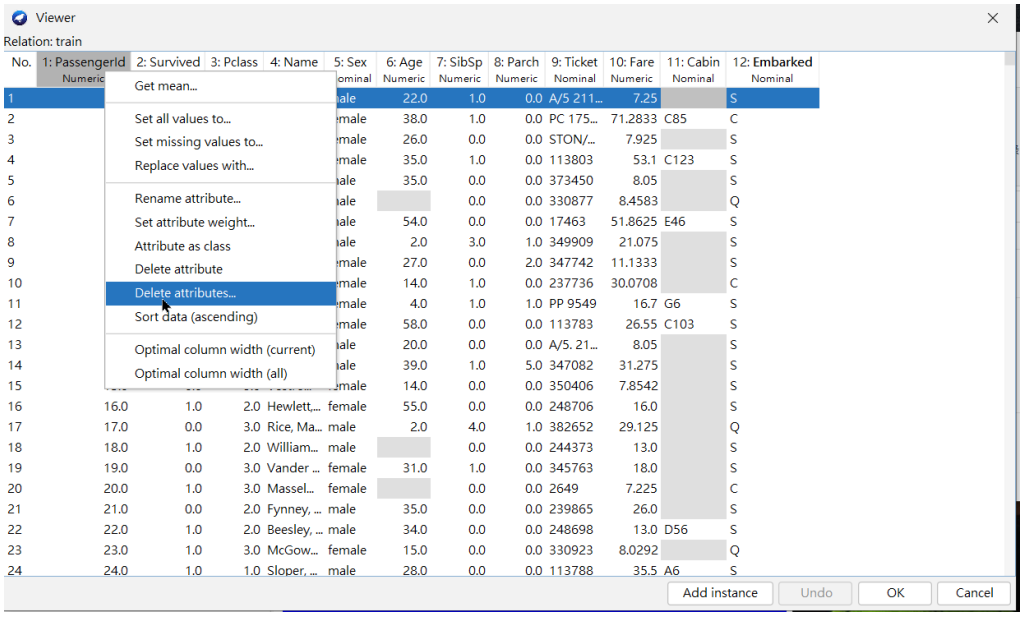
ECT_HW2_107403020

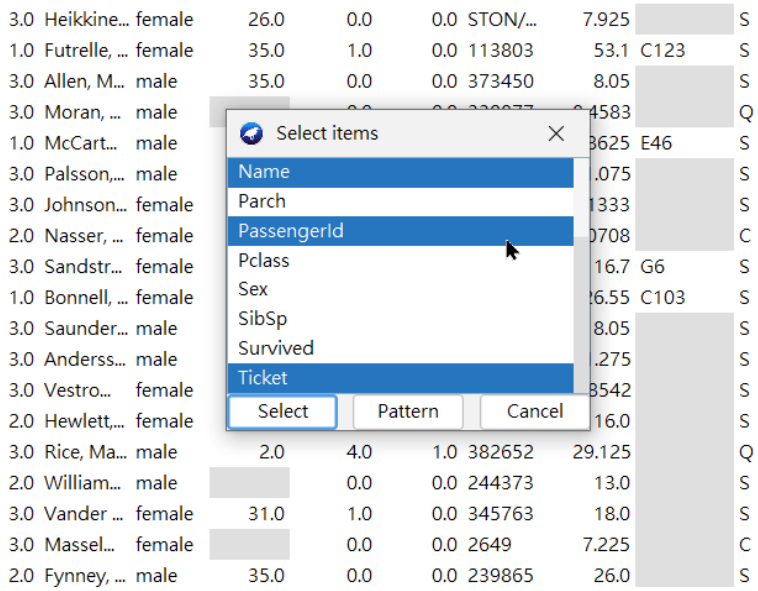
李泳輝

1.1 載入 train.arff，將 PassengerId、Name、Ticket 欄位刪除 (5%)

打開 Edit，點擊右鍵。

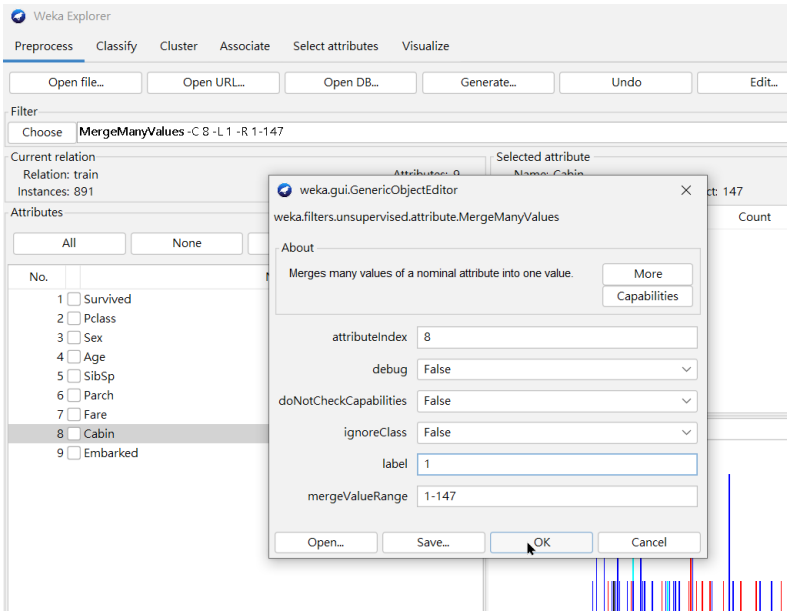


- 選擇欲刪除的 Attributes

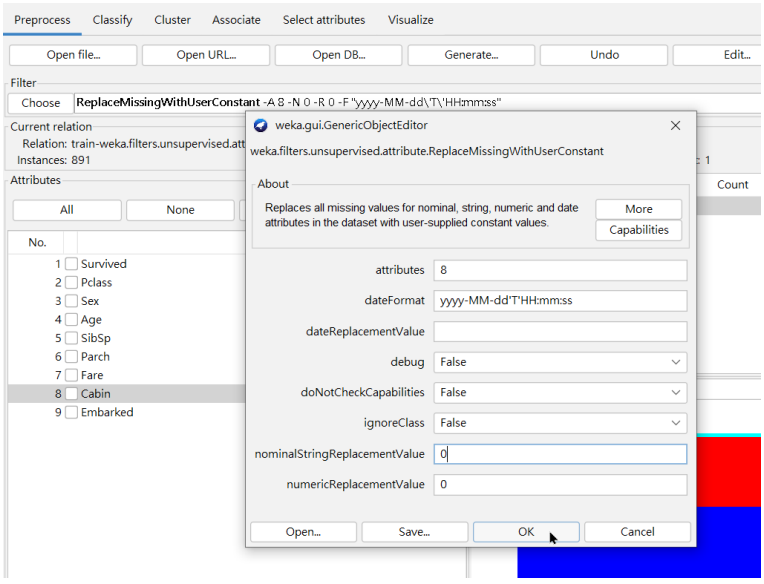


1.2 將 Cabin 的非空值以 1 替代，空值以 0 填入（先用 MergeManyValues 取代非空值，再用 ReplaceMissingWithUserConstant 取代空值） (5%)

- 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

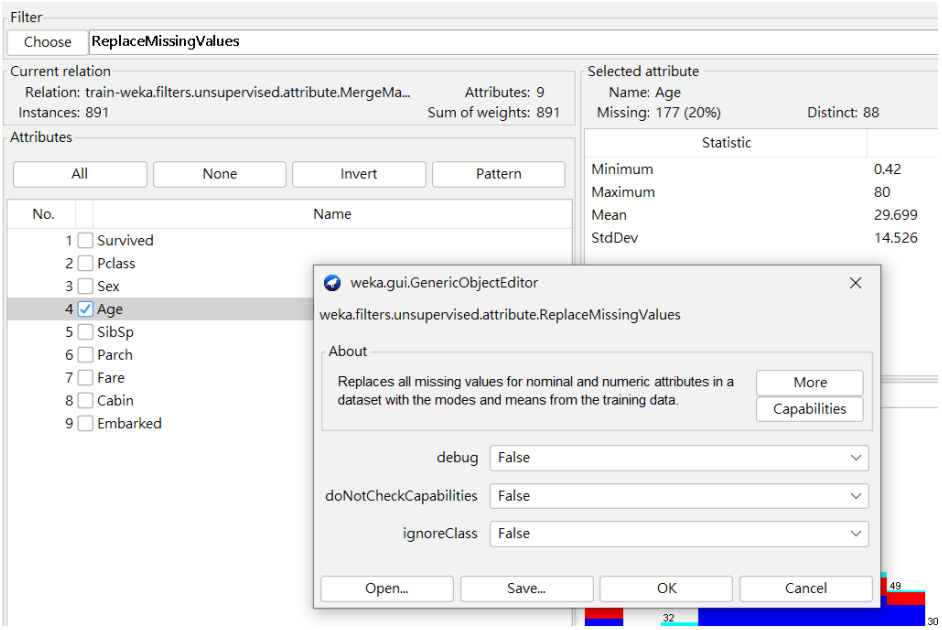
Selected attribute

Name: Cabin　　Type: Nominal

Missing: 0 (0%)　　Distinct: 2　　Unique: 0 (0%)

No.	Label	Count	Weight
1	0	687	687
2	1	204	204

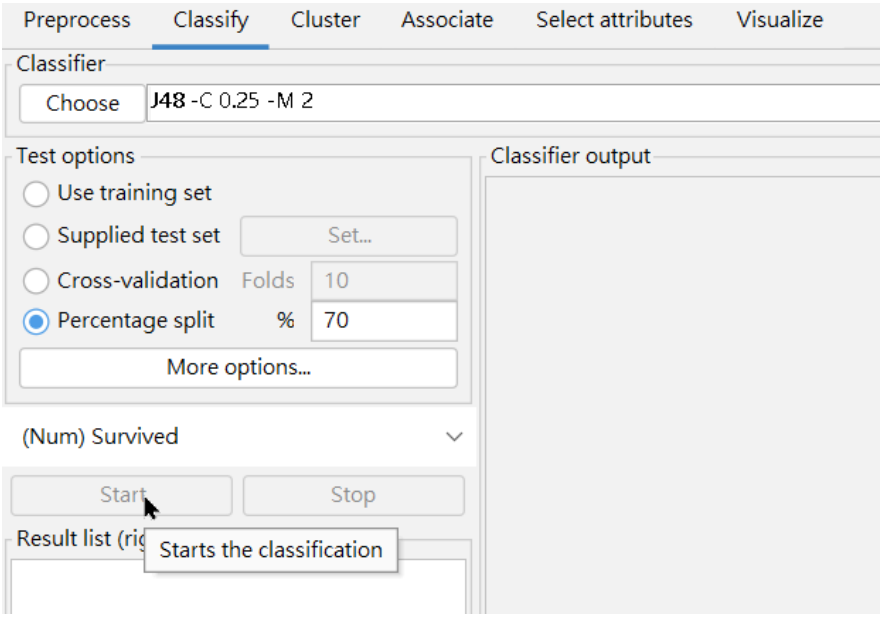
1.3 使用 ReplaceMissingValues 將 Age 的空值以 Age 平均數填入 (5%)



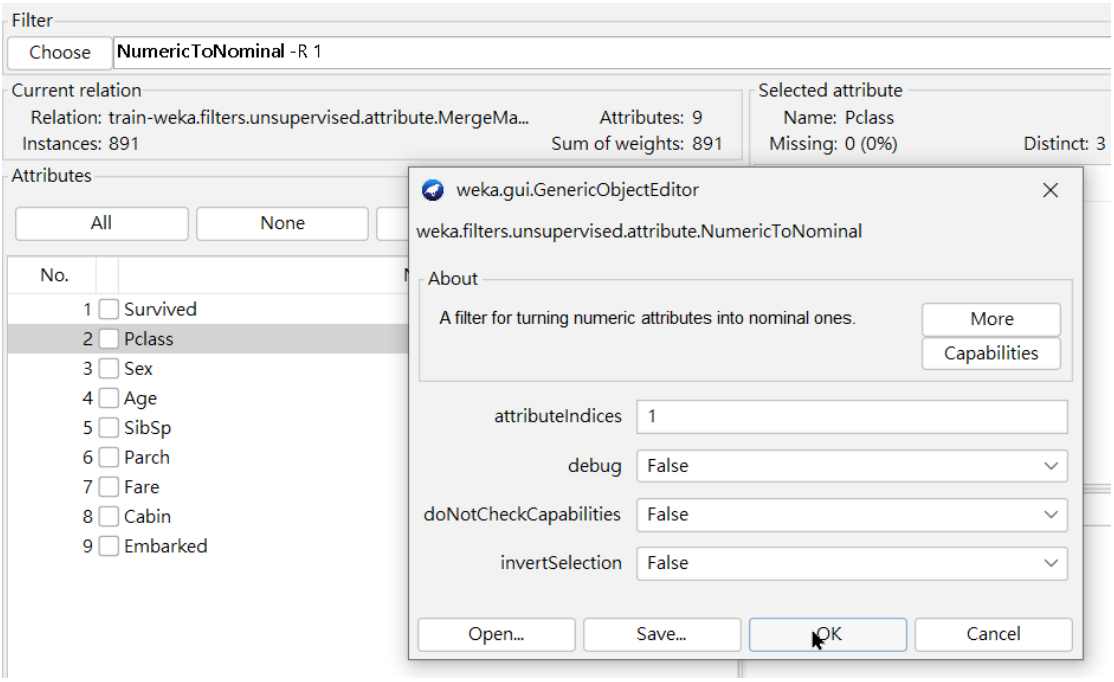
1.4 將 Survived 與 Pclass 轉為 Nominal，並說明為何 Numeric 無法使用在 Decision tree (5%)

- 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。

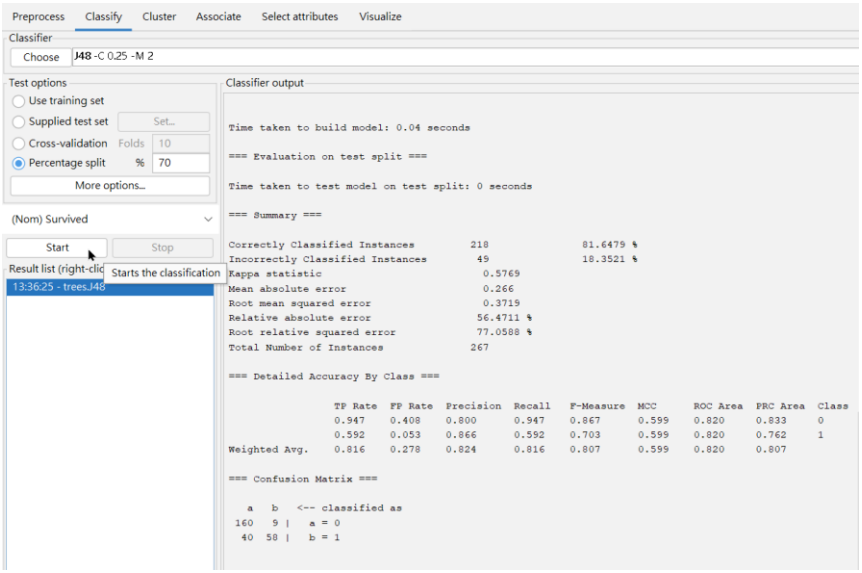


- 結果

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

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%)



- 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 :