Mining Emotional Distress and Mental Health Risk Signals from Reddit Using NLP

Salman Shareef

Master of Science in Applied Data Science

University of St. Thomas

Practicum I – Dr. Follis

December 2025

Abstract

Social media platforms provide a rich source of unstructured data for understanding population-level mental health trends. This study applies large-scale Natural Language Processing (NLP) techniques to over 500,000 Reddit posts related to mental health disorders. The objectives are to (1) identify emotional sentiment trends, (2) uncover latent psychological distress themes using topic modeling, and (3) detect high-risk mental health signals using heuristic risk detection. The results highlight substantial emotional heterogeneity across disorders and reveal critical high-risk content that could inform digital public health surveillance and intervention strategies.

Introduction

Mental health disorders represent a growing public health challenge globally. Traditional surveillance systems rely on clinical diagnoses and self-reported surveys, which often suffer from reporting delays and undercoverage. Social media platforms such as Reddit allow individuals to openly discuss mental health struggles in near real time, making them valuable supplementary data sources for public health analytics.

This project applies NLP-based methods to mental health–related Reddit posts to quantify emotional sentiment, extract dominant psychological themes, and identify high-risk expressions indicative of severe distress or suicidal ideation. The analysis demonstrates how computational text analytics can augment population-level mental health monitoring.

Data Description

The dataset used in this project is the **Mental Disorders Reddit Dataset**, containing:

- **title:** Reddit post title
- **selftext:** Main post content
- **created_utc:** Timestamp
- **over_18:** NSFW indicator
- **subreddit:** Mental health category label

The raw dataset exceeds **214 MB** and contains posts from multiple mental health–focused subreddits including depression, anxiety, bipolar disorder, PTSD, ADHD, and others.

Because of its size, all processing was performed using **chunk-based streaming pipelines** rather than full in-memory loading.

Data Preprocessing

Key preprocessing steps included:

- Removal of deleted and empty posts
- Merging title and body into a unified full_text field
- Lowercasing, punctuation removal, number removal
- Token-length filtering (minimum 15 words)
- Generation of clean_text, word_count, and topic labels

This produced a high-quality cleaned dataset exceeding **500,000 posts**, saved locally for downstream modeling.

Sentiment Analysis

Sentiment was computed using the **VADER (Valence Aware Dictionary for Sentiment Reasoning)** model, well-suited for short informal social media text.

Each post received:

- A continuous **sentiment_score ∈ [–1, +1]**
- A categorical sentiment label:
    - Positive
    - Neutral
    - Negative

Due to memory limitations, the sentiment pipeline was implemented using a **disk-streamed chunk processing system**, writing results directly to disk.

Topic Modeling

Latent Dirichlet Allocation (LDA) was applied to a random 50,000-post sample using TF-IDF features:

- **Vector size:** 8,000 tokens
- **Topics:** 8

Each topic revealed distinct psychological clusters such as:

- Anxiety & panic
- Medication use
- Suicidal ideation
- Therapy and diagnosis
- Social isolation
- Relationship distress

Each post was assigned a dominant topic label for theme-level analysis.

High-Risk Mental Health Signal Detection

A high-risk detection system was built using:

A. Keyword-Based Detection

Flagged phrases including:

- “suicidal”
- “kill myself”
- “end my life”
- “self harm”
- “can’t go on”

B. Sentiment-Based Detection

Strongly negative posts were flagged using:

sentiment_score ≤ –0.6

The final **high-risk flag** was computed using:

high_risk_flag = keyword_flag OR negative_sentiment

Results

Key findings include:

- A measurable proportion of Reddit mental health posts contain **high-risk emotional content**
- Depression and suicide-related subreddits exhibit the **highest distress intensity**
- Certain latent LDA topics show consistently elevated high-risk rates
- Temporal trend analysis reveals fluctuations in distress intensity over time

A summary risk table was saved as:

summary_high_risk_by_topic.csv

Public Health Implications

This study demonstrates that NLP models can support:

- Early-warning mental health surveillance
- Community-level emotional distress monitoring
- Digital triage tools to prioritize high-risk content
- Evidence for online intervention strategies

While this analysis cannot substitute clinical diagnosis, it illustrates how large-scale text analytics can assist public health monitoring in near real time.

Ethical Considerations

All data were publicly available and anonymized. No individual-level intervention or profiling was performed. Risk detection is heuristic and probabilistic, not diagnostic. Future applications must follow strict ethical guidelines for digital mental health monitoring.

Conclusion

This project successfully implemented a **full-scale end-to-end NLP pipeline** including:

- Big data ingestion
- Cleaning & memory-safe processing
- Sentiment modeling
- Topic modeling
- Risk signal detection

The framework demonstrates the practical application of NLP for **mental health surveillance and policy analytics**, fulfilling the technical and research expectations of a Master's-level practicum.