

```
%matplotlib inline
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns; sns.set()
from sklearn.datasets import fetch_20newsgroups
data = fetch_20newsgroups()
data.target_names
```

☞ Downloading 20news dataset. This may take a few minutes.  
 Downloading dataset from <https://ndownloader.figshare.com/files/5975967> (14 MB)

```
['alt.atheism',
 'comp.graphics',
 'comp.os.ms-windows.misc',
 'comp.sys.ibm.pc.hardware',
 'comp.sys.mac.hardware',
 'comp.windows.x',
 'misc.forsale',
 'rec.autos',
 'rec.motorcycles',
 'rec.sport.baseball',
 'rec.sport.hockey',
 'sci.crypt',
 'sci.electronics',
 'sci.med',
 'sci.space',
 'soc.religion.christian',
 'talk.politics.guns',
 'talk.politics.mideast',
 'talk.politics.misc',
 'talk.religion.misc']
```

```
categories = ['alt.atheism', 'comp.graphics', 'comp.os.ms-windows.misc',
              'comp.sys.ibm.pc.hardware',
              'comp.sys.mac.hardware', 'comp.windows.x',
              'misc.forsale', 'rec.autos', 'rec.motorcycles',
              'rec.sport.baseball', 'rec.sport.hockey', 'sci.crypt',
              'sci.electronics', 'sci.med', 'sci.space',
              'soc.religion.christian', 'talk.politics.guns',
              'talk.politics.mideast', 'talk.politics.misc',
              'talk.religion.misc']
```

```
train = fetch_20newsgroups(subset='train', categories=categories)
test = fetch_20newsgroups(subset='test', categories=categories)
print(train.data[5])
```

```
In article <1r1eu1$4t@transfer.stratus.com> cdt@sw.stratus.com (C. D. Tavares) writes:
>In article <1993Apr20.083057.16899@ousrvr.oulu.fi>, dfo@yttoulu.tko.vtt.fi (Foxvog Douglas) writes:
>> In article <1qv87v$4j3@transfer.stratus.com> cdt@sw.stratus.com (C. D. Tavares) writes:
>> >In article <C5n3GI.F8F@ulowell.ulowell.edu>, jrutledg@cs.ulowell.edu (John Lawrence Rutledge) writes:
>>
>> >> The massive destructive power of many modern weapons, makes the
>> >> cost of an accidental or crimial usage of these weapons to great.
>> >> The weapons of mass destruction need to be in the control of
>> >> the government only. Individual access would result in the
>> >> needless deaths of millions. This makes the right of the people
>> >> to keep and bear many modern weapons non-existant.

>> >Thanks for stating where you're coming from. Needless to say, I
>> >disagree on every count.

>> You believe that individuals should have the right to own weapons of
>> mass destruction? I find it hard to believe that you would support a
>> neighbor's right to keep nuclear weapons, biological weapons, and nerve
>> gas on his/her property.
```

>> gas on his/her property.

>> If we cannot even agree on keeping weapons of mass destruction out of  
>> the hands of individuals, can there be any hope for us?

>I don't sign any blank checks.

Of course. The term must be rigidly defined in any bill.

>When Doug Foxvog says "weapons of mass destruction," he means CBW and  
>nukes. When Sarah Brady says "weapons of mass destruction" she means  
>Street Sweeper shotguns and semi-automatic SKS rifles.

I doubt she uses this term for that. You are using a quote allegedly  
from her, can you back it up?

>When John

>Lawrence Rutledge says "weapons of mass destruction," and then immediately  
>follows it with:

>>> The US has thousands of people killed each year by handguns,  
>>> this number can easily be reduced by putting reasonable restrictions  
>>> on them.

>...what does Rutledge mean by the term?

I read the article as presenting first an argument about weapons of mass  
destruction (as commonly understood) and then switching to other topics.  
The first point evidently was to show that not all weapons should be  
allowed, and then the later analysis was, given this understanding, to  
consider another class.

>[cdt@rocket.sw.stratus.com](mailto:cdt@rocket.sw.stratus.com) --If you believe that I speak for my company,  
>OR [cdt@vos.stratus.com](mailto:cdt@vos.stratus.com) write today for my special Investors' Packet...

--

doug foxvog  
[douglas.foxvog@vtt.fi](mailto:douglas.foxvog@vtt.fi)

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.pipeline import make_pipeline
model = make_pipeline(TfidfVectorizer(), MultinomialNB())
model.fit(train.data, train.target)
labels = model.predict(test.data)
```

```
from sklearn.metrics import confusion_matrix
mat = confusion_matrix(test.target, labels)
sns.heatmap(mat.T, square=True, annot=True, fmt='d', cbar=False, xticklabels=train.target_names, yticklabels=train.target_names)
plt.xlabel('true label')
plt.ylabel('predicted label');
```

predicted label	alt.atheism	661	0	0	0	1	0	0	0	0	0	0	0	0	2	0	2	0	0	233
	comp.graphics	025	14	5	3	2	1	1	1	0	0	0	2	4	3	2	0	0	1	0
	comp.os.ms-windows.misc	015	25	8	1	8	1	7	3	0	0	0	0	1	2	0	0	0	0	0
	comp.sys.ibm.pc.hardware	112	45	0	23	13	3	1	3	1	0	0	0	1	7	1	1	0	1	0
	comp.sys.mac.hardware	0	9	3	17	9	8	2	1	2	0	0	1	0	5	1	0	0	0	0
	comp.windows.x	118	9	1	0	29	8	1	0	0	0	0	1	0	3	3	0	0	0	0
	misc.forsale	0	1	0	3	3	1	7	1	4	2	0	0	1	2	1	0	0	2	0
	rec.autos	0	2	2	6	8	0	19	6	10	4	1	3	8	0	2	0	0	1	0
	rec.motorcycles	1	1	1	1	1	4	3	7	0	0	0	7	2	1	1	1	0	0	0
	rec.sport.baseball	1	5	3	0	3	1	4	2	0	8	5	4	0	1	3	0	0	1	0
	rec.sport.hockey	1	2	2	2	1	0	6	2	0	2	8	1	0	2	4	1	0	0	1
	sci.crypt	34	125	19	16	23	5	4	4	0	1	8	7	8	1	6	0	10	2	11
	sci.electronics	0	4	1	13	8	0	12	1	0	0	0	1	2	3	5	1	0	0	0
	sci.med	6	0	0	0	0	1	6	1	0	0	0	0	3	2	2	1	0	0	1
	sci.space	3	6	6	5	2	4	3	3	0	2	1	0	1	6	8	5	2	1	0
	soc.religion.christian	23	523	3	8	10	9	3	8	9	5	3	155	219	8	6	24	35	3	3
	talk.politics.guns	4	4	2	1	3	2	3	4	2	1	0	1	2	6	4	0	84	3	1
	talk.politics.mideast	8	1	0	0	0	0	0	0	0	1	0	0	1	4	0	0	1	84	5
	talk.politics.misc	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	2
	talk.religion.misc	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35

eism  
 hics  
 misc  
 ware  
 ws.x  
 rsale  
 utos  
 ycles  
 eball  
 :crypt  
 nics  
 med  
 pace  
 stan  
 guns  
 least  
 misc

```
def predict_category(s, train=train, model=model):
    pred=model.predict([s])
    return train.target_names[pred[0]]
```

0: sy ip 50 18

predict\_category('Allah')

'alt.atheism'

predict\_category('Space station')

'sci.space'

predict\_category('President of Bangladesh')

'talk.politics.misc'