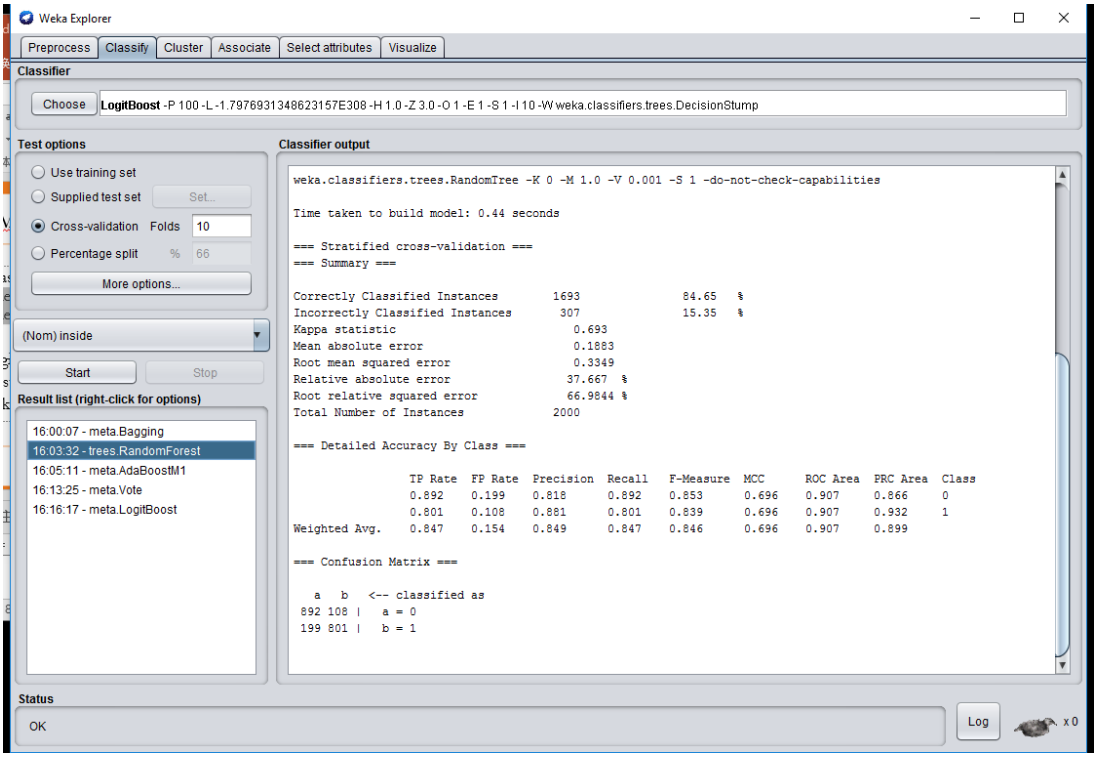
LIS590DT Assignment #5

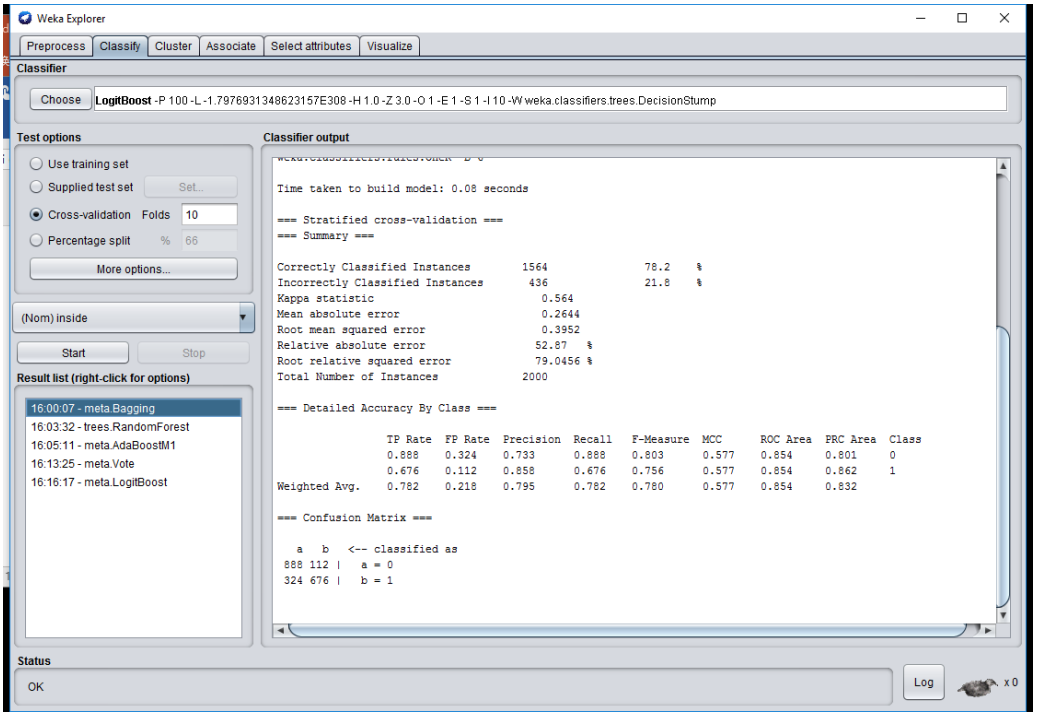
Jialu Wang

1. Circle

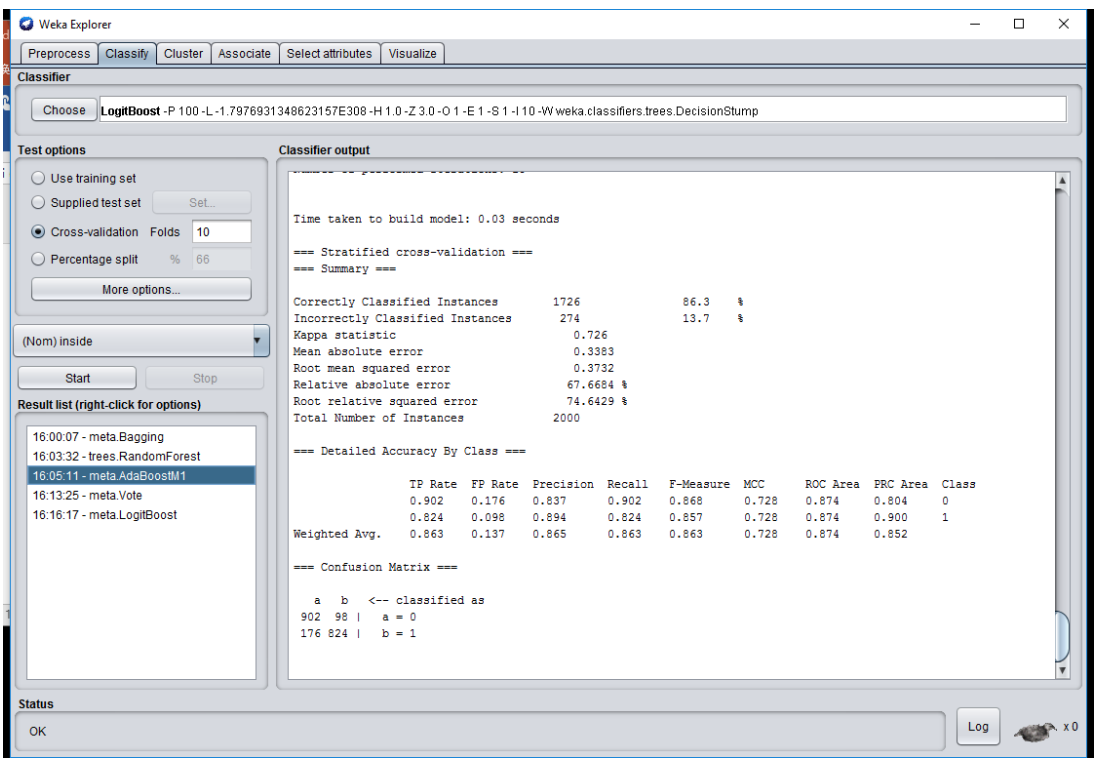
Random Forest



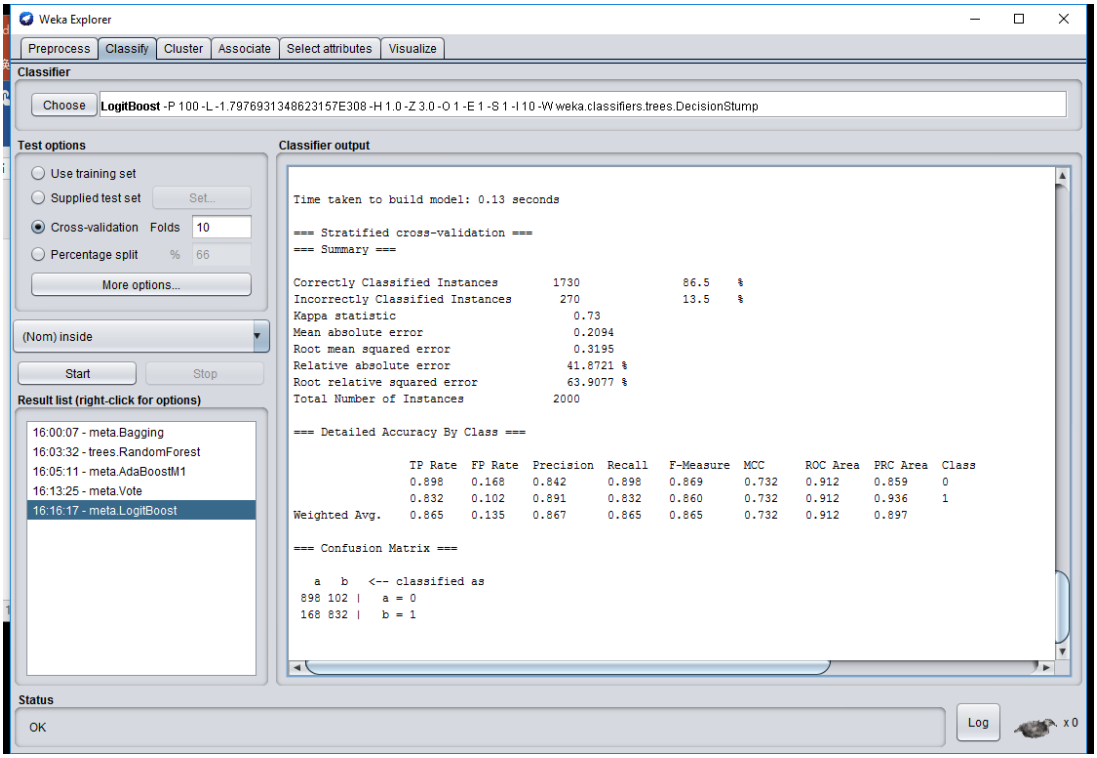
Bagging with oneR



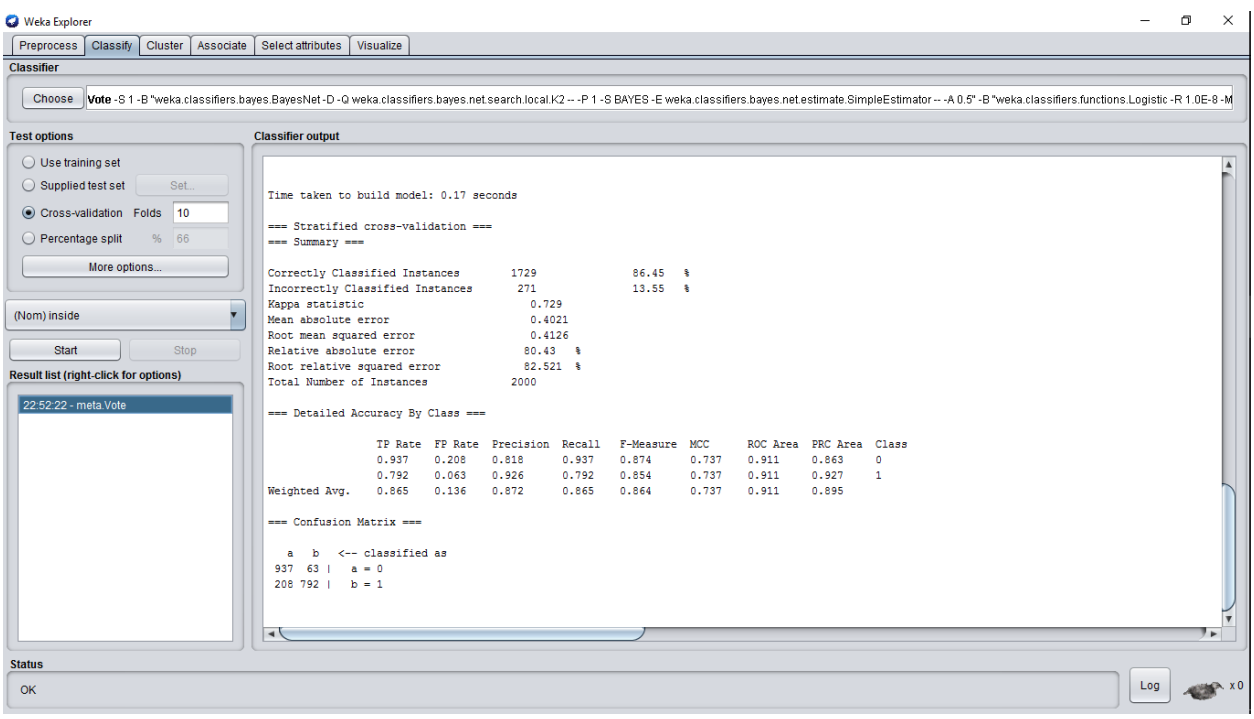
AdaBoostM1



LogitBoost



Vote

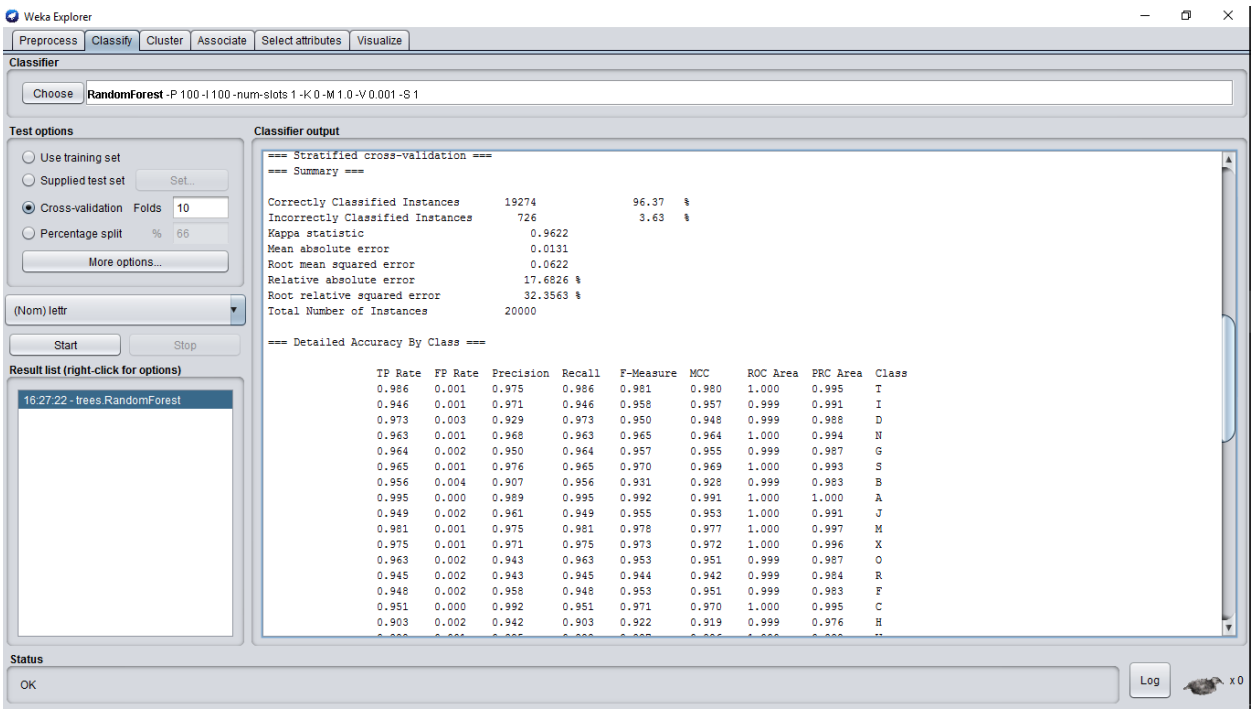



```
Vote combines the probability distributions of these base learners:
        weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.Simp
        weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4
        weka.classifiers.trees.DecisionStump
        weka.classifiers.rules.ZeroR
using the 'Average' combination rule
```

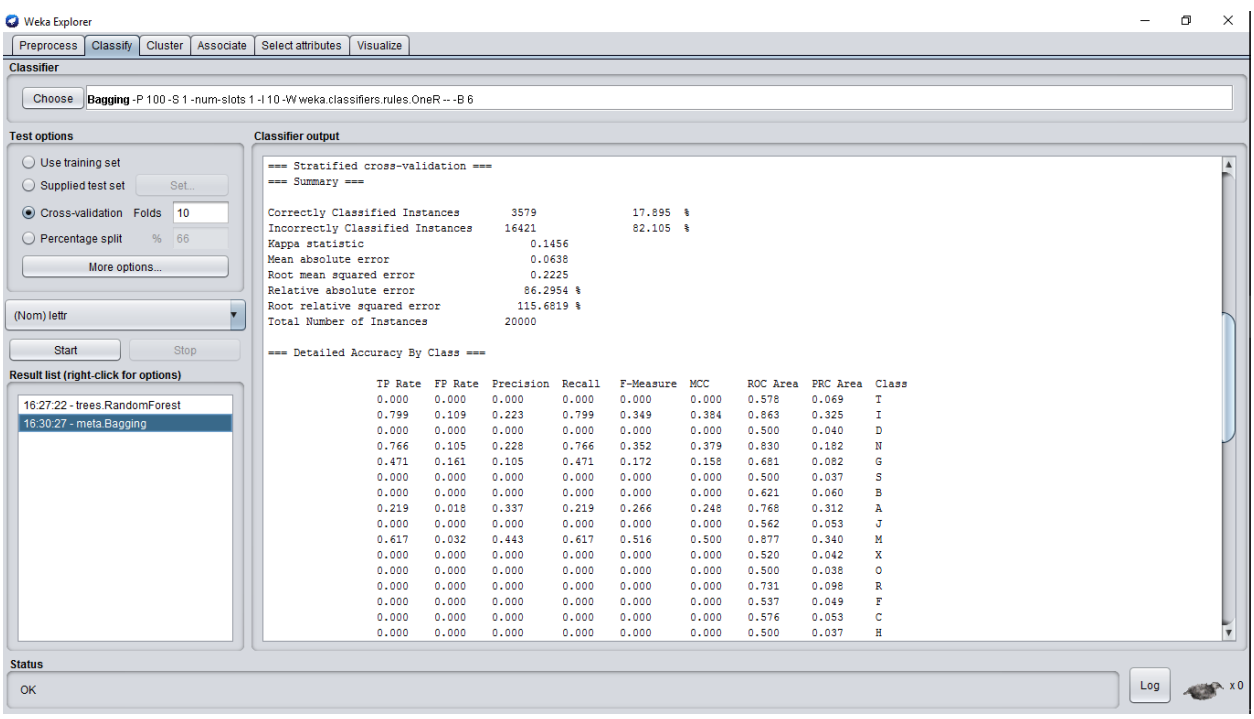
With the dataset circle, LogitBoost provide the highest accuracy rate.

2. OCR

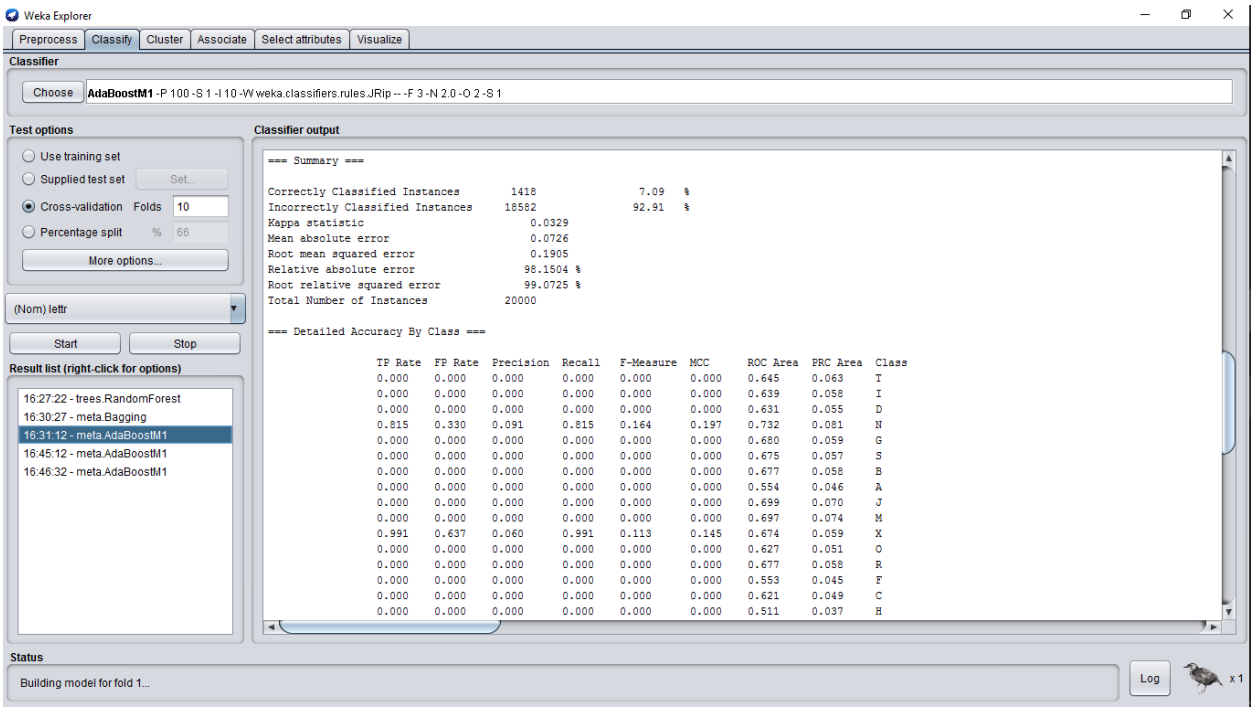
Random Forest



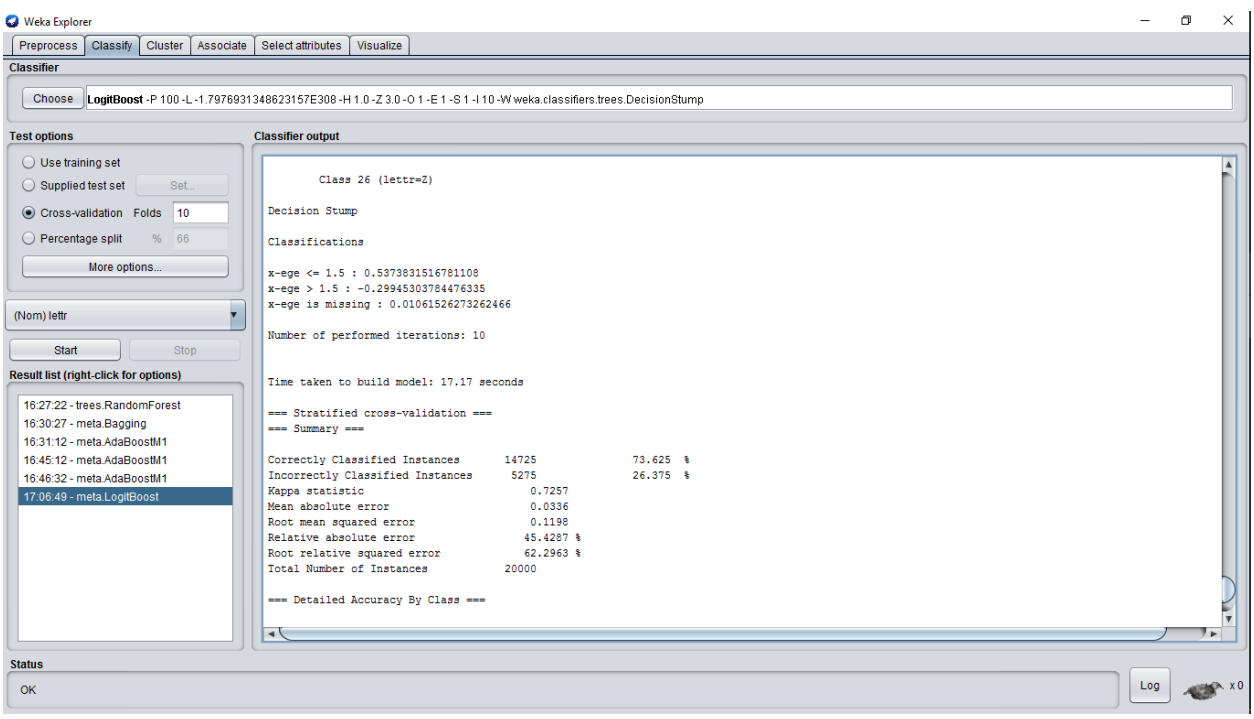
Bagging+oneR



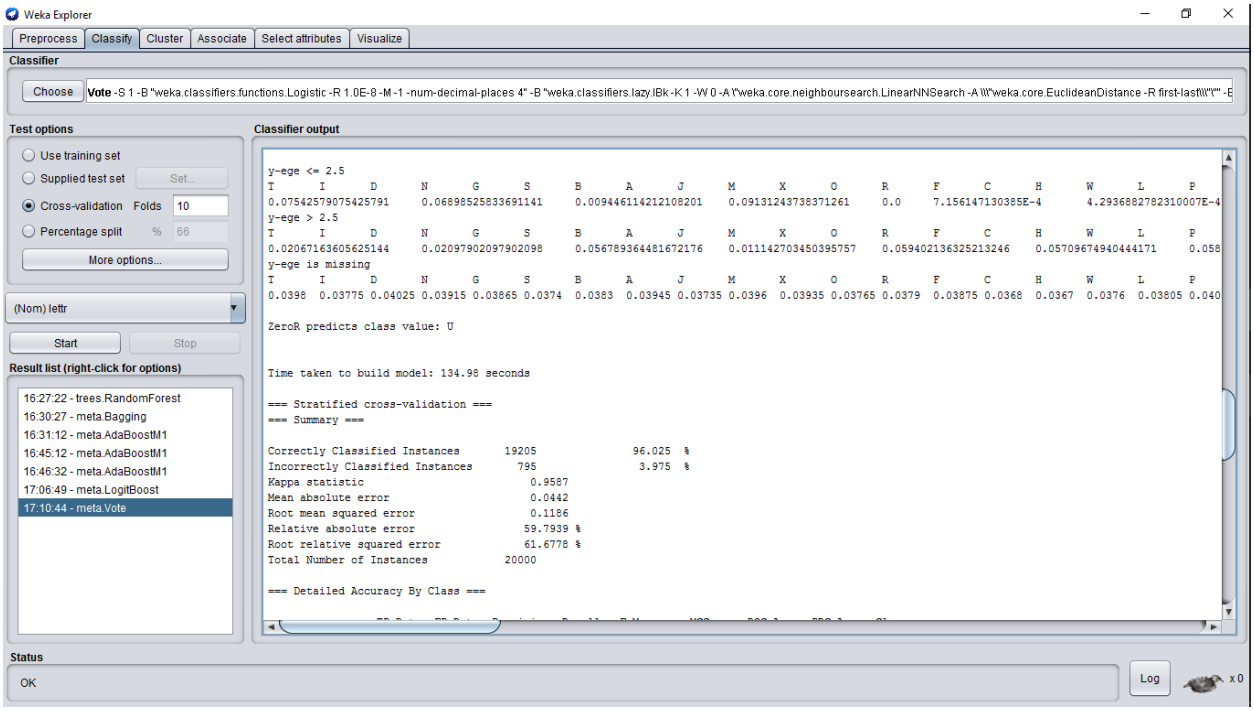
AdaBoost （classifier: decision stump）



LogitBoost



Vote

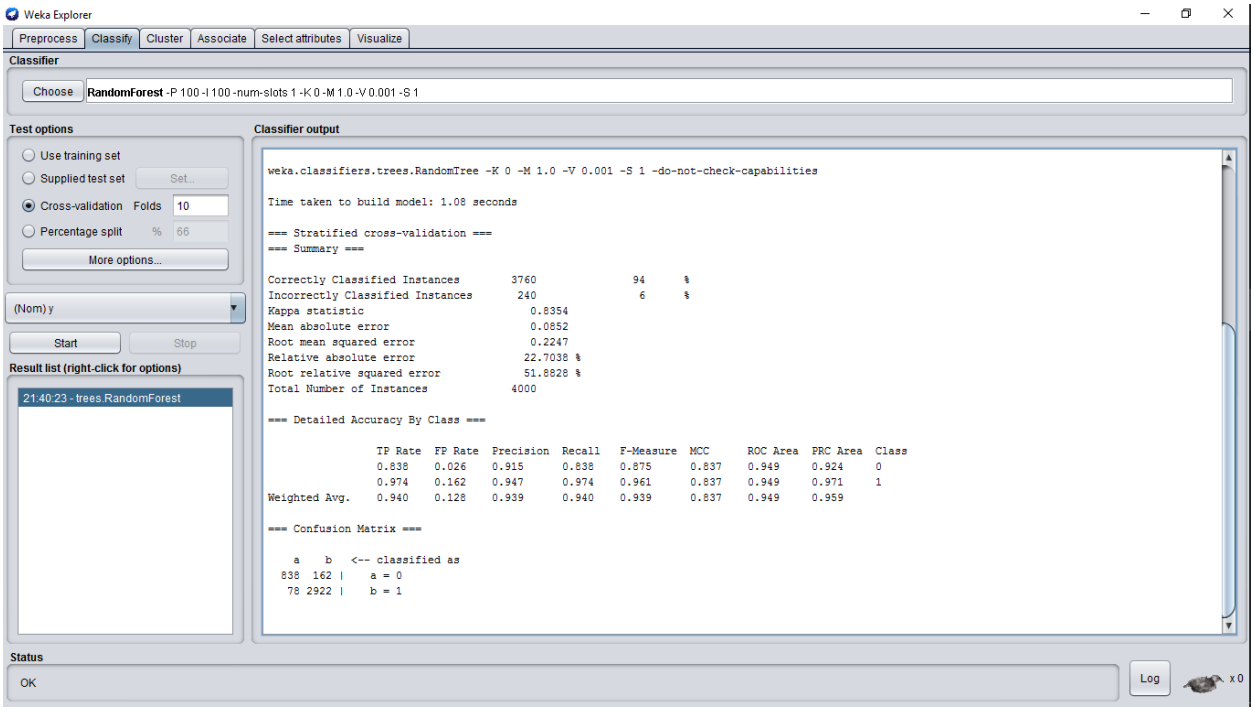



```
Vote combines the probability distributions of these base learners:
        weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4
        weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""
        weka.classifiers.trees.DecisionStump
        weka.classifiers.rules.ZeroR
using the 'Average' combination rule
```

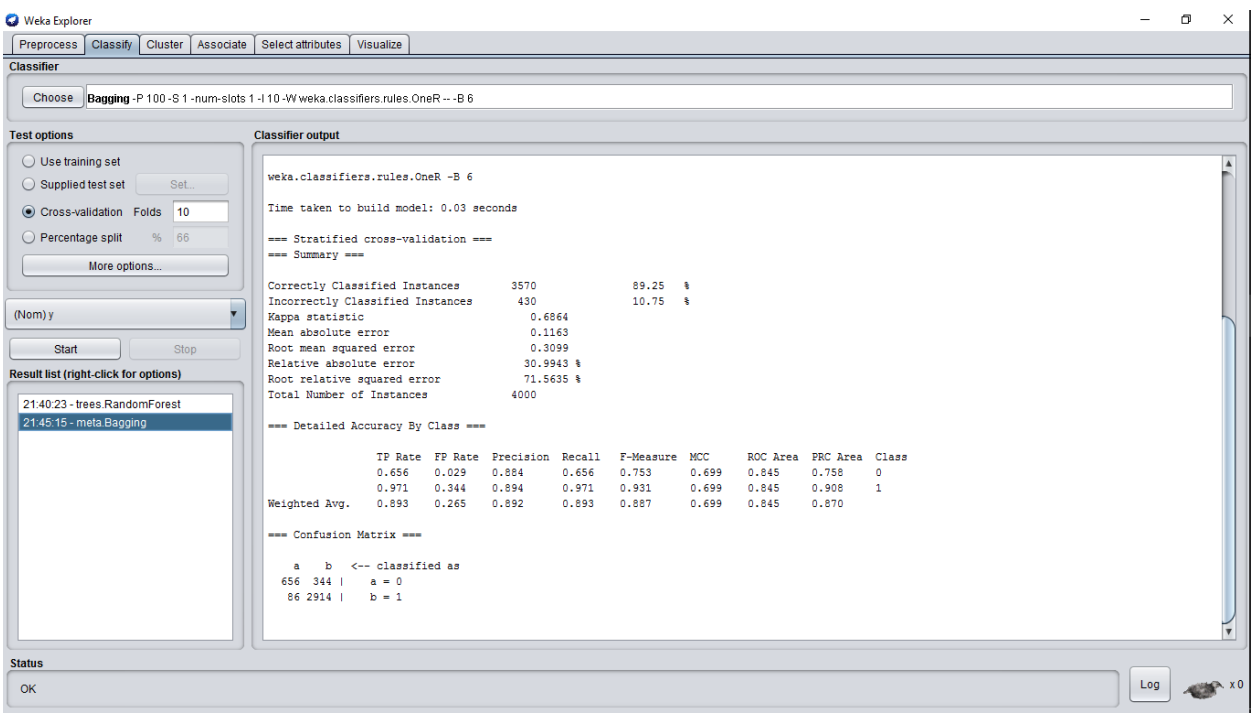
Random Forest provide the highest accuracy. Some methods has really low accuracies and long running times…

3. Eclipse

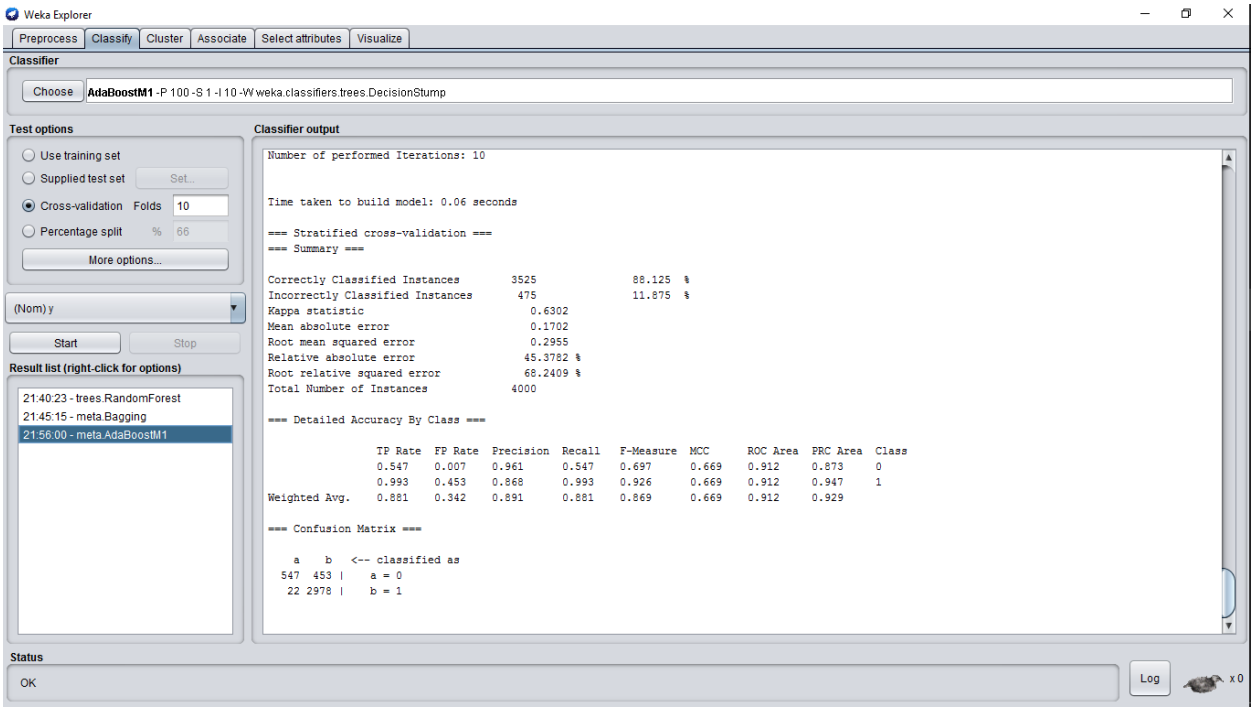
Random Forest



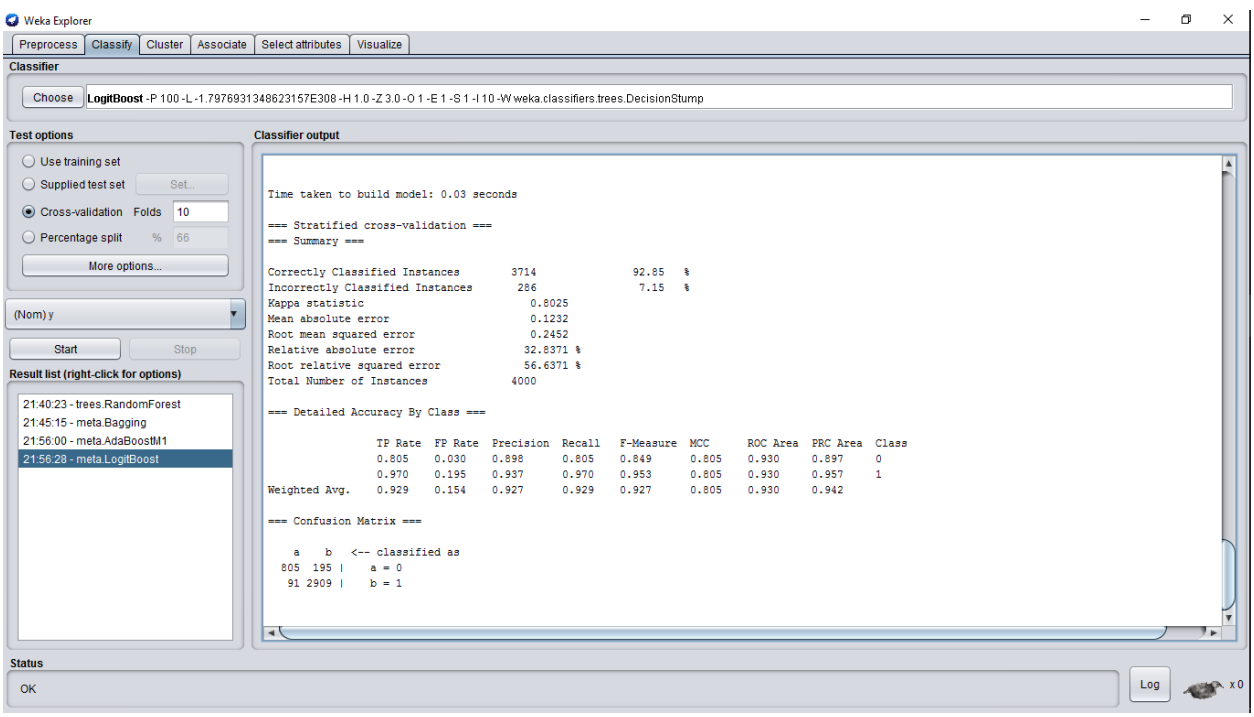
Bagging+oneR



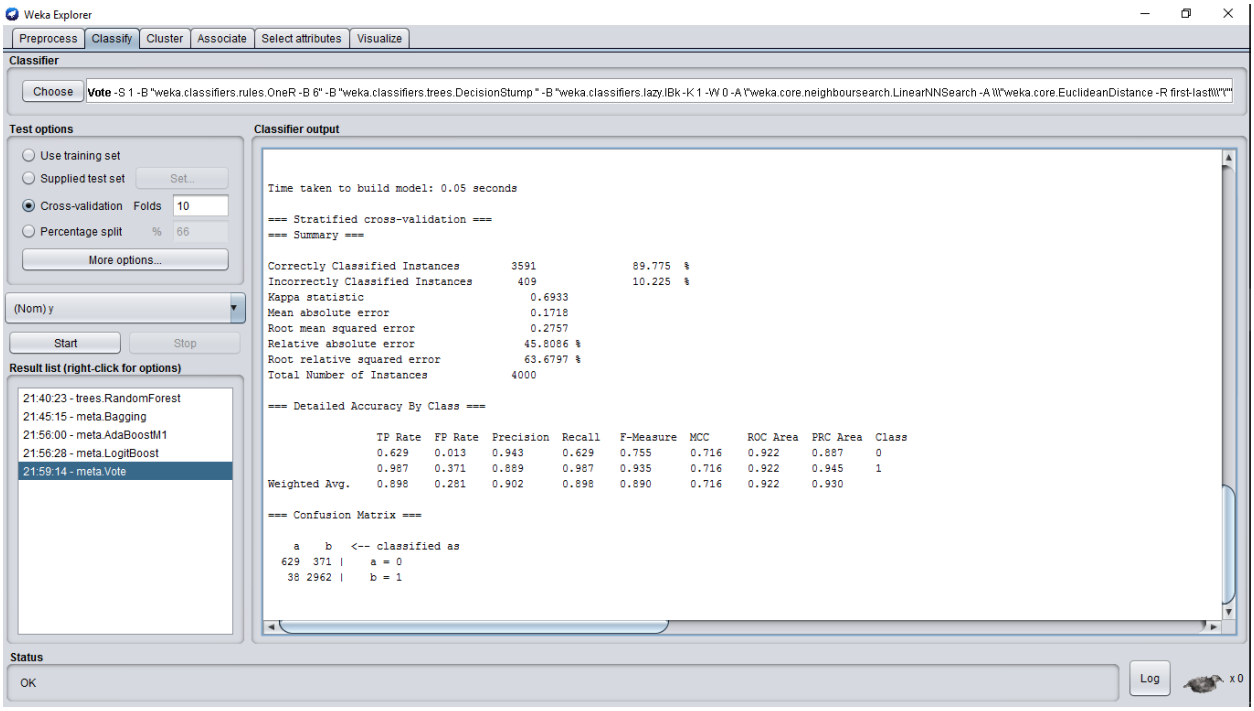
AdaBoost



LogitBoost



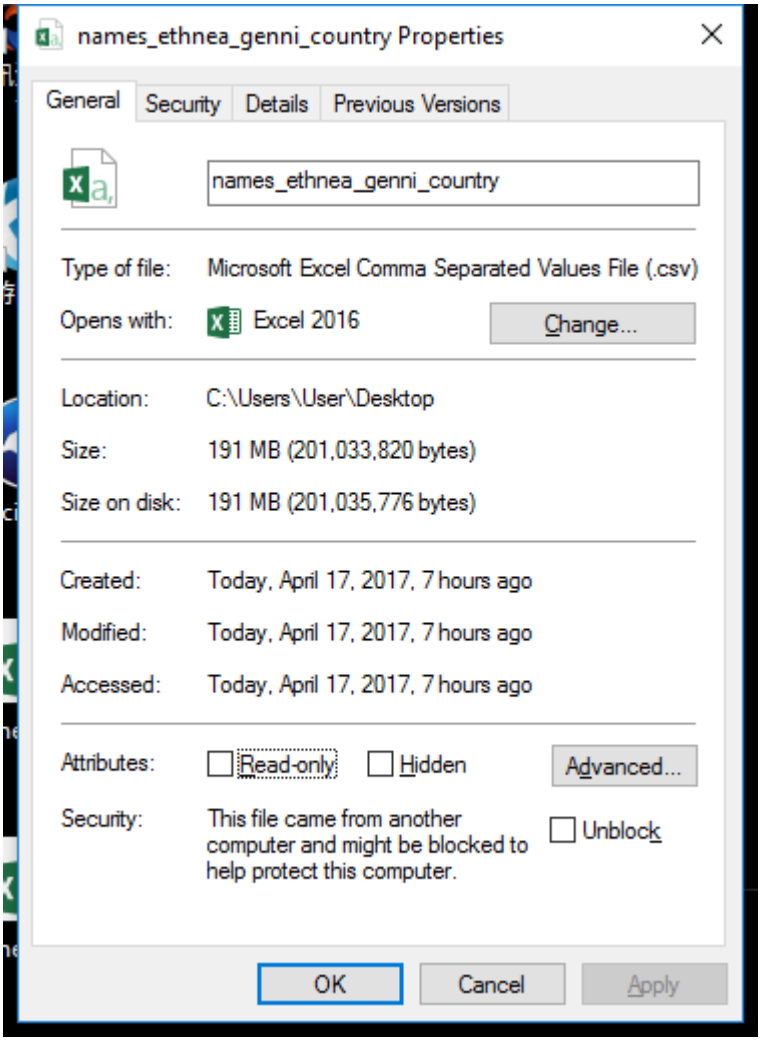
Vote

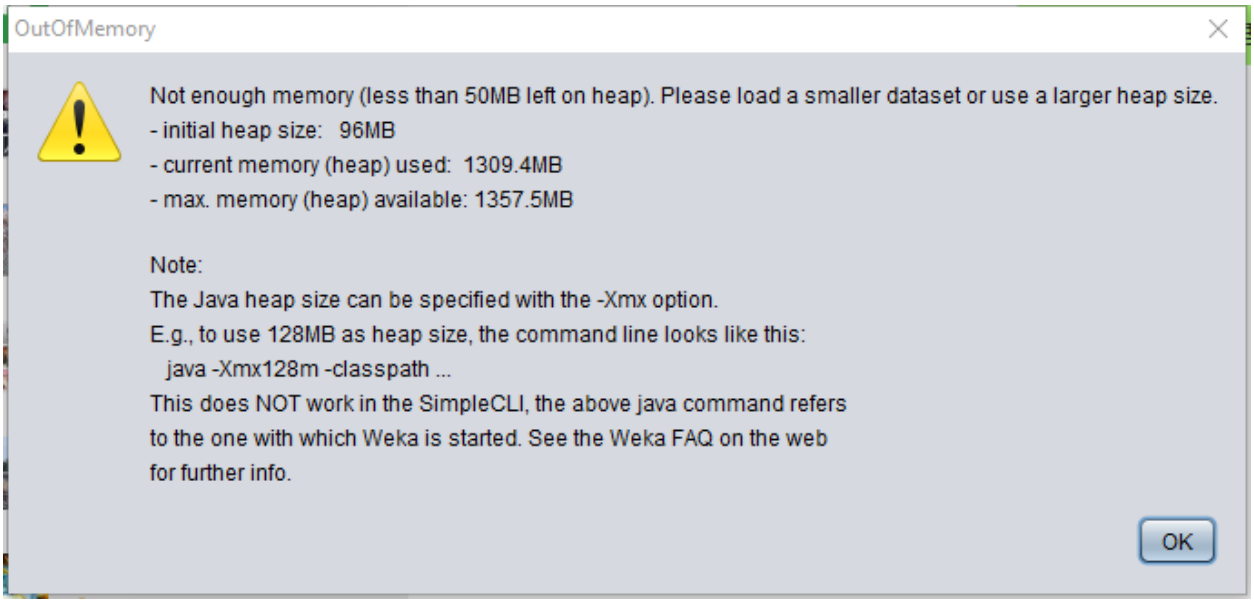



```
Vote combines the probability distributions of these base learners:
        weka.classifiers.rules.OneR -B 6
        weka.classifiers.trees.DecisionStump
        weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""
        weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4
using the 'Average' combination rule
```

Random Forest still has the highest accuracy…

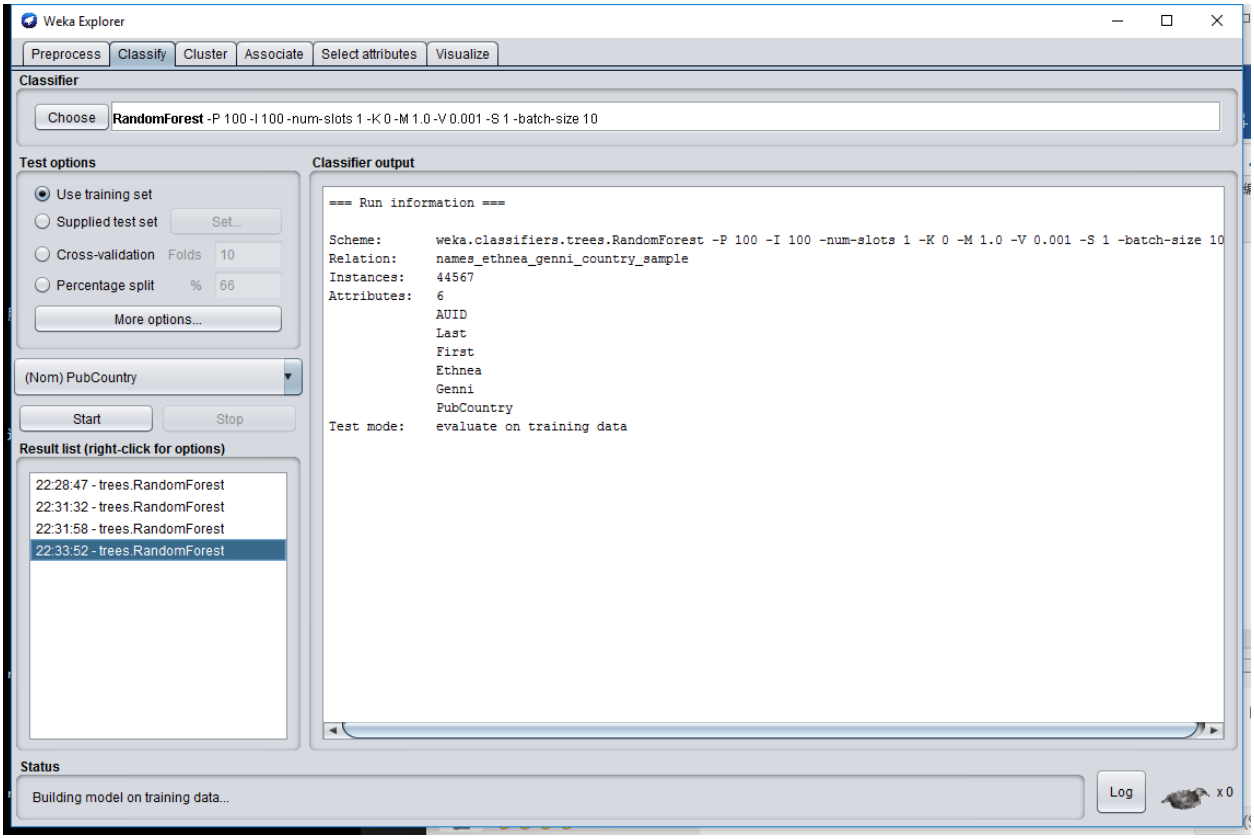
4. names_ethnea_genni_country.csv （cannot reading the file, not enough memory）



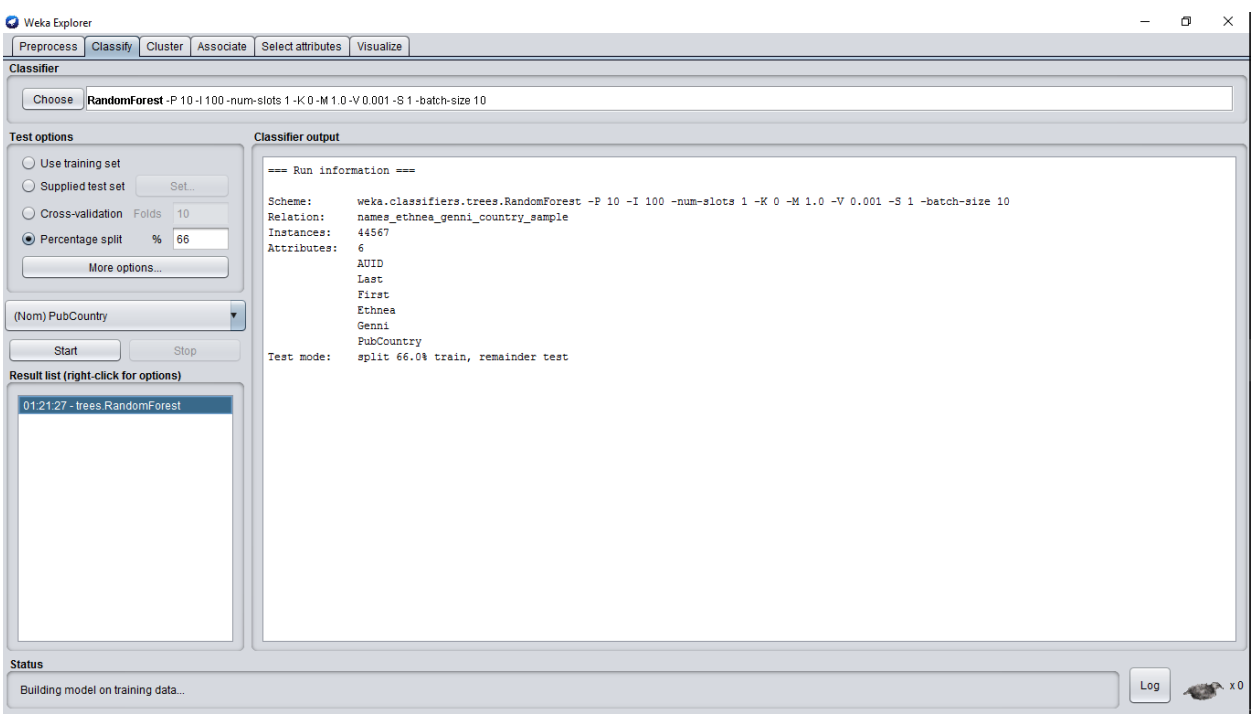


5. names_ethnea_genni_country_sample.csv

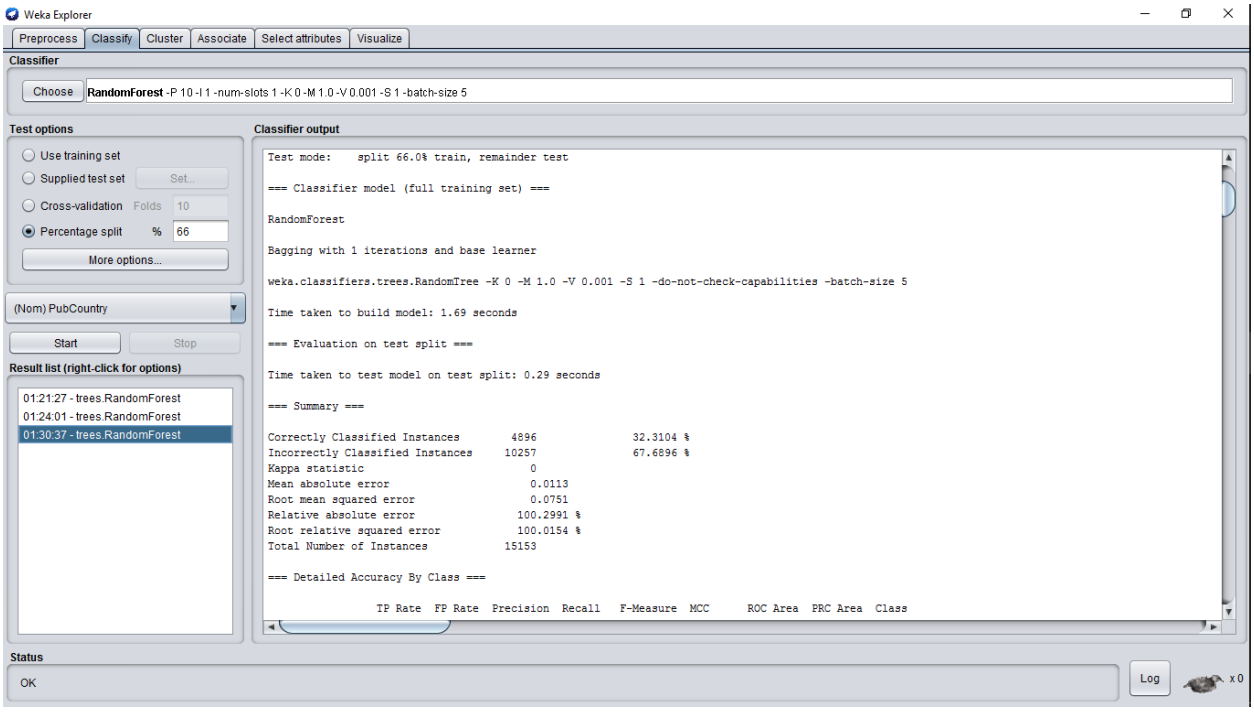
my computer can read the file but cannot run any classifiers…



I adjust the batch size to 10 and it still cannot run

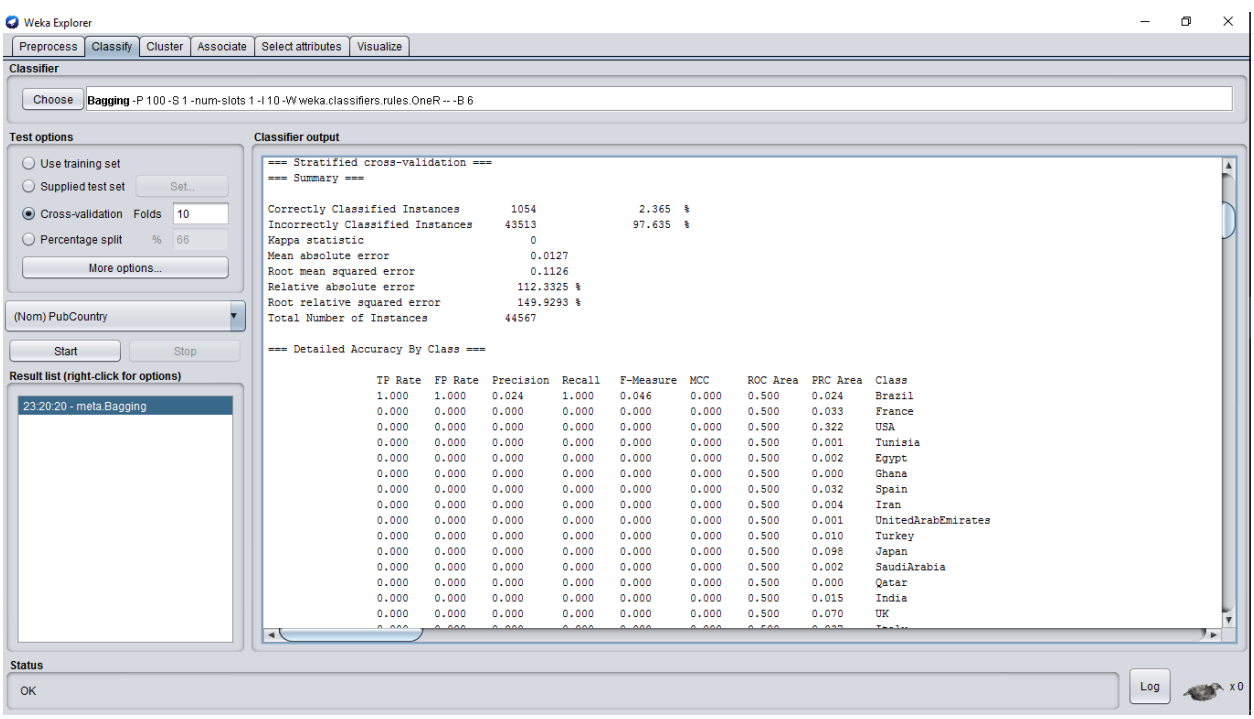


I change the iteration to 1 and it ends like this…



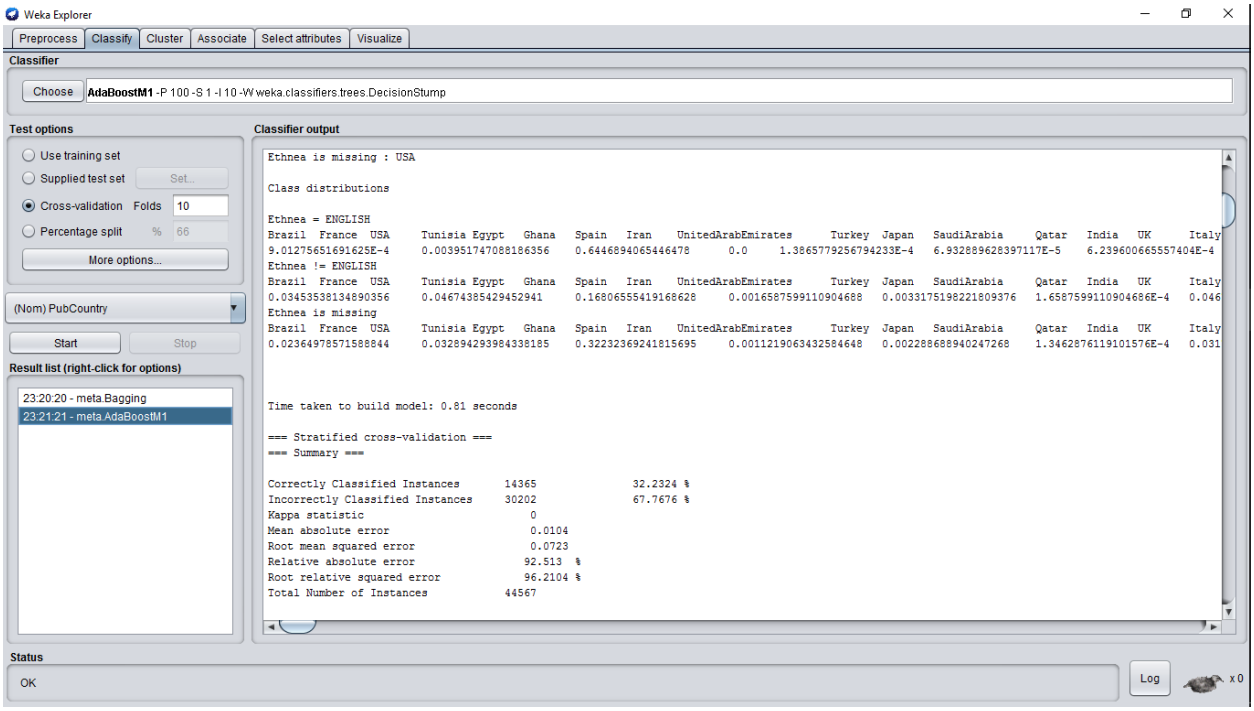
The accuracy becomes low…

Bagging+1R



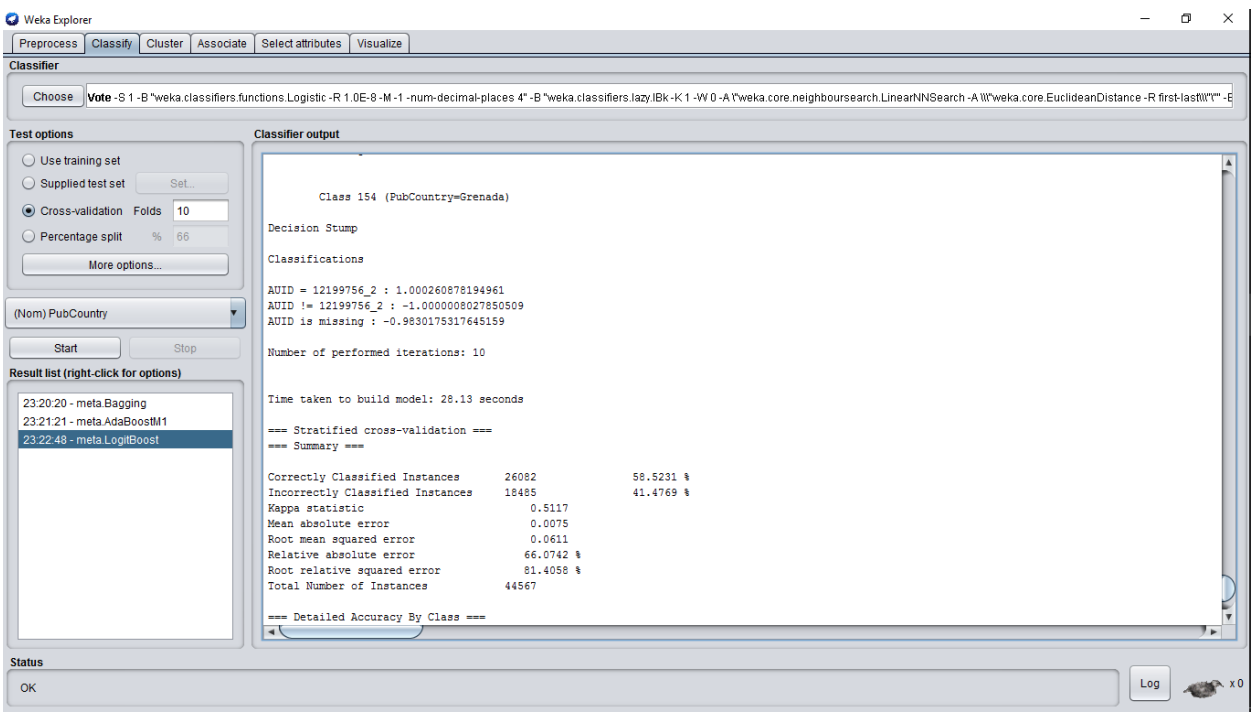
The accuracy is low…

AdaBoostM1



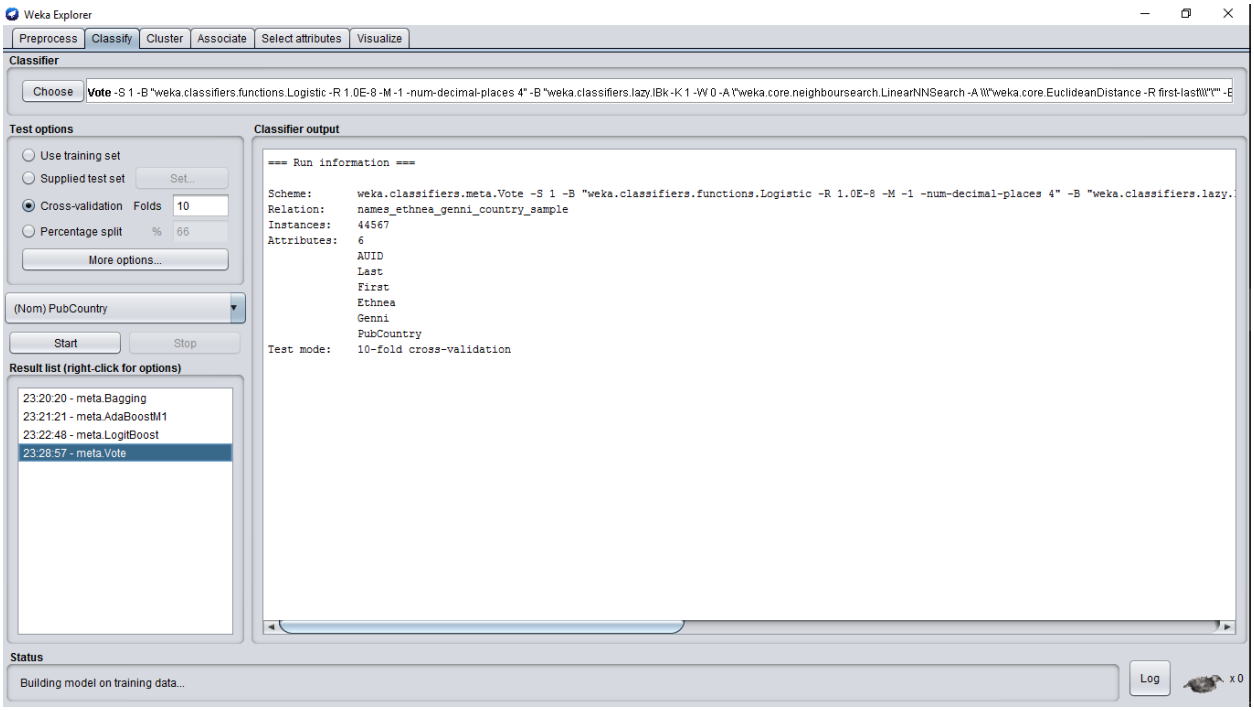
Accuracy higher but still not high enough

LogitBoost



Still not accurate.

Vote



The bird go die again…

Not all classifiers can run on the sample csv, even if they can, the accurate rate is not high as expected.