

114-1 大數據分析與智慧運算作業 5

總分 95/100 ?

電子郵件 *

b1141065@ems.niu.edu.tw

0 分，共 0 分

姓名 *

陳昱安

學號 *

B1141065

題組一

40 分，共 45 分

✓ 決策樹演算法屬於下列何種技術? *

5/5

- ☒ 分類
- ☐ 序列型樣探勘
- ☐ 分群
- ☐ 關聯規則探勘



✓ 資料表的橫向資料又稱為? *

5/5

- ☐ Variable
- ☒ Record
- ☐ Attribute
- ☐ Feature



✓ 資料表的直向資料又稱為? *

5/5

- ☒ Feature
- ☒ Attribute
- ☒ Variable
- ☐ Tuple



✗ 下列敘述何者正確? *

0/5

- ☐ 屬性的Entropy越低，代表該屬性的鑑別能力越高
- ☐ 屬性的Entropy越高，代表該屬性的鑑別能力越低
- ☒ 屬性的Information Gain越高，代表該屬性的鑑別能力越高
- ☐ 屬性的Information Gain越低，代表該屬性的鑑別能力越高



✓ 參考下圖，當目標屬性為buys_computer時，決策樹以何屬性為根節點？* 5/5

Training data tuples from the AllElectronics customer database.

RID	age	income	student	credit_rating	Class: buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 ... 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 ... 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31 ... 40	medium	no	excellent	yes
13	31 ... 40	high	yes	fair	yes
14	>40	medium	no	excellent	no

- ☐ income
- ☐ student
- ☒ age
- ☐ credit_rating



✓ 訓練與測試資料的切割比例通常為？* 5/5

- ☐ 訓練資料20%、測試資料80%
- ☒ 訓練資料70%、測試資料30%
- ☐ 訓練資料30%、測試資料70%
- ☐ 訓練資料40%、測試資料30%



✓ 請問下列何者非決策樹停止分割節點的條件為何? *

5/5

- ☐ 沒有屬性可以使用
- ☐ 節點內所有資料屬於同一類別
- ☐ 節點內的資料量過少
- ☒ 線性分割



✓ 當目標屬性為類別型，可使用下列何種方法評估分類模型? (複選) *

5/5

- ☐ MAE
- ☐ RMSE
- ☒ F-Measure
- ☒ Accuracy



✓ 關於分類技術的描述下列何者錯誤? *

5/5

- ☒ 目標屬性為數值型態
- ☐ 通常會從歷史資料挖一塊資料當作測試資料
- ☐ 測試資料的主要目的是用來評估模型是否準確
- ☐ 訓練資料的主要目的是用來訓練分類器



題組二

20 分，共 20 分

✓ 題組二，表中，分類模型的 Accuracy為何？*

5/5

資料 ID	模型預測結果	測試資料真實答案
1	A	B
2	A	A
3	A	B
4	B	B
5	B	A

☐ 0.3

☐ 0.1

☒ 0.4

☐ 0.8



✓ 題組二，表中，分類模型對 A 的Precision 為何？*

5/5

資料 ID	模型預測結果	測試資料真實答案
1	A	B
2	A	A
3	A	B
4	B	B
5	B	A

- ☐ 1
- ☐ 1/4
- ☐ 1/2
- ☒ 1/3



✓ 題組二，表中，分類模型對 A 的Recall 為何? *

5/5

資料 ID	模型預測結果	測試資料真實答案
1	A	B
2	A	A
3	A	B
4	B	B
5	B	A

☐ 1/4

☐ 1/3

☐ 1

☒ 1/2



✓ 題組二，表中，分類模型對 A 的F-measure 為何? *

5/5

資料 ID	模型預測結果	測試資料真實答案
1	A	B
2	A	A
3	A	B
4	B	B
5	B	A

- ☒ 0.4
- ☐ 0.6
- ☐ 1
- ☐ 10



題組三

35 分，共 35 分

✓ 下列何者為監督式學習方法 *

5/5

- ☐ K-means
- ☐ PCA
- ☐ DBSCAN
- ☒ Decision Tree



✓ 在機器學習中，決策樹是一種用於進行什麼類型的任務？ * 5/5

- ☐ Regression
- ☒ Classification
- ☐ Dimension Reduction
- ☐ Data Visualization



✓ 在決策樹中，節點的分裂依據通常是什麼？ * 5/5

- ☐ 偏差值
- ☐ 權重
- ☐ 標籤
- ☒ 屬性值



✓ 下列何者非決策樹常用的 goodness function * 5/5

- ☐ Information Gain
- ☐ Entropy
- ☐ Gini index
- ☒ Mean absolute error



✓ 資料庫的亂度為?(小數點後兩位四捨五入) *

5/5

Training data tuples from the *AllElectronics* customer database.

<i>RID</i>	<i>age</i>	<i>income</i>	<i>student</i>	<i>credit_rating</i>	<i>Class: buys_computer</i>
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 . . . 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 . . . 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31 . . . 40	medium	no	excellent	yes
13	31 . . . 40	high	yes	fair	yes
14	>40	medium	no	excellent	no

☒ 0.94

☐ 0.96

☐ 0.85

☐ 0.90



✓ Income 的亂度為?(小數點後兩位四捨五入) *

5/5

Training data tuples from the *AllElectronics* customer database.

RID	age	income	student	credit_rating	Class: buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 . . . 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 . . . 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31 . . . 40	medium	no	excellent	yes
13	31 . . . 40	high	yes	fair	yes
14	>40	medium	no	excellent	no

☐ 0.93☐ 0.85☐ 0.89☒ 0.91

✓ Income 的Information Gain為?(小數點後兩位四捨五入) *

5/5

Training data tuples from the *AllElectronics* customer database.

<i>RID</i>	<i>age</i>	<i>income</i>	<i>student</i>	<i>credit_rating</i>	<i>Class: buys_computer</i>
1	≤ 30	high	no	fair	no
2	≤ 30	high	no	excellent	no
3	31 ... 40	high	no	fair	yes
4	> 40	medium	no	fair	yes
5	> 40	low	yes	fair	yes
6	> 40	low	yes	excellent	no
7	31 ... 40	low	yes	excellent	yes
8	≤ 30	medium	no	fair	no
9	≤ 30	low	yes	fair	yes
10	> 40	medium	yes	fair	yes
11	≤ 30	medium	yes	excellent	yes
12	31 ... 40	medium	no	excellent	yes
13	31 ... 40	high	yes	fair	yes
14	> 40	medium	no	excellent	no

☐ 0.15

☐ 0.05

☒ 0.02

☐ 0.12

