International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.008

Mental Health Condition and Sentiment Based Chatbot

PrateekAnand
Department of Data Science and Business Systems
SRM Institute of Science and Technology
Chengalpattu District, Tamil Nadu, India
mathurprateek2001@gmail.com

Meet Desai
Department of Data Science and Business Systems
SRM Institute of Science and Technology
Chengalpattu District, Tamil Nadu, India
meetds8@gmail.com

Dr. SV. ShriBharathi
Assistant Professor, Department of Data Science and Business Systems
SRM Institute of Science and Technology
Chengalpattu District, Tamil Nadu, India
shribharathi01@gmail.com

Dr. N. Arivazhagan
Assistant Professor
Department of Computational Intelligence
SRM Institute of Science and Technology
Chengalpattu District, Tamil Nadu, India
arivazhn@srmist.edu.in

***Abstract*—In today's hectic lifestyle, over time, has arisen the need of looking after one's mental well being. However, many medical, social, financial and personal issues have turned out to be obstacles in timely detection and treatment of mental health diseases which may even cost the life of the sufferer. The presented work proposes a transformer-based mental health condition classification model and a sentiment-oriented response system to interact with the patient, giving human-like replies. Although numerous chatbots have been developed, very few have concentrated on the issues of mental health but that too with irrelevant responses. This approach aims at delivering positive chat system to the patients, using specific Reddit datasets for each category. The classification model resulted in an accuracy upto 93% and the response system has proven to be quite sensible and relevant.**

***Index Terms*—mental health, artificial intelligence, natural language processing, chatbot, transformers, sentiment analysis**

I. INTRODUCTION

During and after the Covid-19 pandemic, the social media caught its attention on a topic existing much before people knew about it and has been more important than people have believed it to be - mental health. While depression and anxiety are the two known conditions, there exist many more mental conditions which may arise due to emotional or psychological reasons and even due to irregularities in habits and lifestyle. Due to very less awareness about mental health issues and their signs and symptoms, the people suffering from many such diseases often become preys to prolonged mental illness, disorders and even suicides.

Due to such reasons, it is quite important to make people aware of their mental well being and so far it has been done by psychiatrists and psychotherapists. However, due to un- availability of a sufficient amount of professionals in this field and the unaffordable consultation fees and treatment charges have restricted people to get a cure of their mental issues andillness. Hence, in this running world of an hectic race, there is an urgent need of educating people about their own psychology and protecting them from the abstract wounds in a feasible manner. With the evolution of Artificial Intelligence and the introduction of chatbots, many individuals and organizations have aimed at providing aid to tensed people who have got a place to put their thoughts down without hesitation and feel a lot relaxed with their problems. However, there are several limitations with the response systems of these chatbots.

Firstly, almost all of them are oriented towards tackling depression and anxiety, putting no light on other mental health conditions. This leaving out leads to low efficacy of these chatbots and the users may get irrelevant responses which may not help them with their issues. Secondly, the way chatbots are designed becomes an obstacle in an interactive conversation and gaining actual feelings out of the users. As mentioned in the Wired article [1], chatbots like Woebot and Wysa though have claimed of being compassionate response systems, do not hold any proper survey report in the first place. Woebot is based on a guided chat system, does not allow the user to type his own words and instead provides input options. Although it is designed well, lacks human inputs and personalized response to each message. Wysa, on the similar ground, presents two different approaches of treatment- self-care and online consultation with a doctor, charging fees for both. Hence, although there are interactive chatbots to tackle mental health patients, their medical accuracy and user satisfaction turn out to be the major limitations. The presented approach, as shown in Figure 1 first classifies the patient into one of the 8 chosen mental health categories and then asks the user for inputs and giving them responses based on their categorized condition as well as the specific message, with a positive sentiment.

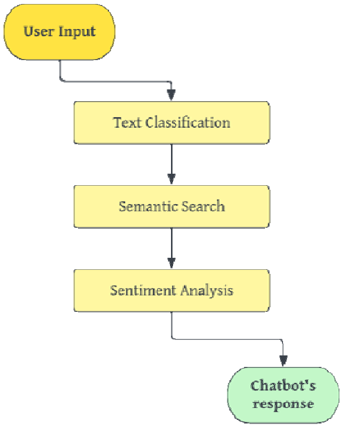

Fig. 1. Structural flow of the proposed work

II. Related Work

As rightly said in the Medical Device Network article [2], most of the mental health chatbots, instead of doing good, turning out to be harmful for their users, their market is unregulated and they do not have government's approval due to the lack of empirical evidence to back their efficacy and usage. The review paper by Alaa et.al. [3], concludes that the majority of the chatbots concentrated on depression and autism. However, there are many other major mental conditions and illnesses which the people should be aware of and the chatbots need to put light on. The presented work aims at dealing with 8 such categories and present an approach so that many more can be added.

Haoet. al. [4] in their work concluded that self-help chatbots helped university students in dealing with depression in a com- paratively better way than bibliotherapy which uses literature to cure the suffering. They further state that the process of conversation matters more than the content and have used statistical methods to illustrate their point. Their chatbot turned out to be much natural one as it involved emotion recognition and Natural Language Understanding(NLU) for interpreting user inputs and then used Natural Language Generation(NLG) with a response database to reply to that input. However, their chatbot was still based on the conventional dialogue management system that involves the user to choose from given options to give his inputs. A major aim of this workis removing this step of dialogue management and generate responses only on the basis of the user inputs, so as to give emphasis to every single input message.

Falguni et.al. [5] in their paper, have presented a chatbot that identifies the user's emotions based on text inputs and classifies each of them as one of the 8 chosen classes, namely joy, happy, shame, disgust, sadness, anger, fear and guilt. Further based on the negative emotions, the user is classified as normal, stressed or depressed. This classification and response system is a result of a pipelined process which involves ISEAR dataset for training and detecting emotion. Next in the pipeline is the classification step which uses CNN, RNN and HAN for categorizing the text into emotion labels and providing corresponding responses. This work takes up a similar approach but instead of categorizing text into emotion labels, classifies the user into one of the 8 mental health conditions based on his initial input.

In the review paper by Nick Boettcher [6], 54 studies and 425 abstracts were screened which dealt with Reddit data collection, focus on mental health conditions, analytics and practical implications. Approximately 63% of the studies based on practice implications suggested the use of Reddit data for professional practice related to human mental health. Some studies even suggested future applications of Machine Learning classification algorithms to help Reddit users access their mental health. As Reddit data appears to be a consid- erably good source of developing models and applications in the field of mental health [7], the Reddit community posts and their corresponding comments were scraped and manipulated to create a dataset for response generation.

Kim et.al. [8] in their paper, have presented a deep learning model to use social media user content and detect mental illness. They have used Reddit community data associated with each of the 6 mental illness classes, namely depression, anxi- ety, bipolar, bipolar disorder, autism and schizophrenia. They have used both XGBoost algorithm and CNN to accomplish the task of binary classification for each class and check if the illness exists or not. For depression they obtained upto 75% accuracy, 78% for anxiety, 90% for bipolar and BPD, 94% for schizophrenia and 96% for autism with their CNN model. The presented work too works on extracted Reddit dataset for the chosen conditions but has taken a step further by categorizing input into 8 categories and instead of binary classification that just checks the presence and absence of an illness, has used a category detection approach.

Anca et.al. [9], in their work have used the SMHD dataset[10] for mental health conditions and employed three deep learning models namely, BERT, RoBERTa and XLNET for disease classification. Similar to the previous paper, they have incorporated binary classification to detect the presence of 9 different conditions, namely depression, schizophrenia, OCD, eating disorder, PTSD, anxiety, BPD, ADHD and autism. Using this approach, they obtained accuracies ranging between 70-80%, each category holding its best accuracy in this range. The proposed approach has used BERT base (uncased) pre- trained model for categorical classification and has got muchbetter overall accuracy.

Taking insights from the cited literature, a novel approach is presented for creating a mental health chatbot involving multiple steps like condition classification, semantic similarity matching and sentiment analysis [11] along with database manipulation and querying. The proposed work aims at creating more personalized chatbot responses and make them motivational at the same time to reduce psychological illness at the grass-roots level itself.

III. Proposed Work

The presented work is divided into two sections - mental health condition categorization and sentiment-based response generation. Both the sections involve data extracted from Reddit communities, each dedicated to one of the 8 selected conditions namely Attention-deficit/hyperactivity disorder (ADHD), Anxiety, Asperger's Syndrome, Bipolar Disorder, Depression, Obsessive-compulsive disorder (OCD), Post-traumatic stress disorder (PTSD) and Schizophrenia.

A. *Mental Health Condition Classifier*

This stage of the presented work deals with classifying the user into one of the eight chosen categories of mental health conditions. For this purpose, a dataset of 7995 Reddit posts is used, which is a combination of around 1000 posts extracted using PRAW (Python Reddit API Wrapper) from each category's Reddit community. This dataset is cleaned during the preprocessing step and all the unavailable posts records are removed. These Reddit posts, at the time of extraction, are the top 1000 posts of their respective subreddits which makes them reliable enough to provide relevance to the training data as well as for responses generated using their comments.

Along with the posts data, corresponding comments are also extracted with their number of upvotes, post-link ID and their text body. In the later stages of the work, these comments are queried from the comments data table using post-link ID as the foreign key, hence mapping each comment to its subreddit post.

In the next step of this stage, the Bidirectional Encoder Representations from Transformers(BERT) base uncased model[12] is used which is a pretrained model trained on a large dataset of English language book corpus. This transformers- based model [13] takes in input ids, token type ids and attention mask [13] which are derived by applying the tokenizer method of this model on the posts body column in the cleaned Reddit posts dataset. The output labels are created as integer values ranging from 0 to 7, each representing a mental health condition. The model is trained on 90% of the dataset, the 10% being the test data, using Adam optimizer, categorical cross entropy loss function and categorical accuracy as the evaluation metrics. Finally, the output of the model is an integer, corresponding to a condition and is fed into the next stage, which is generating responses with respect to the data of this particular condition.

Unlike other works, the proposed approach does not use binary classification as first of all, it achieves an overall accuracy of 93% for category classification and detecting the disease itself is a better approach in case of mental health instead of detecting the presence of disease. This is because, if an input has the textual characteristics of multiple diseases, it is hard to identify the actual dominant illness.

B. *Sentiment-based Response Generation*

This stage deals with generating a response based on four conditions - class, sentiment, sentiment score and relevance. First of all, the class is the mental health condition taken from the output of the previous stage. So, if the associated class is Depression, the Reddit dataset comprising of two data tables - posts and comments, is looked upon and queried. In order to do so, a semantic search based sentence transformers model all-MiniLM-L6-v2 is used. Once the user has briefed about his condition in a few words, it is taken as an initial input to the chatbot. After condition categorization, this input is fed into the semantic similarity model which matches the input to all the posts of the condition's posts dataset. The best match is chosen and the post id is mapped to the comments table's link ids as the dataset is filtered to obtain all the comments associated with the matched post.

In the next step, DistilBERT base uncased finetuned SST- 2 model is used for the sentiment analysis [14] of all the comments in the filtered dataset. This helps to obtain the comments sentiment and a corresponding sentiment score. Since the approach is to provide consoling and motivating responses, all the negative sentiment comments are removed in order to obtain only the positive sentiment comments. Further, in the following step, the filtered comments dataset is sorted based on the positive sentiment score and the number of upvotes each comment has got from the Redditusers. This helps in getting the most relevant response and that too of a positive sentiment. It is important to note that responding with just the most liked comment can go against the primary aim of consolation as its sentiment can be negative and instead of motivating, can even demoralize the patient.

In this way, results for all the other mental health categories can be obtained and human-like responses can be given to every message as these are the replies given by actual humans, the Reddit users. [15] Furthermore, since these responses comprise of enough information required to respond to a particular message from the user, they can be used in the chatbot in stages while rest of the conversations can be handled using conventional conversational flows for chatbots. For example, if the detected mental illness turns out to be OCD for a patient, he can be redirected to a chat flow asking questions and giving responses based on OCD only. These conversations can then be used for extracting data from user responses and on aggregation, can be fed into the response generation module which provides a structural flow to the chatbot as well as informative responses only at the required stages and not making the conversation boring and question- answer based.

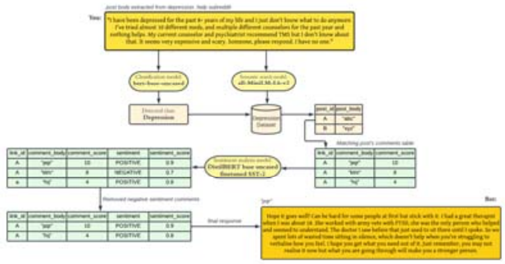

Fig. 2. Complete flow of work presented: a dialogue exchange

TABLE IA COMPARISON TABLE OF PREVIOUS WORKS AND PRESENTED WORK

S.No.	Publication	Mental health conditions	Classification accuracy
1	Falguni	8(emotion labels)	CNN: 75%, RNN and HAN:

	et.al. [5]		75%
2	Kim et.al. [8]	6	75-97% (binary accuracy for each category)
3	Anca et.al. [9]	9	70-81% (binary accuracy for each category)
4	Presented work	8	93% (categorical accuracy)

IV. Evaluation and Results

Upon evaluation of the classification model, the test data is put into play which constitutes the 10% of the Reddit posts dataset. Using categorical accuracy metrics, an accuracy of upto 93% is obtained for categorizing the patient into one of the 8 classes based on his brief about his condition. These results give the required reliability to the response system as the patient is only responded based on his condition. Other than this, this work stands out from other mental health Classification models since it covers more classes than the works done previously and the chosen classes are quite different from one another, which means that the patient does not get same responses in different classes of mental health. A comparative study of the previous works and the proposed work is depicted in the Table I. It can be clearly inferred that this work has more number of categories of mental health and has a sufficiently good enough accuracy in categorizing the patient to one of the eight of these. As illustrated in Figure 2, the initial input gets a relevant response in which the bot is telling its own story to which the user can relate and at the same time get motivated by the phrases used. Hence, the presented approach is able to generatea response to the user input maintaining reliability, relevance and sentiment, all at the same time.

V. Conclusion and Future Work

With the presented work, a chat response system is designed which first asks for an initial brief about the patient's condition and associates his to the corresponding class with an accuracy of upto 93%. On the basis of this assigned condition, a response to the input is generated after querying of Reddit posts and comments dataset based on similarity matching and sentiment analysis.

The work done so far incorporates multiple models for diminishing the limitations of previous chatbots and create a new design for a mental health chatbot. However, as the fields of psychology and mental health are quite vast, such works need regular updates and improvements. For a text-based chatbot, intent based tagging approach can be introduced for a much better interaction. Other than this, although there is a scarcity of mental health dialogue dataset, there is a future scope of creating such corpuses to train conversational models and based the entire chatbot on the basis of psychotherapeutic consultation.

Acknowledgment

We wish to acknowledge Reddit for making its data public for research purposes. We assure the organization, its members and users that we have scraped only its posts and comments text body and have taken no interest in the authors associated with them, hence maintaining the necessary confidentiality and anonymity.

References

[1] G. Browne, "The problem with mental health bots," 10 2022. [Online].Available: https://www.wired.co.uk/article/mental-health-chatbots

[2] G. T. Research, "Therapy chatbots might be worsening your mental health," vol. 10, 2022. [Online]. Available: https://www.medicaldevice-network.com/comment/therapy-chatbots-mental-health

[3] A. Abd-alrazaq, M. Alajlani, A. Alalwan, B. Bewick, P. Gardner, andM. Househ, "An overview of the features of chatbots in mental health: A scoping review," International Journal of Medical Informatics,
vol. 132, p. 103978, 09 2019.

[4] Dhanabalan, S. S., Sitharthan, R., Madurakavi, K., Thirumurugan, A., Rajesh, M., Avaninathan, S. R., & Carrasco, M. F. (2022). Flexible compact system for wearable health monitoring applications. Computers and Electrical Engineering, 102, 108130.

[5] F. Patel, R. Thakore, I. Nandwani, and S. K. Bharti, "Combating depression in students using an intelligent chatbot: A cognitive behavioral therapy," in 2019 IEEE 16th India Council International Conference (INDICON), pp. 1–4, 2019.

[6] N. Boettcher, "Studies of depression and anxiety using reddit as a data source: Scoping review," JMIR Ment Health, vol. 8, no. 11, p. e29487,Nov 2021. [Online]. Available: https://mental.jmir.org/2021/11/e29487

[7] N. S. Kamarudin, G. Beigi, and H. Liu, "A study on mental health discussion through reddit," in 2021 International Conference on Software Engineering Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS- ICOCSIM), pp. 637–643, 2021.

[8] J. Kim, J. Lee, E. Park, and J. Han, "A deep learning model for detecting mental illness from user content on social media," Scientific Reports, vol. 10, 07 2020.

[9] A. Dinu and A.C. Moldovan, "Automatic detection and classification of mental illnesses from general social media texts,", vol. 01, pp. 358–366, 2021.

[10] A. Cohan, B. Desmet, A. Yates, L. Soldaini, S. MacAvaney, andN. Goharian, "Smhd: A large-scale resource for exploring online language usage for multiple mental health conditions," 2018. [Online].Available: https://arxiv.org/abs/1806.05258

[11] Gomathy, V., Janarthanan, K., Al-Turjman, F., Sitharthan, R., Rajesh, M., Vengatesan, K., &Reshma, T. P. (2021). Investigating the spread of coronavirus disease via edge-AI and air pollution correlation. ACM Transactions on Internet Technology, 21(4), 1-10.

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018. [Online]. Available: https://arxiv.org/abs/1810.04805.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones,A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," CoRR, vol. abs/1706.03762, 2017. [Online]. Available: http://arxiv.org/abs/1706.03762

[14] S. Bharathi.Sv and A. Geetha, "Sentiment analysis for effective stock market prediction," International Journal of Intelligent Engineering and Systems, vol. 10, pp. 146–154, 06 2017.

[15] U. Lokala, A. Srivastava, T. G. Dastidar, T. Chakraborty, M. S. Akthar,M. Panahiazar, and A. Sheth, "A computational approach to understand mental health from reddit: Knowledge-aware multitask learning framework," 2022. [Online]. Available: https://arxiv.org/abs/2203.11856