

# 310706034 資管碩一 吳啓玄 資料探勘研究與實務 HW1- Decision Tree

1. 讀取資料後刪除 Book of Death、Death Chapter，結果如下圖：

	Name	Allegiances	Death Year	Book Intro Chapter	Gender	Nobility	GoT	CoK	SoS	FfC	DwD
0	Addam Marbrand	Lannister	NaN	56.0	1	1	1	1	1	1	0
1	Aegon Frey (Jinglebell)	None	299.0	49.0	1	1	0	0	1	0	0
2	Aegon Targaryen	House Targaryen	NaN	5.0	1	1	0	0	0	0	1
3	Adrack Humble	House Greyjoy	300.0	20.0	1	1	0	0	0	0	1
4	Aemon Costayne	Lannister	NaN	NaN	1	1	0	0	1	0	0
...	...	...	...	...	...	...	...	...	...	...	...
912	Zollo	None	NaN	21.0	1	0	0	0	1	0	0
913	Yurkhaz zo Yunzak	None	300.0	47.0	1	0	0	0	0	0	1
914	Yezzan Zo Qaggaz	None	300.0	25.0	1	1	0	0	0	0	1
915	Torwynd the Tame	Wildling	300.0	73.0	1	0	0	0	1	0	0
916	Talbert Serry	Tyrell	300.0	29.0	1	1	0	0	0	1	0

917 rows x 11 columns

2. 利用 Death Year 判斷是否死亡並增加 IsDeath 欄位，結果如下圖：

	Name	Allegiances	Death Year	Book Intro Chapter	Gender	Nobility	GoT	CoK	SoS	FfC	DwD	IsDeath
0	Addam Marbrand	Lannister	NaN	56.0	1	1	1	1	1	1	0	0
1	Aegon Frey (Jinglebell)	None	299.0	49.0	1	1	0	0	1	0	0	1
2	Aegon Targaryen	House Targaryen	NaN	5.0	1	1	0	0	0	0	1	0
3	Adrack Humble	House Greyjoy	300.0	20.0	1	1	0	0	0	0	1	1
4	Aemon Costayne	Lannister	NaN	NaN	1	1	0	0	1	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
912	Zollo	None	NaN	21.0	1	0	0	0	1	0	0	0
913	Yurkhaz zo Yunzak	None	300.0	47.0	1	0	0	0	0	0	1	1
914	Yezzan Zo Qaggaz	None	300.0	25.0	1	1	0	0	0	0	1	1
915	Torwynd the Tame	Wildling	300.0	73.0	1	0	0	0	1	0	0	1
916	Talbert Serry	Tyrell	300.0	29.0	1	1	0	0	0	1	0	1

917 rows x 12 columns

3. 刪除 Death Year、Book Intro Chapter，結果如下圖：

	Name	Allegiances	Gender	Nobility	GoT	CoK	SoS	FfC	DwD	IsDeath
0	Addam Marbrand	Lannister	1	1	1	1	1	1	0	0
1	Aegon Frey (Jinglebell)	None	1	1	0	0	1	0	0	1
2	Aegon Targaryen	House Targaryen	1	1	0	0	0	0	1	0
3	Adrack Humble	House Greyjoy	1	1	0	0	0	0	1	1
4	Aemon Costayne	Lannister	1	1	0	0	1	0	0	0
...	...	...	...	...	...	...	...	...	...	...
912	Zollo	None	1	0	0	0	1	0	0	0
913	Yurkhaz zo Yunzak	None	1	0	0	0	0	0	1	1
914	Yezzan Zo Qaggaz	None	1	1	0	0	0	0	1	1
915	Torwynd the Tame	Wildling	1	0	0	0	1	0	0	1
916	Talbert Serry	Tyrell	1	1	0	0	0	1	0	1

917 rows x 10 columns

4. 檢查目前資料是否存在遺失值，檢查結果如下圖，並未存在任何遺失值

```
Name      0.0
Allegiances 0.0
Gender     0.0
Nobility   0.0
GoT        0.0
CoK        0.0
SoS        0.0
FfC        0.0
DwD        0.0
IsDeath    0.0
dtype: float64
```

5. 刪掉 Name 並將 Allegiances 轉 dummy

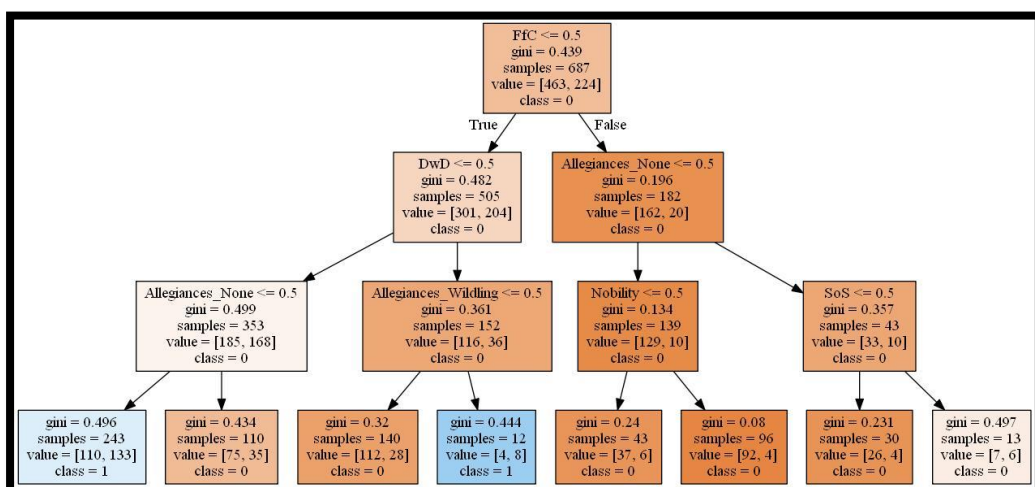
	Gender	Nobility	GoT	CoK	SoS	FfC	DwD	IsDeath	Allegiances_Arryn	Allegiances_Baratheon	...
0	1	1	1	1	1	1	0	0	0	0	...
1	1	1	0	0	1	0	0	1	0	0	...
2	1	1	0	0	0	0	1	0	0	0	...
3	1	1	0	0	0	0	1	1	0	0	...
4	1	1	0	0	1	0	0	0	0	0	...
...	...	...	...	...	...	...	...	...	...	...	...
912	1	0	0	0	1	0	0	0	0	0	...
913	1	0	0	0	0	0	1	1	0	0	...
914	1	1	0	0	0	0	1	1	0	0	...
915	1	0	0	0	1	0	0	1	0	0	...
916	1	1	0	0	0	1	0	1	0	0	...

917 rows x 29 columns

6. 亂數將目前資料拆成訓練集(75%)與測試集(25%)，利用訓練集建構深度為 3 的決策樹，並在測試集上做預測得到以下結果

```
Confusion Matrix:
[[109  40]
 [ 34  47]]
-----
precision: 0.54
recall: 0.58
accuracy: 0.678
```

7. 決策樹圖如下圖：



註：class 即為 IsDeath