

QOCA® AIM 廣達人工智慧醫療雲運算整合平台操作簡易說明 (數值型資料機器學習分析平台)

一、 資料傳輸

***注意事項：**為避免研究資料上傳失敗，上傳資料之前須注意資料格式是否符合以下規定：

1. 檔案建議是 **CSV 檔**(逗號分隔的格式)。若資料是用 SPSS 軟體儲存成「.sav 檔」，建議將檔案匯出或另存成 CSV 檔後，再匯入平台
2. 資料框架(data frame)必須是二維矩陣形式，如下圖所示：

變項(variable)：記錄在資料表頭第一列

No	birth	diagdate	age	weight	sex	education
1	22/5/11	95/1/10	72	56.2	1	0
2	35/4/28	95/1/10	59	65.8	1	2
3	55/12/3	95/1/17	39	70.5	1	3
4	44/11/5	95/1/24	50	66.7	1	2
5	32/5/7	95/2/2	62	50.5	0	1
6	24/3/14	95/2/2	70	50.3	0	0
:	:	:	:	:	:	:

照表頭的變項，依序填入數值

3. 欄位名稱必須是**英文**，名稱開頭不可以是數字、中間不可以有空格，特殊符號只能使用 `_`，以下是建議的命名方式：
如：`diagnosis_date`、`DiagnosisDate`、`ICD_10`

4. 資料中有「**類別型**」資料時，必需將資料「重新編碼」(recoding)
Example 1:

Label coding (binary coding)



二分類資料，建議以「0」、「1」編碼

Example 2:

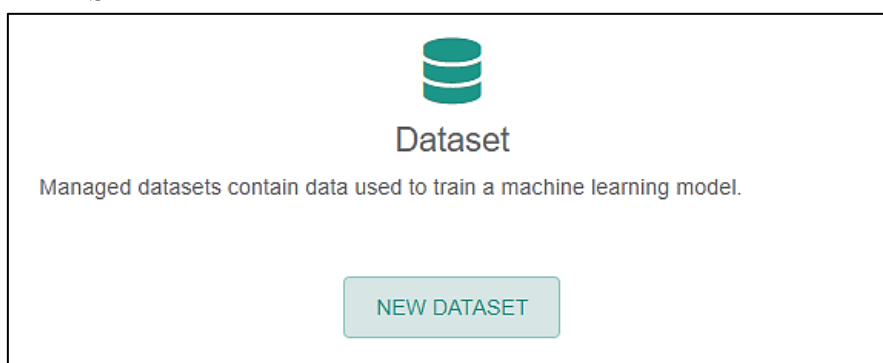
One-hot coding

Severity		Severity_1	Severity_2	Severity_3
1	Recoding →	1	0	0
2		0	1	0
3		0	0	1

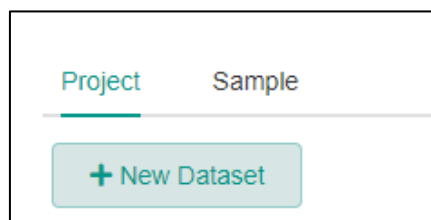
多分類資料(≥ 3 類)，建議以 One-hot coding 編碼

5. 上傳資料：

(1) 點選 NEW DATASET



(2) 點選 + New Dataset



(3) 新增資料：

- i. Project: 選擇您的申請案號(例：CIC_110AIM000)
- ii. Title: 輸入資料檔案名稱
- iii. Type: 資料類型，請選 Structure Data
- iv. Source: 點選 Local File，上傳本機資料
- v. Delimiter: 資料分隔符號，資料為 CSV 檔，請選 comma
- vi. File: 選擇您要上傳的資料
- vii. Save: 儲存並上傳

Create Dataset ✕

* Project CIC_110AIM000

* Title example

Type Structured Data

Source Local File

* Delimiter comma

* File simulated_whas.csv 4...

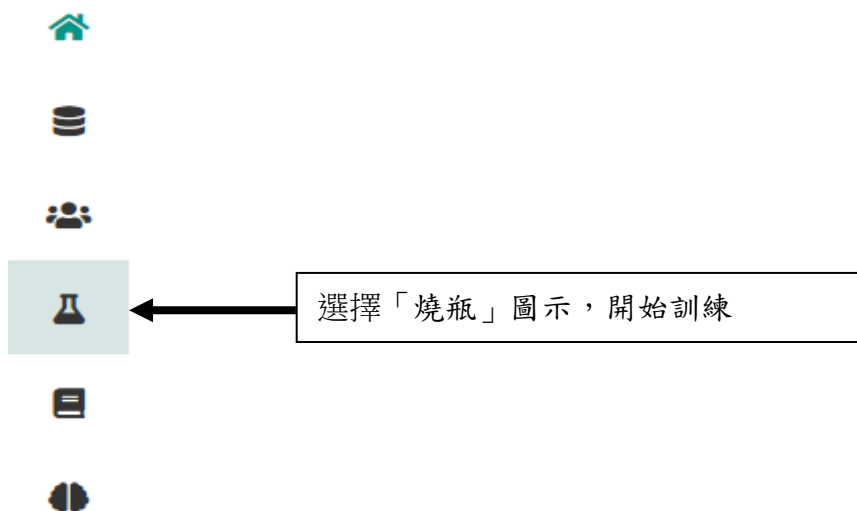
Cancel
Save

(4) 上傳成功畫面

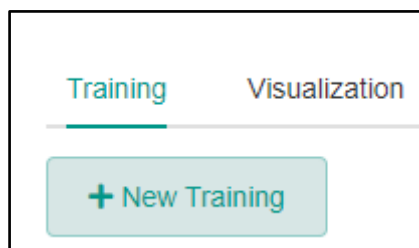
Project		Sample							
#	IRB	Title	Type	Size	Versions	Uploader	Updated Datetime	Operations	
> 1	CIC_110AIM000	example	Structured Data	430.8 KB	1	生醫數據 (CIC)	2021-11-29 10:02:36	ⓘ ↻ 🗑️	

二、 資料分析

1. 開始訓練：點選左側命令列，開始進行模型訓練

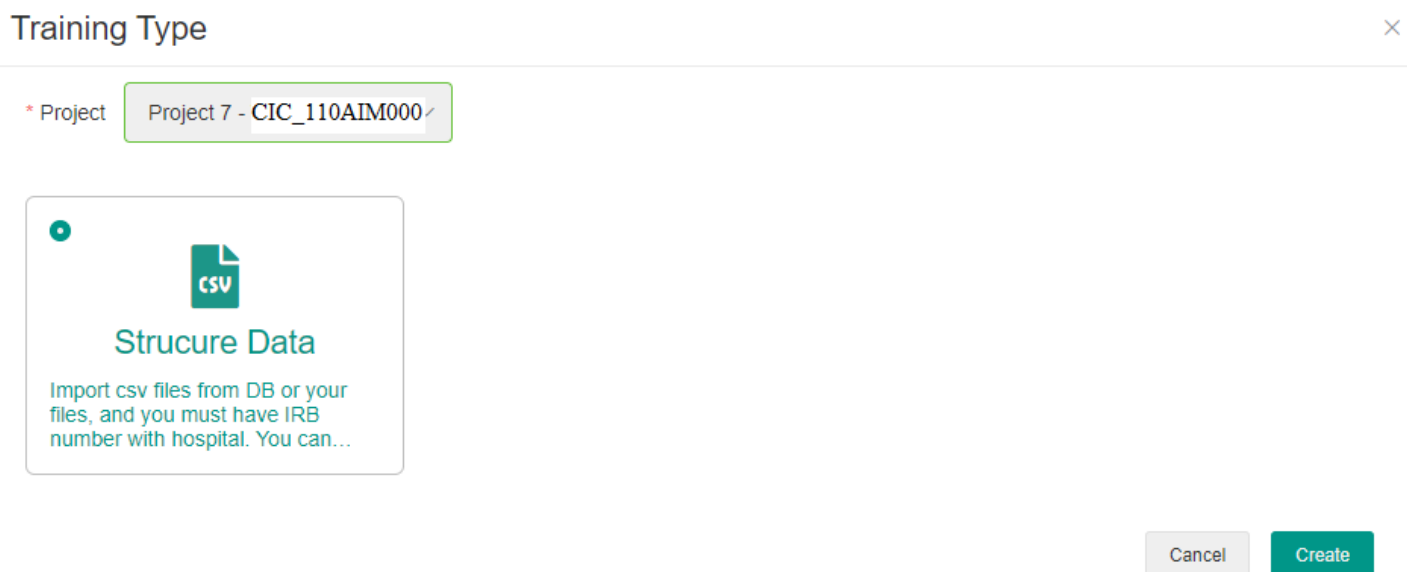


2. 點選 +New Training



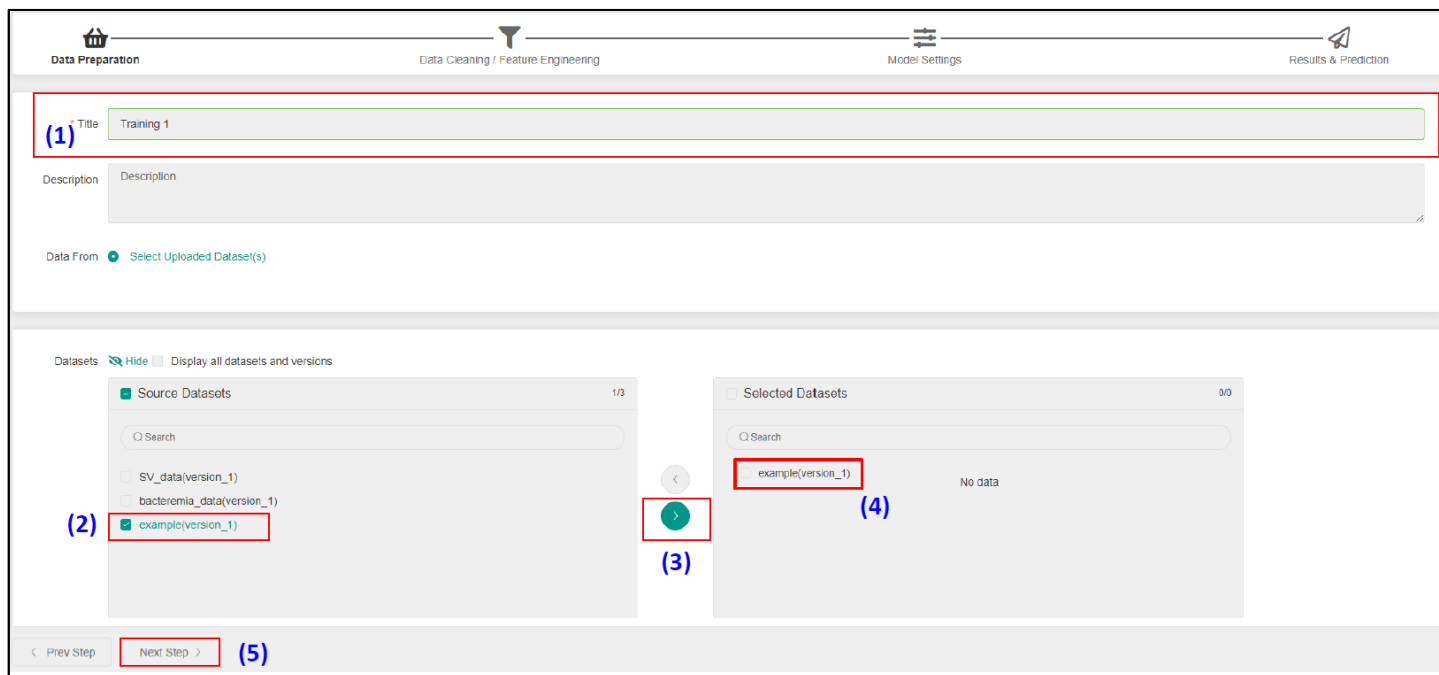
3. Training Type:

- (1) Project: 選擇您的案號(例：CIC_110AIM000)
- (2) 點選 Structure Data
- (3) 點選 Create



4. Data Preparation

- (1) Title：輸入訓練的名稱
- (2) 選擇要分析的資料集
- (3) 點選 「>」
- (4) Select Datasets 出現要分析的資料集
- (5) Next Step 下一步



5. Data Cleaning / Feature Engineering: 資料清理與特徵工程

- (1) 依據上排的功能列，使用者可自行填補遺漏值(選「Missing Values」功能)、資料型別轉換 (選「Data Type Confirmation」功能)與重新編碼(選「Encoding」功能)等

(2) 注意：進行機器學習分析時，「類別型資料」欄位的資料型別必需是「Discrete」的資料型別，且內容為數值；如果是要進行統計分析時，「類別型資料」欄位的資料型別必需是「Categorical」的資料型別。可透過「Data Type Confirmation」功能進行資料型別轉換

#	# ID	# AGE	A GENDER	# HR	# SYSBP	# DIASBP	# BMI	# CVD
1	1	83.0	F	89.0	152.0	78.0	26.0	1
2	2	49.0	F	84.0	120.0	60.0	24.0	1
3	3	70.0	M	83.0	147.0	88.0	22.0	0
4	4	70.0	F	65.0	123.0	76.0	27.0	1
5	5	70.0	F	63.0	135.0	85.0	24.0	1
6	6	70.0	F	76.0	83.0	54.0	23.0	1
7	7	57.0	F	73.0	147.0	131.0	39.0	1
8	8	55.0	F	91.0	147.0	95.0	27.0	1

(3) 準備訓練模型時，記得設定「Training column」，指定 target 和 feature



(4) 指定 Target 和 Feature，並選「Save」

Target	Feature	Name	Data Type	Null Count
<input type="checkbox"/>	<input checked="" type="checkbox"/>	AFB	Discrete	0
<input type="checkbox"/>	<input checked="" type="checkbox"/>	SHO	Discrete	0
<input type="checkbox"/>	<input checked="" type="checkbox"/>	CHF	Discrete	0
<input type="checkbox"/>	<input checked="" type="checkbox"/>	AV3	Discrete	0
<input type="checkbox"/>	<input checked="" type="checkbox"/>	MIORD	Discrete	0
<input type="checkbox"/>	<input checked="" type="checkbox"/>	MITYPE	Discrete	0
<input type="checkbox"/>	<input checked="" type="checkbox"/>	LOS	Discrete	0
<input checked="" type="checkbox"/>	<input type="checkbox"/>	DEATH	Discrete	0

(5) Next step: 下一步

2. Model Settings: 以 Decision Tree 為例

(1) Pre-Processing Setting: 以 5-折交互驗證法為例 (使用者可以選擇其他模型配適方法)

Pre-Processing Setting

* Validation Method

Folds

* Random Seed

(2) Hyper-Parameter Setting: 以 Decision Tree 為例 (使用者可以選擇其他機器學習方法), 使用者可以設定模型的「超參數」(hyper-parameter), 決定模型的複雜程度

Hyper-Parameter Setting

* Model Name

* Max Depth ⓘ

* Min Samples Split

* Min Samples Leaf

* Metrics Index Loss Accuracy Precision Recall F1 Score

(3) 開始訓練

Auto Training

Conduct auto training with current parameters.

準備啟動Container

等待訓練...

100%

100% 訓練完畢

Next Step >

(4) 訓練完成後，點選 Next step 下一步

3. Results / Prediction: 產出報表與結果

(1) 結果: 點選左側「>」標記可展開結果

Data Preparation
Data Cleaning / Feature Engineering
Model Settings
Results & Prediction

Model ID	Class	Creator	Score	Training Model	Training Time	Release	Status	Operations
20211129_111726_5z	Classification	生醫數據 (CIC)	0.9803	DecisionTreeClassifier	3.4511	Un-released	Completed	🔍 ☰ 📄 🖨️ 📥 🗑️

Result

Started Time	2021-11-29 03:17:39	Training Time	3.4511	Feature Weights	[Show]
Model Arguments	[Show]	Accuracy	0.9803	F1 Score	0.9059
Precision	0.9679	Recall	0.8519	Log Loss	0.1092
ROC AUC	0.9853	Confusion Matrix	[Show]	Roc Curve	[Show]
Feature Importances	[Show]	Labels Group Bar	[Show]	Visualizing Decision Trees	[Show]

Parameters

Created Datetime	2021-11-29 11:17	Validation Method	K-Fold Validation	Folds	5
Random Seed	123	Model Name	DecisionTreeClassifier	Metrics Index	loss, accuracy, precision, recall, f1_score
File Name (Ver.)	Training 1.csv (5)	Target	DEATH	Features	AGE GENDER HR SYSBP DIASBP BMI CVD AFB SHO CHF AV3 MIORD MITYPE LOS

(2) 報表: 點選 report 可以呈顯報表

