



INTRODUCTION TO ARTIFICIAL INTELLIGENCE

REPORT LAB 2

Report's writer:

Name: Nguyen Dang Khoa

ID: 19127177

Class: 19clc7

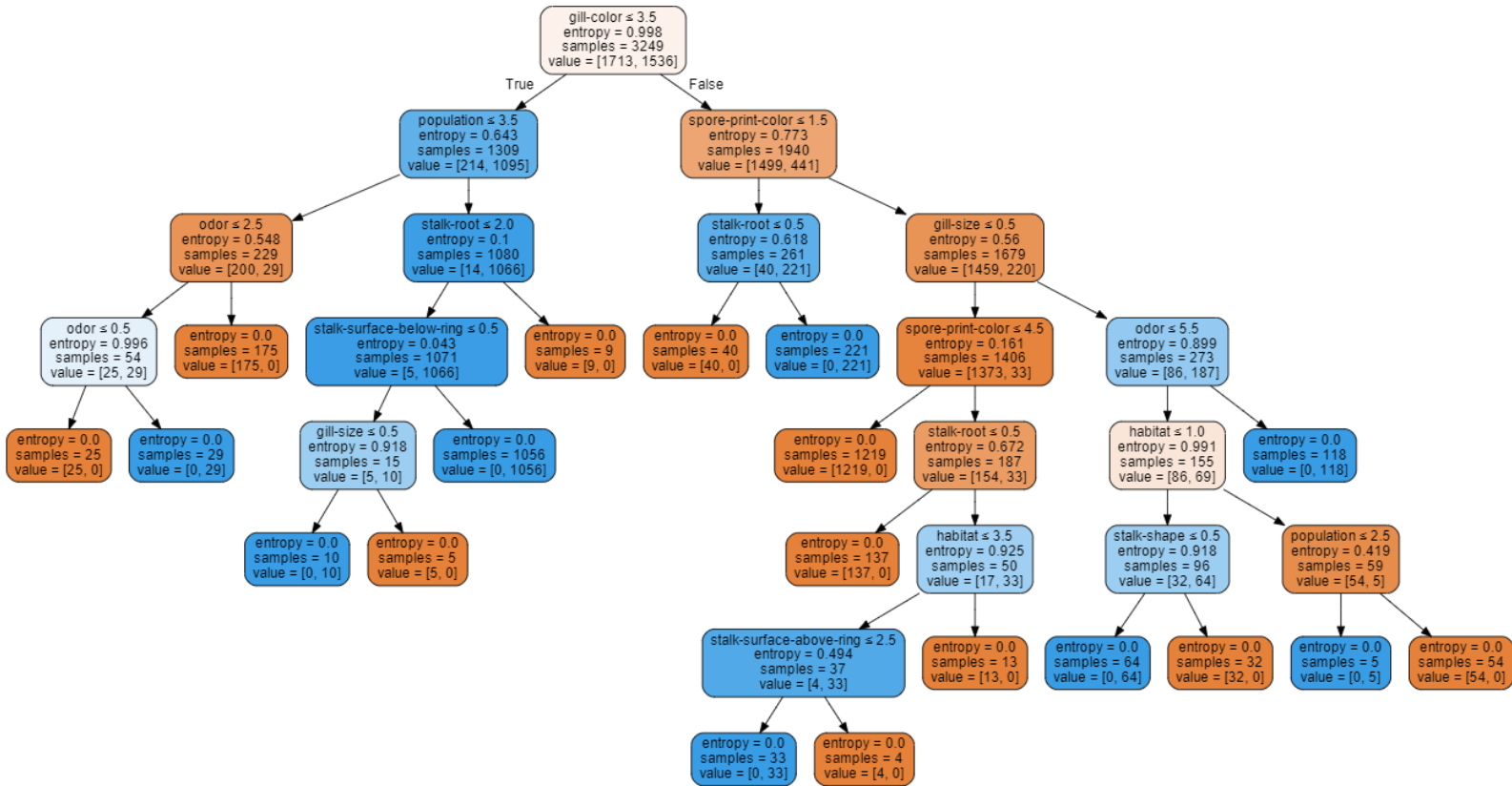
Contents

1. Building the decision tree classifiers:	2
2. Evaluating the decision tree classifiers:.....	6
✓ Classification report:	6
✓ Confusion matrix:.....	7
✓ Interpretation:	9
✓ Comment:	9
3. The depth and accuracy of a decision tree	10
✓ Decision tree	10
✓ Accuracy score report.....	16
✓ Comment:	16
3. REFERENCES	17

The data from the report might be different from results generated from source code, since the data sets are shuffled and split randomly in each execution. The report will use the most recently generated results and still covers the meaning of the data.

1. Building the decision tree classifiers:

- Set 1 (40/60):



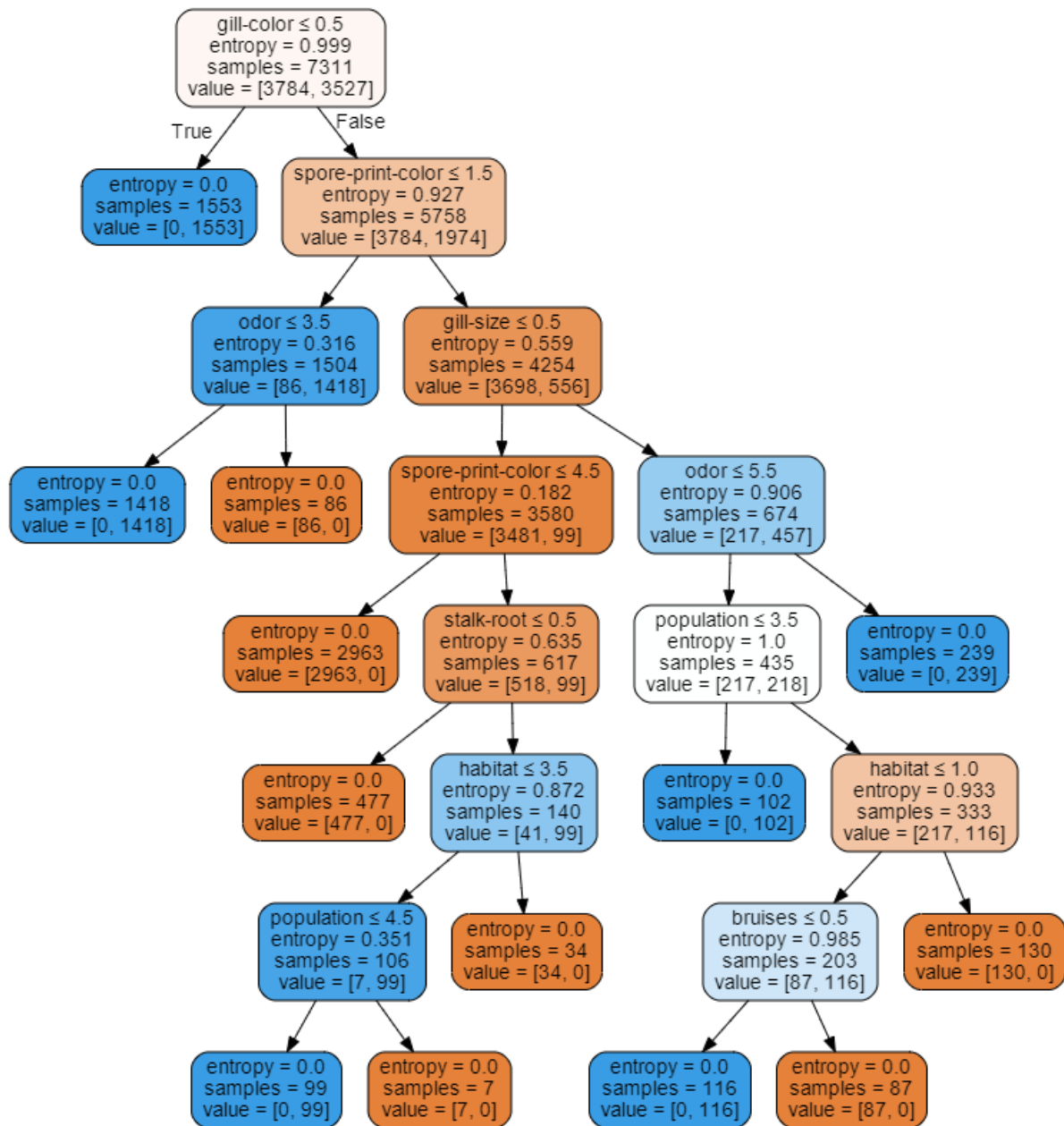
- Set 2 (60/40):



- Set 3 (80/20):



- Set 4 (90/10):



2. Evaluating the decision tree classifiers:

✓ Classification report:

- **Set 1 (40/60):**

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2495
1	1.00	1.00	1.00	2380
accuracy			1.00	4875
macro avg	1.00	1.00	1.00	4875
weighted avg	1.00	1.00	1.00	4875

- **Set 2 (60/40):**

0	1.00	1.00	1.00	1669
1	1.00	1.00	1.00	1581
accuracy			1.00	3250
macro avg	1.00	1.00	1.00	3250
weighted avg	1.00	1.00	1.00	3250

- **Set 3 (80/20):**

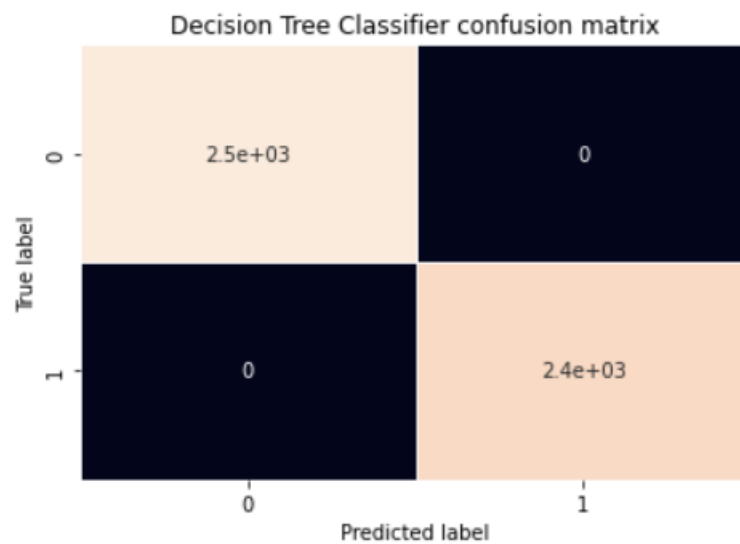
	precision	recall	f1-score	support
0	1.00	1.00	1.00	827
1	1.00	1.00	1.00	798
accuracy			1.00	1625
macro avg	1.00	1.00	1.00	1625
weighted avg	1.00	1.00	1.00	1625

- **Set 4 (90/10):**

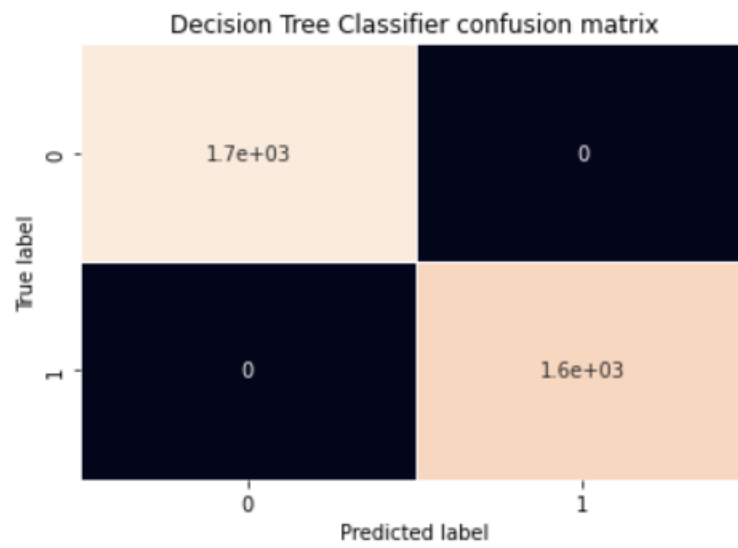
	precision	recall	f1-score	support
0	1.00	1.00	1.00	424
1	1.00	1.00	1.00	389
accuracy			1.00	813
macro avg	1.00	1.00	1.00	813
weighted avg	1.00	1.00	1.00	813

- ✓ **Confusion matrix:**

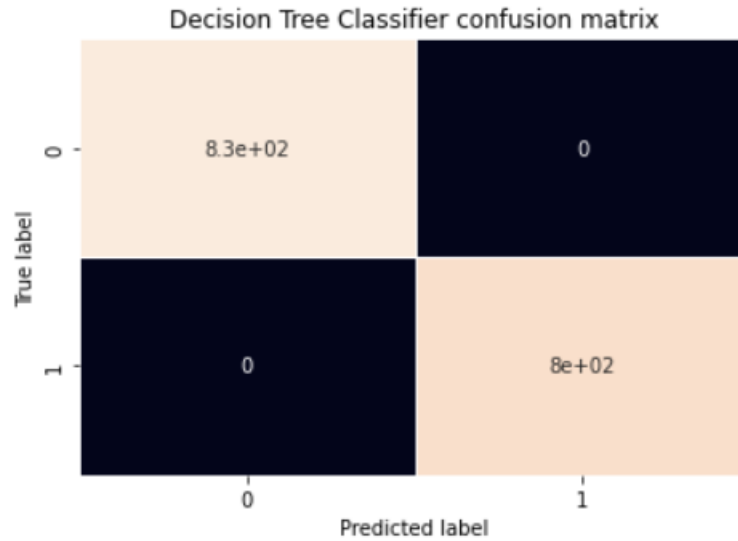
- **Set 1 (40/60):**



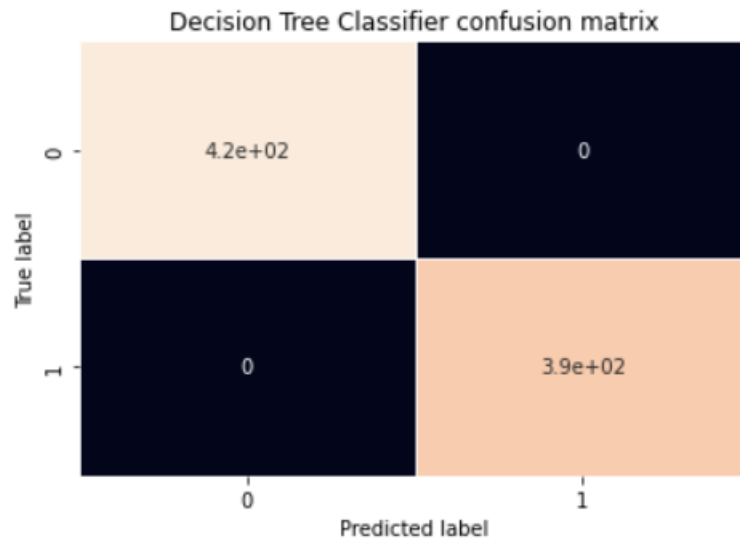
- **Set 2 (6/40):**



- **Set 3 (80/20):**



- **Set 4 (90/10):**



- ✓ **Interpretation:**

- Confusion matrix shows how the data is classified by the model. It presents the number of data classified in each class so that we can clearly see which data from which class usually wrongly classified to another class.
- The classification report is used to measure the quality of predictions. It illustrates how many predictions about mushrooms are true and how many are false.
- For each class, the classification report shows the main metric classification metrics precision, recall, and f1-score. The metrics are calculated by using true and false negative, true and false positive.
- Precision shows the percentage of the predictions are correct.
- Recall is the portion of true positive cases among predicted positive cases.
- F1-score is calculated by using precision and recall, which means it is used when we want to balance between precision and recall.
- From classification report, I will only use precision, accuracy, f1-score and recall to discuss about the mushrooms data.

- ✓ **Comment:**

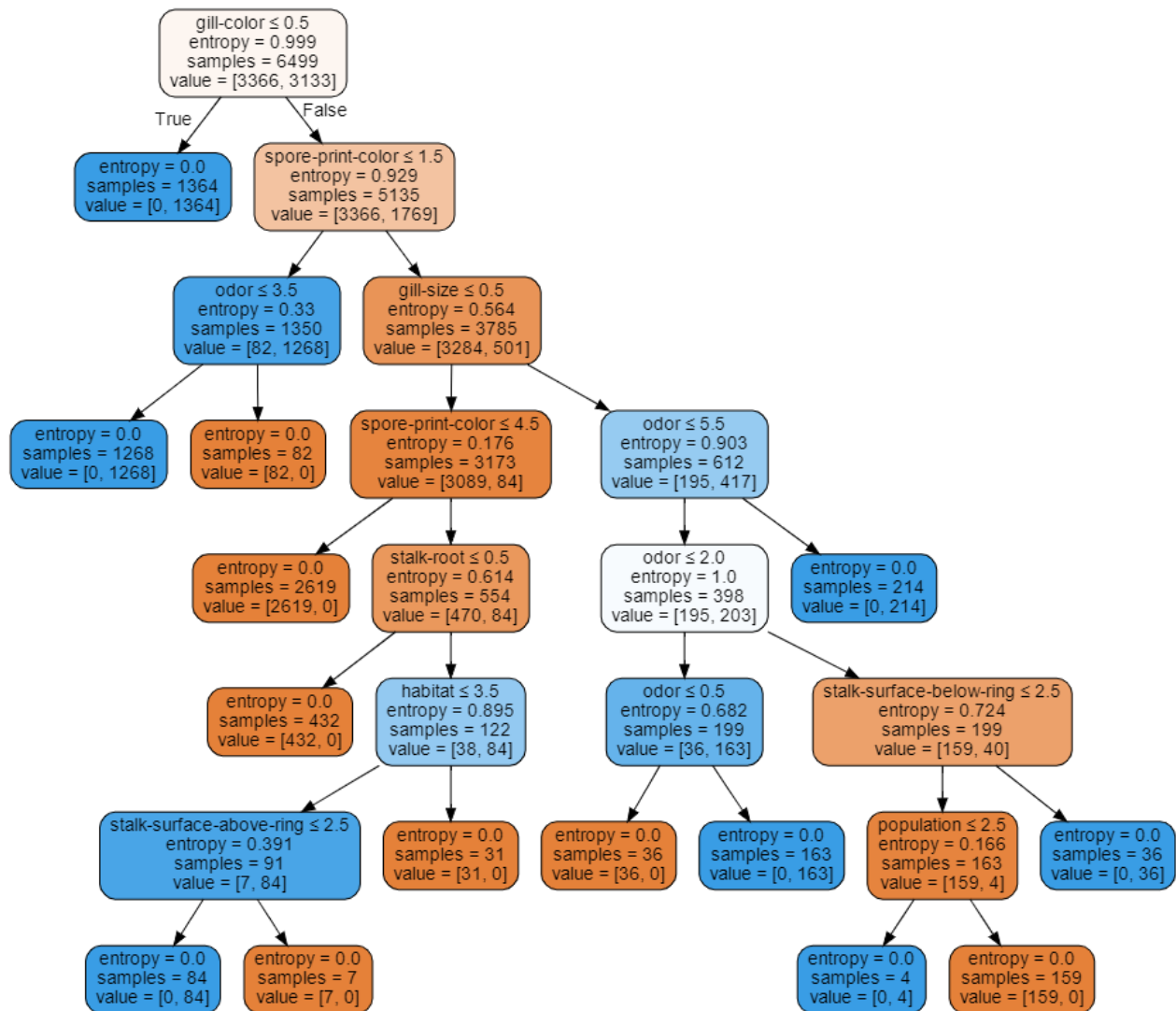
- From the confusion matrix of all data sets, the amount of edible mushrooms and poisonous ones are perfectly classified which means that the predicted data are perfectly matched with the labeled data. As a result, the accuracy from the reports are exactly 1.00. Since the accuracy is 1.00, the number of false positive cases are 0 which is why the precision also becomes 1.00 as well as recall and F1-score.
- The classification reports conclude that the models report mostly the same despite the models are trained and tested in different ways.
- Because the depth of the decision tree is unlimited so the tree grow to their maximum depth in spite of different methods of training and testing. As a result,

the models are really complex, hence overfitting. So the models does not capture any useful patterns which would be used for further predictions.

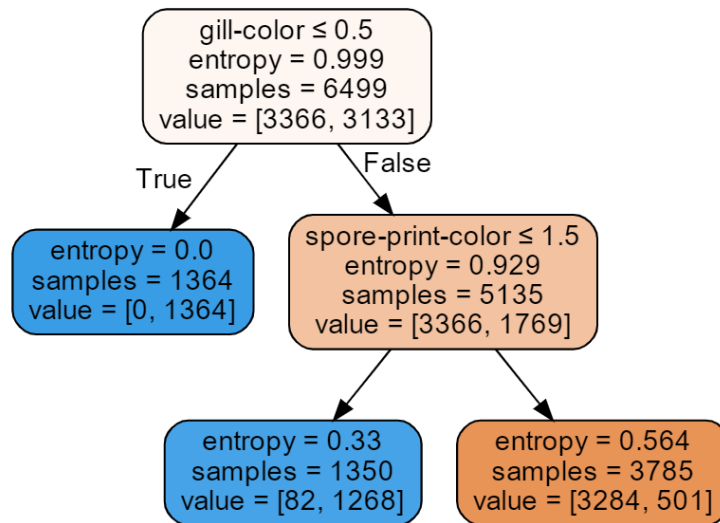
3. The depth and accuracy of a decision tree

✓ Decision tree

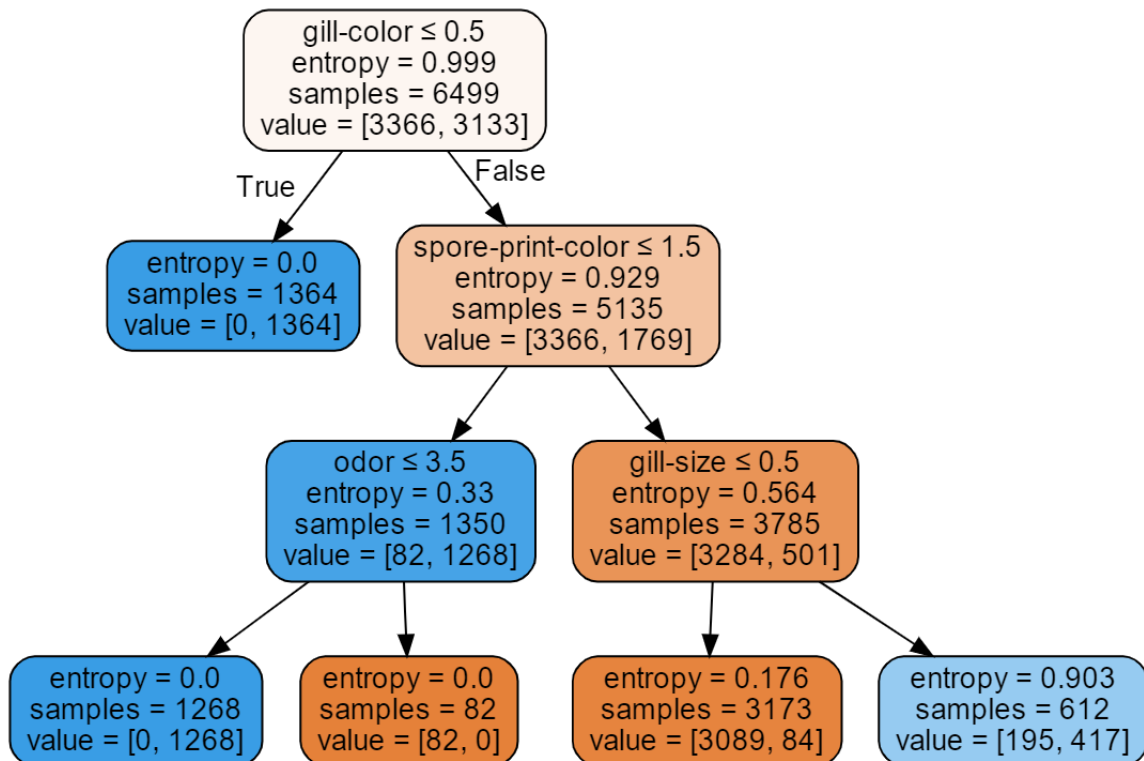
max_depth = None



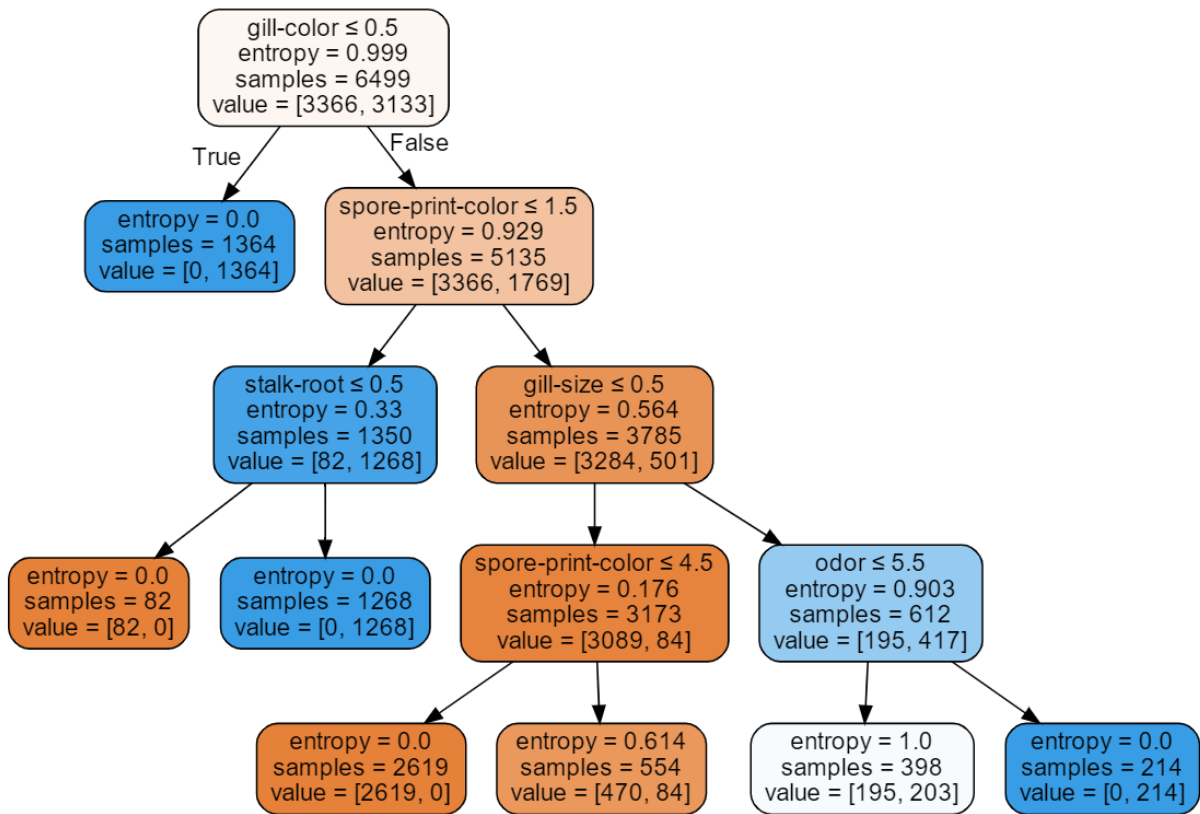
max_depth = 2



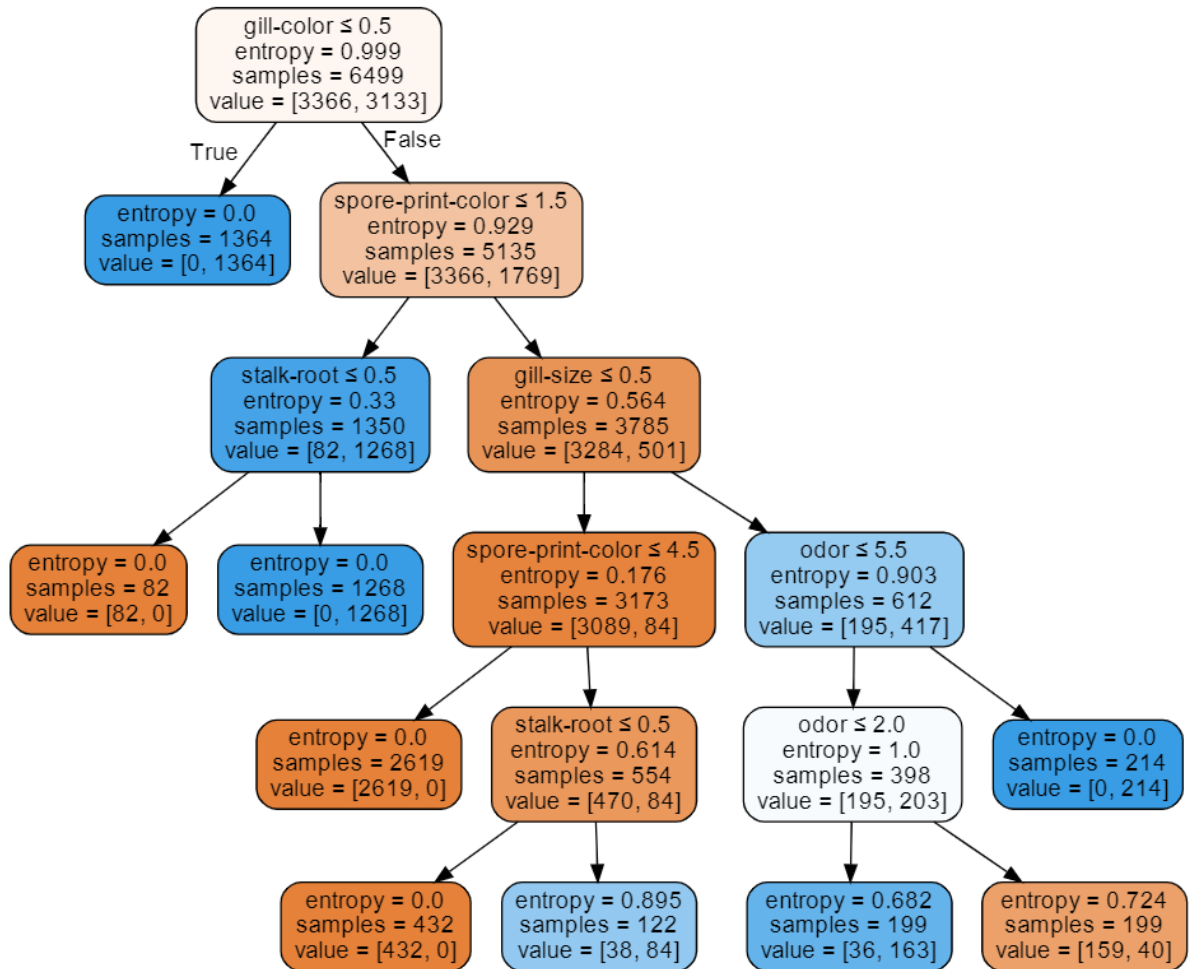
max_depth = 3



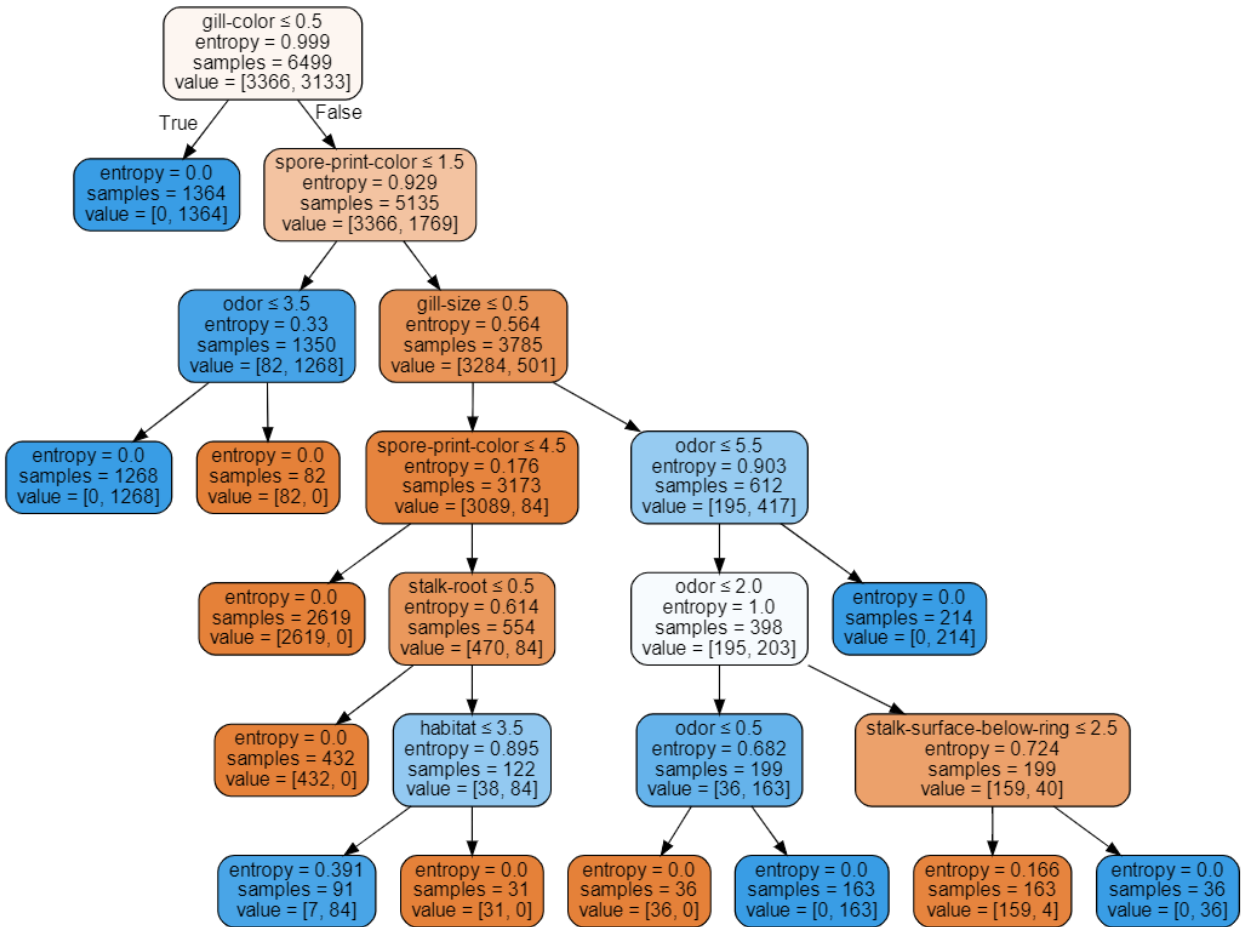
max_depth = 4



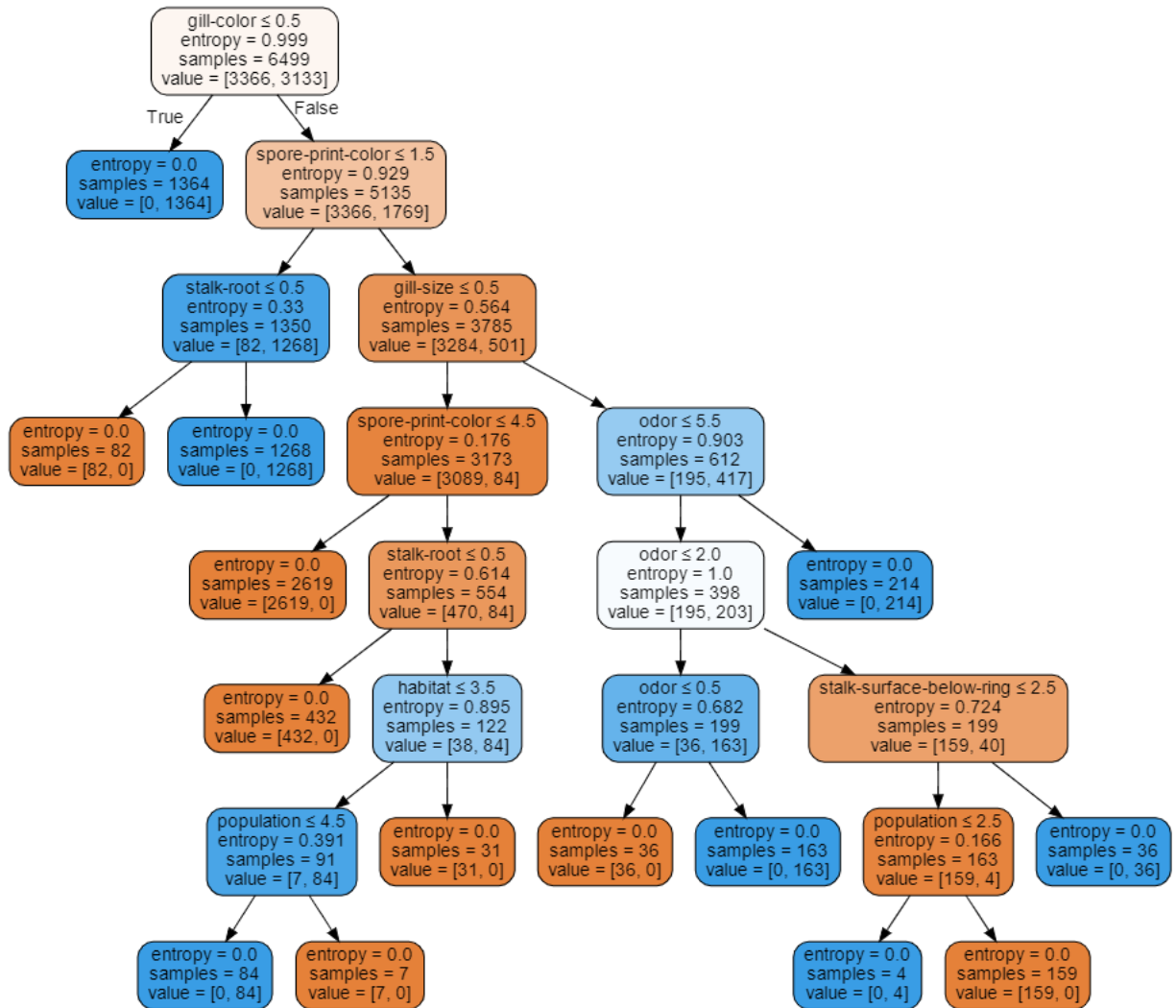
max_depth = 5



max_depth = 6

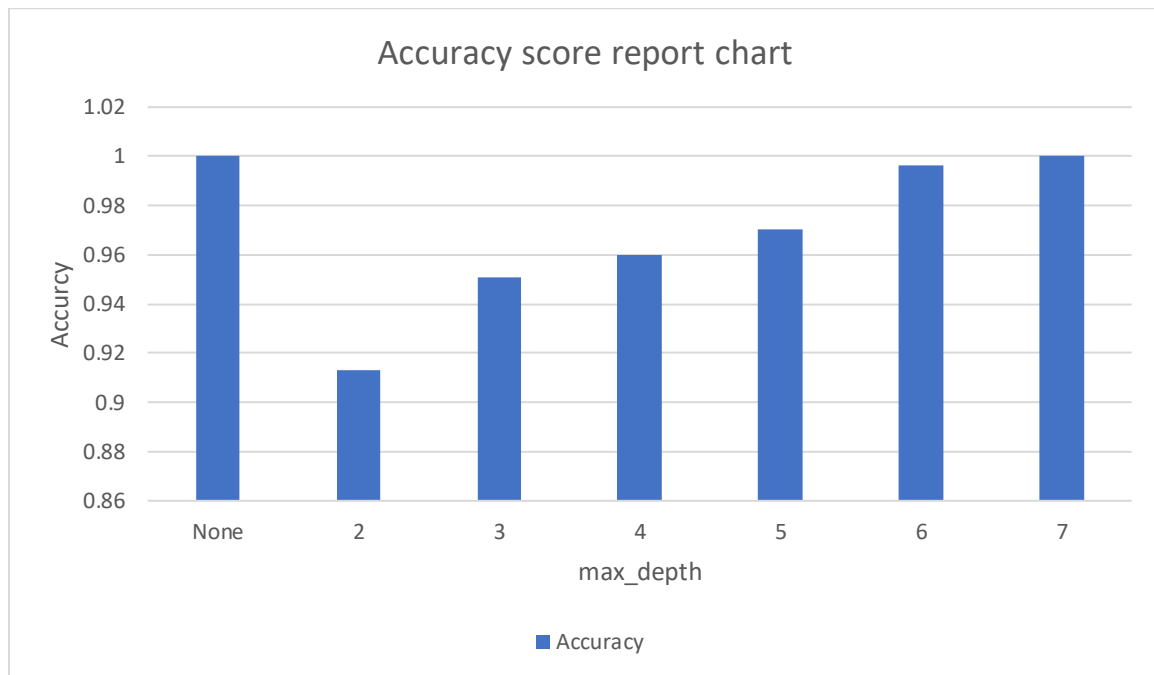


max_depth = 7



✓ Accuracy score report

max_depth	None	2	3	4	5	6	7
Accuracy	1.000	0.913	0.951	0.960	0.970	0.996	1.000



✓ Comment:

- max_depth = None and max_depth = 7 seem to have identical accuracy. We can see that the decision tree of the 2 max_depths are similar in shape, hence the decision tree's maximum depth is 7.
- The max_depth of the decision tree is the maximum depth that we want the tree grows to. The depth of the decision tree seems to increase the complexity of the model so that the accuracy gradually reaches to 1 whenever the max_depth increases.

3. REFERENCES

- [1] - [python - Decision Tree Sklearn -Depth Of tree and accuracy - Stack Overflow](#)
- [2] - [Accuracy, Precision, Recall or F1? | by Koo Ping Shung | Towards Data Science](#)
- [3] - [Understanding the Classification report through sklearn – Muthukrishnan](#)
- [4] - [Analysis and Classification of mushrooms | Kaggle](#)
- [5] - [python - Macro VS Micro VS Weighted VS Samples F1 Score - Stack Overflow](#)