1. 載入customer_churn.csv，列出資料筆數、屬性數量以及每個欄位的空值個數(5%)

資料屬性(numeric or nominal)

1: CustomerID	2: Churn	3: Tenure	4: PreferredLoginDevice	5: CityTier	6: WarehouseToHome	7: PreferredPaymentMode	8: Gender	9: HourSpendOnApp	10: NumberOfDeviceRegistered
Numeric	Numeric	Numeric	Nominal	Numeric	Numeric	Nominal	Nominal	Numeric	Numeric

11: PreferedOrderCat	12: SatisfactionScore	13: MaritalStatus	14: NumberOfAddress	15: Complain	16: OrderAmountHikeFromlastYear	17: CouponUsed	18: OrderCount	19: DaySinceLastOrder	20: CashbackAmount
Nominal	Numeric	Nominal	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric

資料筆數 ：3083

Current relation
Relation: customer_churn
Instances: 3083
Attributes: 20
Sum of weights: 3083

空值個數 (以下欄位是有空值的，如圖)

Selected attribute
Name: Tenure
Missing: 153 (5%)
Distinct: 35
Type: Numeric
Unique: 3 (0%)

Selected attribute
Name: WarehouseToHome
Missing: 154 (5%)
Distinct: 33
Type: Numeric
Unique: 1 (0%)

Selected attribute
Name: HourSpendOnApp
Missing: 150 (5%)
Distinct: 6
Type: Numeric
Unique: 1 (0%)

Selected attribute
Name: OrderAmountHikeFromlastYear
Missing: 131 (4%)
Distinct: 16
Type: Numeric
Unique: 0 (0%)

Selected attribute
Name: CouponUsed
Missing: 126 (4%)
Distinct: 17
Type: Numeric
Unique: 3 (0%)

Selected attribute
Name: OrderCount
Missing: 128 (4%)
Distinct: 16
Type: Numeric
Unique: 0 (0%)

Selected attribute
Name: DaySinceLastOrder
Missing: 166 (5%)
Distinct: 20
Type: Numeric
Unique: 1 (0%)

2. 請刪除重覆多餘的資料(僅保留一筆)，並列出剩餘的資料筆數(5%)

使用RemoveDuplicates

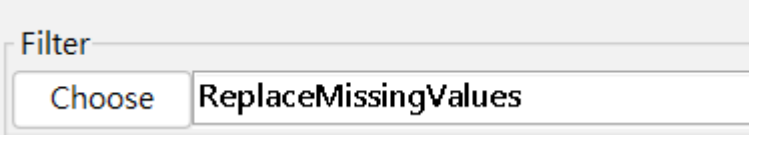
Filter
Choose RemoveDuplicates

資料總數變成3078

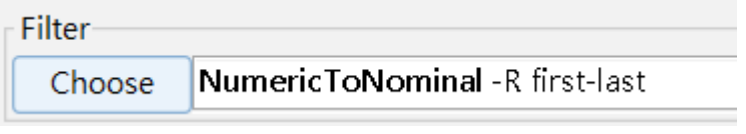
Current relation
Relation: customer_churn-weka.filters.unsupervised.instance.RemoveDuplicates
Instances: 3078
Attributes: 20
Sum of weights: 3078

3. 資料前處理(5%)

使用ReplaceMissingValues 處理空值

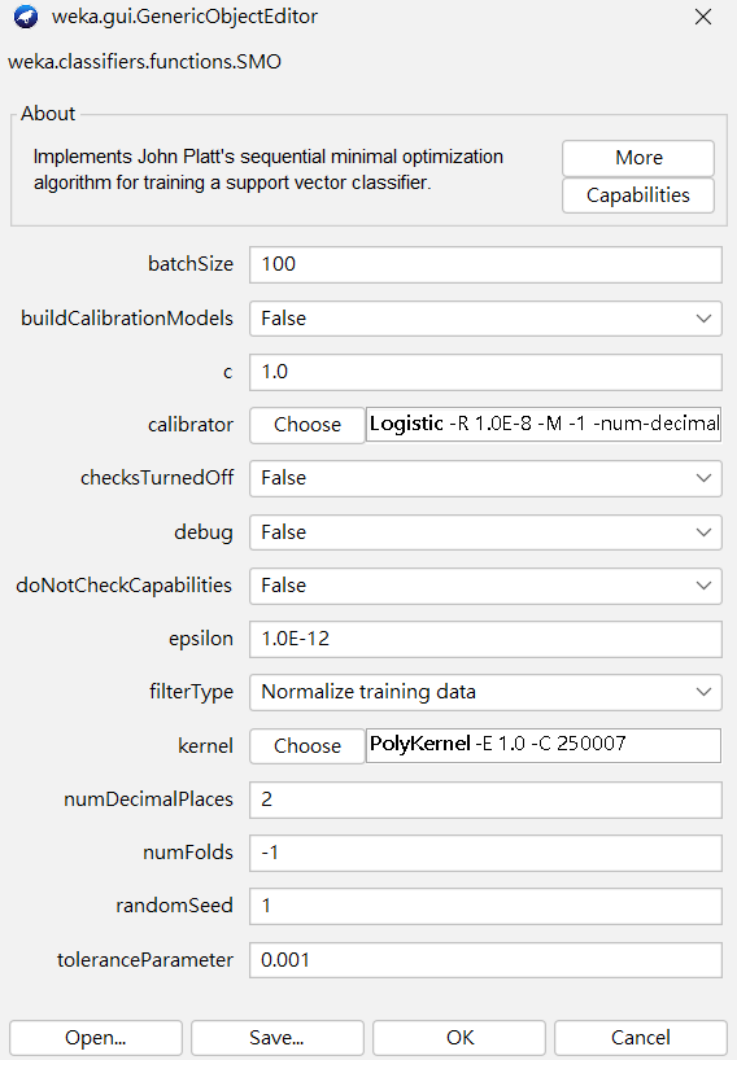


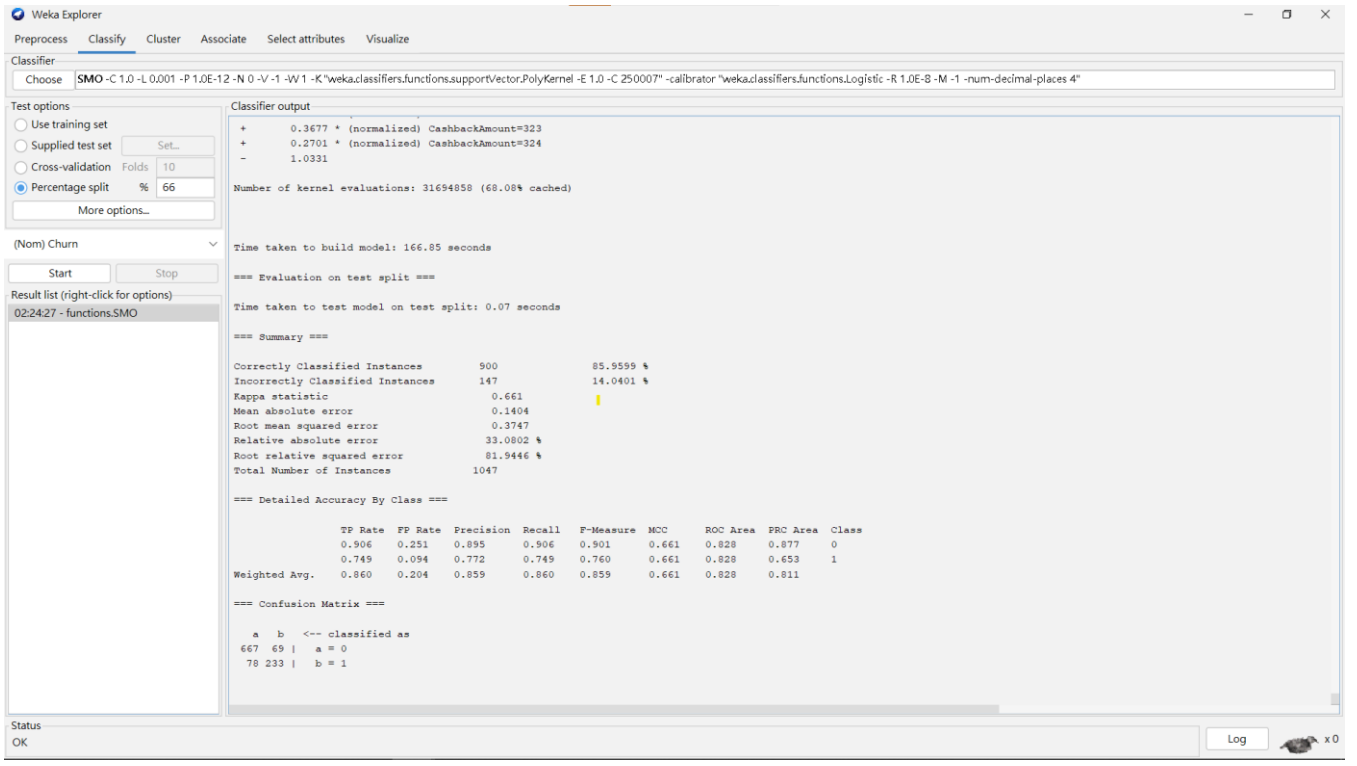
再用NumericToNominal



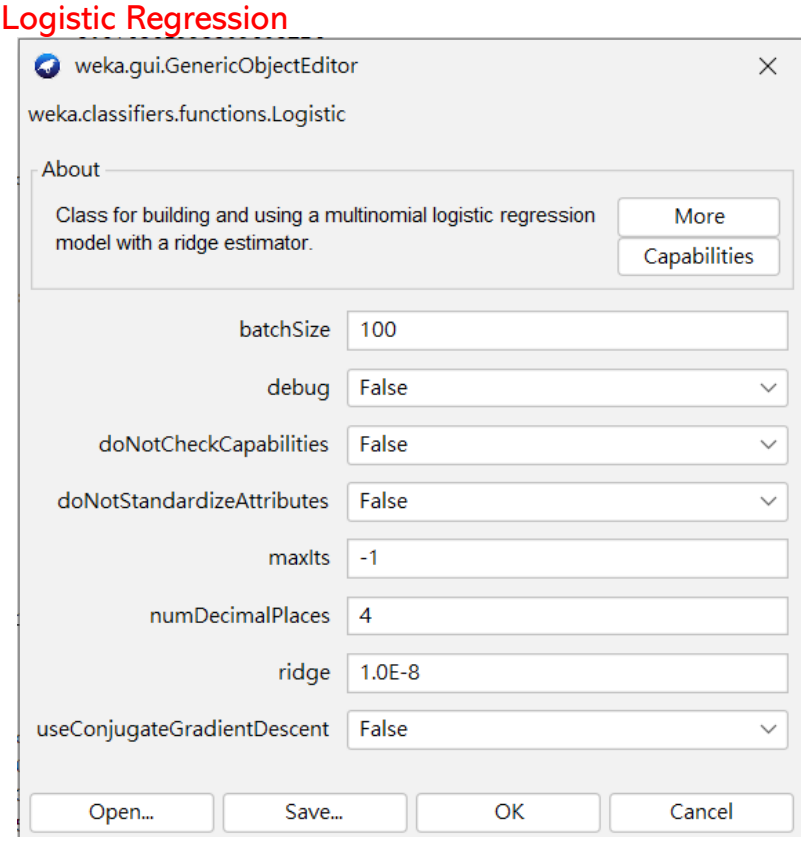
4. 訓練、測試SVM、Logistic Regression、Decision Tree模型，請以Accuracy評估模型表現(10%)

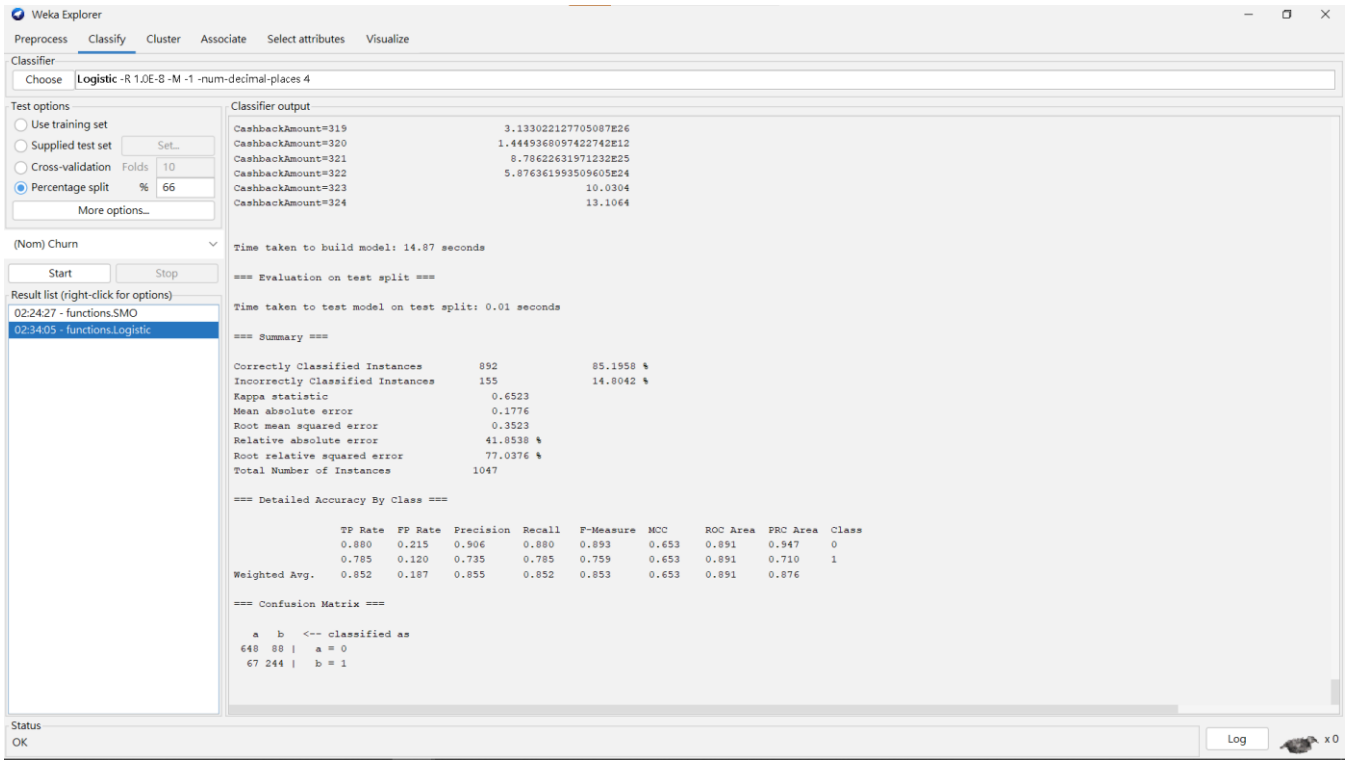
SVM (SMO)結果



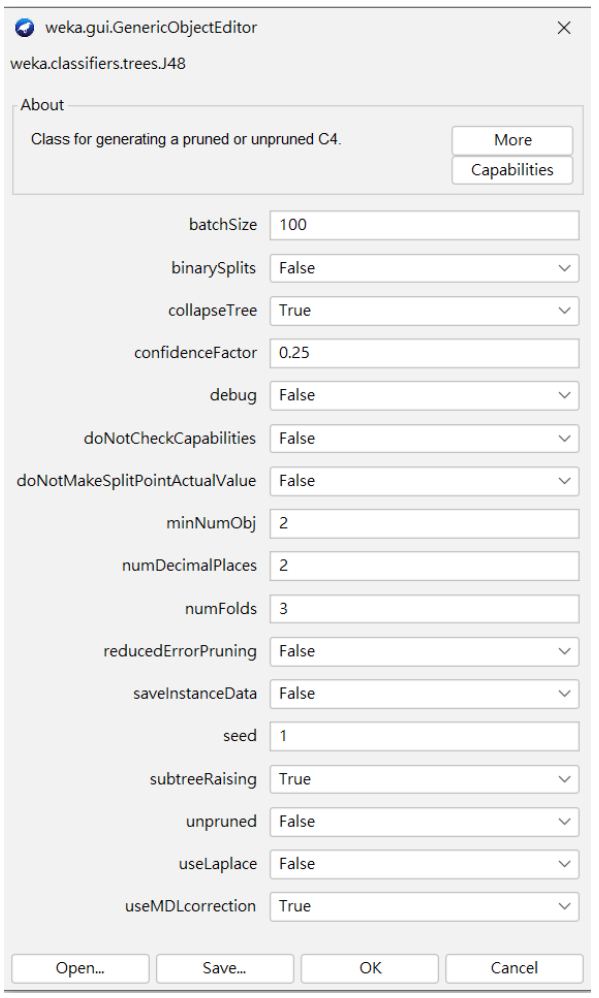


Logistic Regression





Decision Tree



Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose: J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set — Set...
- Cross-validation — Folds 10
- Percentage split — % 66

More options...

(Nom) Churn

Start | Stop

Result list (right-click for options)

- 02:24:27 - functions.SMO
- 02:34:05 - functions.Logistic
- 02:36:11 - trees.J48

Classifier output

```
|   Tenure = 60: 0 (1.0)
|   Tenure = 61: 0 (1.0)

Number of Leaves  :     790

Size of the tree :      839


Time taken to build model: 0.07 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances         884               84.4317 %
Incorrectly Classified Instances       163               15.5683 %
Kappa statistic                          0.6276
Mean absolute error                      0.2261
Root mean squared error                  0.3458
Relative absolute error                 53.2779 %
Root relative squared error             75.6245 %
Total Number of Instances             1047

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.889    0.260    0.890      0.889    0.889      0.628    0.871     0.927     0
                 0.740    0.111    0.737      0.740    0.738      0.628    0.871     0.724     1
Weighted Avg.    0.844    0.216    0.844      0.844    0.844      0.628    0.871     0.867

=== Confusion Matrix ===

   a   b   <-- classified as
 654  82 |   a = 0
  81 230 |   b = 1
```

Status: OK

Log | x 0