In [1]:

```
import numpy as np
import pandas as pd
```

In [2]:

```
df=pd.read_csv("C:/Users/Kadavath Latha/OneDrive/Desktop/stress.csv")
df.head()
```

Out[2]:

	subreddit	post_id	sentence_range	text	id	label	confidence	social_times
0	ptsd	8601tu	(15, 20)	He said he had not felt that way before, sugge...	33181	1	0.8	15216
1	assistance	8lbrx9	(0, 5)	Hey there r/assistance, Not sure if this is th...	2606	0	1.0	15270
2	ptsd	9ch1zh	(15, 20)	My mom then hit me with the newspaper and it s...	38816	1	0.8	15359
3	relationships	7rorpp	[5, 10]	until i met my new boyfriend, he is amazing, h...	239	1	0.6	15164
4	survivorsofabuse	9p2gbc	[0, 5]	October is Domestic Violence Awareness Month a...	1421	1	0.8	15398

5 rows × 116 columns

In [3]:

```
df.describe()
```

Out[3]:

	id	label	confidence	social_timestamp	social_karma	syntax_ari	le
count	2838.000000	2838.000000	2838.000000	2.838000e+03	2838.000000	2838.000000	2
mean	13751.999295	0.524313	0.808972	1.518107e+09	18.262156	4.684272	
std	17340.161897	0.499497	0.177038	1.552209e+07	79.419166	3.316435	
min	4.000000	0.000000	0.428571	1.483274e+09	0.000000	-6.620000	
25%	926.250000	0.000000	0.600000	1.509698e+09	2.000000	2.464243	
50%	1891.500000	1.000000	0.800000	1.517066e+09	5.000000	4.321886	
75%	25473.750000	1.000000	1.000000	1.530898e+09	10.000000	6.505657	
max	55757.000000	1.000000	1.000000	1.542592e+09	1435.000000	24.074231	

8 rows × 112 columns

In [4]:

```
df.isnull().sum()
```

Out[4]:

```
subreddit                    0
post_id                      0
sentence_range               0
text                         0
id                           0
                            ..
lex_dal_avg_pleasantness     0
social_upvote_ratio          0
social_num_comments          0
syntax_fk_grade              0
sentiment                    0
Length: 116, dtype: int64
```

In [17]:

```
import nltk
import re
from nltk. corpus import stopwords
import string
nltk. download( 'stopwords' )
stemmer = nltk. SnowballStemmer("english")
stopword=set (stopwords . words ( 'english' ))

def clean(text):
    text = str(text) . lower()
    text = re. sub('\[.*?\]',' ',text)
    text = re. sub('https?://\S+/www\. \S+', ' ', text)
    text = re. sub('<. *?>+', ' ', text)
    text = re. sub(' [%s]' % re. escape(string. punctuation), ' ', text)
    text = re. sub(' \n',' ', text)
    text = re. sub(' \w*\d\w*' ,' ', text)
    text = [word for word in text. split(' ') if word not in stopword]
    text =" ". join(text)
    text = [stemmer . stem(word) for word in text. split(' ') ]
    text = " ". join(text)
    return text
df [ "text"] = df["text"]. apply(clean)
```

```
[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>

---------------------------------------------------------------------------
LookupError                               Traceback (most recent call last)
~\anaconda4\lib\site-packages\nltk\corpus\util.py in __load(self)
     83                 try:
---> 84                     root = nltk.data.find(f"{self.subdir}/{zip_nam
e}")
     85                 except LookupError:

~\anaconda4\lib\site-packages\nltk\data.py in find(resource_name, paths)
    582     resource_not_found = f"\n{sep}\n{msg}\n{sep}\n"
--> 583     raise LookupError(resource_not_found)
    584

LookupError:
**********************************************************************
  Resource stopwords not found.
  Please use the NLTK Downloader to obtain the resource:

  >>> import nltk
  >>> nltk.download('stopwords')

  For more information see: https://www.nltk.org/data.html (https://www.nlt
k.org/data.html)

  Attempted to load corpora/stopwords.zip/stopwords/

  Searched in:
    - 'C:\\Users\\KADAVATH LATHA/nltk_data'
    - 'C:\\Users\\KADAVATH LATHA\\anaconda4\\nltk_data'
```

```
    - 'C:\\Users\\KADAVATH LATHA\\anaconda4\\share\\nltk_data'
    - 'C:\\Users\\KADAVATH LATHA\\anaconda4\\lib\\nltk_data'
    - 'C:\\Users\\KADAVATH LATHA\\AppData\\Roaming\\nltk_data'
    - 'C:\\nltk_data'
    - 'D:\\nltk_data'
    - 'E:\\nltk_data'
**********************************************************************


During handling of the above exception, another exception occurred:

LookupError                               Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_47988\3462361060.py in <module>
      5 nltk. download( 'stopwords' )
      6 stemmer = nltk. SnowballStemmer("english")
----> 7 stopword=set (stopwords . words ( 'english' ))
      8
      9 def clean(text):

~\anaconda4\lib\site-packages\nltk\corpus\util.py in __getattr__(self, attr)
    119             raise AttributeError("LazyCorpusLoader object has no att
ribute '__bases__'")
    120
--> 121         self.__load()
    122         # This looks circular, but its not, since __load() changes o
ur
    123         # __class__ to something new:

~\anaconda4\lib\site-packages\nltk\corpus\util.py in __load(self)
     84                     root = nltk.data.find(f"{self.subdir}/{zip_nam
e}")
     85                 except LookupError:
---> 86                     raise e
     87
     88         # Load the corpus.

~\anaconda4\lib\site-packages\nltk\corpus\util.py in __load(self)
     79         else:
     80             try:
---> 81                 root = nltk.data.find(f"{self.subdir}/{self.__name}"
)
     82             except LookupError as e:
     83                 try:

~\anaconda4\lib\site-packages\nltk\data.py in find(resource_name, paths)
    581     sep = "*" * 70
    582     resource_not_found = f"\n{sep}\n{msg}\n{sep}\n"
--> 583     raise LookupError(resource_not_found)
    584
    585

LookupError:
**********************************************************************
  Resource stopwords not found.
  Please use the NLTK Downloader to obtain the resource:

  >>> import nltk
  >>> nltk.download('stopwords')

  For more information see: https://www.nltk.org/data.html (https://www.nlt
k.org/data.html)
```

```
  Attempted to load corpora/stopwords

  Searched in:
    - 'C:\\Users\\KADAVATH LATHA/nltk_data'
    - 'C:\\Users\\KADAVATH LATHA\\anaconda4\\nltk_data'
    - 'C:\\Users\\KADAVATH LATHA\\anaconda4\\share\\nltk_data'
    - 'C:\\Users\\KADAVATH LATHA\\anaconda4\\lib\\nltk_data'
    - 'C:\\Users\\KADAVATH LATHA\\AppData\\Roaming\\nltk_data'
    - 'C:\\nltk_data'
    - 'D:\\nltk_data'
    - 'E:\\nltk_data'
**********************************************************************
```

In [14]:

```
import matplotlib. pyplot as plt
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
text = " ". join(i for i in df. text)
stopwords = set (STOPWORDS)
wordcloud = WordCloud( stopwords=stopwords,background_color="white") . generate(text)
plt. figure(figsize=(10, 10) )
plt. imshow(wordcloud )
plt. axis("off")
plt. show( )
```

```
---------------------------------------------------------------------------
ModuleNotFoundError                       Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_47988\3535197414.py in <module>
      1 import matplotlib. pyplot as plt
----> 2 from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
      3 text = " ". join(i for i in df. text)
      4 stopwords = set (STOPWORDS)
      5 wordcloud = WordCloud( stopwords=stopwords,background_color="white")
. generate(text)

ModuleNotFoundError: No module named 'wordcloud'
```

In [7]:

```
from sklearn. feature_extraction. text import CountVectorizer
from sklearn. model_selection import train_test_split

x = np.array (df["text"])
y = np.array (df["label"])

cv = CountVectorizer ()
X = cv. fit_transform(x)
print(X)
xtrain, xtest, ytrain, ytest = train_test_split(X, y,test_size=0.33)
```

```
  (0, 4861)     2
  (0, 8815)     1
  (0, 4765)     1
  (0, 6897)     1
  (0, 4020)     1
  (0, 10235)    2
  (0, 11152)    1
  (0, 1204)     1
  (0, 9918)     1
  (0, 4567)     1
  (0, 8573)     1
  (0, 683)      1
  (0, 9399)     2
  (0, 10557)    1
  (0, 566)      1
  (0, 5210)     1
  (0, 11483)    1
  (0, 8191)     1
  (0, 5179)     1
  (0, 6027)     1
  (0, 6353)     1
  (0, 2767)     1
  (0, 10393)    4
  (0, 6126)     1
  (0, 10828)    2
  :     :
  (2837, 6579)  2
  (2837, 5627)  2
  (2837, 9122)  3
  (2837, 10405) 1
  (2837, 11127) 2
  (2837, 5739)  1
  (2837, 4838)  1
  (2837, 2995)  1
  (2837, 4936)  2
  (2837, 8882)  1
  (2837, 7996)  1
  (2837, 5198)  1
  (2837, 10227) 1
  (2837, 6783)  1
  (2837, 11399) 1
  (2837, 6618)  1
  (2837, 3434)  1
  (2837, 10521) 1
  (2837, 3234)  1
  (2837, 10086) 1
  (2837, 9265)  1
```

```
  (2837, 10519) 1
  (2837, 6497)  1
  (2837, 2959)  1
  (2837, 3728)  1
```

In [8]:

```
from sklearn.naive_bayes import BernoulliNB
model=BernoulliNB()
model.fit(xtrain,ytrain)
```

Out[8]:

```
BernoulliNB()
```

In [9]:

```
user=input("Enter the text")
data=cv.transform([user]).toarray()
output=model.predict(data)
print(output)
```

```
Enter the textjdsgyutriwohdknfhg8y
[0]
```

In [ ]:

```
```