

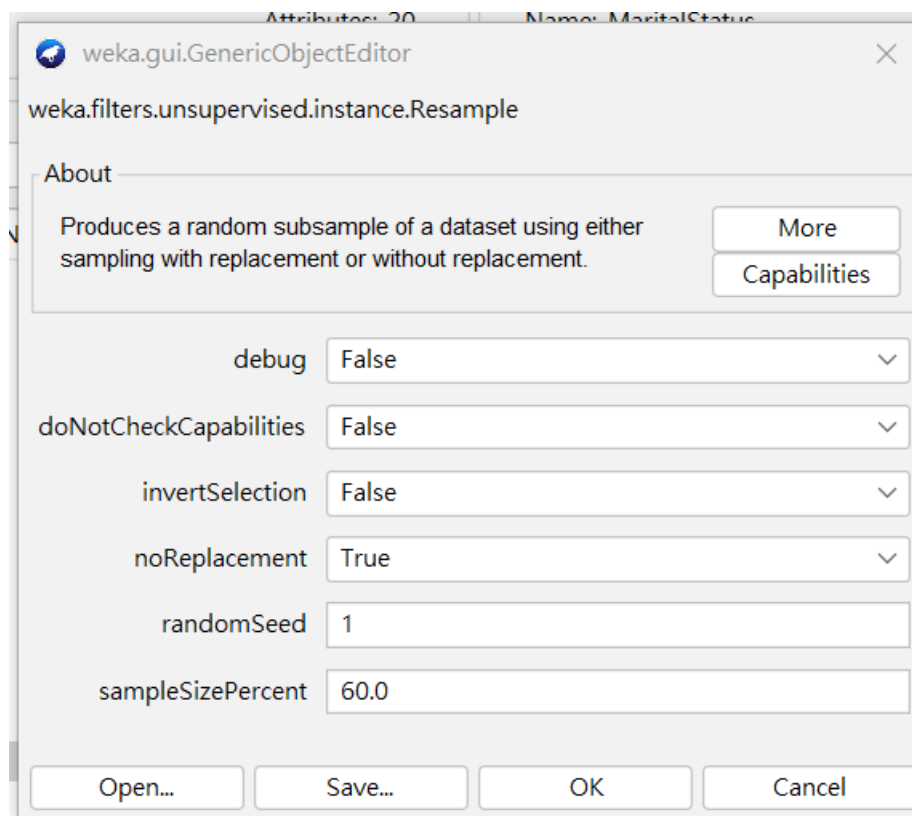
電子商務技術_HW5

108403201 資管三 A 黃名揚

1. 使用 Stratified sampling 從原本的資料集中取 60%的資料

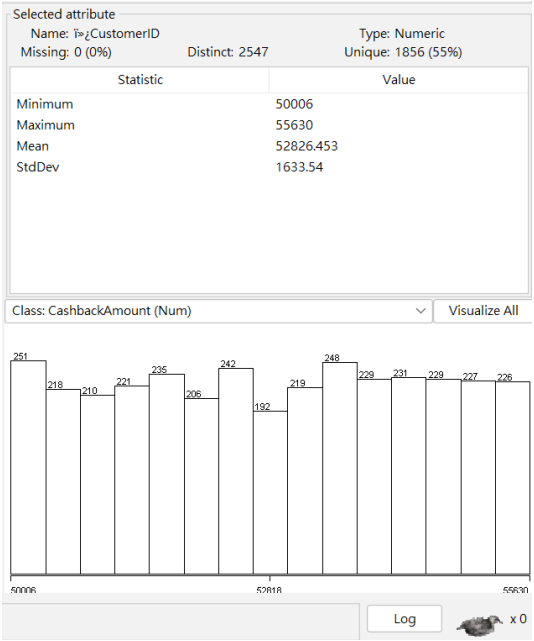
Current relation	
Relation: customer_churn	Attributes: 20
Instances: 5640	Sum of weights: 5640

初始資料量

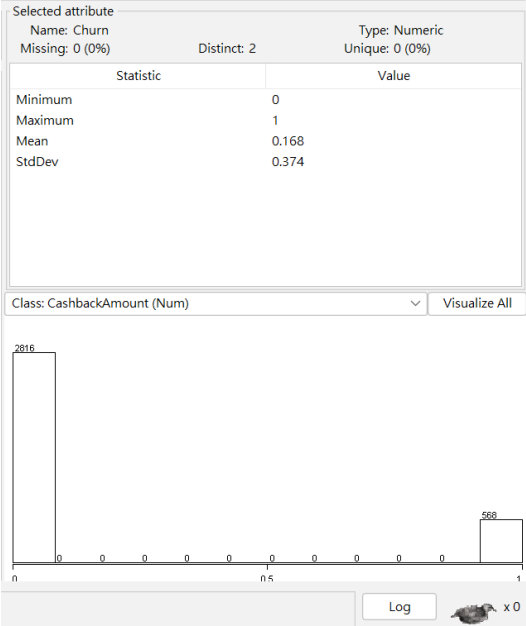


使用 Resample 取 60%的資料

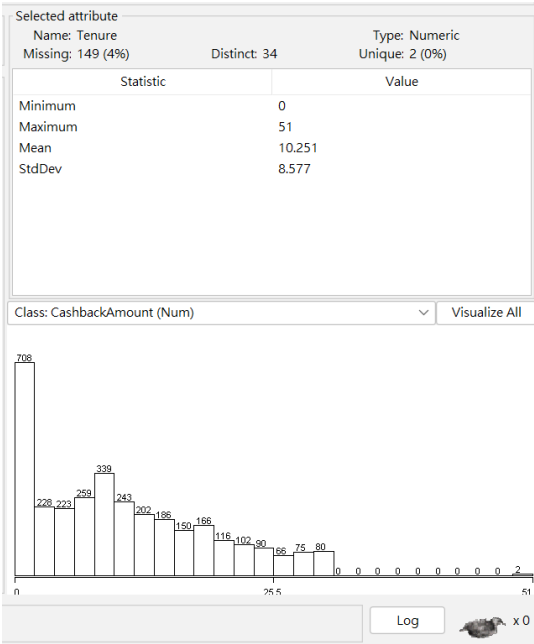
2. 顯示取樣後各類別的資料數量



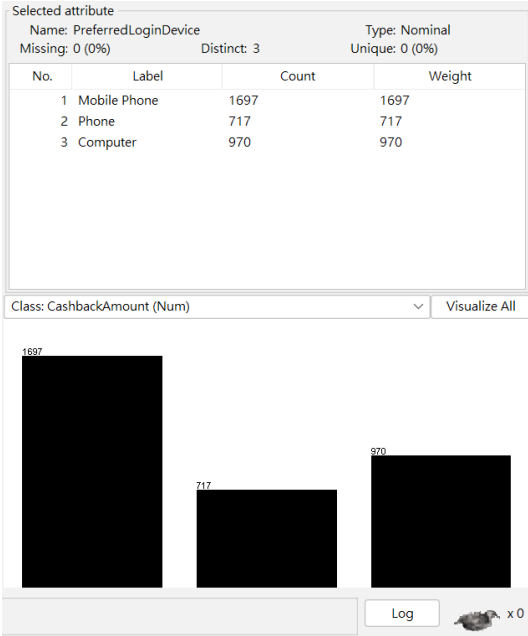
CustomerID



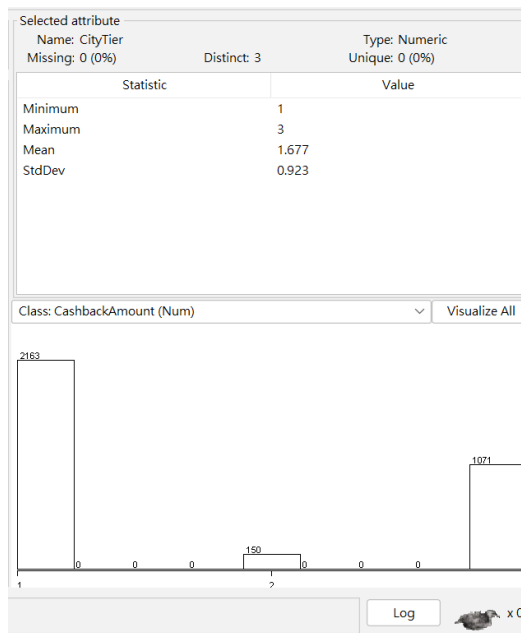
Churn



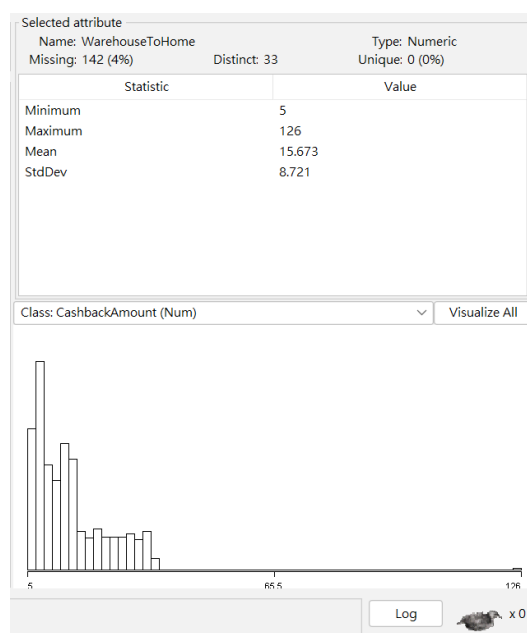
Tenure



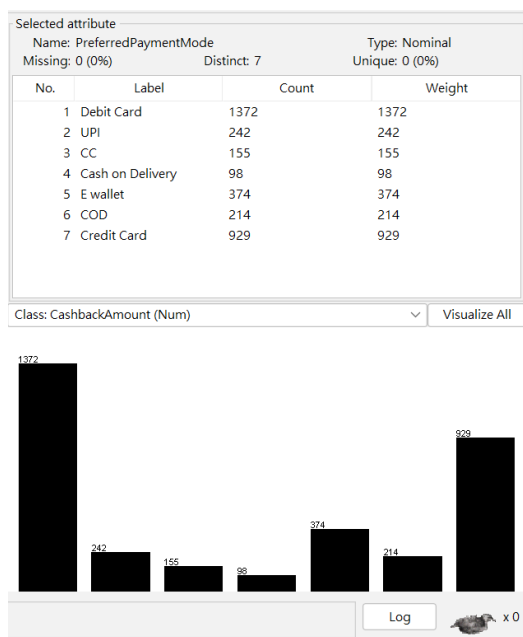
PreferredLoginDevice



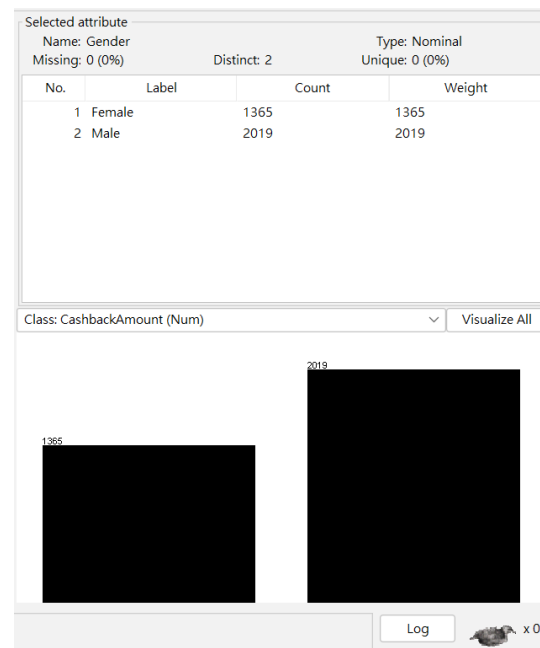
CityTier



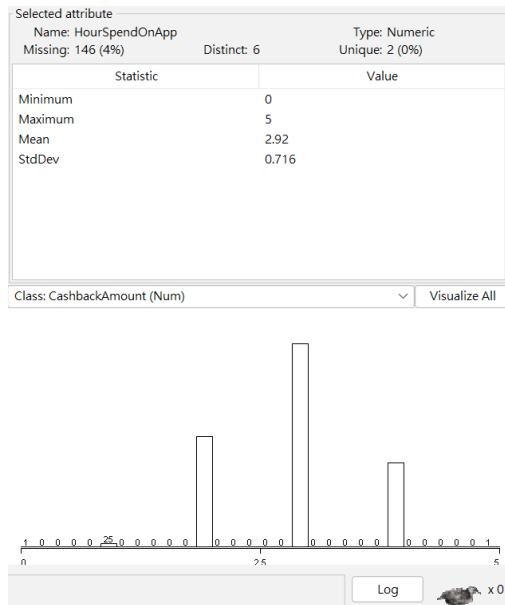
WarehouseToHome



PreferredPaymentMode



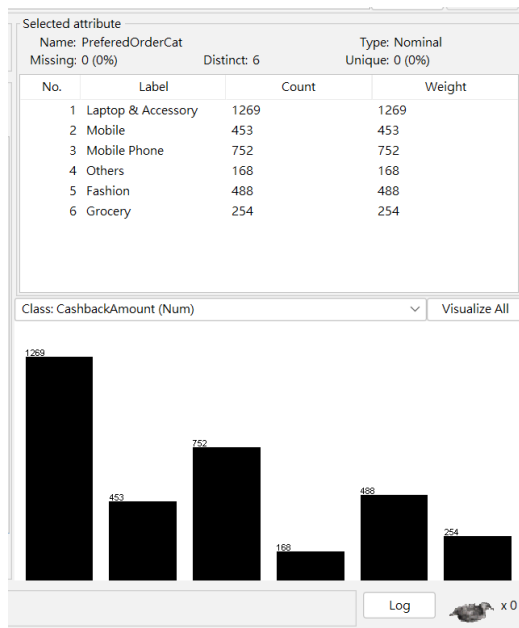
Gender



HourSpendOnApp



NumberOfDeviceRegistered



PreferredOrderCat



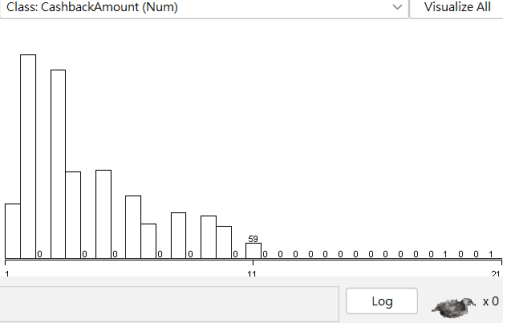
SatisfactionScore

Selected attribute				
Name: MaritalStatus		Type: Nominal		
Missing: 0 (0%)		Unique: 0 (0%)		
Distinct: 3				
No.	Label	Count	Weight	
1	Single	1106	1106	
2	Divorced	528	528	
3	Married	1750	1750	

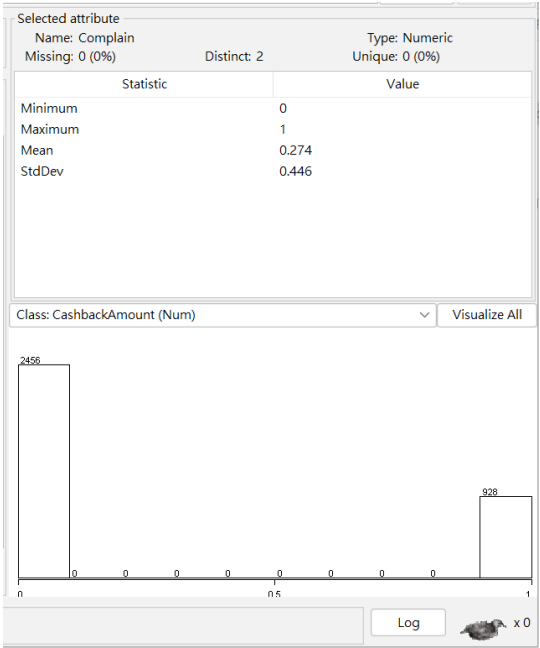


MaritalStatus

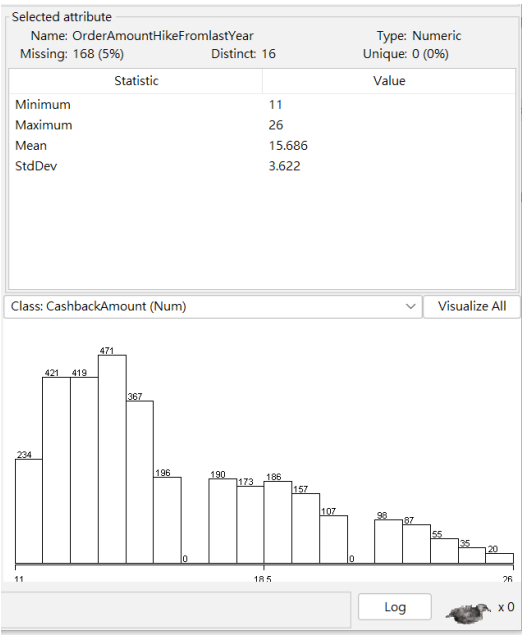
Selected attribute		
Name: NumberOfAddress		Type: Numeric
Missing: 0 (0%)		Unique: 2 (0%)
Distinct: 13		
Statistic	Value	
Minimum	1	
Maximum	21	
Mean	4.29	
StdDev	2.617	



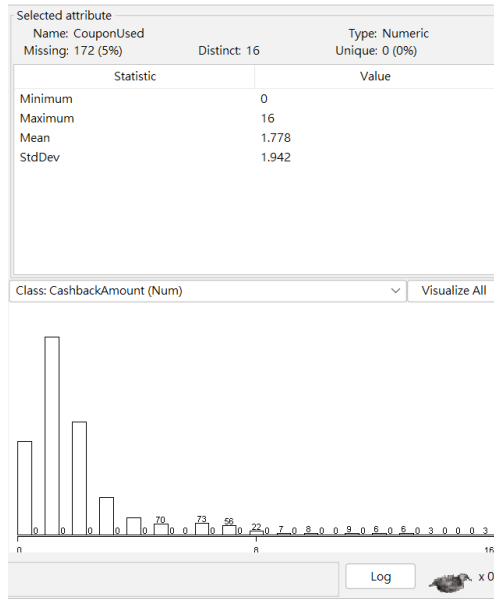
NumberOfAddress



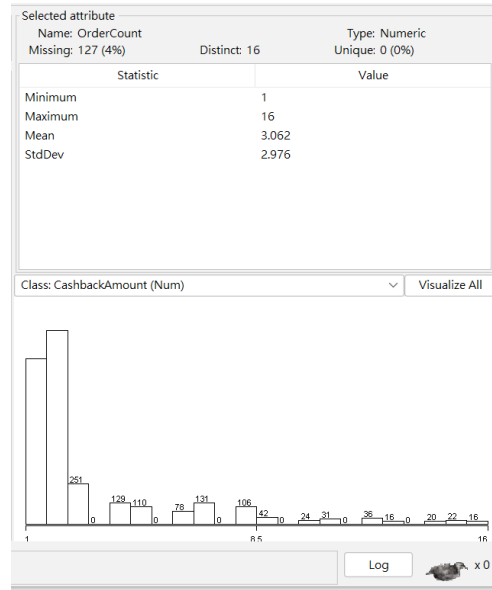
Complain



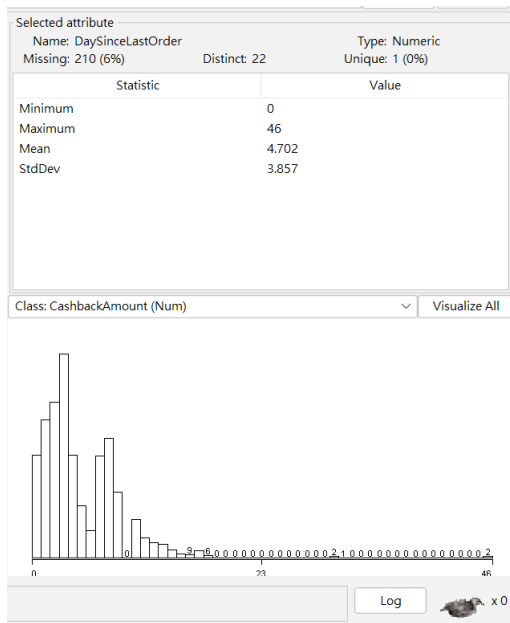
OrderAmountHikeFromlastYear



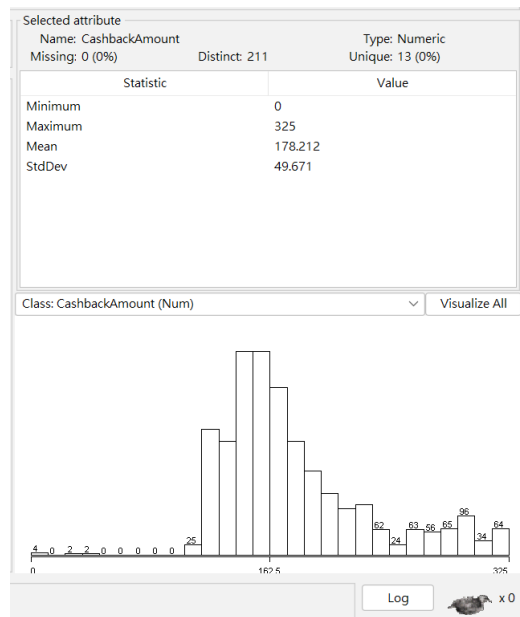
CouponUsed



OrderCount

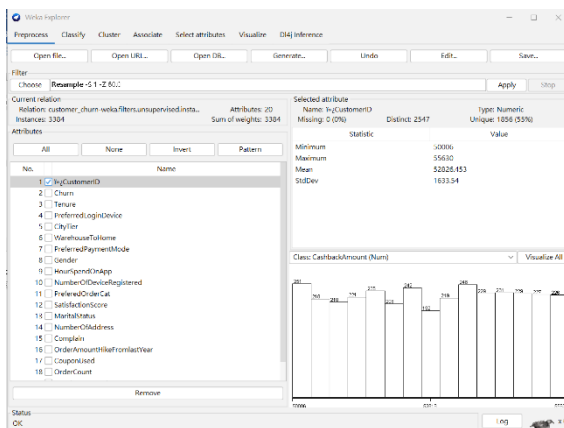


DaySinceOrder

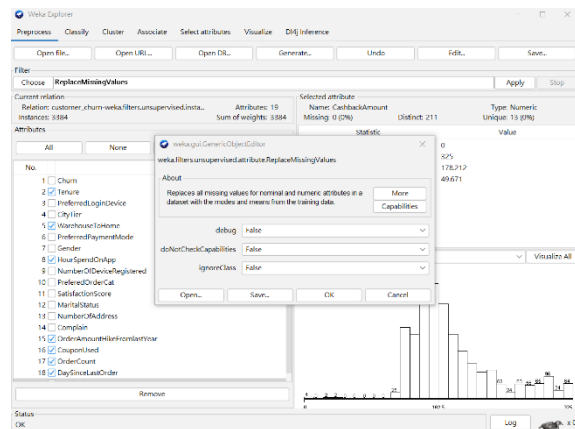


CashbackAmount

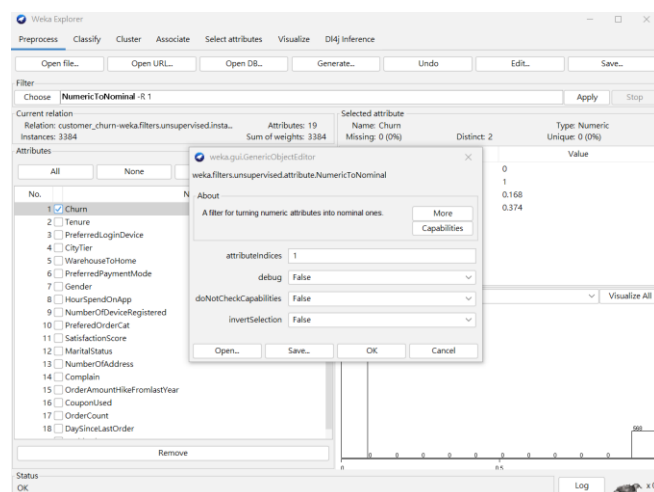
3. 資料前處理，並以 repeated 10 folds cross-validation (重複 10 次) Paired t-test 比較 Logistic Regression 及 SVM 模型



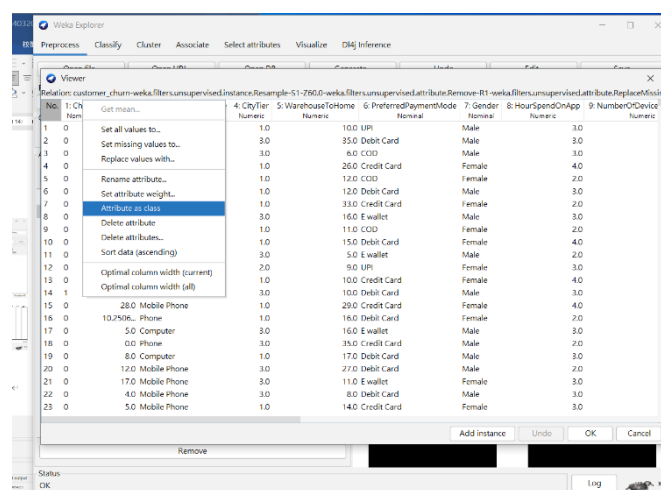
刪除 CustomerID



填補空值



將 Churn 轉為 Nominal



將 Churn 設為 class

4. 根據 weka 的輸出說明結論

Weka Experiment Environment

Setup Run **Analyse**

Source

Got 200 results File... Database... Experiment

Actions

Perform test Save output Open Explorer...

Configure test

Testing with: Paired T-Tester (corrected)

Select rows and cols: Rows Cols Swap

Comparison field: Percent_correct

Significance: 0.05

Sorting (asc.) by: <default>

Test base: Select

Displayed Columns: Select

Show std. deviations: ☐

Output Format: Select

Test output

Tester: weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -result-ma

Analysing: Percent_correct

Datasets: 1

Resultsets: 2

Confidence: 0.05 (two tailed)

Sorted by: -

Date: 2022/5/10 下午5:32

Dataset	(1) function	(2) funct
'customer_churn-weka.filt(100)	89.00	89.88

(✓/ /*) | (0/1/0)

Key:

(1) functions.Logistic '-R 1.0E-8 -M -1 -num-decimal-places 4' 3932117032546553727

(2) functions.SMO '-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K \"functions.supportVect

Result list

- 15:49:01 - Available resultsets
- 15:49:05 - Percent_correct - functions.Logistic '-R 1.0E-8 -M -
- 16:28:17 - Available resultsets
- 16:28:22 - Percent_correct - functions.Logistic '-R 1.0E-8 -M -
- 17:32:04 - Available resultsets
- 17:32:08 - Percent_correct - functions.Logistic '-R 1.0E-8 -M -

信賴區間:0.05

採用雙尾檢定，logistic model 與 SVM model 準確率皆約為 89%，故此訓練集在兩模型的預測下差異不太，無法拒絕虛無假設