

# Audience 站台 1.2 使用說明 (modeling)

黃彥鈞 Weber Huang

Data engineer RD2

# 大綱

- 更新內容
- 操作說明
  - 建立模型任務
  - 模型訓練 / 整備
  - 模型結果1 基礎功能
  - 模型結果2 詞權重操作

# 更新內容

- 專案目的:
  - 重構 模型建立任務 功能至後端 Audience API
  - 重構站台顯示功能，串接 Audience API

舊站台	新站台
訓練、驗證與預測等功能全部包在 audience 站台專案	分離模型訓練主要功能至 Audience API，區隔站台顯示功能
模型訓練和整備產生之資料均儲存於站台資料庫	模型訓練和整備產生之資料改存至後台 Audience result 資料庫
	重構站台功能，串接後台 Audience API
	詳細驗證資料報告修改為檔案下載連結
	新增 term weight 訓練結果下載與上傳功能

# 建立模型任務

模型建立任務

## 任務列表

Show 10 entries

Search:

id	模型名稱	模型類型	使用資料	建立時間	建立者	模型狀態	準確率
3	雜訊關鍵字模型	關鍵字規則	雜訊關鍵字	2021年5月5日	cindy	等待中	0.57
6	男性內文規則式模型	正則表達式比對	男性_內文規則式	2021年6月3日	root	等待中	N/A
7	詞彙權重	詞彙權重模型	雜訊測試	2021年6月4日	root	等待中	0.66
9	男性-舊版八大詞彙權重	詞彙權重模型	「男性-舊版八大」自動建立的任務	2021年7月9日	root	等待中	N/A
10	男性作者名稱	關鍵字規則	男性_作者名關鍵字規則	2021年7月9日	root	完成	N/A

建立模型任務  
方法與舊站台  
相同

ychuang

建立新任務

## 模型建立流程

每種不同的模型都有其各自的建立方式，以下會說明各種類型的建立條件與做法：

- SVM
- 隨機森林
- 關鍵字規則
- 正則表達式比對
- 詞彙權重模型

目前支援這四款模型訓練與整備功能

### 監督式學習模型

此類型的模型藉由帶有標籤的文本資料，讓機器自動學習文本中的特徵，可以使用的資料準備任務為 **監督式學習模型**。

為了訓練出此類型的模型，需要提供人工標記的資料，每種標籤建議數量為 **至少200篇文章**。

建立任務

×

任務名稱：

Test4

描述與定義：

demo

模型類型：

關鍵字規則

特徵欄位：

作者

使用資料：

雜訊關鍵字 (關鍵字規則模型資料)

送出

# 模型訓練/整備

上傳額外測試資料

重新訓練

操作 ▾

Test4

等待中

由於 2022年2月22日 15:36 建立 • demo

模型資訊

模型類型：關鍵字規則

訓練資料

使用資料：#3 雜訊關鍵字

資料數量：

	訓練集	驗證集	測試集
色情	0	0	0
貸款	0	0	0
總計	0	0	0

訓練模型

操作 ▾

模型效果

無論是監督式學習模型或規則模型(關鍵字/正則)，訓練或整備均需要透過訓練模型按鈕。

規則模型每次更新相對應規則列表，都需要再次點擊一次重新訓練，讓規則資料寫入後端。

訓練/整備過的模型(狀態完成)才能被族群貼標任務所使用。

# 模型結果1 基礎功能

Test2

完成

由於 2022年1月27日 16:07 建立。

測試站功能

模型名稱與使用相關資料標籤數量統計

模型資訊

模型類型：隨機森林

訓練資料

使用資料：#1 雜訊測試

資料數量：

	訓練集	驗證集	測試集
一般	3200	400	400
色情	3200	400	400
貸款	3200	400	400
抽獎	3200	400	400
總計	12800	1600	1600

上傳額外測試資料

重新訓練

操作

模型效果

驗證集報告\_120

準確度: 0.25

	precision	recall	f1-score	support
一般	0.26	0.29	0.27	1000
抽獎	0.25	0.22	0.24	1000
色情	0.25	0.29	0.27	1000
貸款	0.25	0.22	0.23	1000
macro avg	0.25	0.25	0.25	4000
weighted avg	0.25	0.25	0.25	4000

驗證細節下載(csv)

測試集報告\_121

驗證與測試資料報告(說明)  
驗證資料細節下載，可以獲取驗證、預測資料集每一筆的答案與預測值

doc_id	content	ground_truth	predict_label
14273	momo好食堂等你來按讚 >>>https://momo.dm/f/ViMf3先到momo好食堂粉絲團按讚再到此貼文按讚+留言+tag2名好友來抽獎 陳為勝林振德〈振振〉楊素真有了【Tefal法國特福】鈦金系列不沾刀具四件組，人人都是特級廚師。	抽獎	抽獎
13368	959助貸網各家銀行車貸利率比較2018一般來說信用貸款的利率落在2%~15%的都有，審核條件關優到核貸利率，因此您要多了解您的情況才能夠找到最適合的銀行。	色情	貸款
13814	優缺點總結： 《留言贈獎活動》☑按讚☑公開分享☑Tag 3名好友+李念庭邱玉青李明智即抽出3名幸運粉絲，大方送【可愛舒適抱枕】(隨機出貨)。	一般	色情
14307	A：因小星星外送茶美眉都是兼差性質，可配合程度也不一定。	貸款	抽獎
13074		貸款	色情
12724		一般	抽獎

# 模型結果2 詞權重操作

詞彙權重列表

新增、修改或刪除其中一筆資料內容

新增 修改 刪除

Show 10 entries

Search:

編號	詞彙	標籤	權重
5757	朋友	色情	1
5758	10	色情	0.98825
5759	momo	色情	0.95653
5760	系列	色情	0.95195
5761	喝茶	色情	0.93738
5762	com	色情	0.92758
5763	玻璃	色情	0.92352
5764	信用卡	色情	0.8472
5765	抽出	色情	0.83033
5766	方案	色情	0.79338

Showing 1 to 10 of 264 entries

Previous 1 2 3 4 5 ... 27 Next

權重資料下載(csv)

上傳額外測試資料 重新訓練 操作

上傳詞彙權重

修改

刪除

請注意，上傳詞彙權重選項會覆寫當前權重列表所有的資料

id	label	term	weight
5757	色情	朋友	1
5758	色情	10	0.98825
5759	色情	momo	0.95653
5760	色情	系列	0.95195
5761	色情	喝茶	0.93738
5762	色情	com	0.92758
5763	色情	玻璃	0.92352
5764	色情	信用卡	0.8472
5765	色情	抽出	0.83033
5766	色情	方案	0.79338

使用者可以在本地端修改內容，若要上傳修改好之檔案，請確保資料格式與欄位相同(必要欄位: label, term, weight)