

Chapter4 - Exercise 2: Play Gof

Cho dữ liệu play golf trong tập tin playgolf_data.xlsx.

Yêu cầu: Hãy đọc dữ liệu từ tập tin này, áp dụng Naive Bayes để thực hiện việc xác định có đi chơi golf hay không dựa trên các thông tin như: 'Outlook', 'Temperature', 'Humidity', 'Wind', 'Play Golf'

Yêu cầu:

1. Hãy chuẩn hóa dữ liệu cho phù hợp
2. Áp dụng Naive Bayes. Tìm kết quả
3. Kiểm tra độ chính xác
4. Xuất/ghi model
5. Đọc model
6. Cho dữ liệu Test: $X_{\text{test}} = [["Overcast", "Cool", "High", "Strong"], ["Sunny", "Cool", "High", "Weak"]] \Rightarrow Y_{\text{test}}$

```
In [1]: import pandas as pd
import numpy as np
```

```
In [2]: df = pd.read_excel('playgolf_data.xlsx', index_col = 0)
df.head()
```

Out[2]:

	Outlook	Temperature	Humidity	Wind	Play Golf
Day					
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes

```
In [3]: features = df.drop("Play Golf", axis=1)
target = df[["Play Golf"]]
```

```
In [4]: from sklearn.preprocessing import LabelEncoder
```

```
In [5]: features = pd.get_dummies(features)
features
```

Out[5]:

	Outlook_Overcast	Outlook_Rain	Outlook_Sunny	Temperature_Cool	Temperature_Hot	Temperature_Mild
Day						
1	0	0	1	0	1	0
2	0	0	1	0	1	0
3	1	0	0	0	1	0
4	0	1	0	0	0	0
5	0	1	0	1	0	0
6	0	1	0	1	0	0
7	1	0	0	1	0	0
8	0	0	1	0	0	0
9	0	0	1	1	0	0
10	0	1	0	0	0	0
11	0	0	1	0	0	0
12	1	0	0	0	0	0
13	1	0	0	0	1	0
14	0	1	0	0	0	0

```
In [6]: from sklearn.utils.validation import column_or_1d
le = LabelEncoder()
target = le.fit_transform(column_or_1d(target))
```

```
In [7]: # Sử dụng BernoulliNB
from sklearn.naive_bayes import BernoulliNB
```

```
In [8]: model1 = BernoulliNB()
model1.fit(features, target)
```

Out[8]: BernoulliNB(alpha=1.0, binarize=0.0, class_prior=None, fit_prior=True)

```
In [9]: # Kiểm tra độ chính xác
print("The prediction accuracy is: ",
      model1.score(features, target)*100, "%")
```

The prediction accuracy is: 92.85714285714286 %

```
In [10]: class_names = model1.classes_  
         class_names
```

```
Out[10]: array([0, 1])
```

```
In [11]: # Từ kết quả trên => chọn model1
```

```
In [12]: # Xuất model  
import pickle  
# Save to file in the current working directory  
pkl_filename = "playgoft_model.pkl"  
with open(pkl_filename, 'wb') as file:  
    pickle.dump(model1, file)
```

```
In [13]: with open(pkl_filename, 'rb') as file:  
         playgoft_model = pickle.load(file)
```

```
In [14]: # Outlook_Overcast Outlook_Rain Outlook_Sunny  
# Temperature_Cool Temperature_Hot Temperature_Mild  
# Humidity_High Humidity_Normal  
# Wind_Strong Wind_Weak  
# _____  
# X_test = ["Overcast", "Cool", "High", "Strong"],  
# ["Sunny", "Cool", "High", "Weak"]]  
X_test = [[1, 0, 0, 1, 0, 0, 1, 0, 1, 0], [0, 0, 1, 1, 0, 0, 1, 0, 0, 1]]  
y_pred = playgoft_model.predict(X_test)  
y_pred
```

```
Out[14]: array([1, 0])
```

```
In [ ]:
```