

Sentiment Analysis for Mental Health Monitoring

Jasser Abdelfattah Mohammed
SRN: 21033101

Abstract

Mental health is a critical concern, especially among students. This project aims to develop a binary classification model for sentiment analysis, evaluating mental health status based on textual data. The model classifies entries as either "**normal**" or "**abnormal**". This paper discusses the dataset, preprocessing steps, model development, and results.

1 Problem Statement

Undiagnosed mental health issues can lead to severe consequences, particularly for students. The objective of this project is to create a sentiment analysis model to evaluate mental health statuses using textual data like tweets or posts. The model categorizes entries into:

- **Normal**
- **Abnormal** (e.g., depression, anxiety, stress)

This system can enable schools to identify students requiring mental health support proactively.

2 Dataset Description

2.1 Data Overview

The dataset contains textual statements related to mental health, labeled with one of seven categories:

- Normal
- Depression
- Suicidal
- Anxiety
- Stress
- Bipolar Disorder

- Personality Disorder

2.2 Data Sources

The data originates from social media platforms such as Twitter and Reddit. Preprocessing ensures its relevance and accuracy. This dataset is suited for:

- Training intelligent mental health chatbots.
- Conducting sentiment analysis.
- Exploring mental health trends.

2.3 Key Features

- **unique_id**: Unique identifier for each entry.
- **Statement**: The textual content of the post.
- **Mental Health Status**: Annotated mental health status of the statement.

3 Approach Explanation

3.1 Data Preprocessing

- **Cleaning**: Removed noise, special characters, and irrelevant data.
- **Tokenization**: Split text into smaller tokens for analysis.

3.2 Exploratory Data Analysis (EDA)

- Analyzed class distribution for balanced representation.
- Created word clouds to visualize common terms.

3.3 Feature Engineering

Mapped mental health statuses into two classes:

- **Normal**
- **Abnormal**

Text features were extracted using:

- **Bag-of-Words (BoW)**
- **TF-IDF (Term Frequency-Inverse Document Frequency)**

3.4 Model Building

Implemented and trained two models:

- **Naive Bayes**

- **Logistic Regression**

3.5 Model Evaluation

Evaluated using:

- **Metrics:** Precision, recall, F1-score, and accuracy.
- Performance was assessed on a held-out test dataset.

4 Results and Discussion

4.1 Naive Bayes Model Results

- **Confusion Matrix:** See fig. 1.

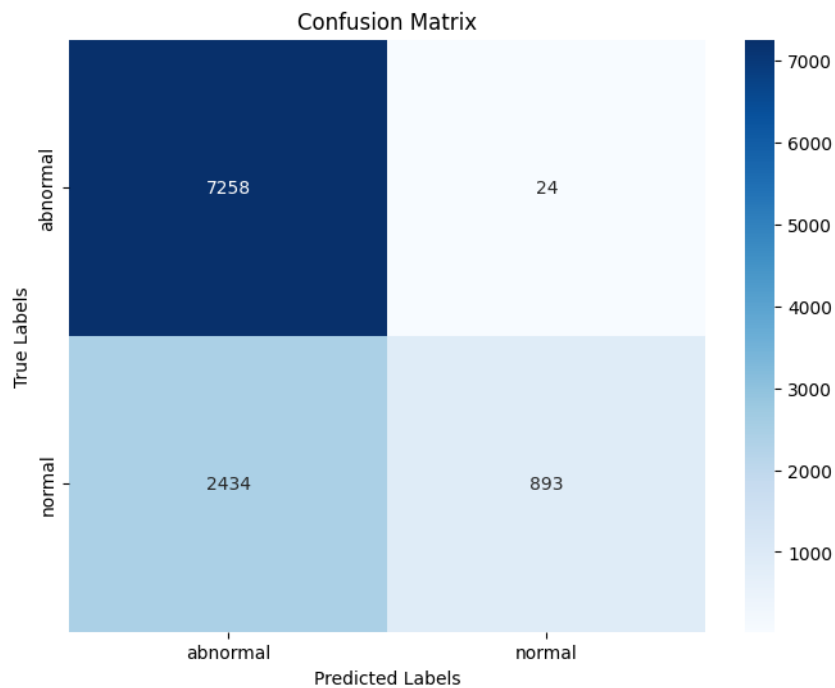


Figure 1: Naive Bayes Confusion Matrix

Accuracy: 77%

4.2 Logistic Regression Model Results

- **Confusion Matrix:** See fig. 2.

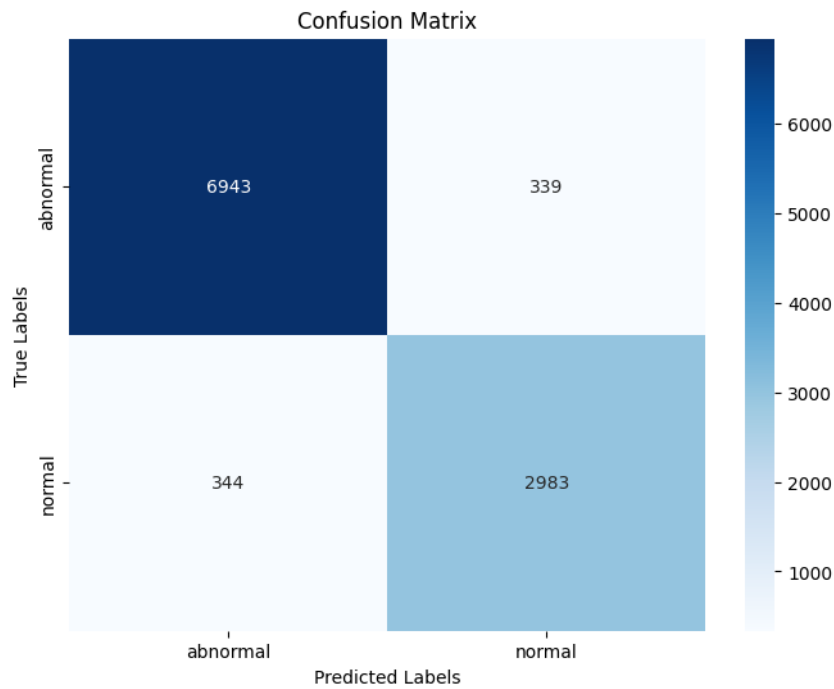


Figure 2: Logistic Regression Confusion Matrix

Accuracy: 94%

5 Performance on Unseen Data

5.1 Naive Bayes Model Predictions

```

Post: I feel like my heart is racing, and I can't catch my breath.
Prediction: abnormal

Post: One moment I'm laughing uncontrollably, and the next I'm crying for no reason.
Prediction: abnormal

Post: Nothing excites me anymore, and even getting out of bed feels like a chore.
Prediction: abnormal

Post: I had a great day at work and treated myself to some ice cream afterward!
Prediction: normal

Post: Sometimes I feel like I'm living someone else's life, and it's confusing.
Prediction: abnormal

Post: The constant deadlines are crushing me; I can't keep up.
Prediction: abnormal

Post: No one would even notice if I disappeared.
Prediction: abnormal

```

Figure 3: Naive Bayes Model Predictions on Unseen Data

5.2 Logistic Regression Model Predictions

```
Post: I feel like my heart is racing, and I can't catch my breath.
Prediction: normal

Post: One moment I'm laughing uncontrollably, and the next I'm crying for no reason.
Prediction: normal

Post: Nothing excites me anymore, and even getting out of bed feels like a chore.
Prediction: abnormal

Post: I had a great day at work and treated myself to some ice cream afterward!
Prediction: normal

Post: Sometimes I feel like I'm living someone else's life, and it's confusing.
Prediction: abnormal

Post: The constant deadlines are crushing me; I can't keep up.
Prediction: normal

Post: No one would even notice if I disappeared.
Prediction: normal
```

Figure 4: Logistic Regression Model Predictions on Unseen Data

5.3 Final Result Predictions Table

Post	Label	Naive_Bayes_Prediction	Logistic_Regression_Prediction
I feel like my heart is racing, and I can't catch my breath.	anxiety	abnormal	normal
One moment I'm laughing uncontrollably, and the next I'm crying for no reason.	bipolar	abnormal	normal
Nothing excites me anymore, and even getting out of bed feels like a chore.	depression	abnormal	abnormal
I had a great day at work and treated myself to some ice cream afterward!	normal	normal	normal
Sometimes I feel like I'm living someone else's life, and it's confusing.	personality disorder	abnormal	abnormal
The constant deadlines are crushing me; I can't keep up.	stress	abnormal	normal
No one would even notice if I disappeared.	suicidal	abnormal	normal

Figure 5: Final Result Predictions Table

6 Conclusion and Next Steps

The Logistic Regression model outperformed Naive Bayes with higher accuracy but exhibited lower performance on unseen data. Both models demonstrated unique strengths:

- **Naive Bayes:** Better performance on unseen data.
- **Logistic Regression:** Higher accuracy on the test set.

6.1 Future Work

1. Incorporate Transformer Models (e.g., BERT, RoBERTa).
2. Explore hybrid model combinations.