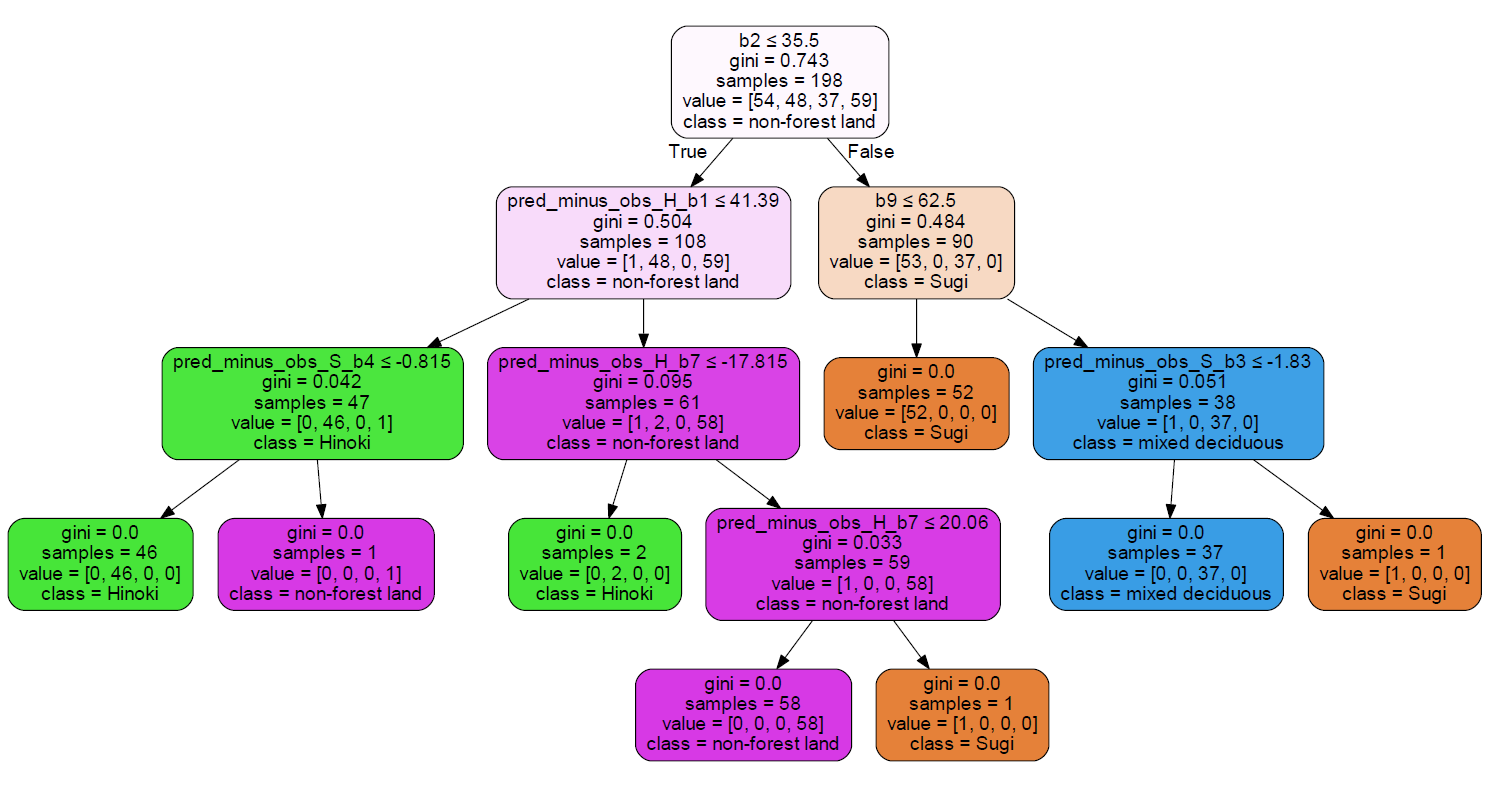
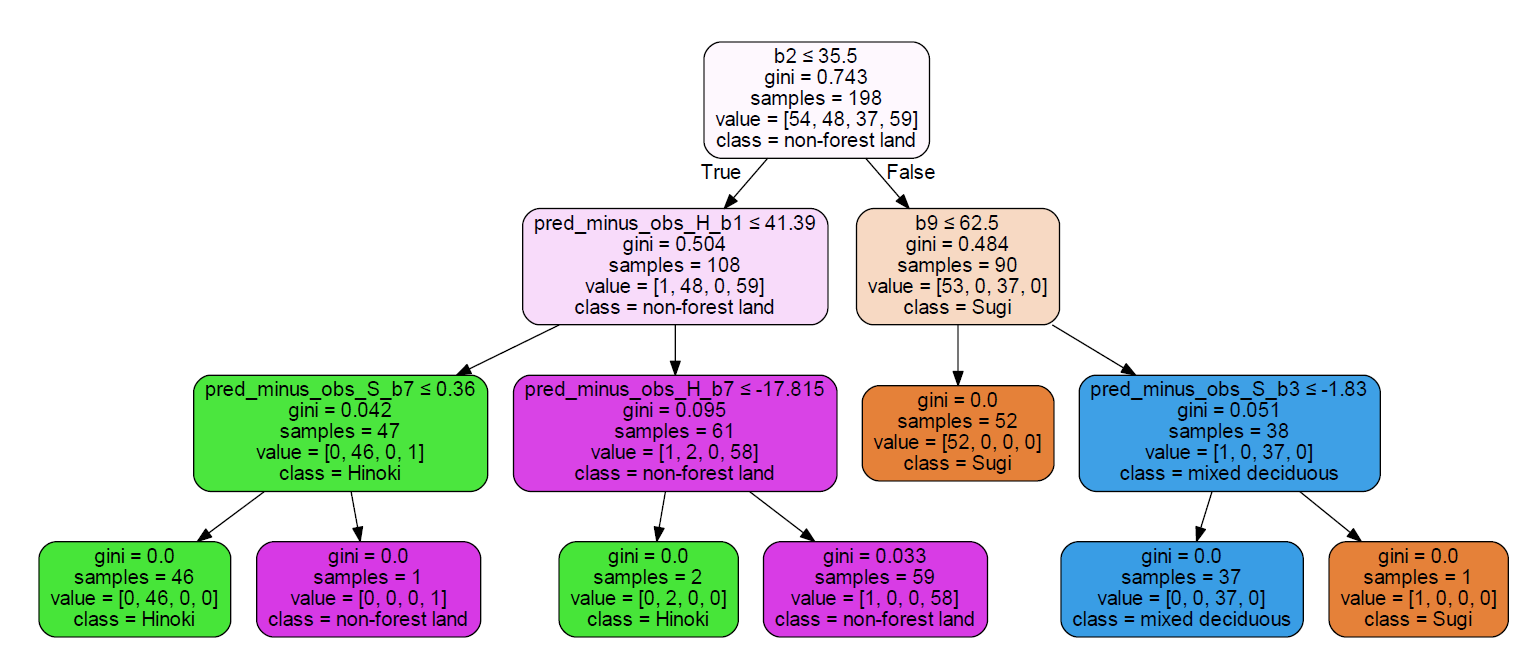
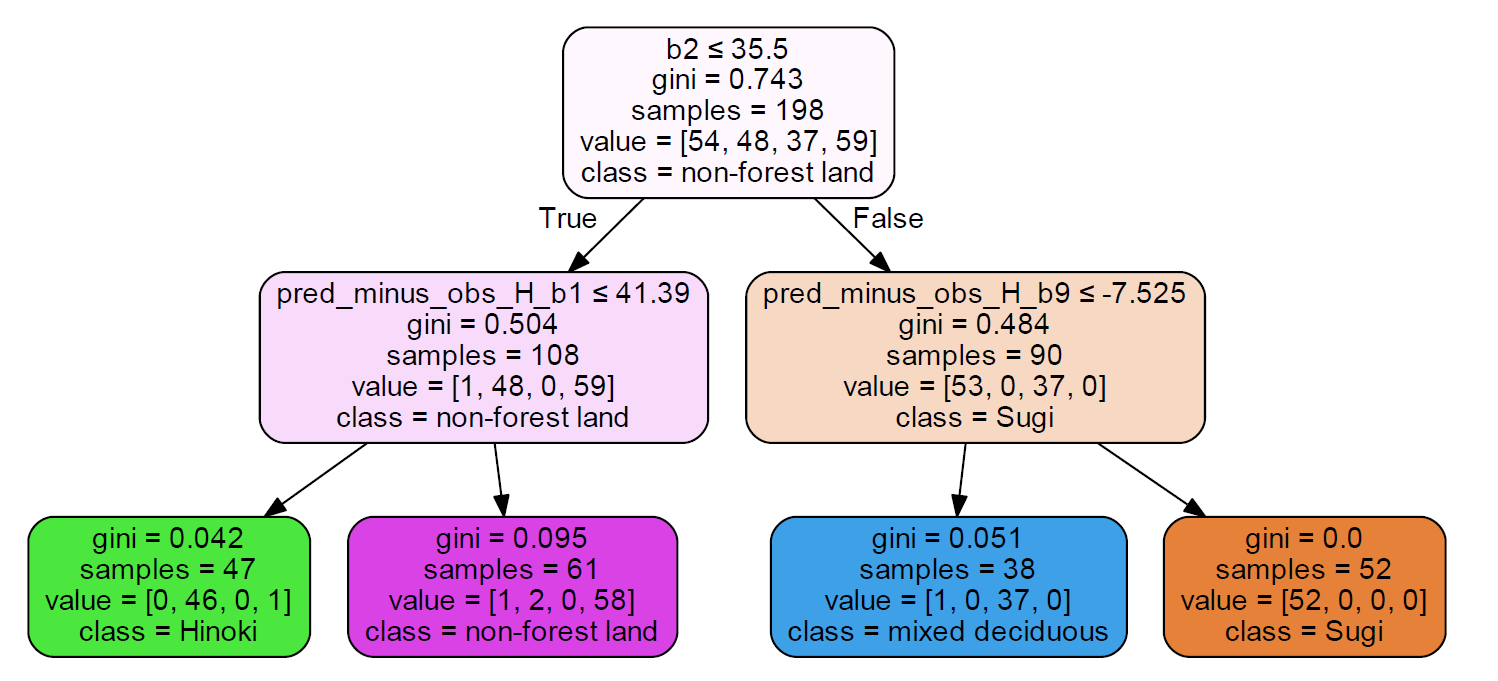
1. Training set partition

*Image 1: Training set using max depth =4*

*Image 2: Training set using max depth =3*

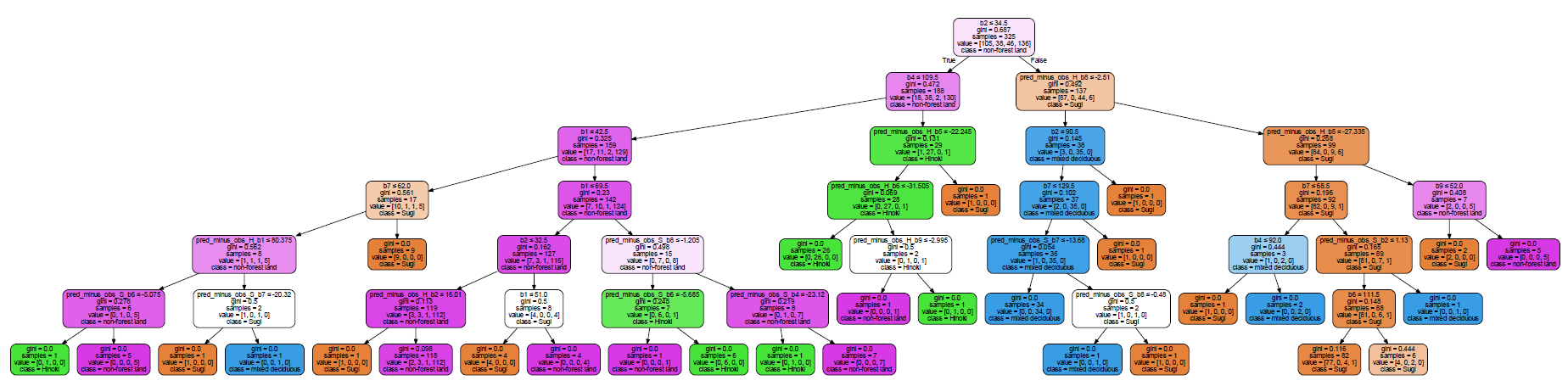


*Image 3: Training set using max depth =2*

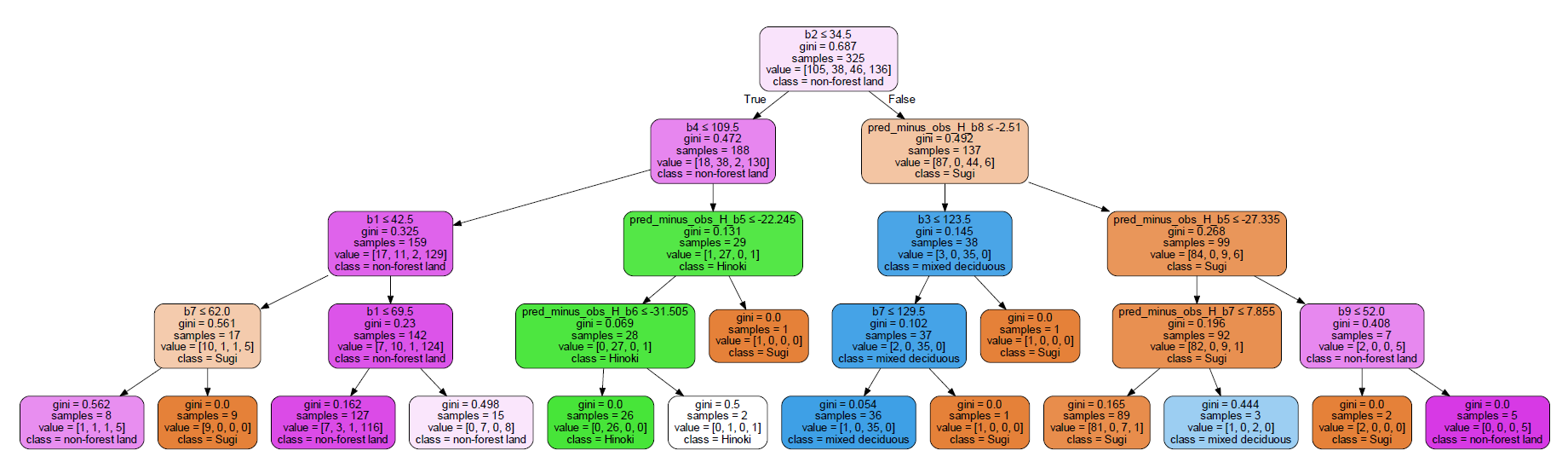
3 decision trees have been generated based on the training set data with 198 samples using the max depth as parameter. Firstly, we can see that the higher the depth, the deeper the tree, the more leaf nodes it has and it includes more information. Focusing on the zero Gini index, the higher depth has more zero Gini index, which summarized as below:

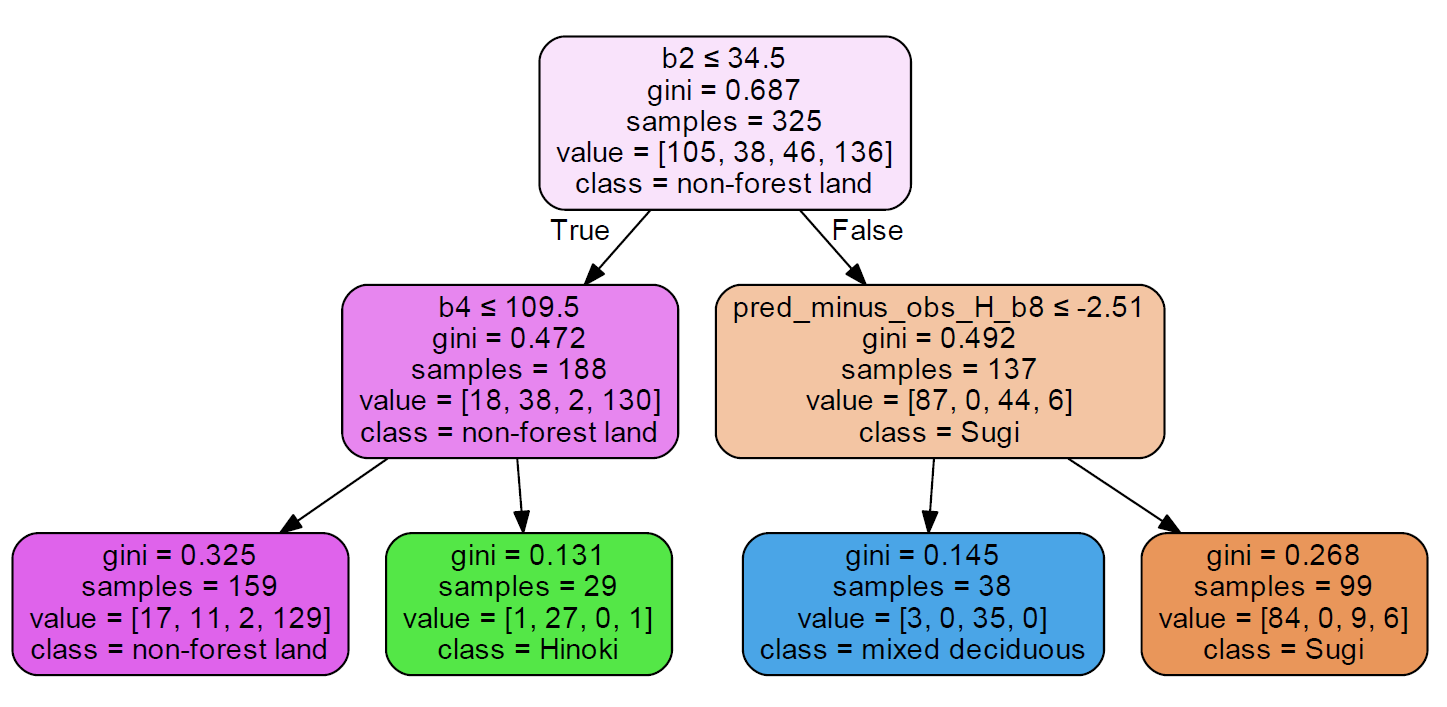
* 4 Max depths 🡪 8 zero Gini index
* 3 Max depths 🡪 6 zero Gini index
* 2 Max depths 🡪 1 zero Gini index

As a result, the higher the depth, the information is more reliable as there are more zero Gini index. In the case of max depths equal of 4, refers to Image 1, Gini index of every end nodes of the tree is equal to zero, which indicates no impurity, all 198 samples are correctly classify. On the other hand, for the case of 2 max depths, refers to Image 3, there are only 1 end nodes has the zero Gini index, which contains of 52 samples, meaning that only 52 samples among 198 samples confirm they belongs to Sugi forest type, but for the other 3 classes, outliers occurred, for example 1 outlier which supposed to be non-forest land occured in Hinoki class among the 46 samples.

(b) Testing set partition

*Image 4:* *Testing set using max depth =6*

*Image 5:* *Testing set using max depth =4*



*Image 6:* *Testing set using max depth =2*

In the case of using testing set which contains 325 samples, we can see that even using the same parameter which is max depth equal to 4, the trees became more complicated, refers to Image 5. the This can be explained because more data sample, more noise will occur. Summary of the zero Gini index obtained from this testing data set is as below:

* 6 Max depths 🡪 26 zero Gini index
* 4 Max depths 🡪 7 zero Gini index
* 2 Max depths 🡪 0 zero Gini index

Although the depth had increased to 6 and there are 26 zero Gini index, but there are just only of 116 samples are correctly classify. Moreover,