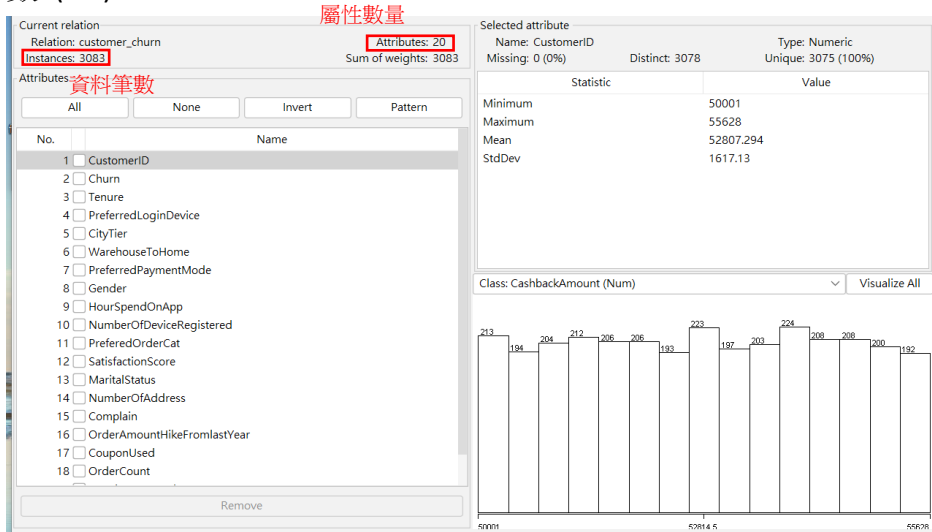


1. 載入 customer_churn.csv，列出資料筆數、屬性數量以及每個欄位的空值個數 (5%)



以下列出含空值的欄位：

Selected attribute 空值個數		Selected attribute	
Name: DaySinceLastOrder	Type: Numeric	Name: HourSpendOnApp	Type: Numeric
Missing: 166 (5%)	Distinct: 20	Missing: 150 (5%)	Distinct: 6
Unique: 1 (0%)		Unique: 1 (0%)	
Statistic	Value	Statistic	Value
Minimum	0	Minimum	0
Maximum	46	Maximum	5
Mean	4.349	Mean	2.946
StdDev	3.609	StdDev	0.719

Selected attribute		Selected attribute	
Name: OrderAmountHikeFromlastYear	Type: Numeric	Name: CouponUsed	Type: Numeric
Missing: 131 (4%)	Distinct: 16	Missing: 126 (4%)	Distinct: 17
Unique: 0 (0%)		Unique: 3 (0%)	
Statistic	Value	Statistic	Value
Minimum	11	Minimum	0
Maximum	26	Maximum	16
Mean	15.714	Mean	1.753
StdDev	3.765	StdDev	1.886

Selected attribute		Selected attribute	
Name: Tenure	Type: Numeric	Name: WarehouseToHome	Type: Numeric
Missing: 153 (5%)	Distinct: 35	Missing: 154 (5%)	Distinct: 33
Unique: 3 (0%)		Unique: 1 (0%)	
Statistic	Value	Statistic	Value
Minimum	0	Minimum	5
Maximum	61	Maximum	126
Mean	9.109	Mean	15.771
StdDev	8.549	StdDev	8.557

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

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%)

Filter

Choose RemoveDuplicates

Current relation

Relation: customer_churn-weka.filters.unsupervised insta... Attributes: 20

Instances: 3078 Sum of weights: 3078

3. 資料前處理 (5%)

3.1 刪除 CustomerID

Viewer

Relation: customer_churn-weka.filters.unsupervised.instance.RemoveDuplicates

No. 1: CustomerID 2: Churn 3: Tenure 4: PreferredLoginDevice 5: CityTier 6: WarehouseToHome 7: PreferredPaymentMode 8: Gender 9: HourSpendOnApp 10: ...

1 50001.0 1.0 4.0 Mobile Phone 3.0 6.0 Debit Card Female 3.0

2 50002.0 1.0 Phone 1.0 8.0 UPI Male 3.0

3 50003.0 1.0 Phone 1.0 30.0 Debit Card Male 2.0

4 50004.0 1.0 0.0 Phone 3.0 15.0 Debit Card Male 2.0

5 50005.0 1.0 0.0 Phone 1.0 12.0 CC Male 3.0

6 50006.0 1.0 0.0 Computer 1.0 22.0 Debit Card Female 3.0

7 50007.0 1.0 Phone 3.0 11.0 Cash on Delivery Male 2.0

8 50008.0 1.0 Phone 1.0 6.0 CC Male 3.0

9 50009.0 1.0 13.0 Phone Male 2.0

10 50010.0 1.0 Phone Male 2.0

11 50011.0 1.0 4.0 Mobile Ph Female 2.0

12 50012.0 1.0 11.0 Mobile Ph Male 3.0

13 50013.0 1.0 0.0 Phone Male 2.0

14 50014.0 1.0 0.0 Phone Male 3.0

15 50015.0 1.0 9.0 Mobile Phone 3.0 15.0 Credit Card Male 3.0

16 50016.0 1.0 Phone 2.0 12.0 UPI Male 3.0

17 50017.0 1.0 0.0 Computer 1.0 12.0 Debit Card Female 3.0

18 50018.0 1.0 0.0 Mobile Phone 3.0 11.0 E wallet Male 2.0

19 50019.0 1.0 0.0 Computer 1.0 13.0 Debit Card Male 3.0

20 50020.0 1.0 19.0 Mobile Phone 1.0 20.0 Debit Card Female 3.0

21 50021.0 1.0 0.0 Mobile Phone 3.0 12.0 Debit Card Male 3.0

22 50022.0 1.0 20.0 Mobile Phone 1.0 29.0 Credit Card Female 3.0

23 50023.0 1.0 Mobile Phone 3.0 28.0 E wallet Male 2.0

Confirm... Do you really want to delete the attribute 'CustomerID'? 是(Y) 否(N)

Add Instance Undo OK Cancel

3.2 清理空值

Filter

Choose ReplaceMissingValues

Current relation

Relation: customer_churn-weka.filters.unsupervised insta... Attributes: 19

Instances: 3078 Sum of weights: 3078

Selected attribute

Name: OrderCount

Missing: 128 (4%) Distinct: 16 Type: Unique

Statistic

Minimum 1

Maximum 16

Mean 2.996

StdDev 2.95

Class: CashbackAmount (Num)

950

Attributes

All None Invert Pattern

No. Name

1 ☐ Churn

2 ☒ Tenure

3 ☐ PreferredLoginDevice

4 ☐ CityTier

5 ☒ WarehouseToHome

6 ☐ PreferredPaymentMode

7 ☐ Gender

8 ☒ HourSpendOnApp

9 ☐ NumberOfDeviceRegistered

10 ☐ PreferredOrderCat

11 ☐ SatisfactionScore

12 ☐ MaritalStatus

13 ☐ NumberOfAddress

14 ☐ Complain

15 ☒ OrderAmountHikeFromlastYear

16 ☒ CouponUsed

17 ☒ OrderCount

18 ☒ DaySinceLastOrder

3.3 feature encoding

Filter

Choose NominalToBinary-R first-last

Current relation

Relation: customer_churn-weka.filters.unsupervised insta... Attributes: 19

Instances: 3078 Sum of weights: 3078

Selected attribute

Name: Churn

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

Statistic

Minimum 0

Maximum 1

Mean 0.308

StdDev 0.462

Class: CashbackAmount (Num)

Visualize All

2100

Attributes

All None Invert Pattern

No. Name

1 ☒ Churn

2 ☐ Tenure

3 ☐ PreferredLoginDevice

4 ☐ CityTier

5 ☐ WarehouseToHome

6 ☐ PreferredPaymentMode

7 ☐ Gender

8 ☐ HourSpendOnApp

9 ☐ NumberOfDeviceRegistered

10 ☐ PreferredOrderCat

11 ☐ SatisfactionScore

12 ☐ MaritalStatus

13 ☐ NumberOfAddress

14 ☐ Complain

15 ☐ OrderAmountHikeFromlastYear

16 ☐ CouponUsed

17 ☐ OrderCount

18 ☐ DaySinceLastOrder

Remove

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

4.1 SVM

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 2.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"]

Test options

☐ Use training set

☐ Supplied test set Set...

☐ Cross-validation Folds 10

☒ Percentage split % 70

More options...

(Nom) Churn

Start Stop

Result list (right-click for options)

- 14:38:10 - trees.J48
- 14:40:50 - functions.Logistic
- 14:42:32 - functions.SMO

Classifier output

Time taken to build model: 15.94 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.03 seconds

=== Summary ===

Correctly Classified Instances	785	85.0489 %
Incorrectly Classified Instances	138	14.9512 %
Kappa statistic	0.6339	
Mean absolute error	0.1495	
Root mean squared error	0.3867	
Relative absolute error	35.4355 %	
Root relative squared error	85.3572 %	
Total Number of Instances	923	

Accuracy

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.895	0.261	0.895	0.895	0.895	0.634	0.817	0.876	0
	0.739	0.105	0.739	0.739	0.739	0.634	0.817	0.620	1

=== Confusion Matrix ===

a	b	classified as	
590	69	a = 0	
69	195	b = 1	

4.2 Logistic Regression

Classifier

Choose [Logistic -R 1.0E-8 -M -1 -num-decimal-places 4]

Test options

☐ Use training set

☐ Supplied test set Set...

☐ Cross-validation Folds 10

☒ Percentage split % 70

More options...

(Nom) Churn

Start Stop

Result list (right-click for options)

- 14:38:10 - trees.J48
- 14:40:22 - functions.SMO
- 14:40:50 - functions.Logistic

Classifier output

Time taken to build model: 0.1 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	781	84.6154 %
Incorrectly Classified Instances	142	15.3846 %
Kappa statistic	0.6317	
Mean absolute error	0.2326	
Root mean squared error	0.3373	
Relative absolute error	55.1196 %	
Root relative squared error	74.4539 %	
Total Number of Instances	923	

Accuracy

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.877	0.231	0.905	0.877	0.891	0.633	0.890	0.946	0
	0.769	0.123	0.715	0.769	0.741	0.633	0.890	0.808	1

=== Confusion Matrix ===

a	b	classified as	
578	81	a = 0	
61	203	b = 1	

4.3 Decision Tree

Classifier

Choose [J48 -C 0.25 -M 2]

Test options

☐ Use training set

☐ Supplied test set Set...

☐ Cross-validation Folds 10

☒ Percentage split % 70

More options...

(Nom) Churn

Start Stop

Result list (right-click for options)

- 14:38:10 - trees.J48
- 14:40:50 - functions.Logistic
- 14:42:32 - functions.SMO

Classifier output

Time taken to build model: 0.18 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	824	89.2741 %
Incorrectly Classified Instances	99	10.7259 %
Kappa statistic	0.7429	
Mean absolute error	0.1283	
Root mean squared error	0.3061	
Relative absolute error	30.4156 %	
Root relative squared error	67.5637 %	
Total Number of Instances	923	

Accuracy

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.910	0.152	0.938	0.910	0.924	0.744	0.927	0.962	0
	0.848	0.090	0.792	0.848	0.819	0.744	0.927	0.783	1

=== Confusion Matrix ===

a	b	classified as	
600	59	a = 0	
40	224	b = 1	