

## 決策樹之糖尿病預測

本樣本（pima-indians-diabetes.csv）是 Pima Indians Diabetes Database，來源於 Kaggle。該樣本資料含有 8 個特徵向量和最後的 Label，標示是否為糖尿病患者，8 個特徵變量分別是懷孕次數、glucose、BP 血壓、skin 皮膚厚度等等，前五筆資料如下。

	pregnant	glucose	bp	skin	insulin	bmi	pedigree	age	label
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

以學習方式建立決策樹辨識機，後輸出針對測試集之 accuracy score，confusion matrix，若題目所給的最大深度 max\_depth≤5 請亦輸出決策樹圖形檔。

### stdin 輸入

測資計三行，前兩行為切割訓練集與測試集參數，分別為 test\_size，與 random\_state。後三行為切割訓練集與測試集參數建立決策樹學習機參數，第三行為子串'entropy'或'gini'，第四行為 max\_depth，第五行為 random\_state。

### Stdin 檔案輸入

pima-indians-diabetes.csv

### stdout 輸出

測試集之 classification report 與 confusion matrix。

### 檔案輸出

若題目所給的最大深度 max\_depth≤5 請輸出決策樹圖形檔輸出決策樹圖形檔案 diabetes\_tree.png。

### 範例 stdin 輸入

```
0.3
0
gini
5
0
```

## 範例 stdout 輸出

Classification report:

	precision	recall	f1-score	support
0	0.81	0.84	0.83	157
1	0.64	0.59	0.62	74
accuracy			0.76	231
macro avg	0.73	0.72	0.72	231
weighted avg	0.76	0.76	0.76	231

Confusion Matrix:

```
[[132  25]
```

```
 [ 30  44]]
```

## 範例圖形輸出

