VIET NAM NATIONAL UNIVERSITY HO CHI MINH CITY
UNIVERSITY OF SCIENCE
FACULTY OF INFORMATION TECHNOLOGY

LAB 02 REPORT
SUBJECT: ARTIFICIAL INTELLIGENCE
DECISION TREE WITH SCIKIT-LEARN

Theory lecturer:
Mrs. Nguyen Ngoc Thao
Instructors:
Mr. Le Ngoc Thanh
Mr. Nguyen Hai Dang
Mrs. Ho Thi Thanh Tuyen
Student - ID:
Pham Thi Anh Phat 20127680

(2021-2022)

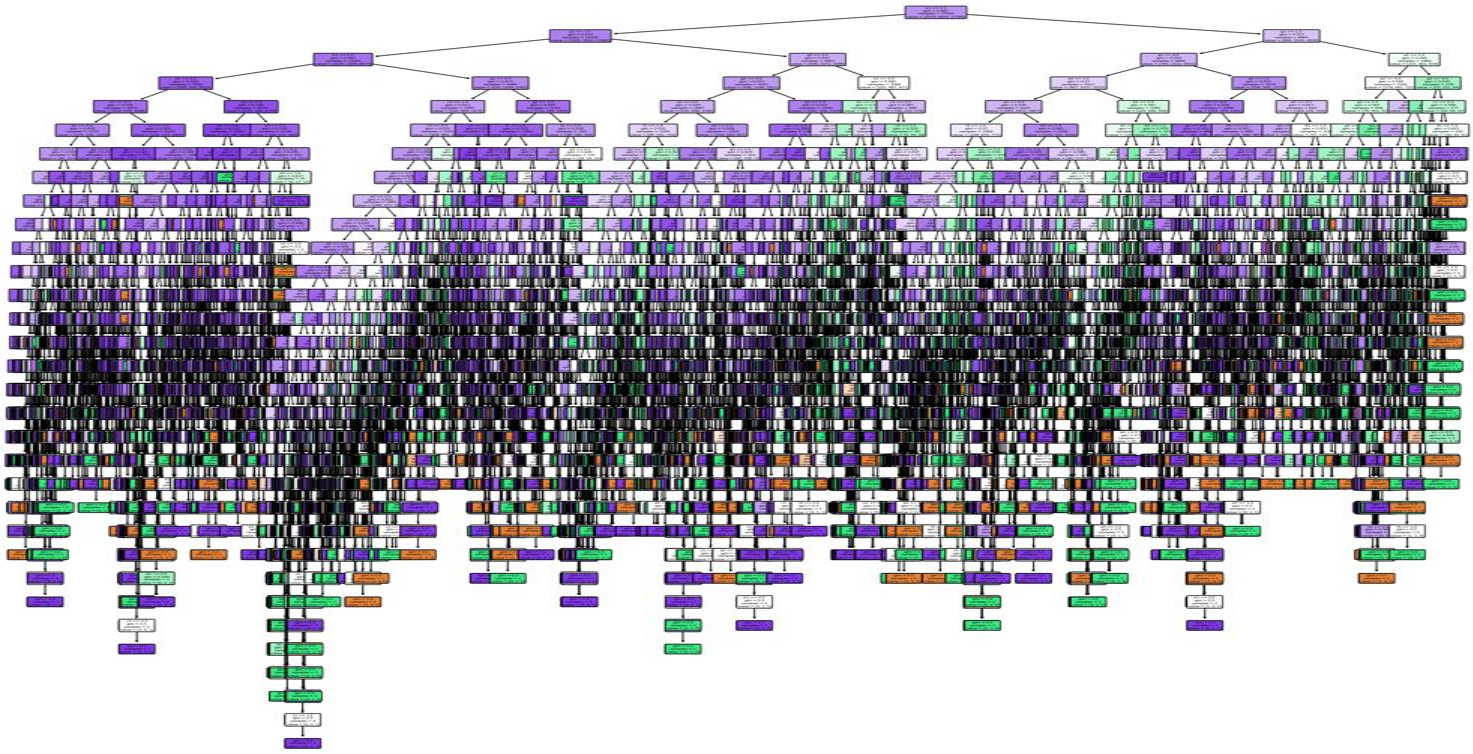
TABLE OF CONTENTS

I. REQUIREMENTS

- Requirement 2, 3 work on for each training sets and test sets with different proportions, so put them together.

➢ **Train/test: 40/60**

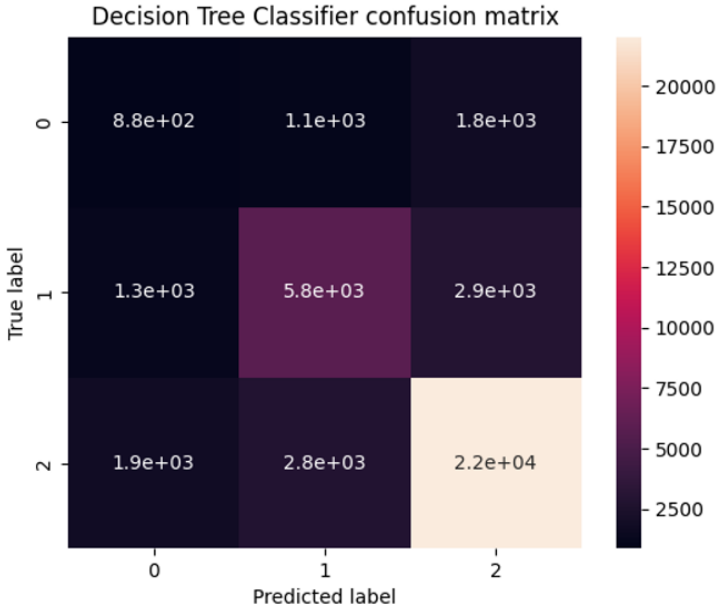
1. *Building the decision tree classifiers*



2. *Evaluating the decision tree classifiers*
   - *Classification report and confusion matrix*

Decision Tree Classifier report

	precision	recall	f1-score	support
0	0.22	0.23	0.22	3846
1	0.59	0.58	0.59	9961
2	0.82	0.82	0.82	26728
accuracy			0.71	40535
macro avg	0.55	0.54	0.54	40535
weighted avg	0.71	0.71	0.71	40535

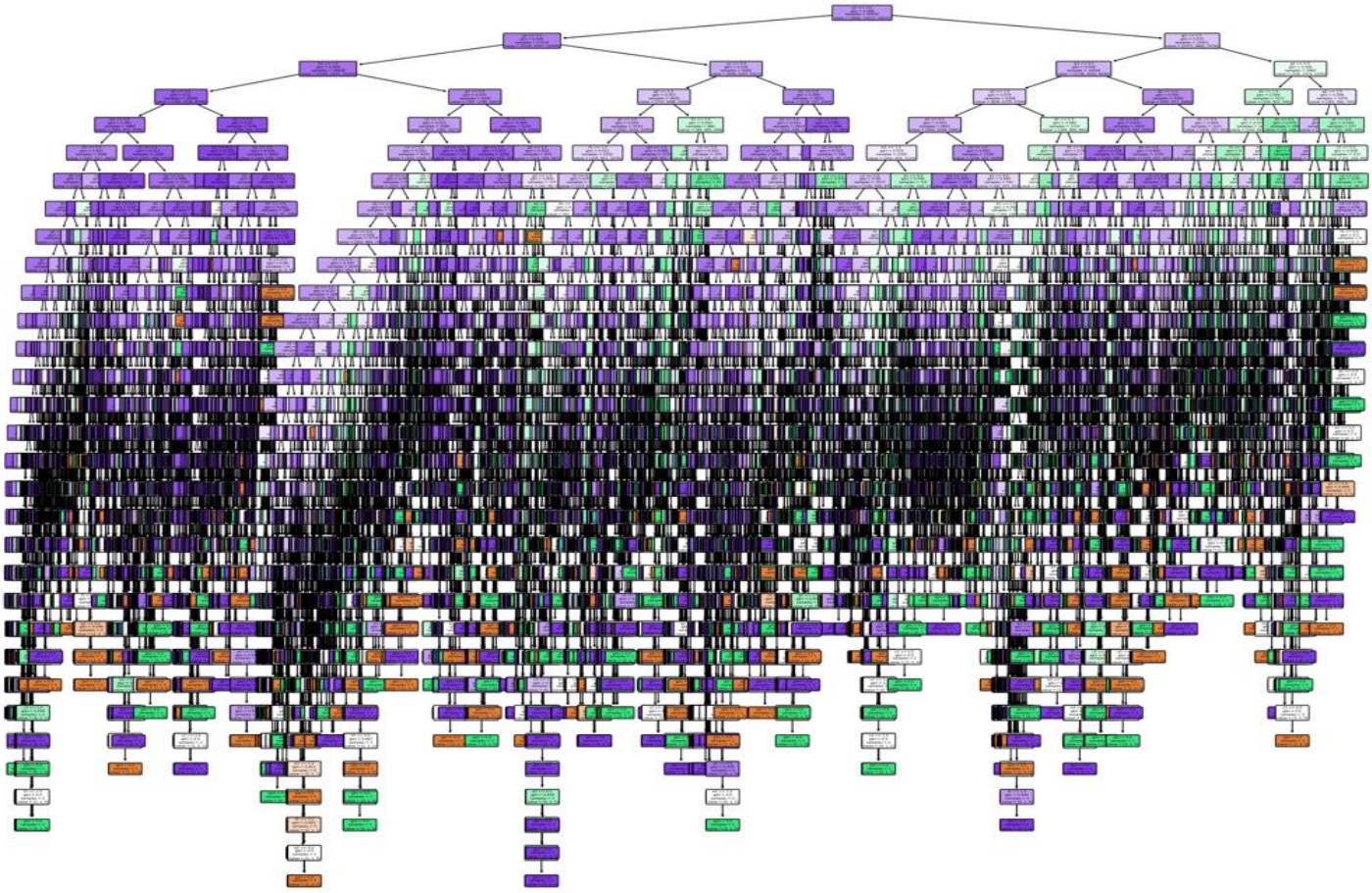


- *Comments*

- The most correctly predicted when predicted label and true label are 2. Because it have the outcome value is highest (the color is light and the precision value is highest – 0.82)
- The most incorrectly predicted when predicted label is 1 and true label is 0 and opposite. (the darkest color and the precision value is lowest – 0.22).
- In this case, the accuracy is lowest (0.71) among four test cases. Because the train takes up less than the test part in the data.

➢ **Train/test: 60/40**

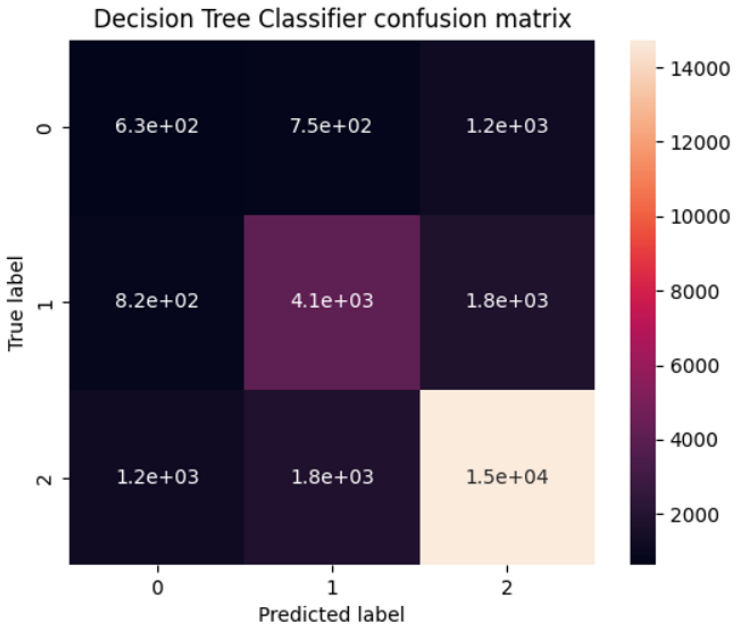
1. *Building the decision tree classifiers*



2. *Evaluating the decision tree classifiers*
    - *Classification report and confusion matrix*

Decision Tree Classifier report

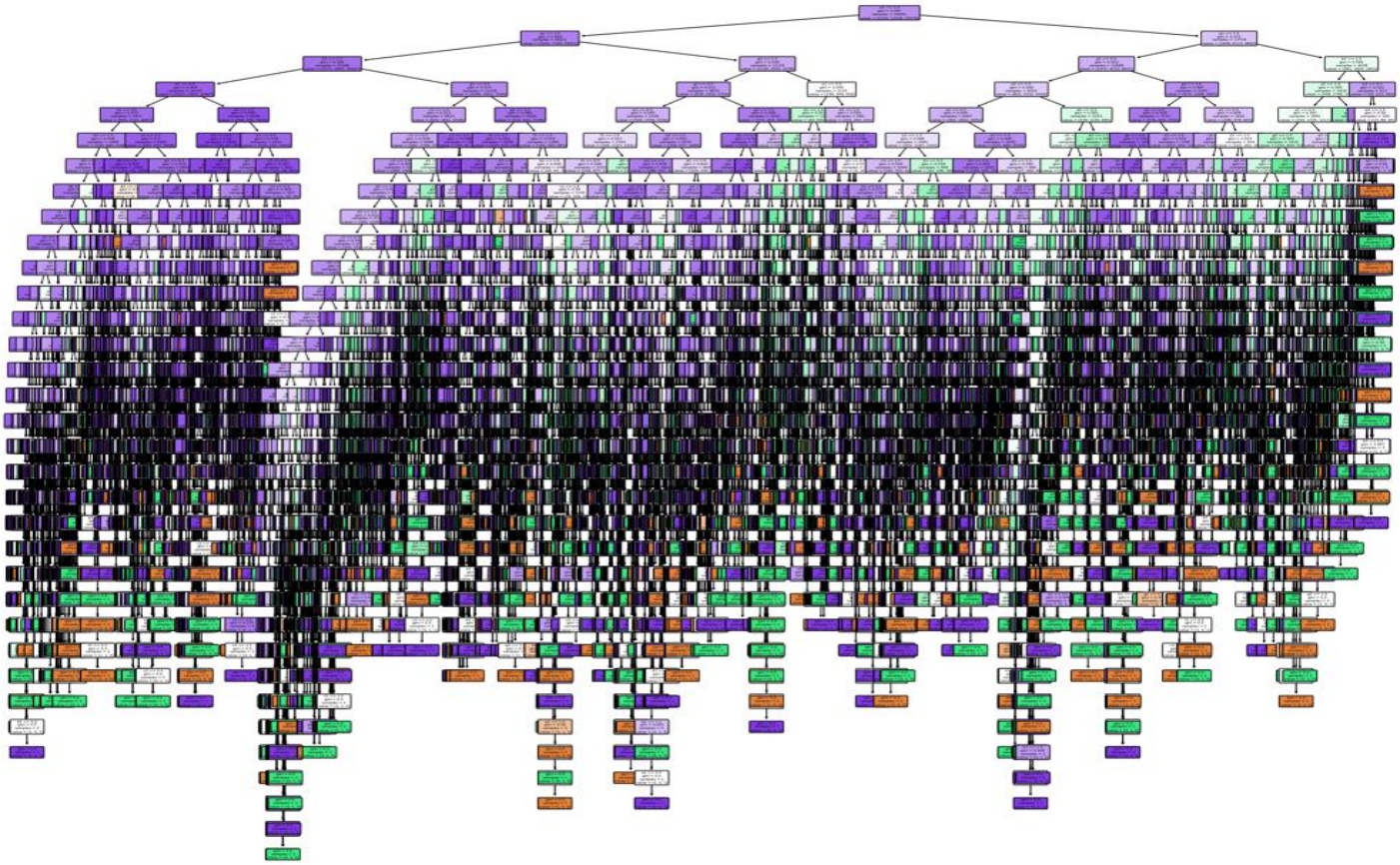
	precision	recall	f1-score	support
0	0.23	0.24	0.24	2559
1	0.62	0.61	0.61	6623
2	0.83	0.83	0.83	17841
accuracy			0.72	27023
macro avg	0.56	0.56	0.56	27023
weighted avg	0.72	0.72	0.72	27023



- *Comments*

- The most correctly predicted when predicted label and true label are 2. Because it have the outcome value is highest (the color is light and the precision value is highest – 0.83)
- The most incorrectly predicted when predicted label is 2 and true label is 0 and opposite. (the darkest color and the precision value is lowest – 0.23).

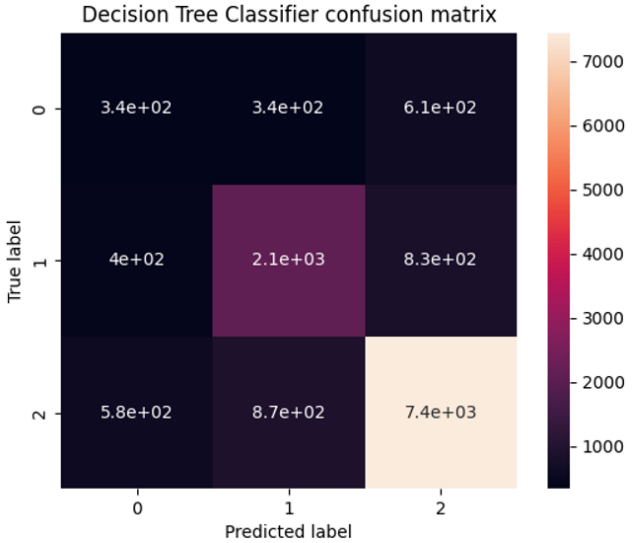
- **Train/test: 80/20**
  1. *Building the decision tree classifiers*



  2. *Evaluating the decision tree classifiers*
     - *Classification report and confusion matrix*

Decision Tree Classifier report

	precision	recall	f1-score	support
0	0.24	0.25	0.24	1278
1	0.64	0.63	0.64	3300
2	0.84	0.84	0.84	8934
accuracy			0.73	13512
macro avg	0.57	0.57	0.57	13512
weighted avg	0.74	0.73	0.74	13512

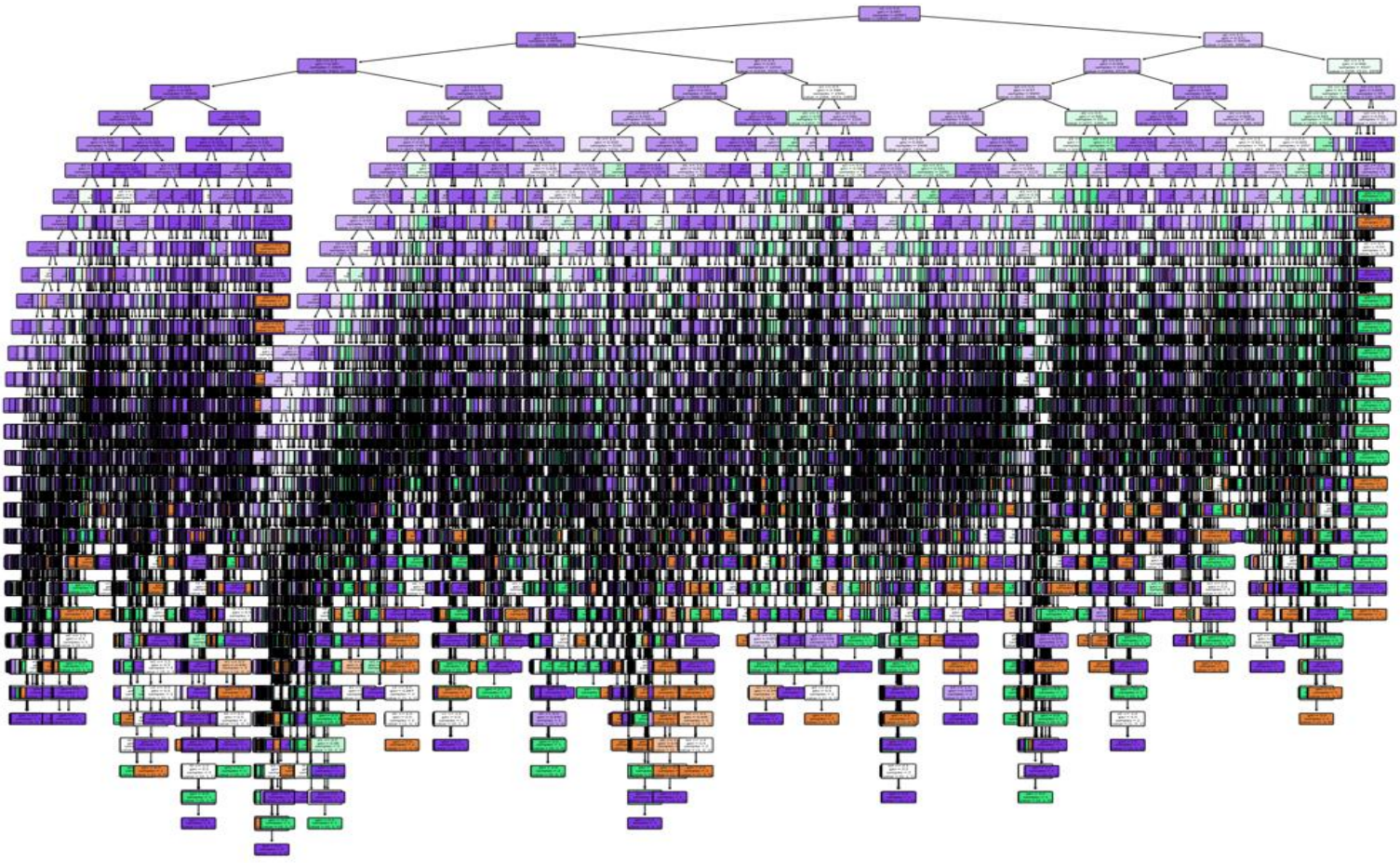


- *Comments*

- The most correctly predicted when predicted label and true label are 2. Because it have the outcome value is highest (the color is light and the precision value is highest – 0.84)
- The most incorrectly predicted when predicted label is 0 and true label is 1 (the darkest color and the precision value is lowest – 0.24).

➢ **Train/test: 90/10**

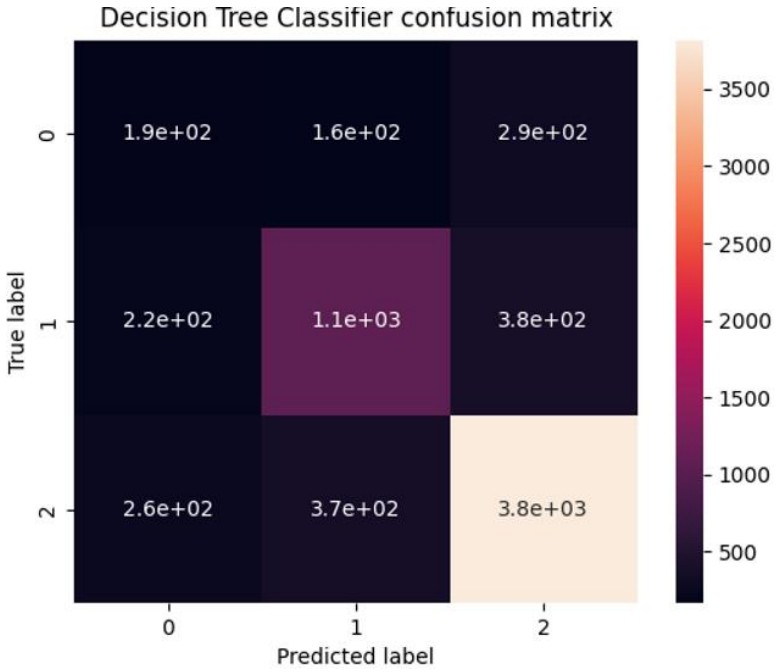
1. *Building the decision tree classifiers*



2. *Evaluating the decision tree classifiers*
    - *Classification report and confusion matrix*

Decision Tree Classifier report

	precision	recall	f1-score	support
0	0.22	0.24	0.23	612
1	0.65	0.64	0.64	1674
2	0.84	0.84	0.84	4470
accuracy			0.74	6756
macro avg	0.57	0.57	0.57	6756
weighted avg	0.74	0.74	0.74	6756

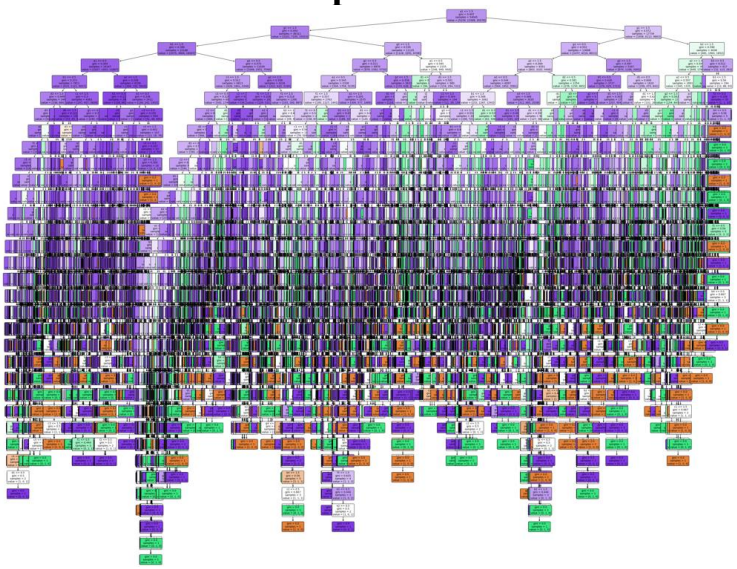


    - *Comments*

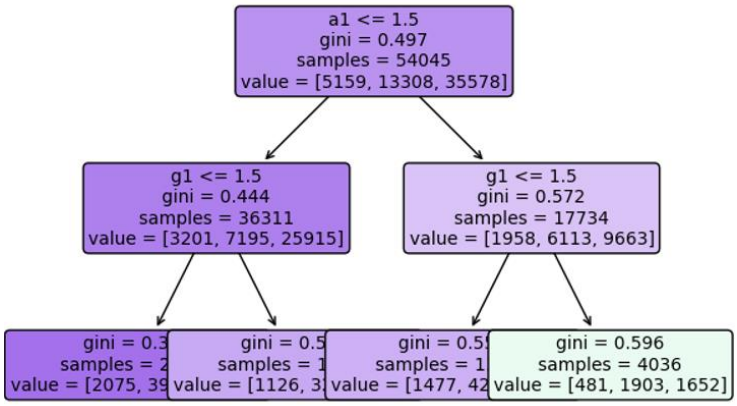
- The most correctly predicted when predicted label and true label are 2. Because it have the outcome value is highest (the color is light and the precision value is highest – 0.84)
- The most incorrectly predicted when predicted label is 1 and true label is 0 and opposite. (the darkest color and the precision value is lowest – 0.22).
- In this case, the accuracy is highest (0.74) and the precision value is improved among four test cases. Because the train part takes up most of the data. (90%)

- **Requirement 4:** ***The depth and accuracy of a decision tree***
  - *Decision tree*

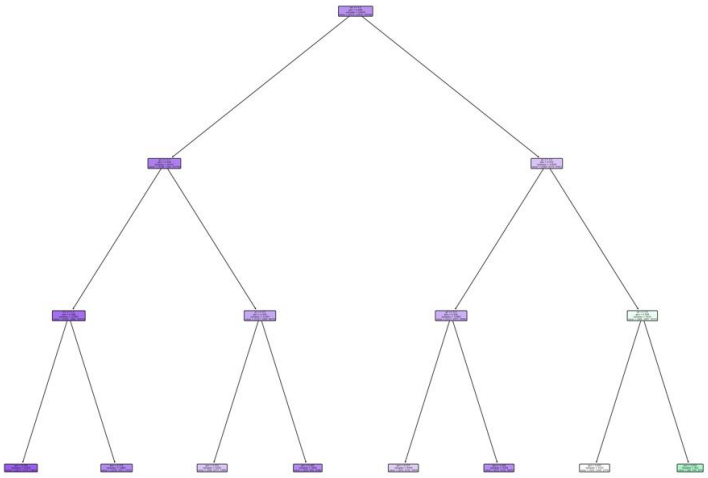
Max depth = None



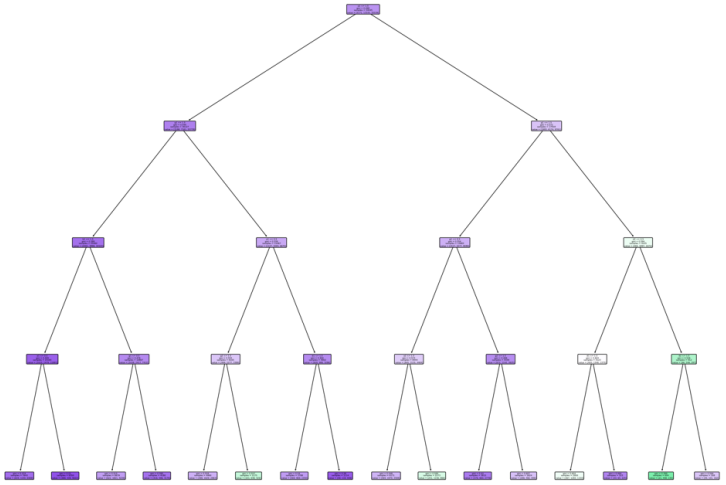
Max depth = 2



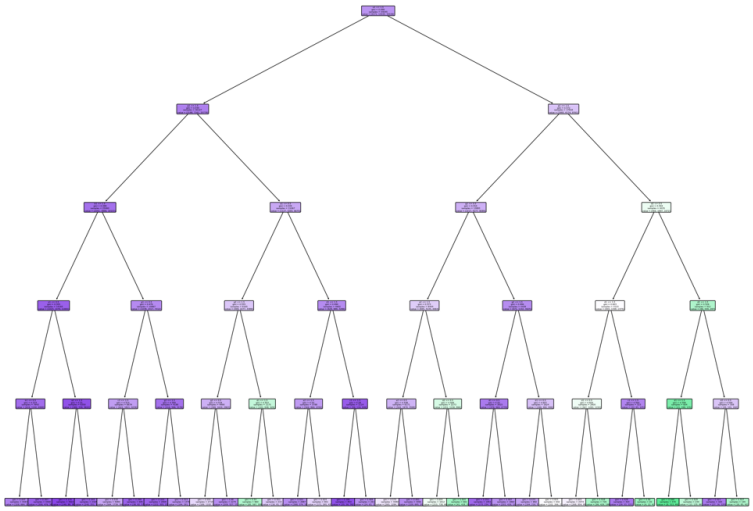
Max depth = 3



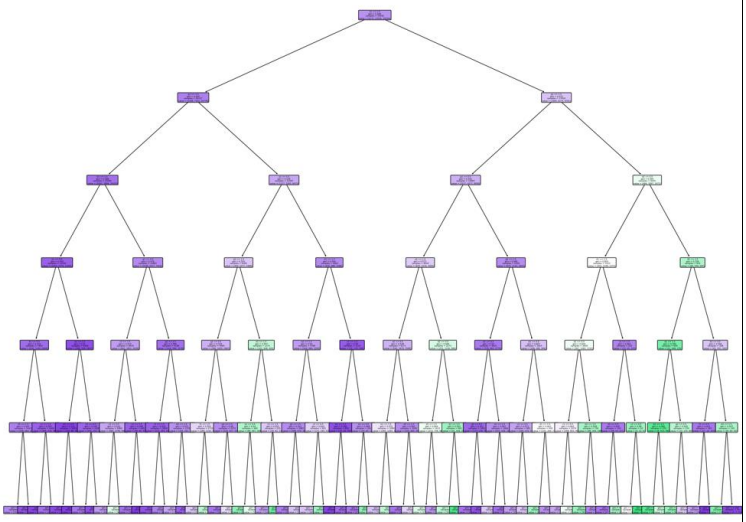
Max depth = 4



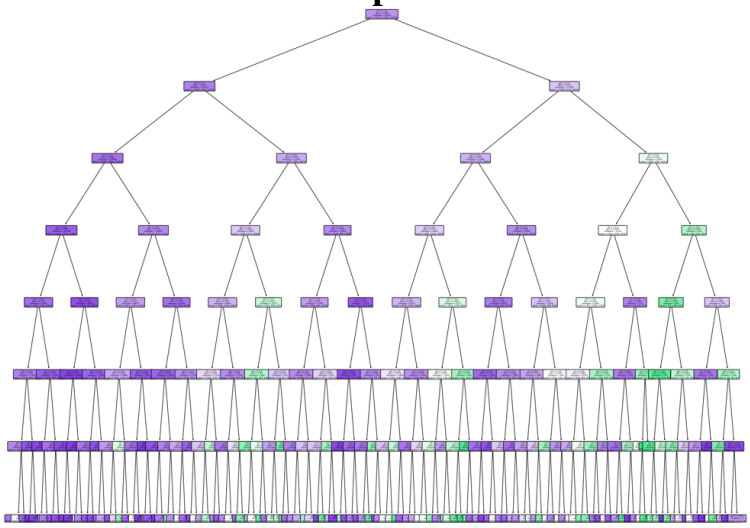
Max depth = 5



Max depth = 6



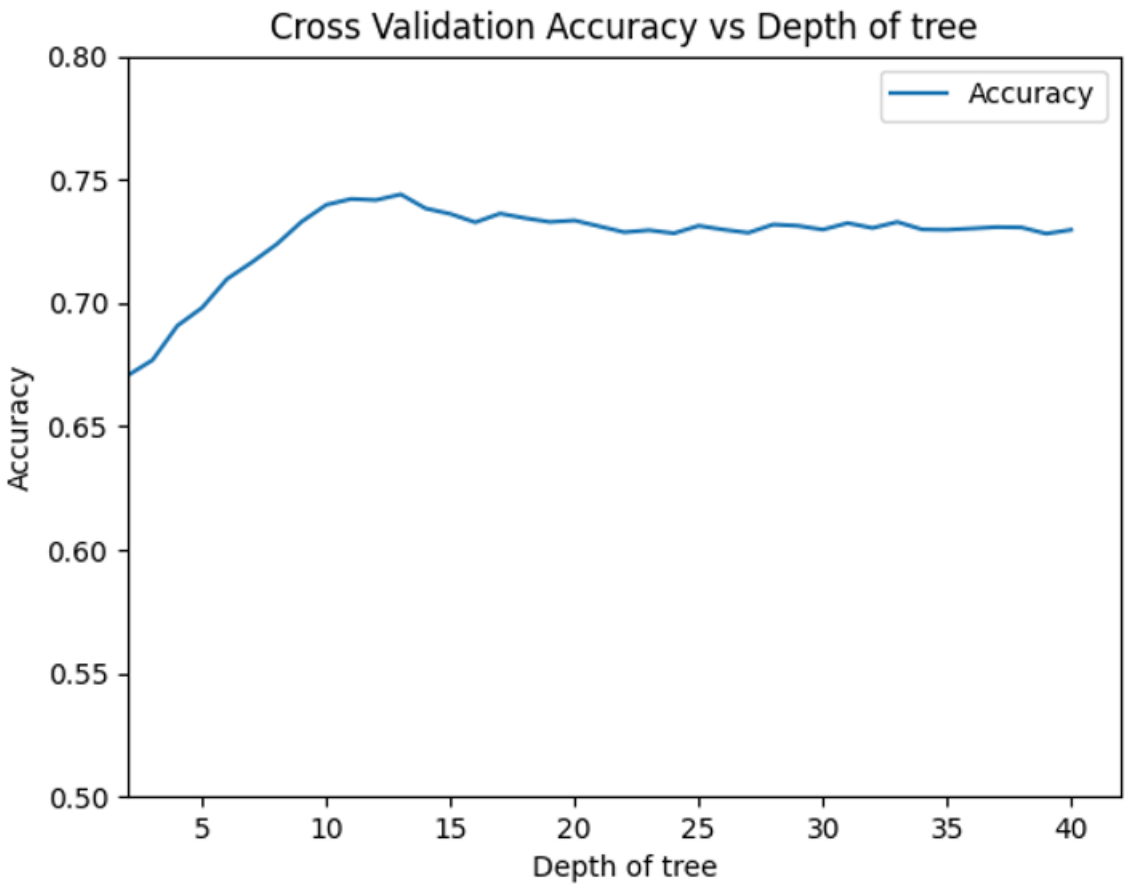
Max depth = 7



- *Accuracy table*

Max_depth	None	2	3	4	5	6	7
Accuracy	0.7370	0.6660	0.6647	0.6789	0.6802	0.6952	0.6972

- *Chart*



- *Comments*

- Accuracy is proportional to the maximum depth, and at a peak when maximum depth is none. (73,7%)
- Accuracy only increases up to a certain threshold and does not increase further.

II. REFERENCES

[1]. **Decision Tree document**

[2]. **How to visualizing decision tree**

[3]. **Decision Tree Classifier example**

[4]. **Analysis and Classification of mushrooms**

III. GRADING

No.	Specifications	Scores(%)	Done
1	Preparing the datasets	20	X
2	Building the decision tree classifiers	20	X
3	Evaluating the decision tree classifiers		
	Classification report and confusion matrix	20	X
	Comments	10	X
4	The depth and accuracy of a decision tree		
	Trees, tables, and charts	20	X
	Comments	10	X
Total		**100**	