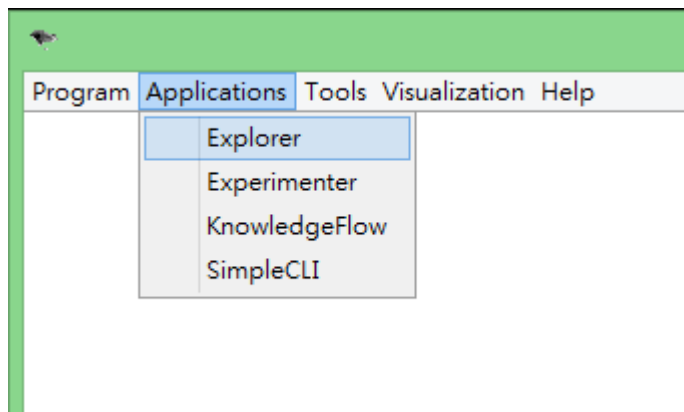
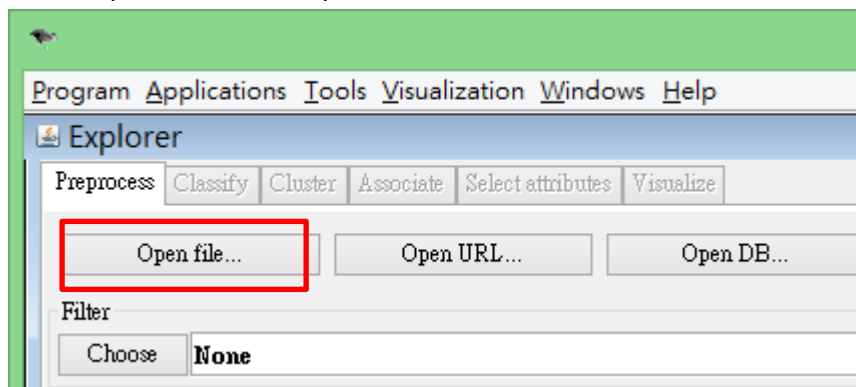


開啟 WEKA 選擇 Applications → Explorer

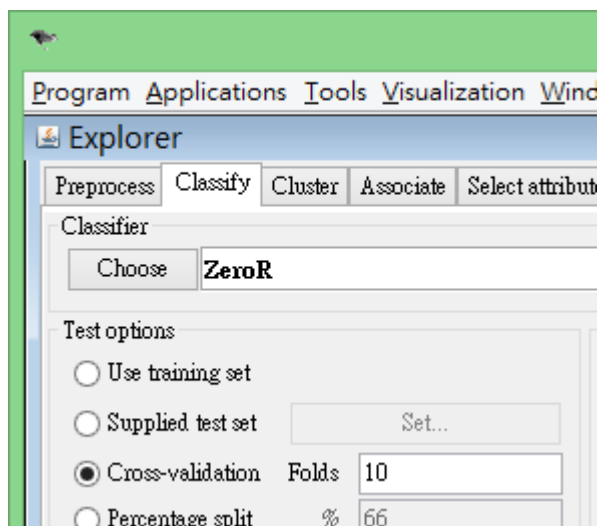


進入 Explorer 後選擇 Open file

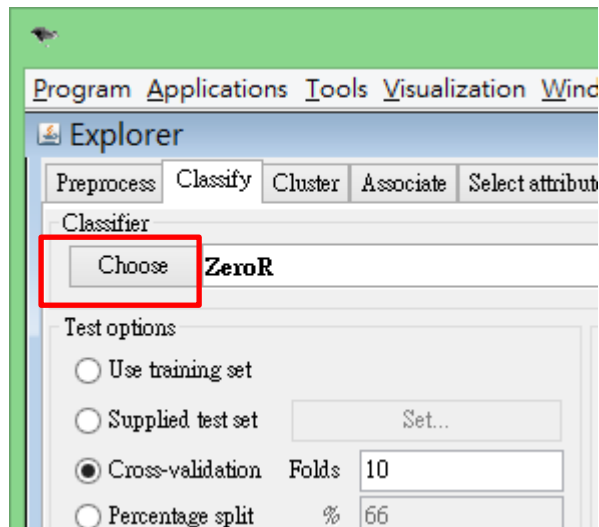


選擇已經整理好的檔案進行載入

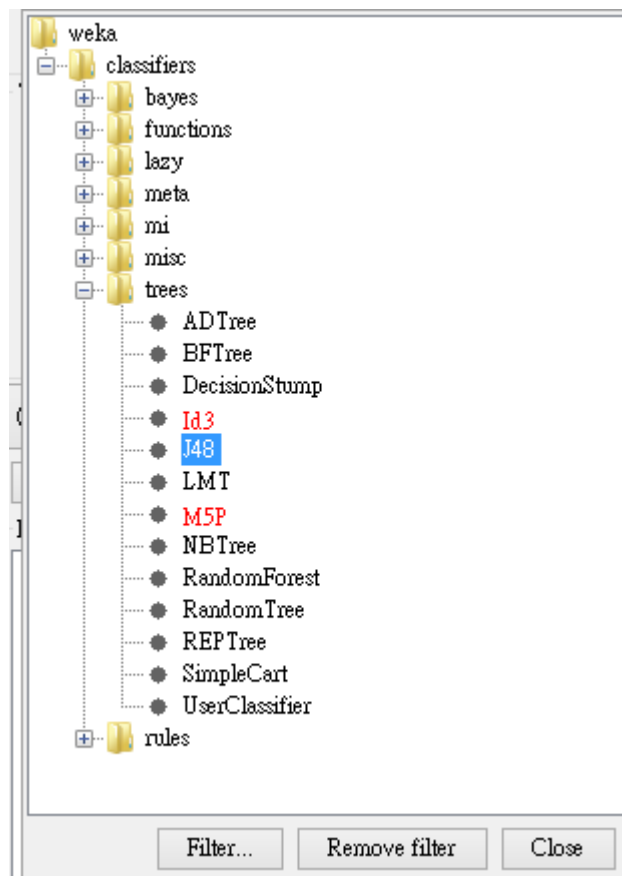
載入檔案後點選 Classify(分類)



接著進行分類法的選擇，點選 choose 按鈕



接著選擇分類樹 J48

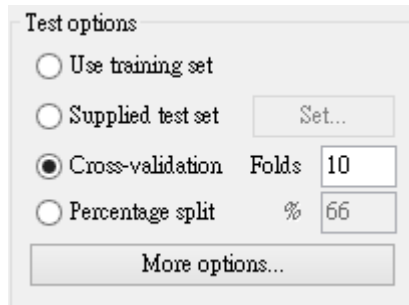


Use Training Set：將訓練資料集，轉成測試資料集

Supplied test set：可以另外設定外來資料做測試資料集

Cross-Validation：交叉驗證，後面框框是設定幾次（一般設定都是 10 次）

Percentage split：將原始資料分割 66% 為訓練，34% 為測試（這是常用的設定）



Test options

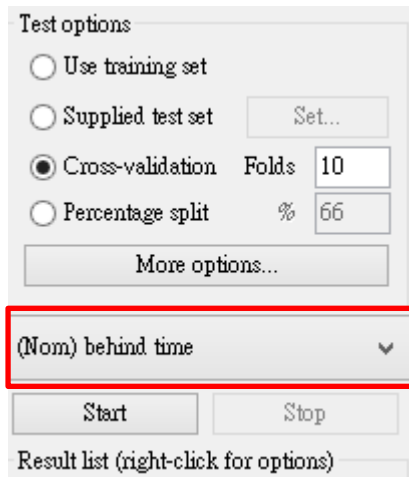
☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

選擇分類的依據之後點選 **Start**



Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

Result list (right-click for options)

執行後會得到一份結果

The screenshot shows the Weka interface with the 'Classifier output' window open. The 'Test options' section on the left shows 'Cross-validation' selected with 'Folds' set to 10. The 'Classifier output' window displays the following text:

```
age
credit rating
occupation
income(year)
loan>deposit
behind time

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

loan>deposit = LOW: NO (78.0/5.0)
loan>deposit = HIGHT
| age <= 57
| | income(year) = MEDIUM
| | | credit rating = GOOD
| | | | age <= 31: YES (8.0)
| | | | age > 31: NO (2.0)
| | | credit rating = MEDIUM: NO (13.0/5.0)
| | | credit rating = BAD: YES (0.0)
| | income(year) = HIGHT: YES (80.0/17.0)
| | income(year) = LOW: YES (1.0)
| age > 57: NO (17.0/3.0)

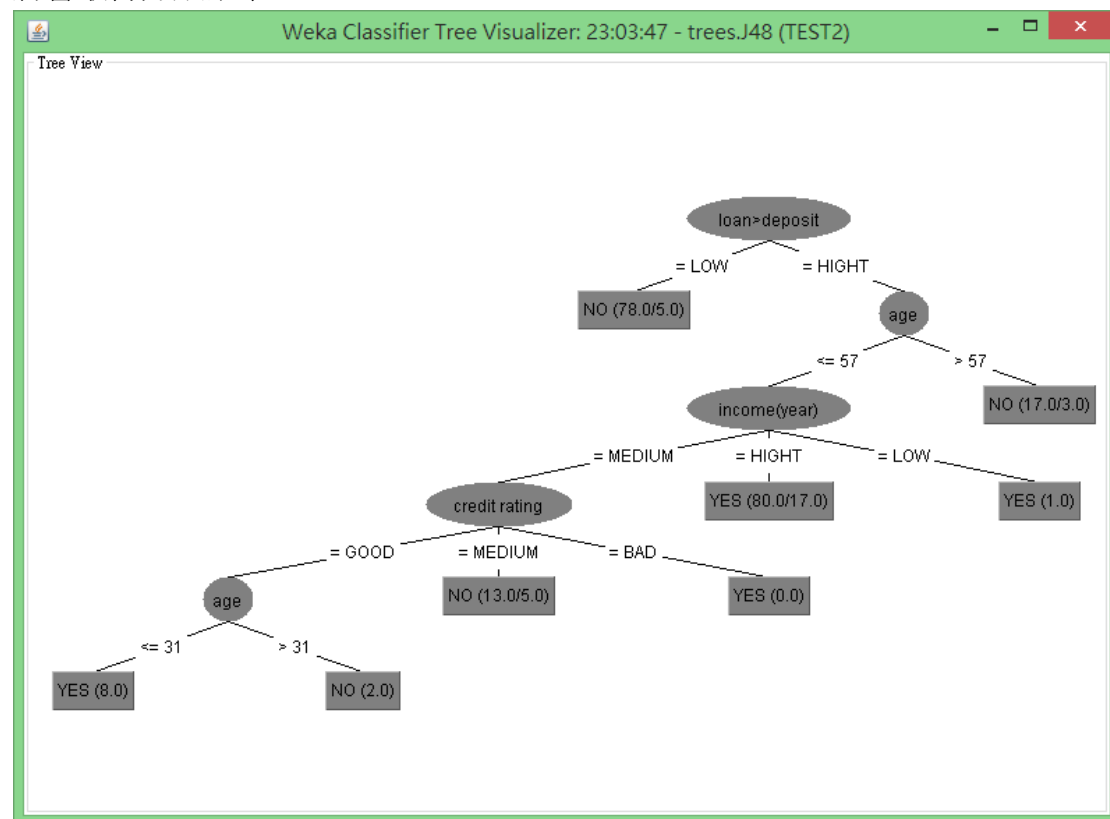
Number of Leaves : 8
Size of the tree : 13
```

對執行結果的選項點滑鼠右鍵→Visualize tree

The screenshot shows the 'Result list (right-click for options)' window with the 'J48 pruned tree' result selected. A right-click context menu is open, showing the following options:

- View in main window
- View in separate window
- Save result buffer
- Delete result buffer
- Load model
- Save model
- Re-evaluate model on current test set
- Visualize classifier errors
- Visualize tree** (highlighted)
- Visualize margin curve
- Visualize threshold curve
- Visualize cost curve

將會取得分類結果



名詞解釋在下頁

解釋分析內容：

Correctly Classified Instances：分類到的屬性為正確的

Incorrectly Classified Instances：分類到的屬性為不正確的

Kappa statistic：是比較兩個或多個觀測者對同一事物觀測結果是否一致，或者是觀測者對同一事物兩次或多次觀測結果是否一致，由於機率造成的一致性和實際觀測的一致性之間的差別大小作為評價基礎的統計指標。

Total Cost：基本上即為 **Incorrectly Classified Instances**，也就是比較後錯誤的部份

TP Rate：正確辨識率

FP Rate：類別錯認率

Accuracy： $(TP+FN)/(TP+TN+FP+FN)$ ，其實也就相當於上面的機率，成功的數目/總樣本數

Precision： $TP/(TP+FP)$ ，其實就是在說，系統說這個「是」的這一類裡($TP+TN$)，有哪些是真的正確的(TP)

Recall： $TP/(TP+FN)$ ，那這個就是在正確的樣本裡面($TP+FP$)，有哪些是系統測試成功的(TP)

F-Measure：評估聚類後的品質好壞

貝氏分類 (Bayes)

- Naive Bayes, 簡單貝氏分類

函式分類 (Functions)

- SMO, 依序最佳化的支持向量機
- Logistic, Logistic 回歸

懶人法 (Lazy)

- IBK, 最鄰近搜尋法
- LBR, 懶人貝氏分類

群體學習 (Meta)

- Adaboost.M1, Adaboost 分類法
- Bagging, Bagging 分類法
- Vote, 多個模型整合投票

多重執行個體 (Multi-Instance, MI)

混合式 (Miscellaneous, Misc)

規則分類 (Rules)

- M5Rule, M5 規則式演算法

決策樹 (Tree)

- J48, Weka 裡頭的 C4.5 演算法
- LTM, Logistic 回歸模型決策樹
- M5P, M5 演算法加強版
- Decision Stump, 單層決策樹
- Random Forest, 隨機森林演算法

參考資料(2013/12/20)：

元智大學-資訊工程系：<http://www.cse.yzu.edu.tw/project/9846.doc>

Android 刷機症候群：http://123android.blogspot.tw/2011/12/weka_24.html