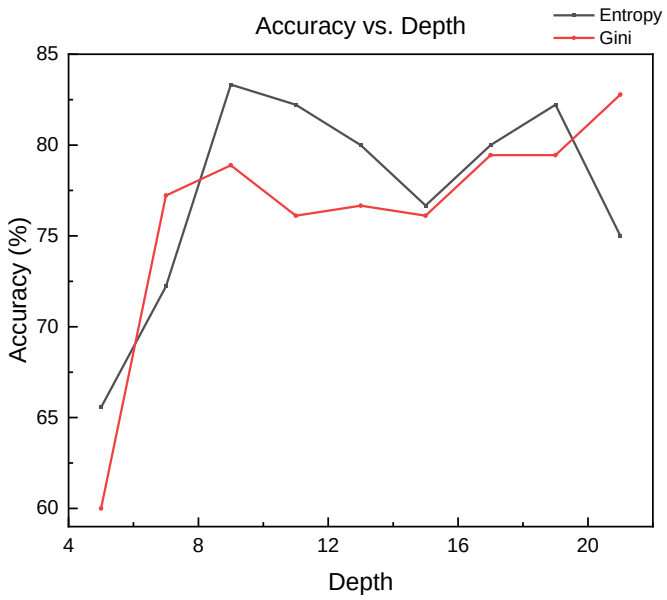
**Decision Tree:**

1. Initialize: Set the maximum depth and the loss function.
2. Depth: The tree will stop growing when the one of the following events occur.
   a. If the depth is larger than the self-defined maximum depth.
   b. If they belong to the same class.
3. Loss: Both “Entropy” and “Gini” could be applied.

$$IG(D_p, f) = I(D_p) - \frac{N_{left}}{N_p} I(D_{left}) - \frac{N_{right}}{N_p} I(D_{right})$$

   a. Entropy: $I(t) = -\sum_{i=1}^{c} p(i|t) \log_2 p(i|t)$
   b. Gini: $I(t) = 1 - \sum_{i=1}^{c} p(i|t)^2$

4. Performance:
   a. Randomly select 20% of the data to be validation data set.
   b. Each depth is performed five times, resulting in an average accuracy.



**Multi-decision trees**

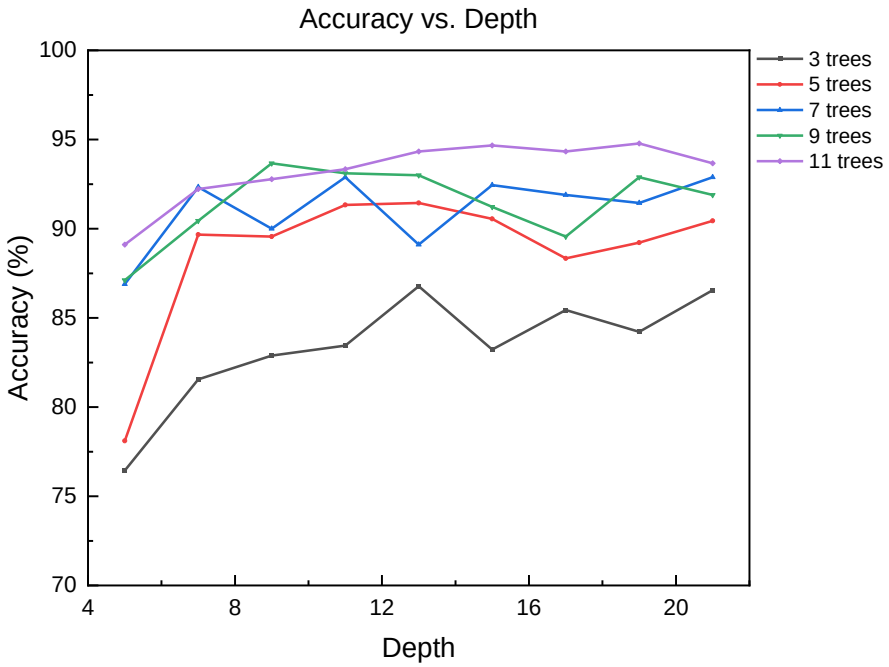
1. Initialize: Set the tree numbers, maximum depth for trees and the loss function.
2. Depth: The tree will stop growing when the one of the following events occur.
   c. If the depth is larger than the self-defined maximum depth.
   d. If they belong to the same class.
3. Loss: “Entropy”, “Gini” and “Mix” could be applied.
   *Mix: some trees apply “Gini”, others apply “Entropy”

4. Training: The training data set would be resampled for each tree in order to increase the diversity (Default: 80%).

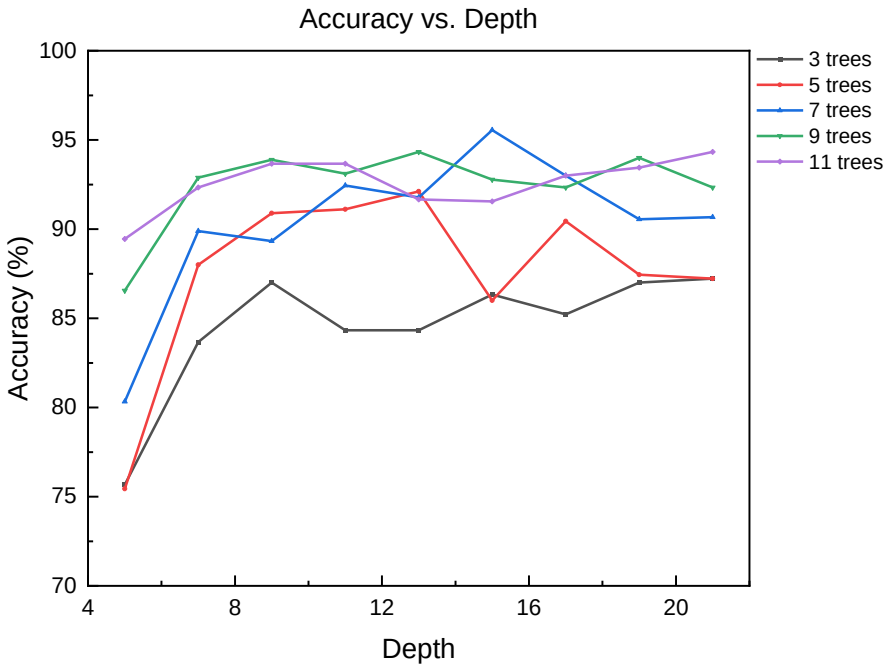
5. Predict: Each tree predicts their own result then the final prediction is determined by voting.

6. Performance:

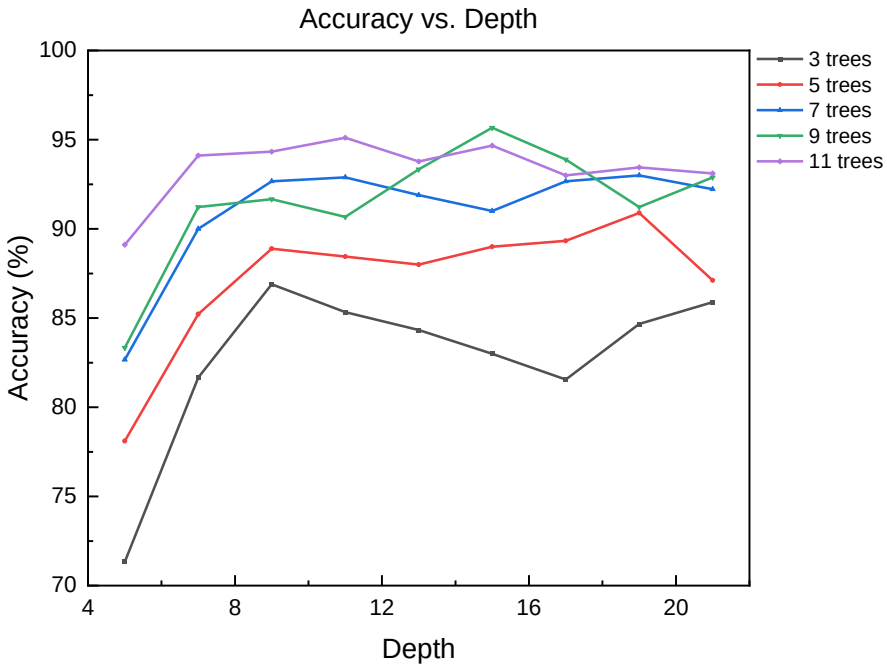
   a. Entropy:



b. Gini:


Accuracy vs. Depth
Accuracy (%)
Depth
100
95
90
85
80
75
70
4
8
12
16
20
3 trees
5 trees
7 trees
9 trees
11 trees

c. Mix:


Accuracy vs. Depth
Accuracy (%)
Depth
100
95
90
85
80
75
70
4
8
12
16
20
3 trees
5 trees
7 trees
9 trees
11 trees

**Conclusion:**

In order to improve the accuracy, we can:

1. Properly increase the depth of the tree but be careful do not overfitting (Validation accuracy decreases).
2. By observing my statistics, if smaller depth (< 16) of the tree is preferred, then I might apply “Entropy”. If larger depth (> 16) of the tree is preferred, then I might apply “Gini”.
3. Multi-trees greatly increase the accuracy.
4. The more trees we create, the more accurate the prediction is.
5. Even though we increase the depth of the trees, the accuracy will not be improved if too many trees are applied. (And very time-cost)
6. Last, The “Mix” method which includes both “Entropy” and “Gini” make the 95% accuracy with 11 trees and the maximum depth < 12, which performs the best among those cases.