```
import pandas as pd
import nltk
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import BernoulliNB

df = pd.read_csv('/content/federalist.csv')
df['author'] = df.author.astype('category')
result = df.head(10)
print(result)
print(df['author'].value_counts())
```

```
      author                                               text
0  HAMILTON  FEDERALIST. No. 1 General Introduction For the...
1       JAY  FEDERALIST No. 2 Concerning Dangers from Forei...
2       JAY  FEDERALIST No. 3 The Same Subject Continued (C...
3       JAY  FEDERALIST No. 4 The Same Subject Continued (C...
4       JAY  FEDERALIST No. 5 The Same Subject Continued (C...
5  HAMILTON  FEDERALIST No. 6 Concerning Dangers from Disse...
6  HAMILTON  FEDERALIST. No. 7 The Same Subject Continued (...
7  HAMILTON  FEDERALIST No. 8 The Consequences of Hostiliti...
8  HAMILTON  FEDERALIST No. 9 The Union as a Safeguard Agai...
9   MADISON  FEDERALIST No. 10 The Same Subject Continued (...
HAMILTON               49
MADISON                15
HAMILTON OR MADISON    11
JAY                     5
HAMILTON AND MADISON    3
Name: author, dtype: int64
```

```
stop = stopwords.words('english')

pat = r'\b(?:{})\b'.format('|'.join(stop))
df['text'] = df['text'].str.replace(pat, '')
df['text'] = df['text'].str.replace(r'\s+', ' ')
print(df)

train=df.sample(frac=0.8,random_state=1234)
test=df.drop(train.index)
print(train.shape)
print(test.shape)

v = TfidfVectorizer()
x = v.fit_transform(test['text'])
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:4: FutureWarning: The defa
  after removing the cwd from sys.path.
      author                                               text
0   HAMILTON  FEDERALIST. No. 1 General Introduction For Ind...
1        JAY  FEDERALIST No. 2 Concerning Dangers Foreign Fo...
2        JAY  FEDERALIST No. 3 The Same Subject Continued (C...
3        JAY  FEDERALIST No. 4 The Same Subject Continued (C...
4        JAY  FEDERALIST No. 5 The Same Subject Continued (C...
..       ...                                                ...
78  HAMILTON  FEDERALIST No. 79 The Judiciary Continued From...
79  HAMILTON  FEDERALIST No. 80 The Powers Judiciary From Mc...
80  HAMILTON  FEDERALIST. No. 81 The Judiciary Continued, Di...
81  HAMILTON  FEDERALIST No. 82 The Judiciary Continued From...
82  HAMILTON  FEDERALIST No. 83 The Judiciary Continued Rela...

[83 rows x 2 columns]
(66, 2)
(17, 2)
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:5: FutureWarning: The defa
  """
```

```
print(x.toarray())

yy = v.transform(test['text'])
xx = v.transform(train['text'])
print(xx.shape)
print(yy.shape)


c = BernoulliNB()
```

```
[[0.         0.         0.03350468 ... 0.         0.         0.        ]
 [0.         0.         0.         ... 0.         0.02298121 0.        ]
 [0.         0.         0.         ... 0.         0.         0.        ]
 ...
 [0.         0.         0.         ... 0.         0.         0.        ]
 [0.         0.         0.         ... 0.         0.         0.        ]
 [0.         0.         0.         ... 0.         0.         0.        ]]
(66, 4065)
(17, 4065)
```

# New Section

# New Section