



**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-
TPHCM
KHOA CÔNG NGHỆ THÔNG TIN**

**BÁO CÁO
Môn: Nhập môn AI**

Chủ đề: Lab 2 Decision Tree.

Họ tên: Kuo Yung Sheng

MSSV: 21127684

Lớp: 21CLC07

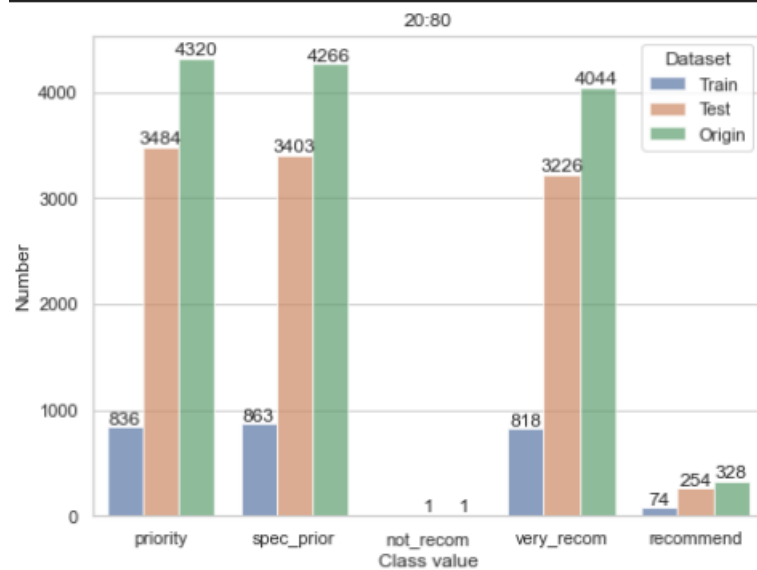
Thành phố Hồ Chí Minh, ngày 19 tháng 6 năm 2023

Mục lục

1.2.1 Preparing the data sets	2
1.2.2 Building the decision tree classifiers.....	2
1.2.3 Evaluating the decision tree classifiers.....	2
1.2.4 Evaluating the decision tree classifiers	4
2 Tài liệu tham khảo.....	4

1.2.1 Preparing the data sets

Dùng `train_test_split` để tách các dữ liệu. Các dữ liệu được tách sẽ bỏ vào array subset. `Sns` and `plt` để visualize dữ liệu.



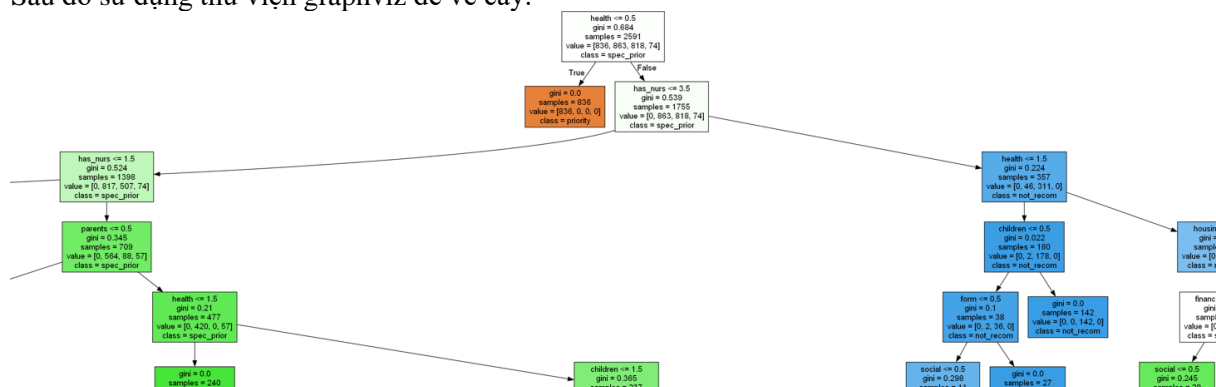
Visualize the distributions of classes.

Vào file 21127684.ipynb để xem visualize của 16 subset.

1.2.2 Building the decision tree classifiers

Sử dụng `label_encoder.fit_transform` để encode dữ liệu về dạng số.

Sau đó sử dụng thư viện graphviz để vẽ cây.



Vào folder decisionTree để xem rõ hơn về cây.

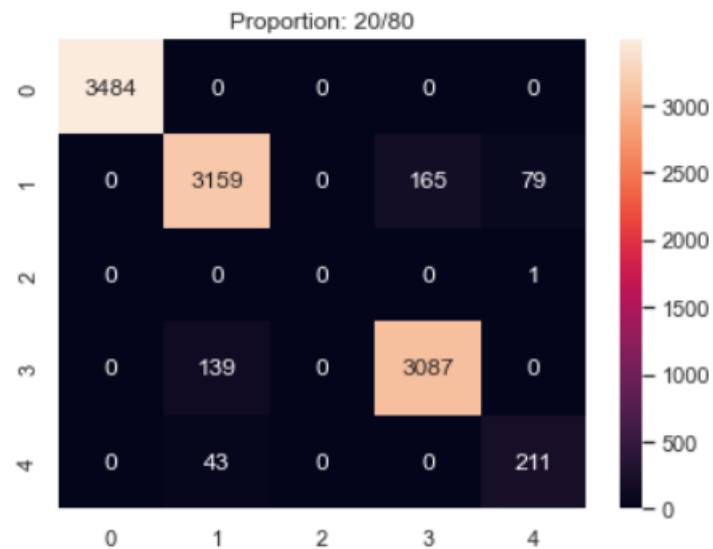
1.2.3 Evaluating the decision tree classifiers

Encode tập test sau đó dùng tập test đó để predict ta thu được kết quả.

Sử dụng `classification_report` dựa trên kết quả `predict` để xuất ra report.

Sử dụng confusion matrix để tính ra array sau đó bỏ array vào sns.heatmap để visualize.

-----Proportion: 20/80-----				
	precision	recall	f1-score	support
priority	1.00	1.00	1.00	3484
spec_prior	0.95	0.93	0.94	3403
not_recom	0.00	0.00	0.00	1
very_recom	0.95	0.96	0.95	3226
recommend	0.73	0.83	0.77	254
accuracy			0.96	10368
macro avg	0.72	0.74	0.73	10368
weighted avg	0.96	0.96	0.96	10368



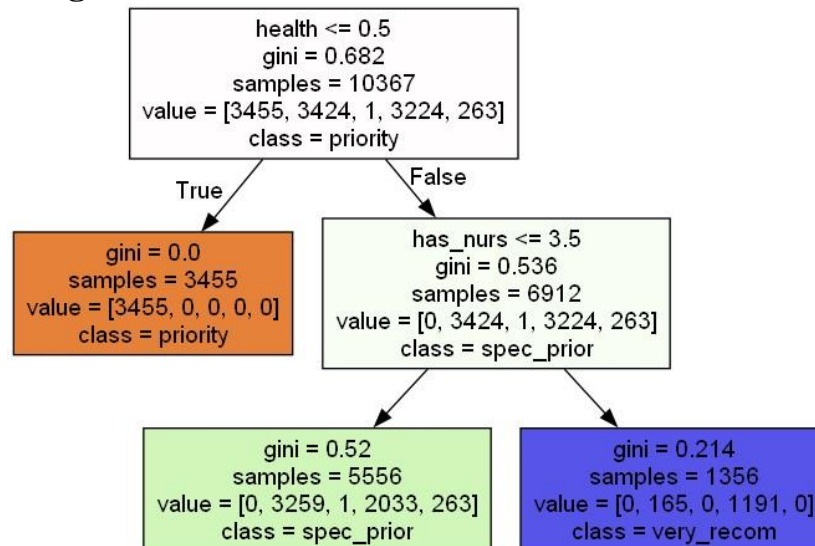
Vào file 21127684.ipynb để xem visualize của 16 subset.

Nhận xét:

-----Proportion: 80/20-----				
	precision	recall	f1-score	support
spec_prior	1.00	1.00	1.00	865
not_recom	0.99	1.00	0.99	842
priority	1.00	0.99	1.00	820
very_recom	0.95	0.97	0.96	65
accuracy			1.00	2592
macro avg	0.99	0.99	0.99	2592
weighted avg	1.00	1.00	1.00	2592

Với tỉ lệ 80/20 thì độ chính xác lên đến 99% và số lượng predict là 2592. Đây là kết quả tốt nhất trong những tỉ lệ.

1.2.4 Evaluating the decision tree classifiers



Ảnh cây vẽ được ở depth = 2

Vào folder DepthAccuracy để xem những cây ở độ sâu khác.

```
-----depth None-----
0.9988425925925926
-----depth 2-----
0.7648533950617284
-----depth 3-----
0.8119212962962963
-----depth 4-----
0.8586998456790124
-----depth 5-----
0.8776041666666666
-----depth 6-----
0.8894675925925926
-----depth 7-----
0.9236111111111112
```

Nhận xét:

Độ chính xác ở depth None cao nhất (không giới hạn độ sâu).

Khi giới hạn độ sâu thì ta thấy rằng độ sâu càng lớn thì độ chính xác càng cao.

2 Tài liệu tham khảo

[https://www.kaggle.com/code/prashant111/decision-tree-classifier-tutorial#2.-Classification-and-Regression-Trees-\(CART\)-](https://www.kaggle.com/code/prashant111/decision-tree-classifier-tutorial#2.-Classification-and-Regression-Trees-(CART)-)

<https://seaborn.pydata.org/generated/seaborn.heatmap.html>

https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.plot.html