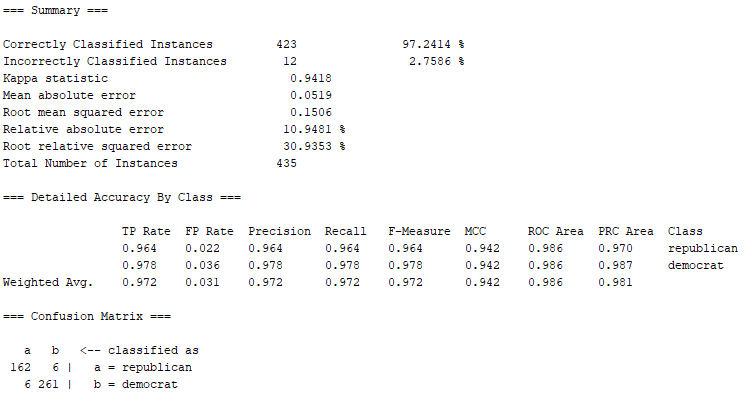
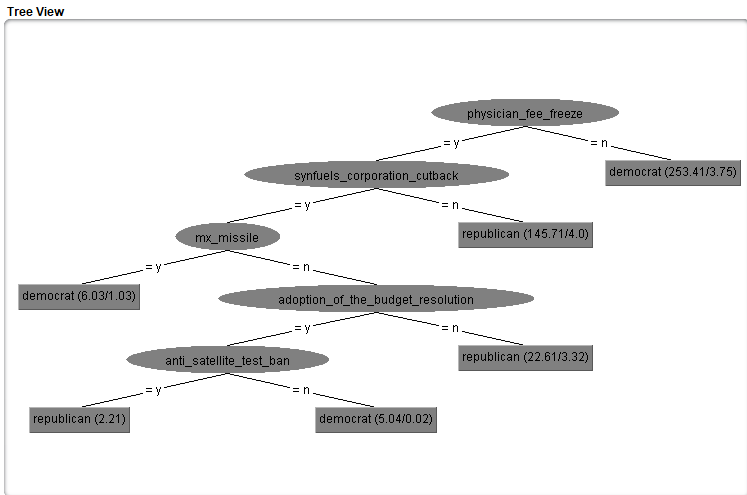
1. **Decision Tree classifier (J48 Classifier):**
   1. **House Votes Dataset:**

Result:

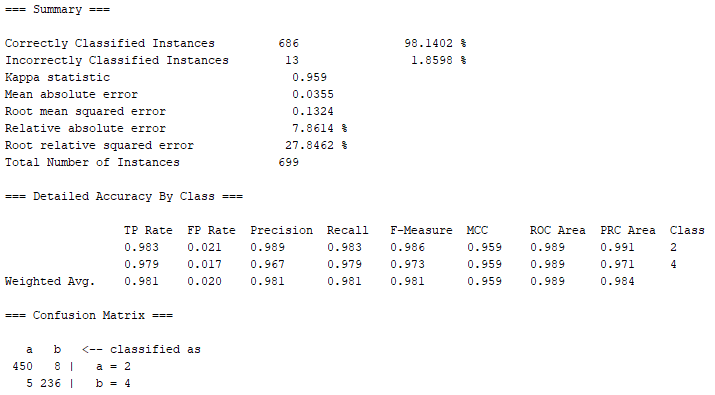


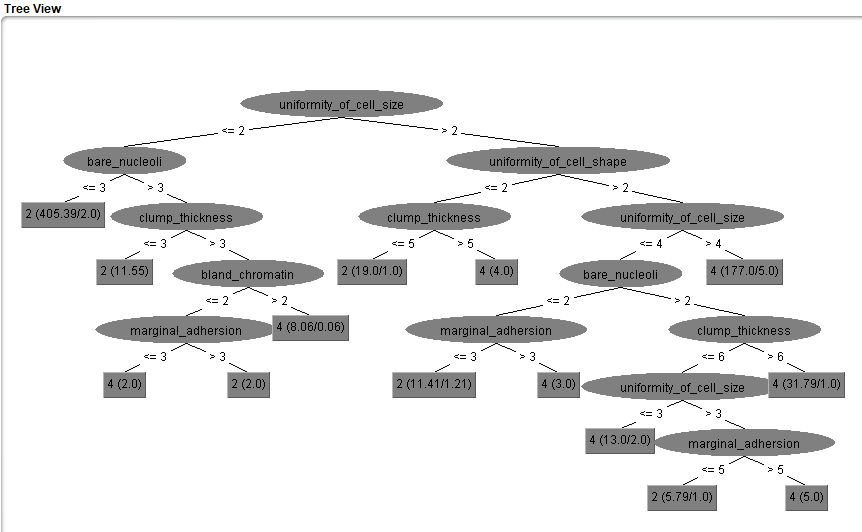


One interesting fact is if a person is against physician fee freeze, then he is a democrat.

* 1. **Breast Cancer Dataset:**

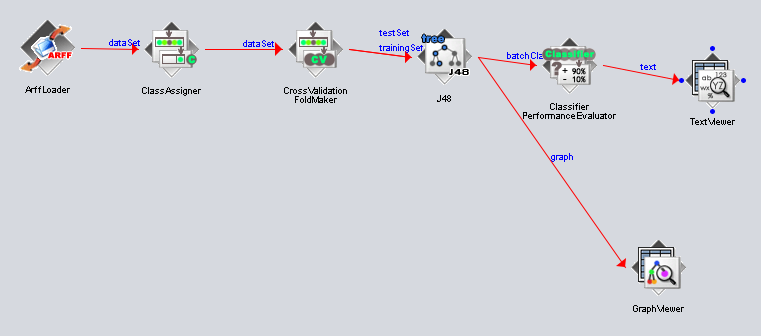
Result:

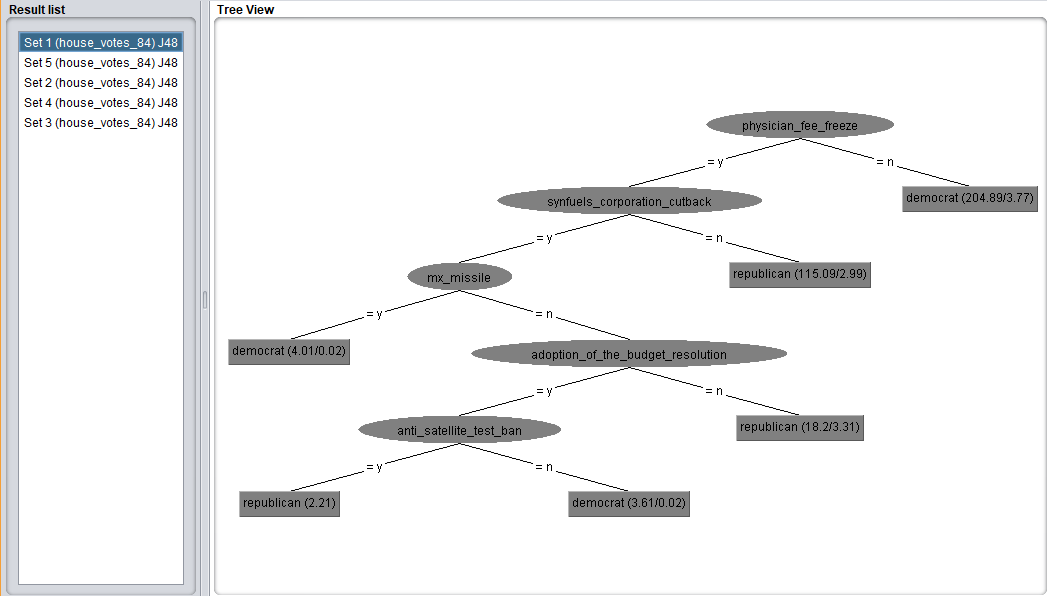




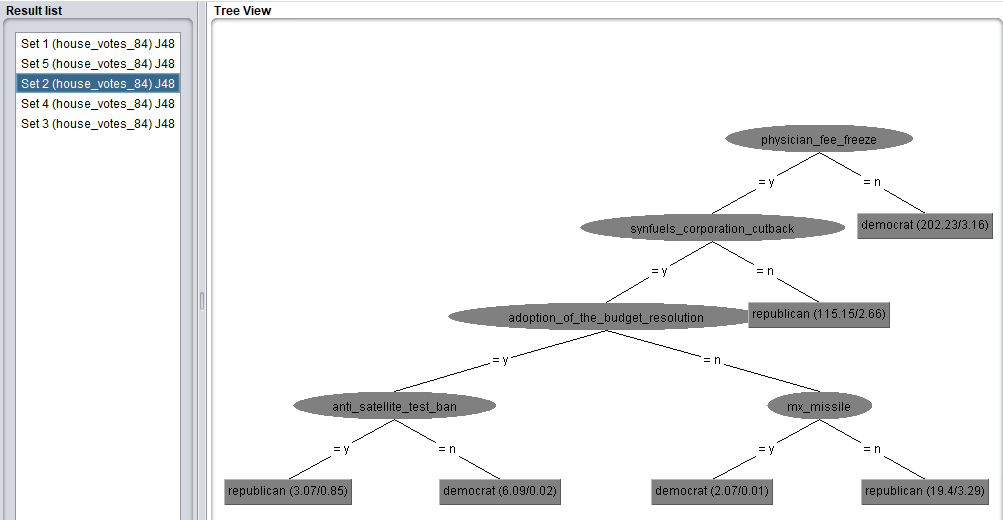
One interesting fact is that if the uniformity of cell size is > 4 then the cancer is class 4 cancer.

1. **5 Fold Cross Validation – House Votes**

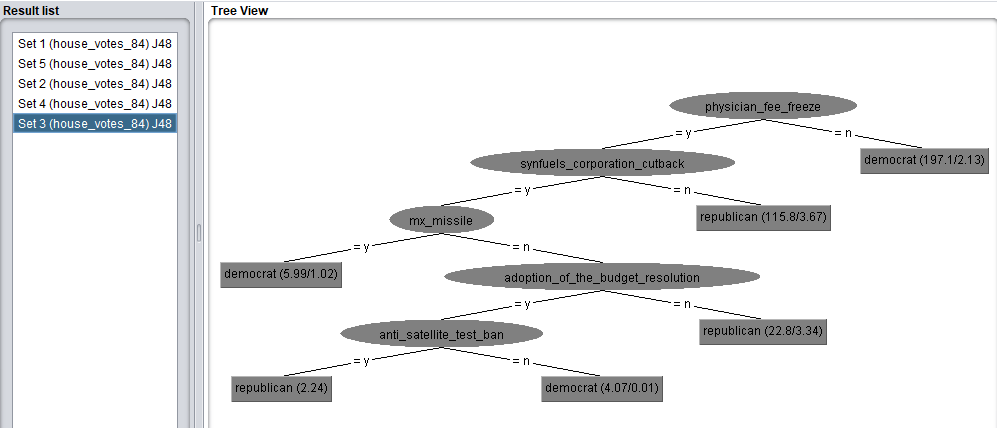
****

**Tree1: **

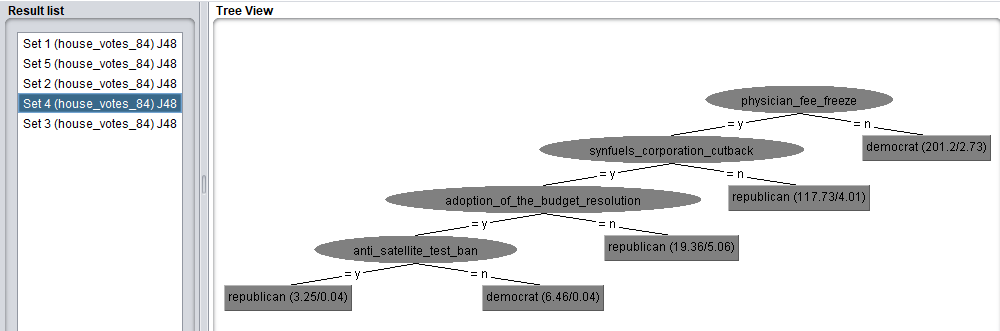
**Tree2:**

****

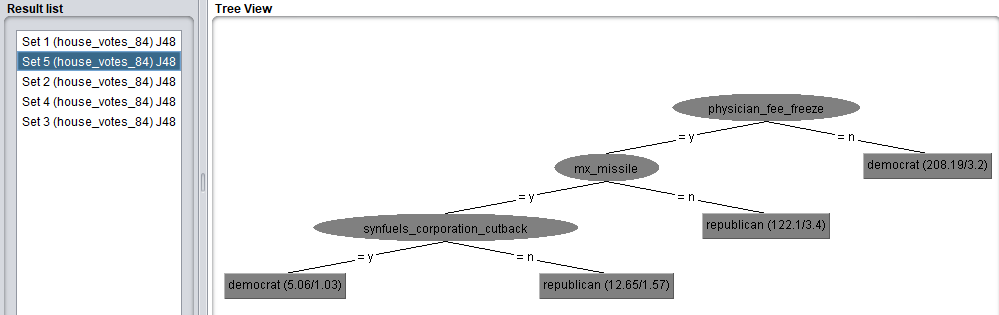
**Tree3:**

****

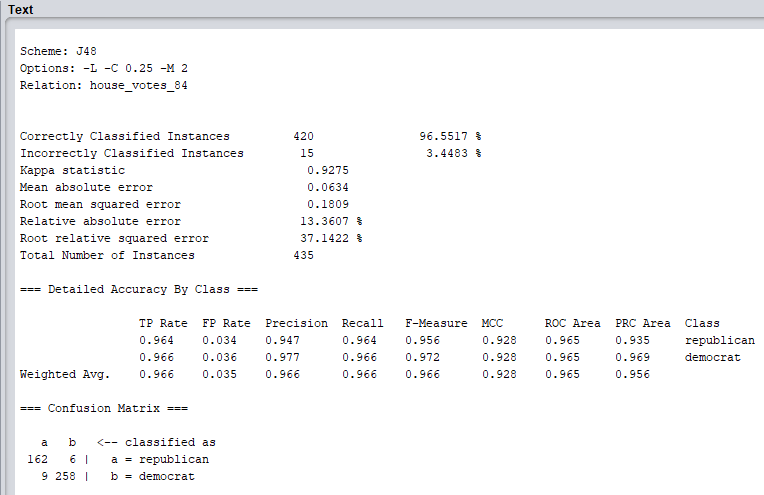
**Tree4:**

****

**Tree5:**

****

**Result:**

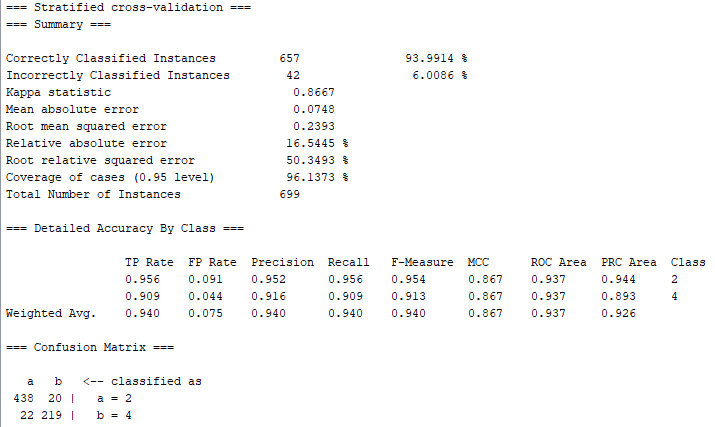
****

Accuracy – 96.55%

Confidence Interval – [0.0175, 0.0515]

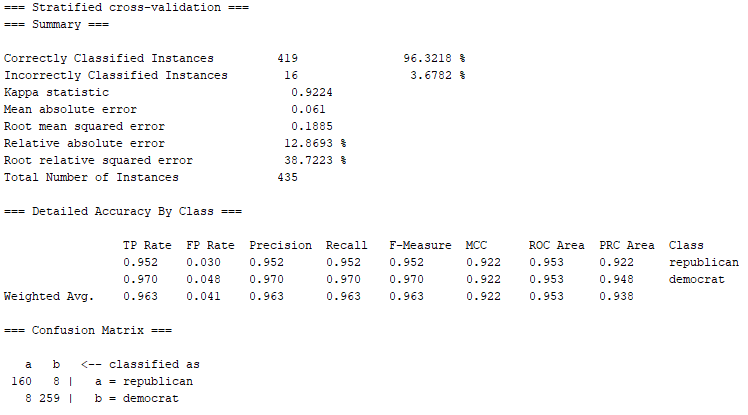
As you can see, the 5 trees differ with each other except tree 1 and 3. Also tree 1 & 3 is same as the tree obtained before.

1. **5-Fold Cross Validation – Breast Cancer Dataset:**



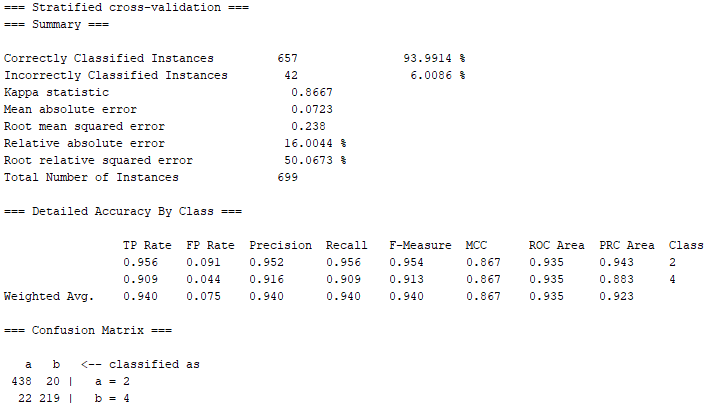
* Accuracy using 5-Fold Validation: 93.99%
* 95% Confidence Interval = [0.0425, 0.0777]

1. **Preprocess using Missing Data Filter:**
   1. House votes dataset:



The accuracy reduced from 96.55 to 96.32

* 1. Breast Cancer Dataset:



Accuracy remained the same.

**References:**

1. Discussed with Nitesh Gupta