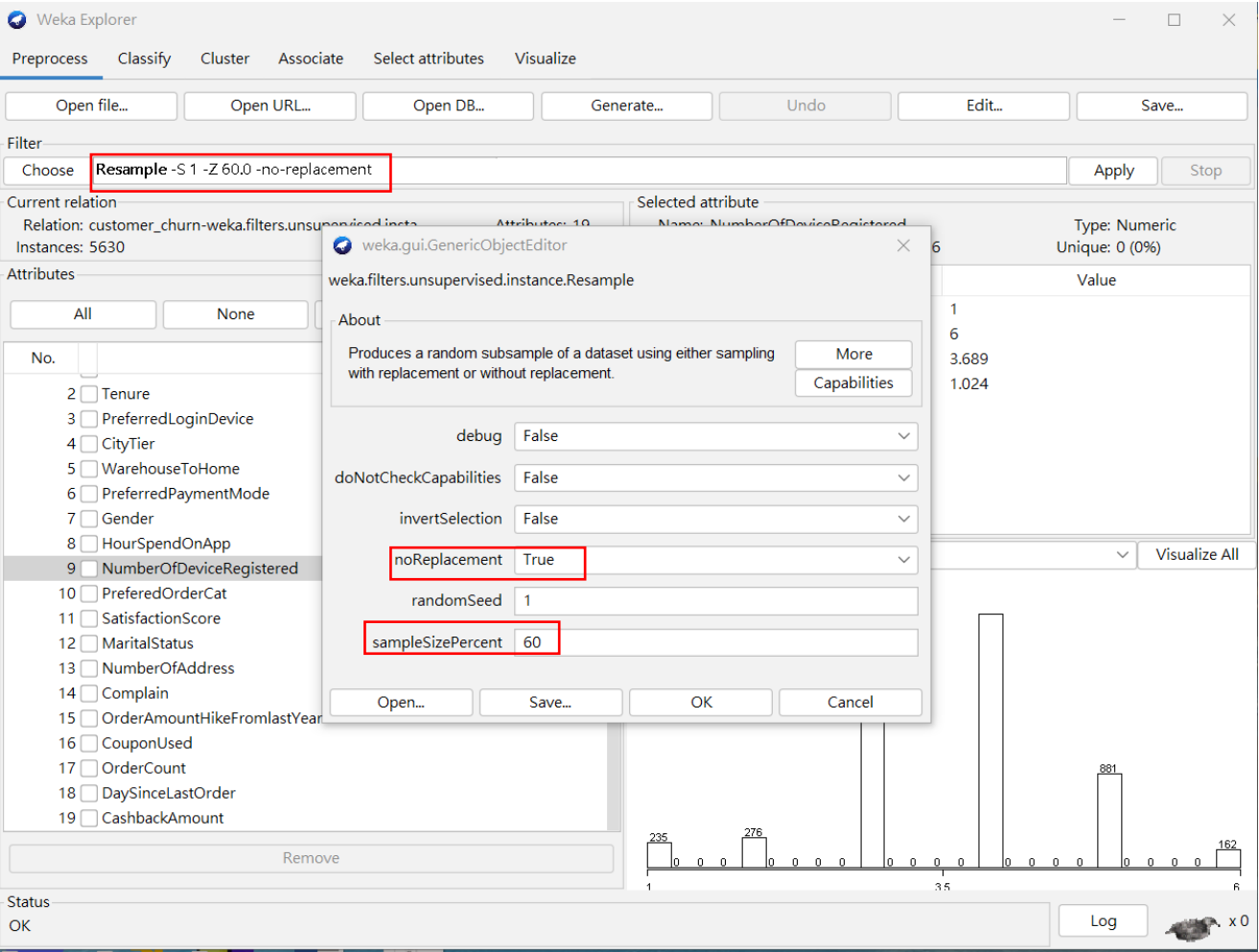


ETC_HW5_107403020

107403020_李泳輝

使用 ETC_HW3 的方法處理 customer_churn.csv

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



2. 顯示取樣後各類別的資料數量

Sampling之前：

Current relation				
Relation: customer_churn-weka.filters.unsupervised.instance.RemoveDuplicates-weka.filters.unsupervised.attribute.Remove-R1-weka.filter...				
Instances: 5630				
Attributes: 19				
Sum of weights: 5630				
Selected attribute				
Name: Churn				
Missing: 0 (0%)				
Distinct: 2				
Type: Nominal				
Unique: 0 (0%)				
No.	Label	Count	Weight	
1	0	4682	4682	
2	1	948	948	

Sampling之後：

Current relation		Selected attribute	
Relation: customer_churn-weka.filters.unsupervised.insta...		Name: Churn	
Instances: 3378		Missing: 0 (0%)	
Attributes: 19		Distinct: 2	
Sum of weights: 3378		Type: Nominal	
		Unique: 0 (0%)	
		No.	Label
		1	0
		2	1
		Count	Weight
		2821	2821
		557	557

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

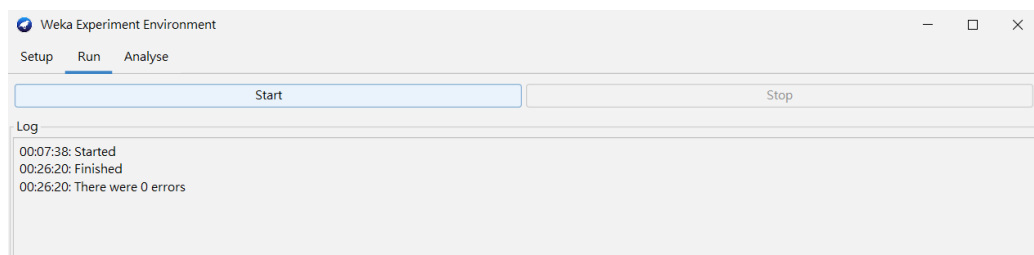
✓ 將Churn設定為要預測的class

Viewer	
Relation: customer_churn-weka.filters.unsupervised.instance.RemoveDuplicates-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.ReplaceMis...	
No.	1: Churn 2: Tenure 3: PreferredLoginDevice 4: CityTier 5: WarehouseToHome 6: PreferredPaymentMode 7: Gender 8: HourSpendOnApp 9: NumberOfDevices
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0

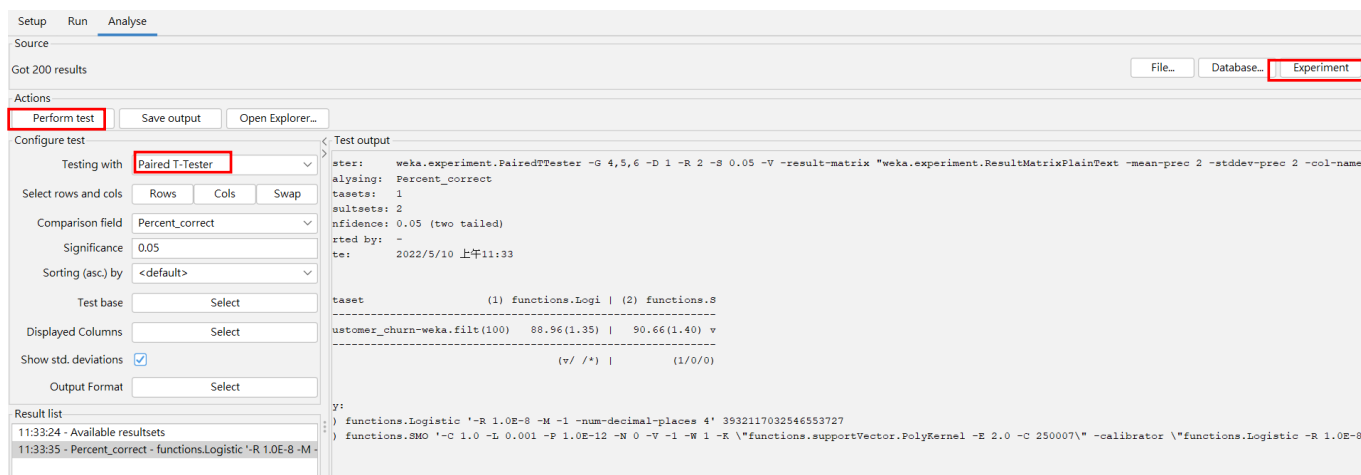
✓ 開啟Experimenter，進行初始設定

Weka Experiment Environment	
Setup Run Analyse	
Experiment Configuration Mode: Simple	
Results Destination	
ARFF file: C:\Users\user\Desktop\電子商務技術\Electronic-Commerce-technology\HW5\customer_churn_stratify.arff	
Experiment Type	
Cross-validation	
Number of folds: 10	
Classification	
Iteration Control	
Number of repetitions: 10	
Data sets first	
Algorithms	
Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	
SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector"	

✓ 執行



✓ 分析結果



4. 根據weka的輸出說明結論

SVM旁的V代表SVM的預測與Logistic Regression有明顯的差異，且預測的準確率結果大於Logistic Regression。