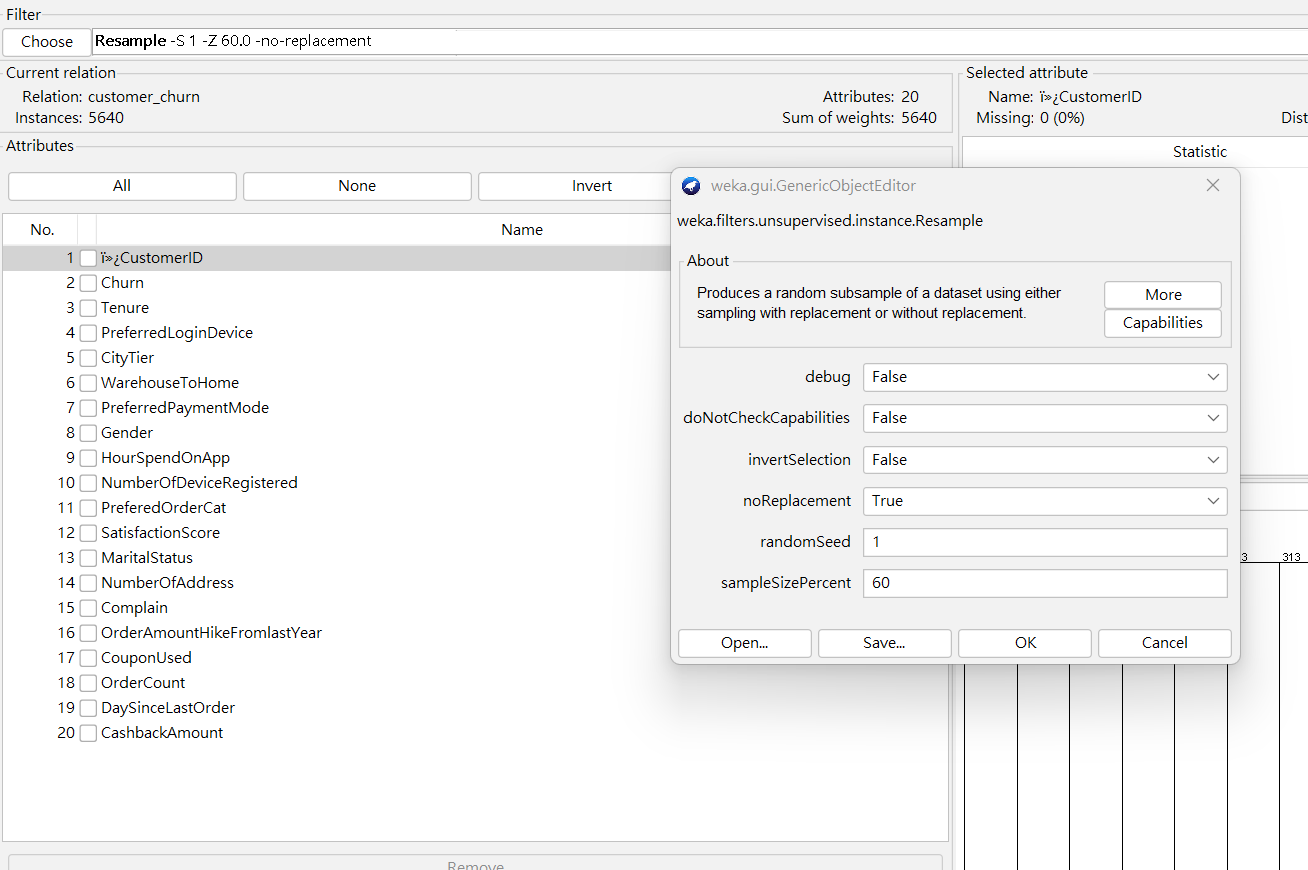
**ECT\_HW5\_109403021\_Weka題**

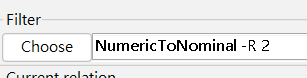
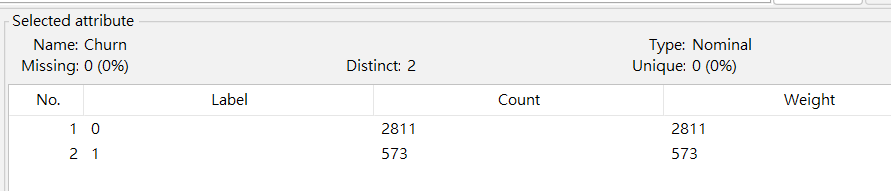
1. **取60%資料**

從Filter選取Resample並將sampleSizePercent調為60後Apply

****

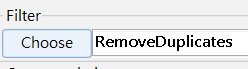
1. **列處churn資料個數**

用NumericToNominal將Churn欄為轉為nominal後可看到各類別資料個數

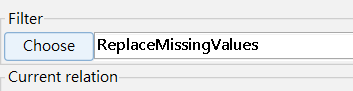
 

1. **資料前處理**

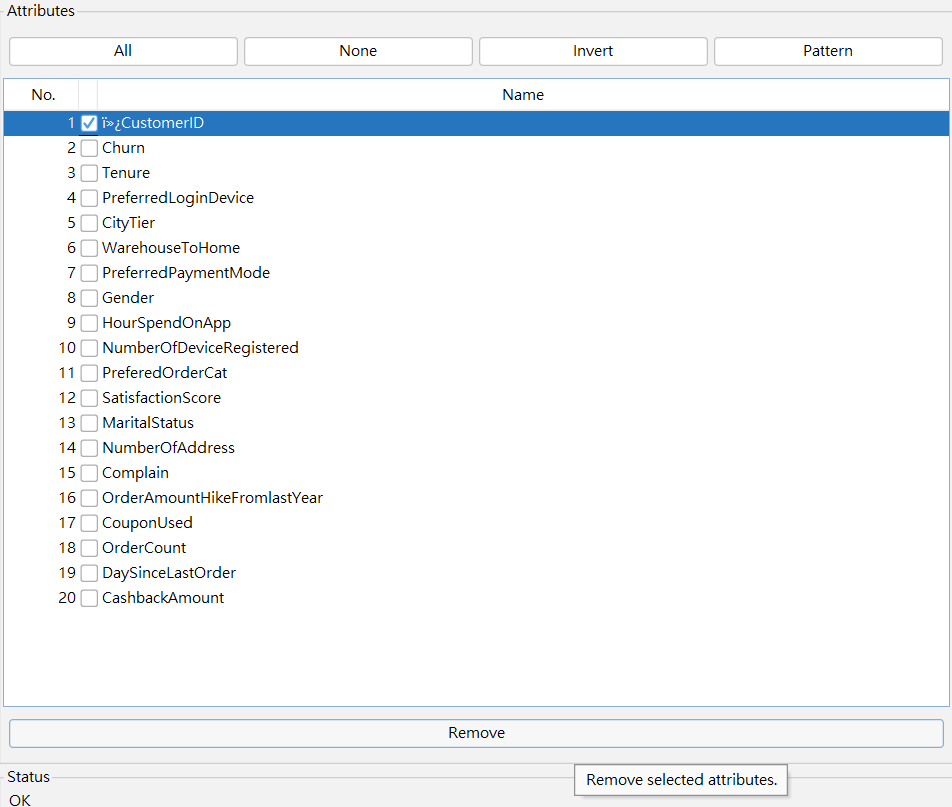
(1)先刪除重複多餘的資料(保留一筆)



(2)用平均數填補空值

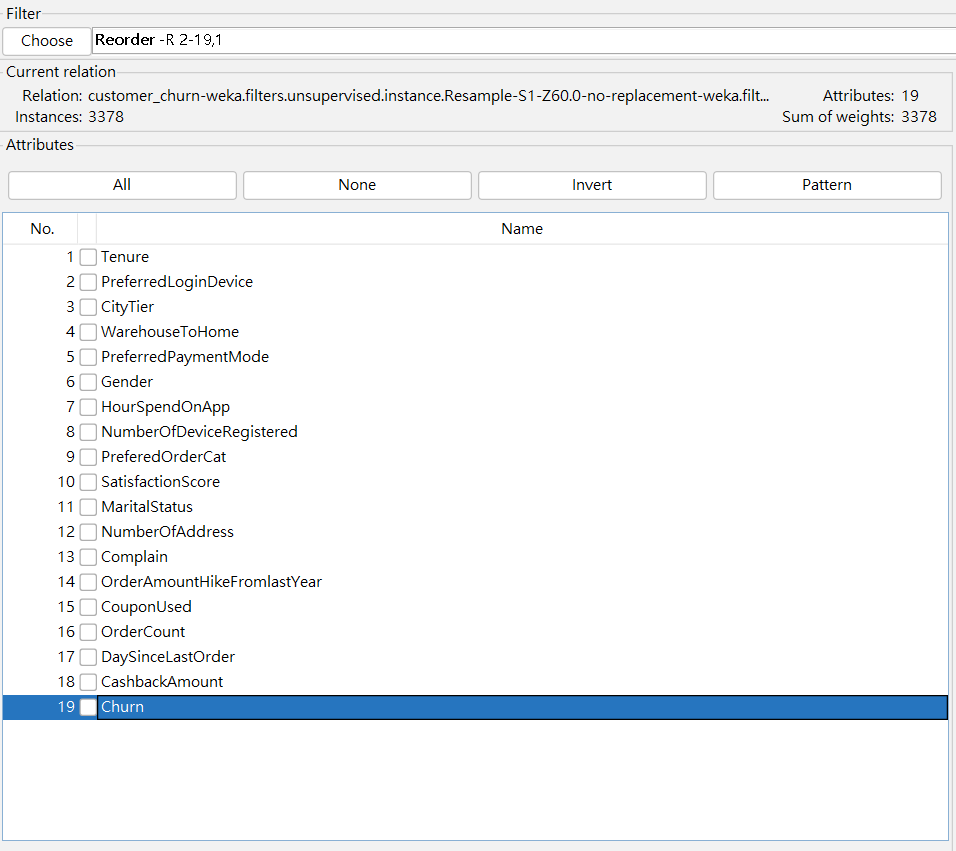


(3)刪除CustomerID欄位

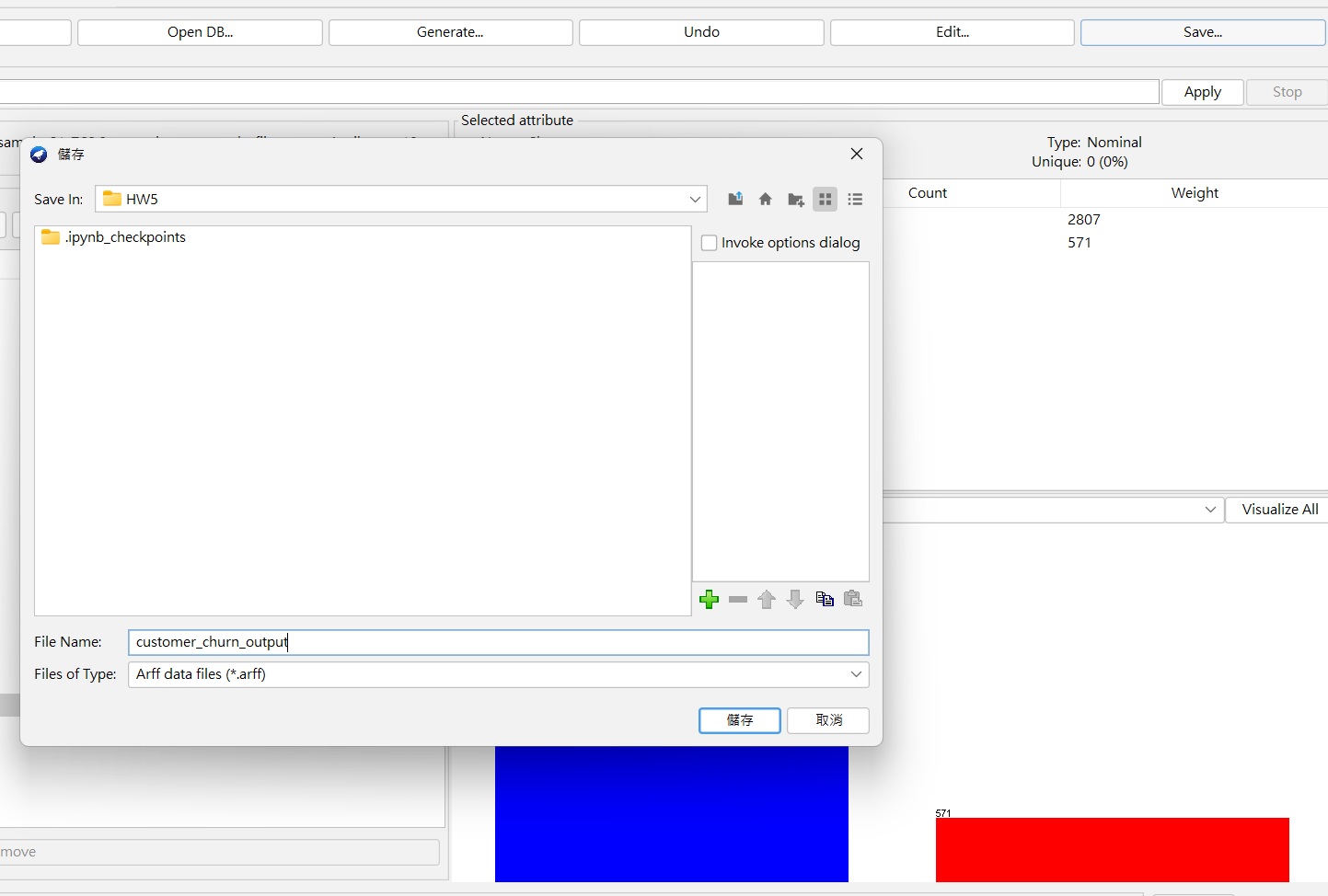


1. **用Experimenter比較兩模型**

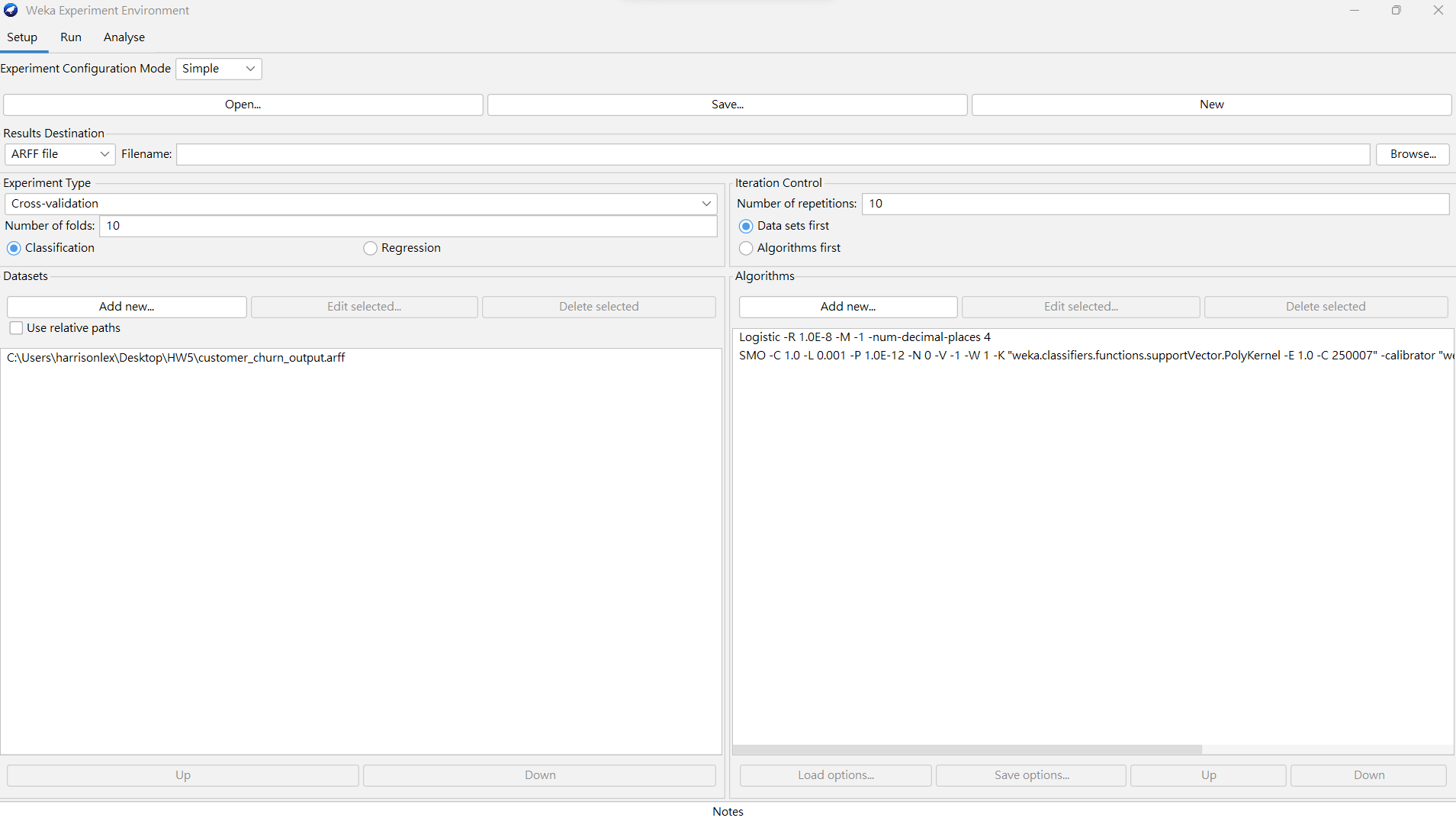
我先把Churn欄位調到最後讓等等experimenter直接識別其為目標變數



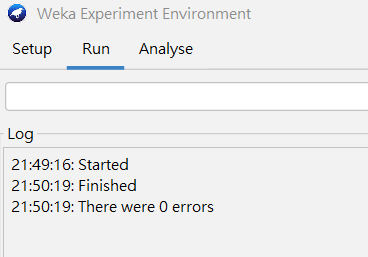
然後按save儲存檔案



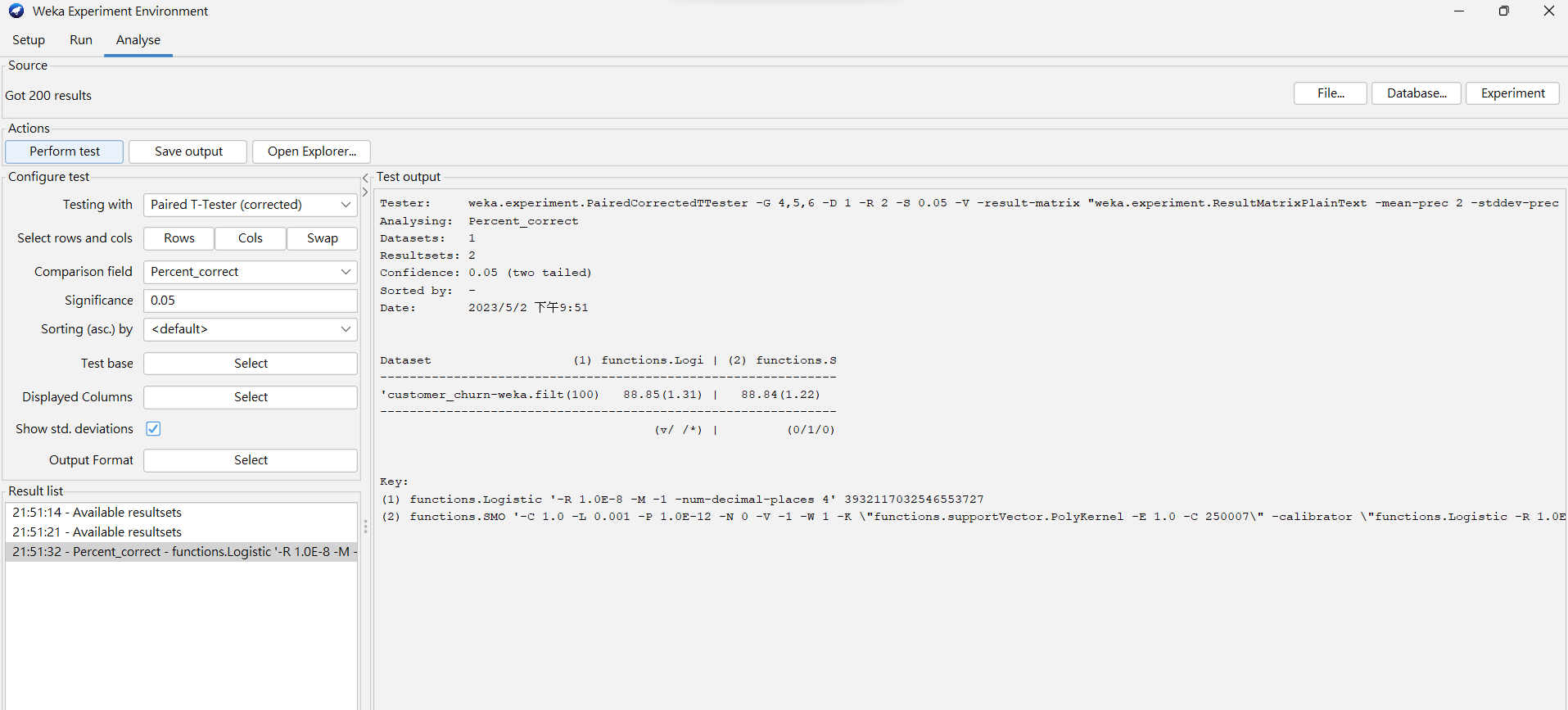
來到Experimenter。Datasets選擇剛輸出的檔案，Algorithms選到Logistic和SMO(兩模型參數我都用default的)，Iteration Control的Number of repetitions設為10，Experiment Type選Cross-validation並Number of folds設為10且選Classification



跑完

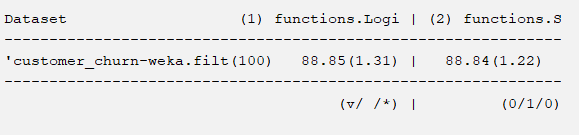


得到結果:

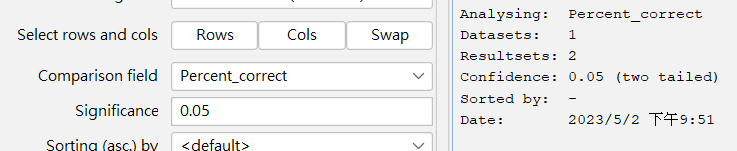


1. **結論**

根據第4題最後得出的t-test結果可以看到:



兩個模型的平均Accuracy分別為88.85和88.84，標準差分別為1.31和1.22。而且兩模型並沒有被標註v或\*(表示某模型顯著較好/較差)，表示兩模型的平均Accuracy並無顯著差異。



由於顯著水準設置為0.05，可得出的結論就是:

在信心水準95%下，此次實驗SVM模型和Logistic Regression表現上並無顯著差異。