

 17

Pegando os arrays do data frame.

Durante esse capítulo, vimos que o nosso algoritmo de classificação recebe dois parâmetros:

```
from sklearn.naive_bayes import MultinomialNB  
modelo = MultinomialNB()  
modelo.fit(treino_dados, treino_marcacoes)
```

Porém, as variáveis `treino_dados` e `treino_marcacoes` são arrays, ou seja, não podemos simplesmente enviar os data frames para o nosso algoritmo. Nesse caso, precisaremos pegar os valores do nossos data frames e transformá-los em arrays. Podemos retornar esses arrays do data frame pedindo os valores dele, nesse caso, `Xdummies.values` e `Ydummies.values`. Porém, observe que o array que retornamos, é justamente o nosso `X` e `Y`. De acordo com essa situação, precisamos deixar explícito quais são os valores são data frames, ou seja, renomear as variáveis referentes aos data frames com o sufixo "df":

```
import pandas as pd  
df = pd.read_csv('buscas.csv')  
X_df = df[['home', 'busca', 'logado']]  
Y_df = df['comprou']  
  
Xdummies_df = pd.get_dummies(X_df).astype(int)  
Ydummies_df = Y_df
```

Agora podemos devolver os arrays de cada um dos dummies para as variáveis `X` e `Y`:

```
X = Xdummies_df.values  
Y = Ydummies_df.values
```

Por fim, imprima os valores de `X` e `Y`. Cole o código e o resultado.

Responda

INserir CÓDIGO		FORMATAÇÃO
<div style="height: 400px; width: 100%;"></div>		

