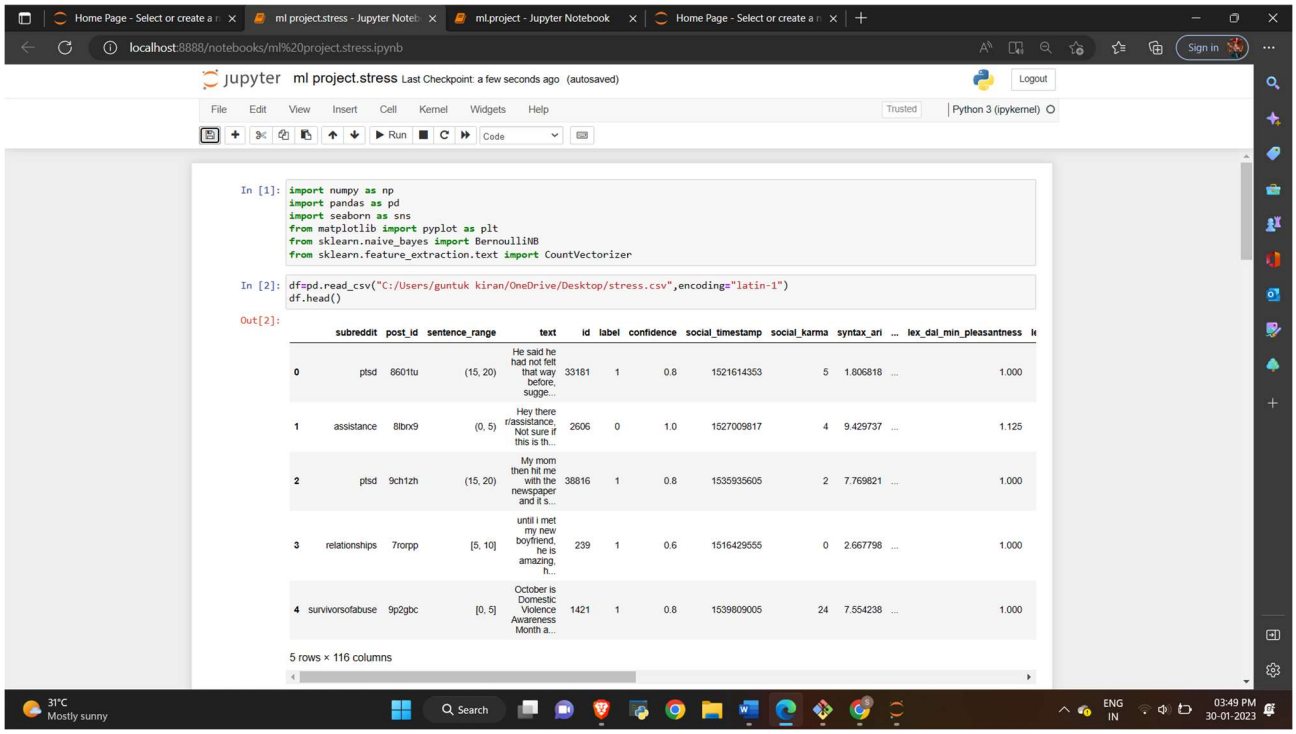
MACHINE LEARNING PROJECT

LITERATURE SURVEY

SI.NO.	TITLE	AUTHOR NAME	PUBLISHED YEAR	REMARKS
1	Machine Learning and IoT for Prediction and Detection of Stress	Mr.Purnendu Shekhar Pandey	2017	The paper focuses on using heart rate as a parameter to detect stress in individuals as well as predicting a person's condition to inform them about their health status. To achieve this goal, the authors propose using Machine Learning and the Internet of Things (IoT). IoT is used to communicate a person's health status to them, whereas ML is used to predict their condition. The goal is to help people recognise when they are unhealthy and take action to improve their health before an acute condition occurs.
2	Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data	Pramod Bobade, Vani M.	2020	In this paper, machine learning and deep learning methods for stress detection are studied. To train the algorithms and identify people who are stressed, the authors employ multimodal physiological data such as heart rate, electrodermal activity, and breathing. The research compares the outcomes produced by each machine learning and deep learning method while discussing their respective efficacies. The authors come to the conclusion that combining multimodal physiological data is useful for stress detection and that deep learning algorithms provide the best results in comparison to traditional machine learning methods.
3	A Decision Tree Optimised SVM Model for Stress Detection using Bio signals	Alana Paul Cruz, Aravind Pradeep, Kavali Riya Sivasankar and Krishnaveni K.S	2020	The paper focuses on the use of decision tree optimization to improve the performance of support vector machine (SVM) models in biosignal stress detection. The authors propose using decision trees to pre-process the biosignal data before training the SVM model with the optimised data. The optimised SVM model's performance is evaluated and compared to traditional SVM models. The proposed decision tree optimised SVM model outperforms traditional SVM models in stress detection using biosignals, according to the results.

4	Stress detection using deep neural networks	Russell Li and Zhandong Liu	2020	The paper focuses on the use of deep neural networks (DNNs) to detect stress. The authors collect physiological and behavioural data from stress-inducing tasks participants and use this data to train DNNs for stress detection. The DNNs' performance is evaluated and compared to that of traditional machine learning algorithms. DNNs replace traditional machine learning algorithms in stress detection, according to the findings. The authors also examine the DNNs' learned features and discover that they capture complex patterns in physiological and behavioural data that indicate stress. The authors conclude that DNNs are effective at detecting stress and have the potential to provide a deeper understanding of the underlying stress mechanisms.
5	Automatic Stress Detection Using Wearable Sensors and Machine Learning: A Review	Shruti Gedam, Sanchita Paul	2020	In this paper, we make an analysis on automatic stress detection using wearable sensors and machine learning. The authors study the different physiological and behavioural cues, such as heart rate, skin conductance, and movement, that have been employed for stress detection. They also discuss the numerous decision trees, support vector machines, and deep learning machine learning algorithms and methods that have been applied to stress detection. The authors offer recommendations for further research in this field after outlining the possible advantages and disadvantages of utilising wearable sensors and machine learning for stress detection.

Screenshots

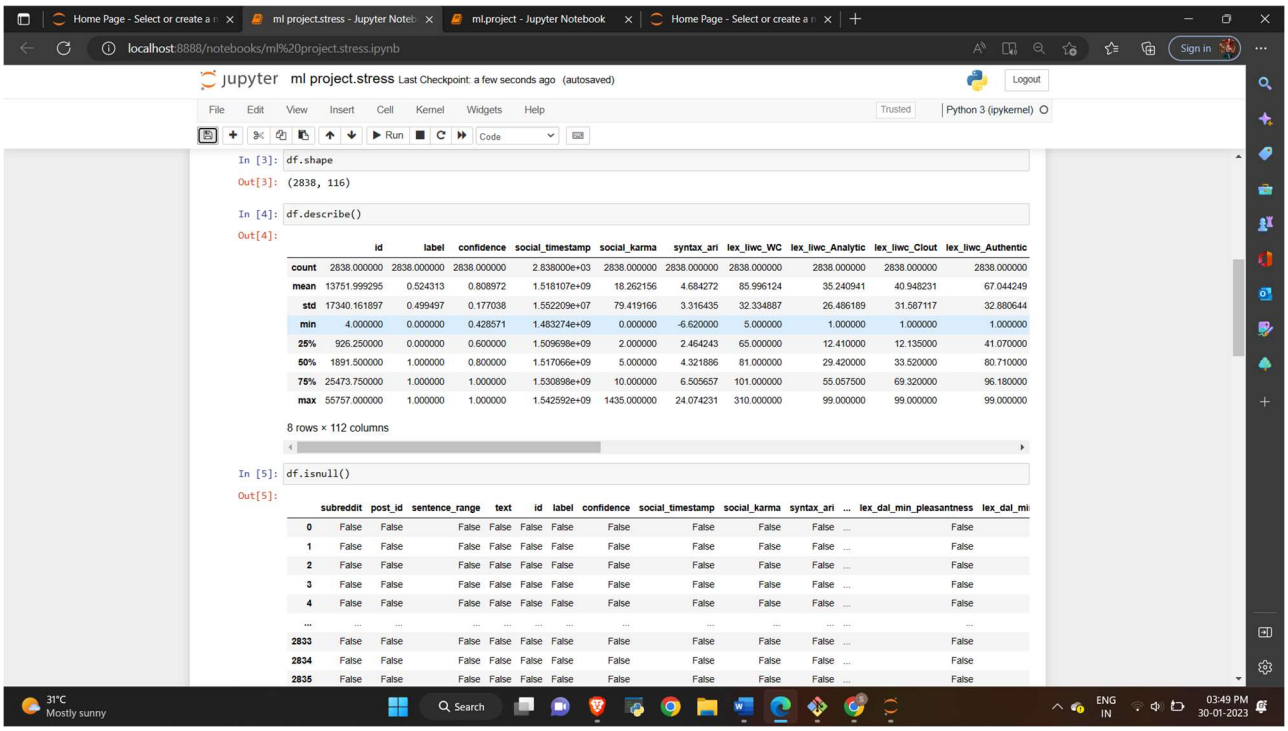


```
In [1]: import numpy as np
        import pandas as pd
        import seaborn as sns
        from matplotlib import pyplot as plt
        from sklearn.naive_bayes import BernoulliNB
        from sklearn.feature_extraction.text import CountVectorizer

In [2]: df=pd.read_csv("C:/Users/guntuk kiran/OneDrive/Desktop/stress.csv",encoding="latin-1")
        df.head()
```

Out[2]:

	subreddit	post_id	sentence_range	text	id	label	confidence	social_timestamp	social_karma	syntax_ari	...	lex_dal_min_pleasantness	le
0	ptsd	8601tu	(15, 20)	He said he had not felt that way before, sugge...	33181	1	0.8	1521614353	5	1.806818	...	1.000	
1	assistance	8lbrx9	(0, 5)	Hey there r/assistance, Not sure if this is th...	2606	0	1.0	1527009817	4	9.429737	...	1.125	
2	ptsd	9ch1zh	(15, 20)	My mom then hit me with the newspaper and it s...	38816	1	0.8	1535935605	2	7.769821	...	1.000	
3	relationships	7rorpp	[5, 10]	until i met my new boyfriend, he is amazing, h...	239	1	0.6	1516429555	0	2.667798	...	1.000	
4	survivorsofabuse	9p2gbc	[0, 5]	October is Domestic Violence Awareness Month a...	1421	1	0.8	1539809005	24	7.554238	...	1.000	

5 rows × 116 columns



```
In [3]: df.shape
Out[3]: (2838, 116)

In [4]: df.describe()
```

Out[4]:

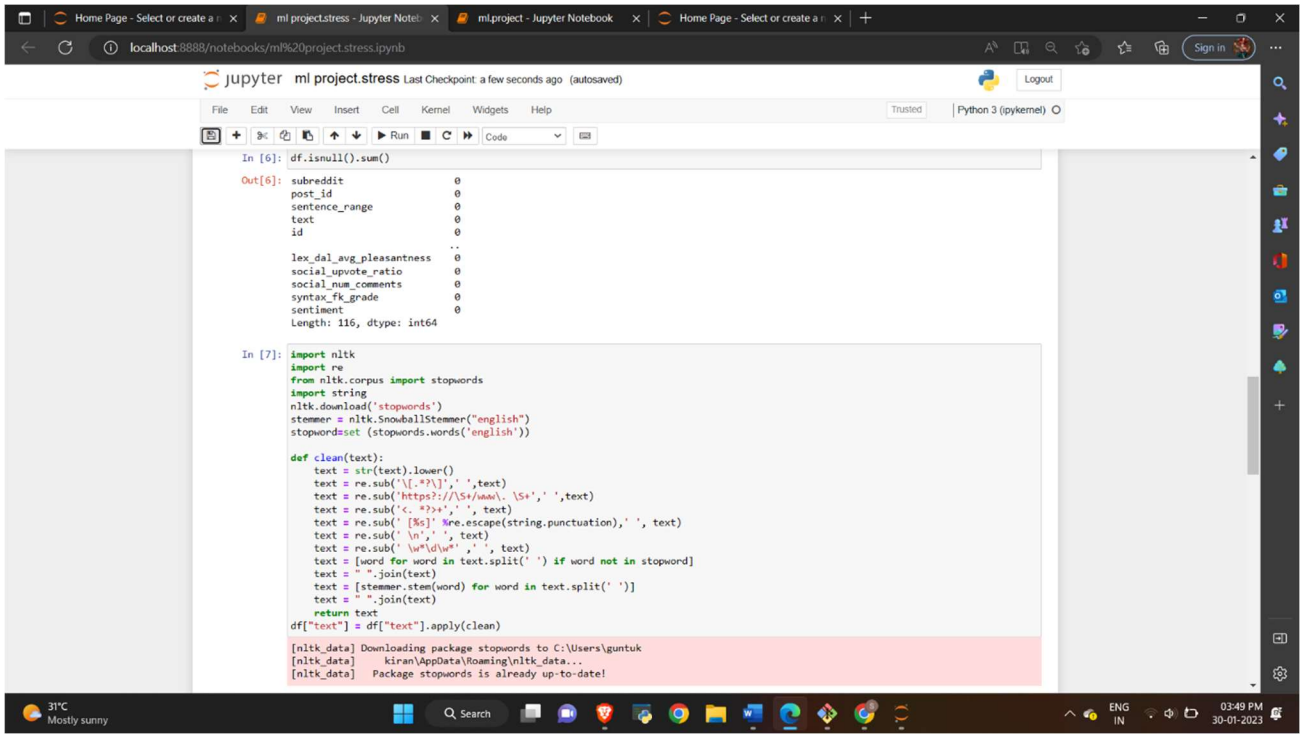
	id	label	confidence	social_timestamp	social_karma	syntax_ari	lex_liwc_WC	lex_liwc_Analytic	lex_liwc_Clout	lex_liwc_Authentic
count	2838.000000	2838.000000	2838.000000	2.838000e+03	2838.000000	2838.000000	2838.000000	2838.000000	2838.000000	2838.000000
mean	13751.999295	0.524313	0.808972	1.518107e+09	18.262156	4.684272	85.996124	35.240941	40.948231	67.044249
std	17340.161897	0.499497	0.177038	1.552209e+07	79.419166	3.316435	32.334887	26.486189	31.587117	32.880644
min	4.000000	0.000000	0.428571	1.483274e+09	0.000000	-6.620000	5.000000	1.000000	1.000000	1.000000
25%	926.250000	0.000000	0.600000	1.509698e+09	2.000000	2.464243	65.000000	12.410000	12.135000	41.070000
50%	1891.500000	1.000000	0.800000	1.517066e+09	5.000000	4.321886	81.000000	29.420000	33.520000	80.710000
75%	25473.750000	1.000000	1.000000	1.530898e+09	10.000000	6.505657	101.000000	55.057500	69.320000	96.180000
max	55757.000000	1.000000	1.000000	1.542592e+09	1435.000000	24.074231	310.000000	99.000000	99.000000	99.000000

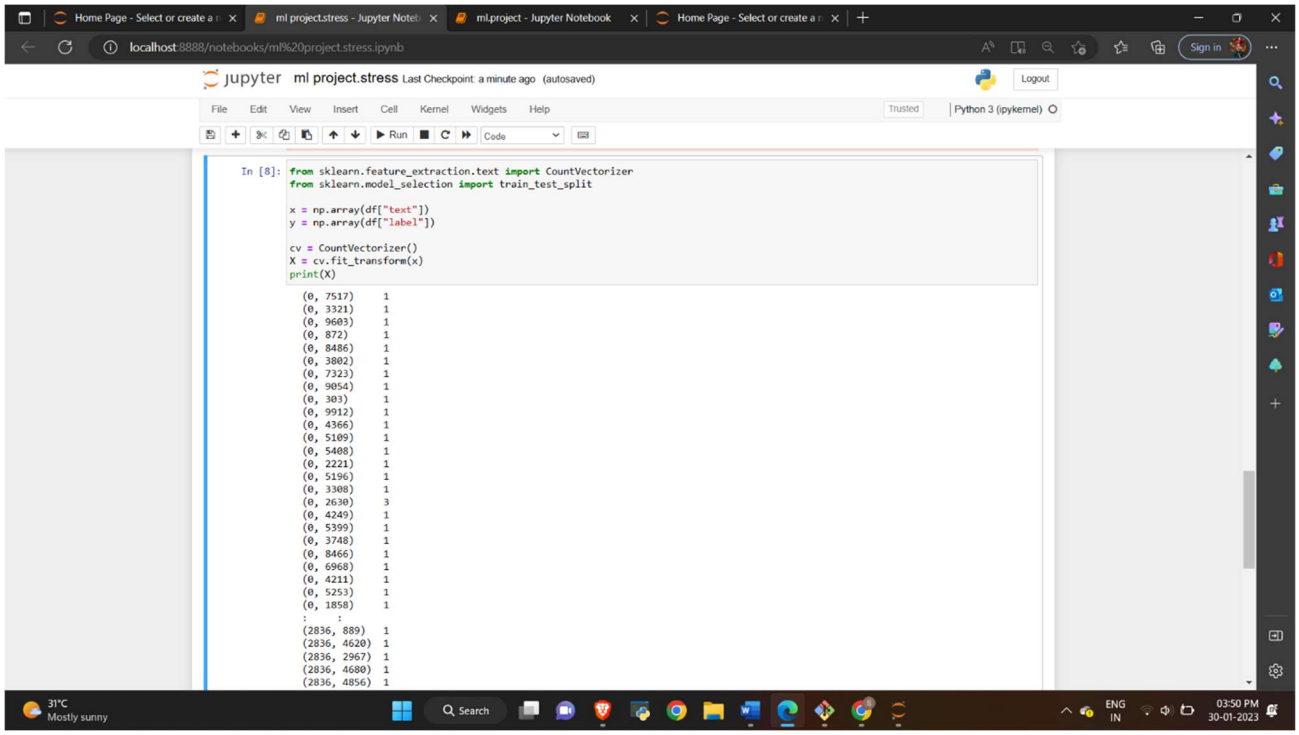
8 rows × 112 columns

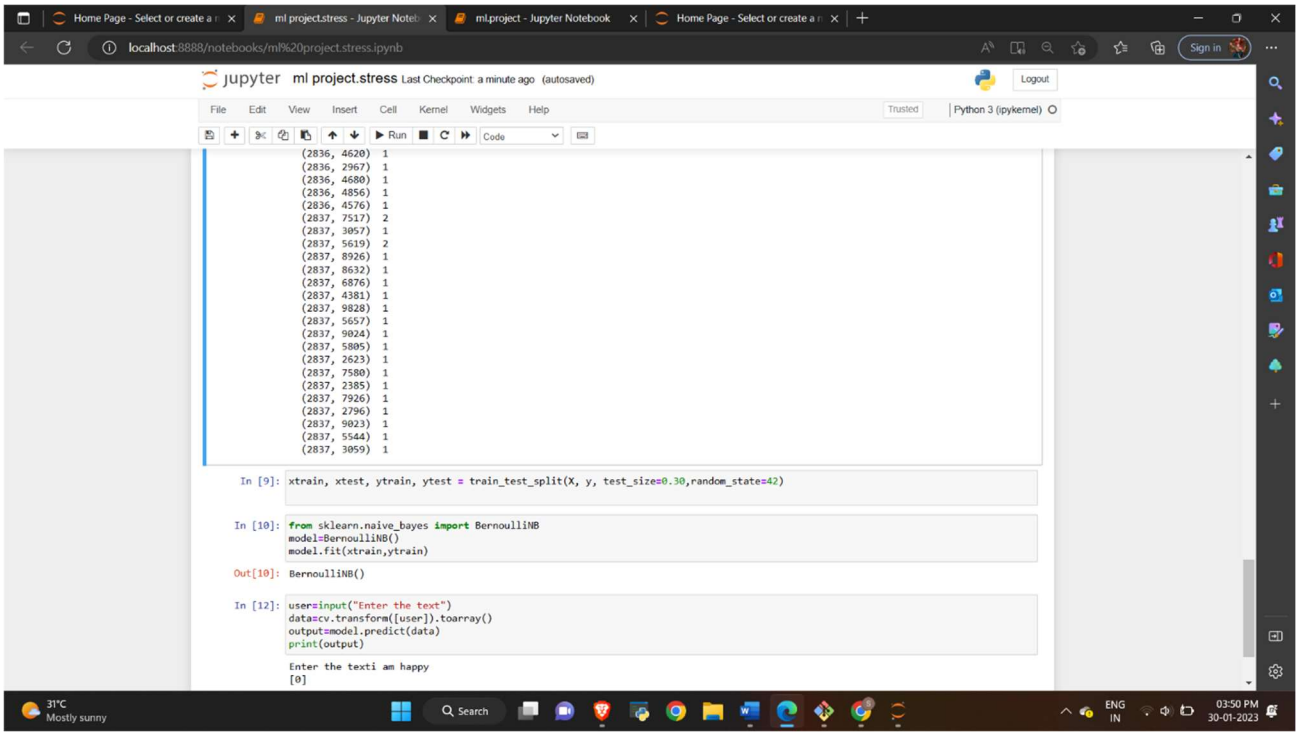
```
In [5]: df.isnull()
```

Out[5]:

	subreddit	post_id	sentence_range	text	id	label	confidence	social_timestamp	social_karma	syntax_ari	...	lex_dal_min_pleasantness	lex_dal_mi
0	False	False	False	False	False	False	False	False	False	False	...	False	
1	False	False	False	False	False	False	False	False	False	False	...	False	
2	False	False	False	False	False	False	False	False	False	False	...	False	
3	False	False	False	False	False	False	False	False	False	False	...	False	
4	False	False	False	False	False	False	False	False	False	False	...	False	
...	...	...	...	...	...	...	...	...	...	...	...	...	
2833	False	False	False	False	False	False	False	False	False	False	...	False	
2834	False	False	False	False	False	False	False	False	False	False	...	False	
2835	False	False	False	False	False	False	False	False	False	False	...	False	


Home Page - Select or create a n
ml project.stress - Jupyter Noteb
ml.project - Jupyter Notebook
Home Page - Select or create a n
localhost:8888/notebooks/ml%20project.stress.ipynb
Sign in
jupyter ml project.stress Last Checkpoint: a few seconds ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [6]: df.isnull().sum()
Out[6]: 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 [7]: 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] Downloading package stopwords to C:\Users\guntuk
[nltk_data] kiran\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
31°C
Mostly sunny
Search
ENG
IN
03:49 PM
30-01-2023


Home Page - Select or create a n
ml project.stress - Jupyter Noteb
ml.project - Jupyter Notebook
Home Page - Select or create a n
localhost:8888/notebooks/ml%20project.stress.ipynb
Sign in
jupyter ml project.stress Last Checkpoint: a minute ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [8]: 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)
(0, 7517) 1
(0, 3321) 1
(0, 9603) 1
(0, 872) 1
(0, 8486) 1
(0, 3802) 1
(0, 7323) 1
(0, 9054) 1
(0, 303) 1
(0, 9912) 1
(0, 4366) 1
(0, 5109) 1
(0, 5408) 1
(0, 2221) 1
(0, 5196) 1
(0, 3308) 1
(0, 2630) 3
(0, 4249) 1
(0, 5399) 1
(0, 3748) 1
(0, 8466) 1
(0, 6968) 1
(0, 4211) 1
(0, 5253) 1
(0, 1858) 1
: :
(2836, 889) 1
(2836, 4620) 1
(2836, 2967) 1
(2836, 4680) 1
(2836, 4856) 1
31°C
Mostly sunny
Search
ENG
IN
03:50 PM
30-01-2023



Code


```
[{"metadata":{"trusted":false},"id":"5714f235","cell_type":"code","sour
ce":"import numpy as np\nimport pandas as pd \nimport seaborn as
sns\nfrom matplotlib import pyplot as plt\nfrom sklearn.naive_bayes
import BernoulliNB\nfrom sklearn.feature_extraction.text import
CountVectorizer","execution_count":1,"outputs":[]},{"metadata":{"truste
d":false},"id":"cc98b769","cell_type":"code","source":"df=pd.read_csv(\
"C:/Users/guntuk kiran/OneDrive/Desktop/stress.csv\",encoding=\"latin-
1\")\ndf.head()","execution_count":2,"outputs":[{"data":{"text/html":"
\n\n    \n      \n      subreddit\n      post_id\n
sentence_range\n      text\n      id\n      label\n      confidence\n
social_timestamp\n      social_karma\n      syntax_ari\n      ...\n
lex_dal_min_pleasantness\n      lex_dal_min_activation\n
lex_dal_min_imagery\n      lex_dal_avg_activation\n
lex_dal_avg_imagery\n      lex_dal_avg_pleasantness\n
social_upvote_ratio\n      social_num_comments\n      syntax_fk_grade\n
sentiment\n    \n  \n  \n    \n      0\n      ptsd\n      8601tu\n
(15, 20)\n      He said he had not felt that way before, sugge...\n
33181\n      1\n      0.8\n      1521614353\n      5\n      1.806818\n
...\n      1.000\n      1.1250\n      1.0\n      1.77000\n
1.52211\n      1.89556\n      0.86\n      1\n      3.253573\n      -
0.002742\n    \n    \n      1\n      assistance\n      8lbrx9\n
(0, 5)\n      Hey there r/assistance, Not sure if this is th...\n
2606\n      0\n      1.0\n      1527009817\n      4\n      9.429737\n
```

...\n   1.125\n   1.0000\n   1.0\n   1.69586\n
1.62045\n   1.88919\n   0.65\n   2\n   8.828316\n
0.292857\n   \n   \n   2\n   ptsd\n   9ch1zh\n   (15,
20)\n   My mom then hit me with the newspaper and it s...\n
38816\n   1\n   0.8\n   1535935605\n   2\n   7.769821\n
...\n   1.000\n   1.1429\n   1.0\n   1.83088\n
1.58108\n   1.85828\n   0.67\n   0\n   7.841667\n
0.011894\n   \n   \n   3\n   relationships\n   7rorpp\n
[5, 10]\n   until i met my new boyfriend, he is amazing, h...\n
239\n   1\n   0.6\n   1516429555\n   0\n   2.667798\n
...\n   1.000\n   1.1250\n   1.0\n   1.75356\n
1.52114\n   1.98848\n   0.50\n   5\n   4.104027\n
0.141671\n   \n   \n   4\n   survivorsofabuse\n   9p2gbc\n
[0, 5]\n   October is Domestic Violence Awareness Month a...\n
1421\n   1\n   0.8\n   1539809005\n   24\n   7.554238\n
...\n   1.000\n   1.1250\n   1.0\n   1.77644\n
1.64872\n   1.81456\n   1.00\n   1\n   7.910952\n   -
0.204167\n   \n \n\n

5 rows × 116 columns

\n
","text/plain":"   subreddit post_id sentence_range \\\n0
ptsd 8601tu   (15, 20)   \n1   assistance 8lbrx9   (0,
5)   \n2   ptsd 9ch1zh   (15, 20)   \n3
relationships 7rorpp   [5, 10]   \n4 survivorsofabuse 9p2gbc
[0, 5]   \n\n   text
id label \\\n0 He said he had not felt that way before, sugge...
33181   1   \n1 Hey there r/assistance, Not sure if this is th...
2606   0   \n2 My mom then hit me with the newspaper and it s...
38816   1   \n3 until i met my new boyfriend, he is amazing, h...
239   1   \n4 October is Domestic Violence Awareness Month a...
1421   1   \n\n   confidence social_timestamp social_karma
syntax_ari ... \\\n0   0.8   1521614353   5
1.806818 ...   \n1   1.0   1527009817   4
9.429737 ...   \n2   0.8   1535935605   2
7.769821 ...   \n3   0.6   1516429555   0
2.667798 ...   \n4   0.8   1539809005   24
7.554238 ...   \n\n   lex_dal_min_pleasantness lex_dal_min_activation
lex_dal_min_imagery \\\n0   1.000
1.1250   1.0   \n1   1.125
1.0000   1.0   \n2   1.000
1.1429   1.0   \n3   1.000
1.1250   1.0   \n4   1.000
1.1250   1.0   \n\n   lex_dal_avg_activation
lex_dal_avg_imagery lex_dal_avg_pleasantness \\\n0
1.77000   1.52211   1.89556   \n1
1.69586   1.62045   1.88919   \n2
1.83088   1.58108   1.85828   \n3
1.75356   1.52114   1.98848   \n4
1.77644   1.64872   1.81456   \n\n
social_upvote_ratio social_num_comments syntax_fk_grade sentiment
\n0   0.86   1   3.253573 -
0.002742 \n1   0.65   2
8.828316   0.292857 \n2   0.67   0
7.841667   0.011894 \n3   0.50   5

```
4.104027   0.141671  \n4                 1.00                    1
7.910952  -0.204167  \n\n[5 rows x 116
columns]"},"execution_count":2,"metadata":{},"output_type":"execute_res
ult"}]},{"metadata":{"trusted":false},"id":"6deb933d","cell_type":"code
","source":"df.shape","execution_count":3,"outputs":[{"data":{"text/pla
in":"(2838,
116)"},"execution_count":3,"metadata":{},"output_type":"execute_result"
}]},{"metadata":{"trusted":false},"id":"d714c405","cell_type":"code","s
ource":"df.describe()","execution_count":4,"outputs":[{"data":{"text/ht
ml":"
\n\n    \n      \n      id\n      label\n      confidence\n
social_timestamp\n      social_karma\n      syntax_ari\n
lex_liwc_WC\n      lex_liwc_Analytic\n      lex_liwc_Clout\n
lex_liwc_Authentic\n      ...\n      lex_dal_min_pleasantness\n
lex_dal_min_activation\n      lex_dal_min_imagery\n
lex_dal_avg_activation\n      lex_dal_avg_imagery\n
lex_dal_avg_pleasantness\n      social_upvote_ratio\n
social_num_comments\n      syntax_fk_grade\n      sentiment\n    \n  \n
\n    \n      count\n      2838.000000\n      2838.000000\n
2838.000000\n      2.838000e+03\n      2838.000000\n      2838.000000\n
2838.000000\n      2838.000000\n      2838.000000\n      2838.000000\n
...\n      2838.000000\n      2838.000000\n      2838.000000\n
2838.000000\n      2838.000000\n      2838.000000\n      2838.000000\n
2838.000000\n      2838.000000\n      2838.000000\n    \n    \n
mean\n      13751.999295\n      0.524313\n      0.808972\n
1.518107e+09\n      18.262156\n      4.684272\n      85.996124\n
35.240941\n      40.948231\n      67.044249\n      ...\n
1.088001\n      1.120099\n      1.000211\n      1.722759\n
1.536400\n      1.879385\n      0.843517\n      9.948555\n
5.448836\n      0.040740\n    \n    \n      std\n      17340.161897\n
0.499497\n      0.177038\n      1.552209e+07\n      79.419166\n
3.316435\n      32.334887\n      26.486189\n      31.587117\n
32.880644\n      ...\n      0.117159\n      0.085227\n      0.006500\n
0.047835\n      0.102971\n      0.058932\n      0.174794\n
21.798032\n      2.535829\n      0.195490\n    \n    \n      min\n
4.000000\n      0.000000\n      0.428571\n      1.483274e+09\n
0.000000\n      -6.620000\n      5.000000\n      1.000000\n
1.000000\n      1.000000\n      ...\n      1.000000\n      1.000000\n
1.000000\n      1.485400\n      1.200000\n      1.561150\n
0.140000\n      0.000000\n      -1.918000\n      -1.000000\n    \n
\n      25%\n      926.250000\n      0.000000\n      0.600000\n
1.509698e+09\n      2.000000\n      2.464243\n      65.000000\n
12.410000\n      12.135000\n      41.070000\n      ...\n
1.000000\n      1.000000\n      1.000000\n      1.691430\n
1.469745\n      1.841782\n      0.750000\n      2.000000\n
3.729973\n      -0.072222\n    \n    \n      50%\n      1891.500000\n
1.000000\n      0.800000\n      1.517066e+09\n      5.000000\n
4.321886\n      81.000000\n      29.420000\n      33.520000\n
80.710000\n      ...\n      1.000000\n      1.142900\n      1.000000\n
1.721430\n      1.530295\n      1.878250\n      0.890000\n
5.000000\n      5.210000\n      0.044821\n    \n    \n      75%\n
25473.750000\n      1.000000\n      1.000000\n      1.530898e+09\n
10.000000\n      6.505657\n      101.000000\n      55.057500\n
69.320000\n      96.180000\n      ...\n      1.142900\n      1.142900\n
1.000000\n      1.751760\n      1.596030\n      1.916243\n
1.000000\n      10.000000\n      6.855217\n      0.166667\n    \n    \n
max\n      55757.000000\n      1.000000\n      1.000000\n
```

```
1.542592e+09\n      1435.000000\n      24.074231\n      310.000000\n
99.000000\n      99.000000\n      99.000000\n      ...\n
1.900000\n      1.500000\n      1.200000\n      2.007400\n
2.066670\n      2.158490\n      1.000000\n      416.000000\n
21.198919\n      1.000000\n    \n  \n\n
```

8 rows × 112 columns

```
\n
","text/plain":"                 id        label   confidence
social_timestamp  social_karma  \\\ncount   2838.000000  2838.000000
2838.000000      2.838000e+03   2838.000000   \nmean   13751.999295
0.524313     0.808972      1.518107e+09     18.262156   \nstd
17340.161897     0.499497     0.177038      1.552209e+07     79.419166
\nmin        4.000000     0.000000     0.428571      1.483274e+09
0.000000   \n25%      926.250000     0.000000     0.600000
1.509698e+09      2.000000   \n50%     1891.500000     1.000000
0.800000      1.517066e+09      5.000000   \n75%    25473.750000
1.000000     1.000000      1.530898e+09     10.000000   \nmax
55757.000000     1.000000     1.000000      1.542592e+09   1435.000000
\n\n        syntax_ari  lex_liwc_WC  lex_liwc_Analytic  lex_liwc_Clout
\\\ncount  2838.000000  2838.000000        2838.000000     2838.000000
\nmean      4.684272    85.996124          35.240941       40.948231
\nstd       3.316435    32.334887          26.486189       31.587117
\nmin      -6.620000     5.000000           1.000000        1.000000
\n25%       2.464243    65.000000          12.410000       12.135000
\n50%       4.321886    81.000000          29.420000       33.520000
\n75%       6.505657   101.000000          55.057500       69.320000
\nmax      24.074231   310.000000          99.000000       99.000000
\n\n       lex_liwc_Authentic  ...  lex_dal_min_pleasantness  \\\ncount
2838.000000  ...               2838.000000   \nmean
67.044249  ...                  1.088001   \nstd             32.880644
...                  0.117159   \nmin              1.000000  ...
1.000000   \n25%             41.070000  ...                   1.000000
\n50%             80.710000  ...                  1.000000   \n75%
96.180000  ...                  1.142900   \nmax             99.000000
...                  1.900000   \n\n       lex_dal_min_activation
lex_dal_min_imagery  lex_dal_avg_activation  \\\ncount
2838.000000          2838.000000             2838.000000   \nmean
1.120099             1.000211                1.722759   \nstd
0.085227             0.006500                0.047835   \nmin
1.000000             1.000000                1.485400   \n25%
1.000000             1.000000                1.691430   \n50%
1.142900             1.000000                1.721430   \n75%
1.142900             1.000000                1.751760   \nmax
1.500000             1.200000                2.007400   \n\n
lex_dal_avg_imagery  lex_dal_avg_pleasantness  social_upvote_ratio
\\\ncount          2838.000000               2838.000000
2838.000000   \nmean             1.536400                 1.879385
0.843517   \nstd              0.102971                 0.058932
0.174794   \nmin              1.200000                 1.561150
0.140000   \n25%              1.469745                 1.841782
0.750000   \n50%              1.530295                 1.878250
0.890000   \n75%              1.596030                 1.916243
1.000000   \nmax              2.066670                 2.158490
1.000000   \n\n       social_num_comments  syntax_fk_grade    sentiment
```

```
\ncount          2838.000000      2838.000000  2838.000000  \nmean
9.948555         5.448836     0.040740  \nstd              21.798032
2.535829     0.195490  \nmin               0.000000        -1.918000
-1.000000  \n25%               2.000000         3.729973    -0.072222
\n50%               5.000000         5.210000     0.044821  \n75%
10.000000         6.855217     0.166667  \nmax             416.000000
21.198919     1.000000  \n\n[8 rows x 112
columns]"},"execution_count":4,"metadata":{},"output_type":"execute_res
ult"}]},{"metadata":{"trusted":false},"id":"824b7ff8","cell_type":"code
","source":"df.isnull()","execution_count":5,"outputs":[{"data":{"text/
html":"
\n\n    \n      \n      subreddit\n      post_id\n
sentence_range\n      text\n      id\n      label\n      confidence\n
social_timestamp\n      social_karma\n      syntax_ari\n      ...\n
lex_dal_min_pleasantness\n      lex_dal_min_activation\n
lex_dal_min_imagery\n      lex_dal_avg_activation\n
lex_dal_avg_imagery\n      lex_dal_avg_pleasantness\n
social_upvote_ratio\n      social_num_comments\n      syntax_fk_grade\n
sentiment\n    \n  \n  \n    \n      0\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      ...\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n    \n    \n      1\n      False\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      ...\n      False\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n    \n    \n      2\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n      ...\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n    \n    \n
3\n      False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n      ...\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n    \n    \n
4\n      False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n      ...\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n    \n    \n
...\n      ...\n      ...\n      ...\n      ...\n      ...\n      ...\n
...\n      ...\n      ...\n      ...\n      ...\n      ...\n      ...\n
...\n      ...\n      ...\n      ...\n      ...\n      ...\n      ...\n
...\n    \n    \n      2833\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      ...\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n    \n    \n      2834\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      ...\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n    \n    \n      2835\n      False\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      ...\n      False\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n    \n    \n      2836\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n      ...\n
```

```
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n    \n    \n
2837\n      False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n      ...\n
False\n      False\n      False\n      False\n      False\n
False\n      False\n      False\n      False\n      False\n    \n  \n\n
```

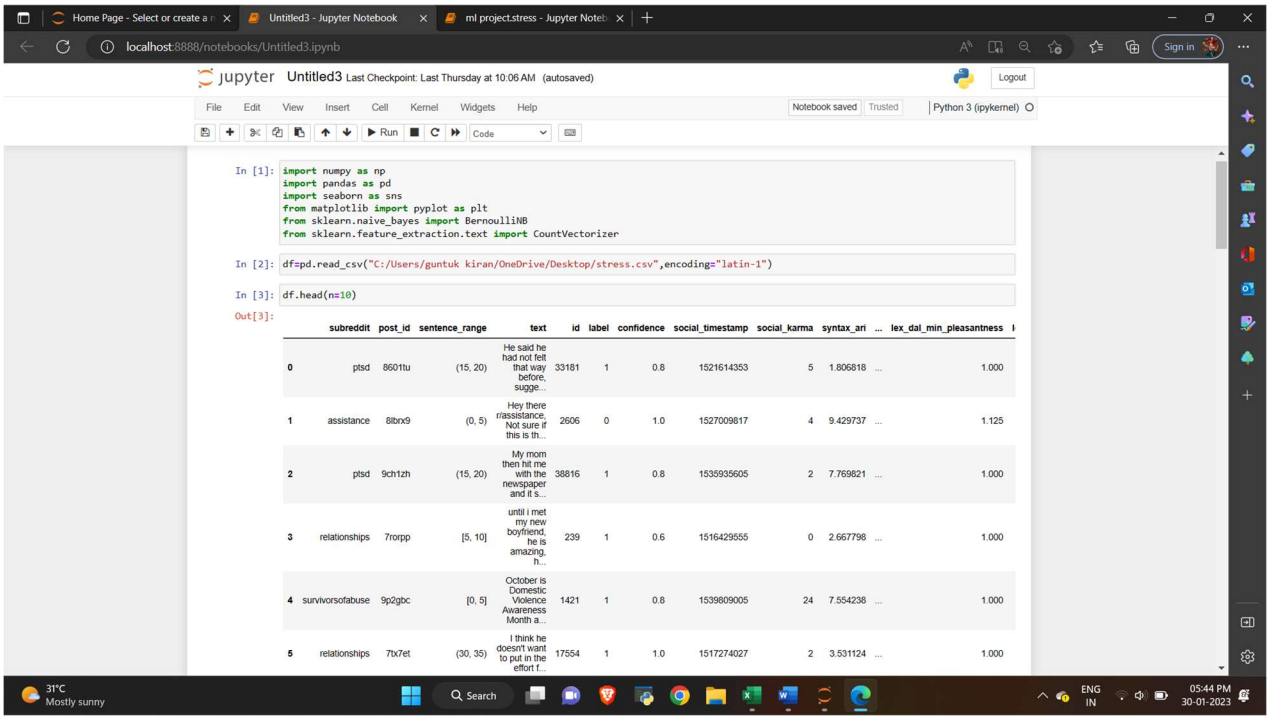
2838 rows × 116 columns

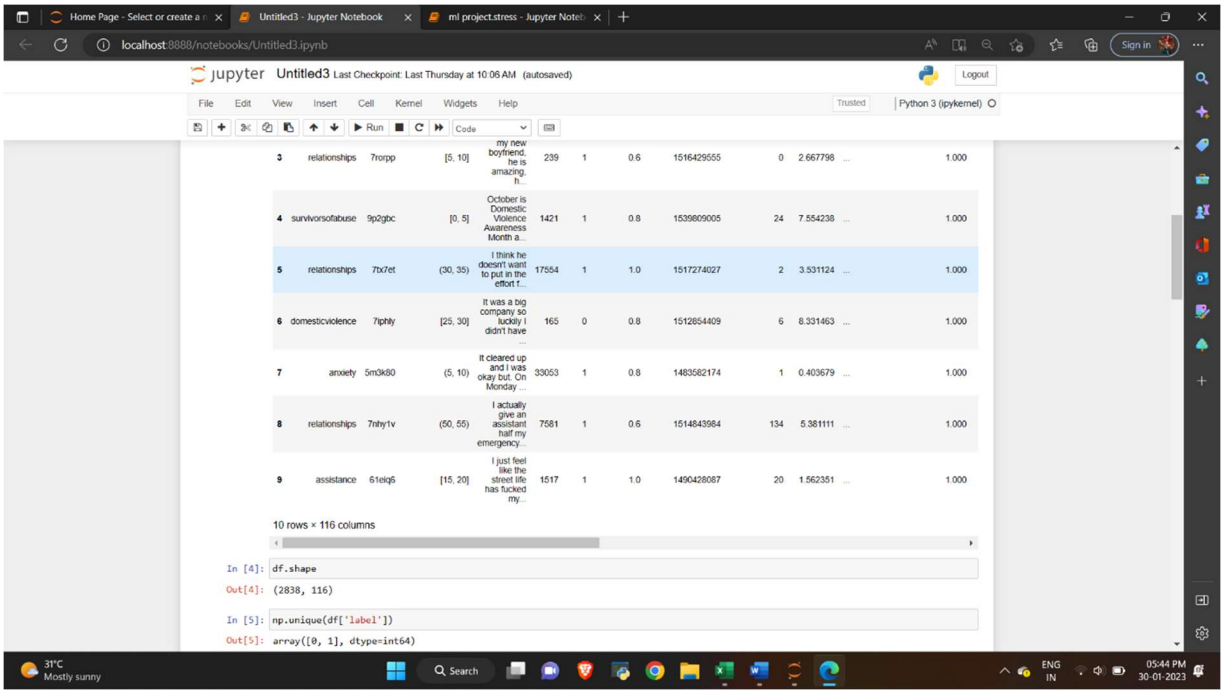
```
\n
","text/plain":"      subreddit  post_id  sentence_range   text     id
label  confidence  \\\n0         False    False           False  False
False  False       False   \n1         False    False           False
False  False  False       False   \n2         False    False
False  False  False  False       False   \n3         False    False
False  False  False  False       False   \n4         False    False
False  False  False  False       False   \n...         ...      ...
...    ...    ...    ...          ...   \n2833      False    False
False  False  False  False       False   \n2834      False    False
False  False  False  False       False   \n2835      False    False
False  False  False  False       False   \n2836      False    False
False  False  False  False       False   \n2837      False    False
False  False  False  False       False   \n\n      social_timestamp
social_karma  syntax_ari  ...  \\\n0                False         False
False  ...   \n1               False         False       False  ...
\n2               False         False       False  ...   \n3
False         False       False  ...   \n4                False
False       False  ...   \n...                ...           ...
...  ...   \n2833            False         False       False  ...
\n2834            False         False       False  ...   \n2835
False         False       False  ...   \n2836            False
False       False  ...   \n2837            False         False
False  ...   \n\n      lex_dal_min_pleasantness  lex_dal_min_activation
lex_dal_min_imagery  \\\n0                        False
False                False   \n1                       False
False                False   \n2                       False
False                False   \n3                       False
False                False   \n4                       False
False                False   \n...                       ...
...                  ...   \n2833                    False
False                False   \n2834                    False
False                False   \n2835                    False
False                False   \n2836                    False
False                False   \n2837                    False
False                False   \n\n      lex_dal_avg_activation
lex_dal_avg_imagery  lex_dal_avg_pleasantness  \\\n0
False                False                      False   \n1
False                False                      False   \n2
False                False                      False   \n3
False                False                      False   \n4
False                False                      False   \n...
...                  ...                        ...   \n2833
False                False                      False   \n2834
False                False                      False   \n2835
False                False                      False   \n2836
False                False                      False   \n2837
```

```
False                False                     False   \n\n
social_upvote_ratio  social_num_comments  syntax_fk_grade  sentiment
\n0                   False                False            False
False  \n1                    False                False
False      False  \n2                   False               False
False      False  \n3                   False               False
False      False  \n4                   False               False
False      False  \n...                   ...                 ...
...        ...  \n2833               False               False
False      False  \n2834                 False               False
False      False  \n2835                 False               False
False      False  \n2836                 False               False
False      False  \n2837                 False               False
False      False  \n\n[2838 rows x 116
columns]"},"execution_count":5,"metadata":{},"output_type":"execute_res
ult"}]},{"metadata":{"trusted":false},"id":"53858da3","cell_type":"code
","source":"df.isnull().sum()","execution_count":6,"outputs":[{"data":{
"text/plain":"subreddit                  0\npost_id
0\nsentence_range              0\ntext                        0\nid
0\n                           ..\nlex_dal_avg_pleasantness
0\nsocial_upvote_ratio         0\nsocial_num_comments
0\nsyntax_fk_grade             0\nsentiment
0\nLength: 116, dtype:
int64"},"execution_count":6,"metadata":{},"output_type":"execute_result
"}]},{"metadata":{"trusted":false},"id":"732d07a1","cell_type":"code","
source":"import nltk\nimport re\nfrom nltk.corpus import
stopwords\nimport string\nnltk.download('stopwords')\nstemmer =
nltk.SnowballStemmer(\"english\")\nstopword=set
(stopwords.words('english'))\n\ndef clean(text):\n    text =
str(text).lower()\n    text = re.sub('\\[.*?\\]',' ',text)\n    text =
re.sub('https?://\\S+/www\\. \\S+',' ',text)\n    text = re.sub('+','
', text)\n    text = re.sub(' [%s]' %re.escape(string.punctuation),' ',
text)\n    text = re.sub(' \\n',' ', text)\n    text = re.sub('
\\w*\\d\\w*' ,' ', text)\n    text = [word for word in text.split(' ')
if word not in stopword]\n    text = \" \".join(text)\n    text =
[stemmer.stem(word) for word in text.split(' ')]\n    text = \"
\".join(text)\n    return text\ndf[\"text\"] =
df[\"text\"].apply(clean)","execution_count":7,"outputs":[{"name":"stde
rr","output_type":"stream","text":"[nltk_data] Downloading package
stopwords to C:\\Users\\guntuk\n[nltk_data]
kiran\\AppData\\Roaming\\nltk_data...\n[nltk_data]   Package stopwords
is already up-to-
date!\n"}]},{"metadata":{"trusted":false},"id":"d9645164","cell_type":"
code","source":"from sklearn.feature_extraction.text import
CountVectorizer\nfrom sklearn.model_selection import
train_test_split\n\nx = np.array(df[\"text\"])\ny =
np.array(df[\"label\"])\n\ncv = CountVectorizer()\nX =
cv.fit_transform(x)\nprint(X)","execution_count":8,"outputs":[{"name":"
stdout","output_type":"stream","text":"  (0, 7517)\t1\n  (0, 3321)\t1\n
(0, 9603)\t1\n  (0, 872)\t1\n  (0, 8486)\t1\n  (0, 3802)\t1\n  (0,
7323)\t1\n  (0, 9054)\t1\n  (0, 303)\t1\n  (0, 9912)\t1\n  (0,
4366)\t1\n  (0, 5109)\t1\n  (0, 5408)\t1\n  (0, 2221)\t1\n  (0,
5196)\t1\n  (0, 3308)\t1\n  (0, 2630)\t3\n  (0, 4249)\t1\n  (0,
5399)\t1\n  (0, 3748)\t1\n  (0, 8466)\t1\n  (0, 6968)\t1\n  (0,
4211)\t1\n  (0, 5253)\t1\n  (0, 1858)\t1\n  :\t:\n  (2836, 889)\t1\n
(2836, 4620)\t1\n  (2836, 2967)\t1\n  (2836, 4680)\t1\n  (2836,
4856)\t1\n  (2836, 4576)\t1\n  (2837, 7517)\t2\n  (2837, 3057)\t1\n
```

```
(2837, 5619)\t2\n  (2837, 8926)\t1\n  (2837, 8632)\t1\n  (2837,
6876)\t1\n  (2837, 4381)\t1\n  (2837, 9828)\t1\n  (2837, 5657)\t1\n
(2837, 9024)\t1\n  (2837, 5805)\t1\n  (2837, 2623)\t1\n  (2837,
7580)\t1\n  (2837, 2385)\t1\n  (2837, 7926)\t1\n  (2837, 2796)\t1\n
(2837, 9023)\t1\n  (2837, 5544)\t1\n  (2837,
3059)\t1\n"}]},{"metadata":{"trusted":false},"id":"057cd828","cell_type
":"code","source":"xtrain, xtest, ytrain, ytest = train_test_split(X,
y,
test_size=0.30,random_state=42)\n","execution_count":9,"outputs":[]},{"
metadata":{"trusted":false},"id":"82a61628","cell_type":"code","source"
:"from sklearn.naive_bayes import
BernoulliNB\nmodel=BernoulliNB()\nmodel.fit(xtrain,ytrain)","execution_
count":10,"outputs":[{"data":{"text/plain":"BernoulliNB()"},"execution_
count":10,"metadata":{},"output_type":"execute_result"}]},{"metadata":{
"trusted":false},"id":"fa5569a3","cell_type":"code","source":"user=inpu
t(\"Enter the
text\")\ndata=cv.transform([user]).toarray()\noutput=model.predict(data
)\nprint(output)","execution_count":12,"outputs":[{"name":"stdout","out
put_type":"stream","text":"Enter the texti am happy\n[0]\n"}]}]
```


Sample code on stress file





Home Page - Select or create a n | Untitled3 - Jupyter Notebook | ml project.stress - Jupyter Noteb

localhost:8888/notebooks/Untitled3.ipynb

jupyter Untitled3 Last Checkpoint: Last Thursday at 10:06 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

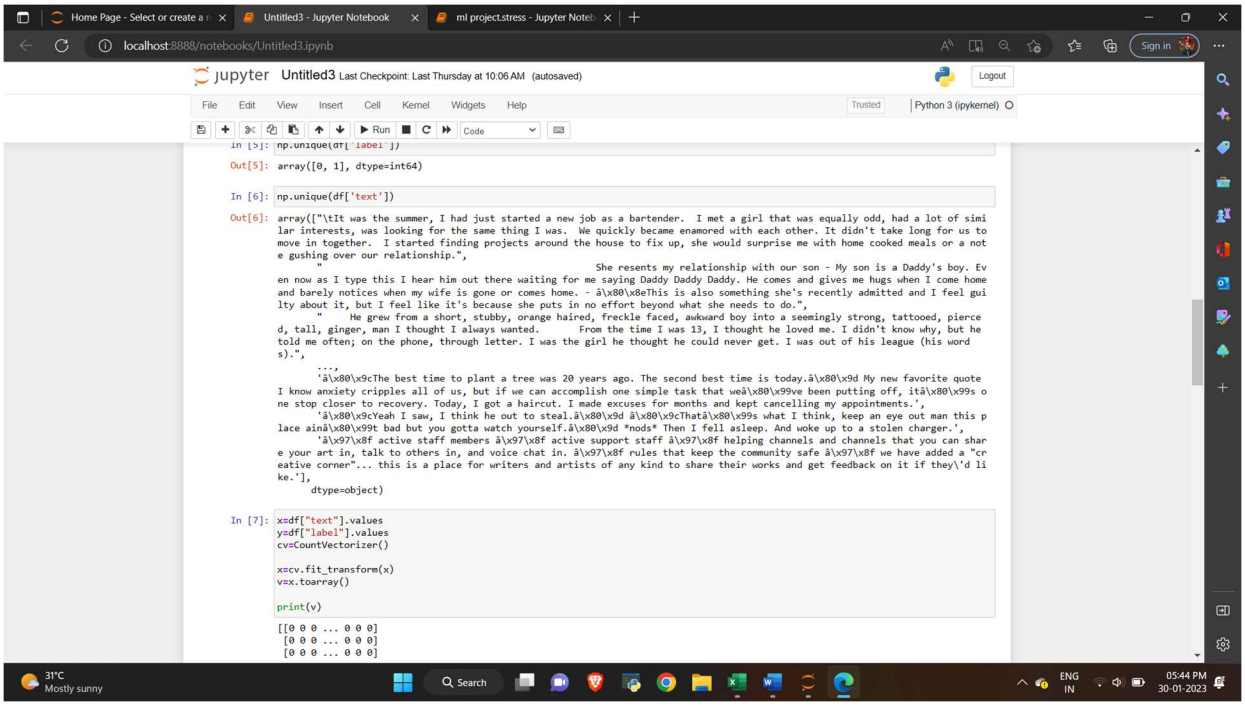
Run Code

3	relationships	7rorpp	[5, 10]	my new boyfriend, he is amazing, h...	239	1	0.6	1516429555	0	2.667798	...	1.000
4	survivorsofabuse	9p2gbc	[0, 5]	October is Domestic Violence Awareness Month a...	1421	1	0.8	1539809005	24	7.554238	...	1.000
5	relationships	7tx7et	(30, 35)	I think he doesn't want to put in the effort f...	17554	1	1.0	1517274027	2	3.531124	...	1.000
6	domesticviolence	7iphly	[25, 30]	It was a big company so luckily I didn't have ...	165	0	0.8	1512854409	6	8.331463	...	1.000
7	anxiety	5m3k80	(5, 10)	It cleared up and I was okay but. On Monday ...	33053	1	0.8	1483582174	1	0.403679	...	1.000
8	relationships	7nhy1v	(50, 55)	I actually give an assistant half my emergency...	7581	1	0.6	1514843984	134	5.381111	...	1.000
9	assistance	61eiq6	[15, 20]	I just feel like the street life has fucked my...	1517	1	1.0	1490428087	20	1.562351	...	1.000

10 rows × 116 columns

```
In [4]: df.shape
Out[4]: (2838, 116)

In [5]: np.unique(df['label'])
Out[5]: array([0, 1], dtype=int64)
```

31°C Mostly sunny | Search | ENG IN | 05:44 PM 30-01-2023



Home Page - Select or create a n | Untitled3 - Jupyter Notebook | ml project.stress - Jupyter Noteb

localhost:8888/notebooks/Untitled3.ipynb

jupyter Untitled3 Last Checkpoint: Last Thursday at 10:06 AM (autosaved)

Logout

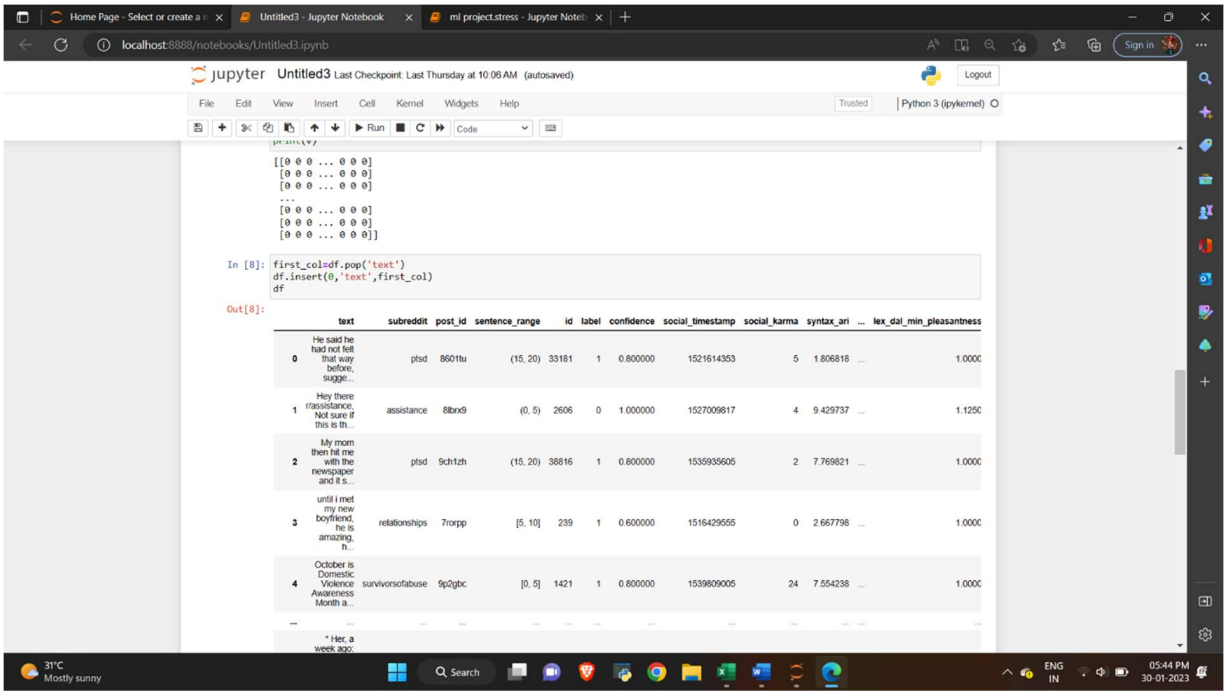
File Edit View Insert Cell Kernel Widgets Help

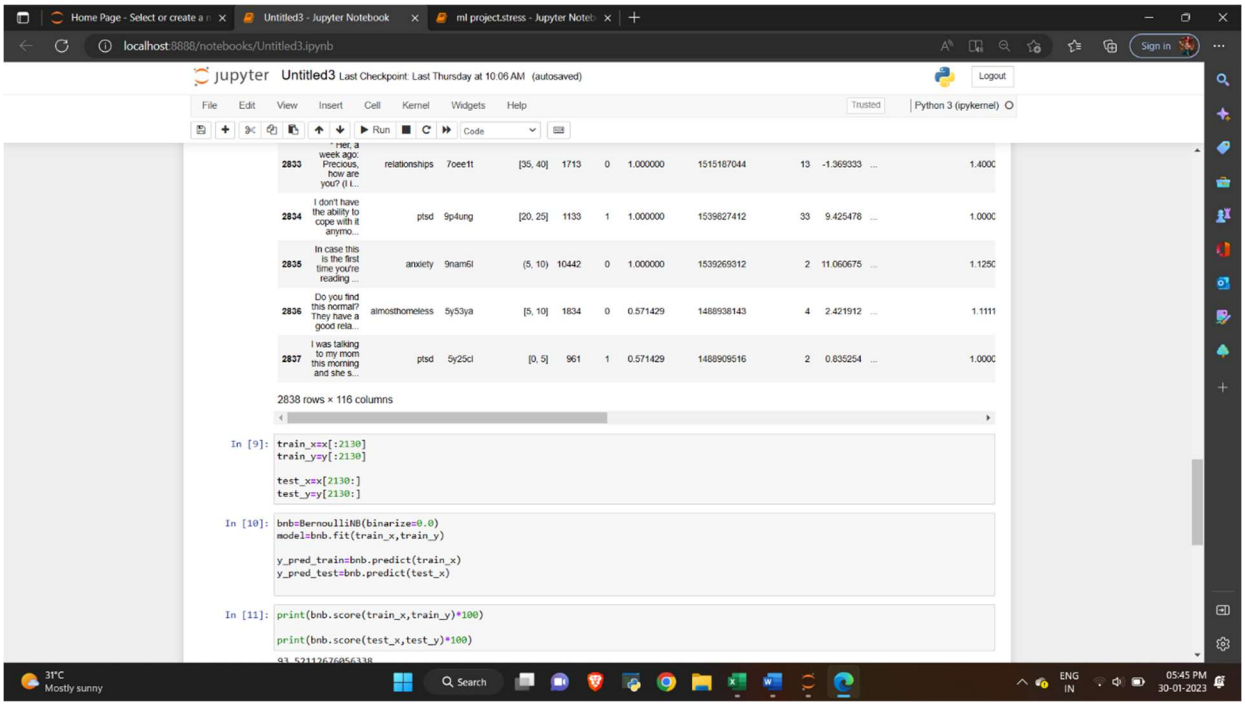
Trusted | Python 3 (ipykernel)

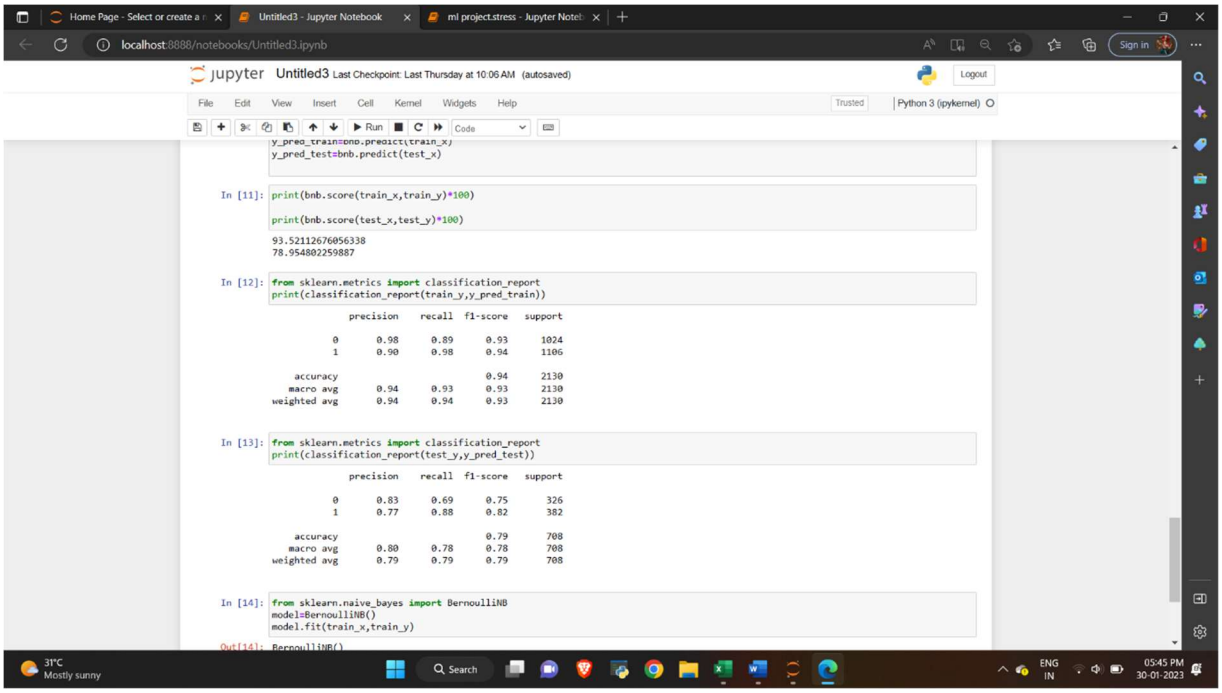
Run Code

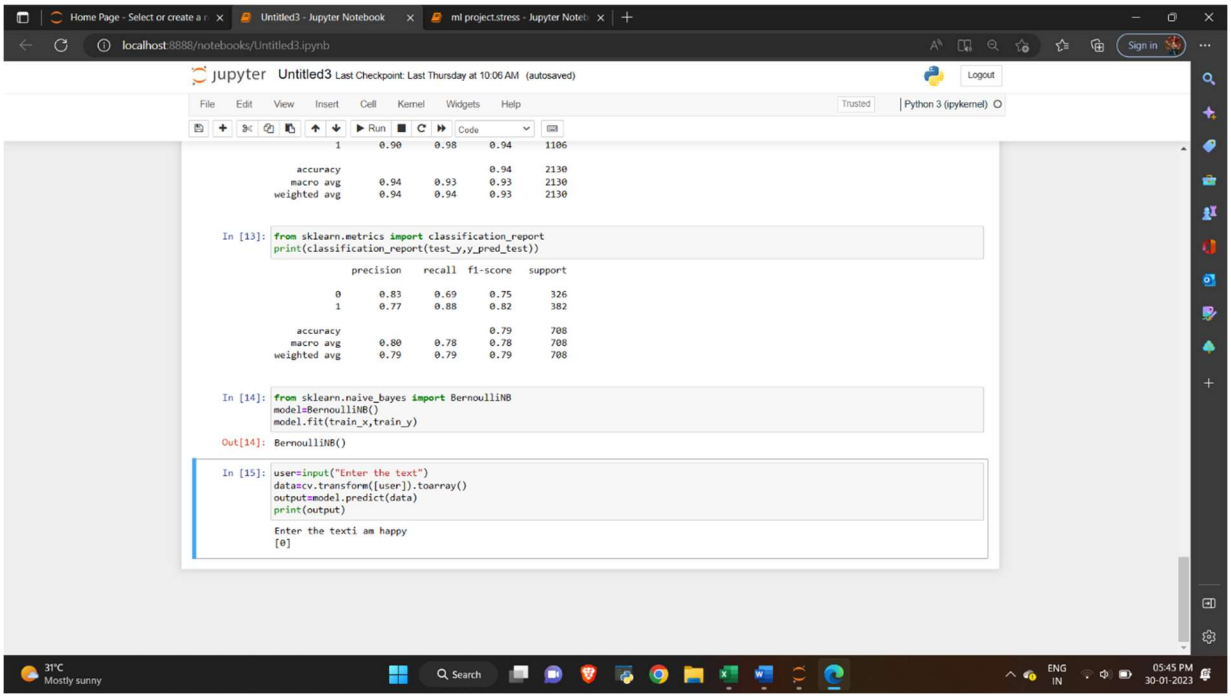
```
In [5]: np.unique(df['label'])
Out[5]: array([0, 1], dtype=int64)

In [6]: np.unique(df['text'])
Out[6]: array(["\tIt was the summer, I had just started a new job as a bartender.  I met a girl that was equally odd, had a lot of simi
        lar interests, was looking for the same thing I was.  We quickly became enamored with each other. It didn't take long for us to
        move in together.  I started finding projects around the house to fix up, she would surprise me with home cooked meals or a not
        e gushing over our relationship.",
               "                                                She resents my relationship with our son - My son is a Daddy's boy. Ev
        en now as I type this I hear him out there waiting for me saying Daddy Daddy Daddy. He comes and gives me hugs when I come home
        and barely notices when my wife is gone or comes home. - â\x80\x8eThis is also something she's recently admitted and I feel gui
        lty about it, but I feel like it's because she puts in no effort beyond what she needs to do.",
               "      He grew from a short, stubby, orange haired, freckle faced, awkward boy into a seemingly strong, tattooed, pierce
        d, tall, ginger, man I thought I always wanted.       From the time I was 13, I thought he loved me. I didn't know why, but he
        told me often; on the phone, through letter. I was the girl he thought he could never get. I was out of his league (his word
        s).",
               ...,
               'â\x80\x9cThe best time to plant a tree was 20 years ago. The second best time is today.â\x80\x9d My new favorite quote
        I know anxiety cripples all of us, but if we can accomplish one simple task that weâ\x80\x99ve been putting off, itâ\x80\x99s o
        ne stop closer to recovery. Today, I got a haircut. I made excuses for months and kept cancelling my appointments.',
               'â\x80\x9cYeah I saw, I think he out to steal.â\x80\x9d â\x80\x9cThatâ\x80\x99s what I think, keep an eye out man this p
        lace ainâ\x80\x99t bad but you gotta watch yourself.â\x80\x9d *nods* Then I fell asleep. And woke up to a stolen charger.',
               'â\x97\x8f active staff members â\x97\x8f active support staff â\x97\x8f helping channels and channels that you can shar
        e your art in, talk to others in, and voice chat in. â\x97\x8f rules that keep the community safe â\x97\x8f we have added a "cr
        eative corner"... this is a place for writers and artists of any kind to share their works and get feedback on it if they\'d li
        ke.'],
              dtype=object)

In [7]: x=df["text"].values
        y=df["label"].values
        cv=CountVectorizer()

        x=cv.fit_transform(x)
        v=x.toarray()

        print(v)

[[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
```

31°C Mostly sunny | Search | ENG IN | 05:44 PM 30-01-2023


Home Page - Select or create a n
Untitled3 - Jupyter Notebook
ml project.stress - Jupyter Noteb
localhost:8888/notebooks/Untitled3.ipynb
jupyter Untitled3 Last Checkpoint: Last Thursday at 10:06 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
[[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 0 0 ... 0 0 0]
[0 0 0 ... 0 0 0]]
In [8]:
first_col=df.pop('text')
df.insert(0,'text',first_col)
df
Out[8]:
text subreddit post_id sentence_range id label confidence social_timestamp social_karma syntax_ari ... lex_dal_min_pleasantness
0 He said he had not felt that way before, sugge... ptsd 8601tu (15, 20) 33181 1 0.800000 1521614353 5 1.806818 ... 1.0000
1 Hey there r/assistance, Not sure if this is th... assistance 8ibrx9 (0, 5) 2606 0 1.000000 1527009817 4 9.429737 ... 1.1250
2 My mom then hit me with the newspaper and it s... ptsd 9ch1zh (15, 20) 38816 1 0.800000 1535935605 2 7.769821 ... 1.0000
3 until i met my new boyfriend, he is amazing, h... relationships 7rorpp [5, 10] 239 1 0.600000 1516429555 0 2.667798 ... 1.0000
4 October is Domestic Violence Awareness Month a... survivorsofabuse 9p2gbc [0, 5] 1421 1 0.800000 1539809005 24 7.554238 ... 1.0000
... ...
* Her, a week ago:
31°C
Mostly sunny
Search
ENG
IN
05:44 PM
30-01-2023


Home Page - Select or create a n
Untitled3 - Jupyter Notebook
ml project.stress - Jupyter Noteb
localhost:8888/notebooks/Untitled3.ipynb
jupyter Untitled3 Last Checkpoint: Last Thursday at 10:06 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
2833 * Her, a week ago: Precious, how are you? (I l... relationships 7oee1t [35, 40] 1713 0 1.000000 1515187044 13 -1.369333 ... 1.4000
2834 I don't have the ability to cope with it anymo... ptsd 9p4ung [20, 25] 1133 1 1.000000 1539827412 33 9.425478 ... 1.0000
2835 In case this is the first time you're reading ... anxiety 9nam6l (5, 10) 10442 0 1.000000 1539269312 2 11.060675 ... 1.1250
2836 Do you find this normal? They have a good rela... almosthomeless 5y53ya [5, 10] 1834 0 0.571429 1488938143 4 2.421912 ... 1.1111
2837 I was talking to my mom this morning and she s... ptsd 5y25cl [0, 5] 961 1 0.571429 1488909516 2 0.835254 ... 1.0000
2838 rows × 116 columns
In [9]:
train_x=x[:2130]
train_y=y[:2130]
test_x=x[2130:]
test_y=y[2130:]
In [10]:
bnb=BernoulliNB(binarize=0.0)
model=bnb.fit(train_x,train_y)
y_pred_train=bnb.predict(train_x)
y_pred_test=bnb.predict(test_x)
In [11]:
print(bnb.score(train_x,train_y)*100)
print(bnb.score(test_x,test_y)*100)
93.52112676056338
31°C
Mostly sunny
Search
ENG
IN
05:45 PM
30-01-2023


Home Page - Select or create a r
Untitled3 - Jupyter Notebook
ml project.stress - Jupyter Noteb
localhost:8888/notebooks/Untitled3.ipynb
Sign in
jupyter Untitled3 Last Checkpoint: Last Thursday at 10:06 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
y_pred_test=bnb.predict(test_x)
In [11]:
print(bnb.score(train_x,train_y)*100)
print(bnb.score(test_x,test_y)*100)
93.52112676056338
78.954802259887
In [12]:
from sklearn.metrics import classification_report
print(classification_report(train_y,y_pred_train))
precision recall f1-score support
0 0.98 0.89 0.93 1024
1 0.90 0.98 0.94 1106
accuracy 0.94 2130
macro avg 0.94 0.93 0.93 2130
weighted avg 0.94 0.94 0.93 2130
In [13]:
from sklearn.metrics import classification_report
print(classification_report(test_y,y_pred_test))
precision recall f1-score support
0 0.83 0.69 0.75 326
1 0.77 0.88 0.82 382
accuracy 0.79 708
macro avg 0.80 0.78 0.78 708
weighted avg 0.79 0.79 0.79 708
In [14]:
from sklearn.naive_bayes import BernoulliNB
model=BernoulliNB()
model.fit(train_x,train_y)
Out[14]:
BernoulliNB()
31°C
Mostly sunny
Search
ENG
IN
05:45 PM
30-01-2023


Home Page - Select or create a r
Untitled3 - Jupyter Notebook
ml project.stress - Jupyter Noteb
localhost:8888/notebooks/Untitled3.ipynb
Sign in
jupyter Untitled3 Last Checkpoint: Last Thursday at 10:06 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
1 0.90 0.98 0.94 1106
accuracy 0.94 2130
macro avg 0.94 0.93 0.93 2130
weighted avg 0.94 0.94 0.93 2130
In [13]:
from sklearn.metrics import classification_report
print(classification_report(test_y,y_pred_test))
precision recall f1-score support
0 0.83 0.69 0.75 326
1 0.77 0.88 0.82 382
accuracy 0.79 708
macro avg 0.80 0.78 0.78 708
weighted avg 0.79 0.79 0.79 708
In [14]:
from sklearn.naive_bayes import BernoulliNB
model=BernoulliNB()
model.fit(train_x,train_y)
Out[14]:
BernoulliNB()
In [15]:
user=input("Enter the text")
data=cv.transform([user]).toarray()
output=model.predict(data)
print(output)
Enter the texti am happy
[0]
31°C
Mostly sunny
Search
ENG
IN
05:45 PM
30-01-2023

CODE

```
[{"metadata":{"trusted":true},"id":"c0ce0666","cell_type":"code","sourc
e":"import numpy as np\nimport pandas as pd \nimport seaborn as
sns\nfrom matplotlib import pyplot as plt\nfrom sklearn.naive_bayes
import BernoulliNB\nfrom sklearn.feature_extraction.text import
CountVectorizer","execution_count":1,"outputs":[]},{"metadata":{"truste
d":true},"id":"7a661285","cell_type":"code","source":"df=pd.read_csv(\"
C:/Users/guntuk kiran/OneDrive/Desktop/stress.csv\",encoding=\"latin-
1\")","execution_count":2,"outputs":[]},{"metadata":{"trusted":true},"i
d":"bc22bbab","cell_type":"code","source":"df.head(n=10)","execution_co
unt":3,"outputs":[{"output_type":"execute_result","execution_count":3,"
data":{"text/plain":"          subreddit post_id sentence_range  \\\n0
ptsd  8601tu       (15, 20)   \n1        assistance  8lbrx9         (0,
5)   \n2             ptsd  9ch1zh       (15, 20)   \n3
relationships  7rorpp        [5, 10]   \n4  survivorsofabuse  9p2gbc
[0, 5]   \n5     relationships  7tx7et       (30, 35)   \n6
domesticviolence  7iphly       [25, 30]   \n7           anxiety  5m3k80
(5, 10)   \n8     relationships  7nhy1v       (50, 55)   \n9
assistance  61eiq6       [15, 20]   \n\n
text     id  label  \\\n0  He said he had not felt that way before,
sugge...  33181      1   \n1  Hey there r/assistance, Not sure if this
is th...   2606      0   \n2  My mom then hit me with the newspaper and
it s...  38816      1   \n3  until i met my new boyfriend, he is
amazing, h...    239      1   \n4  October is Domestic Violence
Awareness Month a...   1421      1   \n5  I think he doesn't want to
put in the effort f...  17554      1   \n6  It was a big company so
luckily I didn't have ...    165      0   \n7    It cleared up and I
was okay but. On Monday ...  33053      1   \n8  I actually give an
assistant half my emergency...   7581      1   \n9  I just feel like
the street life has fucked my...   1517      1   \n\n   confidence
social_timestamp  social_karma  syntax_ari  ...  \\\n0         0.8
1521614353             5    1.806818  ...   \n1         1.0
1527009817             4    9.429737  ...   \n2         0.8
1535935605             2    7.769821  ...   \n3         0.6
1516429555             0    2.667798  ...   \n4         0.8
1539809005            24    7.554238  ...   \n5         1.0
1517274027             2    3.531124  ...   \n6         0.8
1512854409             6    8.331463  ...   \n7         0.8
1483582174             1    0.403679  ...   \n8         0.6
1514843984           134    5.381111  ...   \n9         1.0
1490428087            20    1.562351  ...   \n\n
lex_dal_min_pleasantness  lex_dal_min_activation  lex_dal_min_imagery
\\\n0                    1.000                  1.1250
1.0   \n1                    1.125                  1.0000
1.0   \n2                    1.000                  1.1429
1.0   \n3                    1.000                  1.1250
1.0   \n4                    1.000                  1.1250
1.0   \n5                    1.000                  1.1818
1.0   \n6                    1.000                  1.1250
1.0   \n7                    1.000                  1.1429
1.0   \n8                    1.000                  1.0000
1.0   \n9                    1.000                  1.1250
1.0   \n\n   lex_dal_avg_activation  lex_dal_avg_imagery
lex_dal_avg_pleasantness  \\\n0                 1.77000
1.52211                  1.89556   \n1                 1.69586
1.62045                  1.88919   \n2                 1.83088
```

```
1.58108                    1.85828   \n3                 1.75356
1.52114                    1.98848   \n4                 1.77644
1.64872                    1.81456   \n5                 1.70415
1.44396                    1.82046   \n6                 1.66985
1.57757                    1.89926   \n7                 1.69549
1.44860                    1.88963   \n8                 1.73571
1.71471                    1.93607   \n9                 1.74956
1.67353                    1.89320   \n\n   social_upvote_ratio
social_num_comments  syntax_fk_grade  sentiment  \n0
0.86                     1         3.253573  -0.002742  \n1
0.65                     2         8.828316   0.292857  \n2
0.67                     0         7.841667   0.011894  \n3
0.50                     5         4.104027   0.141671  \n4
1.00                     1         7.910952  -0.204167  \n5
0.75                     5         4.338429  -0.550000  \n6
0.88                    29         8.394244   0.277381  \n7
1.00                     0         1.028654  -0.079630  \n8
0.94                    43         5.991049  -0.272917  \n9
0.80                     6         2.851935  -0.117262  \n\n[10 rows x
116 columns]","text/html":"
\n\n    \n      \n      subreddit\n      post_id\n
sentence_range\n      text\n      id\n      label\n      confidence\n
social_timestamp\n      social_karma\n      syntax_ari\n      ...\n
lex_dal_min_pleasantness\n      lex_dal_min_activation\n
lex_dal_min_imagery\n      lex_dal_avg_activation\n
lex_dal_avg_imagery\n      lex_dal_avg_pleasantness\n
social_upvote_ratio\n      social_num_comments\n      syntax_fk_grade\n
sentiment\n    \n  \n  \n    \n      0\n      ptsd\n      8601tu\n
(15, 20)\n      He said he had not felt that way before, sugge...\n
33181\n      1\n      0.8\n      1521614353\n      5\n      1.806818\n
...\n      1.000\n      1.1250\n      1.0\n      1.77000\n
1.52211\n      1.89556\n      0.86\n      1\n      3.253573\n      -
0.002742\n    \n    \n      1\n      assistance\n      8lbrx9\n
(0, 5)\n      Hey there r/assistance, Not sure if this is th...\n
2606\n      0\n      1.0\n      1527009817\n      4\n      9.429737\n
...\n      1.125\n      1.0000\n      1.0\n      1.69586\n
1.62045\n      1.88919\n      0.65\n      2\n      8.828316\n
0.292857\n    \n    \n      2\n      ptsd\n      9ch1zh\n      (15,
20)\n      My mom then hit me with the newspaper and it s...\n
38816\n      1\n      0.8\n      1535935605\n      2\n      7.769821\n
...\n      1.000\n      1.1429\n      1.0\n      1.83088\n
1.58108\n      1.85828\n      0.67\n      0\n      7.841667\n
0.011894\n    \n    \n      3\n      relationships\n      7rorpp\n
[5, 10]\n      until i met my new boyfriend, he is amazing, h...\n
239\n      1\n      0.6\n      1516429555\n      0\n      2.667798\n
...\n      1.000\n      1.1250\n      1.0\n      1.75356\n
1.52114\n      1.98848\n      0.50\n      5\n      4.104027\n
0.141671\n    \n    \n      4\n      survivorsofabuse\n      9p2gbc\n
[0, 5]\n      October is Domestic Violence Awareness Month a...\n
1421\n      1\n      0.8\n      1539809005\n      24\n      7.554238\n
...\n      1.000\n      1.1250\n      1.0\n      1.77644\n
1.64872\n      1.81456\n      1.00\n      1\n      7.910952\n      -
0.204167\n    \n    \n      5\n      relationships\n      7tx7et\n
(30, 35)\n      I think he doesn't want to put in the effort f...\n
17554\n      1\n      1.0\n      1517274027\n      2\n      3.531124\n
...\n      1.000\n      1.1818\n      1.0\n      1.70415\n
1.44396\n      1.82046\n      0.75\n      5\n      4.338429\n      -
```

0.550000\n   \n   \n   6\n   domesticviolence\n   7iphly\n
[25, 30]\n   It was a big company so luckily I didn't have ...\n
165\n   0\n   0.8\n   1512854409\n   6\n   8.331463\n
...\n   1.000\n   1.1250\n   1.0\n   1.66985\n
1.57757\n   1.89926\n   0.88\n   29\n   8.394244\n
0.277381\n   \n   \n   7\n   anxiety\n   5m3k80\n   (5,
10)\n   It cleared up and I was okay but. On Monday ...\n
33053\n   1\n   0.8\n   1483582174\n   1\n   0.403679\n
...\n   1.000\n   1.1429\n   1.0\n   1.69549\n
1.44860\n   1.88963\n   1.00\n   0\n   1.028654\n   -
0.079630\n   \n   \n   8\n   relationships\n   7nhy1v\n
(50, 55)\n   I actually give an assistant half my emergency...\n
7581\n   1\n   0.6\n   1514843984\n   134\n   5.381111\n
...\n   1.000\n   1.0000\n   1.0\n   1.73571\n
1.71471\n   1.93607\n   0.94\n   43\n   5.991049\n   -
0.272917\n   \n   \n   9\n   assistance\n   61eiq6\n
[15, 20]\n   I just feel like the street life has fucked my...\n
1517\n   1\n   1.0\n   1490428087\n   20\n   1.562351\n
...\n   1.000\n   1.1250\n   1.0\n   1.74956\n
1.67353\n   1.89320\n   0.80\n   6\n   2.851935\n   -
0.117262\n   \n \n\n

10 rows × 116 columns

\n
"},"metadata":{}}]},{"metadata":{"trusted":true},"id":"b13ac533","cell_
type":"code","source":"df.shape","execution_count":4,"outputs":[{"outpu
t_type":"execute_result","execution_count":4,"data":{"text/plain":"(283
8,
116)"},"metadata":{}}]},{"metadata":{"trusted":true},"id":"967ba2d4","c
ell_type":"code","source":"np.unique(df['label'])","execution_count":5,
"outputs":[{"output_type":"execute_result","execution_count":5,"data":{
"text/plain":"array([0, 1],
dtype=int64)"},"metadata":{}}]},{"metadata":{"trusted":true},"id":"9959
3882","cell_type":"code","source":"np.unique(df['text'])","execution_co
unt":6,"outputs":[{"output_type":"execute_result","execution_count":6,"
data":{"text/plain":"array([\"\\tIt was the summer, I had just started
a new job as a bartender. I met a girl that was equally odd, had a lot
of similar interests, was looking for the same thing I was. We quickly
became enamored with each other. It didn't take long for us to move in
together. I started finding projects around the house to fix up, she
would surprise me with home cooked meals or a note gushing over our
relationship.\",\n   \"
She resents my relationship with our son - My son is a Daddy's boy.
Even now as I type this I hear him out there waiting for me saying
Daddy Daddy Daddy. He comes and gives me hugs when I come home and
barely notices when my wife is gone or comes home. - â\\x80\\x8eThis is
also something she's recently admitted and I feel guilty about it, but
I feel like it's because she puts in no effort beyond what she needs to
do.\",\n   \"   He grew from a short, stubby, orange haired,
freckle faced, awkward boy into a seemingly strong, tattooed, pierced,
tall, ginger, man I thought I always wanted.   From the time I was
13, I thought he loved me. I didn't know why, but he told me often; on
the phone, through letter. I was the girl he thought he could never
get. I was out of his league (his words).\",\n   ...,\n
'â\\x80\\x9cThe best time to plant a tree was 20 years ago. The second

```
best time is today.â\\x80\\x9d My new favorite quote I know anxiety
cripples all of us, but if we can accomplish one simple task that
weâ\\x80\\x99ve been putting off, itâ\\x80\\x99s one stop closer to
recovery. Today, I got a haircut. I made excuses for months and kept
cancelling my appointments.',\n       'â\\x80\\x9cYeah I saw, I think
he out to steal.â\\x80\\x9d â\\x80\\x9cThatâ\\x80\\x99s what I think,
keep an eye out man this place ainâ\\x80\\x99t bad but you gotta watch
yourself.â\\x80\\x9d *nods* Then I fell asleep. And woke up to a stolen
charger.',\n       'â\\x97\\x8f active staff members â\\x97\\x8f active
support staff â\\x97\\x8f helping channels and channels that you can
share your art in, talk to others in, and voice chat in. â\\x97\\x8f
rules that keep the community safe â\\x97\\x8f we have added a
\"creative corner\"... this is a place for writers and artists of any
kind to share their works and get feedback on it if they\\'d like.'],\n
dtype=object)"},"metadata":{}}]},{"metadata":{"trusted":true},"id":"8b0
7fee7","cell_type":"code","source":"x=df[\"text\"].values\ny=df[\"label
\"].values\ncv=CountVectorizer()\n\nx=cv.fit_transform(x)\nv=x.toarray(
)\n\nprint(v)","execution_count":7,"outputs":[{"output_type":"stream","
text":"[[0 0 0 ... 0 0 0]\n [0 0 0 ... 0 0 0]\n [0 0 0 ... 0 0 0]\n
...\n [0 0 0 ... 0 0 0]\n [0 0 0 ... 0 0 0]\n [0 0 0 ... 0 0
0]]\n","name":"stdout"}]},{"metadata":{"trusted":true},"id":"13a5bd97",
"cell_type":"code","source":"first_col=df.pop('text')\ndf.insert(0,'tex
t',first_col)\ndf","execution_count":8,"outputs":[{"output_type":"execu
te_result","execution_count":8,"data":{"text/plain":"
text         subreddit  \\\n0     He said he had not felt that way
before, sugge...              ptsd   \n1     Hey there r/assistance,
Not sure if this is th...        assistance   \n2     My mom then hit
me with the newspaper and it s...              ptsd   \n3     until i
met my new boyfriend, he is amazing, h...     relationships   \n4
October is Domestic Violence Awareness Month a...  survivorsofabuse
\n...                                                  ...
...   \n2833  * Her, a week ago: Precious, how are you? (I i...
relationships   \n2834  I don't have the ability to cope with it
anymo...              ptsd   \n2835  In case this is the first time
you're reading ...           anxiety   \n2836  Do you find this normal?
They have a good rela...    almosthomeless   \n2837  I was talking to
my mom this morning and she s...              ptsd   \n\n     post_id
sentence_range     id  label  confidence  social_timestamp  \\\n0
8601tu       (15, 20)  33181      1    0.800000        1521614353   \n1
8lbrx9         (0, 5)   2606      0    1.000000        1527009817   \n2
9ch1zh       (15, 20)  38816      1    0.800000        1535935605   \n3
7rorpp        [5, 10]    239      1    0.600000        1516429555   \n4
9p2gbc         [0, 5]   1421      1    0.800000        1539809005
\n...      ...            ...    ...    ...         ...
...   \n2833  7oee1t       [35, 40]   1713      0    1.000000
1515187044   \n2834  9p4ung       [20, 25]   1133      1    1.000000
1539827412   \n2835  9nam6l        (5, 10)  10442      0    1.000000
1539269312   \n2836  5y53ya        [5, 10]   1834      0    0.571429
1488938143   \n2837  5y25cl         [0, 5]    961      1    0.571429
1488909516   \n\n      social_karma  syntax_ari  ...
lex_dal_min_pleasantness  \\\n0               5    1.806818  ...
1.0000   \n1                4    9.429737  ...
1.1250   \n2                2    7.769821  ...
1.0000   \n3                0    2.667798  ...
1.0000   \n4               24    7.554238  ...
1.0000   \n...            ...         ...  ...
...   \n2833           13   -1.369333  ...                    1.4000
```

```
\n2834            33    9.425478  ...                    1.0000
\n2835             2   11.060675  ...                    1.1250
\n2836             4    2.421912  ...                    1.1111
\n2837             2    0.835254  ...                    1.0000   \n\n
lex_dal_min_activation  lex_dal_min_imagery  lex_dal_avg_activation
\\\n0                     1.1250                  1.0
1.77000   \n1                    1.0000                  1.0
1.69586   \n2                    1.1429                  1.0
1.83088   \n3                    1.1250                  1.0
1.75356   \n4                    1.1250                  1.0
1.77644   \n...                     ...                  ...
...   \n2833                 1.0000                  1.0
1.71133   \n2834                 1.0000                  1.0
1.65003   \n2835                 1.1250                  1.0
1.79768   \n2836                 1.1429                  1.0
1.71642   \n2837                 1.0000                  1.0
1.68891   \n\n      lex_dal_avg_imagery  lex_dal_avg_pleasantness
social_upvote_ratio  \\\n0                 1.52211
1.89556                  0.86   \n1                 1.62045
1.88919                  0.65   \n2                 1.58108
1.85828                  0.67   \n3                 1.52114
1.98848                  0.50   \n4                 1.64872
1.81456                  1.00   \n...                   ...
...                  ...   \n2833             1.45301
2.00304                  0.84   \n2834             1.56842
1.81527                  0.96   \n2835             1.49074
1.92286                  1.00   \n2836             1.57627
1.89972                  0.75   \n2837             1.44615
1.89707                  0.76   \n\n      social_num_comments
syntax_fk_grade  sentiment  \n0                      1
3.253573  -0.002742  \n1                      2         8.828316
0.292857  \n2                      0         7.841667   0.011894  \n3
5         4.104027   0.141671  \n4                      1
7.910952  -0.204167  \n...                    ...               ...
...  \n2833                   16         0.254444   0.552066  \n2834
6         8.640664  -0.220370  \n2835                   1
9.951524   0.045455  \n2836                    7         4.036765
0.159722  \n2837                   2         2.412000   0.016667
\n\n[2838 rows x 116 columns]","text/html":"
\n\n    \n      \n      text\n      subreddit\n      post_id\n
sentence_range\n      id\n      label\n      confidence\n
social_timestamp\n      social_karma\n      syntax_ari\n      ...\n
lex_dal_min_pleasantness\n      lex_dal_min_activation\n
lex_dal_min_imagery\n      lex_dal_avg_activation\n
lex_dal_avg_imagery\n      lex_dal_avg_pleasantness\n
social_upvote_ratio\n      social_num_comments\n      syntax_fk_grade\n
sentiment\n    \n  \n  \n    \n      0\n      He said he had not felt
that way before, sugge...\n      ptsd\n      8601tu\n      (15, 20)\n
33181\n      1\n      0.800000\n      1521614353\n      5\n
1.806818\n      ...\n      1.0000\n      1.1250\n      1.0\n
1.77000\n      1.52211\n      1.89556\n      0.86\n      1\n
3.253573\n      -0.002742\n    \n    \n      1\n      Hey there
r/assistance, Not sure if this is th...\n      assistance\n
81brx9\n      (0, 5)\n      2606\n      0\n      1.000000\n
1527009817\n      4\n      9.429737\n      ...\n      1.1250\n
1.0000\n      1.0\n      1.69586\n      1.62045\n      1.88919\n
0.65\n      2\n      8.828316\n      0.292857\n    \n    \n      2\n
```

My mom then hit me with the newspaper and it s...\n ptsd\n 9ch1zh\n (15, 20)\n 38816\n 1\n 0.800000\n 1535935605\n 2\n 7.769821\n ...\n 1.0000\n 1.1429\n 1.0\n 1.83088\n 1.58108\n 1.85828\n 0.67\n 0\n 7.841667\n 0.011894\n \n \n 3\n until i met my new boyfriend, he is amazing, h...\n relationships\n 7rorpp\n [5, 10]\n 239\n 1\n 0.600000\n 1516429555\n 0\n 2.667798\n ...\n 1.0000\n 1.1250\n 1.0\n 1.75356\n 1.52114\n 1.98848\n 0.50\n 5\n 4.104027\n 0.141671\n \n \n 4\n October is Domestic Violence Awareness Month a...\n survivorsofabuse\n 9p2gbc\n [0, 5]\n 1421\n 1\n 0.800000\n 1539809005\n 24\n 7.554238\n ...\n 1.0000\n 1.1250\n 1.0\n 1.77644\n 1.64872\n 1.81456\n 1.00\n 1\n 7.910952\n -0.204167\n \n \n ...\n \n \n 2833\n * Her, a week ago: Precious, how are you? (I i...\n relationships\n 7oee1t\n [35, 40]\n 1713\n 0\n 1.000000\n 1515187044\n 13\n -1.369333\n ...\n 1.4000\n 1.0000\n 1.0\n 1.71133\n 1.45301\n 2.00304\n 0.84\n 16\n 0.254444\n 0.552066\n \n \n 2834\n I don't have the ability to cope with it anymo...\n ptsd\n 9p4ung\n [20, 25]\n 1133\n 1\n 1.000000\n 1539827412\n 33\n 9.425478\n ...\n 1.0000\n 1.0000\n 1.0\n 1.65003\n 1.56842\n 1.81527\n 0.96\n 6\n 8.640664\n -0.220370\n \n \n 2835\n In case this is the first time you're reading ...\n anxiety\n 9nam6l\n (5, 10)\n 10442\n 0\n 1.000000\n 1539269312\n 2\n 11.060675\n ...\n 1.1250\n 1.1250\n 1.0\n 1.79768\n 1.49074\n 1.92286\n 1.00\n 1\n 9.951524\n 0.045455\n \n \n 2836\n Do you find this normal? They have a good rela...\n almosthomeless\n 5y53ya\n [5, 10]\n 1834\n 0\n 0.571429\n 1488938143\n 4\n 2.421912\n ...\n 1.1111\n 1.1429\n 1.0\n 1.71642\n 1.57627\n 1.89972\n 0.75\n 7\n 4.036765\n 0.159722\n \n \n 2837\n I was talking to my mom this morning and she s...\n ptsd\n 5y25cl\n [0, 5]\n 961\n 1\n 0.571429\n 1488909516\n 2\n 0.835254\n ...\n 1.0000\n 1.0000\n 1.0\n 1.68891\n 1.44615\n 1.89707\n 0.76\n 2\n 2.412000\n 0.016667\n \n \n\n

2838 rows × 116 columns

\n
"},"metadata":{}}]},{"metadata":{"trusted":true},"id":"bcc4ea1b","cell_type":"code","source":"train_x=x[:2130]\ntrain_y=y[:2130]\n\ntest_x=x[2130:]\ntest_y=y[2130:]","execution_count":9,"outputs":[]},{"metadata":{"trusted":true},"id":"f2d83459","cell_type":"code","source":"bnb=BernoulliNB(binarize=0.0)\nmodel=bnb.fit(train_x,train_y)\n\ny_pred_train=bnb.predict(train_x)\ny_pred_test=bnb.predict(test_x)\n","execution_count":10,"outputs":[]},{"metadata":{"trusted":true},"id":"3b1eaa04","cell_type":"code","source":"print(bnb.score(train_x,train_y)*100)\n\nprint(bnb

.score(test_x,test_y)*100)","execution_count":11,"outputs":[{"output_type":"stream","text":"93.52112676056338\n78.954802259887\n","name":"stdout"}]},{"metadata":{"trusted":true},"id":"72e48198","cell_type":"code","source":"from sklearn.metrics import classification_report\nprint(classification_report(train_y,y_pred_train))","execution_count":12,"outputs":[{"output_type":"stream","text":"   precision   recall f1-score   support\n\n   0   0.98   0.89   0.93   1024\n   1   0.90   0.98   0.94   1106\n\n   accuracy   0.94   2130\n   macro avg   0.94   0.93   0.93   2130\nweighted avg   0.94   0.94   0.93   2130\n\n","name":"stdout"}]},{"metadata":{"trusted":true},"id":"20439ab1","cell_type":"code","source":"from sklearn.metrics import classification_report\nprint(classification_report(test_y,y_pred_test))","execution_count":13,"outputs":[{"output_type":"stream","text":"   precision   recall f1-score   support\n\n   0   0.83   0.69   0.75   326\n   1   0.77   0.88   0.82   382\n\n   accuracy   0.79   708\n   macro avg   0.80   0.78   0.78   708\nweighted avg   0.79   0.79   0.79   708\n\n","name":"stdout"}]},{"metadata":{"trusted":true},"id":"d3719504","cell_type":"code","source":"from sklearn.naive_bayes import BernoulliNB\nmodel=BernoulliNB()\nmodel.fit(train_x,train_y)","execution_count":14,"outputs":[{"output_type":"execute_result","execution_count":14,"data":{"text/plain":"BernoulliNB()"},"metadata":{}}]},{"metadata":{"trusted":true},"id":"c9ce7d6e","cell_type":"code","source":"user=input(\"Enter the text\")\ndata=cv.transform([user]).toarray()\noutput=model.predict(data)\nprint(output)","execution_count":15,"outputs":[{"output_type":"stream","name":"stdout","text":"Enter the texti am happy\n[0]\n"}]}]

----------→THE END←----------