In [4]:
```
import numpy as np
import pandas as pd
```

In [6]:
```
df=pd.read_csv('B:\KARTHICK\stress.csv')
df.head()
```

Out[6]:

	subreddit	post_id	sentence_range	text	id	label	confidence	social_timestamp	social_ka
0	ptsd	8601tu	(15, 20)	He said he had not felt that way before, sugge...	33181	1	0.8	1521614353	
1	assistance	8lbrx9	(0, 5)	Hey there r/assistance, Not sure if this is th...	2606	0	1.0	1527009817	
2	ptsd	9ch1zh	(15, 20)	My mom then hit me with the newspaper and it s...	38816	1	0.8	1535935605	
3	relationships	7rorpp	[5, 10]	until i met my new boyfriend, he is amazing, h...	239	1	0.6	1516429555	
4	survivorsofabuse	9p2gbc	[0, 5]	October is Domestic Violence Awareness Month a...	1421	1	0.8	1539809005	

5 rows × 116 columns

In [8]:
```
df.describe()
```

Out[8]:

	id	label	confidence	social_timestamp	social_karma	syntax_ari	lex_liwc_WC	lex
count	2838.000000	2838.000000	2838.000000	2.838000e+03	2838.000000	2838.000000	2838.000000	
mean	13751.999295	0.524313	0.808972	1.518107e+09	18.262156	4.684272	85.996124	
std	17340.161897	0.499497	0.177038	1.552209e+07	79.419166	3.316435	32.334887	
min	4.000000	0.000000	0.428571	1.483274e+09	0.000000	-6.620000	5.000000	
25%	926.250000	0.000000	0.600000	1.509698e+09	2.000000	2.464243	65.000000	
50%	1891.500000	1.000000	0.800000	1.517066e+09	5.000000	4.321886	81.000000	
75%	25473.750000	1.000000	1.000000	1.530898e+09	10.000000	6.505657	101.000000	
max	55757.000000	1.000000	1.000000	1.542592e+09	1435.000000	24.074231	310.000000	

8 rows × 112 columns

In [10]:
```
df.isnull().sum()
```

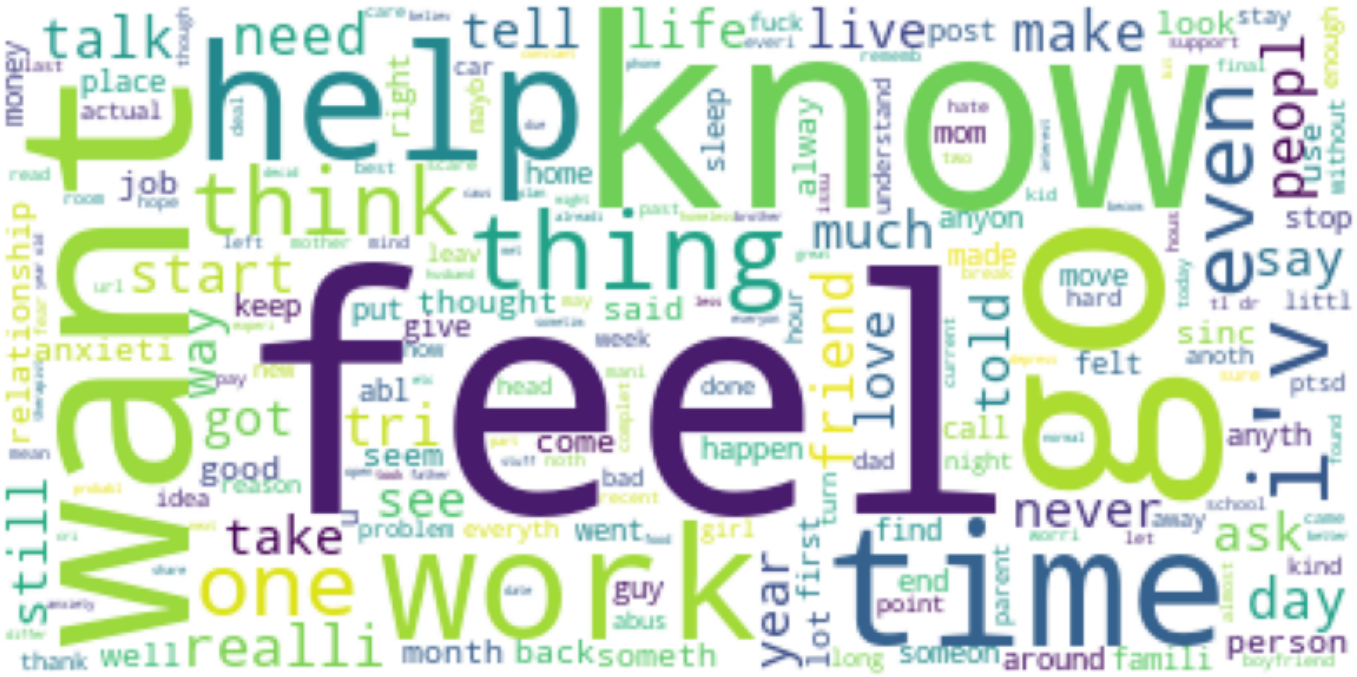
Out[10]:
```
subreddit                    0
post_id                      0
sentence_range               0
text                         0
id                           0
                            ..
lex_dal_avg_pleasantness     0
social_upvote_ratio          0
social_num_comments          0
syntax_fk_grade              0
sentiment                    0
Length: 116, dtype: int64
```

In [12]:
```
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()  #returns a string where all characters are lower case. S
    text = re. sub('\[.*?\]',' ',text)  #substring and returns a string with replaced va
    text = re. sub('https?://\S+/www\. \S+', ' ', text)#whitespace char with pattern
    text = re. sub('<. *?>+', ' ', text)#special char enclosed in square brackets
    text = re. sub(' [%s]' % re. escape(string. punctuation), ' ', text)#eliminate punct
    text = re. sub(' \n',' ', text)
    text = re. sub(' \w*\d\w*' ,' ', text)#word character ASCII punctuation
    text = [word for word in text. split(' ') if word not in stopword]  #removing stopwo
    text =" ". join(text)
    text = [stemmer . stem(word) for word in text. split(' ') ]#remove morphological aff
    text = " ". join(text)
    return text
df [ "text"] = df["text"]. apply(clean)
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\GOWRI\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

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( )
```



In [16]:

```
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, 7346)     1
  (0, 3226)     1
  (0, 9392)     1
  (0, 814)      1
  (0, 8301)     1
  (0, 3697)     1
  (0, 7155)     1
  (0, 8846)     1
  (0, 252)      1
  (0, 9683)     1
  (0, 4250)     1
  (0, 4980)     1
  (0, 5272)     1
  (0, 2141)     1
  (0, 5066)     1
  (0, 3213)     1
  (0, 2543)     3
  (0, 4135)     1
  (0, 5263)     1
  (0, 3644)     1
  (0, 8281)     1
  (0, 6803)     1
  (0, 4097)     1
  (0, 5121)     1
  (0, 1782)     1
  :     :
  (2836, 830)   1
  (2836, 4502)  1
  (2836, 2875)  1
  (2836, 4562)  1
  (2836, 4731)  1
  (2836, 4458)  1
  (2837, 7346)  2
  (2837, 2966)  1
  (2837, 5479)  2
  (2837, 8722)  1
  (2837, 8443)  1
  (2837, 6714)  1
  (2837, 4265)  1
  (2837, 9579)  1
  (2837, 5515)  1
  (2837, 8819)  1
  (2837, 5659)  1
  (2837, 2537)  1
  (2837, 7409)  1
  (2837, 2302)  1
  (2837, 7746)  1
  (2837, 2706)  1
  (2837, 8818)  1
  (2837, 5405)  1
  (2837, 2968)  1
```

In [18]:
```
from sklearn.naive_bayes import BernoulliNB
model=BernoulliNB()
model.fit(xtrain,ytrain)
```

Out[18]:
```
BernoulliNB()
```

In [19]:
```
user=input("Enter the text")
data=cv.transform([user]).toarray()
output=model.predict(data)
print(output)
```

Enter the textThese past couple of months have been the worst. My anxiety has gotten so bad itâ€™s effecting my sleep and relationship. Iâ€™ve become so paranoid about my health as well. I donâ€™t feel like me anymore and I just feel scared all the time now over every little thing. I donâ€™t have money to see a therapist either...
[1]