ML TUTORIAL

Methodology for Project: Detecting Stress, Anxiety, and Depression from Voice Tone and Text Responses

Group members-

- Monika K. - 1NT22IS097
- Palak Dattatraya Kota - 1NT22IS110
- Prathyusha N. B. - 1NT22IS119
- Purabh Singh - 1NT22IS122

Methodology:

1. Data Collection

a. Dataset Used:

We utilized the DAIC-WOZ (Distress Analysis Interview Corpus - Wizard of Oz) dataset. This is a publicly available, widely accepted dataset for mental health detection tasks.

b. Source Type:

The dataset is from a secondary source, provided by the University of Southern Californias Institute for Creative Technologies.

c. Data Description:

- Audio recordings of human participants in virtual interviews.
- Text transcripts of interview responses.
- PHQ-8 depression scores, which serve as labels for classification.

d. Data Reliability:

- The dataset is expert-annotated, peer-reviewed, and has been used extensively in academic research.
- It offers multimodal real-world data, making it highly reliable for our study.

2. Data Analysis and Processing

a. Preprocessing Techniques:

- **Text:**
  - Lowercasing
  - Tokenization
  - Stopword removal
  - Lemmatization
  - BERT embeddings
- **Audio:**
  - MFCC extraction
  - Pitch and energy analysis
  - Spectrogram generation (Librosa)
- **Labels:**
  - PHQ-8 score thresholds converted to binary or multi-class labels

b. Tools and Frameworks:

- Programming: Python (Jupyter Notebooks)
- Text Processing: NLTK, SpaCy
- Audio Processing: Librosa
- Modelling: TensorFlow, Keras, HuggingFace Transformers, PyTorch
- Evaluation: Scikit-learn
- Visualization: Matplotlib, Seaborn

c. Handling Missing or Inconsistent Data:

- Remove audio-text mismatches and corrupted files.
- Discard samples with missing PHQ-8 scores.
- For minor feature gaps, apply KNN or mean imputation where appropriate.

3. Modelling and Experimentation

a. Problem Statement:

To classify individuals as likely experiencing stress, anxiety, or depression based on voice tone and textual responses using multimodal deep learning techniques.

b. Model Architecture:

Multimodal Deep Learning Pipeline:

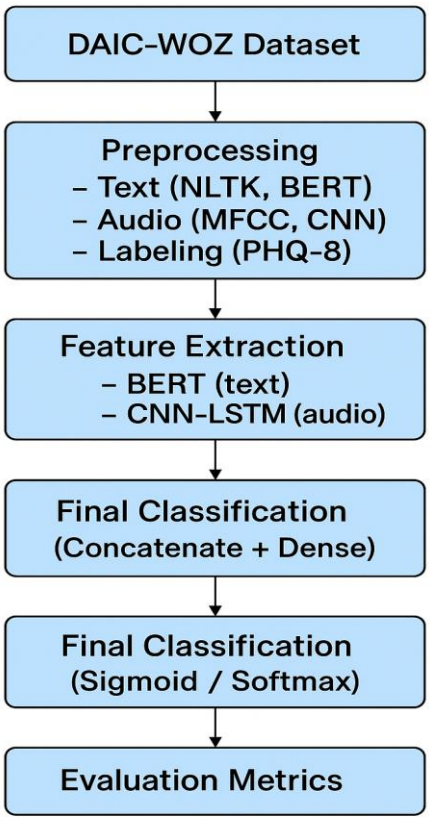
- Text: BERT or BiLSTM to extract semantic embeddings from tokenized text
- Audio: CNN-LSTM or wav2vec 2.0 to capture temporal and tonal features
- Fusion: Concatenation of text and audio features followed by dense neural layers
- Output: Final classification using Sigmoid (binary) or Softmax (multi-class)

c. Training Details:

- **Train/Val/Test:** 70% / 15% / 15%
- **Epochs:** 2530 with early stopping
- **Batch Size:** 32
- **Optimizer:** Adam
- **Loss Function:** Binary Crossentropy
- **Regularization:** Dropout (0.30.5), L2 weight decay
- **Metrics:** Accuracy, Precision, Recall, F1-Score, ROC-AUC

4. Visual Representation of Methodology

a. Data Flow Diagram: Simplified System Workflow



b. Architecture Diagram: Detailed Model Pipeline Architecture