

Problem Statement

Mental health issues such as anxiety, depression, loneliness, and suicidal thoughts are increasingly discussed in online communities. However, due to the vast volume of user-generated content, it is challenging to automatically identify the underlying causes or warning signs behind these mental health struggles.

The goal of this project is to analyze Reddit posts from key mental health subreddits (r/anxiety, r/depression, r/mentalhealth, r/suicidewatch, and r/lonely) to detect potential causes or risk factors contributing to mental health issues, such as drug and alcohol use, trauma and stress, early life experiences, and personality traits.

By developing a text classification model trained on this dataset, the project aims to:

- Automatically categorize posts based on underlying mental health causes.
- Support early identification of risk factors through language patterns.
- Contribute to improved understanding and digital mental health analysis.

Models Trained:

- BERT

Category	BERT Accuracy
Drug and Alcohol	0.87
Early Life	0.80
Personality	0.72
Trauma and Stress	0.60
Overall Accuracy	0.75

BERT was trained for 5 epochs before showing signs of overfitting. When trained on 10 epochs it showed significant overfitting with a training loss of ~0.01 while the testing loss increased to 1.4 which was higher than what we began with in epoch 1.

BERT Classifier

Metric	Drug and Alcohol	Early Life	Personality	Trauma and Stress	Macro Avg	Weighted Avg
Precision	0.97	0.82	0.58	0.67	0.76	0.75
Recall	0.87	0.80	0.72	0.60	0.75	0.75
F1-Score	0.92	0.81	0.64	0.63	0.75	0.76
Support	45	40	40	40	165	165
Accuracy						0.75

The BERT classifier achieved an overall accuracy of 75%, with particularly strong performance on the Drug and Alcohol and Early Life categories. Slightly lower F1-scores were observed for Personality and Trauma and Stress, likely due to overlapping linguistic cues between these classes.

Conclusion

In this study, we developed and evaluated BERT to classify Reddit posts according to potential causes of mental health issues. The model achieved an overall accuracy of 75% and demonstrated consistent per-class performance, particularly in the Drug and Alcohol and Early Life categories.

The performance of BERT can be attributed to its bidirectional language modeling, which captures contextual information effectively, enabling better understanding of nuanced language in mental health discussions. These results highlight the potential of transformer-based models to accurately identify underlying causes of mental health issues from textual data, providing a useful tool for research and digital mental health analysis.