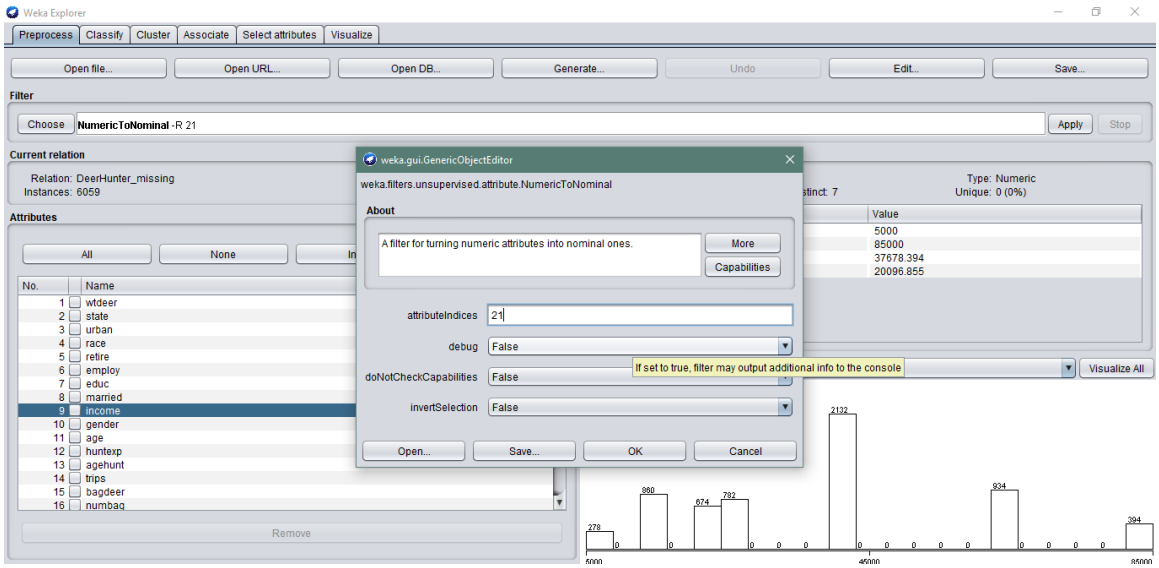
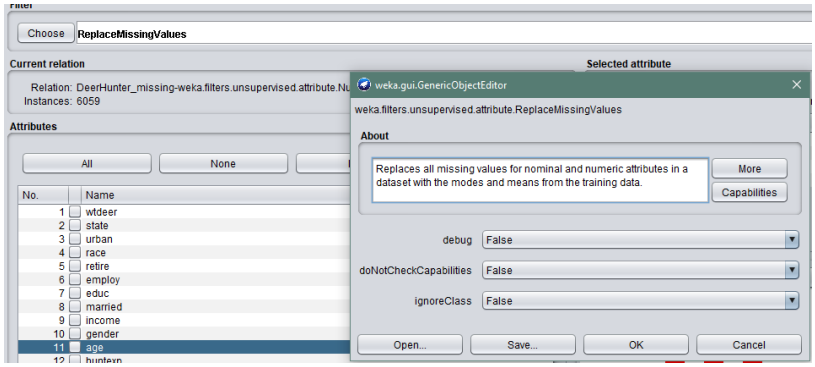
1. 利用Weka對DeerHunter.arff進行分類，需要包含J48、Naïve Bays兩種方法，依序完成以下步驟及問題: 資料前處理：須包含以下動作，並說明該動作的含義，如何在 Weka 中實現及實 現順序

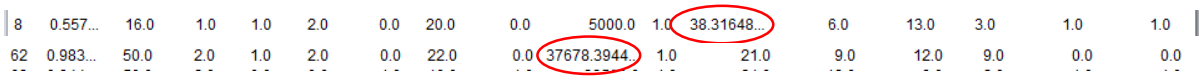
由於 J48 和 Naïve Bayes 不能夠使用於 Numeric instances，所以要先將資料做轉換為 Nominal 的形態。

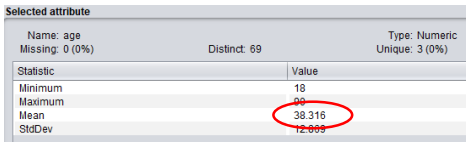


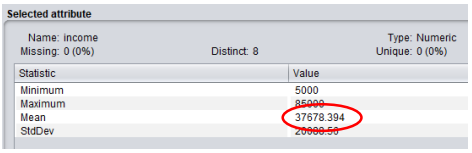
(a) Replace Missing Value(需要列出補上的值為何)(10%)



Take advantage of this following filter to replace missing value with specific attribute **mean**. (Income: 5 instances, Age: 5 instances.)



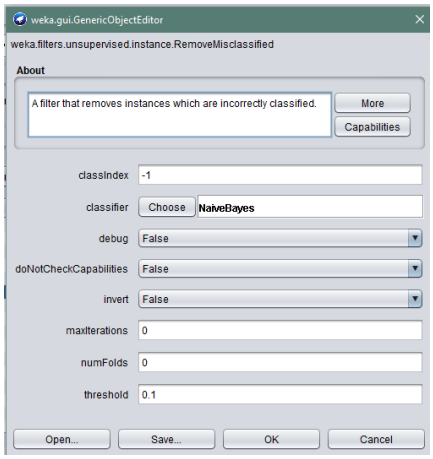




Fill in the missing values of age and income with attribute mean, which is 39.316 and 37678.394 respectively.

The missing value of age is No. 8, 1493, 3712, 5101, and 5954, on the other hand, of age is No. 62, 1360, 3438, 5171, and 5579.

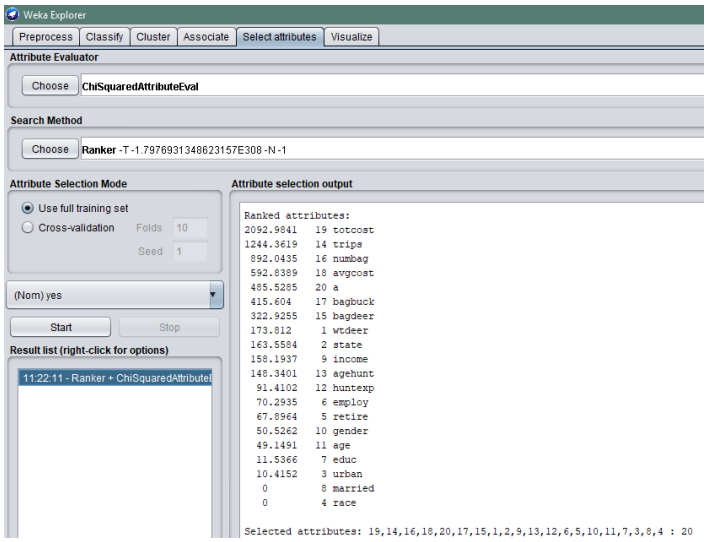
(b) Outlier Detection& Remove(10%)



Take advantage with filter RemoveMisclassified, which then set the classifier as NaiveBayes, to delete those wrongly classified instances.

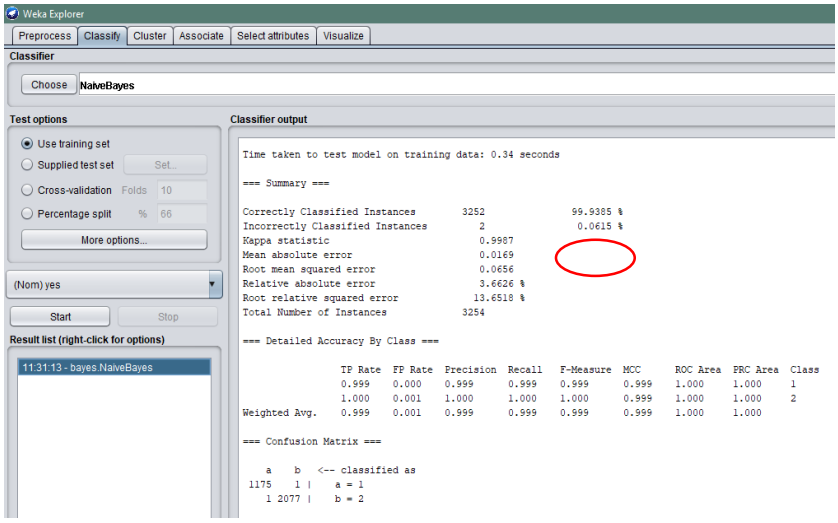
(c) Attribute Selection，請篩選出適合屬性(10%)

建模：附圖並詳細說明每一步驟執行原因、執行過程、執行結果。



Redundant attributes may be detected by ChiSquaredAttributeEval, so that the larger Chi-square value, the more likely are related. Delete those attributes value that is 0.

(d) 分類結果正確率需達到 95%以上，說明最佳模型所使用的演算法及各項參數設定意義(20%)



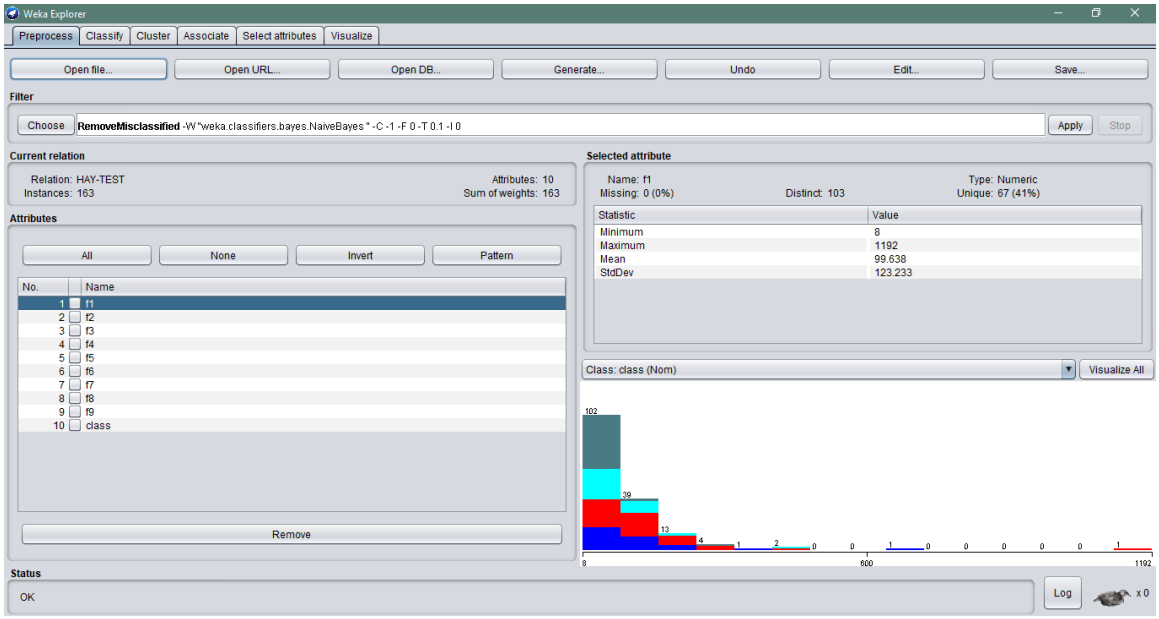
Classify dataset with Naïve Bayes Classifier, then get the result of 99.9385% correction rate. 在前處理時補上 missing values 以及刪除錯誤 classified 的 instances，所以正確率高。

2. 利用Weka 對 hay_train.arff 進行分類，需要包含 Multilayer Perceptron、 J48 及 Naive Bayes 三種方法，依序完成以下步驟及問題： 資料前處理：需包含以下動作，並說明該動作的含義，如何在 Weka 中實現及實 現順序（注意：資料集 hay-train.arff 及 hay-test.arff 皆須進行資料前處理）

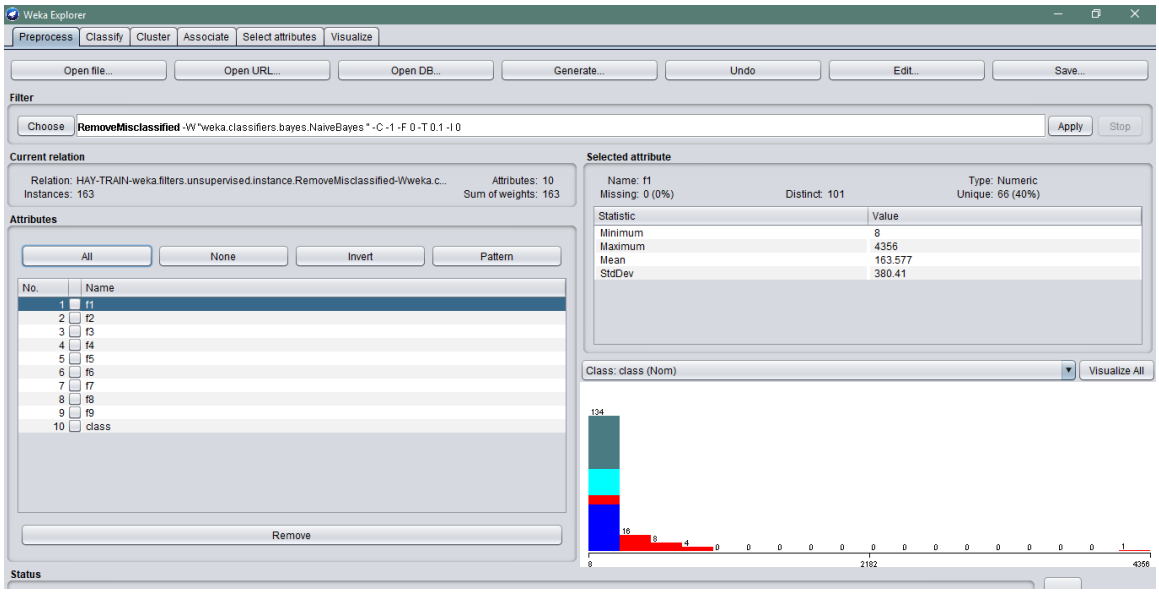
(a) Outlier Detection & Remove，使用RemoveMissclassified，分類器 設定為 Naive Bayes(10%)

使用 RemoveMisClassified，且以 Naïve Bayes 進行前處理，delete wrongly classified instances.

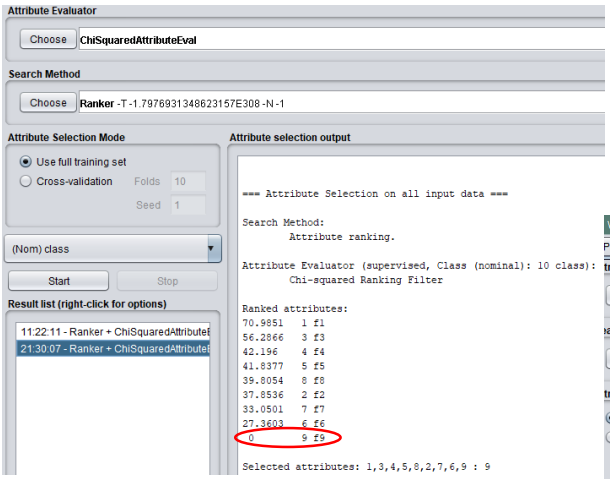
Hay-test:



Hay-train:

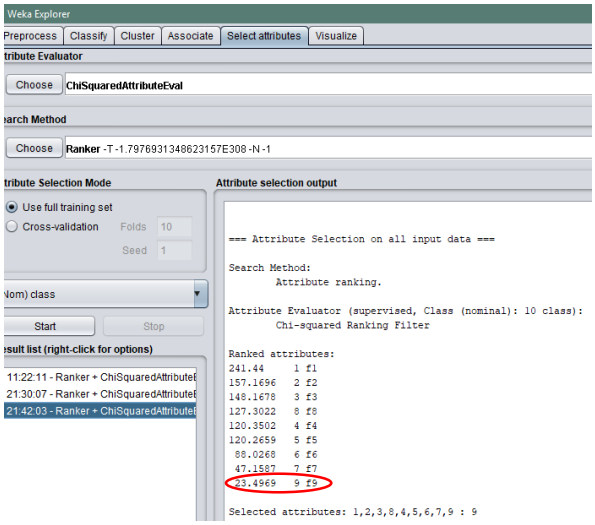


(b) Attribute Selection，請篩選出適合屬性(10%)



← Hay-test:

Filter with ChiSquaredAttributeEval and delete 0 related attribute, which is f9.



Hay-train: →

Filter with ChiSquaredAttributeEval and delete least related attribute, which is f9.

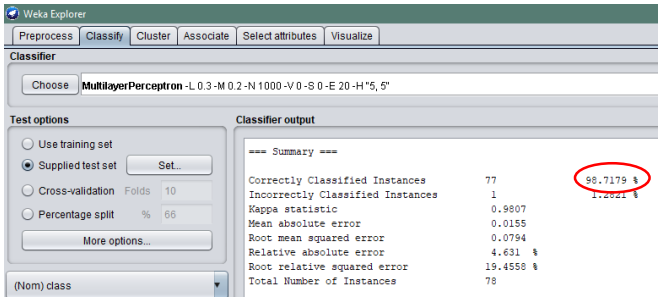
(c) 處理完上述步驟後，分別另存新檔為 hay-train-done.arff 及 haytest-done.arff(10%)

建模:附圖並詳細說明每一步驟執行原因、執行過程、執行結果。其中 Test Option 設定為 Supplied test set，並匯入 hay-test-done.arff

(d) 列出分類器結果中正確率超過 80%以上的模型，以及其參數設定(20%)

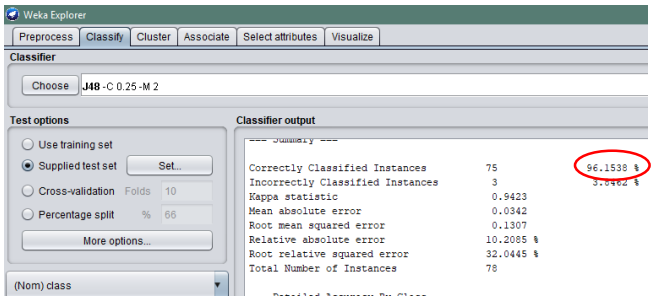
Multiple Perceptron:

Hidden layers 設爲 5,5，即爲兩層各 5 個 node，training time 設爲 1000 次，得正確率 98.7179%。



J48:

Seed 設爲 1，得正確率 96.1538%。



Naïve Bayes:

得正確率 98.7179%。