Journal of Computer Science

Original Research Paper

Analysis of Mental Health Counseling Conversation Using Natural Language Processing

[1]Saad Ahmed, [2]Sidrah Khurshid, [3]Muhammad Imran, [4]Muhammad Shoaib Siddiqui, [2]Saman Hina and [5]Munad Ahmed

[1]Department of Computer Science, Iqra University, Karachi, Sindh, Pakistan
[2]Department of Computer Science and Information Technology, NED University of Engineering and Technology Karachi, Sindh, Pakistan
[3]Department of Metallurgical Engineering, NED University of Engineering and Technology Karachi, Sindh, Pakistan
[4]Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah, Saudi Arabia
[5]Department of Research, MSM360, Pakistan

Article history
Received: 13-09-2023
Revised: 30-10-2023
Accepted: 28-12-2023

Corresponding Author:
Saad Ahmed
Department of Computer Science, Iqra University, Karachi, Sindh, Pakistan
Email: saadahmed@iqra.edu.pk

Abstract: One of the most significant public health challenges of our day is mental illness. Despite the benefits of psychotherapy and counseling, our understanding of conducting effective counseling conversations has been limited due to a lack of high-quality data with labeled results. This research presents a quantitative analysis of relatively good-quality data scraped from an online counseling forum. The dataset comprises questions related to various mental illnesses from actual patients and the responses from professional, certified therapists. Through graphical representations, we visualize the correlation between various linguistic aspects of conversations with conversation outcomes. We further apply certain language models, including the pre-trained BERT model, to analyze the quality of therapist responses. The results are then compared to identify effective conversational strategies contributing to improved outcomes. The novelty of this study lies in the mathematical explanations of Language models, making it a valuable resource for readers seeking a deep understanding of machine learning techniques. Additionally, it provides practical implementation guidance for the BERT model, enhancing its usability in real-world scenarios related to mental health challenges.

Keywords: Mental Health, Counseling, Psychotherapy, Depression, Therapist, Chatbot, NLP

Introduction

One of the most rapidly rising public health challenges of our day is mental illness. As a serious public health concern worldwide, mental illness affects 43.6 million individuals (18.1 percent) in the United States alone in any given year (NIH, 2023). WHO reports that depression alone impacts more than 264 million people worldwide (Collaborators, 2018). Moreover, the beginning of the COVID-19 pandemic has resulted in a major economic downturn, negatively affecting many people’s mental health and creating new barriers for those who already suffer from mental illness and substance use disorders (Stuckler *et al*., 2010). According to the John Tung Foundation survey, more than one-fifth of participants were dissatisfied with the changes brought about by the pandemic. Stress (38%), anxiety (23%), nervousness (22%) and panic (9%) were the top four negative emotions experienced by participants, which can lead to long-term psychological disorders such as hopelessness, depression, and anxiety if not treated promptly which may result in suicidal tendencies (Trappey *et al*., 2022; Ye *et al*., 2021).

Counseling and psychotherapy have emerged as helpful treatments for people suffering from mental health problems. However, in such trying times, seeking physical therapy can be challenging. Sometimes the patients are too hesitant, or they are unable to bear the expenses of therapy. The ratio of mental health nurses, psychiatric social workers, therapists, and psychiatrists to patients is one to 10,000, even in wealthy countries (Kislay, 2020).

Consequently, tech firms have developed AI-powered programs designed to serve as the primary mental health care providers for patients without compromising their privacy or anonymity. Programs that are tailored to each individual have

Science Publications

been created to actively monitor patients, provide activities that enhance users' well-being, and be accessible to listen at any time and from any place (Woodward *et al*., 2020). Researchers have begun to investigate Natural Language Processing (NLP) methods for assessing the nature of counseling encounters by researching factors such as mirroring, empathy, and reflective listening to increase their understanding of counseling practice and the conversations between therapists and their patients. NLP is a field that combines linguistics, Artificial Intelligence (AI), and computer science to enable computers to comprehend, analyze, and approximate human speech creation. Woebot, Wysa, Joyable, and Talkspace are a few examples of chatbots that can perform mental health assessments using natural conversation and are available as Android/iOS apps or websites (Dey and Desai, 2022).

In many circumstances, psychotherapy and counseling can effectively address mental health disorders (WHO, 2023). A wealth of information about our behavior, beliefs, mood, and general well-being is available to us thanks to the increasing digitization of our lives (Coppersmith *et al*., 2017). This information can help patients by providing them with some insight into their lives outside of the clinical setting. This research aims to describe a large-scale, quantitative analysis of text-message-based communication between patients and therapists using various NLP tools and techniques. While the lack of large-scale data with labeled outcomes of the conversations has limited our understanding of how to assess the quality of counseling sessions, this research aims to fill that gap. Making use of Natural Language Processing (NLP) techniques and Machine Learning (ML) algorithms, we will evaluate the quality of counseling sessions between a therapist and a patient.

A multitude of research papers from the past six to seven years was thoroughly studied. It appears that NLP has not been widely used in the domain of mental health and associated illnesses. However, the few studies that were conducted have produced significant findings that can be useful in improving the care and treatment of such illnesses.

In one of the studies, a large-scale analysis of counseling conversations was performed by Althoff *et al*. (2016), producing quantitative results. In their research, they investigate anonymous counseling interactions from a non-profit organization that offers free crisis intervention via SMS messaging. As all exchanges between the two dialogue partners are completely observed, Conversation analysis works especially well with text-based counseling talks. Sequence-based conversation models, message grouping, language model comparisons, and word frequency analysis influenced by psycho-linguistics were among the techniques used.

The work produced unique computational discourse analysis tools suitable for large-scale datasets. The findings could contribute to better counselor training and the development of real-time therapy quality monitoring and answer suggestion support tools. In the future, real-time monitoring of counseling quality could be used to validate the findings. In another study conducted by Bertagnolli, counseling data was scraped from www.counselchat.com (Bertagnolli, 2020). They trained a topic classifier using SVM on TF-IDF features. Unfortunately, the model performance on the validation set did not look so good. The limitations of this study included a small dataset and limited access to good therapist-patient interactions/conversations.

Pérez-Rosas *et al*. introduce a new dataset of counseling conversations collected from public web sources (Pérez-Rosas *et al*., 2018). With this dataset, they intend to address the issue of a scarcity of psychotherapy data for NLP applications, as most existing psychotherapy corpora are restricted from public access due to ethical and privacy concerns. From YouTube and Vimeo, they have collected high- and low-quality counseling conversations using specific keywords like “motivational interviewing, good counseling, bad counseling” etc.

Semantic word classes from the LIWC lexicon, semantic word-class scoring by Pérez-Rosas *et al*. (2018), and linguistic cues are among the employed techniques to create a computational model that forecasts the overall quality. With the use of the dataset acquired by Kislay (2020), text-based classifiers that can distinguish between high- and low-quality counseling linguistically and predict the general quality of a counseling conversation can be developed. The gap in the research includes unclear grounds for manual classification of high- and low-quality sessions during the collection of data. In a study conducted during the COVID-19 pandemic, Low *et al*. used the Reddit mental health dataset with posts from 826,961 unique users from 2018-2020 (Low *et al*., 2020). To examine trends in 90 text-derived features, including sentiment analysis, personal pronouns, and semantic categories, they used regression. To identify issues on Reddit before and during the outbreak, they employed unsupervised techniques like topic modeling and unsupervised clustering. The study identified at-risk individuals, showed patterns in the linguistic manifestations of various mental health disorders, and showed how concerns were distributed throughout Reddit.

Tewari *et al*. have conducted a comprehensive survey of Mental Health Chatbots that use NLP (Tewari *et al*., 2021). They address the usage of NLP in psychotherapy and conduct a broad study of existing systems by comparing chatbot responses to a set of predetermined user inputs pertaining to well-being and mental health concerns. The overall methodology used in the development of such chatbots involves fundamental NLP techniques such as word embeddings, sentiment analysis, and models such as the sequence-to-sequence model and attention mechanism. The study was based on the dataset obtained from the Cornell movie dialog corpus and open subtitles corpus. The use of cutting-edge technology Natural language technologies combined with psychotherapy can result in tools that can fill gaps in the delivery of mental health care to a large extent. Before approving any clinical use, they must be studied and attempted on a broad scale and viable outcomes must be documented.

Materials and Methods

Dataset

The data for this research was scraped from Counsel Chat, an online free-of-cost platform where anyone suffering from mental illnesses can post a question based on their actual condition and receive a relevant response from a certified therapist or counselor located in a nearby area. This website provides a safe space for real mental health patients to connect with a professional counselor anywhere, anytime. The data acquired does not include an entire conversation between the patients and their therapists; rather, it just includes individual talk-turns between them. Most of the questions on the website were similar in text or nature. Therefore, some counselors used a single response for many similar queries. We did not clean the dataset for such duplications.

The site has 307 therapist members, most of whom are on the West Coast of the United States (Washington, Oregon, California). Ph.D. grade psychologists, social workers, and licensed mental health counselors are among those with this certification. Another interesting fact about this online forum is that it allows the community members (including the one who posted the question) to "upvote" or like a therapist's response that they find enlightening or helpful. The focus area of our research is the prediction of these upvotes through models trained on the dataset. Our dataset is of relatively good-quality responses from certified counselors to mental health queries from actual patients. The final CSV contains eight columns/attributes that are described in Table 1. Additionally, the dataset was randomly split into "train" and "test" labels in a 14:86 ratio (Table 2 and Fig. 1).

Data Analysis

After cleaning the data for redundancies and irrelevant information, we ran a comprehensive analysis on it to understand the meaningful trends and patterns in disguise. We have visualized the findings through multiple charts and graphs.

Table 1: The columns/attributes of our dataset, along with their descriptions

Column/Attribute	Description
Question ID	A unique question identifier
Question Title	Question title
Question text	Question body
Topic	of query
Therapist info	Name and field of specialization of each therapist
Answer text	Therapist's response
Upvotes	Number of upvotes on every answer text
Split	"test" or "train" label for each sample

Table 2: The picot table of the "split" column quantifies the distribution

Split	Count of split
Train	1839
Test	290
Total	2129

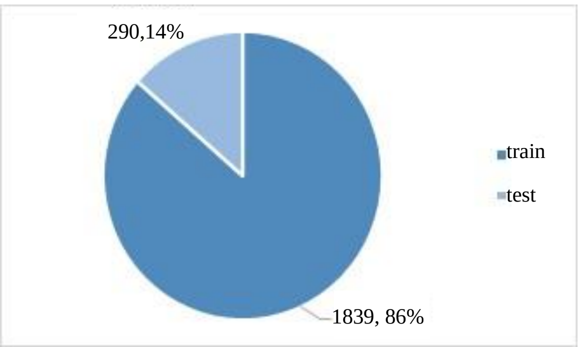

Fig. 1: A pie chart visualizing the "train" and "test" distribution

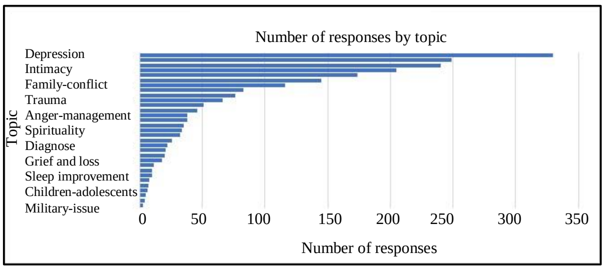

Fig. 2: The number of responses per topic

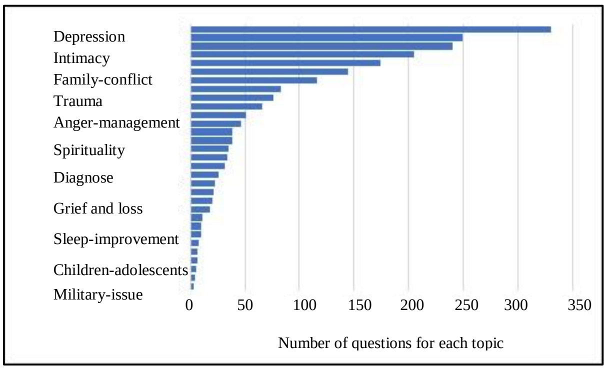

Fig. 3: Frequency distribution of topics

There are 31 topics on the website, out of which "depression" is the most discussed topic, with answers ranging from 317. It appears that the least popular topic on the forum is related to "military issues," with only three answered queries (Fig. 2).

The topic-wise distribution of patient questions and therapist responses can be seen in Figs. 3-4. We have also represented the distribution of upvotes (likes received by a therapist response) as a logarithmic graph in Fig. 5.

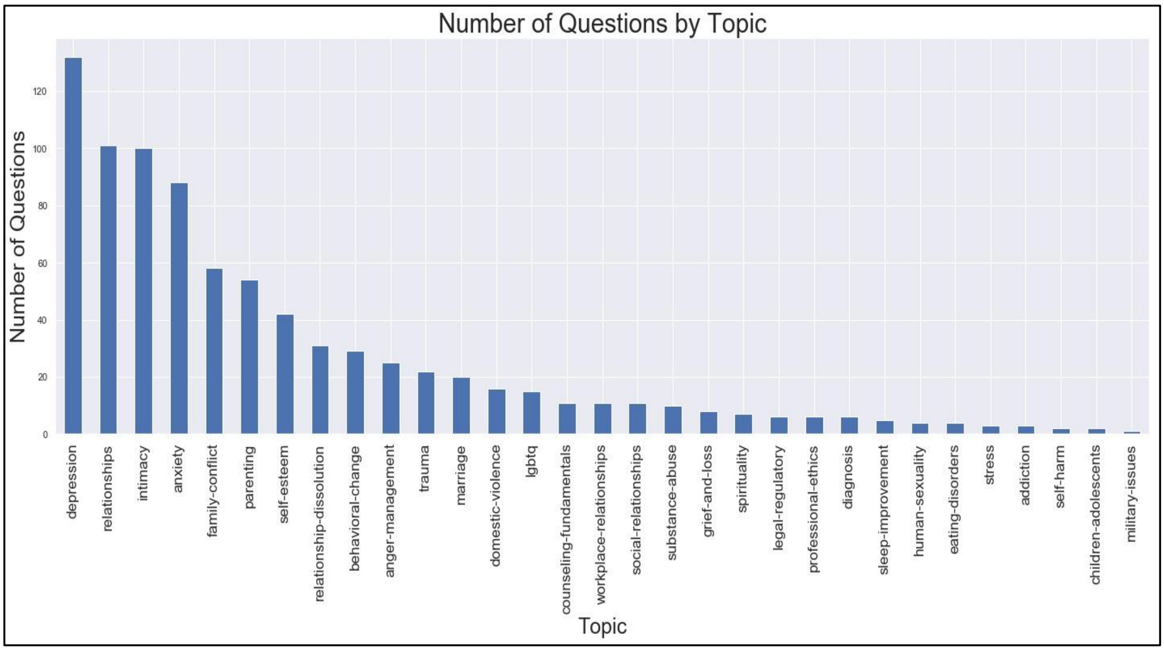

Fig. 4: The number of questions per topic

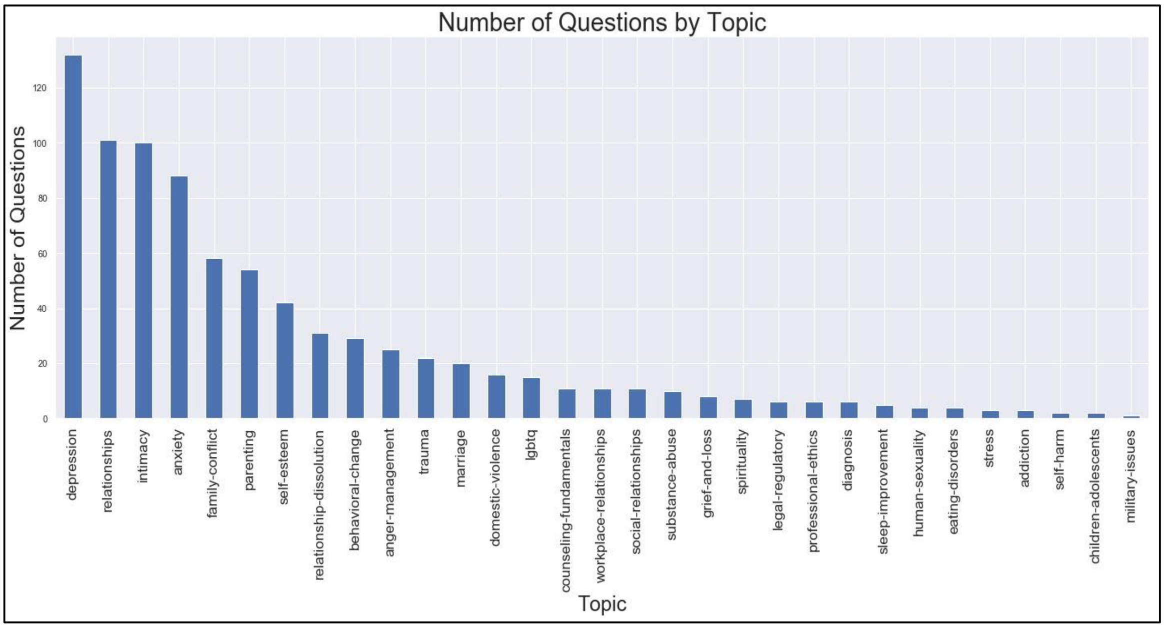

Fig. 5: Distribution of upvotes on a logarithmic scale

TF-IDF

TF-IDF, or Term Frequency (TF) Inverse Dense Frequency (IDF), is a method for determining the meaning of word-based statements. It makes up for the shortcomings of the bag of words method, which helps categorize texts or help machines understand words represented by numbers (Madan, 2019). It is a score that the machine records in which it analyses the words used in a sentence and compares their usage to the words used in the overall document. In other words, it's a value that emphasizes each word's importance throughout the text. In other words, it's a value that emphasizes each word's importance throughout the text. TF-IDF is computed as:

TF = Occurrence of a word in a document/number of words in a document

IDF = Log number of documents/number of documents that include the word

Language Models

Now comes the part where we have to determine the best-fit language models for our dataset. We select three algorithms: Na¨ıve Bayes, SVM, and BERT, to predict the upvotes for therapist responses in the test dataset. After executing them, we will analyze the results and evaluate their performances in comparison with each other. These models are discussed below.

Naive Bayes

A class of probabilistic algorithms known as the "naive" Bayes classifier algorithms uses the Bayes theorem under the "naive" assumption that any two features are conditionally independent of one another. The Bayes theorem calculates the probability *P*(*c*-*x*), where, *x* is the given instance that needs to be identified and expresses some particular qualities and c is the class of likely outcomes:

$$P(c\mid x)=\frac{P(x\mid c)=P(c)}{P(x)} \tag{1}$$

Natural Language Processing (NLP) applications typically employ Naive Bayes to predict tags of texts. For a given text, the model calculates the likelihood of each tag and then reports the tag with the highest likelihood (AMNBNLPP, 2023).

Support Vector Machine (SVM)

Applications for regression and classification can make use of the supervised machine learning method known as the support vector machine. Nonetheless, the majority of its uses are in categorization applications. A large-margin classifier is a vector space-based machine learning technique known as a Support Vector Machine (SVM). Its objective is to find a judgment border between two classes that are as far away from every point in the training data as feasible, potentially eliminating specific points from the analysis as noise or outliers. Every data example is represented graphically in an n-dimensional space (where n is the number of features), with a given coordinate representing the value of each feature. Next, classification is achieved by identifying the hyper-plane that most effectively separates the two groups (CWKC, 2009; Ray, 2023).

BERT (Bidirectional Encoder Representations from Transformers)

In 2018, google research researchers introduced an open-source natural language processing model called Bidirectional Encoder Representations from Transformers (BERT). By building context with the surrounding information, BERT is designed to help computers grasp meaningless words in the text. The BERT framework can be further improved by using question-and-answer datasets after it has been pre-trained on Wikipedia text. Its core is a deep learning model called Transformers, wherein every output element is connected to every input element and weightings between them are dynamically established according to these relationships. BERT's bidirectional capacity has been pre-trained using two distinct but related NLP tasks: Predicting the next sentence and Making Use of Masked Language Models (BERT Model, 2024; Lutkevich, 2020).

Results and Discussion

For evaluation of the performance of our models, we employed the following metrics:

- Accuracy: The percentage of correct predictions made by the model
- Precision: The proportion of relevant occurrences among the retrieved examples
- Recall: The percentage of all relevant instances that were retrieved
- Confusion matrix: The number of correct and wrong predictions made by each class is summarized in the form of a table

Tables 3-5 summarize and quantify the performance of our models. It can be observed that Naive Bayes and SVM both perform considerably well with F1-scores of 0.87-0.76 and 0.88-0.76, relatively. However, our BERT model had a comparatively lower F1-Score and Accuracy for either class, indicating that it did not fit well on our dataset. Training a BERT model is computationally intensive and requires a lot of memory to run, so training or fine-tuning a BERT model can take a long time, especially when working with large datasets like ours. Using a smaller training dataset may result in a faster execution but can impact the accuracy of the results. By using large datasets and more powerful computers, it is expected that BERT's Model performance will improve significantly. which we will be working on in our future work. Our findings contribute to actionable strategies linked with successful therapy. In addition, our trained models can be effectively leveraged to train a real-time chatbot that is capable of generating reasonable responses to mental health queries.

Table 3: Classification matrix of the Naive Bayes model

Class	F1-score	Precision	Recall	Support	Accuracy
0	0.87	0.82	0.92	72	0.83
1	0.76	0.84	0.69	45	0.81

Table 4: Classification matrix of the SVM model

Class	F1-score	Precision	Recall	Support	Accuracy
0	0.88	0.57	0.85	86	0.83
1	0.76	0.63	0.79	77	0.82

Table 5: Classification matrix of the BERT model

Class	F1-score	Precision	Recall	Support	Accuracy
0	0.65	0.57	0.76	87	0.59
1	0.50	0.63	0.42	86	0.59

Conclusion

The subject of mental health agents and chatbots is expanding rapidly. The use of cutting-edge Natural language technologies in conjunction with psychotherapy can result in tools that can help to cover gaps in the delivery of mental health care. However, they must be studied and tried on a broad scale and acceptable outcomes must be proven before any clinical use is approved.

The inclusion of visualizations illustrates the concept of upvotes for psychotherapy-related questions asked by counselors which help improve the mental conditions faced by patients. This adds another layer of clarity to the explanation, making it more accessible for readers to grasp this important aspect of our research work.

We hope that our study will inspire future generations of crisis intervention tools and counselors. Our findings may aid in the improvement of counselor training and the development of real-time counseling quality monitoring and answer suggestion assistance technologies.

Acknowledgment

We would like to show our gratitude to NED University Karachi, Pakistan for allowing us to use computer labs during this study. This study was performed according to the university research ethics protocols. In addition to this, we also wish to thank all staff members who supported us during this research study.

Funding Information

The authors have not received any financial support or funding to report.

Data Availability Statement

The text data can be made available on request. Since patients have a right to privacy, identifying information (including patients' images, names, initials, or hospital numbers) cannot be shared publicly.

Author's Contributions

Saad Ahmed: Provided essential guidance and oversight throughout the project. Designed the research plan and organized the study, coordinated the data analysis, contributed to the written of the manuscript.

Sidrah Khurshid: Designed the research planed and organized the study. Participated in all experiments, coordinated the data analysis, and contributed to the written of the manuscript.

Muhammad Imran and Munad Ahmed: Participated in all experiments, coordinated the data analysis, and contributed to the writing of the manuscript.

Muhammad Shoaib Siddiqui: Coordinated the data analysis and contributed to the writing of the manuscript. managed the administrative aspects of the project.

Saman Hina: Designed the research plan and organized the study contributed to the writing of the manuscript.

Ethics

This study is an original research work, and the lead author confirms that all co-authors have reviewed and endorsed the manuscript without any ethical concerns.

Conflicts of Interest

The authors declare no conflict of interest.

References

Althoff, T., Clark, K., & Leskovec, J. (2016). Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics*, *4*, 463-476. https://doi.org/10.1162/tacla00111

AMNBNLPP. (2023). Applying Multinomial Naïve Bayes to NLP Problems. https://www.geeksforgeeks.org/applying-multinomial-naive-bayes-to-nlp-problems/.

Bertagnolli, N. (2020). Counsel chat: Bootstrapping high-quality therapy data. *by towardsdatascience. com. URL: https://towardsdatascience. com/c ounsel-chat-bootstrapping-high-quality-therapy-data-971b 419f33da (cit. on pp. 72, 96)*. https://huggingface.co/datasets/nbertagnolli/counsel-chat

BERT Model. (2024). Explanation of BERT Model-NLP. https://www.geeksforgeeks.org/explanation-of-bert-model-nlp/

Coppersmith, G., Hilland, C., Frieder, O., & Leary, R. (2017, February). Scalable mental health analysis in the clinical whitespace via natural language processing. In *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)* (pp. 393-396). IEEE. https://doi.org/10.1109/BHI.2017.7897288

Collaborators, G. B. D. (2018). Global, regional and national incidence, prevalence years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. https://hdl.handle.net/2381/45609

CWKC. (2009). Choosing what kind of classifier to use. https://nlp.stanford.edu/IR-book/html/htmledition/choosing-what-kind-of-classifier-to-use-1.html

Dey, J., & Desai, D. (2022). NLP-Based Approach for Classification of Mental Health Issues using LSTM and GloVe Embeddings. https://ijarsct.co.in/Paper2296.pdf

Kislay, K. (2020, October 19). Chatbots in mental health. friendly but not too friendly. *Analytics India Magazine*. https://analyticsindiamag.com/chatbots-in-mental-health-friendly-but-not-too-friendly/

Low, D. M., Rumker, L., Talkar, T., Torous, J., Cecchi, G., & Ghosh, S. S. (2020). Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19: Observational study. *Journal of Medical Internet Research*, *22*(10), e22635. https://doi.org/10.2196/22635

Lutkevich, B. (2020). BERT language model. https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model

Madan, R. (2019, November 27). TF-IDF/Term Frequency Technique: Easiest explanation for Text classification in NLP using Python (Chatbot training on words). https://medium.com/analytics-vidhya/tf-idf-term-frequency-technique-easiest-explanation-for-text-classification-in-nlp-with-code-8ca3912e58c3

NIH. (2023). Mental Illness. *National Institute of Mental Health*. https://www.nimh.nih.gov/health/statistics/mental-illness

Pérez-Rosas, V., Sun, X., Li, C., Wang, Y., Resnicow, K., & Mihalcea, R. (2018, May). Analyzing the quality of counseling conversations: The tell-tale signs of high-quality counseling. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. https://aclanthology.org/L18-1591.pdf

Ray, S. (2023). Learn How to Use Support Vector Machines (SVM) for Data Science. https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/

Stuckler, D., Basu, S., & McDaid, D. (2010). Depression amidst depression: Mental health effect of the ongoing recession. http://eprints.lse.ac.uk/id/eprint/61816

Trappey, A. J., Lin, A. P., Hsu, K. Y., Trappey, C. V., & Tu, K. L. (2022). Development of an empathy-centric counseling chatbot system capable of sentimental dialogue analysis. *Processes*, *10*(5), 930. https://doi.org/10.3390/pr10050930

Tewari, A., Chhabria, A., Khalsa, A. S., Chaudhary, S., & Kanal, H. (2021, April). A survey of mental health chatbots using NLP. In *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*. https://doi.org/10.2139/ssrn.3833914

Woodward, K., Kanjo, E., Brown, D. J., McGinnity, T. M., Inkster, B., Macintyre, D. J., & Tsanas, A. (2020). Beyond mobile apps: A survey of technologies for mental well-being. *IEEE Transactions on Affective Computing*, *13*(3), 1216-1235. https://doi.org/10.1109/TAFFC.2020.3015018

WHO. (2023, March 31). Depressive disorder (depression). *World Health Organisation*. https://www.who.int/news-room/fact-sheets/detail/depression

Ye, Y. X., Dai, Y. J., Xie, B. T., & Jian, D. K. (2021). Survey of Life Changes and Mood during the COVID-19 Epidemic.