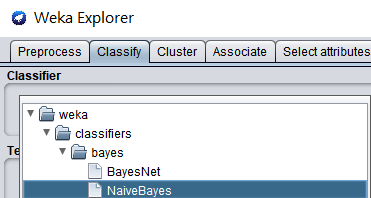
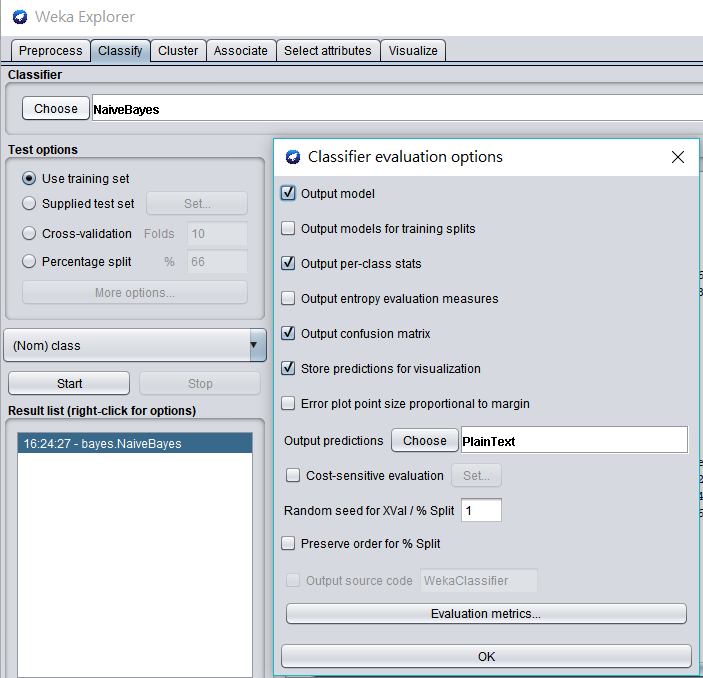
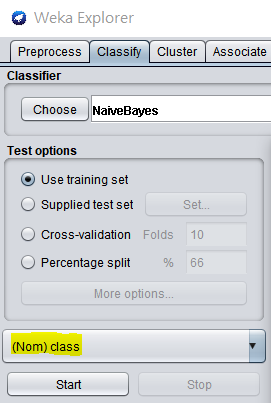
1. 重要步驟截圖及說明
2. 打開weka的Explore之後，在Classify Panel中的Classifier選取「NaiveBayes」。



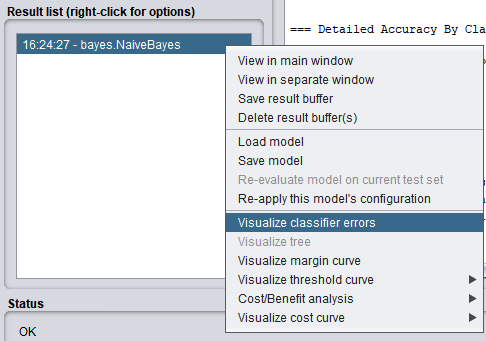
1. 在Test options中勾選「Use training set」，並在點選「More options」後，在「Output predictions欄位選擇「PlainText」。



1. 在螢光筆處選擇「(Nom)class」，並點選「Start」開始分析。

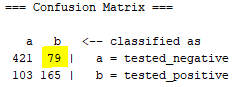


1. 若要利用 Visualize Classifier Errors，則右鍵Result list中執行的結果，並選取「Visualize Classifier Errors」。

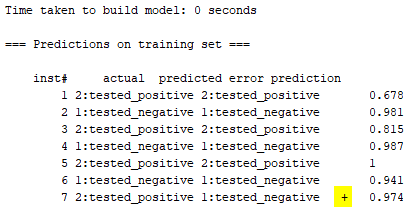


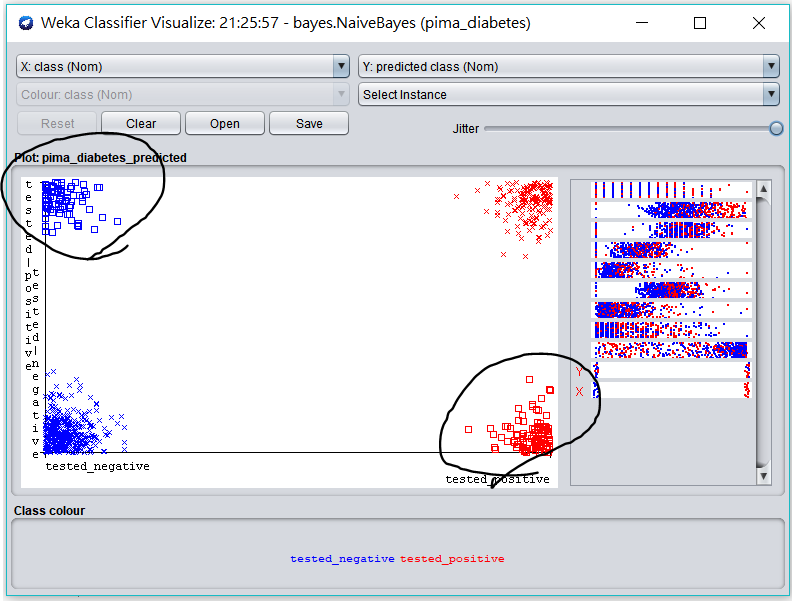
1. 錯誤率為23.6979% 

由題，「Test dataset instances被分類到 tested\_negative class，但實際上屬於tested\_positive class」，代表Confusion matrix中的「false negative (FN)」，由圖可知總共有79筆Test dataset instances被分類到 FN，可得百分比為 79/768 \* 100% = 10.2864%

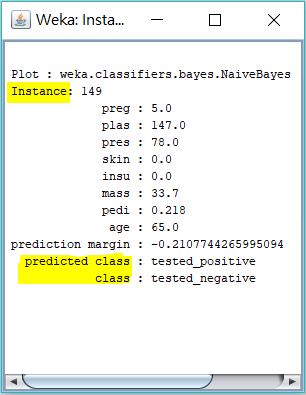


1. 欄位 error 出現“+”代表欄位 actual 被錯誤地預測為欄位predicted 的情況；以inst7 為例，代表 tested\_positive 被錯誤地預測為 tested\_negative。

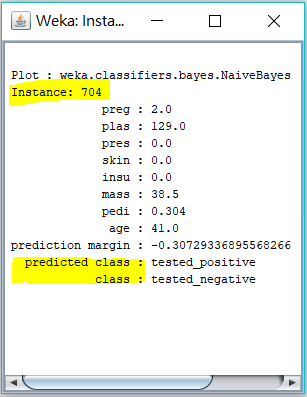




1. 若點取左上角的藍色方框，可得預測錯誤的資料點的詳細資訊（如下圖）

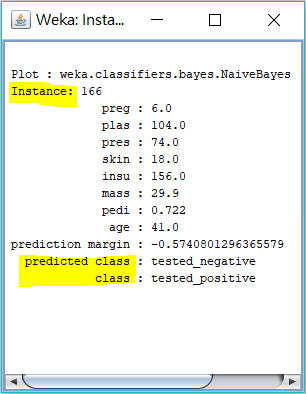


可由instance得知為第149筆資料，預測的class為tested\_positive，但實際上為tested\_negative，屬於Confusion matrix中的 FN。



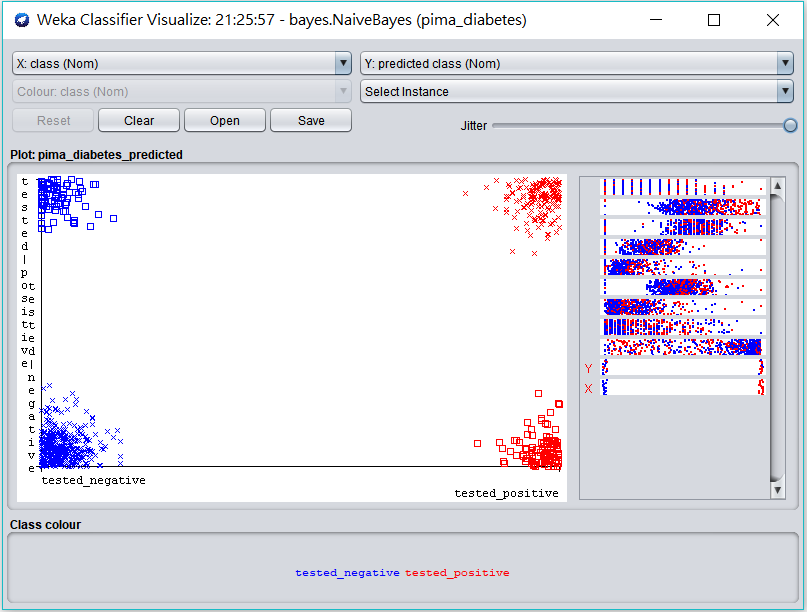
可由instance得知為第704筆資料，預測的class為tested\_positive，但實際上為tested\_negative，屬於Confusion matrix中的 FN。

1. 若點取右上角的紅色方框，可得預測錯誤的資料點的詳細資訊（如下圖）



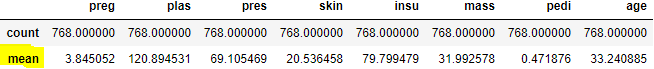
可由instance得知為第166筆資料，預測的class為tested\_ negative，但實際上為tested\_ positive，屬於Confusion matrix中的 FP。



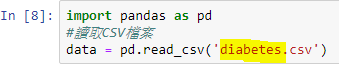


由(c)可知，當x軸代表class(Nom)、y軸代表predicted class(Nom)時，左上角的資料代表Confusion matrix中的FN，右下角代表Confusion matrix中的FP；左下角的資料皆為tested\_negative，因此屬於Confusion matrix中的TN，而右上角的資料皆為tested\_positive，因此屬於Confusion matrix中的TP。

1. 由圖所示， preg平均值為3.845052，plas平均值為120.894531，pres平均值為68.105469，skin平均值為20.536458，insu平均值為79.799479，mass平均值為31.992578，pedi平均值為0.471876，age平均值為33.240885



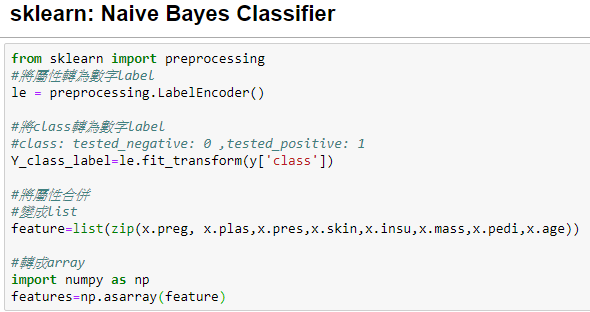
1. 開啟Anaconda Navigator的 Jupyter。
2. 將Naïve Bayes\_weather\_example.ipynb 與 diabetes.csv 上傳到Jupyter，並開啟Naïve Bayes\_weather\_example.ipynb。
3. 將weather改成diabetes。



1. 填入所需的x:input和y:output。

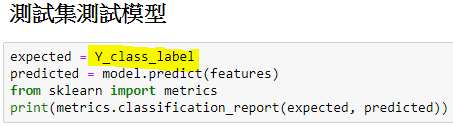


1. 由於其餘屬性皆為數字，所以只將class轉為數字label、並將屬性合併變成list

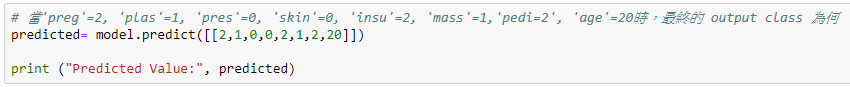


1. 變更參數名稱

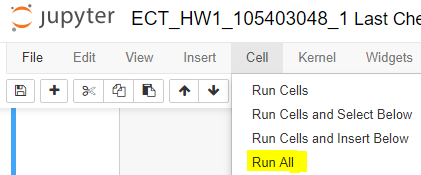




1. 依(e)題，變更最後的預測參數



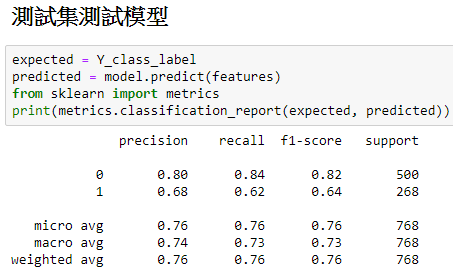
1. 按下Run all執行所有cell的程式，即可得所有cell之output。



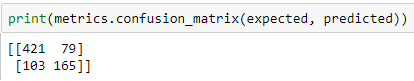
1. Precision(P) = ，表示的是預測為正的樣本中有多少是真正的正樣本。

Recall(R) = ，表示的是樣本中的正例有多少是被預測正確了。

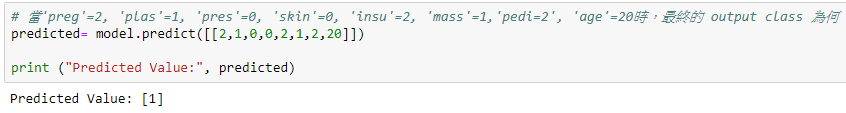
F1-Score = ， F1值即Precision和Recall的調和均值。



1. 共有421筆資料屬於TN，79筆資料屬於FP，103筆資料屬於FN，165筆資料屬於TP。



1. 結果為1，代表tested\_positive



1. Python相較於Weka而言，應用較為廣泛，可以再cell中執行更多的預測情況，如題(e)所示；但Weka提供較佳的用戶者體驗，操作介面較簡單且直覺。