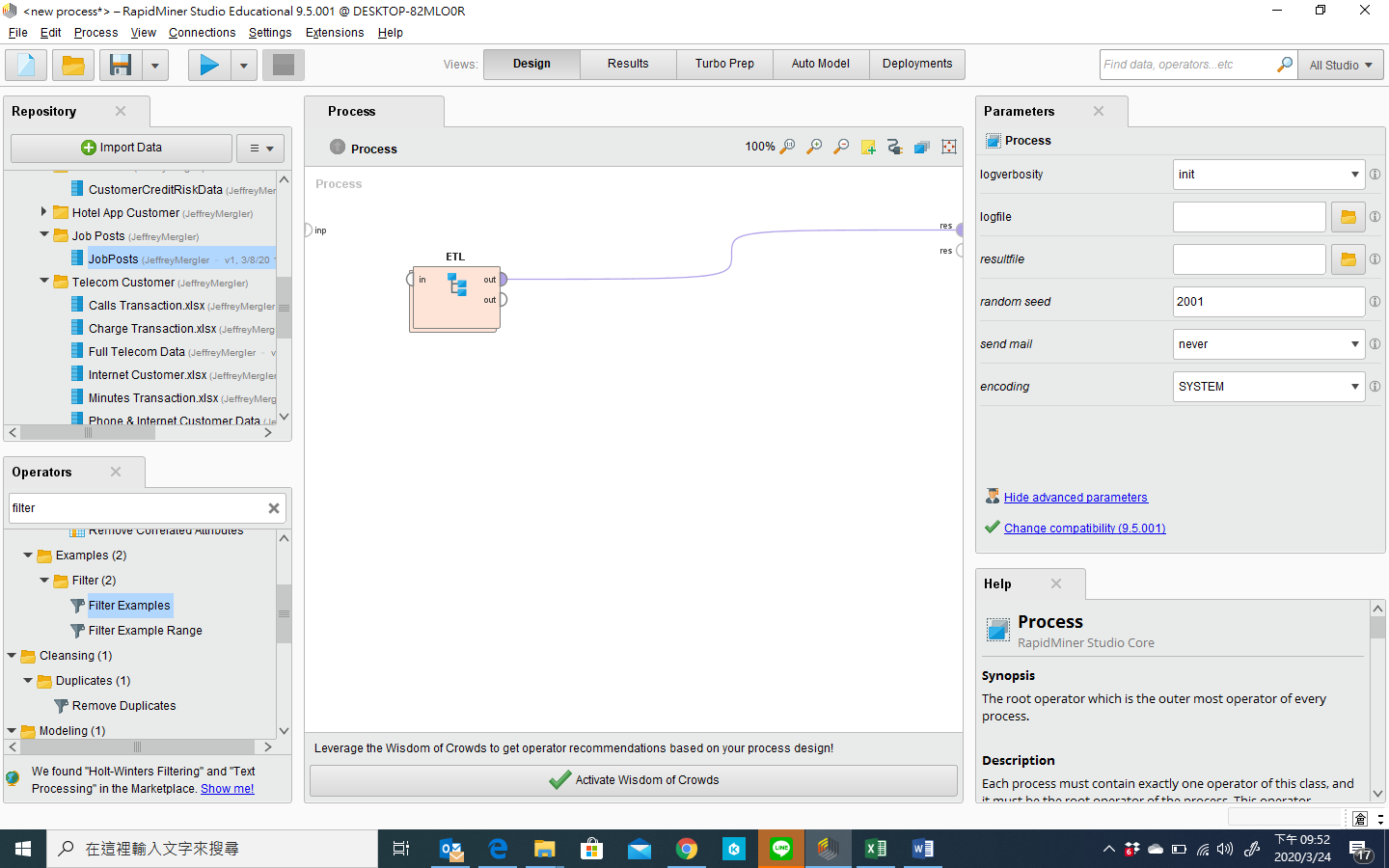
作業2: 看我用方程式算命

是否很期待呢,我們終於要開始第一個分析專案了。

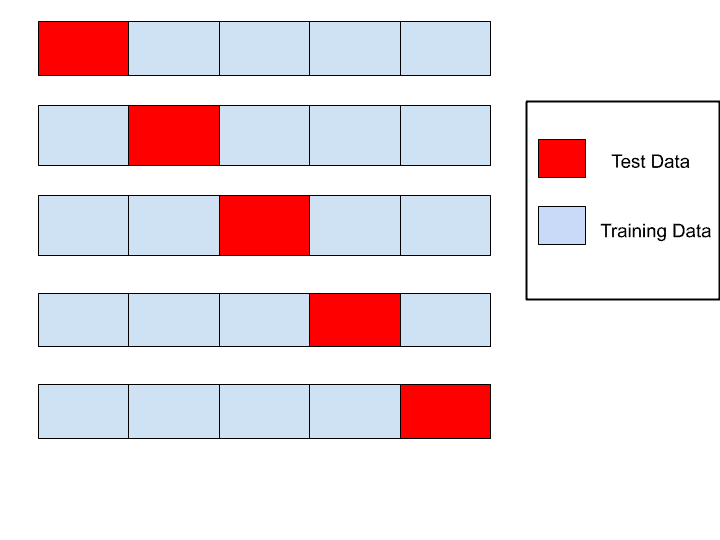
講到分析，不得不來了解一下大名鼎鼎的狠角色：迴歸方程式.

以下作業請先讀入baston檔案,並開始分析:

1. 讀入檔案後,請大概說明一下您對於資料的了解(10%),並進行「必要」之資料前處理,並請將相關前處理存於ETL子流程如下圖(請說明您做了什麼,為什麼) (10%) **→ 請寫出敘述統計 （EX.平均值、是否為常態分配）**



2. 請找出cross validation, 並利用它進行模型的建置.何謂cross validation? (10%)



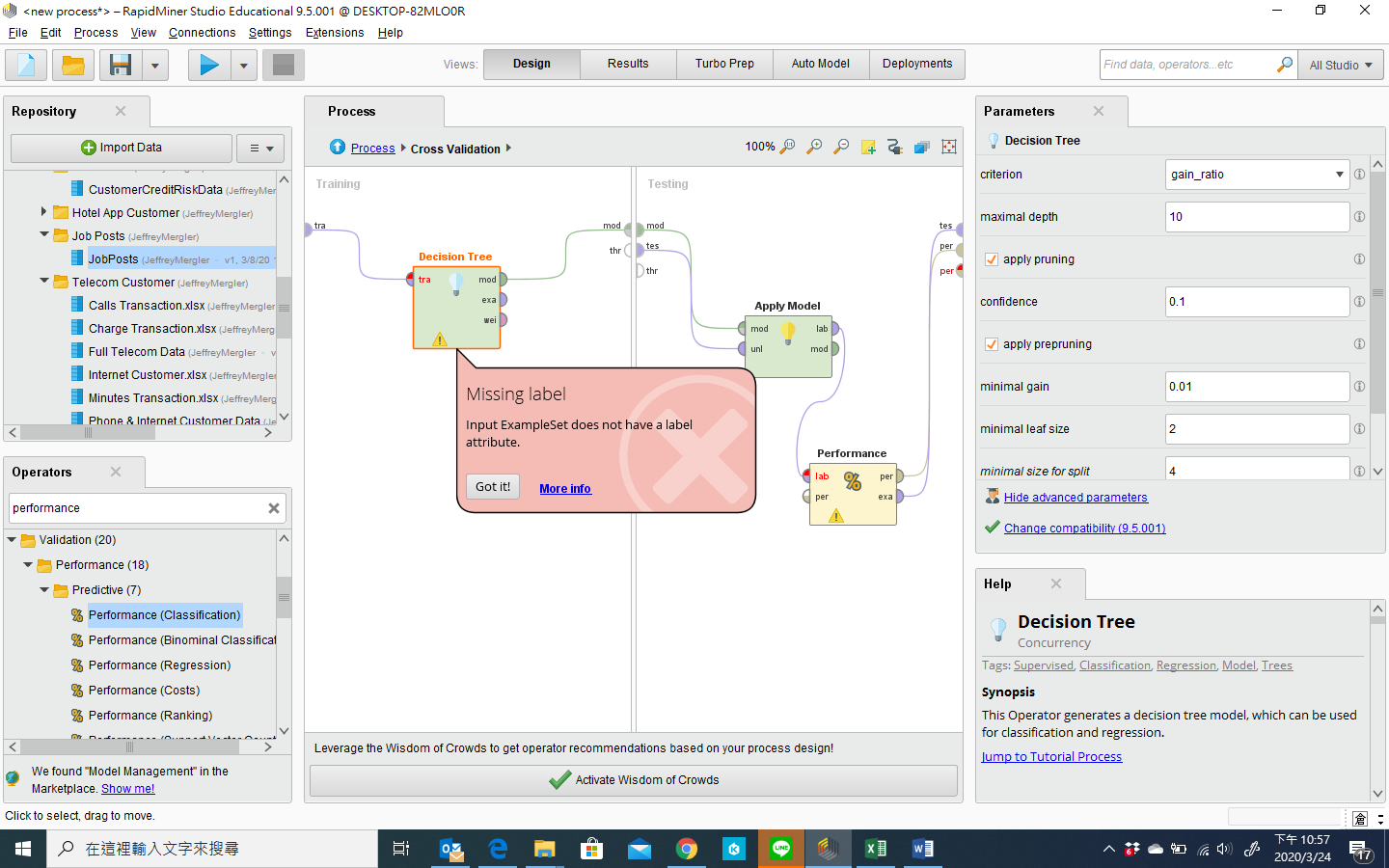
Cross Validation：

EX. 跑五次，第一次用第一批次資料做測試集、第二次用第二批次資料做測試集…

留一法交叉验证 Leave-One-Out Cross Validation：只有對或錯  
→假設只有十筆資料，切分成十個批次(number of folds=10)，也就是一個批次只有一筆，預測準確率僅100%或0%，沒有中間值(EX.90%)。

3. 請在cross validation中進行回歸的建構，並請貼出您的流程圖，並請說明相關之參數設定 (10%)

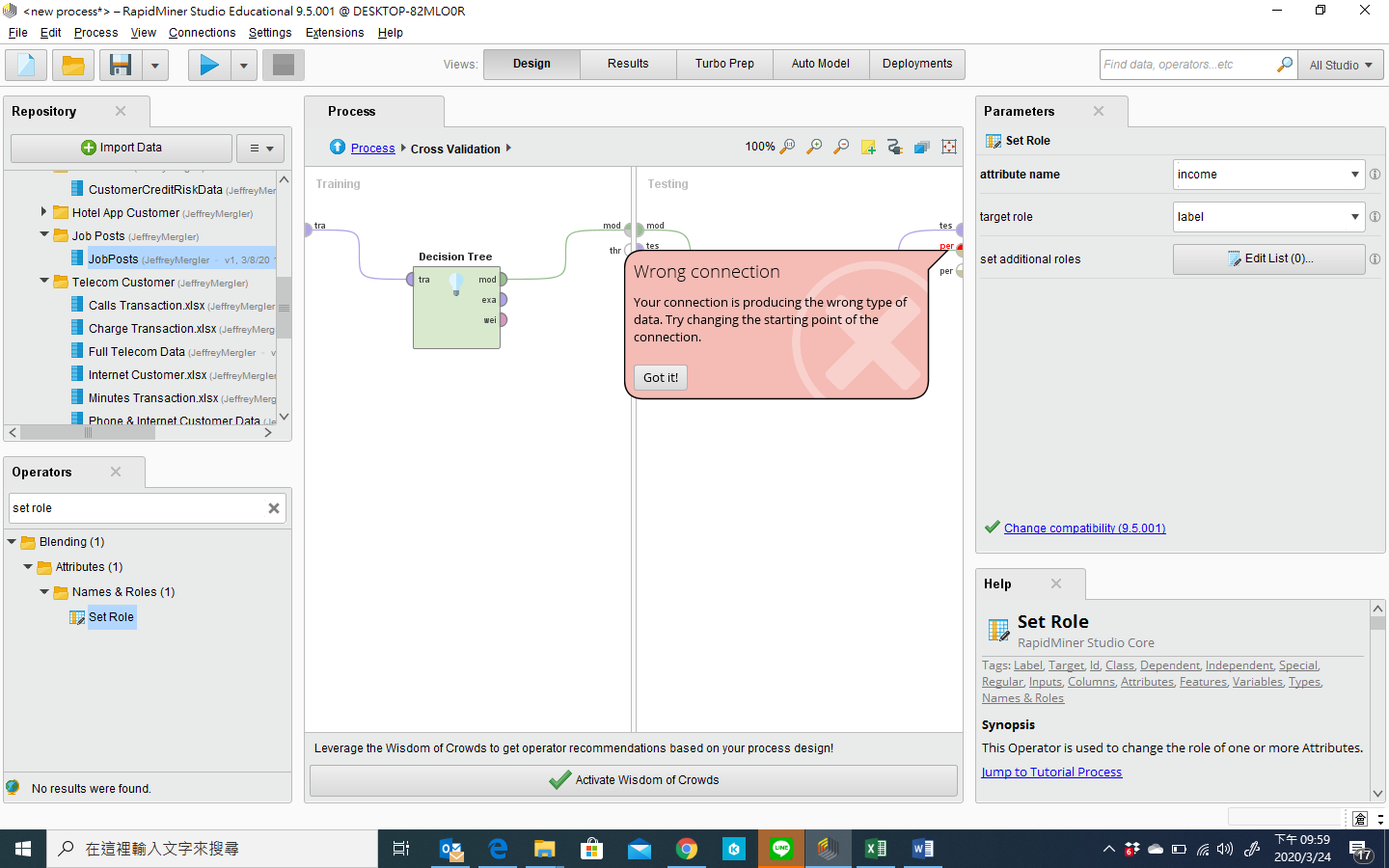
4. 請問以下錯誤訊息是指? 該如何解決(5%)



沒有跟模型說要預測哪一個欄位；

要利用Set Role設定標籤（Label）

5. 請問以下錯誤訊息是指? 該如何解決(5%)



6. 若您的模型在建構過程中有上述問題，請試著解決並以MEDV為預測標的，貼出您的決策樹,並加以說明您的方程式意涵(10%)

7 請列出您的~~混肴矩陣~~,並請加以說明其意涵，除此之外，我還希望您能列出您模型的下述４個指標，並請簡略說明他們的計算方式及含意(10%) **→ 請列出2-4個**



Accuracy、ClassificationError→分類問題才有

因此只要顯示root mean squared error及squared error

一張含有 文字 的圖片

自動產生的描述

但是,您用了幾乎全部的變數.幾乎全部的變數. 幾乎全部的變數.

〔Feature Selection〕

8.以下我們將限制您的變數使用量，

8-1請問若只讓您使用5個變數進行預測,您會選擇(請列出變數名稱)? (5%)

8-2請問若只讓您使用3個變數進行預測,您會選擇(請列出變數名稱)? (5%)

8-3您的選擇依據是? (5%) 變數變少後,正確率的變化是? (5%)

9.最後，在５個變數的限制下，請調整參數，找出一條正確率最高的迴歸方程式，並請說明您試過那些參數的調整？ (10%)

請將您的process存檔為學號-1, 如: 106AB001\_1.rmp檔