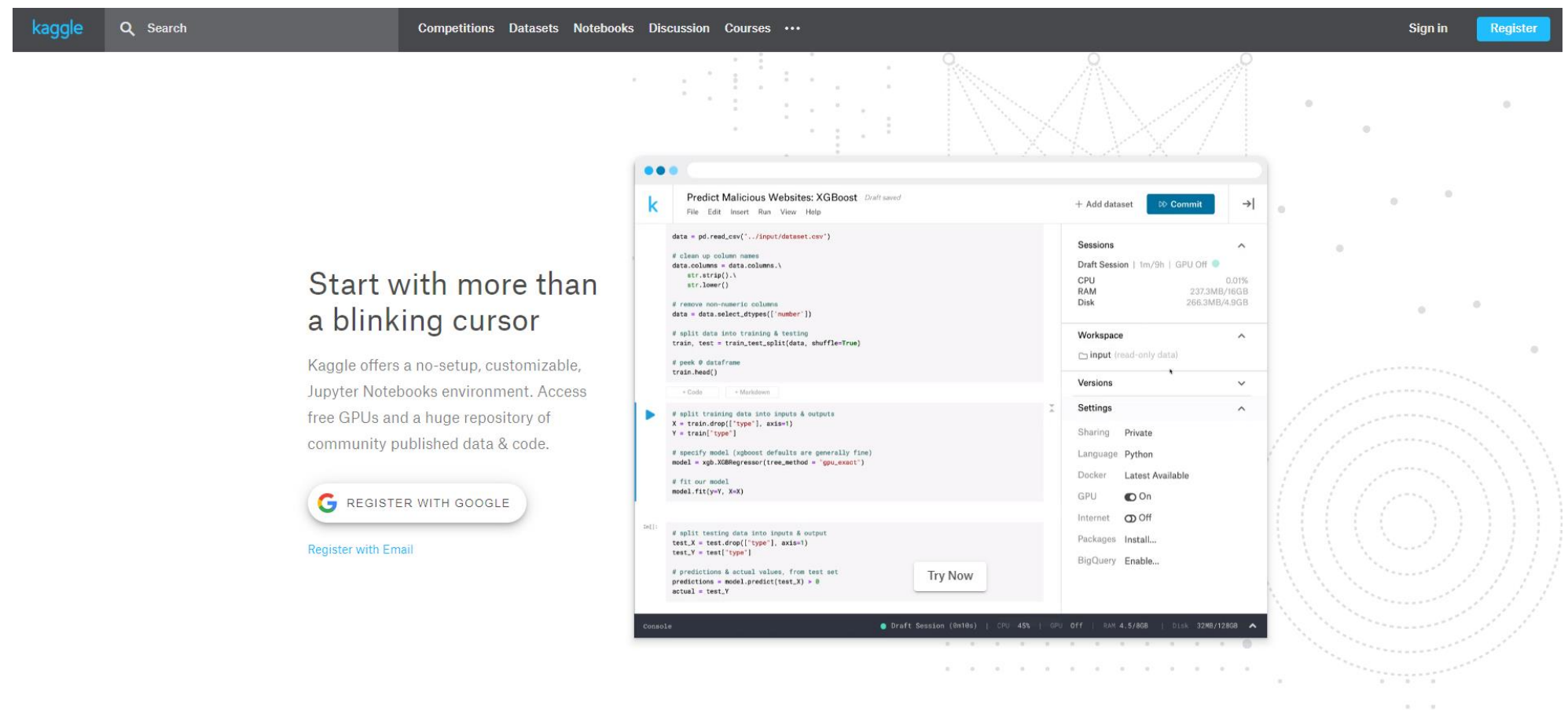
Kaggle 教學

楊鎔篷 吳承翰

Kaggle 簡介

Kaggle 是一個資料科學競賽平台，每個人都可在其上發布資料集舉辦比賽，也可以參與其他人舉辦的比賽，期望透過競賽排名，集合眾人力量找出最好的模型。

(<https://www.kaggle.com/>)



The image displays the Kaggle website interface, featuring a dark header with the Kaggle logo, a search bar, and navigation links for Competitions, Datasets, Notebooks, Discussion, and Courses. The main content area is titled "Start with more than a blinking cursor" and describes the platform's features: a no-setup, customizable Jupyter Notebook environment with free GPUs and a large repository of community data and code. A "REGISTER WITH GOOGLE" button and a "Register with Email" link are provided. The right sidebar shows session details, workspace, versions, and settings. The central area displays a Jupyter Notebook titled "Predict Malicious Websites: XGBoost" with Python code for data cleaning, splitting, and model training. A "Try Now" button is visible at the bottom right of the notebook area.

Start with more than a blinking cursor

Kaggle offers a no-setup, customizable, Jupyter Notebooks environment. Access free GPUs and a huge repository of community published data & code.

[REGISTER WITH GOOGLE](#)

[Register with Email](#)

Predict Malicious Websites: XGBoost

```
data = pd.read_csv('../input/dataset.csv')

# clean up column names
data.columns = data.columns.\
    str.strip().\
    str.lower()

# remove non-numeric columns
data = data.select_dtypes(['number'])

# split data into training & testing
train, test = train_test_split(data, shuffle=True)

# peek @ dataframe
train.head()

# split training data into inputs & outputs
X = train.drop(['type'], axis=1)
Y = train['type']

# specify model (xgboost defaults are generally fine)
model = xgb.XGBRegressor(tree_method='gpu_exact')

# fit our model
model.fit(y=Y, X=X)

# split testing data into inputs & output
test_X = test.drop(['type'], axis=1)
test_Y = test['type']

# predictions & actual values, from test set
predictions = model.predict(test_X)
actual = test_Y
```

Sessions
Draft Session | 1m/9h | GPU Off
CPU 0.01%
RAM 237.3MB/16GB
Disk 266.3MB/4.9GB

Workspace
input (read-only data)

Versions

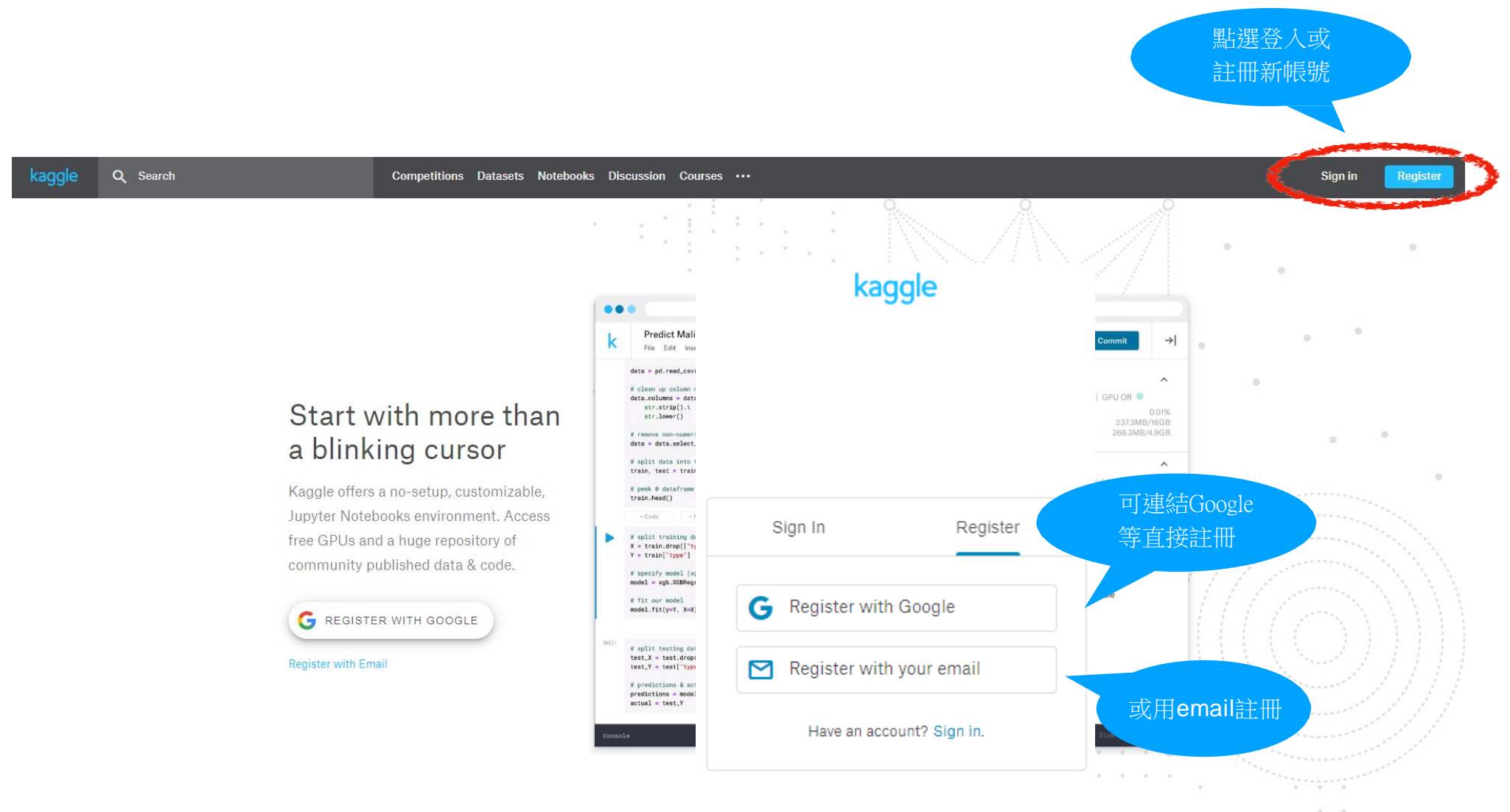
Settings
Sharing Private
Language Python
Docker Latest Available
GPU On
Internet Off
Packages Install...
BigQuery Enable...

Console
Draft Session (9m18s) | CPU 45% | GPU Off | RAM 4.5/8GB | DISK 32MB/12GB

帳號註冊

- 點選首頁右上角的Sign in或Register，選擇登入或註冊新帳號：

註冊：連結google帳號 or 使用email註冊。

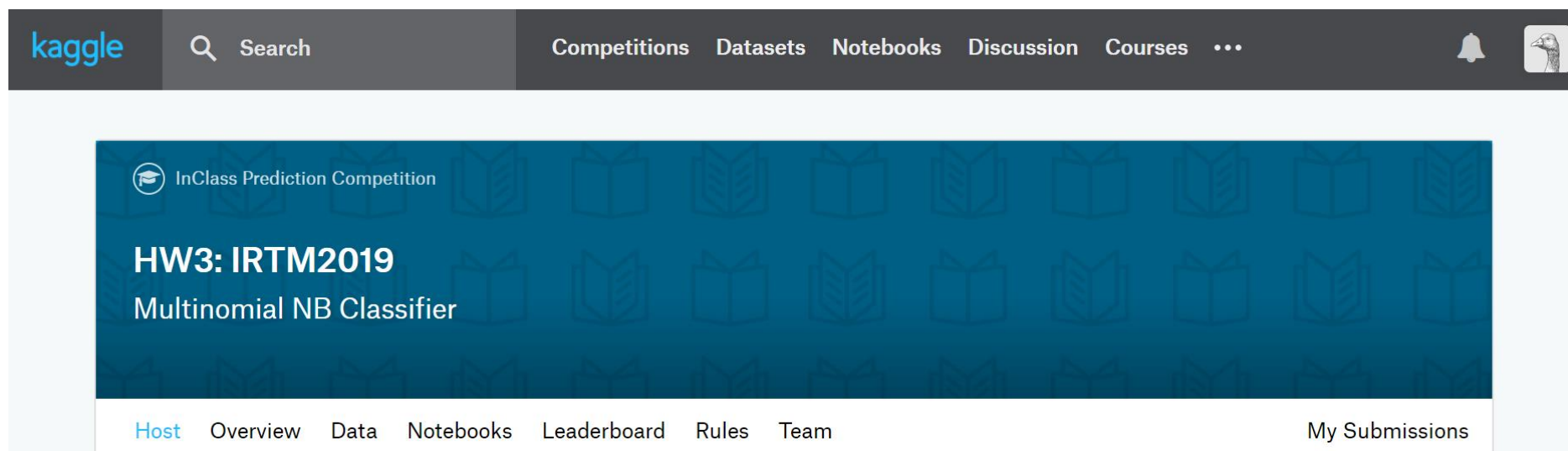


Kaggle 競賽

- kaggle競賽分為兩類：general、inclass。
- 作業三採inclass方式，即只有知道連結的人才能進入比賽。
- 希望同學不要任意把連結給其他非修課相關的人！
- 連結如下：

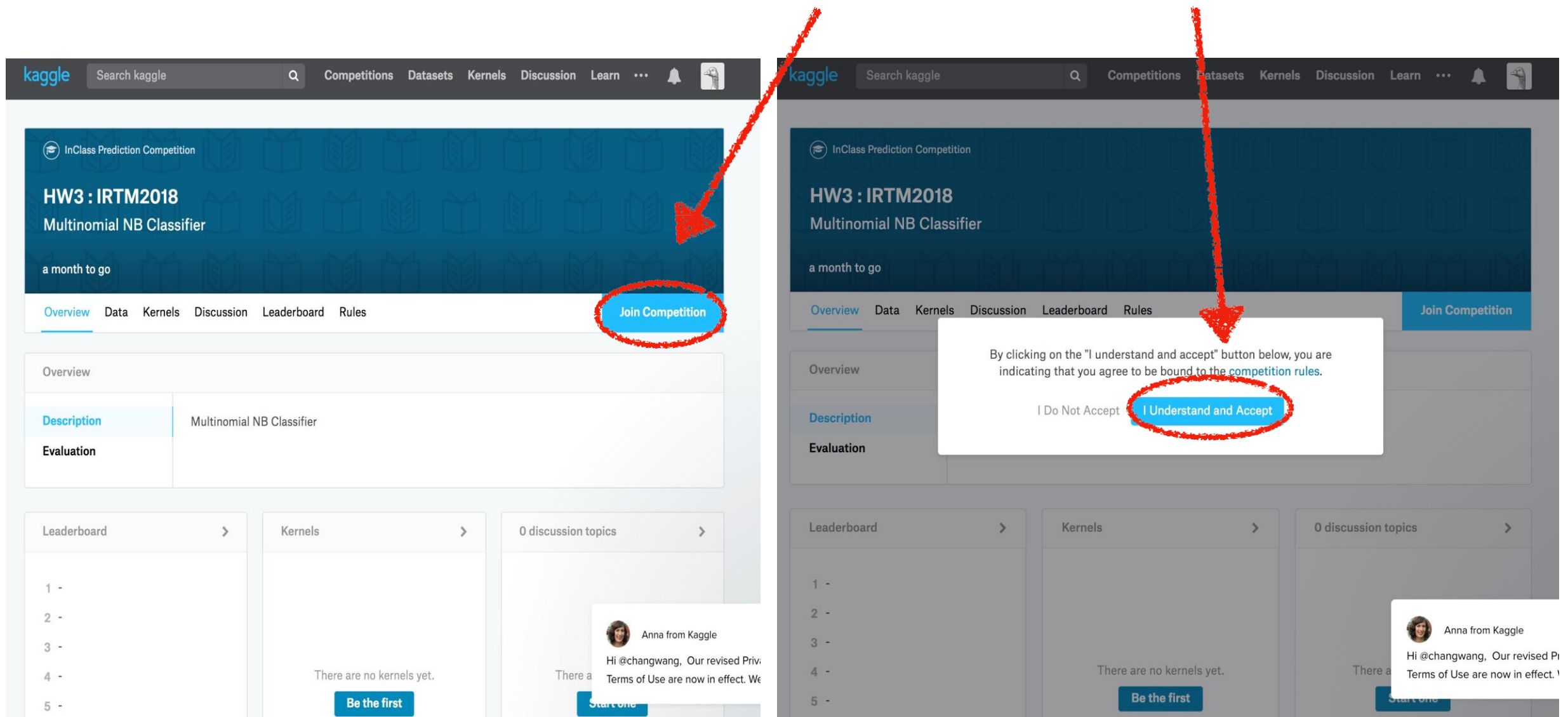
<https://www.kaggle.com/t/65ff9ec3c9a64a19bd458419dcb6b5b3>

(如果尚未登入，點選連結會導向登入頁面，但登入後有可能直接跳回首頁，此時建議先登入再點選連結，就可順利進入比賽頁面)



參加競賽

- 成功進入競賽畫面後，點選右上角的Join Competition，再同意競賽規則即可。



上傳注意事項

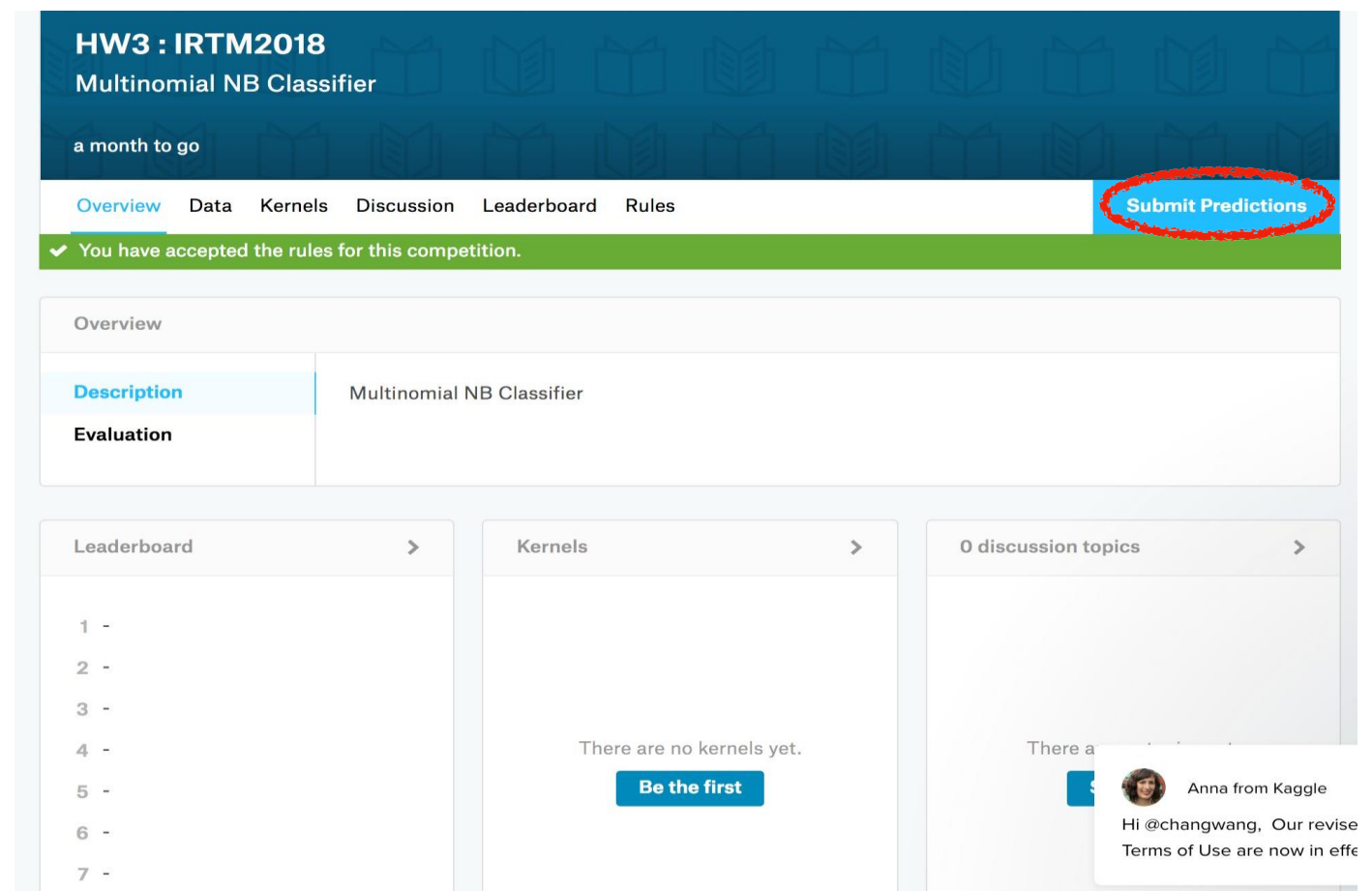
- 成功參加比賽後，頁面右上角即會變成Submit Predictions

- 關於上傳結果有以下事項需要注意：

- 注意上傳格式！

- 注意每日上傳次數上限5次！

- 請更改隊名為學號！



上傳格式

- 請使用**CSV**檔，其他檔會無法上傳。
- 需有兩欄，分別為Id和Value。
- Id直接用docid即可，Value為分類結果。

sub

Id	Value
1	0
2	0
3	0
4	0
5	0

計分方式

- 排名採用micro-averaged F1-score
 - 右圖結果為0.6

sub	
Id	Value
1	1
2	2
3	3
4	0
5	0

sol	
Id	Value
1	1
2	2
3	3
4	4
5	5

上傳預測結果

- 點選右上角Submit Predictions，就進入上傳頁面。
- 若上傳等很久，可以重開頁面再上傳，未上傳成功不會計入上傳次數。
- 設次數限制是怕大家無限上傳去試出答案。

The screenshot shows the submission interface with several annotations in blue speech bubbles and red circles:

- 點選可以更改隊名**: Points to the `TA_test` team name in the header.
- 剩餘上傳次數和reset時間**: Points to the text "You have 5 submissions remaining today. This resets 14 hours from now (00: 00 UTC)".
- 拖入檔案**: Points to the "Upload Submission File" button.
- 等待上傳完成**: Points to the progress bar showing 100% completion.
- Make Submission**: A red circle highlights the "Make Submission" button at the bottom.

The interface includes the following elements:

- Header**: "Make a submission for `TA_test`"
- Status Bar**: "You have 5 submissions remaining today. This resets 14 hours from now (00: 00 UTC)." (Note: "resets" is misspelled as "reset" in the image).
- Step 1: Upload submission file**
 - Upload Submission File**: A button with an upload icon.
 - Progress Bar**: Shows "ans.csv (33 B)" with a status of "Complete" at "100%" and a size of "33 B".
 - File Format**: "Your submission should be in CSV format. You can upload this in a zip/gz/rar/7z archive, if you prefer."
 - Number of Predictions**: "We expect the solution file to have 5 prediction rows. This file should have a header row. Please see sample submission file on the [data page](#)."
- Step 2: Describe submission**
 - Rich Text Editor**: Includes a toolbar with bold, italic, link, unlink, code, image, list, and other formatting options. A note says "Styling with Markdown supported".
 - Description Box**: A text area with the placeholder "Briefly describe your submission."
- Make Submission**: A blue button at the bottom.

預測結果評分

a month to go

Overview Data Kernels Discussion **Leaderboard** Rules [Submit Predictions](#)

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
sub.csv	14 minutes ago	0 seconds	0 seconds	0.00000

Complete

[Jump to your position on the leaderboard](#)

Public Leaderboard Private Leaderboard

This leaderboard is calculated with approximately 40% of the test data.
The final results will be based on the other 60%, so the final standings may be different.

[Raw Data](#) [Refresh](#)

#	Δ1w	Team Name	Kernel	Team Members	Score ?	Entries	Last
1	new	TA_test			0.00000	1	14m

Your Best Entry ↑

Your submission scored 0.00000, which is not an improvement of your best score. Keep trying!

Anna from Kaggle
Hi @changwang, Our revised F

等待系統計算分數

顯示你目前的排名

有時候系統負荷過大，可能評分很久，可以嘗試重開頁面，看看是否有評分成功，即有顯示出新的分數，若失敗的話不會有紀錄，也就不會用到上傳次數，重新再上傳即可。

更改隊名

Manage Team



Team Name

TA_test

Save Team Name

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


Team Members



schin-gu-ba (you)

Leader

Invite Others

 Merge with other teams or invite users to your team by their team name

You are unable to merge with other teams on this competition.

輸入新隊名

原本註冊的display name

- kaggle提供組隊功能，但這次有限制只能一人一隊。
- 為了方便掌握大家的上傳結果，隊名開頭請一律用學號，底線後則不限制。
- 例: r06725123_xxxx
- 輸入新隊名後按Save Team Name即可