mentalhealthprediction

August 17, 2024

```
[1]: import pandas as pd
     data=pd.read_csv('Combined Data.csv',index_col=0)
     data.head()
[1]:
                                                statement
                                                            status
                                               oh my gosh Anxiety
     1 trouble sleeping, confused mind, restless hear... Anxiety
     2 All wrong, back off dear, forward doubt. Stay ... Anxiety
     3 I've shifted my focus to something else but I'... Anxiety
     4 I'm restless and restless, it's been a month n... Anxiety
[2]: data = data.dropna()
[3]: import numpy as np
     from sklearn.model_selection import train_test_split
     from sklearn.feature_extraction.text import TfidfVectorizer
     from sklearn.ensemble import RandomForestClassifier
     from sklearn.metrics import classification_report
     from imblearn.over_sampling import SMOTE
[4]: X_train, X_test, y_train, y_test = train_test_split(data['statement'],__

data['status'], test_size=0.2, random_state=42)

[5]: vectorizer = TfidfVectorizer(max_features=5000)
     X train vectorized = vectorizer.fit transform(X train)
     X_test_vectorized = vectorizer.transform(X_test)
[6]: smote = SMOTE(random_state=42)
     X_train_resampled, y_train_resampled = smote.fit_resample(X_train_vectorized,__
      →y_train)
[7]: clf = RandomForestClassifier(n_estimators=100, random_state=42)
     clf.fit(X_train_resampled, y_train_resampled)
[7]: RandomForestClassifier(random_state=42)
```

```
[8]: y_pred = clf.predict(X_test_vectorized)
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
Anxiety	0.78	0.74	0.76	755
Bipolar	0.78	0.67	0.70	527
•				
Depression	0.65	0.70	0.67	3016
Normal	0.86	0.94	0.90	3308
Personality disorder	0.93	0.49	0.64	237
Stress	0.73	0.48	0.58	536
Suicidal	0.64	0.62	0.63	2158
accuracy			0.74	10537
macro avg	0.78	0.66	0.71	10537
weighted avg	0.74	0.74	0.74	10537

```
[13]: new_sentences = ["oh my gosh", "Everything is great today"]
new_sentences_vectorized = vectorizer.transform(new_sentences)
predictions = clf.predict(new_sentences_vectorized)
print(predictions)
```

['Anxiety' 'Normal']