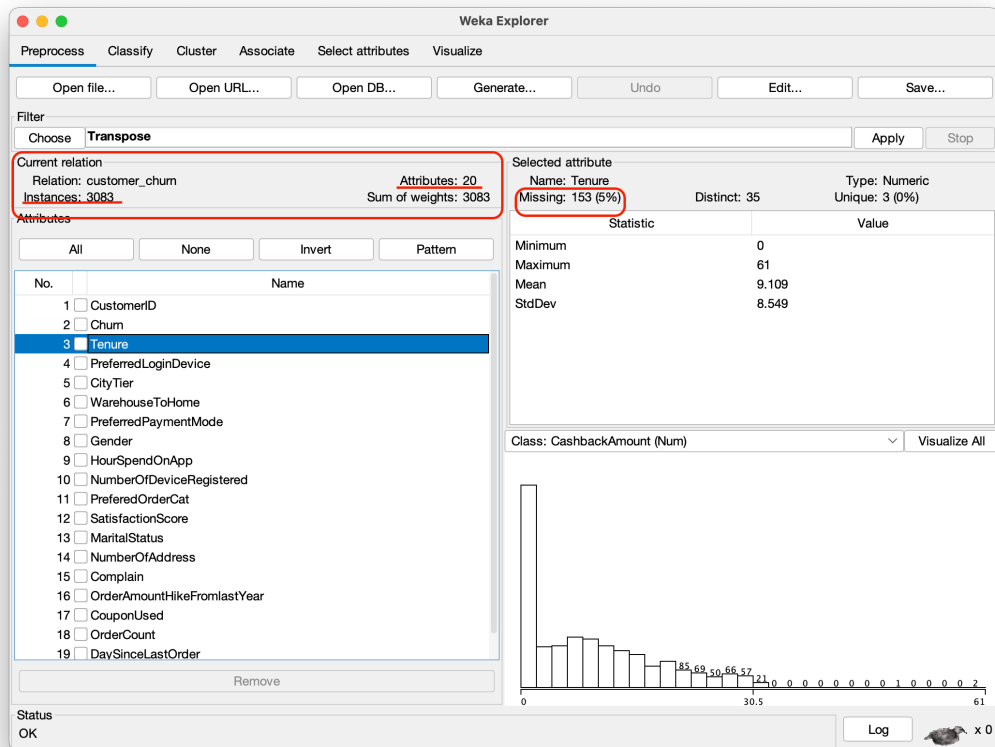


ECT_HW3_第三部分

1.



左方為資料筆數(3083)、屬性數量(20)，右方為該欄位空值個數（此圖為Tenure欄位，其餘以此類推）下方圖列出所有欄位具空值的情況：

各欄位空值數量：

```
Out[3]: CustomerID          0
        Churn              0
        Tenure             153
        PreferredLoginDevice 0
        CityTier           0
        WarehouseToHome     154
        PreferredPaymentMode 0
        Gender              0
        HourSpendOnApp       150
        NumberOfDeviceRegistered 0
        PreferredOrderCat    0
        SatisfactionScore    0
        MaritalStatus        0
        NumberOfAddress      0
        Complain             0
        OrderAmountHikeFromlastYear 131
        CouponUsed           126
        OrderCount           128
        DaySinceLastOrder    166
        CashbackAmount       0
        dtype: int64
```

2.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter
Choose **RemoveDuplicates** Apply Stop

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

Attributes
All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> CustomerID
2	<input type="checkbox"/> Churn
3	<input type="checkbox"/> Tenure
4	<input type="checkbox"/> PreferredLoginDevice
5	<input type="checkbox"/> CityTier
6	<input type="checkbox"/> WarehouseToHome
7	<input type="checkbox"/> PreferredPaymentMode
8	<input type="checkbox"/> Gender
9	<input type="checkbox"/> HourSpendOnApp
10	<input type="checkbox"/> NumberOfDeviceRegistered
11	<input type="checkbox"/> PreferredOrderCat
12	<input type="checkbox"/> SatisfactionScore
13	<input type="checkbox"/> MaritalStatus
14	<input type="checkbox"/> NumberOfAddress
15	<input type="checkbox"/> Complain
16	<input type="checkbox"/> OrderAmountLikeFromlastYear
17	<input type="checkbox"/> CouponUsed
18	<input type="checkbox"/> OrderCount
19	<input type="checkbox"/> DaySinceLastOrder

Remove

Status
OK

Selected attribute
Name: CustomerID Type: Numeric
Missing: 0 (0%) Distinct: 3078 Unique: 3078 (100%)

Statistic	Value
Minimum	50001
Maximum	55628
Mean	52811.623
StdDev	1614.867

Class: CashbackAmount (Num) Visualize All

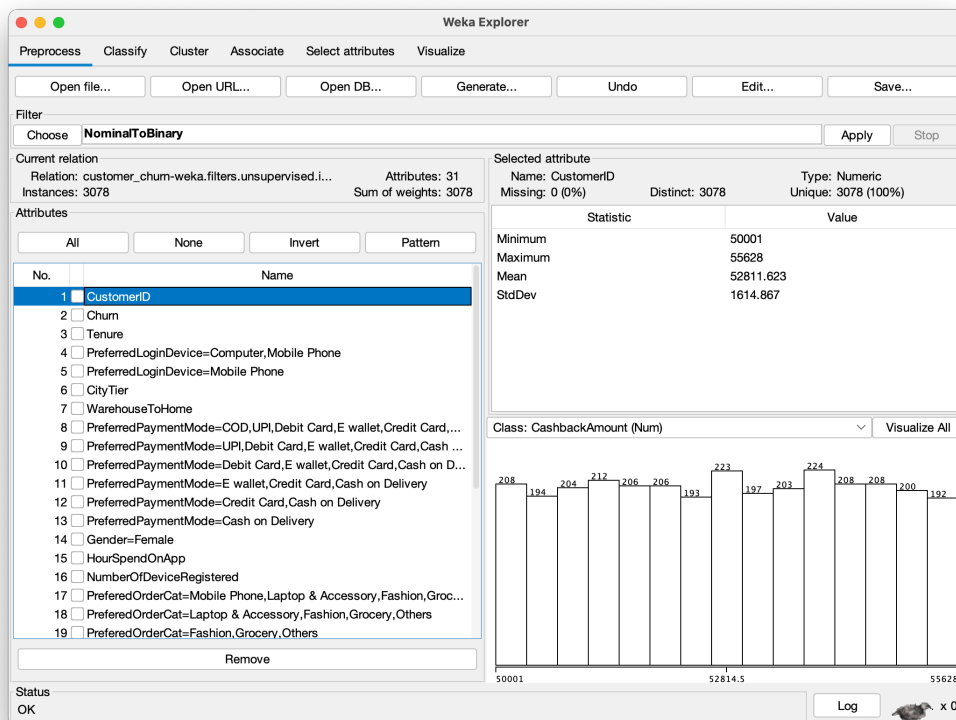
208 194 204 212 206 206 193 223 197 203 224 208 208 200 192

50001 52814.5 55628

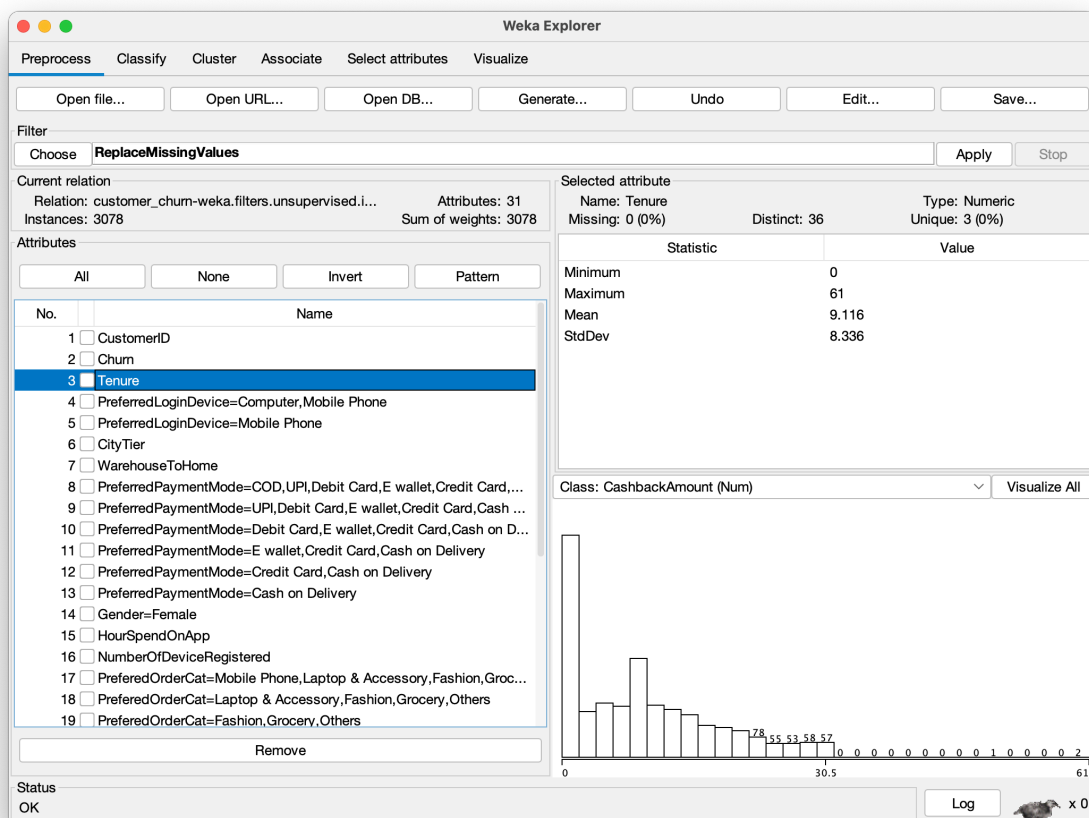
Log x 0

剩餘資料筆數：3078

3. 先將Nominal資料改成Binary

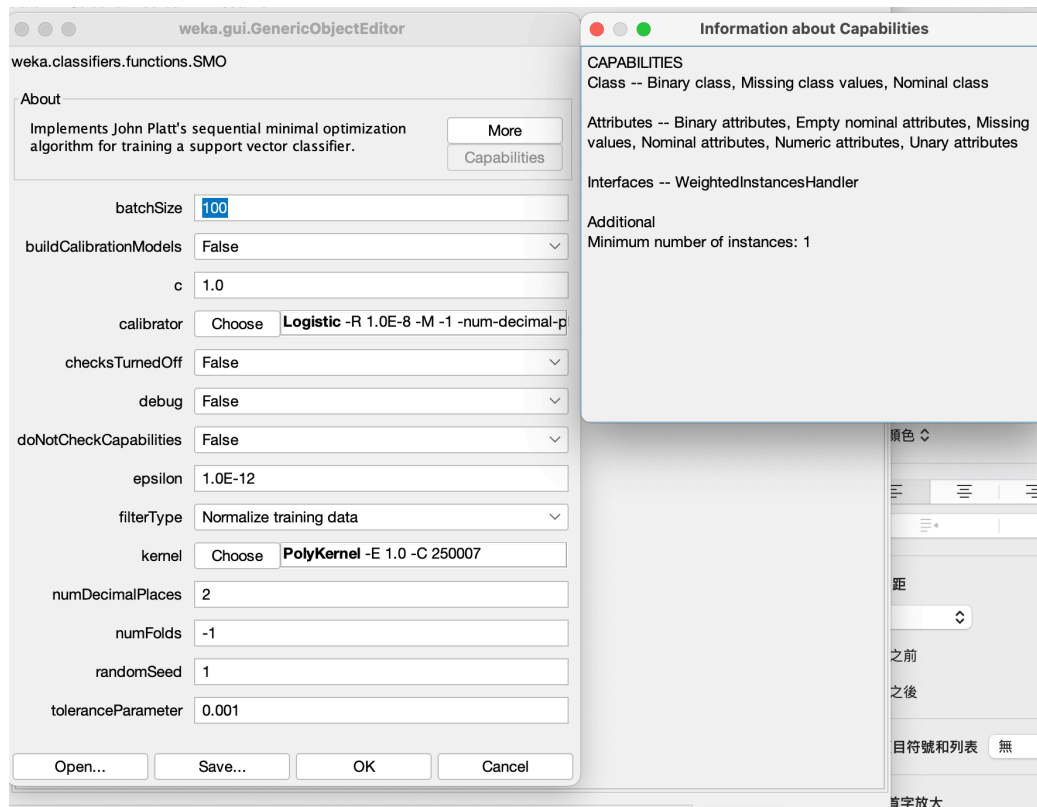


再將空值以該欄位平均數帶入

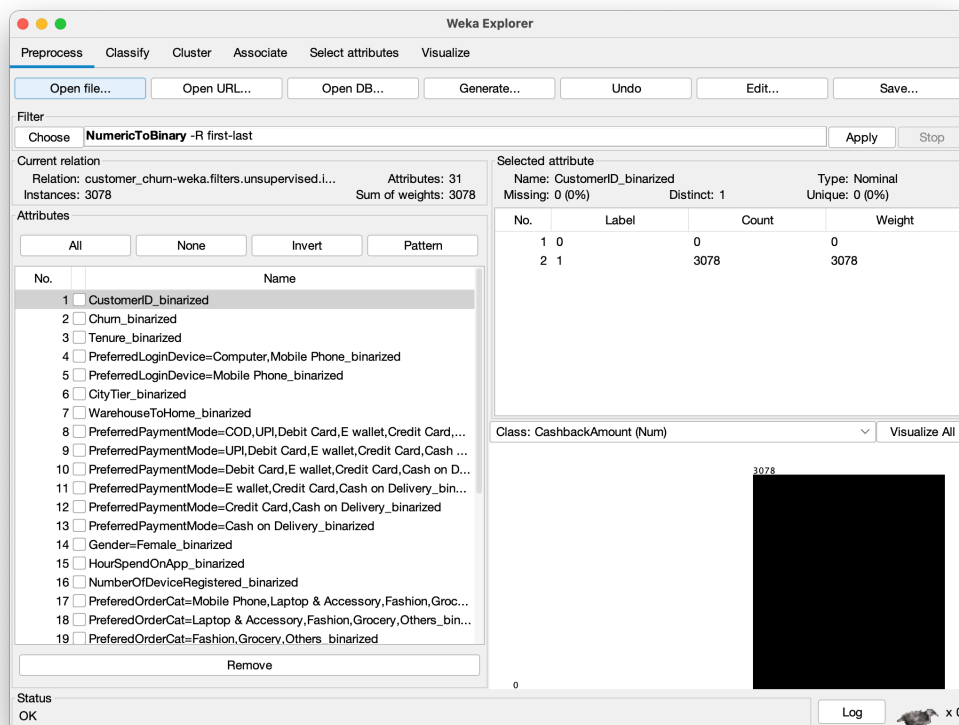


4.

因SMO 只能處理Binary , Nominal



故將資料全部轉成binary type



跑SVM(SMO)結果：

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.funct'.

Test options:

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

(Nom) Churn_binarized

Start Stop

Result list (right-click for options)

- 16:46:08 - functions.SMO

Classifier output

```
+ -0.2878 * (normalized) CashbackAmount
+ 1.0936

Number of kernel evaluations: 5218632 (68.003% cached)

Time taken to build model: 0.53 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances      835      79.7517 %
Incorrectly Classified Instances    212      20.2483 %
Kappa statistic                    0.4731
Mean absolute error                 0.2025
Root mean squared error             0.45
Relative absolute error             47.7075 %
Root relative squared error         98.4078 %
Total Number of Instances          1047

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.914    0.479    0.819     0.914    0.864     0.484    0.718    0.809     0
Weighted Avg.   0.521    0.086    0.720     0.521    0.604     0.484    0.718    0.517     1

=== Confusion Matrix ===

  a  b  <-- classified as
673 63 | a = 0
149 162 | b = 1
```

Status: OK Log x 0

跑Logistic Regression結果：

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'J48 -C 0.25 -M 2'.

Test options:

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

(Nom) Churn_binarized

Start Stop

Result list (right-click for options)

- 16:46:08 - functions.SMO
- 17:02:03 - functions.SimpleLogistic
- 17:03:16 - functions.Logistic
- 17:04:02 - trees.J48

Classifier output

```
[MaritalStatus=Married,Divorced_binarized=1] * -0.42 +
[MaritalStatus=Divorced_binarized=1] * 0.11 +
[Complain_binarized=1] * 0.69 +
[CouponUsed_binarized=1] * 0.21 +
[DaySinceLastOrder_binarized=1] * -0.32

Time taken to build model: 0.59 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances      829      79.1786 %
Incorrectly Classified Instances    218      20.8214 %
Kappa statistic                    0.4614
Mean absolute error                 0.3053
Root mean squared error             0.388
Relative absolute error             71.9305 %
Root relative squared error         84.8538 %
Total Number of Instances          1047

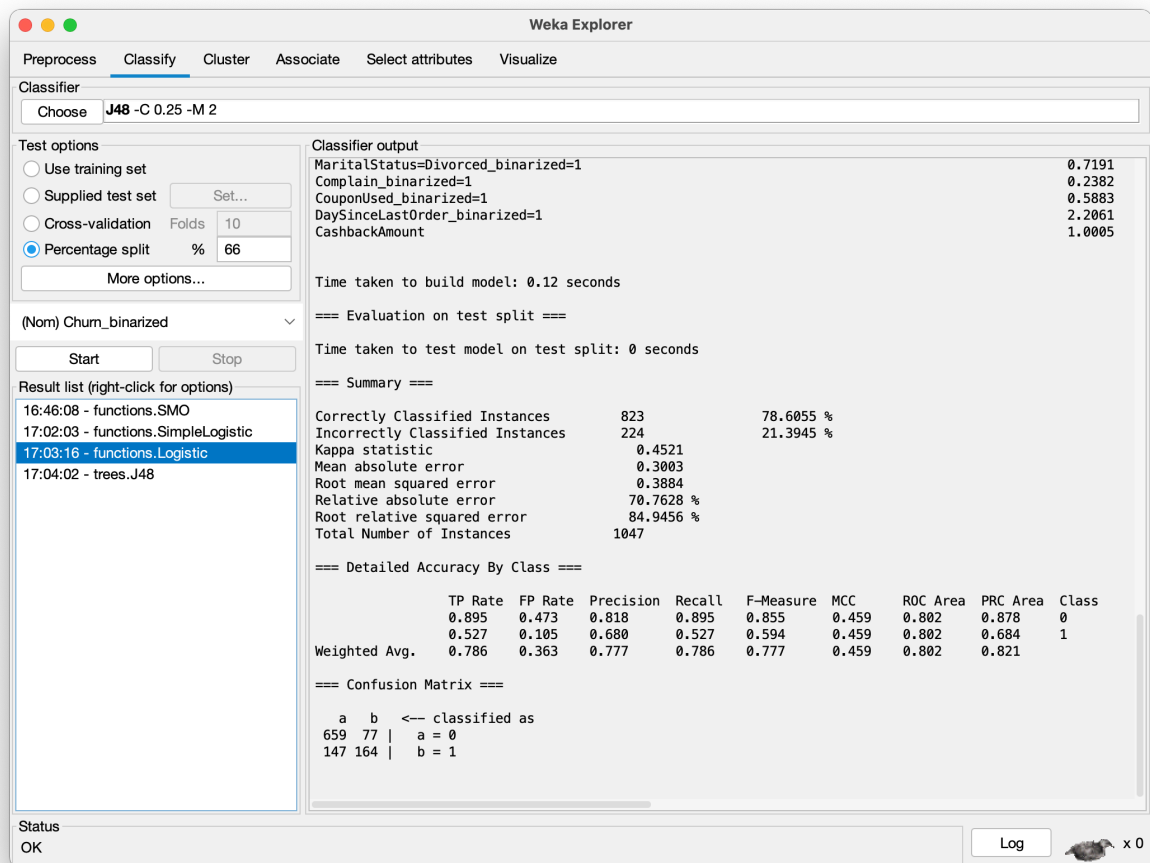
=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.906    0.479    0.817     0.906    0.860     0.471    0.801    0.873     0
Weighted Avg.   0.521    0.094    0.701     0.521    0.598     0.471    0.801    0.683     1

=== Confusion Matrix ===

  a  b  <-- classified as
667 69 | a = 0
149 162 | b = 1
```

Status: OK Log x 0



跑Decision Tree(J48)結果：

