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: 模型 進確 ↑↓ 模型類型 ↑↓ 狀態 **率** ↑↓ 模型名稱 使用資料 ↑↓ 建立時間 ↑↓ 建立者 雜訊關鍵字模型 關鍵字規則 雜訊關鍵字 2021年5月5 cindv 等待中 0.57 男性內文規則式模 正則表達式 男性_內文規則式 2021年6月3 等待中 root N/A 比對 詞彙權重 詞彙權重模 雜訊測試 2021年6月4 等待中 0.66 root 型 日 男性-舊版八大詞彙 「男性-舊版八大」自動建立的任 詞彙權重模 2021年7月9 root 等待中 N/A 型 日 權重 務 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 root 完成 N/A \Box

建立模型任務 方法與舊站台 相同



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

款模型訓練與 整備功能

目前支援這四

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。

王務名稱:		
Test4		
苗述與定義:		
demo		
莫型類型:		

雜訊關鍵字(關鍵字規則模型資料)

建立任務

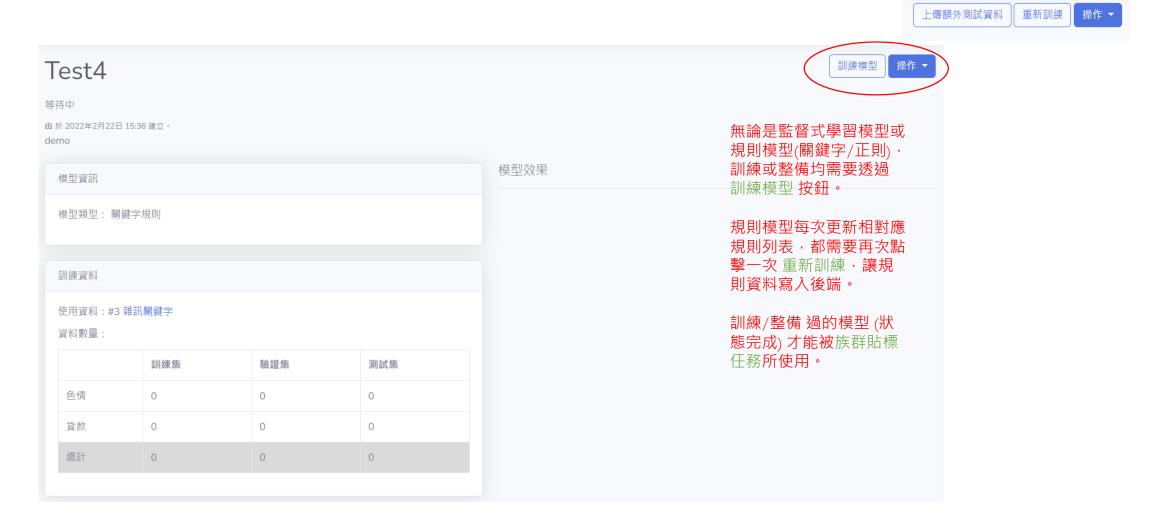
作者

使用資料:

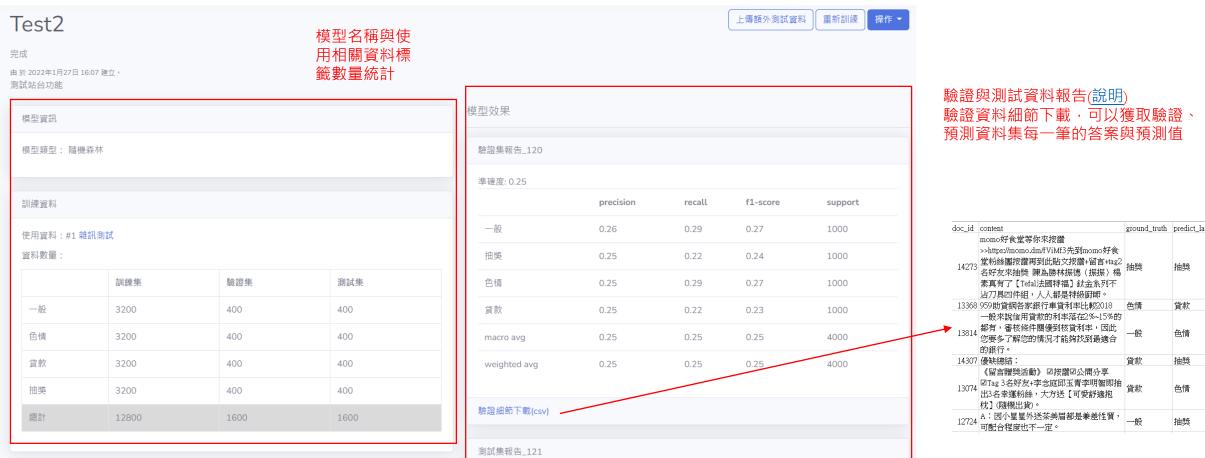
送出

×

模型訓練/整備



模型結果1基礎功能



loc_id	content	ground_truth	predict_label
14273	momo好食堂等你來按讚 >>https://momo.dm/YiMf3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法國特溫】欽金系列不 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	色情	貸款
13814	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。	一般	色情
14307	優缺總結:	貸款	抽獎
13074	《留言贈獎活動》 図按讚四公開分享 図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)。	貸款	色情
12724	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。	一般	抽獎

模型結果2 詞權重操作

