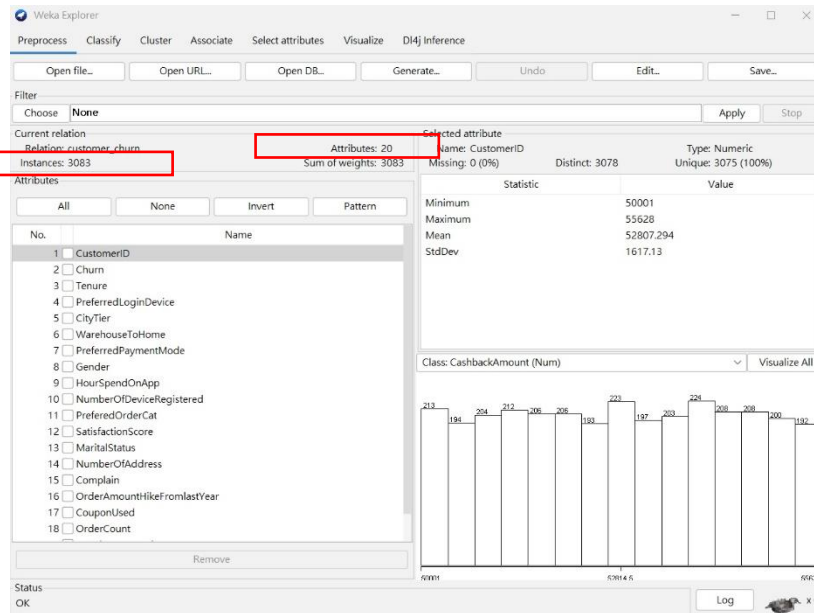


電子商務技術_HW3

108403201 資管三 A 黃名揚

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



(1)資料筆數及屬性個數

Selected attribute		
Name: CustomerID	Type: Numeric	
Missing: 0 (0%)	Distinct: 3078	Unique: 3075 (100%)

CustomerID 空值:0

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

Churn 空值:0

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

Tenure 空值:153

Selected attribute		
Name: CityTier	Type: Numeric	
Missing: 0 (0%)	Distinct: 3	Unique: 0 (0%)

CityTier 空值:0

Selected attribute		
Name: PreferredLoginDevice	Type: Nominal	
Missing: 0 (0%)	Distinct: 3	Unique: 0 (0%)

PreferredLoginDevice 空值:0

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

HourSpendOnApp 空值:150

Selected attribute		
Name: PreferredPaymentMode	Type: Nominal	
Missing: 0 (0%)	Distinct: 7	Unique: 0 (0%)

PreferredPaymentMode 空值:0

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

WareHouseToHome 空值:154

Selected attribute		
Name: NumberOfDeviceRegistered	Type: Numeric	
Missing: 0 (0%)	Distinct: 6	Unique: 0 (0%)

NumberOfDeviceRegistered 空值:0

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

Gender 空值:0

Selected attribute		
Name: SatisfactionScore	Type: Numeric	
Missing: 0 (0%)	Distinct: 5	Unique: 0 (0%)

SatisfactionScore 空值:0

Selected attribute		
Name: PreferredOrderCat	Type: Nominal	
Missing: 0 (0%)	Distinct: 6	Unique: 0 (0%)

PreferredOrderCat 空值:0

Selected attribute		
Name: MaritalStatus	Type: Nominal	
Missing: 0 (0%)	Distinct: 3	Unique: 0 (0%)

MaritalStatue 空值:0

Selected attribute		
Name: Complain	Type: Numeric	
Missing: 0 (0%)	Distinct: 2	Unique: 0 (0%)

Complain 空值:0

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

CouponUsed 空值:126

Selected attribute		
Name: CashbackAmount	Type: Numeric	
Missing: 0 (0%)	Distinct: 212	Unique: 20 (1%)

CashbackAmount 空值:0

Selected attribute		
Name: NumberOfAddress	Type: Numeric	
Missing: 0 (0%)	Distinct: 14	Unique: 3 (0%)

NumberOfAddress 空值:0

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

OrderAmountHikeFromlastYear

空值:131

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

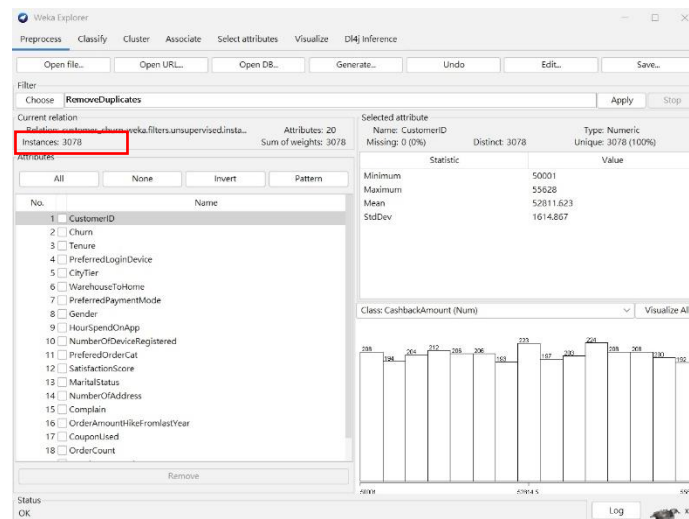
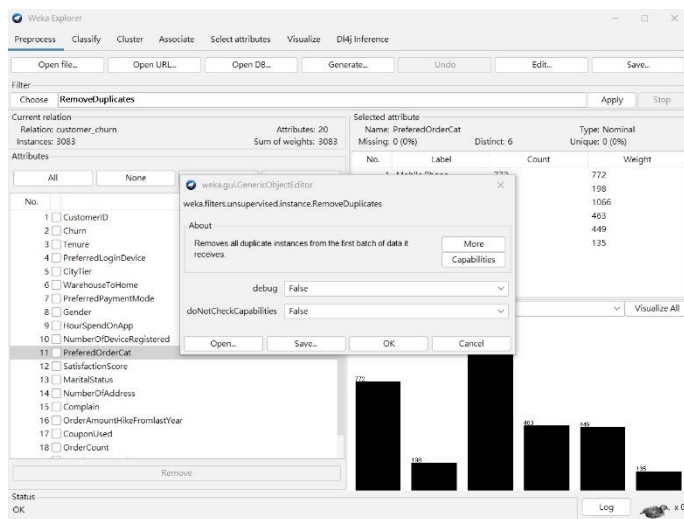
OrderCount 空值:128

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

DaySinceLastOrder 空值:166

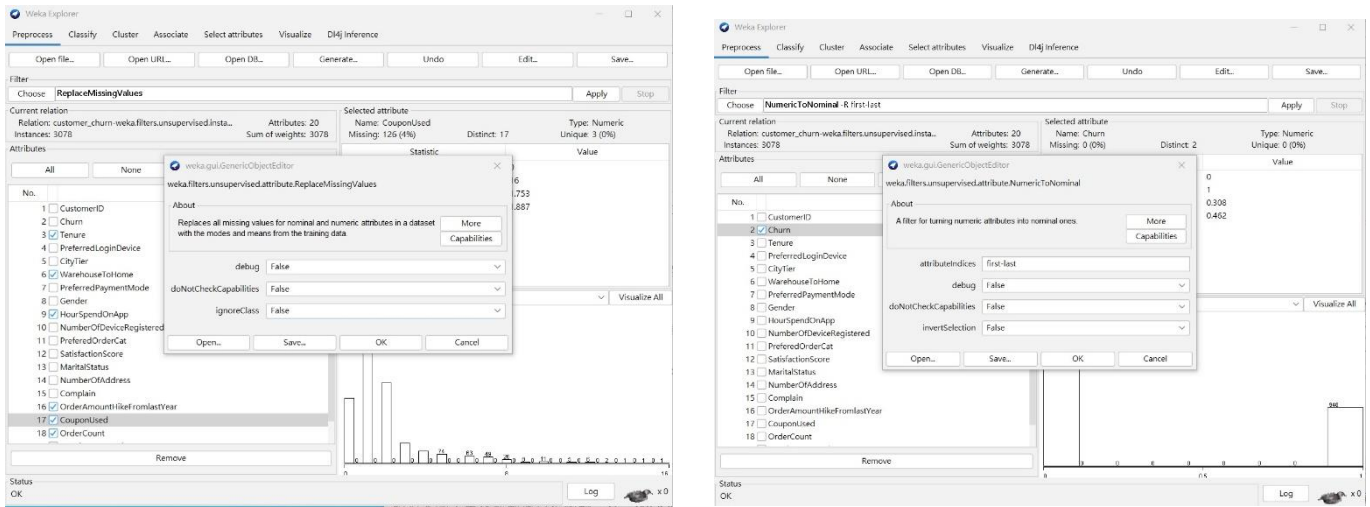
(2)每個欄位的空值個數

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



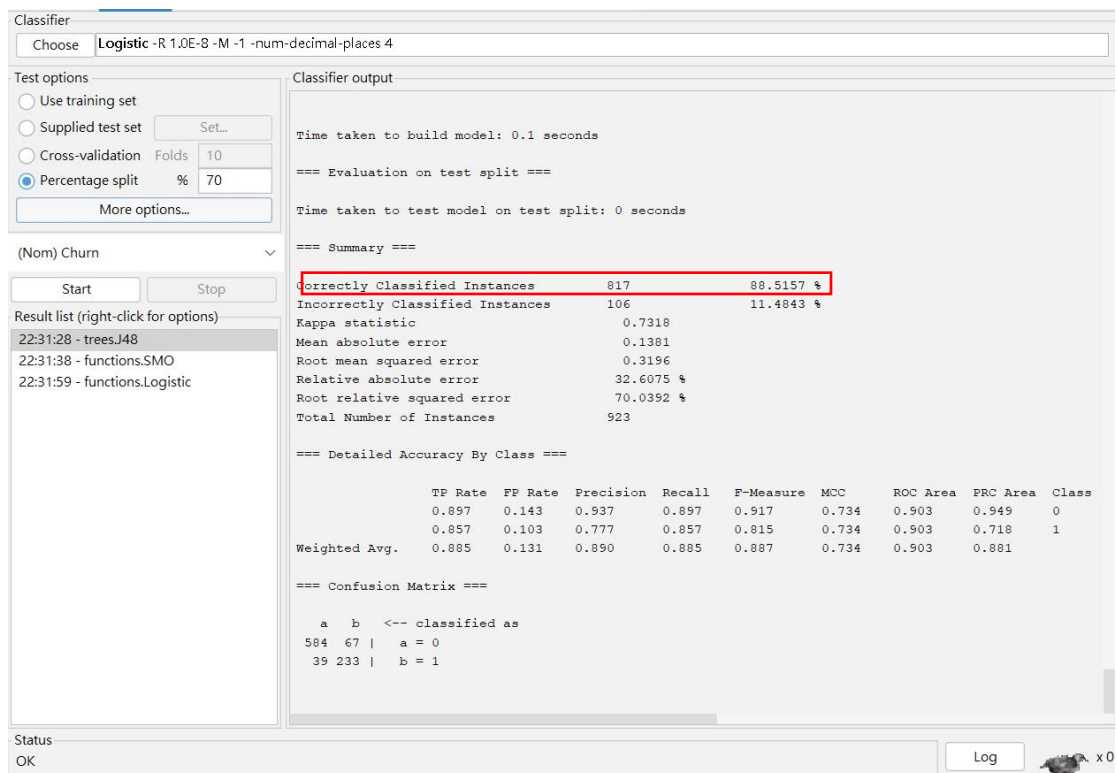
(3)用 RemoveDuplicate 刪除重複資料，剩餘 3078 筆資料

3. 資料前處理



(3)用 ReplaceMissingValues 處理空值，並將用 NumericToNominal 將 churn 轉為 nominal

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



(4)DecisionTree 模型，Accuracy:88.5157

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)

22:31:28 - trees.J48

22:31:38 - functions.SMO

22:31:59 - functions.Logistic

Classifier output

Time taken to build model: 0.35 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.02 seconds

=== Summary ===

Correctly Classified Instances	765	82.8819 %
Incorrectly Classified Instances	158	17.1181 %
Kappa statistic	0.5828	
Mean absolute error	0.2389	
Root mean squared error	0.3478	
Relative absolute error	56.3988 %	
Root relative squared error	76.215 %	
Total Number of Instances	923	


=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.888	0.313	0.872	0.888	0.880	0.583	0.881	0.938	0
	0.688	0.112	0.719	0.688	0.703	0.583	0.881	0.783	1
Weighted Avg.	0.829	0.253	0.827	0.829	0.828	0.583	0.881	0.893	

=== Confusion Matrix ===

a b <-- classified as

578	73	a = 0
85	187	b = 1

Status OK 

(5) Logistic Regression 模型，Accuracy:82.8819

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)

22:31:28 - trees.J48

22:31:38 - functions.SMO

22:31:59 - functions.Logistic

Classifier output

Time taken to build model: 3.21 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	758	82.1235 %
Incorrectly Classified Instances	165	17.8765 %
Kappa statistic	0.5591	
Mean absolute error	0.1788	
Root mean squared error	0.4228	
Relative absolute error	42.2045 %	
Root relative squared error	92.6583 %	
Total Number of Instances	923	


=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.891	0.346	0.861	0.891	0.875	0.560	0.773	0.844	0
	0.654	0.109	0.715	0.654	0.683	0.560	0.773	0.570	1
Weighted Avg.	0.821	0.276	0.818	0.821	0.819	0.560	0.773	0.763	

=== Confusion Matrix ===

a b <-- classified as

580	71	a = 0
94	178	b = 1

Status OK 

(6) SVM 模型，Accuracy:82.1235