

ECT_HW2_107403020

李泳輝

1.1 載入 train.arff，將 PassengerId、Name、Ticket 欄位刪除 (5%)
打開 Edit，點擊右鍵。

Viewer

Relation: train

No.	1: PassengerId	2: Survived	3: Pclass	4: Name	5: Sex	6: Age	7: SibSp	8: Parch	9: Ticket	10: Fare	11: Cabin	12: Embarked
	Numeric	Numeric	Numeric	Nominal	Nominal	Numeric	Numeric	Numeric	Nominal	Numeric	Nominal	Nominal
1		Get mean...			male	22.0	1.0	0.0	A/5 211...	7.25		S
2		Set all values to...			female	38.0	1.0	0.0	PC 175...	71.2833	C85	C
3		Set missing values to...			female	26.0	0.0	0.0	STON/...	7.925		S
4		Replace values with...			female	35.0	1.0	0.0	113803	53.1	C123	S
5					female	35.0	0.0	0.0	373450	8.05		S
6		Rename attribute...			male		0.0	0.0	330877	8.4583		Q
7		Set attribute weight...			male	54.0	0.0	0.0	17463	51.8625	E46	S
8		Attribute as class			male	2.0	3.0	1.0	349909	21.075		S
9		Delete attribute			female	27.0	0.0	2.0	347742	11.1333		S
10		Delete attributes...			female	14.0	1.0	0.0	237736	30.0708		C
11		Sort data (ascending)			female	4.0	1.0	1.0	PP 9549	16.7	G6	S
12		Optimal column width (current)			female	58.0	0.0	0.0	113783	26.55	C103	S
13		Optimal column width (all)			male	20.0	0.0	0.0	A/5, 21...	8.05		S
14					male	39.0	1.0	5.0	347082	31.275		S
15					female	14.0	0.0	0.0	350406	7.8542		S
16	16.0	1.0	2.0	Hewlett...	female	55.0	0.0	0.0	248706	16.0		S
17	17.0	0.0	3.0	Rice, Ma...	male	2.0	4.0	1.0	382652	29.125		Q
18	18.0	1.0	2.0	William...	male		0.0	0.0	244373	13.0		S
19	19.0	0.0	3.0	Vander ...	female	31.0	1.0	0.0	345763	18.0		S
20	20.0	1.0	3.0	Massel...	female		0.0	0.0	2649	7.225		C
21	21.0	0.0	2.0	Fynney, ...	male	35.0	0.0	0.0	239865	26.0		S
22	22.0	1.0	2.0	Beesley, ...	male	34.0	0.0	0.0	248698	13.0	D56	S
23	23.0	1.0	3.0	McGow...	female	15.0	0.0	0.0	330923	8.0292		Q
24	24.0	1.0	1.0	Sloper, ...	male	28.0	0.0	0.0	113788	35.5	A6	S

Add instance Undo OK Cancel

● 選擇欲刪除的 Attributes

3.0	Heikkine...	female	26.0	0.0	0.0	STON/...	7.925		S
1.0	Futrelle, ...	female	35.0	1.0	0.0	113803	53.1	C123	S
3.0	Allen, M...	male	35.0	0.0	0.0	373450	8.05		S
3.0	Moran, ...	male				330877	8.4583		Q
1.0	McCart...	male				330877	8.4583	E46	S
3.0	Palsson,...	male					21.075		S
3.0	Johnson...	female					11.1333		S
2.0	Nasser, ...	female					30.0708		C
3.0	Sandstr...	female					16.7	G6	S
1.0	Bonnell, ...	female					26.55	C103	S
3.0	Saunders...	male					8.05		S
3.0	Anderss...	male					31.275		S
3.0	Vestro...	female					7.8542		S
2.0	Hewlett...	female					16.0		S
3.0	Rice, Ma...	male	2.0	4.0	1.0	382652	29.125		Q
2.0	William...	male			0.0	0.0	244373	13.0	S
3.0	Vander ...	female	31.0	1.0	0.0	345763	18.0		S
3.0	Massel...	female			0.0	0.0	2649	7.225	C
2.0	Fynney, ...	male	35.0	0.0	0.0	239865	26.0		S

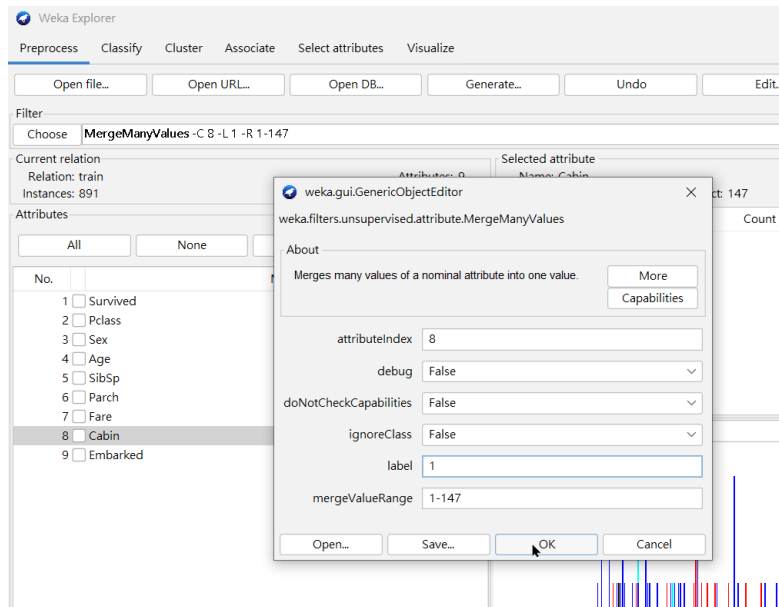
Select items

- Name
- Parch
- PassengerId
- Pclass
- Sex
- SibSp
- Survived
- Ticket

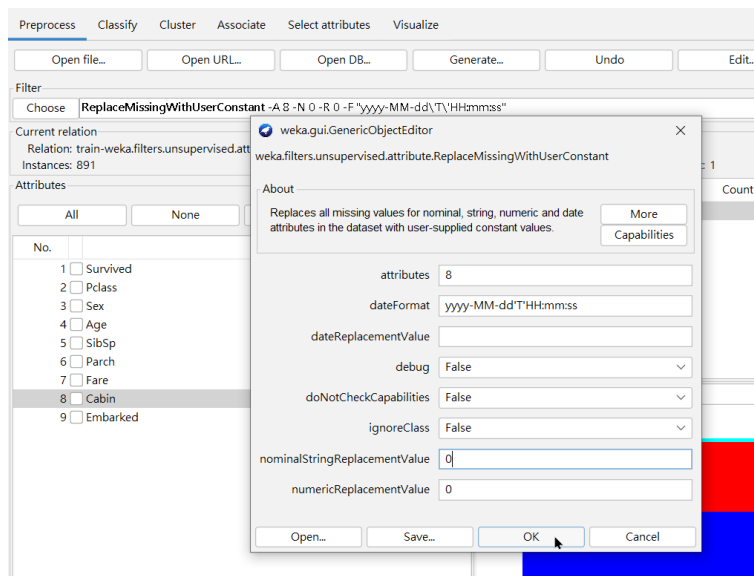
Select Pattern Cancel

1.2 將 Cabin 的非空值以 1 替代，空值以 0 填入（先用 MergeManyValues 取代非空值，再用 ReplaceMissingWithUserConstant 取代空值） (5%)

- 用 MergeManyValues 以 1 取代 Cabin 的非空值



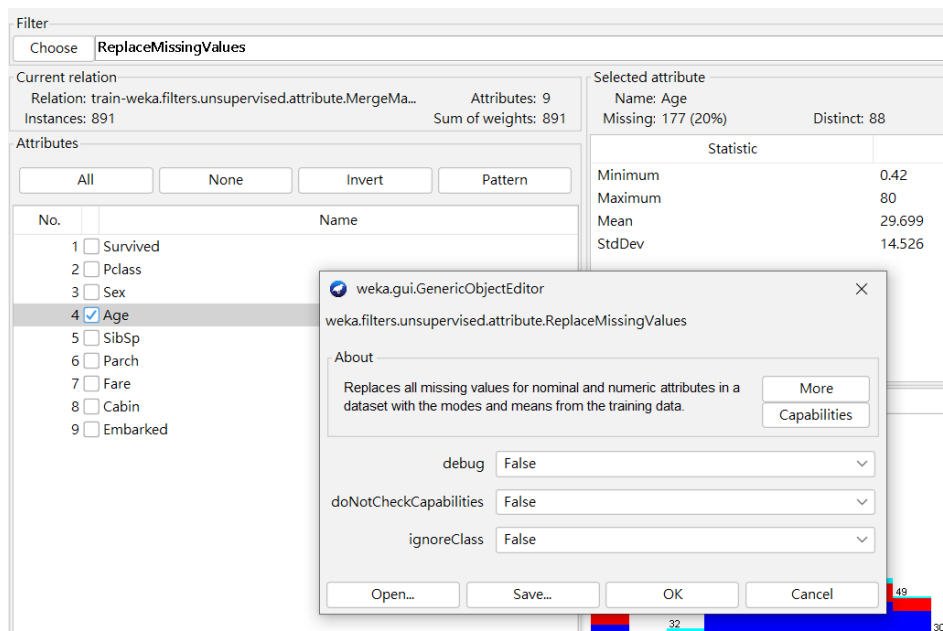
- 用 ReplaceMissingWithUserConstant 以 0 取代 Cabin 的空值



- 執行結果

Selected attribute			
Name: Cabin		Type: Nominal	
Missing: 0 (0%)		Distinct: 2	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	0	687	687
2	1	204	204

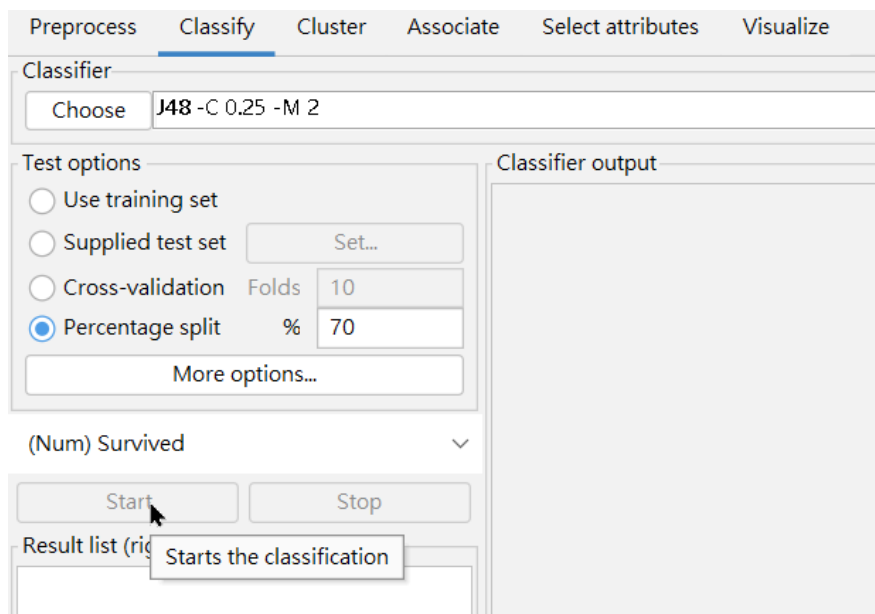
1.3 使用 ReplaceMissingValues 將 Age 的空值以 Age 平均數填入 (5%)



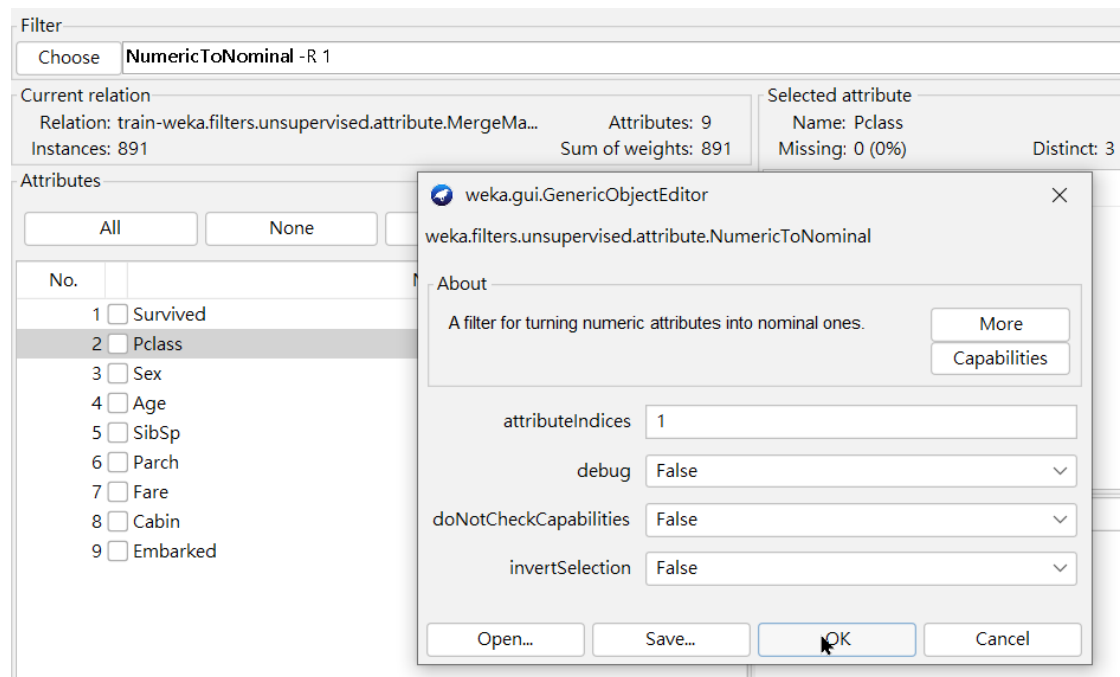
1.4 將 Survived 與 Pclass 轉為 Nominal，並說明為何 Numeric 無法使用在 Decision tree (5%)

- numeric 和 categorical data 無法直接使用 Decision tree，必須將其轉成 nominal data。

下圖為未轉成 Nominal data 執行 J48 的狀況



- 使用 NumericToNominal 將 Survived 轉為 Nominal，Pclass 同理。



- 結果

Relation: train-weka.filters.unsupervised.attribute.MergeManyValues-C8-L1-R1-147-weka.filters.unsupervised.attribute.NumericToNominal

No.	1: Survived Nominal	2: Pclass Nominal	3: Sex Nominal	4: Age Numeric	5: SibSp Numeric	6: Parch Numeric	7: Fare Numeric	8: Cabin Nominal	9: Embarked Nominal
1	0	3	male	22.0	1.0	0.0	7.25	0	S
2	1	1	female	38.0	1.0	0.0	71.2833	1	C
3	1	3	female	26.0	0.0	0.0	7.925	0	S
4	1	1	female	35.0	1.0	0.0	53.1	1	S
5	0	3	male	35.0	0.0	0.0	8.05	0	S
6	0	3	male	29.699...	0.0	0.0	8.4583	0	Q
7	0	1	male	54.0	0.0	0.0	51.8625	1	S
8	0	3	male	2.0	3.0	1.0	21.075	0	S
9	1	3	female	27.0	0.0	2.0	11.1333	0	S

1.5 以 70% 切割訓練資料，使用 J48 對 Survived 進行分類，並截圖分類準確率、混淆矩陣及視覺化的 Decision tree (10%)

The screenshot shows the Weka Classifier window with the J48 classifier selected. The test options are set to 'Percentage split' at 70%. The classifier output shows the following summary:

```
Time taken to build model: 0.04 seconds
=== Evaluation on test split ===
Time taken to test model on test split: 0 seconds
=== Summary ===
Correctly Classified Instances      218          81.6479 %
Incorrectly Classified Instances    49           18.3521 %
Kappa statistic                    0.5769
Mean absolute error                 0.266
Root mean squared error             0.3719
Relative absolute error             56.4711 %
Root relative squared error        77.0588 %
Total Number of Instances          267

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	FRC Area	Class
	0.947	0.408	0.800	0.947	0.867	0.599	0.820	0.833	0
	0.592	0.053	0.866	0.592	0.703	0.599	0.820	0.762	1
Weighted Avg.	0.816	0.278	0.824	0.816	0.807	0.599	0.820	0.807	

Below the summary is the Confusion Matrix:

```
=== Confusion Matrix ===
 a  b  <-- classified as
160  9 | a = 0
 40 58 | b = 1
```

● 準確率：

Correctly Classified Instances	218	81.6479 %
Incorrectly Classified Instances	49	18.3521 %

● 混淆矩陣：

```
=== Confusion Matrix ===
 a  b  <-- classified as
160  9 | a = 0
 40 58 | b = 1
```

● 視覺化 Decision Tree：

