<> Code   ⊙ Issues   ⑂ Pull requests   •••

Phase2 /   Main.py   in   `main`   Cancel changes   Commit changes...

Edit   Preview

Spaces ⇕   4 ⇕   No wrap ⇕

```python
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

from nltk.corpus import stopwords
import nltk
nltk.download('stopwords')

from sklearn.pipeline import Pipeline

from sklearn.naive_bayes import BernoulliNB , MultinomialNB , GaussianNB

from sklearn.metrics import accuracy_score

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))



filepath = '/kaggle/input/sms-spam-collection-dataset/spam.csv'
data_import = pd.read_csv(filepath , encoding = 'ISO-8859-1')
data_import.head()
df =  data_import.drop(['Unnamed: 2' , 'Unnamed: 3' , 'Unnamed: 4'] , axis = 1)
df.head()
sw = stopwords.words('english')

def stopword(text) :
    txt = [word.lower() for word in text.split() if word.lower() not in sw]
    return txt

df['v2'] = df['v2'].apply(stopword)

df.head()

from nltk.stem.snowball import SnowballStemmer
```

Use `Control + Shift + m` to toggle the `tab` key moving focus. Alternatively, use `esc` then `tab` to move to the next interactive element on the page.

```python
39
40    def stemming(text) :
41        text = [ss.stem(word) for word in text if word.split()]
42        return "".join(text)
43
44    df['v2'] = df['v2'].apply(stemming)
45
46
47    df.head()
```