



Exercise 7: Credit Scoring ¶

Cho dữ liệu Credit Scoring trong tập tin credit_scoring.txt.

Yêu cầu: Hãy đọc dữ liệu từ tập tin này, áp dụng Naive Bayes để thực hiện việc xác định quyết định trong tính điểm tín dụng là Y hay N dựa trên các thông tin như BUSAGE, DAYSDELQ :

Tạo X_train, X_test, y_train, y_test từ dữ liệu đọc được với tỷ lệ dữ liệu test là 0.2

1. Áp dụng thuật toán Naive Bayes
2. Tìm kết quả
3. Kiểm tra độ chính xác
4. Với x_new = [[20, 28], [50, 59]] kết quả là Y hay N

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

```
In [2]: # đường dẫn tập tin: http://saedsayad.com/datasets/credit_scoring.txt
df = pd.read_csv('credit_scoring.txt', sep="\t")
df.head()
```

Out[2]:

	BUSAGE	DAYSDELQ	DEFAULT
0	87	2	N
1	89	2	N
2	90	2	N
3	90	2	N
4	101	2	N

```
In [3]: features = df[["BUSAGE", "DAYSDELQ"]]
target = df[["DEFAULT"]]
```

```
In [4]: features.head()
```

Out[4]:

	BUSAGE	DAYSDELQ
0	87	2
1	89	2
2	90	2
3	90	2
4	101	2

In [5]: `target.head()`

Out[5]:

	DEFAULT
0	N
1	N
2	N
3	N
4	N

In [6]: `from sklearn.model_selection import train_test_split`
`X_train, X_test, y_train, y_test = train_test_split(features, target, test_size=0.`

In [7]: `from sklearn.naive_bayes import GaussianNB`
`import numpy as np`
`from sklearn.utils.validation import column_or_1d`

#Create a Gaussian Classifier

`model = GaussianNB()`

Train the model using the training sets

`model.fit(features, target)`

c:\program files\python36\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
`y = column_or_1d(y, warn=True)`

Out[7]: `GaussianNB(priors=None)`

In [8]: `import numpy as np`
`# Kiểm tra độ chính xác`
`print("The prediction accuracy is: ", model.score(features,np.array(target))*100,"`

The prediction accuracy is: 87.0 %

In [9]: `class_names = model.classes_`
`class_names`

Out[9]: `array(['N', 'Y'], dtype='<U1')`

In [10]: `X_test = [[20, 28], [50,59]]`
`y_pred = model.predict(X_test)`
`y_pred`

Out[10]: `array(['N', 'Y'], dtype='<U1')`

In []:

