

THE REPORT ON BINARY CLASSIFICATION

A Decision Tree classifier was the first thing we used, and it originally exhibited a strong bias toward the "Depression" classification. After implementing class balancing strategies, we employed logistic regression to rectify this imbalance. The recall for "mental-health" stayed low (0.29) despite the enhanced class representation, suggesting that it may be confused with other subtypes.

We conducted binary classification tasks comparing "mental-health" with each distinct subgroup in order to do additional research:

- Depression vs Mental Health
- Anxiety vs Mental Health
- Autism vs Mental Health
- BPD vs Mental Health
- Bipolar vs Mental Health
- Schizophrenia vs Mental Health

The classifier's ability to differentiate between "mental-health" and each category was reflected in the performance score generated by each model.

DECISION TREE

```
02]: from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report, confusion_matrix

dt_model = DecisionTreeClassifier(max_depth=10, random_state=42)
dt_model.fit(X_train, y_train)

y_pred_dt = dt_model.predict(X_test)
print(classification_report(y_test, y_pred_dt))
```

	precision	recall	f1-score	support
Anxiety	0.74	0.68	0.71	17246
BPD	0.89	0.38	0.53	7638
autism	0.89	0.43	0.58	1428
bipolar	0.79	0.40	0.53	8295
depression	0.68	0.95	0.79	51679
mentalhealth	0.31	0.01	0.02	7874
schizophrenia	0.80	0.20	0.32	3498
accuracy			0.70	97658
macro avg	0.73	0.43	0.50	97658
weighted avg	0.69	0.70	0.65	97658

LOGISTIC REGRESSION

```
103]: lr_model = LogisticRegression(max_iter=1000, class_weight = 'balanced', random_state=42)
lr_model.fit(X_train, y_train)

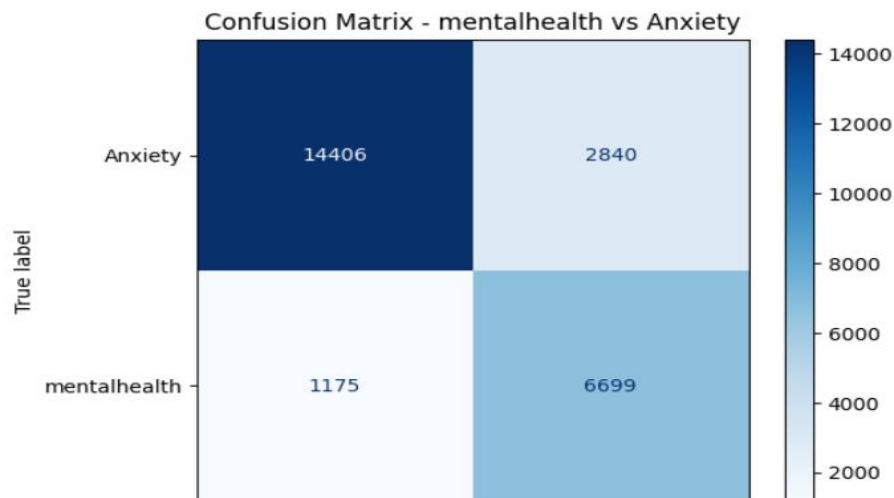
y_pred_lr = lr_model.predict(X_test)
print(classification_report(y_test, y_pred_lr))
```

	precision	recall	f1-score	support
Anxiety	0.77	0.76	0.76	17246
BPD	0.55	0.71	0.62	7638
autism	0.46	0.84	0.60	1428
bipolar	0.64	0.67	0.65	8295
depression	0.89	0.70	0.79	51679
mentalhealth	0.29	0.46	0.36	7874
schizophrenia	0.37	0.71	0.49	3498
accuracy			0.69	97658
macro avg	0.57	0.69	0.61	97658
weighted avg	0.75	0.69	0.71	97658

BINARY CLASSIFICATION

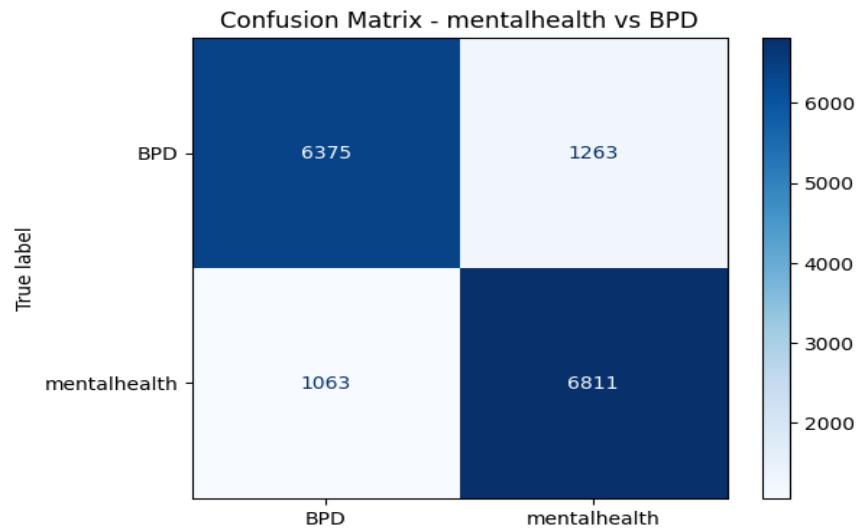
- MENTAL HEALTH VS ANXIETY

	precision	recall	f1-score	support
Anxiety	0.92	0.84	0.88	17246
mentalhealth	0.70	0.85	0.77	7874
accuracy			0.84	25120
macro avg	0.81	0.84	0.82	25120
weighted avg	0.85	0.84	0.84	25120



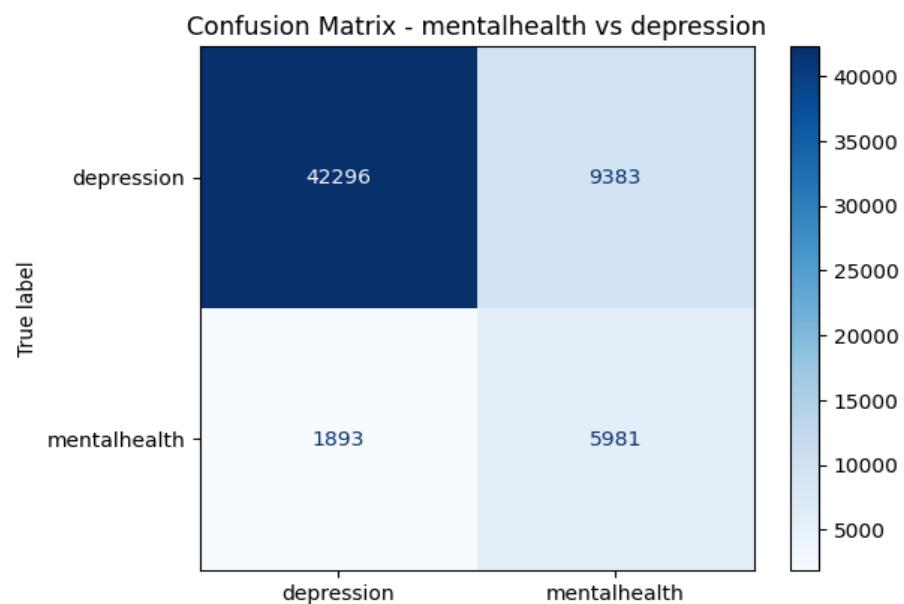
- MENTAL HEALTH VS BPD

	precision	recall	f1-score	support
BPD	0.86	0.83	0.85	7638
mentalhealth	0.84	0.86	0.85	7874
accuracy			0.85	15512
macro avg	0.85	0.85	0.85	15512
weighted avg	0.85	0.85	0.85	15512



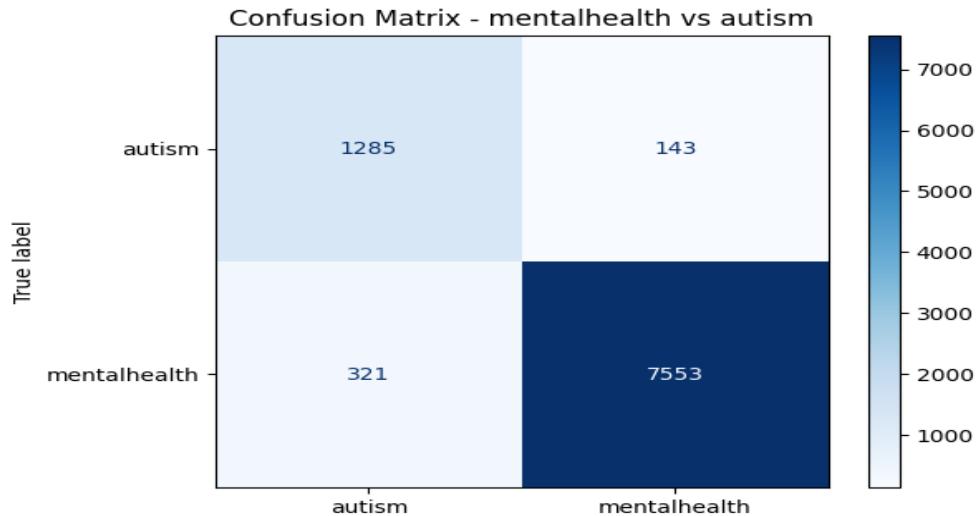
- MENTAL HEALTH VS DEPRESSION

	precision	recall	f1-score	support
depression	0.96	0.82	0.88	51679
mentalhealth	0.39	0.76	0.51	7874
accuracy			0.81	59553
macro avg	0.67	0.79	0.70	59553
weighted avg	0.88	0.81	0.83	59553



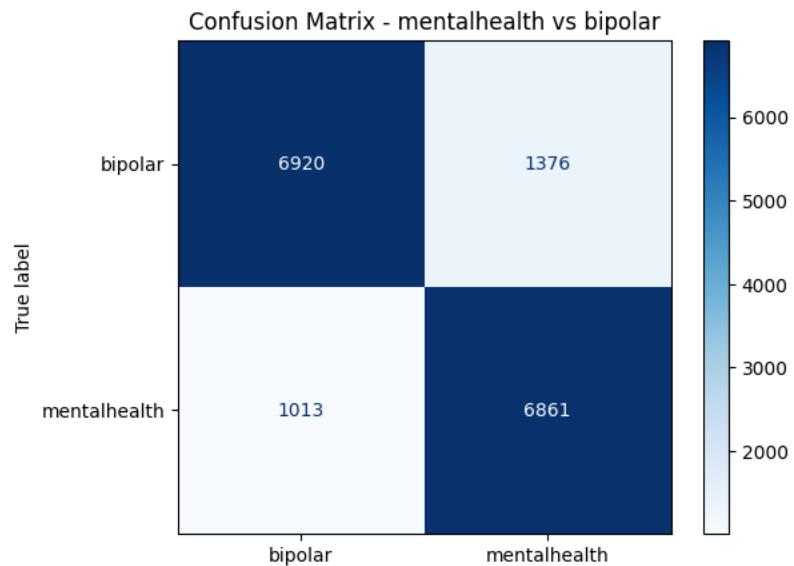
- MENTAL HEALTH VS AUTISM

	precision	recall	f1-score	support
autism	0.80	0.90	0.85	1428
mentalhealth	0.98	0.96	0.97	7874
accuracy			0.95	9302
macro avg	0.89	0.93	0.91	9302
weighted avg	0.95	0.95	0.95	9302



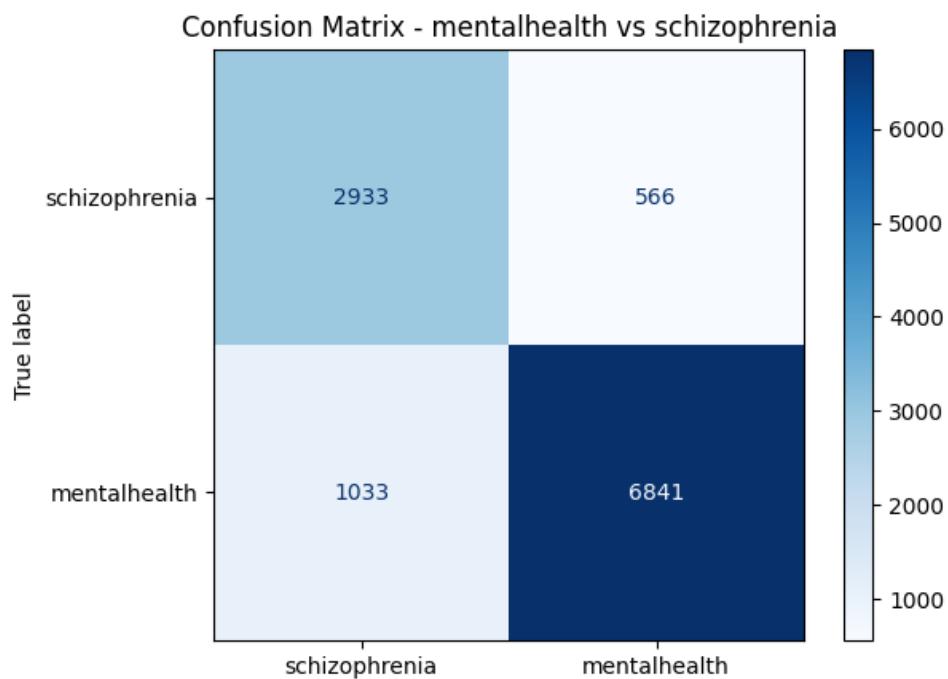
- MENTAL HEALTH VS BIPOLAR

	precision	recall	f1-score	support
bipolar	0.87	0.83	0.85	8296
mentalhealth	0.83	0.87	0.85	7874
accuracy			0.85	16170
macro avg	0.85	0.85	0.85	16170
weighted avg	0.85	0.85	0.85	16170



- MENTAL HEALTH VS SCHIZOPHRENIA

	precision	recall	f1-score	support
schizophrenia	0.74	0.84	0.79	3499
mentalhealth	0.92	0.87	0.90	7874
accuracy			0.86	11373
macro avg	0.83	0.85	0.84	11373
weighted avg	0.87	0.86	0.86	11373



FINAL INTERPRETATION OF THE VALUES:

SUBTYPE	MENTAL HEALTH SCORE	SUBTYPE SCORE
AUTISM	0.97	0.85
SCHIZOPHRENIA	0.90	0.79
BPD	0.85	0.85
BIPOLAR	0.85	0.85
ANXIETY	0.77	0.88
DEPRESSION	0.51	0.88

INTERPRETATION & INSIGHTS:

- The summary table uses F1 scores and qualitative observations to show how the "mental-health" class compares to each distinct subtype:
- With the highest F1 score for "mental-health" (0.97), autism clearly distinguishes itself from the other two divisions. The model was able to distinguish between them with ease.
- With a mental-health F1 of 0.90, schizophrenia likewise shows good separation, indicating that it is distinct from the general mental-health classification.
- The F1 values for bipolar disorder and BPD are the same (0.85 for both classes), suggesting a very balanced classification—the model identified either class with similar ease or difficulty.
- Anxiety's F1 score is 0.88, while mental-health's is 0.77, indicating a significant degree of confusion. This implies that the two share some characteristics.
- With the lowest F1 score for mental-health (0.51), depression comes out as having significant feature overlap and high levels of misunderstanding with the more general mental-health category. This is consistent with previous findings of bias toward depression shown in the Decision Tree model.
- Overall, the observations and F1 scores show that while some subtypes—such as schizophrenia and autism—are easily distinguished from normal mental health, others—particularly anxiety and depression—have more overlapping qualities, making classification more difficult.

DOUBTS:

1. SHOULD THE MENTAL HEALTH CLASS BE MERGED WITH DEPRESSION AND ANXIETY, AS THEY HAVE MORE SIMILARITIES AND THEN FURTHER CLASSIFIED?
2. IF WE DO SO WE ARE LOSING MAJOR PART OF THE DATASET.
3. IS IT A GOOD IDEA TO ELIMINATE “MENTAL HEALTH” CLASS?