

電商技術 HW3 第三部分

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

customer_churn.csv 原始資料共3083筆、屬性數20:

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

以下為每個欄位的空值個數：

1. CustomerID 無空值

Selected attribute	
Name: CustomerID	Type: Numeric
Missing: 0 (0%)	Distinct: 3078
	Unique: 3075 (100%)
Statistic	Value
Minimum	50001
Maximum	55628
Mean	52807.294
StdDev	1617.13

2. Churn 無空值

Selected attribute	
Name: Churn	Type: Numeric
Missing: 0 (0%)	Distinct: 2
	Unique: 0 (0%)
Statistic	Value
Minimum	0
Maximum	1
Mean	0.308
StdDev	0.462

3. Tenure 有153筆空值

Selected attribute		
Name: Tenure		Type: Numeric
Missing: 153 (5%)	Distinct: 35	Unique: 3 (0%)
Statistic	Value	
Minimum	0	
Maximum	61	
Mean	9.109	
StdDev	8.549	

4. PreferredLoginDevice 無空值

Selected attribute			
Name: PreferredLoginDevice		Type: Nominal	
Missing: 0 (0%)		Distinct: 3	Unique: 0 (0%)
No.	Label	Count	Weight
1	Computer	928	928
2	Mobile Phone	1458	1458
3	Phone	697	697

5. CityTier 無空值

Selected attribute		
Name: CityTier		Type: Numeric
Missing: 0 (0%)	Distinct: 3	Unique: 0 (0%)
Statistic	Value	
Minimum	1	
Maximum	3	
Mean	1.682	
StdDev	0.925	

6. WarehouseToHome 154筆空值

Selected attribute	
Name: WarehouseToHome	Type: Numeric
Missing: 154 (5%)	Distinct: 33
	Unique: 1 (0%)
Statistic	Value
Minimum	5
Maximum	126
Mean	15.771
StdDev	8.557

7. PreferredPaymentMode 無空值

Selected attribute			
Name: PreferredPaymentMode		Type: Nominal	
Missing: 0 (0%)		Distinct: 7	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	Debit Card	1255	1255
2	Credit Card	789	789
3	E wallet	368	368
4	CC	147	147
5	COD	225	225
6	UPI	222	222
7	Cash on Delivery	77	77

8. Gender 無空值

Selected attribute			
Name: Gender		Type: Nominal	
Missing: 0 (0%)		Distinct: 2	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	Male	1854	1854
2	Female	1229	1229

9. HourSpendOnApp 150筆空值

Selected attribute		
Name: HourSpendOnApp		Type: Numeric
Missing: 150 (5%)	Distinct: 6	Unique: 1 (0%)
Statistic	Value	
Minimum	0	
Maximum	5	
Mean	2.946	
StdDev	0.719	

10. NumberOfDeviceRegistered 無空值

Selected attribute		
Name: NumberOfDeviceRegistered		Type: Numeric
Missing: 0 (0%)	Distinct: 6	Unique: 0 (0%)
Statistic	Value	
Minimum	1	
Maximum	6	
Mean	3.75	
StdDev	1.005	

11. PreferredOrderCat 無空值

Selected attribute			
Name: PreferredOrderCat			Type: Nominal
Missing: 0 (0%)		Distinct: 6	Unique: 0 (0%)
No.	Label	Count	Weight
1	Mobile Phone	772	772
2	Grocery	198	198
3	Laptop & Accessory	1066	1066
4	Mobile	463	463
5	Fashion	449	449
6	Others	135	135

12. SatisfactionScore 無空値

Selected attribute	
Name: SatisfactionScore	Type: Numeric
Missing: 0 (0%)	Distinct: 5
	Unique: 0 (0%)
Statistic	Value
Minimum	1
Maximum	5
Mean	3.134
StdDev	1.382

13. MaritalStatus 無空値

Selected attribute			
Name: MaritalStatus		Type: Nominal	
Missing: 0 (0%)		Distinct: 3	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	Single	1102	1102
2	Married	1547	1547
3	Divorced	434	434

14. NumberOfAddress 無空値

Selected attribute	
Name: NumberOfAddress	Type: Numeric
Missing: 0 (0%)	Distinct: 14
	Unique: 3 (0%)
Statistic	Value
Minimum	1
Maximum	21
Mean	4.335
StdDev	2.663

15. Complain 無空值

Selected attribute		
Name: Complain		Type: Numeric
Missing: 0 (0%)	Distinct: 2	Unique: 0 (0%)
Statistic	Value	
Minimum	0	
Maximum	1	
Mean	0.328	
StdDev	0.47	

16. OrderAmountHikeFromlastYear 131筆空值

Selected attribute		
Name: OrderAmountHikeFromlastYear		Type: Numeric
Missing: 131 (4%)	Distinct: 16	Unique: 0 (0%)
Statistic	Value	
Minimum	11	
Maximum	26	
Mean	15.714	
StdDev	3.765	

17. CouponUsed 126筆空值

Selected attribute		
Name: CouponUsed		Type: Numeric
Missing: 126 (4%)	Distinct: 17	Unique: 3 (0%)
Statistic	Value	
Minimum	0	
Maximum	16	
Mean	1.753	
StdDev	1.886	

18. OrderCount 128筆空值

Selected attribute		
Name: OrderCount		Type: Numeric
Missing: 128 (4%)	Distinct: 16	Unique: 0 (0%)
Statistic	Value	
Minimum	1	
Maximum	16	
Mean	2.994	
StdDev	2.948	

19. DaySinceLastOrder 166筆空值

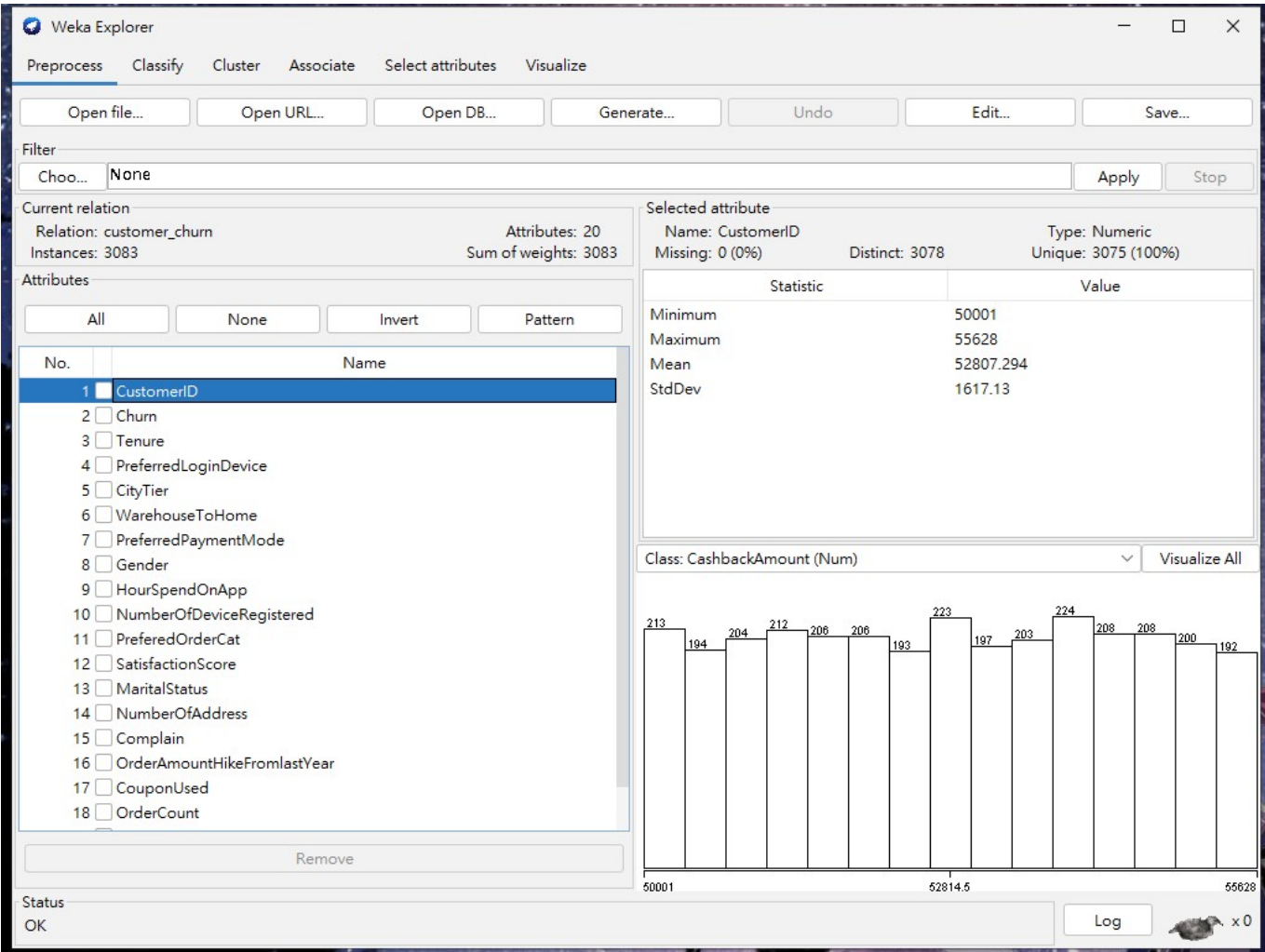
Selected attribute		
Name: DaySinceLastOrder		Type: Numeric
Missing: 166 (5%)	Distinct: 20	Unique: 1 (0%)
Statistic	Value	
Minimum	0	
Maximum	46	
Mean	4.349	
StdDev	3.609	

20. CashbackAmount 無空值

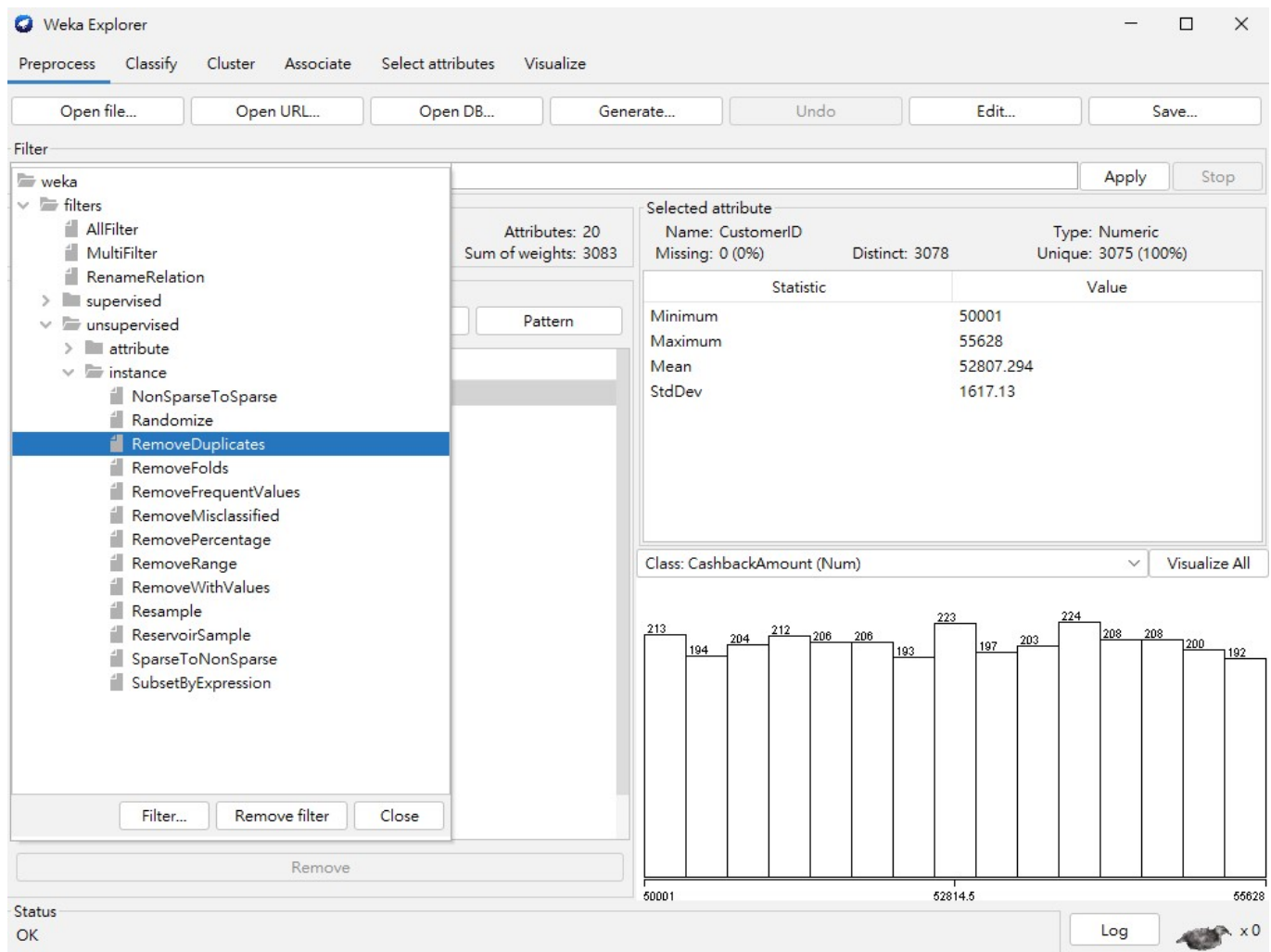
Selected attribute		
Name: CashbackAmount		Type: Numeric
Missing: 0 (0%)	Distinct: 212	Unique: 20 (1%)
Statistic	Value	
Minimum	0	
Maximum	324	
Mean	174.904	
StdDev	47.829	

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

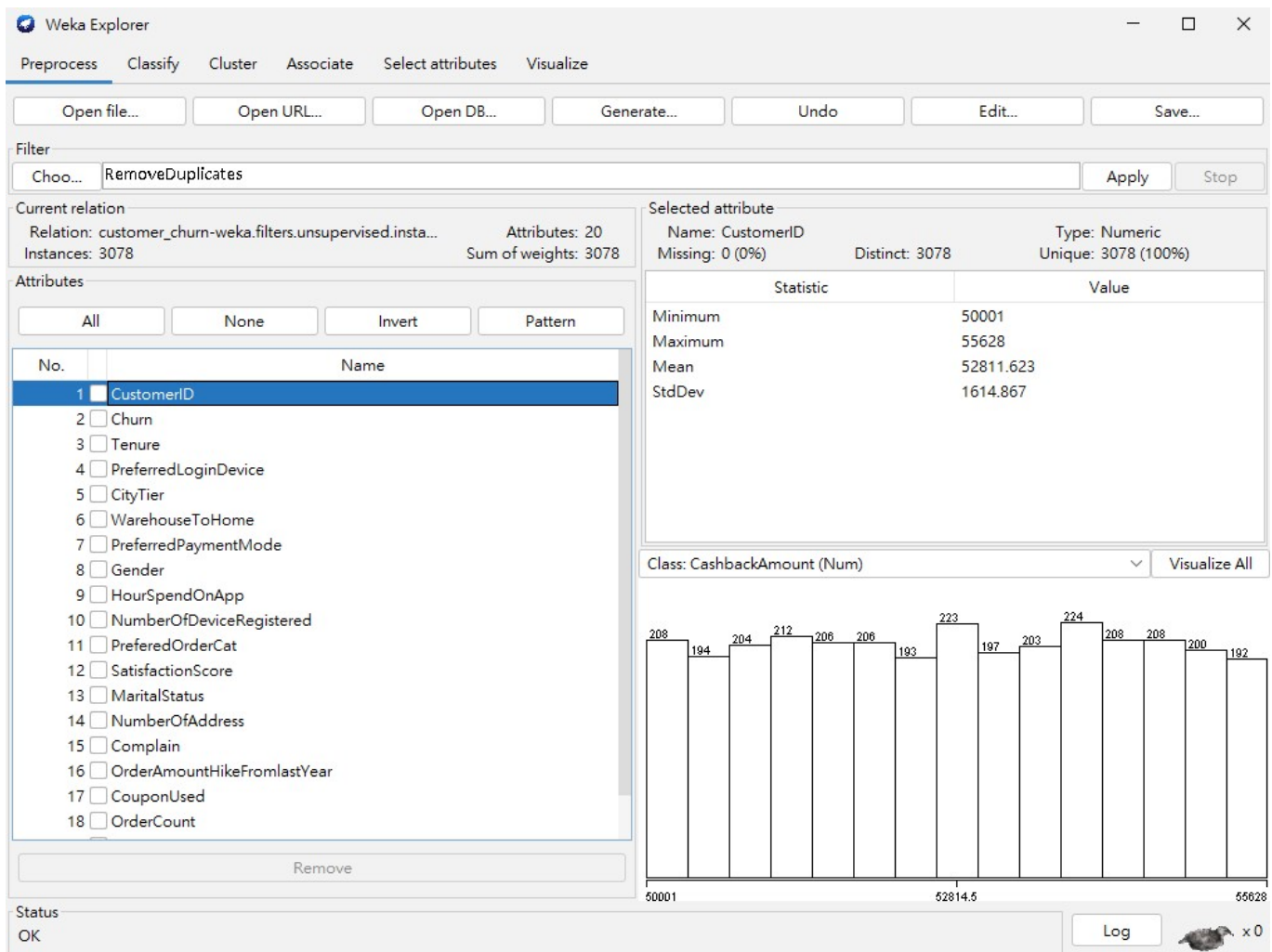
有重複的 CustomerID 表示有重複的紀錄，必須將其刪除：



透過 filter 找到 RemoveDuplicates 刪去重複資料：



Apply 之後，可以觀察到已經沒有重複的 CustomerID，亦即沒有重複的資料：

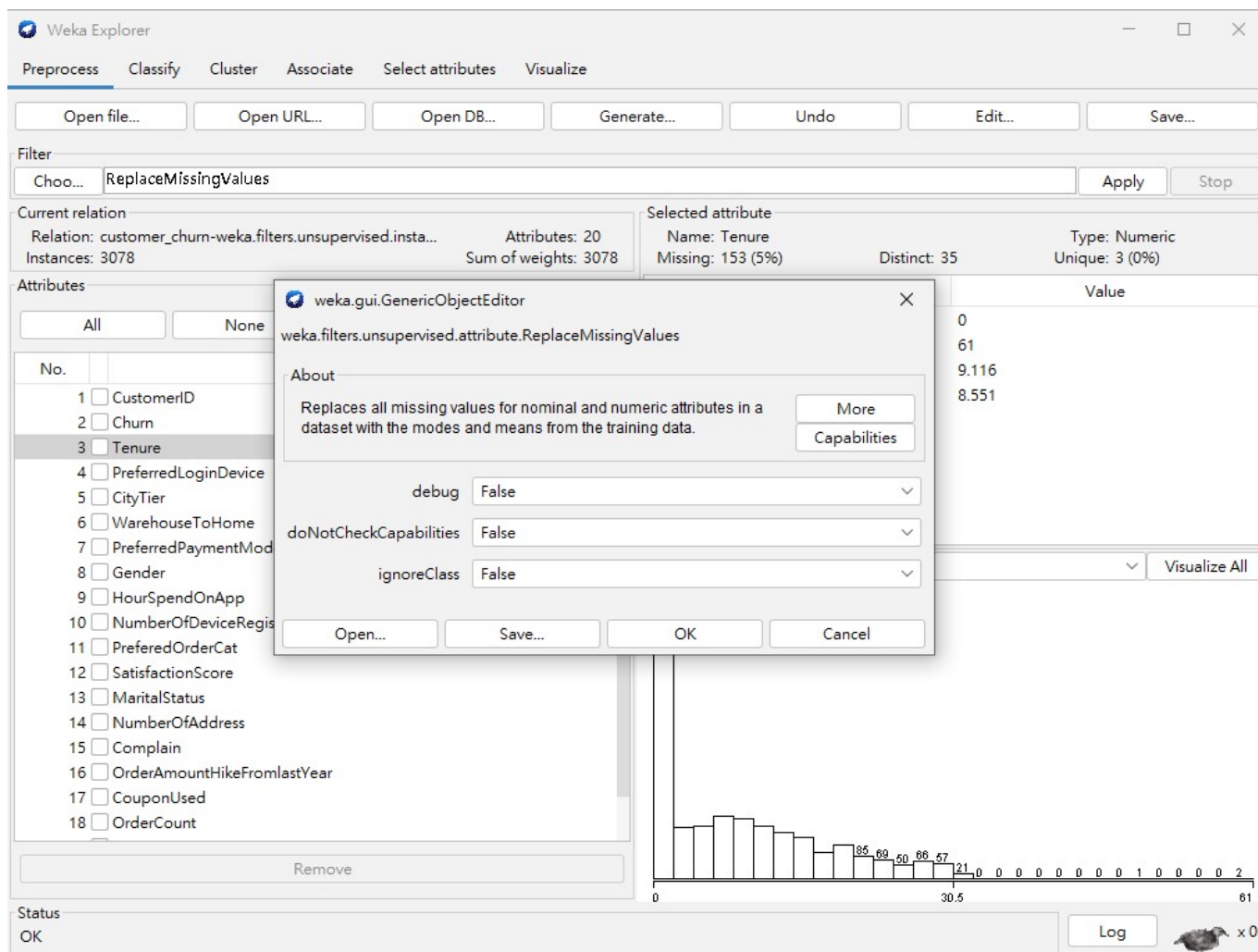


3.資料前處理

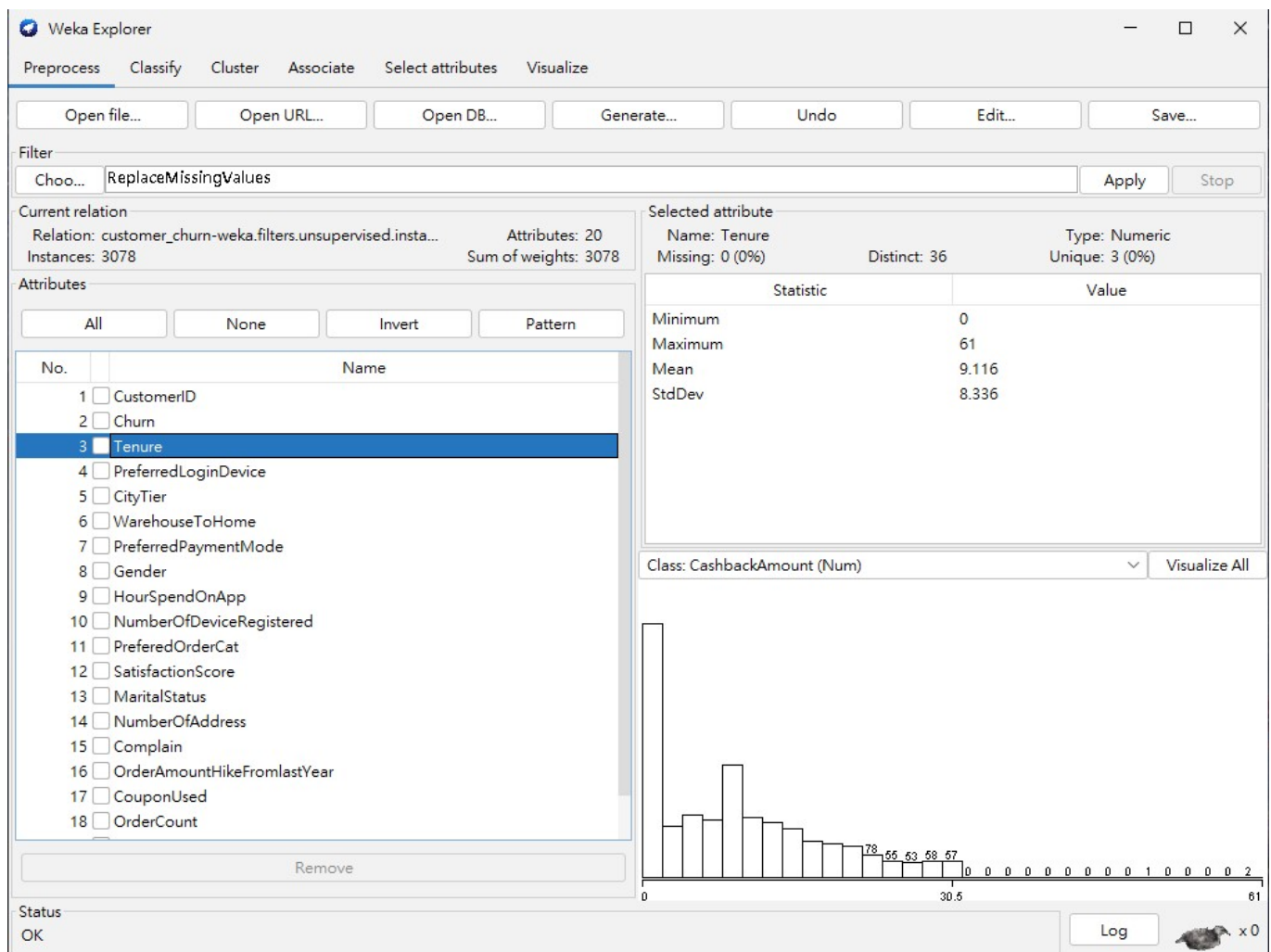
填補 Missing Value

nominal 以 mode 填補 missing value

numeric 以 mean 填補 missing value



以 Tenure 為例，missing value 都不見了，其餘屬性也是如此



接下來將 categorical 屬性值都轉成 numeric value，我選擇的是 filter 中的 NominalToBinary，在屬性值不是二元可分的情況，其實就跟 Python 的 `pandas.get_dummies` 很像，把屬性變得更多了。

Weka Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Open file...Open URL...Open DB...Generate...UndoEdit...Save...

FilterChooseNominalToBinary -R 2,4,7,8,11ApplyStop

Current relationRelation: customer_churn-weka.filters.unsupervised.attribute.NominalToBinaryInstances: 3078

AttributesAllNone

No.	
1	<input 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"/> OrderAmountHikeFromlastYear
17	<input type="checkbox"/> CouponUsed
18	<input type="checkbox"/> OrderCount

Remove

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.NominalToBinary

About

Converts all nominal attributes into binary numeric attributes.

MoreCapabilities

attributeIndices2,4,7,8,11

binaryAttributesNominalFalse

debugFalse

doNotCheckCapabilitiesFalse

invertSelectionFalse

spreadAttributeWeightFalse

transformAllValuesFalse

Open...Save...OKCancel

Type: NumericUnique: 3078 (100%)

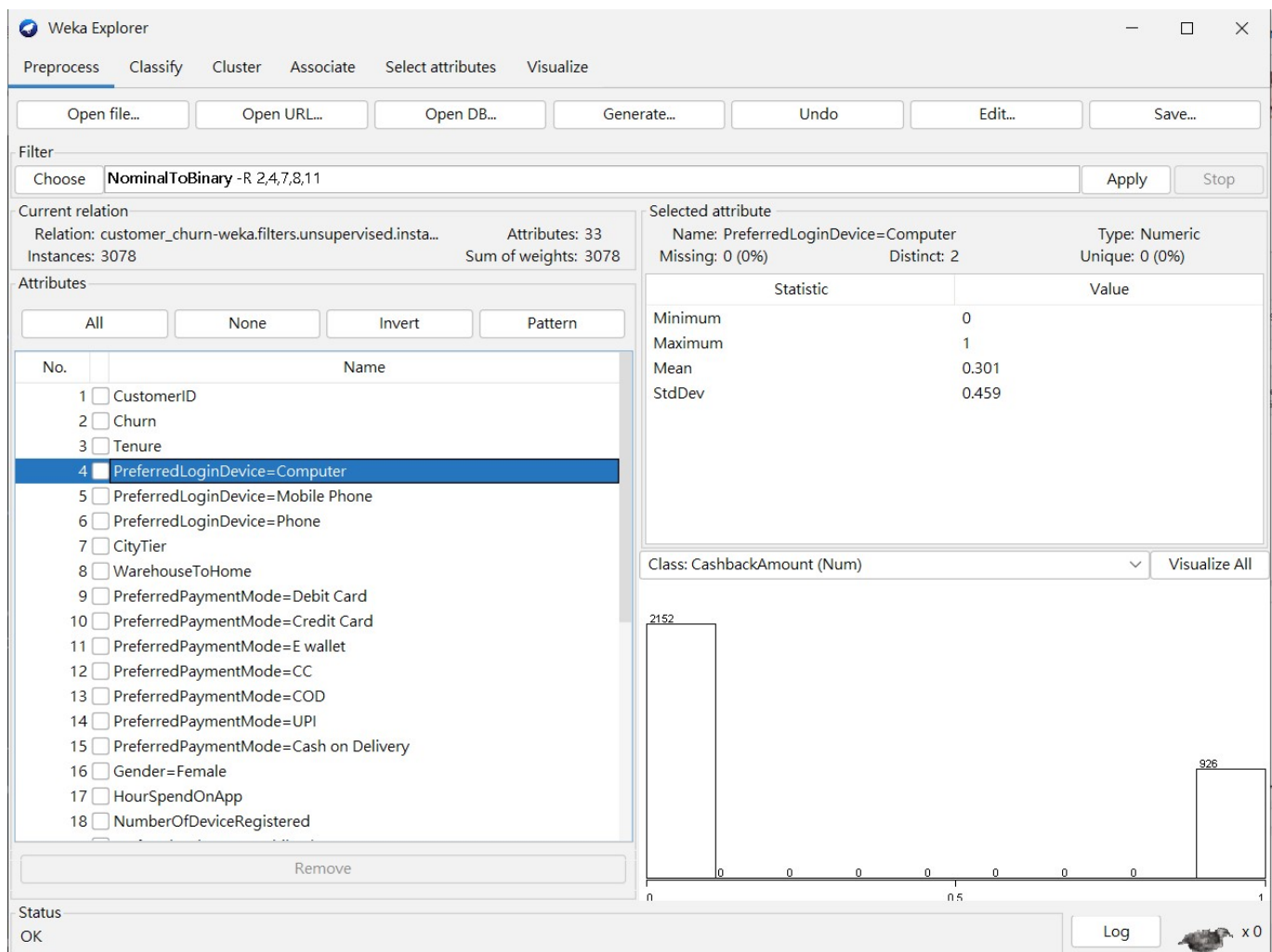
Value

500015562852811.6231614.867

Visualize All

Value	Frequency
50001	23
52814.5	197
55628	203
52811.623	224
1614.867	208
	208
	200
	192

StatusOKLogx 0



最後，把 churn 用 filter 中的 NumericToNominal 就可以套用 SVM、Logistic Regression、Decision Tree 來預測 churn 了。

Weka Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choose

NumericToNominal -R 2

Apply

Stop

Current relation

Relation: customer_churn-weka.filters.unsupervised.insta...

Instances: 3078

Attributes: 33

Sum of weights: 3078

Selected attribute

Name: PreferredLoginDevice=Computer

Missing: 0 (0%)

Distinct: 2

Type: Numeric

Unique: 0 (0%)

Attributes

All

None

No.	
1	<input type="checkbox"/> CustomerID
2	<input type="checkbox"/> Churn
3	<input type="checkbox"/> Tenure
4	<input checked="" type="checkbox"/> PreferredLoginDevice=Computer
5	<input type="checkbox"/> PreferredLoginDevice=Mobile
6	<input type="checkbox"/> PreferredLoginDevice=Phone
7	<input type="checkbox"/> CityTier
8	<input type="checkbox"/> WarehouseToHome
9	<input type="checkbox"/> PreferredPaymentMode=Debit
10	<input type="checkbox"/> PreferredPaymentMode=Credit
11	<input type="checkbox"/> PreferredPaymentMode=Electronic wallet
12	<input type="checkbox"/> PreferredPaymentMode=Credit Card
13	<input type="checkbox"/> PreferredPaymentMode=Cash on Delivery
14	<input type="checkbox"/> PreferredPaymentMode=UPI
15	<input type="checkbox"/> PreferredPaymentMode=Cash on Delivery
16	<input type="checkbox"/> Gender=Female
17	<input type="checkbox"/> HourSpendOnApp
18	<input type="checkbox"/> NumberOfDeviceRegistered

Remove

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.NumericToNominal

About

A filter for turning numeric attributes into nominal ones.

More

Capabilities

attributeIndices

2

debug

False

doNotCheckCapabilities

False

invertSelection

False

Open...

Save...

OK

Cancel

Value

0

1

0.301

0.459

Visualize All

Status

OK

Log

x 0

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

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

Filter: Choose **NumericToNominal -R 2** Apply Stop

Current relation
Relation: customer_churn-weka.filters.unsupervised insta... Attributes: 33
Instances: 3078 Sum of weights: 3078

Attributes

All None Invert Pattern

No.	Name
1	<input type="checkbox"/> CustomerID
2	<input checked="" type="checkbox"/> Churn
3	<input type="checkbox"/> Tenure
4	<input type="checkbox"/> PreferredLoginDevice=Computer
5	<input type="checkbox"/> PreferredLoginDevice=Mobile Phone
6	<input type="checkbox"/> PreferredLoginDevice=Phone
7	<input type="checkbox"/> CityTier
8	<input type="checkbox"/> WarehouseToHome
9	<input type="checkbox"/> PreferredPaymentMode=Debit Card
10	<input type="checkbox"/> PreferredPaymentMode=Credit Card
11	<input type="checkbox"/> PreferredPaymentMode=E wallet
12	<input type="checkbox"/> PreferredPaymentMode=CC
13	<input type="checkbox"/> PreferredPaymentMode=COD
14	<input type="checkbox"/> PreferredPaymentMode=UPI
15	<input type="checkbox"/> PreferredPaymentMode=Cash on Delivery
16	<input type="checkbox"/> Gender=Female
17	<input type="checkbox"/> HourSpendOnApp
18	<input type="checkbox"/> NumberOfDeviceRegistered

Remove

Status: OK

Selected attribute
Name: Churn
Missing: 0 (0%) Distinct: 2 Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	0	2130	2130
2	1	948	948

Class: CashbackAmount (Num) Visualize All

Log x 0

4.訓練、測試 SVM、Logistic Regression、Decision Tree 模型，請以 Accuracy 評估 模型表現

(模型皆為預設未調整，Training Data: 66%)

未用 **NominalToBinary** 前各模型準確率：

SVM：判斷正確為 82.9035%

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **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.functions.Logistic"**

Test options

☐ Use training set

☐ Supplied test set

☐ Cross-validation Folds

☒ Percentage split %

(Nom) Churn

Result list (right-click for options)

- 10:09:20 - trees.J48
- 10:09:47 - functions.Logistic
- 10:11:00 - functions.SMO
- 10:12:10 - functions.SMO
- 10:13:27 - functions.SMO
- 10:14:25 - functions.Logistic
- 10:14:45 - trees.J48
- 10:15:01 - trees.J48
- 10:15:31 - functions.Logistic
- 10:15:47 - functions.SMO
- 10:34:48 - functions.SMO
- 10:35:25 - functions.SMO
- 10:36:09 - functions.SMO
- 10:36:36 - functions.Logistic
- 10:37:01 - trees.J48
- 10:37:19 - trees.J48
- 10:37:36 - trees.J48
- 10:37:49 - trees.J48
- 10:45:34 - functions.SMO**

Classifier output

Time taken to build model: 2.44 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	868	82.9035 %
Incorrectly Classified Instances	179	17.0965 %
Kappa statistic	0.5809	
Mean absolute error	0.171	
Root mean squared error	0.4135	
Relative absolute error	40.2814 %	
Root relative squared error	90.4249 %	
Total Number of Instances	1047	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.895	0.328	0.866	0.895	0.880	0.582	0.784	0.849	0
	0.672	0.105	0.731	0.672	0.700	0.582	0.784	0.589	1
Weighted Avg.	0.829	0.262	0.826	0.829	0.827	0.582	0.784	0.772	


=== Confusion Matrix ===

```

a  b  <-- classified as
659  77 |  a = 0
102 209 |  b = 1

```

Status

OK  x 0

Logistic Regression : 判斷正確為 83.0946%

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **Logistic -R 1.0E-8 -M -1 -num-decimal-places 4**

Test options

☐ Use training set

☐ Supplied test set

☐ Cross-validation Folds

☒ Percentage split %

(Nom) Churn

Result list (right-click for options)

- 10:09:47 - functions.Logistic
- 10:11:00 - functions.SMO
- 10:12:10 - functions.SMO
- 10:13:27 - functions.SMO
- 10:14:25 - functions.Logistic
- 10:14:45 - trees.J48
- 10:15:01 - trees.J48
- 10:15:31 - functions.Logistic
- 10:15:47 - functions.SMO
- 10:34:48 - functions.SMO
- 10:35:25 - functions.SMO
- 10:36:09 - functions.SMO
- 10:36:36 - functions.Logistic
- 10:37:01 - trees.J48
- 10:37:19 - trees.J48
- 10:37:36 - trees.J48
- 10:37:49 - trees.J48
- 10:45:34 - functions.SMO
- 10:47:25 - functions.Logistic

Classifier output

Time taken to build model: 0.1 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	870	83.0946 %
Incorrectly Classified Instances	177	16.9054 %
Kappa statistic	0.5902	
Mean absolute error	0.2374	
Root mean squared error	0.3479	
Relative absolute error	55.927 %	
Root relative squared error	76.0817 %	
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	0.305	0.873	0.889	0.881	0.591	0.881	0.938	0
	0.695	0.111	0.725	0.695	0.709	0.591	0.881	0.784	1
Weighted Avg.	0.831	0.248	0.829	0.831	0.830	0.591	0.881	0.892	


=== Confusion Matrix ===

```

a  b  <-- classified as
654 82 | a = 0
95 216 | b = 1

```

Status

OK  x0

Decision Tree : 判斷正確為 87.5836%

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 v

Start Stop

Result list (right-click for options)

- 10:12:10 - functions.SMO
- 10:13:27 - functions.SMO
- 10:14:25 - functions.Logistic
- 10:14:45 - trees.J48
- 10:15:01 - trees.J48
- 10:15:31 - functions.Logistic
- 10:15:47 - functions.SMO
- 10:34:48 - functions.SMO
- 10:35:25 - functions.SMO
- 10:36:09 - functions.SMO
- 10:36:36 - functions.Logistic
- 10:37:01 - trees.J48
- 10:37:19 - trees.J48
- 10:37:36 - trees.J48
- 10:37:49 - trees.J48
- 10:45:34 - functions.SMO
- 10:47:25 - functions.Logistic
- 10:48:33 - trees.J48
- 10:48:41 - trees.J48

Classifier output

Time taken to build model: 0.05 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	917	87.5836 %
Incorrectly Classified Instances	130	12.4164 %
Kappa statistic	0.7054	
Mean absolute error	0.1529	
Root mean squared error	0.3274	
Relative absolute error	36.0176 %	
Root relative squared error	71.5958 %	
Total Number of Instances	1047	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.905	0.193	0.917	0.905	0.911	0.706	0.887	0.929	0
	0.807	0.095	0.782	0.807	0.794	0.706	0.887	0.745	1
Weighted Avg.	0.876	0.164	0.877	0.876	0.876	0.706	0.887	0.874	

=== Confusion Matrix ===

```

a  b  <-- classified as
666  70 |  a = 0
 60 251 |  b = 1

```

Status: OK

Log x 0

用 **NominalToBinary** 後各模型準確率：

SVM：判斷正確為 82.9035%

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **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.functions.Logistic"

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)

- 10:13:27 - functions.SMO
- 10:14:25 - functions.Logistic
- 10:14:45 - trees.J48
- 10:15:01 - trees.J48
- 10:15:31 - functions.Logistic
- 10:15:47 - functions.SMO
- 10:34:48 - functions.SMO
- 10:35:25 - functions.SMO
- 10:36:09 - functions.SMO
- 10:36:36 - functions.Logistic
- 10:37:01 - trees.J48
- 10:37:19 - trees.J48
- 10:37:36 - trees.J48
- 10:37:49 - trees.J48
- 10:45:34 - functions.SMO
- 10:47:25 - functions.Logistic
- 10:48:33 - trees.J48
- 10:48:41 - trees.J48
- 10:57:51 - functions.SMO**

Classifier output

Time taken to build model: 1.81 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	868	82.9035 %
Incorrectly Classified Instances	179	17.0965 %
Kappa statistic	0.5809	
Mean absolute error	0.171	
Root mean squared error	0.4135	
Relative absolute error	40.2814 %	
Root relative squared error	90.4249 %	
Total Number of Instances	1047	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.895	0.328	0.866	0.895	0.880	0.582	0.784	0.849	0
	0.672	0.105	0.731	0.672	0.700	0.582	0.784	0.589	1
Weighted Avg.	0.829	0.262	0.826	0.829	0.827	0.582	0.784	0.772	

=== Confusion Matrix ===

```

a  b  <-- classified as
659 77 | a = 0
102 209 | b = 1

```

Status

OK Log x 0

Logistic Regression : 判斷正確為 83.0946%

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **Logistic -R 1.0E-8 -M -1 -num-decimal-places 4**

Test options

☐ Use training set

☐ Supplied test set

☐ Cross-validation Folds 10

☒ Percentage split % 66

(Nom) Churn

Result list (right-click for options)

- 10:14:25 - functions.Logistic
- 10:14:45 - trees.J48
- 10:15:01 - trees.J48
- 10:15:31 - functions.Logistic
- 10:15:47 - functions.SMO
- 10:34:48 - functions.SMO
- 10:35:25 - functions.SMO
- 10:36:09 - functions.SMO
- 10:36:36 - functions.Logistic
- 10:37:01 - trees.J48
- 10:37:19 - trees.J48
- 10:37:36 - trees.J48
- 10:37:49 - trees.J48
- 10:45:34 - functions.SMO
- 10:47:25 - functions.Logistic
- 10:48:33 - trees.J48
- 10:48:41 - trees.J48
- 10:57:51 - functions.SMO
- 10:58:43 - functions.Logistic**

Classifier output

Time taken to build model: 0.09 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	870	83.0946 %
Incorrectly Classified Instances	177	16.9054 %
Kappa statistic	0.5902	
Mean absolute error	0.2374	
Root mean squared error	0.3479	
Relative absolute error	55.927 %	
Root relative squared error	76.0817 %	
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	0.305	0.873	0.889	0.881	0.591	0.881	0.938	0
	0.695	0.111	0.725	0.695	0.709	0.591	0.881	0.784	1
Weighted Avg.	0.831	0.248	0.829	0.831	0.830	0.591	0.881	0.892	


=== Confusion Matrix ===

```

a  b  <-- classified as
654 82 | a = 0
95 216 | b = 1

```

Status

OK 

Decision Tree : 判斷正確為 90.4489%

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

☐ Cross-validation Folds

☒ Percentage split %

(Nom) Churn

Result list (right-click for options)

- 10:14:45 - trees.J48
- 10:15:01 - trees.J48
- 10:15:31 - functions.Logistic
- 10:15:47 - functions.SMO
- 10:34:48 - functions.SMO
- 10:35:25 - functions.SMO
- 10:36:09 - functions.SMO
- 10:36:36 - functions.Logistic
- 10:37:01 - trees.J48
- 10:37:19 - trees.J48
- 10:37:36 - trees.J48
- 10:37:49 - trees.J48
- 10:45:34 - functions.SMO
- 10:47:25 - functions.Logistic
- 10:48:33 - trees.J48
- 10:48:41 - trees.J48
- 10:57:51 - functions.SMO
- 10:58:43 - functions.Logistic
- 10:59:25 - trees.J48**

Classifier output

Time taken to build model: 0.12 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	947	90.4489 %
Incorrectly Classified Instances	100	9.5511 %
Kappa statistic	0.7791	
Mean absolute error	0.1122	
Root mean squared error	0.2843	
Relative absolute error	26.4338 %	
Root relative squared error	62.1756 %	
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	0.100	0.956	0.906	0.930	0.782	0.945	0.973	0
	0.900	0.094	0.802	0.900	0.848	0.782	0.945	0.827	1
Weighted Avg.	0.904	0.098	0.910	0.904	0.906	0.782	0.945	0.929	

=== Confusion Matrix ===

```

a  b  <-- classified as
667 69 | a = 0
31 280 | b = 1

```

Status

OK x 0

結論：用 NominalToBinary 的前後，SVM 與 Logistic Regression 的準確率沒有影響，但 Decision Tree 的準確率有微幅上升。