

Automated Mental Health Concern Analysis Using Natural Language Processing

Project Abstract

This project presents an advanced Natural Language Processing (NLP) system designed to automatically extract, classify, and analyse mental health concerns from user-provided text input. The system employs a comprehensive pipeline architecture that integrates five key components: sentiment analysis, keyword extraction, concern classification, intensity scoring, and temporal analysis.

System Architecture

A[User Input] --> B[Polarity Detection]
A --> C[Keyword Extraction]
C --> D[Concern Classification]
D --> E[Intensity Scoring]
B & D & E --> F[Timeline Analysis]
F --> G[Mental Health Insights]

Key Components

- Polarity Detection**
 - RoBERTa-based sentiment analysis model
 - Three-way classification: positive, negative, neutral
 - Confidence scoring for reliability assessment
- Keyword Extraction (NER)**
 - Hybrid approach combining PhraseMatcher and pattern matching
 - Context-aware extraction of mental health expressions
 - Training on domain-specific datasets
- Concern Classification**
 - TF-IDF vectorization with Random Forest classifier
 - Mapping to predefined mental health categories
 - Multi-label classification capabilities
- Intensity Scoring**
 - 1-10 severity scale based on linguistic markers
 - Context-aware intensity adjustment
 - Consideration of temporal patterns
- Timeline Analysis**
 - Progression tracking over time
 - Detection of significant mental state shifts
 - Trend analysis and early warning system

Technical Stack

- Python 3.8+
- Transformers (RoBERTa)
- SpaCy for NER
- Scikit-learn for classification
- Pandas for data processing