NLP PROJECT

ANALYZING THERAPY SESSIONS

MEET THE TEAM:

SAMEERA NALLAMOTHU – SE22UARI150
KANNIKANTI CHANDANA PRIYA- SE22UARI067
SUSMITHA MANTHENA – SE22UARI170
SACHETHAN PILLI – SE22UARI199

INTRODUCTION:

- **Aim**:
  - To analyze therapy sessions using Natural Language Processing (NLP) techniques for behavior classification and session summarization.
- **Objectives**:
  - Utilize DistilBERT for classifying conversational patterns.
  - Implement an automatic summarizer to provide concise summaries of therapy sessions.
- **Motivation**:
  - Provide therapists with insights into session dynamics and improve the efficiency of therapy session reviews.
  - Help therapists focus on key points and better manage their sessions.

DATASET:

Training Dataset:

- Real-world therapy dialogues used for training the models.
- We utilized reduced_dataset.csv, which includes samples of text data labeled with behavior categories.
- Example entry:
- "I feel frustrated because of my workload." → *Label:* "Frustration."

	subreddit	post
0	guns	Bersa Thunder 22: Long-term reliability? Hi, f...
1	guns	Anybody have any experience with the diamondba...
2	guns	Old & New // Colt 1911 & MCX Rattler I...
3	guns	Start working in a gun shop next week. Need so...
4	guns	Cheap targets for plinking? I'm headed to the ...
...	...	...
1024070	addiction	Anyone relate I have relapsed a lot even after...
1024071	addiction	Will my husband ever stop using coke? My husba...
1024072	addiction	Im 15 and cant stop stealing First i just want...
1024073	addiction	Heroin overdose, medical doctor advice? I had ...
1024074	addiction	How much does drug rehab cost? How much does r...

DATASET:

Testing Dataset:

•The testing was conducted on a separate Hugging Face dataset, mental_health_counseling_conversations.

•It features real-world scenarios with *Context* (user queries) and *Response* (counselor replies).

Context string · lengths 25 2.7k	Response string · lengths 0 32.7k
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless…	If everyone thinks you're worthless, then maybe you need to find new people to hang out with.Seriously, the social context in whic…
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless…	Hello, and thank you for your question and seeking advice on this. Feelings of worthlessness is unfortunately common. In fact, most…
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless…	First thing I'd suggest is getting the sleep you need or it will impact how you think and feel. I'd look at finding what is going…
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless…	Therapy is essential for those that are feeling depressed and worthless. When I work with those that are experiencing concerns…
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless…	I first want to let you know that you are not alone in your feelings and there is always someone there to help. You can alway…
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless…	Heck, sure thing, hun!Feelings of 'depression' have a deeply-rooted base in physical structures that may not be functioning…

DATA PREPROCESSING:

Steps Taken:

- **Text Normalization** : Converted text to lowercase to ensure uniformity.
- **Noise Removal** : Removed punctuation, numbers, and non-alphabetical characters to reduce noise in the data.
- **Stemming and Stopword Removal** :

1. Applied stemming (e.g., running → run) using NLTK's PorterStemmer .
2. Eliminated common words (e.g., the, and, is) that do not add significant meaning.

- **Label Encoding** : Used Label Encoder to transform categorical labels into numeric representations for model compatibility.
- Example: "Stress" → 0, "Frustration" → 1.
- **Challenges Addressed** : Ensuring preprocessing does not lead to excessive information loss, as context is vital for classification tasks.

MODELS EXPLORED:

Explored ModelsBERT (Bidirectional Encoder Representations from Transformers)

- Why Considered : Standard for many NLP tasks . Strong generalization for classification and text summarization.
- Limitations : Computationally intensive . Challenging for real-time or resource-constrained setups.

RoBERTa (Robustly Optimized BERT)

- Why Considered : Enhanced pre-training compared to BERT . **DistilBERT**
- Why Considered :It's a smaller and faster version of BERT but still gives almost the same accuracy. Great for tasks where speed and efficiency matter.
- Strengths : Fast training and inference times .

T5 (Text-to-Text Transfer Transformer)

- Why Considered : Unified framework for NLP tasks (classification, summarization).Strong performance in generating summaries.
- Limitations : Overhead in handling multi-task learning . Complex for single-domain tasks like therapy analysis.

GPT-Based Models

- Why Considered : Excels in language understanding and generation tasks . Good at conversational and dialogue-based summarization.
- Limitations : Tendency to generate verbose outputs . High computational resource demands . Model Selection

Final Choice: DistilBERTSelected for its balance of performance, speed, and computational efficiency . Ideal for scaling the solution to larger datasets or real-time analysis

.

MODEL ARCHITECTURE:

About DistilBERT :

- A distilled version of the popular BERT model. It's a smaller and faster version of BERT but still gives almost the same accuracy. Great for tasks where speed and efficiency matter.
- Designed specifically for tasks requiring speed and efficiency without sacrificing accuracy.

Sequence Classification :

- Fine-tuned for multi-class behavior classification.
- Uses transformer-based architecture for understanding relationships within sentences.

Why DistilBERT Works Well:

- Pre-trained on a large corpus, allowing it to understand complex sentence structures and nuances.

TRAINING PIPELINE:

Implementation:

- Created a custom Dataset class for tokenizing input text and preparing it for the DistilBERT model.
- Utilized Hugging Face's Trainer API for efficient model training and hyperparameter optimization.

Key Parameters :

- Learning Rate: 5e-5 for stable and effective weight updates.
- Batch Size: 100, balancing memory constraints and speed.
- Epochs: 3, to prevent overfitting while ensuring adequate learning.

Model Outputs :

- Fine-tuned DistilBERT model capable of behavior classification.
- Saved tokenizer and label encoder for easy deployment.

Challenges :

- High memory usage during training due to batch size.Limited token length (max_len=10) restricted the model's ability to process longer sentences.

OUTPUT AFTER TRAINING THE MODEL:

```
Evaluation results: {'eval_loss': 1.7744554281234741, 'eval_runtime': 36.9013, 'eval_samples_per_second': 1110.069,
'eval_steps_per_second': 11.111, 'epoch': 0.9993898718730934}
                                                     post        predicted
151830  figur share may help ya httpswwwredditcomrcoro...       depression
90592   much ira invest mutual fund im finish grad sch...  personalfinance
186469  fight fear futur turn year oldest child famili...       depression
64692   bodili ach mental agoni past week ive sever ac...       depression
110473  agre worker believ ive misclassifi option new ...  personalfinance
...                                                    ...             ...
7872    ne flsa question partner work sport busi midwe...  personalfinance
11011   one realli want help ive struggl depress anxie...       depression
37091   im move anoth contin start new job feel someth...       depression
193699  ten year schizoversari around time first start...       depression
129011         lone joke im lone even duolingo send email           lonely

[40963 rows x 2 columns]
```

TESTING PIPELINE:

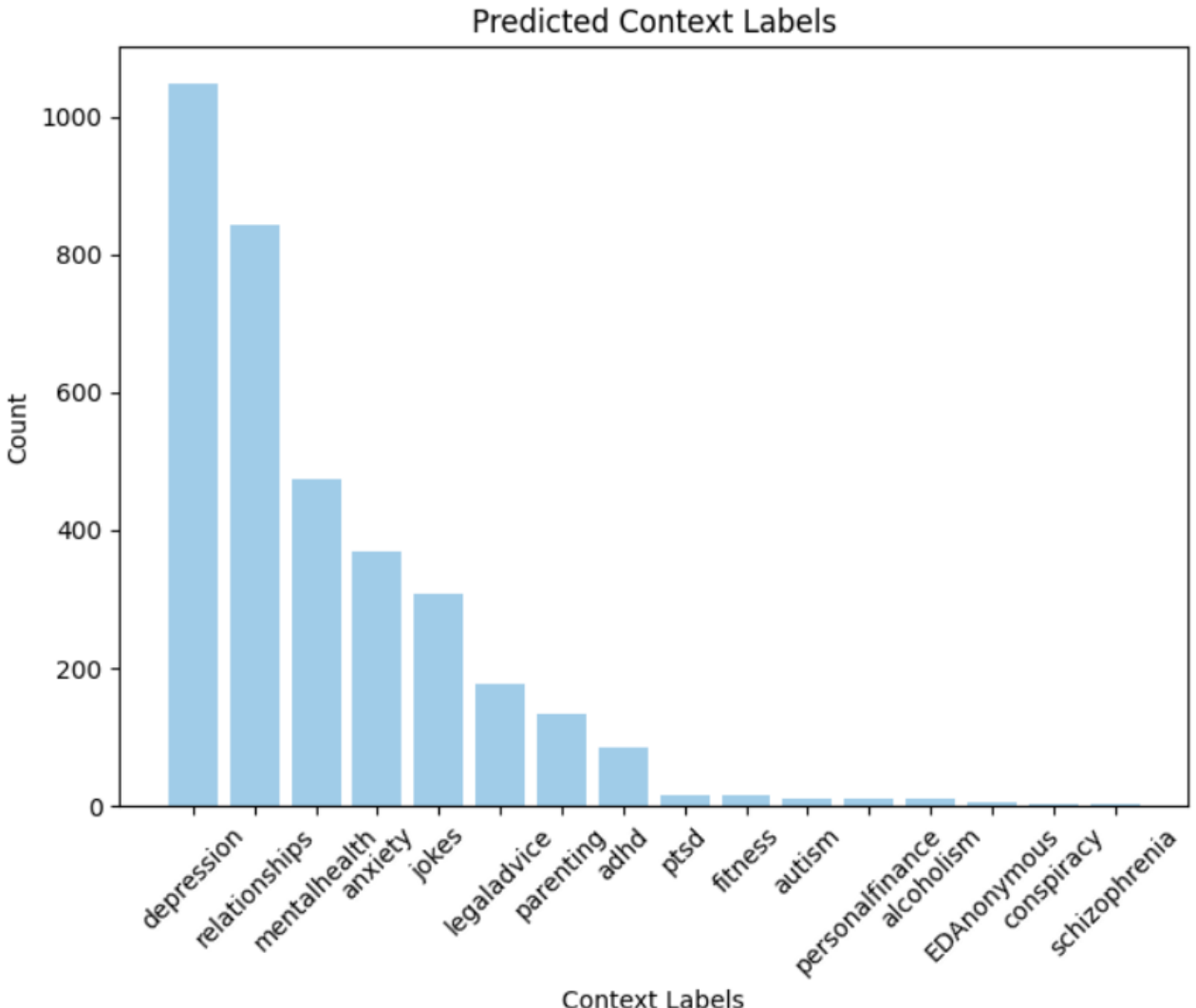
- **Process Overview:**
  - Reloaded the fine-tuned model, tokenizer, and label encoder to test on new data.
  - Input: Text pairs consisting of *Context* and *Response*.
  - Output: Predicted behavior labels for both inputs.
- **Evaluation:**
  - Conducted visual analysis using bar plots to examine the distribution of predicted labels.
- **Example Prediction:**
  - *Context:* "I feel like giving up." → Predicted Label: *Despair.*
  - *Response:* "It's okay to feel that way; let's find solutions together." → Predicted Label: *Support.*

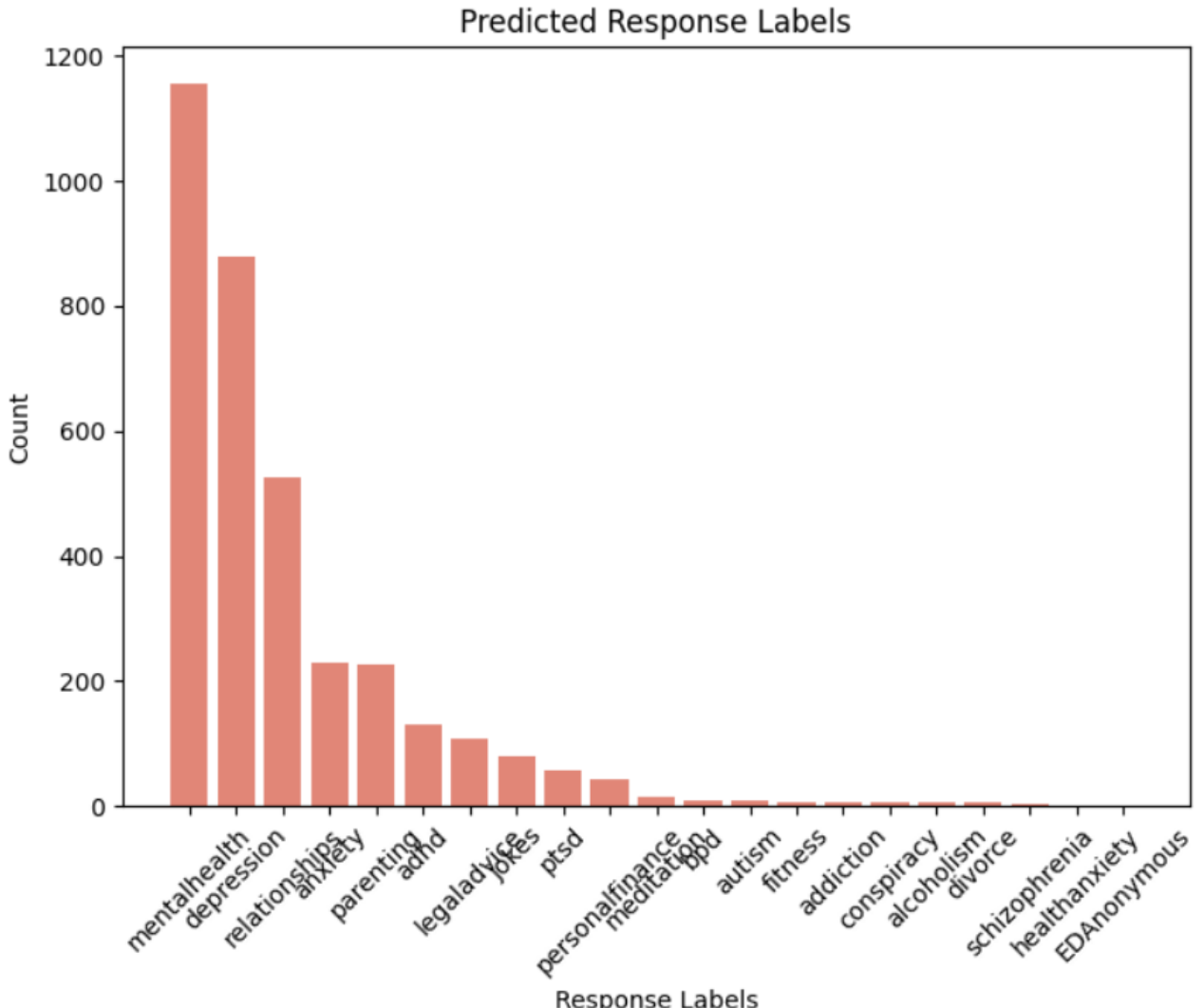
OUTPUT AFTER TESTING TRAINED MODEL:

	Context	predicted_context	Response	predicted_response
0	I'm going through some things with my feelings...	depression	If everyone thinks you're worthless, then mayb...	depression
23	I have so many issues to address. I have a his...	mentalhealth	Let me start by saying there are never too man...	mentalhealth
70	I have been feeling more and more down for ove...	anxiety	Answers about our inner lives are most success...	anxiety
72	I'm facing severe depression and anxiety and I...	anxiety	Have you used meditation or hypnosis? Relaxing...	adhd
81	How can I get to a place where I can be conten...	depression	Your question is a fascinating one!As humans w...	mentalhealth

OUTPUT GRAPHS FOR TRAIN AND TEST DATASET:

The bar graphs show the distribution of predicted labels for context and response categories.





Session Summarization: Approach and Results:

Approach:

Model Selection:

- Utilized BART (Bidirectional and Auto-Regressive Transformer), a pre-trained model specifically designed for abstractive summarization tasks.

Data Preparation:

- Combined Context and Response fields from the dataset to create a single input text for summarization.
- Applied preprocessing to ensure input quality for the summarizer.

Implementation Steps:

- Created a Python-based pipeline using the transformers library
- Summarized therapy conversations sequentially, storing outputs in a text file for further analysis.
- Setup environment for efficient and isolated execution using virtual environments

Summarization Workflow:

- Input: Merged text from "Context" and "Response.“
- Output: Concise summaries highlighting key points and recurring themes.

Results:

Performance:

- Successfully reduced session transcripts to approximately 20-30% of their original size while retaining critical information.
- Enabled therapists to quickly identify core issues, such as critical emotions, recurring themes, and notable client progress.

Example output:

Original Context

- Context (Patient's Statement):
  "I often feel overwhelmed and disconnected from the people around me. Even in social settings, I struggle to engage and feel like I don’t belong. These feelings have been affecting my confidence, making it hard to speak up or share my thoughts."

- Response (Therapist's Reply):
  "It’s natural to feel this way when your confidence is low. However, it’s important to focus on small, achievable steps to reconnect with those around you. Start by identifying one person you trust and open up to them about your feelings. Over time, these efforts can help you rebuild your sense of belonging and confidence. Remember, you are not alone, and support is always available.“

Generated Summary

- The patient struggles with feelings of disconnection and low confidence, particularly in social settings. The therapist suggests rebuilding trust and confidence through small, actionable steps, such as confiding in a trusted individual and seeking support.

OUTPUT:

- **Dataset:** The dataset used in this project was accessed from https://huggingface.co/datasets/Amod/mental_health_counseling_conversations
- **Output:**
- Link (click on it to access):

summaries.txt

```
Conversation 1 Summary:
The social context in which a person lives is a big influence in self-esteem. If
everyone thinks you're worthless, then maybe you need to find new people to hang
out with. There are many inspirational messages you can find in social media.

Conversation 2 Summary:
I'm going through some things with my feelings and myself. I barely sleep and I
do nothing but think about how I'm worthless. I've never tried or contemplated
suicide. How can I change my feeling of being worthless to everyone?

Conversation 3 Summary:
CBT is good for individuals dealing with depression, anxiety, toxic
relationships, stress, self esteem, codependency, etc. CBT helps with gaining a
better awareness of how your thought process influences your beliefs.

Conversation 4 Summary:
Many of the symptoms you have described are consistent with a person who is
dealing with depression. Depression is a treatable condition. If these symptoms
have persisted for more than two weeks, then it is a good idea to seek
professional help.

Conversation 5 Summary:
I barely sleep and I do nothing but think about how I'm worthless and how I
shouldn't be here. I've always wanted to fix my issues, but I never get around to
it. How can I change my feeling of being worthless to everyone? People who feel
worthless were, in one way or another, told that they are worthless.

Conversation 6 Summary:
I barely sleep and I do nothing but think about how I'm worthless and how I
shouldn't be here. I've never tried or contemplated suicide. How can I change my
feeling of being worthless to everyone? I'm glad you are interested in changing
```

RESULTS:

- **Prediction Results:**
  - Bar plots showed the distribution of predicted labels for both *Context* and *Response*.
  - Results reflected the model's ability to distinguish behaviors based on linguistic cues
  - Successfully reduced session transcripts to 25-30% of their original size.
- **Observations:**
  - High accuracy in identifying negative emotions like *Stress* and *Frustration*.
  - Slight confusion in closely related labels due to limited training data.
- **Key Insights:**
  - Visualization helped identify areas where the model performed well and where it needed improvement.
  - Summaries retained essential context and captured the core issues discussed.

CONCLUSION:

- **Project Outcomes**:
  - Successfully developed a system combining behavior classification and session summarization for therapy sessions.
  - The summarizer enhances session reviews by providing quick summaries of key points.
- **Implications**:
  - This system helps therapists by improving their ability to track and reflect on therapy outcomes effectively.
  - It also offers potential for large-scale mental health studies, improving care strategies.
- **Future Scope**:
  - More data sources and real-time analysis could broaden the system's impact, potentially aiding therapists and researchers in real-time therapy settings.