

# 競賽測試集上傳須知

## 競賽說明

本次競賽共會提供兩筆測試集，第一筆測試集會在初賽開始時 (2021/05/10 12:00) 公布，共 10000 筆音檔，其預測成績會公布在 public leaderboard 上。第二筆測試集會在初賽結束前 (2021/06/08 12:00) 公布，共 20000 筆音檔，其預測成績會在比賽結束後隔天 (2021/06/11 12:00) 公布在 private leaderboard 上，晉級決賽之依據將以 private leaderboard 之排名以及繳交之審核資料為主。

## 測試集說明

| 測試集    | 數量    | 公佈日期       | 音檔資訊                               |
|--------|-------|------------|------------------------------------|
| 第一筆測試集 | 10000 | 2021/05/10 | 單聲道 (Mono), 8kHz, 16 bits PCM, wav |
| 第二筆測試集 | 20000 | 2021/06/08 | 單聲道 (Mono), 8kHz, 16 bits PCM, wav |

## 上傳格式

上傳時需將測試之結果存成 csv 檔，需符合以下之格式

```
Filename,Barking,Howling,Crying,COSmoke,GlassBreaking,Other  
[Filename],[prob_0],[prob_1],[prob_2],[prob_3],[prob_4],[prob_5]
```

第一列依序為 Filename, Barking, Howling, Crying, COSmoke, GlassBreaking, Other。  
從第二列開始為測試資料的檔名及各類別之預測機率，相加需等於1。以下為範例：

```
1 Filename,Barking,Howling,Crying,COSmoke,GlassBreaking,Other  
2 public_00001,0.73,0.04,0.03,0.05,0.14,0.01  
3 public_00002,0.03,0.01,0.01,0.06,0.88,0.01  
4 public_00003,0.03,0.01,0.01,0.06,0.88,0.01  
5 public_00004,0.03,0.01,0.01,0.06,0.88,0.01  
6 public_00005,0.03,0.01,0.01,0.06,0.88,0.01
```

註：上傳總筆數需為 30000 筆，前 10000 筆為 public test data 的預測結果，後 20000 筆為 private test data 的預測結果，在 private test data 釋出前可先用 sample\_submission 裡的資料代替。

## 評分方式

Prevalence-weighted average of Area Under ROC Curve (AUC) for the multi-class problem (One-vs-rest)

$$AUC_{total} = \sum_{c_i \in C} AUC(c_i) \cdot p(c_i)$$

- C 為全部類別的集合，在此競賽中， $C = \{\text{'類別1'}, \text{'類別2'}, \text{'類別3'}, \text{'類別4'}, \text{'類別5'}, \text{'類別6'}\}$ 。
- Scheme 選擇 One-vs-rest。
- Average method 選擇 prevalence-weighted，會根據每個類別的資料數量進行加權平均。