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

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
x = pd.read_csv("/content/drive/MyDrive/health.csv")
```

```
x.head()
```

	text	label
0	I recently went through a breakup and she said...	depression
1	I do not know how to navigate these feelings, ...	depression
2	So I have been with my bf for 5 months , and h...	depression
3	I am so exhausted of this. Just when I think I...	SuicideWatch
4	I have been severly bullied since i was 5 till...	depression

```
x.shape
```

```
(20366, 2)
```

```
x.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20366 entries, 0 to 20365
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   text    20365 non-null  object
 1   label   20363 non-null  object
dtypes: object(2)
memory usage: 318.3+ KB
```

```
x.dropna()
```

	text	label
0	I recently went through a breakup and she said...	depression
1	I do not know how to navigate these feelings, ...	depression
2	So I have been with my bf for 5 months , and h...	depression
3	I am so exhausted of this. Just when I think I...	SuicideWatch
4	I have been severly bullied since i was 5 till...	depression
...	...	...
20361	I took 50mg of seroquel a few hours after I dr...	SuicideWatch
20362	that is what has happened to me last week. And...	depression
20363	Ever just feel alone in a house full of people...	depression
20364	Politicians. Neighbors. Corporations. Society....	depression
20365	I feel like I am just existing, but for what. ...	depression

20363 rows × 2 columns

```
from sklearn.model_selection import train_test_split
```

```
df = pd.DataFrame(x)
```

```
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics import accuracy_score, classification_report,f1_score
from sklearn.naive_bayes import MultinomialNB
import re
import pandas as pd
```

```
x['text'].dtypes
```

```
dtype('O')
```

```
def preprocess_text(text):
    if isinstance(text, str):
        text = text.lower()
        text = re.sub(r"[^a-zA-Z\s]", "", text)
    return text
```

```
x['text'] = x['text'].apply(preprocess_text)
```

```
x.dropna(subset=['text'],inplace=True)
x['text'].fillna('',inplace=True)
```

```
X = x['text']
y = x['label']
```

```
X_train,X_test,y_train,y_test= train_test_split(X,y,test_size=0.2,random_state=42)
```

```
vectorizer = CountVectorizer()
X_train_vectorized = vectorizer.fit_transform(X_train)
X_test_vectorized = vectorizer.transform(X_test)
```

```
y.value_counts()
```

```
depression      10371
SuicideWatch     9992
```

```
Name: label, dtype: int64
```

```
y=pd.get_dummies(y)
```

```
y=y.drop(['depression'],axis=1)
```

```
y
```

	SuicideWatch
0	0
1	0
2	0
3	1
4	0
...	...
20361	1
20362	0
20363	0
20364	0
20365	0

20365 rows × 1 columns

```
x = x.join(y)
```

```
x
```

	text	label	SuicideWatch
0	i recently went through a breakup and she said...	depression	0
1	i do not know how to navigate these feelings n...	depression	0
2	so i have been with my bf for months and he ...	depression	0
3	i am so exhausted of this just when i think i ...	SuicideWatch	1
4	i have been severly bullied since i was till ...	depression	0
...	...	...	...
20361	i took mg of seroquel a few hours after i dran...	SuicideWatch	1
20362	that is what has happened to me last week and ...	depression	0
20363	ever just feel alone in a house full of people...	depression	0
20364	politicians neighbors corporations society cul...	depression	0
20365	i feel like i am just existing but for what i ...	depression	0

20365 rows × 3 columns

```
x=x.drop(['label'],axis=1)
```

```
x.head()
```

	text	SuicideWatch
0	i recently went through a breakup and she said...	0
1	i do not know how to navigate these feelings n...	0
2	so i have been with my bf for months and he ...	0
3	i am so exhausted of this just when i think i ...	1
4	i have been severly bullied since i was till ...	0

```
x_train,x_test,y_train,y_test= train_test_split(x,y,test_size=0.2,random_state=42)
```

```
vectorizer = CountVectorizer()
X_train_vectorized = vectorizer.fit_transform(X_train)
X_test_vectorized = vectorizer.transform(X_test)
```

```
nb_classifier = MultinomialNB()
nb_classifier.fit(X_train_vectorized,y_train)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/utils/validation.py:1143: DataConversio
  y = column_or_1d(y, warn=True)
```

MultinomialNB

MultinomialNB()

```
y_pred = nb_classifier.predict(X_test_vectorized)
```

```
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:",accuracy*100)
```

```
Accuracy: 73.14019150503314
```

```
f1 = f1_score(y_test, y_pred)
print("F1 Score:", f1)
```

```
F1 Score: 0.738527724665392
```

```
# Print classification report
print("Classification Report:")
print(classification_report(y_test, y_pred))
```

```
Classification Report:
              precision    recall  f1-score   support

           0       0.74      0.71      0.72      2027
           1       0.72      0.76      0.74      2046

    accuracy                           0.73      4073
   macro avg       0.73      0.73      0.73      4073
weighted avg       0.73      0.73      0.73      4073
```

```
import joblib
joblib.dump(nb_classifier, 'naivebayes_model.joblib')
```

```
['naivebayes_model.joblib']
```

```
loaded_model = joblib.load('naivebayes_model.joblib')
```

```
joblib.dump(vectorizer, 'vectorizer.joblib')
```

```
['vectorizer.joblib']
```

```
loaded_model = joblib.load('vectorizer.joblib')
```