
Machine Learning Final Project

— ntu.mlta@gmail.com —

Outline

- Projects
- Schedule
- Scoring
- Policy
- Report
- Github
- Bonus

Projects: Kaggle & Driven Data

- 俄羅斯房地產
[[Competition Site](#)] [[Slides](#)]
- 登革熱案例數預測
[[Competition Site](#)] [[Slides](#)]
- 抽水機狀況預測
[[Competition Site](#)] [[Slides](#)]

Schedule

DATE	TODO
5/4	Final Project Releases
5/14 21:00:00	Team Makeup (Stage One) Release Simple Baseline
5/18	5/19 停修截止
5/25 21:00:00	Beyond Simple Baseline (Stage Two) Release Strong Baseline
6/1	Keep on moving moving.
6/8 21:00:00	Beyond Strong Baseline (Stage Three)
6/15 21:00:00	Top 10% Teams are Chosen
06/22	Top 10% Teams Presentation
06/28 (Wed) 21:00:00	Platform Ranking Deadline
06/30 21:00:00	Github (Report, src) Deadline
06/30 23:30:00	組內互評 Deadline

Scoring (45/40)

- Report : 20
- Team Makeup before Stage One (5/14): 3
- Beyond Simple Baseline before Stage Two (5/25): 2
- Beyond Strong Baseline before Stage Three (6/8): 3
- Beyond Simple Baseline before Platform Rank Deadline (6/28): 2
- Beyond Strong Baseline before Platform Rank Deadline (6/28): 3
- Ranking: 4
- Github: 4
- 組內互評: 4

Scoring details

- Team makeup (3%)
 - 填好組隊表單 (2~4人)
 - 在Driven Data/Kaggle 比賽上組成一隊 (包含所有的組員)
 - 上傳任意答案, 使自己組別出現在 leaderboard上面, Before Stage One 5/14 21:00:00
 - Driven Data平台命名不可重複, 請同學在隊名後面加上任意一個 A~Z或0~9的字母
- 如果你 “Beyond Simple Baseline before **Stage Two**” 代表你也會 “Beyond Simple Baseline before Driven Data Rank Deadline” 所以你會得到 2+3 分。同理Strong baseline。
- 可同時進行多個題目, 但助教只會依照最後繳交報告的**一個**題目評分。
 - 如果在題目 A “Beyond Simple Baseline before Stage Two” 但是你最後繳交的報告是題目 B, 則你不會拿到題目 A的任何分數, 一切以題目 B為準。
 - 中途換題目會使你拿不到前面 Stage的分數, 建議大家盡早決定一個題目。
 - 組隊的分數要在網站中完成組隊才算成功, 因此每個比賽的組隊是獨立計分。
- 提醒: 可以先在每一個題目都組隊完成, 如果等到公布Simple baseline才組隊選擇題目會害你拿不到組隊的分數。

Scoring details (cont)

- Ranking (4%)
 - beyond 30% : **4**
 - beyond 50% : **3**
 - beyond 70% : **2**
 - beyond sample submission : **1**
- 所有的Kaggle/Driven Data分數都是依據Public Set。
(beyond baseline、ranking判定等)

Policy

- Kaggle/Driven Data Team Name:
“ NTU_隊長學號_隊伍名稱 ”
隊長學號請小寫
- Team formed before stage one
- Feel free to use any toolkit.
- Abide by the rules of the competition.
- No cheating.
- Team sheet:
<https://goo.gl/forms/dWrGWf1Mkk46sHzn1>

Report -- Rules

- 12 號字
- 單行間距
- 至少 6 頁, 至多 12 頁 (含圖表)
- 報告中不須附上 code
- 請以中文撰寫

Report -- Contents

- **請清楚標示你選的題目 (1 or 0)**
- Team name, members and your work division (1)
- Preprocessing/Feature Engineering (3)
- Model Description (At least two different models) (7)
- Experiments and Discussion (8)

Github -- what it should contain?

- Team leader's github account !
- Only master, no branch.
- Directory "final" should contain:
 - a directory "src" with the source codes you write
 - a PDF file "Report.pdf" which is your final project report

ML2016/final

```
|
|----- /src
|         |----- codes
|         |_____ ...
|
|----- Report.pdf
|
|_____ any other files
```

BONUS

- If you are the top 10% team in the kaggle competition, you are welcomed to give a presentation on 06/22
- Presentation time: 10 minutes
- Bonus score for presentation: 2 point
- The presentation would be voted.

The top three presentation teams can get additional bonus score:

1st -> 3

2nd -> 2

3rd -> 1

Deadline

- Satge ONE deadline: 5/14 21:00:00
 - Stage TWO deadline: 5/25 21:00:00
 - Stage THREE deadline: 6/08 21:00:00
 - Platform deadline: 6/28 (三) 21:00:00
 - Github deadline: 6/30 (五) 21:00:00
 - 互評 deadline: 6/30 (五) 23:30:00
-
- Platform Score會依照每一個stage及此投影片所公布的時間為準，不需理會Kaggle/Driven Data上設定之deadline
 - Late : <https://goo.gl/uNuQ64>
(70% discount per day. At most 48 hours late)

QAO

Question and Questions?