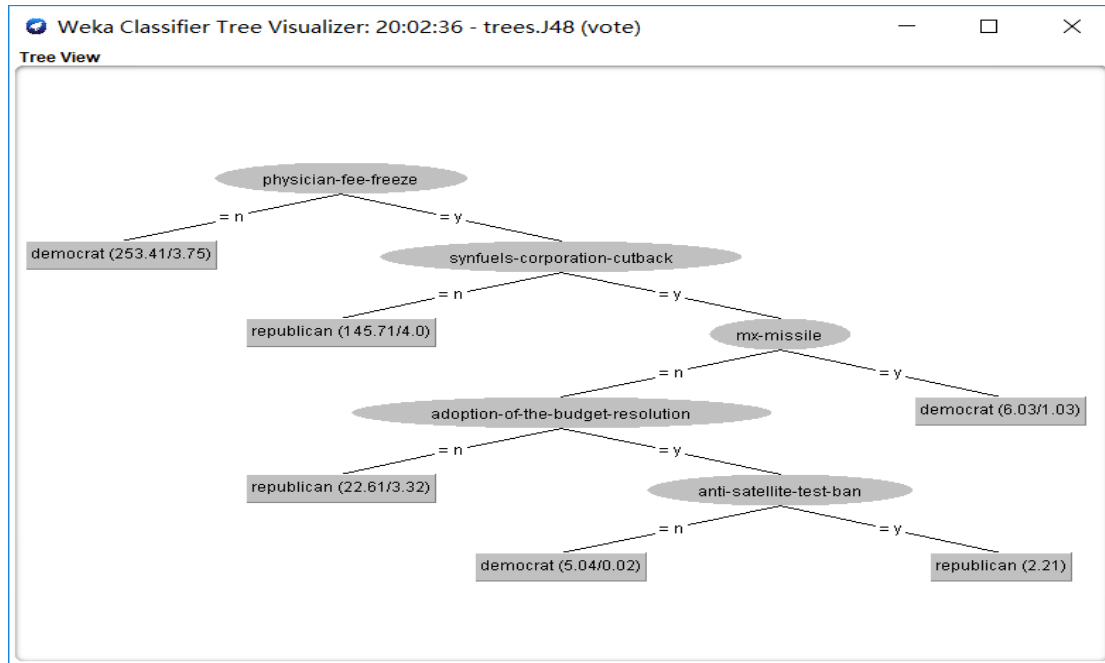


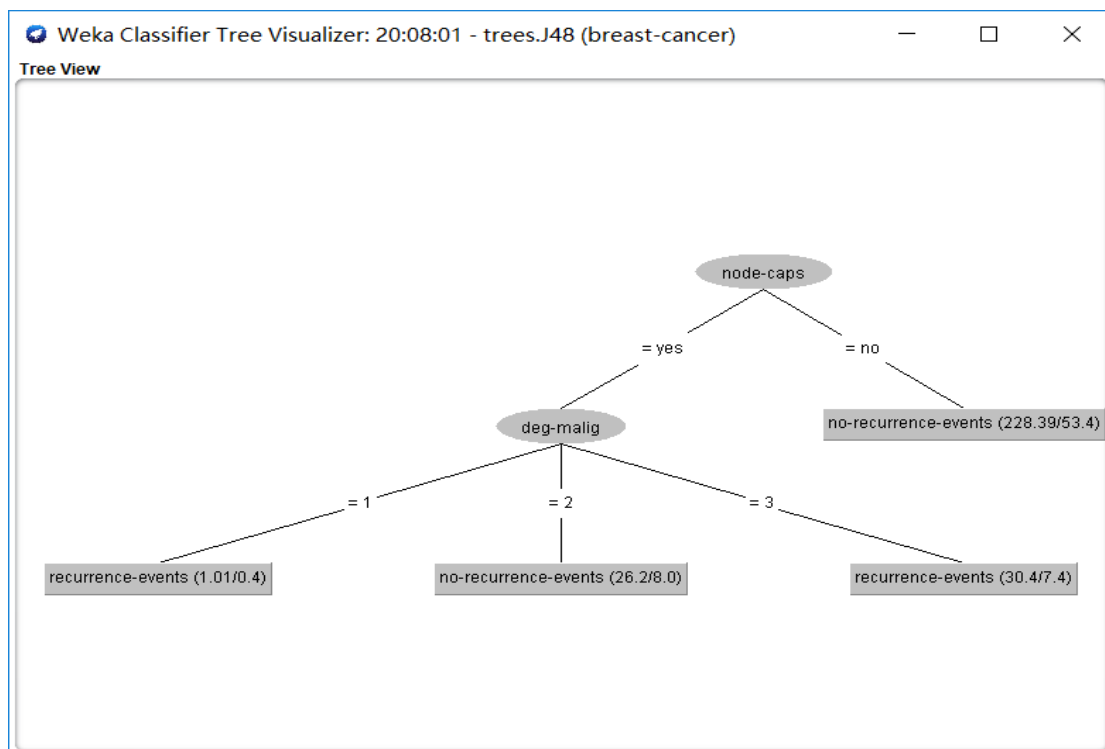
Lab2 Report

1. Learn decision tree classifier

Congressional Voting Records Data Set:



Breast Cancer Wisconsin (Original) Data Set:



2. The accuracy of the Decision Tree classifier on the Breast Cancer Wisconsin

Using 5-fold cross validation:

Classifier

Choose **J48 -C 0.25 -M 2**

Test options

☐ Use training set
☐ Supplied test set **Set...**
☒ Cross-validation Folds **5**
☐ Percentage split % **66**
More options...

(Nom) Class

Start **Stop**

Result list (right-click for options)

- 20:02:36 - trees.J48
- 20:06:37 - trees.J48
- 20:07:56 - trees.J48
- 20:08:01 - trees.J48
- 20:11:28 - trees.J48**

Classifier output

```

Size of the tree :      6

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      212           74.1259 %
Incorrectly Classified Instances    74           25.8741 %
Kappa statistic                    0.2288
Mean absolute error                 0.3726
Root mean squared error            0.4435
Relative absolute error            89.0412 %
Root relative squared error        97.0395 %
Total Number of Instances         286

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Cla
               0.960   0.776   0.745     0.960   0.839     0.287   0.582    0.728   no-
               0.224   0.040   0.704     0.224   0.339     0.287   0.582    0.444   rec
Weighted Avg.   0.741   0.558   0.733     0.741   0.691     0.287   0.582    0.643

=== Confusion Matrix ===
  a  b  <-- classified as
+---+
103  81  a = 0, b = 1

```

Status

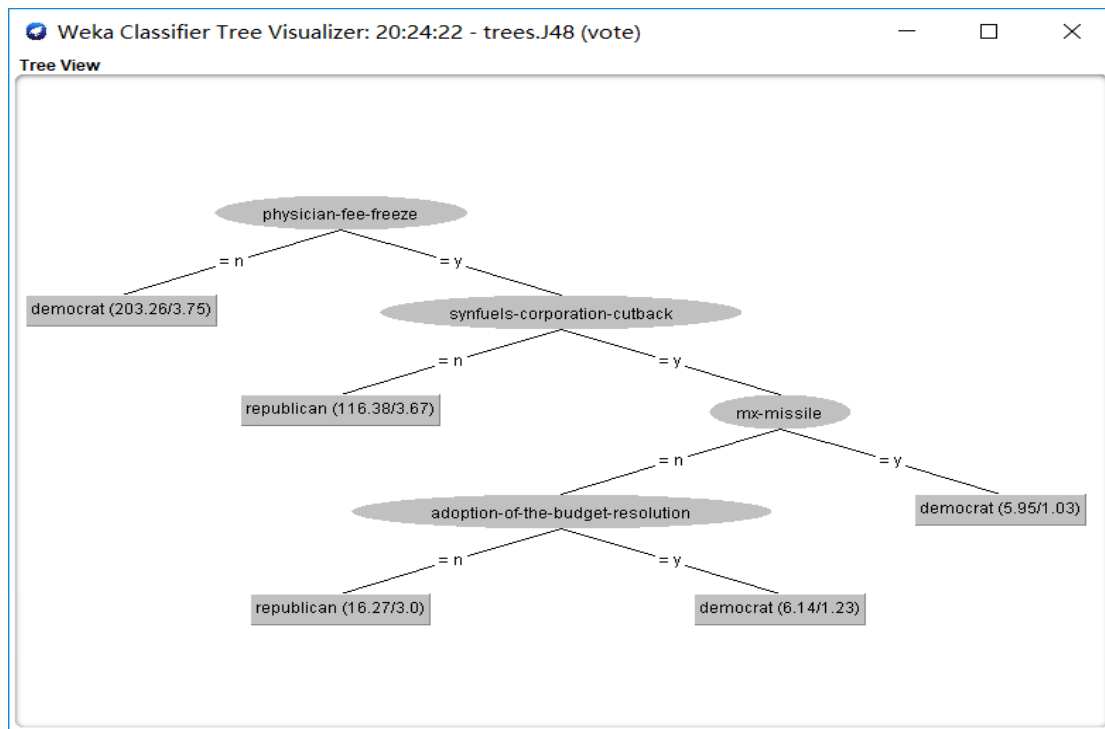
OK **Log**

Accuracy: 74.1259%

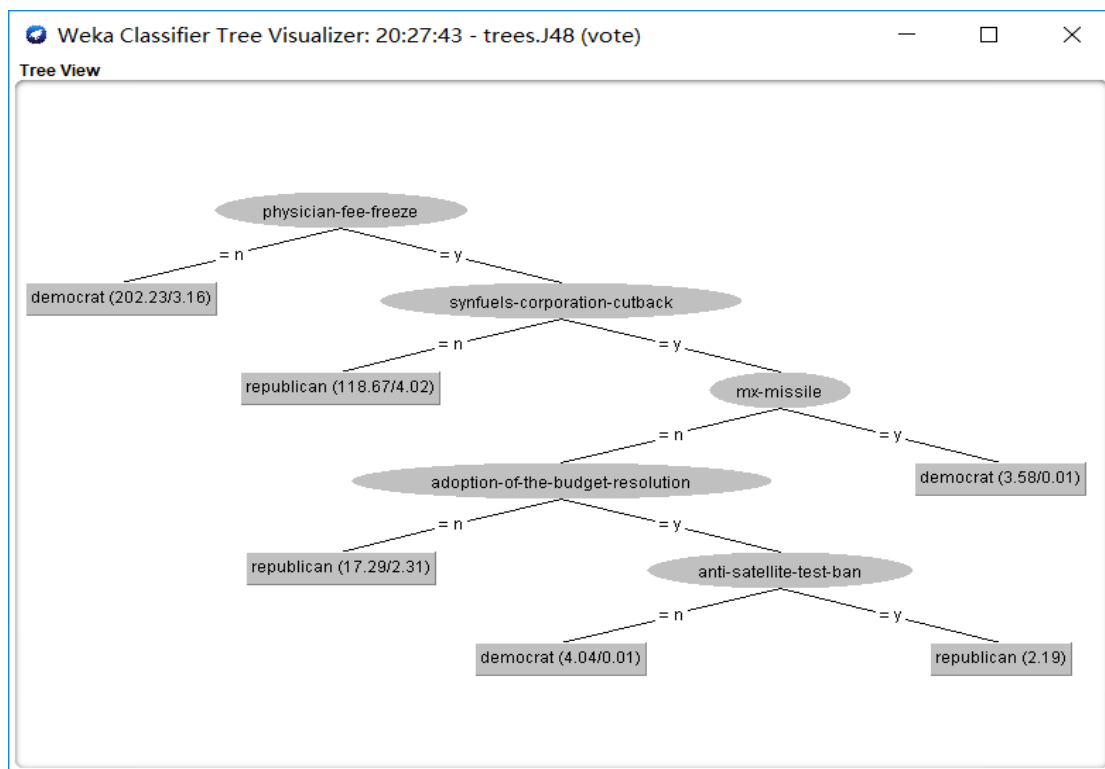
95% confidence interval: $[\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}] = [0.6921, 0.7935]$

3. Perform the 5-fold cross-validation experiments with the Congressional Voting Records Data

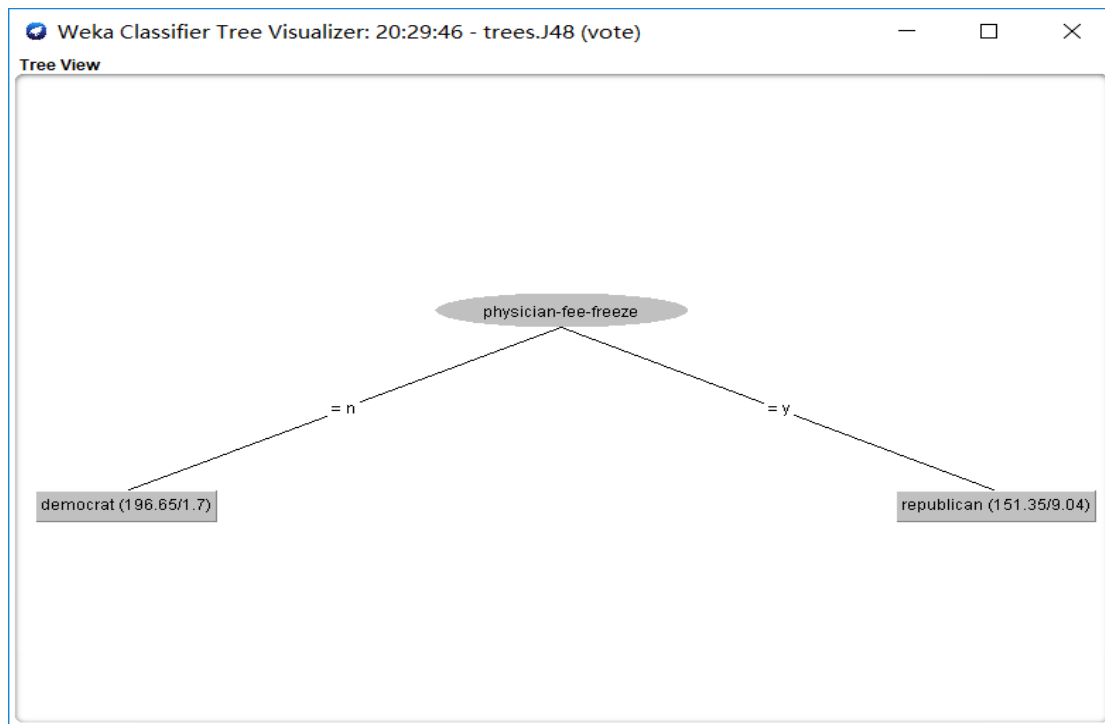
Test Set1: Accuracy: 98.8506%



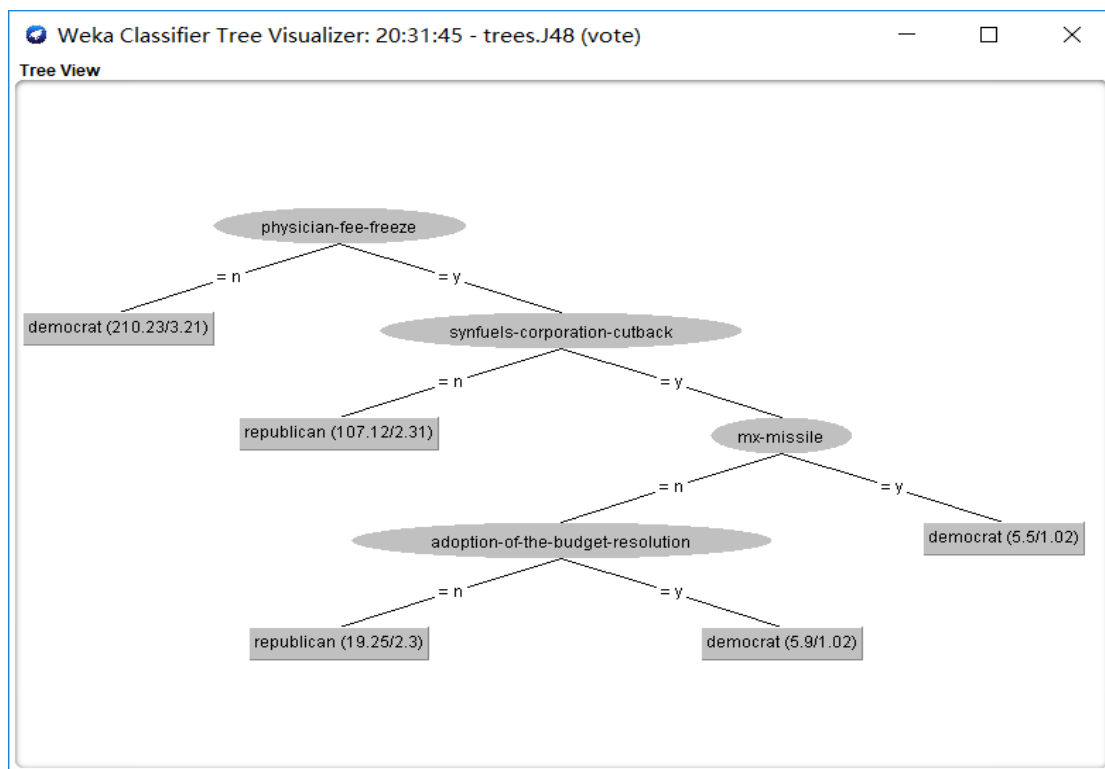
Test Set2: Accuracy: 96.5517%



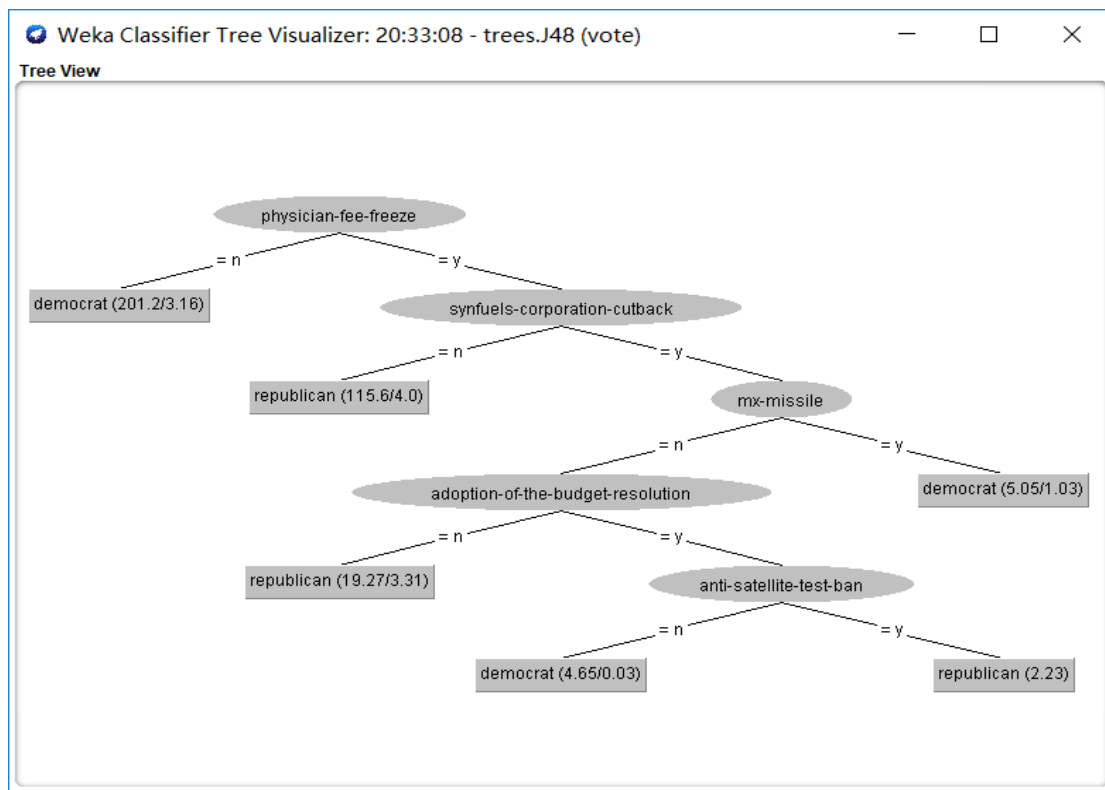
Test Set3: Accuracy: 88.5057%



Test Set4: Accuracy: 95.4023%



Test Set5: Accuracy: 98.8506%



They are different with each other and with the tree constructed using all the data!

Average: (98.8506% + 96.5517% + 88.5057% + 95.4023% + 98.8506%) / 5

= 95.6322%

5-fold cross-validation: 96.5517%

95% confidence interval: $[\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}] = [0.9367, 0.9753]$