

▾ Fake News DEtection

By : Sampurna Dey

Linkedin: www.linkedin.com/in/sampurna-dey-060573129

Github: <https://github.com/sampurna-project>

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
import re
import string
```

+ Code

+ Text

```
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import LogisticRegression
```

```
fake_data = pd.read_csv("Fake.csv")
real_data = pd.read_csv("True.csv")
```

fake_data

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017
...
23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016
...

real_data

	title	text	subject	date
--	-------	------	---------	------

```
fake_data['label'] = 0
real_data['label'] = 1
```

```
news_data = pd.concat([fake_data, real_data], ignore_index=True)
# ...
```

```
news_data
```

	title	text	subject	date	label
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
...
	'Fully committed' NATO	BRUSSELS (Reuters) -		August 22	

```
news_data = news_data.drop(['title', 'subject', 'date'], axis=1)
```

```
news_data
```

	text	label
0	Donald Trump just couldn t wish all Americans ...	0
1	House Intelligence Committee Chairman Devin Nu...	0
2	On Friday, it was revealed that former Milwauk...	0
3	On Christmas day, Donald Trump announced that ...	0
4	Pope Francis used his annual Christmas Day mes...	0
...
44893	BRUSSELS (Reuters) - NATO allies on Tuesday we...	1
44894	LONDON (Reuters) - LexisNexis, a provider of l...	1
44895	MINSK (Reuters) - In the shadow of disused Sov...	1
44896	MOSCOW (Reuters) - Vatican Secretary of State ...	1
44897	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	1

```
44898 rows x 2 columns
```

```
def remove_punctuations(text):
    text = re.sub('[^a-zA-Z]', ' ', text)
    text = re.sub(r'\s+[a-zA-Z]\s+', ' ', text)
    text = re.sub(r'\s+', ' ', text)
    return text
```

```
news_data['text'] = news_data['text'].apply(remove_punctuations)
```

```
news_data['text'] = news_data['text'].apply(lambda x: x.lower())
```

```
import nltk
```

```
nltk.download('stopwords')
```

```
[nltk_data] Error loading stopwords: <urlopen error [WinError 10060] A
[nltk_data] connection attempt failed because the connected party
[nltk_data] did not properly respond after a period of time, or
```

```
[nltk_data]      established connection failed because connected host
[nltk_data]      has failed to respond>
False
```

```
from nltk.corpus import stopwords
```

```
tfidf = TfidfVectorizer()
```

```
x = tfidf.fit_transform(news_data['text'])
```

```
x_train, x_test, y_train, y_test = train_test_split(x, news_data['label'], test_size=0.2, random_state=42)
```

```
nb = MultinomialNB()
```

```
nb.fit(x_train, y_train)
```

```
▼ MultinomialNB
MultinomialNB()
```

```
y_pred_nb = nb.predict(x_test)
```

```
print('Accuracy score of Naive Bayes:', accuracy_score(y_test, y_pred_nb))
```

```
Accuracy score of Naive Bayes: 0.9360801781737194
```

```
print('Confusion matrix of Naive Bayes:\n', confusion_matrix(y_test, y_pred_nb))
```

```
Confusion matrix of Naive Bayes:
[[4502  231]
 [ 343 3904]]
```

```
lr = LogisticRegression()
```

```
lr.fit(x_train, y_train)
```

```
▼ LogisticRegression
LogisticRegression()
```

```
y_pred_lr = lr.predict(x_test)
```

```
print('Accuracy score of Logistic Regression:', accuracy_score(y_test, y_pred_lr))
```

```
Accuracy score of Logistic Regression: 0.9863028953229399
```

```
print('Confusion matrix of Logistic Regression:\n', confusion_matrix(y_test, y_pred_lr))
```

```
Confusion matrix of Logistic Regression:
[[4665   68]
 [   55 4192]]
```

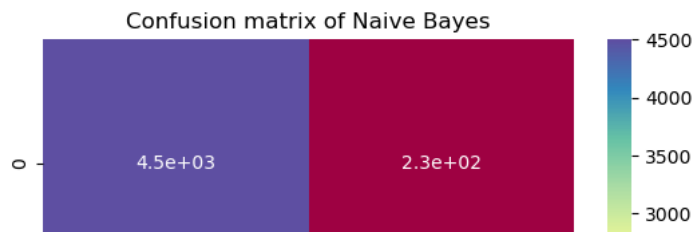
```
sns.heatmap(confusion_matrix(y_test, y_pred_nb), annot=True, cmap='Spectral')
```

```
plt.title('Confusion matrix of Naive Bayes')
```

```
plt.xlabel('Predicted labels')
```

```
plt.ylabel('True labels')
```

```
plt.show()
```



```
sns.heatmap(confusion_matrix(y_test, y_pred_lr), annot=True, cmap='mako')
plt.title('Confusion matrix of Logistic Regression')
plt.xlabel('Predicted labels')
plt.ylabel('True labels')
```

Text(50.72222222222214, 0.5, 'True labels')

