Mental Health Analysis ChatBot

NLP MINI PROJECT

INTRODUCTION

This paper presents a tailored mental health chatbot for addressing the acute mental health challenges . Leveraging AI and NLP, the chatbot, fine-tuned with insights from diverse social media, offers non-judgmental support and early diagnosis. Evaluation, using golden responses, examines its effectiveness and identifies areas for improvement, contributing to a nuanced understanding of its capabilities.

DATA SET

Dataset Sources:

- Hugging Face and Kaggle: Curated mental health conversations.
- Social Media Integration: Enriched from five distinct sources.
  - DR (Depression Recognition): Reddit posts for detecting signs of depression.
  - Dreddit (Stress Detection): Reddit dataset focusing on stress detection.
  - SAD (Stress Cause Detection): SMS conversations dedicated to detecting stress causes.
  - Irf (Interpersonal Risk Factors): Extracted from Reddit, detecting interpersonal risk factors.
  - MultiWD (Wellness Dimensions): Reddit dataset focusing on wellness dimensions.

Methodology

- **Data Merging:** Strategically merged for uniformity using Instruction finetuning.
- **Fine-tuning Parameters:** Utilized QLoRA, bitsandbytes configurations influencing attention dimensions, alpha parameters, dropout probabilities, and 4-bit precision.
- **Training Process:** Loaded datasets, tokenizers, pre-trained Llama 2 7B model, incorporated QLoRA and LoRA configurations, and executed training with SFTTrainer module.
- **Post Fine-tuning Evaluation:** Assessed correctness, quality, and label inferencing to validate the chatbot's proficiency in recognizing and addressing mental health concerns.

Evaluation:

- **Response Generation:**
  - Rigorous testing of fine-tuned model.
  - Comparison with golden responses.
  - Metrics: correctness, quality, label inferencing.
  - Ethical adherence for transparent reporting.
- **Correctness:**
  a. **Accuracy:**
    - Used BERT Sequence Classifiers (>90% accuracy).
    - Ensured precise identification of mental health concerns.

  b. **F1 Score:**
    - Balanced metric (precision and recall).
    - Enriched understanding, considering false positives and negatives.
-

Quality

1. **ROGUE Score:**
    - **Automated content similarity evaluation (n-gram overlap).**
2. **BLEU Score:**
    - **Precision-oriented metric measuring n-gram overlap.**
3. **BERT Score:**
    - **Bidirectional context understanding for language comprehension.**
4. **BART Score:**
    - **Specialized in abstractive summarization tasks.**
    - **Gauges similarity considering structure and fluency.**

RESULTS

TABLE I
CORRECTNESS EVALUATION SCORES

Data set	Accuracy	F1 score
DR	68	66.8
SAD	62	61.1
MultiWD	71	71.56
DreadDit	74	71.4
Irf	74	73.09

TABLE II
QUALITY EVALUATION SCORES

Data set	ROGUE-L	BLEU
DR	0.35	0.22
SAD	0.34	0.24
MultiWD	0.49	0.35
DreadDit	0.33	0.24
Irf	0.39	0.28

TABLE III
QUALITY EVALUATION SCORES

Data set	BERT	BART
DR	0.87	-2.96
SAD	0.89	-2.93
MultiWD	0.92	-2.30
DreadDit	0.91	-2.62
Irf	0.90	-2.66

RESULTS

TABLE I
CORRECTNESS EVALUATION SCORES

Data set	Accuracy	F1 score
DR	0	0
SAD	62	61.1
MultiWD	71	71.56
DreadDit	74	71.4
Irf	74	73.09

TABLE II
QUALITY EVALUATION SCORES

Data set	ROGUE-L	BLEU
DR	0	0
SAD	0.34	0.24
MultiWD	0.49	0.35
DreadDit	0.33	0.24
Irf	0.39	0.28

TABLE III
QUALITY EVALUATION SCORES

Data set	BERT	BART
DR	0	0
SAD	0.89	-2.93
MultiWD	0.92	-2.30
DreadDit	0.91	-2.62
Irf	0.90	-2.66