Project Proposal

Multimodal Emotion Detection for Early Mental Health Insights

Jessey Morales Trejo & Bryan Duran

A. Project Description

Rising rates of anxiety and depression among young adults highlight the need for intelligent systems that can detect emotional distress early. Social media often reveals subtle signs of mental strain that go unnoticed until conditions worsen. This project uses transformer-based and multimodal fusion models to analyze existing Instagram and Reddit data, aiming to detect multiple co-occurring emotions through a regression approach. We will focus on interpretability, recall on minority emotions, and visualization of emotional cues, while limiting our scope to modeling, evaluation, and ethical analysis without collecting live data or performing medical diagnoses.

B. Dataset

Instagram comments on posts about celebrity mental health comments are labelled for sentiment (positive/neutral/negative) and stigma:
https://zenodo.org/records/11202766?utm_source

Investigating COVID-19's Impact on Mental Health: Trend and Thematic Analysis of Reddit Users' Discourse:
Article: https://www.jmir.org/2023/1/e46867/
Dataset here: https://github.com/Sallyzhu/RedditImpact

C. Methods

We will use a Transformer-based text encoder (e.g., BERT or RoBERTa) to capture contextual semantics and emotional cues from Reddit and Instagram posts. If additional metadata (e.g., post timing or emoji use) is available, we plan to integrate it through a dense-layer fusion module or an attention-based late-fusion approach for final emotion prediction.

Transformer models excel at understanding nuanced language in social media text, capturing subtle emotional context and sentiment shifts. Late-fusion mechanisms allow flexible integration of multiple feature types, supporting more accurate emotion and mental health state detection.

D. Your Contribution

This project introduces a multimodal fusion approach that integrates text, audio, and metadata to capture richer emotional context in social-media communication. We will apply fine-tuning and architectural adjustments, including class-weighted losses and attention tweaks, to improve recall for minority emotions. To ensure transparency and ethical alignment, we will employ explainability techniques such as attention visualization and SHAP to highlight the model's key linguistic and behavioral influences.

Deliverables include trained Transformer-based and multimodal models, reproducible code, and visualizations of emotion attribution. Success criteria are improved F1-score and recall over baseline Transformer models, plus interpretable outputs that reveal emotional patterns and provide actionable insights for early mental-health detection and research.