

**BÀI TẬP THỰC HÀNH**  
**MÔN TRÍ TUỆ NHÂN TẠO**  
**BÀI TOÁN PHÂN LỚP VỚI CÂY QUYẾT ĐỊNH**

**Bài tập 1:** Cho dữ liệu như sau:

Age	Experience	Rank	Nationality	Go
36	10	9	UK	NO
42	12	4	USA	NO
23	4	6	N	NO
52	4	4	USA	NO
43	21	8	USA	YES
44	14	5	UK	NO
66	3	7	N	YES
35	14	9	UK	YES
52	13	7	N	YES
35	5	9	N	YES
24	3	5	USA	NO
18	3	7	UK	YES
45	9	9	UK	YES

- Tiến hành đọc file csv vào DataFrame
- Hiển thị lại dữ liệu đã đọc
- Hiển thị toàn bộ các loại dữ liệu cho cột Nationality (giá trị không trùng nhau)

```
df_nation = df['Nationality'].unique()
print(df_nation)
```

- Hiển thị toàn bộ các loại dữ liệu cho cột Go (giá trị không trùng nhau) (tương tự câu c)
- Thay đổi giá trị cho Nationality theo thứ tự: UK: 0, USA: 1, N: 2

```
map_value_nation = {}
index = 0
for nation in df_nation:
    new_value = {nation:index}
    map_value_nation.update(new_value)
    index = index + 1
print(map_value_nation)
```

```
df['Nationality'] = df['Nationality'].map(map_value_nation)
print(df)
```

Kết quả thu được:

	Age	Experience	Rank	Nationality	Go
0	36	10	9	0	NO
1	42	12	4	1	NO
2	23	4	6	2	NO
3	52	4	4	1	NO
4	43	21	8	1	YES
5	44	14	5	0	NO
6	66	3	7	2	YES
7	35	14	9	0	YES
8	52	13	7	2	YES
9	35	5	9	2	YES
10	24	3	5	1	NO
11	18	3	7	0	YES
12	45	9	9	0	YES

- f. Thay đổi giá trị cho Go theo thứ tự: Yes: 1, No: 0  
Thực hiện tương tự câu e và kết quả thu được cuối cùng:

	Age	Experience	Rank	Nationality	Go
0	36	10	9	0	0
1	42	12	4	1	0
2	23	4	6	2	0
3	52	4	4	1	0
4	43	21	8	1	1
5	44	14	5	0	0
6	66	3	7	2	1
7	35	14	9	0	1
8	52	13	7	2	1
9	35	5	9	2	1
10	24	3	5	1	0
11	18	3	7	0	1
12	45	9	9	0	1

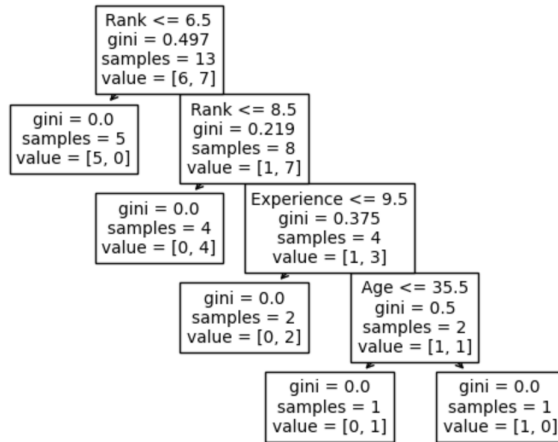
- g. Chia DataFrame X (feature columns) gồm các cột: Age, Experience, Rank, Nationality, DataFrame Y (target column) gồm cột Go.  
h. Dùng DecisionTreeClassifier từ scikit-learn để huấn luyện.  
i. Hiển thị cây dự đoán đã huấn luyện.

```

|--- feature_2 <= 6.50
|   |--- class: 0
|--- feature_2 > 6.50
|   |--- feature_2 <= 8.50
|   |   |--- class: 1
|   |--- feature_2 > 8.50
|   |   |--- feature_1 <= 9.50
|   |   |   |--- class: 1
|   |   |--- feature_1 > 9.50
|   |   |   |--- feature_0 <= 35.50
|   |   |   |   |--- class: 1
|   |   |   |--- feature_0 > 35.50
|   |   |       |--- class: 0

```

Và



Có thể lưu kết quả cây quyết định bằng matplotlib:

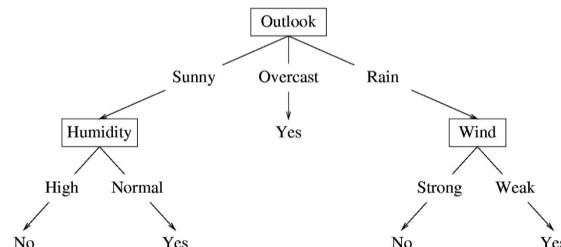
```
plt.figure()
tree.plot_tree(dtrees, filled=True)
plt.savefig('tree.png', format='png', bbox_inches = "tight")
```

- Tiến hành kiểm thử với case:  $[[40, 10, 7, 1]]$  và kiểm tra kết quả có phải thuộc class 1 (Yes) hay không?
- Tiến hành chia dữ liệu phân lớp theo phương pháp Hold out (2/3 training và 1/3 testing). Tiến hành xây dựng lại cây và kiểm thử kèm tính toán độ chính xác của mô hình. So sánh kết quả xây dựng cây từ scikit-learn với việc mô phỏng tự tính toán.

**Bài tập 2:** Chuẩn bị tập dữ liệu như sau:

Predictors				Response
Outlook	Temperature	Humidity	Wind	Class
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

- Tiến hành tạo cây quyết định DecisionTree và so sánh với kết quả:



- Kiểm thử với case:  
 $\langle outlook=sunny, temperature=hot, humidity=high, wind=weak \rangle$

--Hết--