
frontier
tech
leaders
A new generation
of tech specialists

Mental Health Detection from Social Media Posts

By Group 12

- Phyo Ko Ko
- Aung Thu Phyo
- Su Mon Hlaing
- Moe Thet Hmue
- Nang Mon Kham

Frontier tech leaders

Background

Context : Mental health is a critical component of global well-being. Platforms like Facebook and X are primary forums for personal expression where users share emotions, creating a massive data stream where distress signals are often lost.

Problem: Traditional manual support and reactive reporting tools cannot handle the volume of millions of daily posts. Subtle linguistic patterns of distress often go unnoticed until a crisis becomes apparent.

Solution Necessity: An automated NLP-driven approach is required to analyze text at scale to bridge the gap between expression of distress and professional support.

Frontier tech leaders

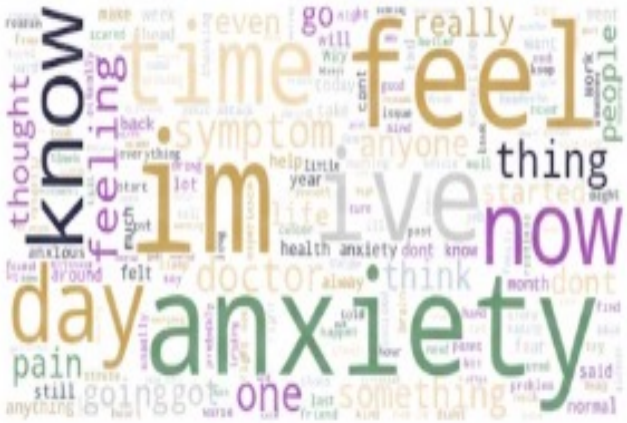
Objectives

❖ **Real Time Detection** : Automatically classify and detect mental helath conditions from social medial posts

❖ Leverages machine learning models to classify mental health conditions from real-time social media text data.

❖ **Contextual Analysis**

Utilize fine-tuned DistilBert Model to identify patterns related to mental health states and provide fast, reliable deployed cloud API



Frontier tech leaders

SDG Relation

By supporting early awareness of mental-health risk patterns, the system aims to assist organizations in promoting mental well-being and prevention, while avoiding harmful automation.

By leveraging modern NLP models and cloud-based APIs, the system demonstrates how AI can be applied to analyze large-scale social-media data. The project emphasizes scalable, modular design and supports innovation in digital monitoring tools.

frontier tech leaders

Data

Frontier tech leaders

Data

- Source : Primary dataset from Kaggle (Mental Health Sentiment Analysis)

  Characteristics:
  - Short, informal text
  - Emotionally expressive language
  - Multiple mental-health categories

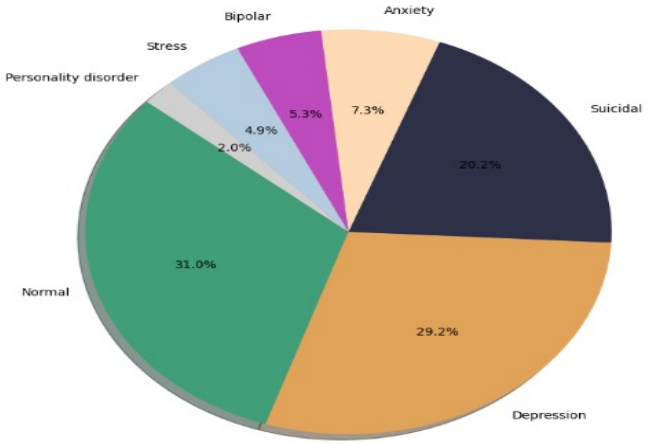
- Live social-media posts (scraped via API for deployment testing)

Data Source Link : https://www.kaggle.com/code/annastasy/mental-health-sentiment-analysis-nlp-ml/input

Frontier tech leaders

Data Preprocessing

- Text Cleaning and Normalization
- Removal of empty data / invalid Posts
- Translation of non-english to enlgish for modal compatibility
- Used Random Over Sampler for unbalanced data
- Preservation of original text for human review and transparency

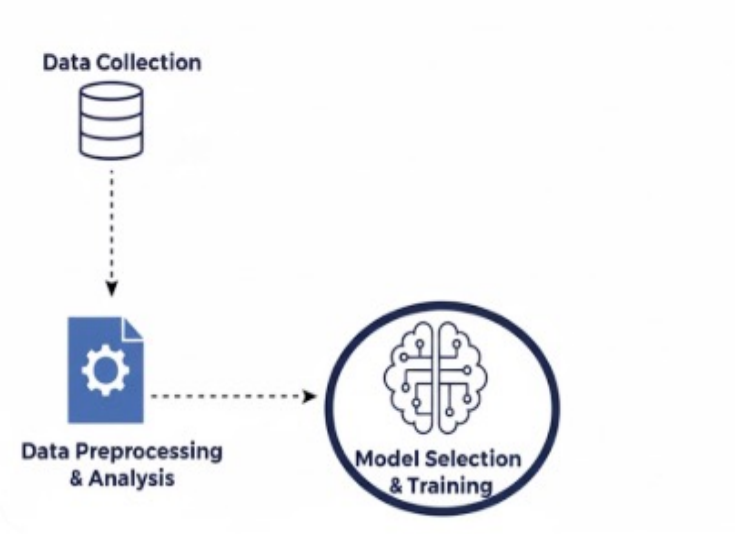


frontier tech leaders

Model

Frontier tech leaders

Model Selection



Baseline Approach

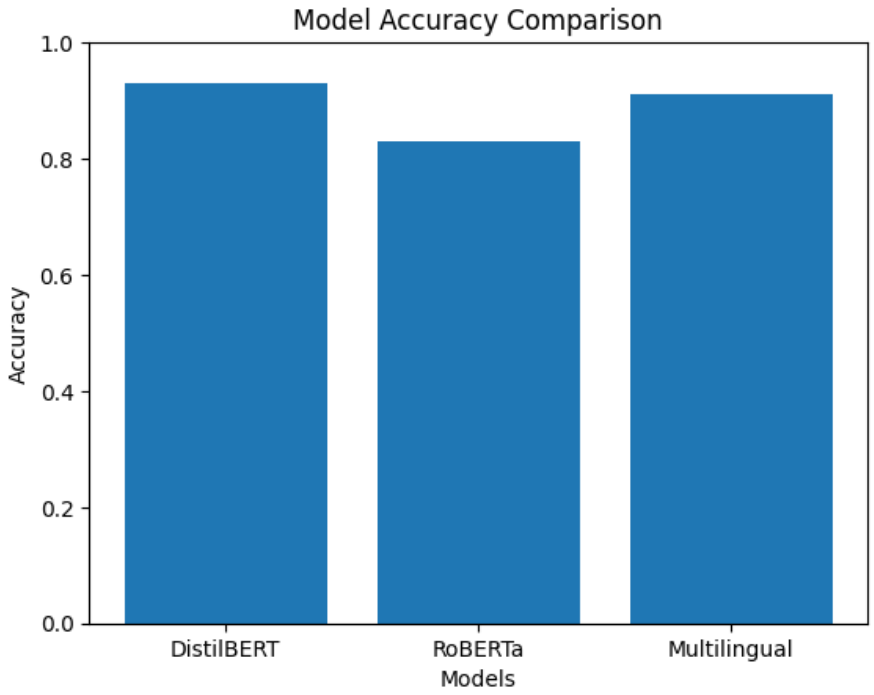
- Initially experimented with classic models (Logistic Regression, XGBoost) using TF-IDF features
- Limitation: Failed to capture context (e.g., distinguishing "not happy" from "happy") due to lack of sequence awareness

Selected Architecture

- DistilBERT (distilbert-base-uncased)
- Contextual Intelligence: Uses self-attention to understand the relationship between words across the entire sentence.

Frontier tech leaders

Models Comparison



❖ Efficiency: Chosen for its superior balance of high accuracy (~93%) and faster inference speed compared to larger models like RoBERTa.

Frontier tech leaders

Model Training and testing

- **Training Configuration:**
  - **Tokenization**: Switched to WordPiece to break words into meaningful sub-words, preserving semantic value better than stemming .
  - **Optimization**: Utilized Optuna to verify optimal learning rates 1×10^{-5} to 5×10^{-5}
  - **Execution:** Fine-tuned for 2 epochs with a batch size of 16 using the Hugging Face Trainer API.

- **Testing Protocol:**
  - **Data Split:** Evaluated on a 20% held-out test set strictly unseen during training to ensure valid performance metrics.
  - **Primary Metric:** Prioritized Weighted F1-Score to ensure a safe balance between Precision (accuracy of flags) and Recall (catching true cases).

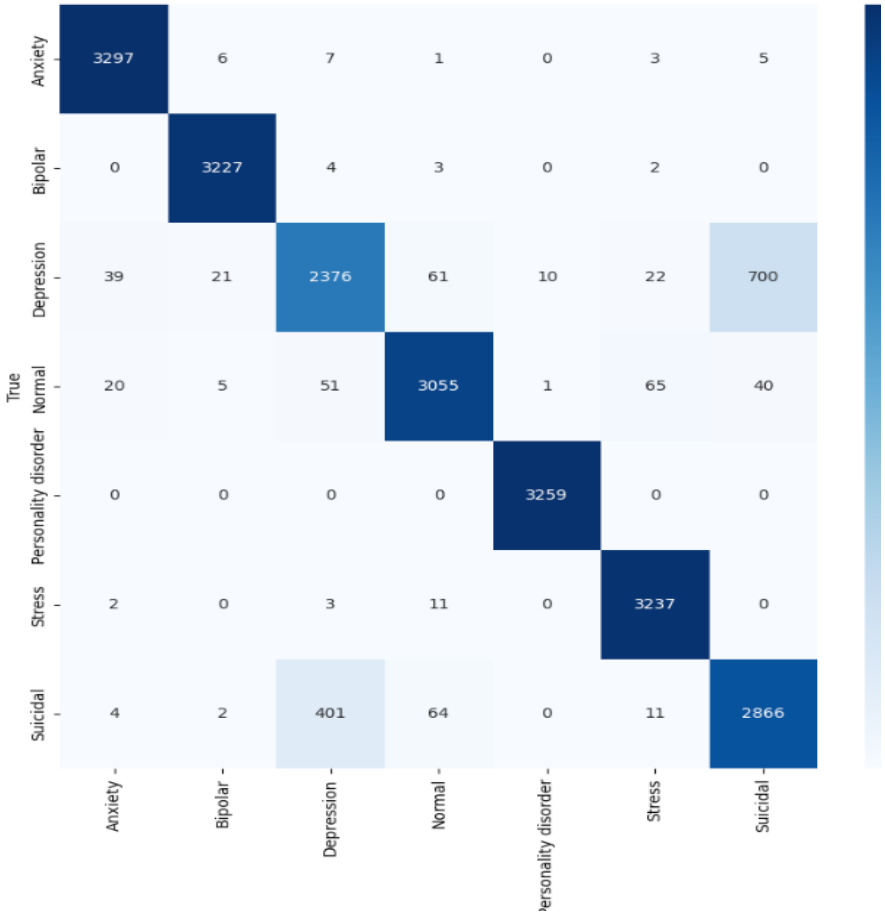
frontier tech leaders

Result

Frontier tech leaders

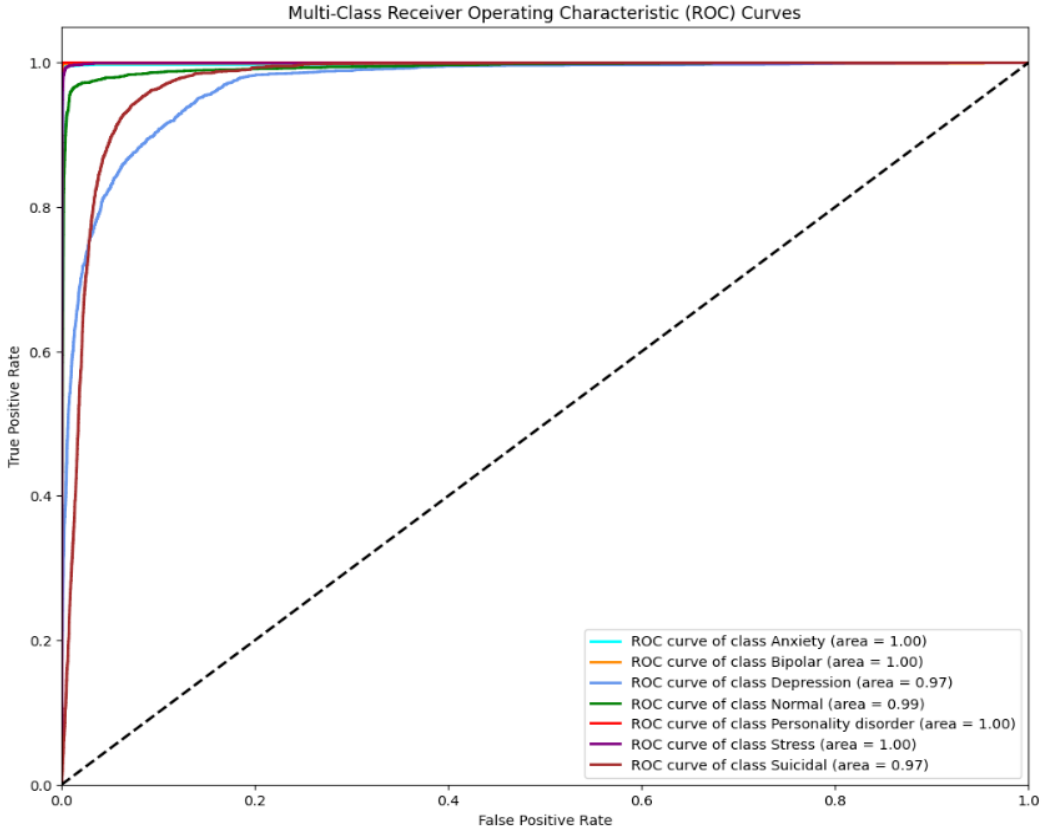
Evaluation Results

Category	Precision	Recall	F1 -Score
Anxiety	0.98	0.99	0.99
Bipolar	0.99	1.00	0.99
Depression	0.84	0.74	0.78
Normal	0.96	0.94	0.95
Personality Disorder	1.00	1.00	1.00
Stress	0.97	1.00	0.98
Sucidial	0.79	0.86	0.82
Test Accuracy		0.93	



Frontier tech leaders

Evaluation Results



frontier tech leaders

Deployment

Frontier tech leaders

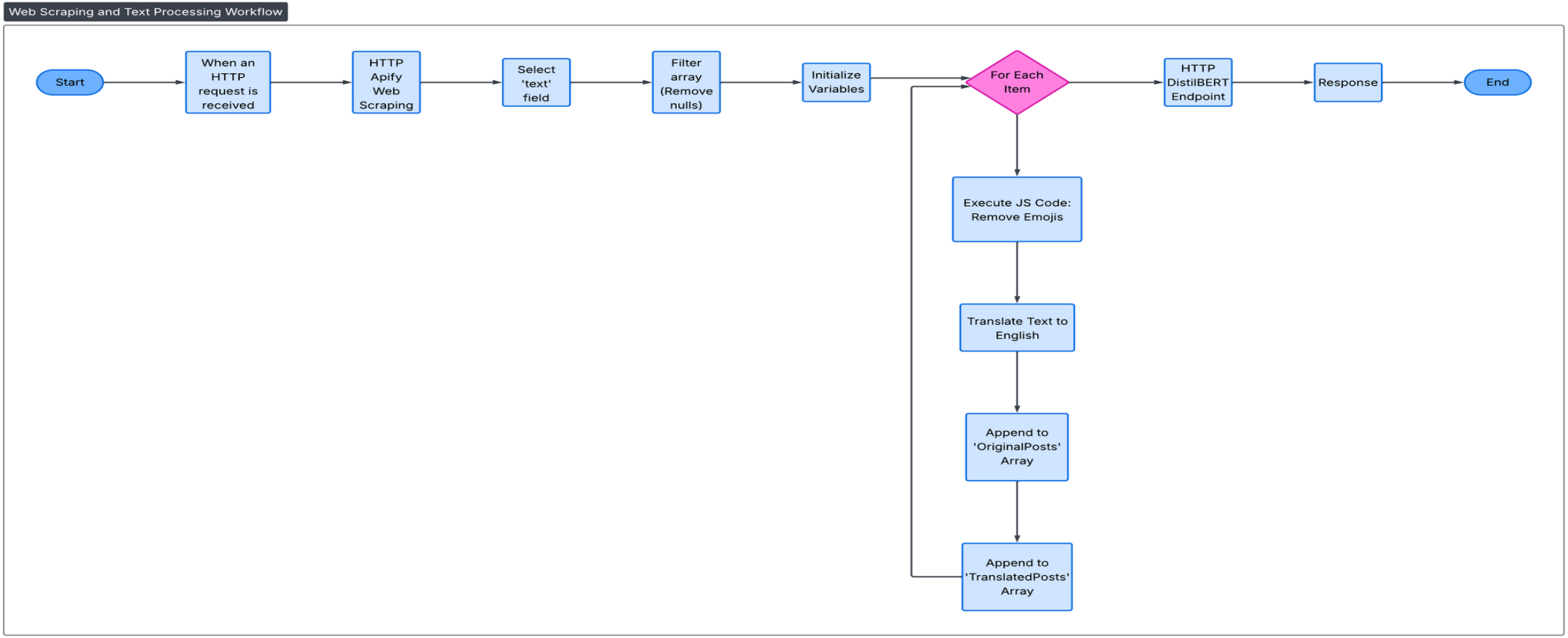
Deployment

- **Platform:** Deployed on Azure Machine Learning using a Managed Online Endpoint for real-time inference.

- **Workflow:**
  - **Input:** Logic App receives a post URL.
  - **Scoring:** score.py script loads the model and tokenizes text.
  - **Safety:** Inference runs on a secure Standard_DS2_v2 instance.
  - **Output:** JSON response with predicted status and probability.

frontier tech leaders

Deployment



Frontier tech leaders

Conclusion

- **Bilingual Resource:** Created a custom 2,000-sample dataset in English and Myanmar (Combined_Data_2000_en_my.csv), addressing the low-resource language gap.
- **Proven Performance:** Successfully implemented a Transformer-based model for effective cross-lingual text classification.
- **Social Impact:** Enables automated monitoring of social-media text to identify potentially harmful mental-health content, supporting SDG 3 (Good Health and Well-Being) through early risk detection and SDG 9 (Industry, Innovation, and Infrastructure) by applying AI-driven NLP systems to social platforms.

Frontier tech leaders

Futurework

- **Data Expansion:** Scale dataset to cover slang, dialects, and both Zawgyi/Unicode fonts.
- **Deployment:** Launch as a real-time Web API or mobile app for instant predictions.
- **Multimodal Model:** Integrate Speech-to-Text to analyze voice messages (popular in Myanmar).
- **Granularity:** Upgrade from simple sentiment (Positive/Negative) to specific emotion detection (Joy, Anger, Fear).

Thank you!

frontier tech leaders