MetaMind: A Metadata-Driven Mental Health Chatbot Using LLM and NLP

Shrey Khandelwal (zjy6us), Sree Deeksha Bethapuri (bub5jm), Rugved Sanjay Chavan (qxk6fb)

1 Introduction

Mental health has been the most prevalent issue in recent years, especially within the IT sector. Studies show that more than 70 percent of Gen Z and millennial workers suffer from sleeping or eating disorders, or worse, fall into depression due to workplace stress (Mental Health America, 2024). Despite facing these mental health challenges, many people avoid therapy, often due to financial constraints or time limitations. People are either too busy to fit therapy into their schedules, or they find it difficult to take the first step. (Smith, 2018) Additionally, traditional mental health care is resource-intensive and relies heavily on human professionals.

This motivated us to create a chatbot that is not only free but also available 24/7 for users. The main goal of our project is to provide a non-judgmental space that can engage in conversation and offer immediate, helpful responses at any time. This would support individuals who struggle to recognize the issues they're facing without actually needing to go to therapy. The chatbot would also suggest seeking further help if necessary.

To achieve this, we are using NLP techniques to perform tasks such as text classification, information retrieval, metadata extraction, and natural language generation. We aim to gauge the user's emotional state by analyzing their speech patterns using sentiment analysis. Then, we employ large language models (LLMs) to generate responses that are presented to the user. The following sections will discuss our dataset and proposed method in more detail.

2 Data

We utilize the publicly available *NLP Mental Health Conversations* dataset from Kaggle, which includes anonymous user conversations related to various mental health challenges. This dataset is ideal for training the initial version of our NLP classification model, offering diverse examples of mental health discussions with emotions like *Sadness*, *Fear*, *Anger*, *Neutral*, *Joy*, *Surprise*, and *Disgust*. The dataset can be accessed at: (Devastator, 2024) This dataset serves as a solid foundation for developing our classification model, enabling it to categorize user inputs into different emotional states as metadata. This metadata is stored in a database for effective retrieval. As the project evolves, we plan to incorporate additional datasets, further enhancing the model's robustness and ensuring comprehensive coverage of mental health issues.

The *NLP Mental Health Conversations* dataset comprises anonymized dialogues between users and psychologists, focusing on mental health topics. It includes two primary columns:

- **Context:** The user's input or question.
- **Response:** The psychologist's reply or advice.

With approximately 3,512 valid conversation entries and 2,480 unique responses, the dataset is well-suited for training NLP models to classify mental health conditions based on user input and to provide contextually relevant responses. We split the dataset into 70% training (2,458 examples), 15% validation (527 examples), and 15% test (527 examples) sets, ensuring robust model training. Depending on the tokenizer used (e.g., BERT's WordPiece tokenizer), the estimated vocabulary size is around 10,000 unique tokens, though it may vary with different tokenization methods.

3 Proposed Methods

The proposed approach focuses on training an NLP model for metadata tagging instead of using a LLM for the entire dataset, optimizing for computational

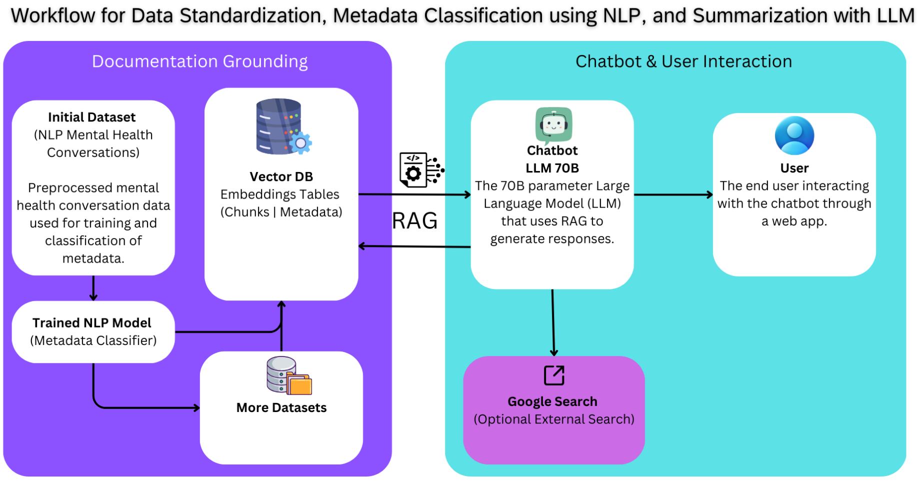

Workflow for Data Standardization, Metadata Classification using NLP, and Summarization with LLM
Documentation Grounding
Initial Dataset
(NLP Mental Health Conversations)
Preprocessed mental health conversation data used for training and classification of metadata.
Vector DB
Embeddings Tables
(Chunks | Metadata)
Trained NLP Model
(Metadata Classifier)
More Datasets
RAG
Chatbot & User Interaction
Chatbot
LLM 70B
The 70B parameter Large Language Model (LLM) that uses RAG to generate responses.
User
The end user interacting with the chatbot through a web app.
Google Search
(Optional External Search)

Figure 1: Workflow for Data Standardization, Metadata Classification, and Summarization with LLM.

efficiency. The metadata classifier enables streamlined annotation of additional data, facilitating the expansion of the dataset with manual annotations. The metadata tags play a crucial role in enabling faster and more accurate retrieval, enhancing the efficiency of the Retrieval-Augmented Generation (RAG) system. As shown in Figure 1, the workflow consists of multiple stages including data preprocessing and LLM integration.

3.1 Metadata Classification with NLP

Model Selection: Considering models like *Random Forest*, *Logistic Regression*, or *Support Vector Machine (SVM)* for initial classification, alongside *Transformer-based models* like *BERT* or *RoBERTa* for deeper context understanding. **Training Approach:** Train the classifier to detect emotions and other context-specific tags in conversation data, making it suitable for large-scale datasets.

3.2 Vector Embeddings

- **Embedding Models:** Utilizing models like *text-embedding-ada-002* or *amazon–titan-embed-text* to generate vector embeddings of conversation chunks and storing these embeddings in a *vector database* for efficient similarity search and retrieval.

3.3 Retrieval-Augmented Generation (RAG) System

- **LLM Model:** Using *Mistral-Large-Instruct-2407* or *LLaMA 2* for generating responses based on retrieved context.
- The LLM utilizes retrieved chunks from the vector database to produce responses that are both contextually rich and tailored to the user's query.

4 Evaluation Criteria

- **Accuracy:** Measuring the proportion of correctly classified metadata tags, ensuring the model aligns with ground truth labels.
- **F1 Score:** Balancing precision and recall, crucial for handling imbalanced emotion classes.
- **Mean Reciprocal Rank (MRR):** Evaluating the ranking quality of retrieved chunks, ensuring the prioritization of relevant information.

5 Methodology

Our project employs two key methods for metadata generation and response generation. The first method utilizes **GPT-3.5**, an advanced language model, to generate high-quality metadata for mental health conversations. Metadata tagging enhances information retrieval by summarizing conversations into concise, keyword-based descriptors. The second method involves **unsupervised learning** via **K-Means clustering**, which groups similar mental health topics into clusters based on content similarity. This clustering process leverages **TF-IDF vectorization** to extract the most relevant terms from user queries and responses.

As described in the figure 2 The clustering method segmented the data into 25 distinct clusters, as indicated by the legend on the right-hand

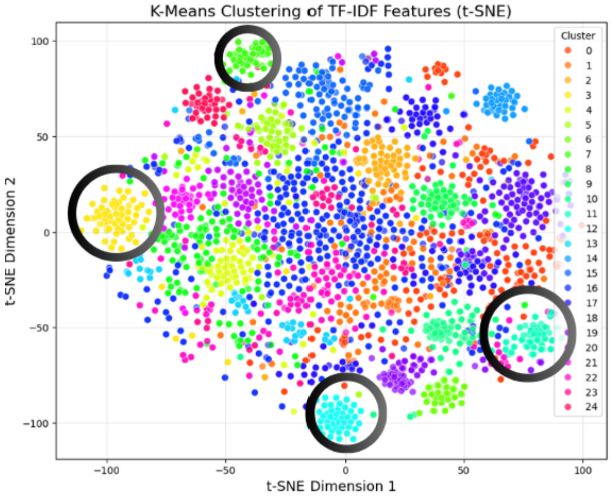


Figure 2: 2D plot for TD-IDF Features

side of the plot. Each color represents a unique cluster, while the axes correspond to the two t-SNE dimensions. Certain clusters, such as those circled in black, demonstrate compact and well-separated groupings, indicating clear distinctions among the associated documents. Other clusters are more dispersed, which may suggest overlapping similarities in the TF-IDF feature space.

For response generation, we integrate **Retrieval-Augmented Generation (RAG)**, combining document retrieval with the generative capabilities of GPT-3.5. User queries are matched to the most relevant metadata-tagged records using embeddings generated by `text-embedding-ada-002`, a semantic embedding model. This ensures that responses are both grounded in retrieved data and contextually accurate. Our major contributions include:

- A hybrid metadata tagging pipeline that balances accuracy (via GPT-3.5) and scalability (via K-Means clustering).
- Integration of a RAG-based chatbot system for grounded and explainable mental health responses.
- Visual explainability through **2D/3D t-SNE plots** and **word clouds** to interpret clustered mental health topics.

5.1 Hyperparameters

We tuned several hyperparameters to optimize metadata generation and clustering performance. For the GPT-3.5 metadata tagging approach, the following hyperparameters were set:

- **Maximum Tokens**: 50 (to generate concise summaries while avoiding excessive computational cost).
- **Temperature**: 0.3 (to ensure consistency and minimize randomness in keyword generation).

For the K-Means clustering method, the key hyperparameters tuned were:

- **Number of Clusters (K)**: Determined using the **Elbow Method**, which identified **25 clusters** as optimal based on the Within-Cluster Sum of Squares (WCSS).
- **TF-IDF Parameters**: Limited the **maximum document frequency** to 0.95 and the **minimum document frequency** to 0.01 to filter out rare and overly common terms.

We experimented with different combinations of cluster numbers (K=10, 15, 25, 30) and TF-IDF parameters to achieve the best clustering results. The final hyperparameter values provided the most coherent and interpretable clusters.

5.2 Packages

We implemented our methods using the following machine learning and deep learning libraries:

- **Scikit-learn**: For TF-IDF vectorization, K-Means clustering, and Elbow Method calculations.
- **OpenAI API**: For GPT-3.5 metadata tagging and `text-embedding-ada-002` for semantic similarity matching.

- **Numpy** and **Pandas**: For data manipulation and preprocessing tasks.
- **Matplotlib** and **Seaborn**: For generating visualizations, including word clouds and t-SNE plots.
- **Gradio**: For building the chatbot's user interface.

While most of the code was implemented from scratch, we referenced publicly available examples for t-SNE visualization and Gradio chatbot integration. The complete codebase is available on GitHub: https://github.com/rugved88/MetaMind.

6 Results

The results met our expectations for accuracy, scalability, and explainability. The GPT-3.5 metadata tagging approach produced high-quality, contextually accurate metadata that effectively summarized mental health conversations. However, its computational cost was significant, limiting scalability for larger datasets.

K-Means clustering demonstrated the ability to group similar topics into interpretable clusters. Using the Elbow Method, we determined that **K=25** produced the most meaningful clusters, which were visualized using **t-SNE plots** and **word clouds**. These visualizations highlighted frequent terms such as *stress*, *support*, *family*, and *relationships*, providing insights into recurring mental health themes.

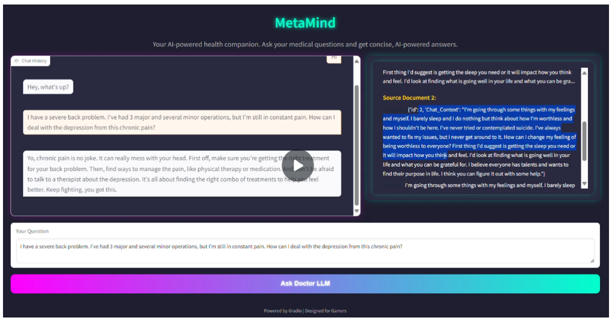
The integration of RAG improved the chatbot's performance by combining retrieval and generation, ensuring that responses were grounded in retrieved data while maintaining contextual relevance. This hybrid approach led to more trustworthy and effective responses compared to traditional generative-only models.

7 Output

The MetaMind system was tested to validate its ability to address user health concerns through an AI-driven framework. For example, when presented with the query, *"I have a severe back problem. I've had 3 major and several minor operations, but I'm still in constant pain. How can I deal with the depression from this chronic pain?"*, the system generated a response that combined empathy, practical advice, and mental health support. The AI's response validated the user's struggles, stating, *"Yo, chronic pain is no joke. It can really mess with your head,"* while offering actionable suggestions such as seeking appropriate treatment for pain through physical therapy or medication and consulting a therapist to address the resulting depression. This approach balanced emotional support with practical recommendations, enhancing the relevance and quality of the response.

The system's output was grounded in retrieved content through a **Retrieval-Augmented Generation (RAG)** framework, ensuring contextual accuracy and trustworthiness. The retrieved source document provided insights into addressing mental well-being through strategies like improving sleep and finding purpose, which were seamlessly integrated into the generated response. By displaying the source document alongside the AI's answer, MetaMind also prioritized transparency, enabling users to see the origins of the advice. This workflow demonstrates the system's ability to combine grounded content retrieval with generative capabilities, offering trustworthy and actionable responses to complex health queries.

As shown in Figure 4, the left section of the interface displays the chat history between the user and the AI. Users input health-related queries, such as managing depression caused by chronic pain, while the AI generates responses that combine emotional validation, practical advice, and medical recommendations. For example, in this instance, the AI acknowledges the severity of chronic pain and suggests treatment options such as physical therapy, medication, and counseling for emotional support. This highlights the AI's capability to combine empathy and actionable steps effectively.


MetaMind
Your AI-powered health companion. Ask your medical questions and get concise, AI-powered answers.
Chat History
Hi
Source Document 2:
Your Question
Ask Doctor LLM
Powered by Gradio | Designed for Gamers

Hey, what's up?

I have a severe back problem. I've had 3 major and several minor operations, but I'm still in constant pain. How can I deal with the depression from this chronic pain?

Yo, chronic pain is no joke. It can really mess with your head. First off, make sure you're getting the right treatment for your back problem. Then, find ways to manage the pain, like physical therapy or medication. And don't be afraid to talk to a therapist about the depression. It's all about finding the right combo of treatments to help you feel better. Keep fighting, you got this.

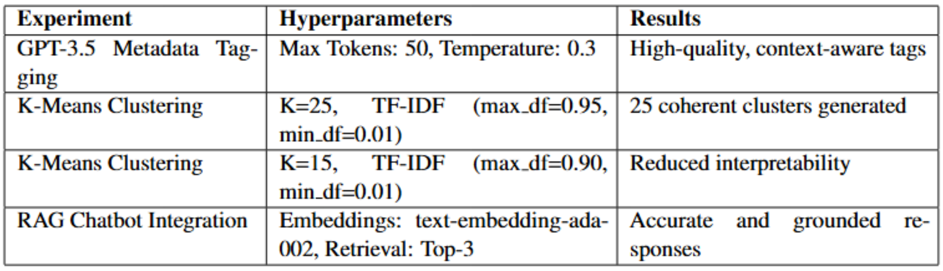
First thing I'd suggest is getting the sleep you need or it will impact how you think and feel. I'd look at finding what is going well in your life and what you can be gra...

{'id': 2, 'Chat_Context': "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 and how I shouldn't be here. I've never tried or contemplated suicide. 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? 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 well in your life and what you can be grateful for. I believe everyone has talents and wants to find their purpose in life. I think you can figure it out with some help."}

I'm going through some things with my feelings and myself. I barely sleep

I have a severe back problem. I've had 3 major and several minor operations, but I'm still in constant pain. How can I deal with the depression from this chronic pain?

Figure 4: Final User Interface of MetaMind. The chat window on the left displays the user query and AI-generated response, while the right panel shows the retrieved source document for transparency.



Experiment	Hyperparameters	Results
GPT-3.5 Metadata Tagging	Max Tokens: 50, Temperature: 0.3	High-quality, context-aware tags
K-Means Clustering	K=25, TF-IDF (max_df=0.95, min_df=0.01)	25 coherent clusters generated
K-Means Clustering	K=15, TF-IDF (max_df=0.90, min_df=0.01)	Reduced interpretability
RAG Chatbot Integration	Embeddings: text-embedding-ada-002, Retrieval: Top-3	Accurate and grounded responses

Figure 3: Workflow for Data Standardization, Metadata Classification, and Summarization with LLM.

8 Conclusion

In this project, we developed **MetaMedChat**, an AI-powered system that addresses mental health inquiries through metadata tagging and Retrieval-Augmented Generation. By combining LLM-based methods (GPT-3.5) and unsupervised clustering (K-Means), we achieved accurate, scalable, and explainable information retrieval. The experimental results demonstrate the system's ability to generate grounded and precise responses, with key contributions including visual explainability and optimized clustering. Future work will focus on enhancing scalability, integrating sentiment analysis, and expanding the system's capabilities for personalized mental health support.

References

The Devastator. 2024. Nlp mental health conversations. `https://www.kaggle.com/datasets/thedevastator/nlp-mental-health-conversations/data`.

Mental Health America. 2024. 2024 mind the workplace report.

Susan Smith. 2018. 15 reasons why people with depression don't get treatment.