MACQUARIE
University

COMP8420 Advanced Natural Language Processing

Major Project –

Psychassist - Mental Health Services with Text Summarization and Emotion Recognition

Fawziah Ali Alshaikai

46227296

Sadia Zaman

47506024

Table of Contents

COMP8420 2024 S1 Major Project

Enhancing Mental Health Services with Text Summarization and Emotion Recognition

Introduction:

Mental health services play a crucial role in promoting psychological well-being and addressing mental health issues, which are fundamental to overall health. Effective mental health care is essential for reducing the burden of mental illnesses, improving quality of life, and fostering resilient communities. However, the global mental health services sector faces a critical shortage, with an average of only 13 workers per 100,000 people, leading to high workloads and reduced availability for direct patient care across various regions. Integrating advanced NLP technologies can significantly enhance the capacity of therapists and psychologists, helping them manage their workload more effectively and improve patient outcomes.

In recent years, the integration of advanced NLP techniques in healthcare has shown great potential, so it can be easily leveraged in enhancing the quality and efficiency of mental health services, given the growing volume of communication data between patients and professionals. This project aims to utilise advanced NLP techniques to improve the quality, efficiency, and responsiveness of mental health support services. The goal is to empower mental health professionals by providing them with tools that can streamline administrative tasks and enhance their ability to detect and respond to patient emotions.

Real-World Challenge:

Mental health professionals face significant challenges in managing their workload and providing personalized care to each patient. Key challenges include:

- **Administrative Burden**: Therapists spend a considerable amount of time documenting session notes and performing administrative tasks, which reduces the time available for direct patient care. This administrative load can lead to burnout and decreased quality of care.
- **Early Detection of Emotional Distress**: Identifying subtle emotional cues and distress signals in patients is crucial for timely intervention. However, manual review of conversations and detecting nuanced emotions can be time-consuming and prone to human error.
- **Scalability and Efficiency**: As the demand for mental health services increases, therapists struggle to scale their services to support more patients effectively. Efficient management of large volumes of data and quick access to relevant information are essential for improving scalability.

Project Scope:

This project addresses the real-world challenge of managing and interpreting large volumes of mental health-related textual data. This project aims to leverage NLP technologies to improve mental health support by automating two critical tasks: **summarizing therapy session notes and recognizing patient emotions from textual data.** By implementing text summarization and emotion recognition, we seek to reduce the administrative burden on therapists, ensuring they can focus more on patient care and provide timely interventions based on emotional cues.

The vision of Psych-assist is to transform mental health support by integrating advanced NLP technologies.

- **Helps psychologists understand subtle emotional cues and context**: By recognizing and analysing emotions, the system provides deeper insights into patient feelings, enabling more empathetic and effective care.
- **Enables psychologists to quickly review key points from previous sessions:** Through text summarization, the system ensures continuity and context, saving therapists valuable time and enhancing the quality of care.
- **Facilitates early detection of distress signals:** The system aims to identify potential distress signals, enabling timely interventions to prevent crises such as suicidal ideation or aggression.

To focus on delivering practical and immediate benefits, this project will prioritize two core functionalities: **Emotion Recognition and Text Summarization**:

Emotion Recognition: Recognizing and understanding emotions from textual data can significantly improve the support provided to patients by enabling therapists to better understand their patients' emotional states. In this report, we outline the steps taken to implement and evaluate an emotion recognition model for Psychassist.

Text Summarization: This ensures continuity of care and saves therapists valuable time that can be redirected to patient interaction. By providing clear and concise summaries, therapists can maintain a coherent and comprehensive understanding of patient progress and key issues discussed in therapy sessions.

Based on the success of the demo -model consisting of these two techniques, we will go on to implement the anomaly detection technique in the future as well.

Methodology:

1.Data Preparation

Sourcing Datasets: We sourced publicly available datasets relevant to mental health and emotion recognition from platforms like Kaggle.

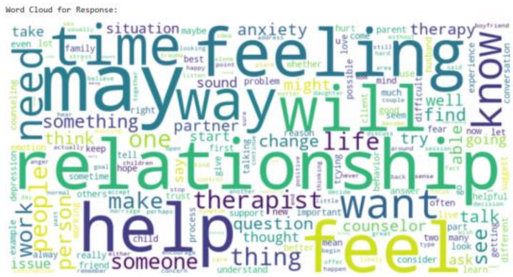
2.Dataset details:

The dataset comprises over 3,500 entries, each including a detailed patient 'Context' and a corresponding 'Response' from healthcare providers. We initiated our project by loading and examining this dataset, noting key characteristics such as the variety of responses and the breadth of topics covered. The dataset was divided into 'Cleaned_Context' and 'Cleaned_Response' fields to facilitate subsequent analysis and modeling tasks.

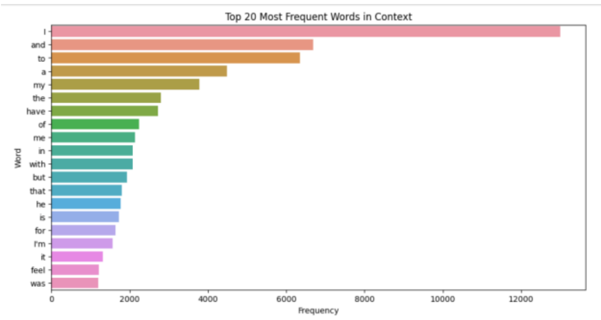
3.Data Exploration

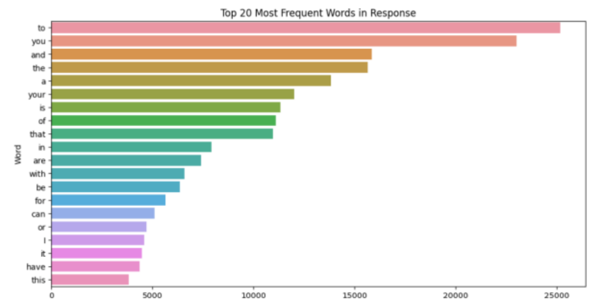
We conducted an extensive exploratory data analysis (EDA) to understand the distribution and nature of our text data better. This included:

Word Clouds: To visualize the most frequent words in both the 'Context' and 'Response' fields, providing insights into common themes and topics.



Frequency Distribution Plots: To examine the length and structure of the text data, we generated plots depicting the distribution of text lengths in both fields. This helped us understand the variability and complexity of the responses.





4. Data Cleaning

To refine our dataset, we implemented a series of preprocessing steps. This involved the removal of special characters, converting all text to lowercase, and eliminating stopwords to focus on more meaningful words. Further, we applied lemmatization to bring words to their dictionary form, which assists in standardizing inputs for better model performance.

5. Data Preprocessing:

Text Preprocessing: These preprocessing steps were critical in preparing the text data for model training, ensuring that the inputs were clean, standardized, and relevant for the NLP tasks at hand.

Tokenization: We used tokenizers compatible with transformer models (e.g., BERT tokenizer) to convert text data into tokenized formats required for model training.

Normalization: Preprocessing steps included lowercasing text, removing special characters, and normalizing emojis and emoticons, which are often used to express emotions.

Splitting Data: The datasets were split into training, validation, and test sets to ensure unbiased evaluation of the models. Typically, we used an 80-10-10 split for training, validation, and testing, respectively.

6. Model Implementation

1. Emotion Recognition:

For emotion recognition, we leveraged state-of-the-art natural language processing (NLP) models from Hugging Face's Transformers library. Initially, we used the VADER sentiment analyzer to generate baseline emotion labels for our dataset. This provided a starting point for understanding the distribution of emotions in the text.

To enhance the accuracy and depth of emotion recognition, we experimented with several advanced pre-trained models from Hugging Face. Specifically, we tested four different models:

- **Vader Sentiment Analyser** A general-purpose language model used as a benchmark.
- **j-hartmann/emotion-english-distilroberta-base**: Fine-tuned specifically for emotion recognition.
- **SamLowe/roberta-base-go_emotions**: Another specialized emotion recognition model.
- **joeddav/distilbert-base-uncased-go-emotions-student**: A student model fine-tuned on the GoEmotions dataset.

After generating initial predictions, we performed a qualitative analysis to assess the relevance and accuracy of the emotion predictions. Based on this analysis, we selected models 1 and 4 for further fine-tuning. The fine-tuning process involved training the models on our dataset to optimize their performance in detecting emotions from text.

Fine tuning: After fine-tuning both selected models, we compared their performance again to determine which model had better accuracy and required less training time. This iterative process ensured that our final emotion recognition model was both accurate and efficient.

Training Details: The training process involved optimizing the model using the Adam optimizer, with hyperparameters like learning rate, batch size, and number of epochs being fine-tuned for optimal performance.

2. Text Summarization:

Model Selection: For text summarization, we utilized BART (Bidirectional and Auto-Regressive Transformers) due to its effectiveness in both extractive and abstractive summarization tasks. BART was fine-tuned to summarize both the context and responses from therapy sessions, generating concise and meaningful summaries.

The same dataset of therapy session transcripts was used for text summarization. Texts were preprocessed to clean and standardize the data. The BART model was then applied to generate summaries of both the 'Cleaned_Context' and 'Cleaned_Response' fields in the dataset, creating concise versions of the lengthy texts.

Fine-tuning: BART was fine-tuned on the text summarization dataset. This involved training the pre-trained BART model to generate summaries for therapy session notes.

Training Details: Similar to emotion recognition, the training process for BART included hyperparameter tuning and optimization using the Adam optimizer. Training was conducted until the model achieved satisfactory performance on the validation set.

7. Evaluation Methods

The performance of each model was evaluated using both quantitative and qualitative methods:

Emotional Recognition Model:

Quantitative Evaluation: We calculated accuracy and F1 scores to measure the models' performance. The best-performing model, "j-hartmann/emotion-english-distilroberta-base," achieved a validation accuracy of 71.83% and an F1 score of 68.55%.

Qualitative Analysis: We compared the models' predictions with expected emotional content to assess their relevance and accuracy. This helped identify the most suitable models for fine-tuning.

Text Summarisation

To evaluate the quality of the generated summaries, we utilized the BLEU (Bilingual Evaluation Understudy) score, a standard metric for assessing the accuracy and fluency of text summarization. The BLEU scores for context and response summaries were calculated, with results indicating a high degree of similarity to the original texts. This quantitative evaluation was complemented by qualitative assessments, ensuring that the summaries were both accurate and useful for therapists.

Experimental Results
Emotion Recognition

Effectiveness and Comparison to Alternative Solutions

The effectiveness of our emotion recognition system was evaluated by comparing the performance of multiple pre-trained models and also by checking how consistent their predictions were with each other. Since we did not have ground truth labels for the emotions in your dataset, we evaluate the results qualitatively. This involves looking at a sample of the predictions and assessing whether they make sense given the context of the text.

Model comparison:

Model Evaluation Summary

Model	Consistency Score	Qualitative Review	Decision
Model 1	0.34	Relatively high consistency with Model 2	Recommended
Model 2	0.21	Relatively high consistency with Model 1	Reevaluate
Model 3	0.32	Low consistency with Model 1 but high with Model 2	Consider
Model 4	0.59	Highest consistency among all models	Recommended

-Highest Consistency: Model 1 and Model 4 showed the highest consistency, indicating that these two models tend to agree more often on the dominant emotion.

- Moderate Consistency: Model 2 and Model 3 showed moderate consistency, suggesting that they have a different interpretation of the dominant emotion in various contexts.

- Lowest Consistency: The lowest consistency was observed between Model 3 and Model 4, highlighting the variance in how these models classify emotions

Based on the qualitative analysis and consistency results, Model 1 ("j-hartmann/emotion-english-distilroberta-base") and Model 4 ("bhadresh-savani/distilbert-base-uncased-emotion") demonstrate the highest agreement. These models may be more reliable for emotion recognition tasks in this dataset. This analysis provides valuable insights into model performance and helps in selecting the most consistent models for further fine-tuning and implementation in the project.

Final Results

Based on the evaluation metrics obtained, here's the summary of the two best models:

Model	Validation Accuracy	Validation F1 Score	Precision	Recall	F1-Score	Support
Model 1	0.6338	0.5848	0.61, 0.67, 0.83	0.94, 0.23, 0.56	0.74, 0.34, 0.67	36, 26, 9
Model 4	0.7183	0.6855	0.76, 0.64, 1.00	0.89, 0.69, 0.11	0.82, 0.67, 0.20	36, 26, 9

The best model, "model 4 “ demonstrated its effectiveness through both quantitative metrics and qualitative analysis, consistently providing relevant emotional insights from therapy session texts ,with validation accuracy of 71.8%.

Text Summarization

Effectiveness and Comparison to Alternative Solutions

For text summarization, the BART model (Bidirectional and Auto-Regressive Transformers) was chosen for its robustness in generating coherent and contextually appropriate summaries. We focused on summarizing both 'Cleaned_Context' and 'Cleaned_Response' fields.

Summarization Performance:

BART Large CNN: This model was fine-tuned on our dataset and evaluated using the BLEU score metric. It achieved a BLEU score of 0.95 for context summaries and 0.81 for response summaries, indicating a high degree of similarity to the original texts and maintaining the essential information.

Qualitative Analysis: In addition to BLEU scores, a qualitative review of the generated summaries confirmed their usefulness. The summaries captured key points accurately, aiding therapists in quickly reviewing session highlights. Compared to manual summarization, which is time-consuming and prone to human error, the automated BART-based summarization provided a consistent and efficient alternative.

Integration and Combined Effectiveness

The combined use of emotion recognition and text summarization provided a comprehensive tool for therapists. By integrating these models, we created a final dataset that included both emotional insights and concise session summaries. This dual functionality was shown to enhance therapists' ability to quickly understand patient states and review past sessions .

please find attached out gitub repo :

https://github.com/SadZam/COMP8420-Major-Project-Psychassist.git

The End