

Mental Health Detection from Social Media Posts

Group members

Aung Thu Phyoo

Su Mon Hlaing

Phyo Ko Ko

Moe Thet Hmue

Nang Mon Kham

Capstone Project – Abstract, Literature Review, Data, and Technology Review

1. Abstract

The goal is to develop a mental health detection model that can be integrated into social media platforms or external tools to help promote awareness, enable early intervention, and support mental well-being initiatives.

2. Literature Review

In recent years, the application of natural language processing (NLP) and machine learning techniques to detect mental health-related signals from social media and other textual sources has grown rapidly. For example, Natural language processing applied to mental illness detection a narrative review **[1]** provide a comprehensive review of 399 studies spanning a decade, and identify key trends: in particular a shift from traditional feature-based classifiers toward deep-learning architectures, and an increasing focus on interpretability and varied data sources (clinical notes, interviews, social-media posts).

Complementing this broader survey, the work titled A deep learning model for detecting mental illness from user content on social media **[2]** offers an empirical demonstration using data collected from online communities (e.g., Reddit “subreddits” focused on specific disorders) the authors constructed several binary classification models for disorders such as depression, anxiety, bipolar disorder, borderline personality disorder, schizophrenia and autism. Their deep-learning approach achieved high accuracy in identifying whether a user’s posts indicated a specific mental-disorder forum membership.

More specifically focusing on suicidal risk, the review Suicidal ideation detection on social media using machine learning: A review **[3]** compiles evidence on machine-learning approaches to detect suicidal ideation in online text, highlighting both algorithmic methods (e.g., multilayer neural networks, ensemble classifiers) and the ethical imperatives of deploying AI in mental-health contexts.

3. Data Description

This study utilizes the Sentiment Analysis for Mental Health dataset [4], which provides a comprehensive and structured collection of mental health related textual statements. The dataset was curated by aggregating multiple publicly available mental-health datasets and applying extensive cleaning procedures to ensure uniformity and usability for natural language processing tasks.

- **Data Type:** Text data stored in CSV format.
- **Language:** Primarily English for the initial model training phase.
 - Later, a translation algorithm (e.g., Google Translate API or MarianMT) will be applied to expand the dataset to Burmese and other regional languages.
- **Data Overview:**
The dataset consists of statements tagged with one of the following seven mental health statuses:
 - Normal
 - Depression
 - Suicidal
 - Anxiety
 - Stress
 - Bi-Polar
 - Personality Disorder

This approach ensures scalability and cross-language adaptability, making the system useful for Myanmar's multilingual context.

4. Technology Review

The project will experiment with both Machine Learning (ML) and Deep Learning (DL) techniques to determine which achieves better accuracy and contextual understanding.

4.1 Machine Learning Models

- **Preprocessing Steps:**
 - Remove emojis, punctuation, and URLs
 - Tokenize text and remove stop words

- Normalize case and clean symbols
- Apply translation for multilingual analysis
- **Support Vector Machine (SVM):** Effective for binary and multi-class text classification using TF-IDF features.
- **XGBoost:** An optimized gradient boosting algorithm that performs well with structured text features and handles imbalanced data efficiently.

4.2 Deep Learning Models

- **LSTM (Long Short-Term Memory):** A type of Recurrent Neural Network (RNN) that can capture long-term dependencies and sequential emotional patterns in text.
- **BERT (Bidirectional Encoder Representations from Transformers):** A state-of-the-art NLP model capable of understanding the contextual meaning of words in both directions. Fine-tuning BERT for mental health detection is expected to yield high accuracy and better emotional context recognition.

The performance of these models will be compared using evaluation metrics such as accuracy, precision, recall, and F1-score, and the best-performing model will be selected for final implementation.

5. Technology Stack

Component	Tools / Libraries
Programming Language	Python
ML Frameworks	scikit-learn, XGBoost
DL Frameworks	TensorFlow, PyTorch, Hugging Face Transformers
Text Processing	NLTK, spaCy
Data Handling	pandas, NumPy
Visualization	Matplotlib, Seaborn
Deployment (optional)	Docker and Flask or Streamlit for web app integration

References :

1. Zhang, T., Schoene, A.M., Ji, S. & Ananiadou, S. (2022) 'Natural language processing applied to mental illness detection: a narrative review', *npj Digital Medicine*, 5, p. 46. doi: 10.1038/s41746-022-00589-7.
<https://www.nature.com/articles/s41746-022-00589-7>
2. Kim, J., Lee, J., Park, E., Han, J. & Sungkyunkwan University (2020) 'A deep learning model for detecting mental illness from user content on social media', *Scientific Reports*, 10(1), p. 11846. doi: 10.1038/s41598-020-68764-y.
<https://www.nature.com/articles/s41598-020-68764-y>
3. Sheng, Z. (n.d.) 'Suicidal ideation detection on social media using machine learning: A review'. [online] Available at:
<https://arxiv.org/abs/2201.10515>
4. Sarkar, S. (n.d.) *Sentiment analysis for mental health* [dataset]. Kaggle. Available at:
<https://www.kaggle.com/datasets/suchintikasarkar/sentiment-analysis-for-mental-health/data>