

# Mental Health Sentiment Analysis Using NLP

Group Members: Divyamol Madathilparambil Sunikumar  
Sruthy Ajitha Kumari Somasekharan Nair

## 1. Introduction

Mental health challenges are increasingly being expressed through digital text on social media and messaging platforms. This project explores the use of Natural Language Processing (NLP) and machine learning to analyze the sentiment of such messages and classify them as Positive, Neutral, or Negative. The objective is to provide an automated tool that can support early mental health assessment through sentiment recognition.

## 2. Methodology

## 2.1 Data Collection

The dataset used contains text messages labeled with sentiments relevant to mental health. Each message is categorized as Positive, Neutral, or Negative based on its emotional tone.

Dataset: GoEmotions by Google Research (58k+ comments with 27 emotion labels).

## 2.2 Preprocessing

Text preprocessing involves several key steps:

- Lowercasing
- Removing URLs, hashtags, and punctuation
- Removing numbers
- Tokenization
- Stopword removal using NLTK
- Lemmatization using WordNetLemmatizer

## 2.3 Feature Extraction

TF-IDF (Term Frequency–Inverse Document Frequency) was used to convert the cleaned text into numerical feature vectors. This method captures the importance of each word relative to the document and the overall corpus.

## 2.4 Model Training

A Logistic Regression model was trained on the TF-IDF feature vectors. The model was chosen for its efficiency and effectiveness in text classification tasks. After training, the model, TF-IDF vectorizer, and label encoder were saved using pickle.

## 3. Streamlit Web Application

A Streamlit web application was developed to make the sentiment classifier accessible. The application allows users to input a message and returns the predicted sentiment. It uses the trained model and preprocessing pipeline to ensure consistency in predictions.

## 4. Results

The model showed promising results in classifying short text messages. Examples include:

- "I feel better today" → Positive
- "I don't know what to do anymore" → Negative
- "I went for a walk" → Neutral

The app provides a fast and easy interface for real-time sentiment analysis.

## 5. Challenges and Future Work

Some limitations faced in this project include:

- Difficulty handling sarcasm and irony
- Inability to process emojis and slang effectively
- Limited training data for complex emotional expressions

Future work could involve:

- Using BERT or transformer-based models for better contextual understanding
- Expanding the dataset with real-world social media data
- Including emoji and slang interpretation mechanisms

## 6. Model Evaluation and Comparison

The model was evaluated using Accuracy, Precision, Recall, and F1-Score. A comparison of different feature extraction techniques is shown below. Among them, both Bag-of-Words and TF-IDF achieved similar performance, while Word2Vec performed slightly worse.

Feature Type	Accuracy	Precision	Recall	F1-Score
Bag-of-Words	0.63	0.64	0.63	0.63
TF-IDF	0.63	0.64	0.63	0.63
Word2Vec (pretrained)	0.58	0.59	0.58	0.58

## 7. Conclusion – Best Embedding Technique

After evaluating multiple feature extraction techniques, TF-IDF was selected as the best-performing embedding method. It demonstrated the highest F1-score, especially for the positive sentiment class, and showed consistent performance across all classes.

- TF-IDF strikes a strong balance across all sentiment classes, achieving the highest F1-score for the positive class (0.70).
- Word2Vec, although semantically rich, underperformed due to the simple averaging strategy used. It could benefit from deeper architectures like LSTMs or BERT.

Final Decision: TF-IDF combined with Logistic Regression is selected for deployment. It offers a good trade-off between performance and computational efficiency, making it suitable for real-time sentiment analysis.