International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Prediction of a Novel Rule-Based Chatbot Approach (RCA) using Natural Language Processing Techniques

Smita Rath[1*], Adyasha Pattanayak[1], Sashikanta Tripathy[1], Sushree Bibhuprada B. Priyadarshini[2], Anjela Tripathy[3], Shipra Tanvi [4]

Submitted: 28/04/2023 **Revised**: 27/06/2023 **Accepted**: 06/07/2023

Abstract: The management of a mental health of a person has been made possible in recent years by a variety of virtual assistants. This paper uses a Rule-based Chatbot to provide a quick description of a mental state of a human. An AI-powered computer programme known as a Chatbot can mimic human interaction via voice commands, text dialogues, or a combination of the two. This AI function, also referred to as a Rule-based Chatbot, can be included in and used with well-known messaging platforms. The natural language toolkit is used in the implementation of this Chatbot. Natural language processing (NLP), a branch of AI, is used by Chatbot powered by AI to improve the user experience. These NLP-based Chatbot, also known as virtual agents or intelligent virtual assistants, support human agents by managing time-consuming and repetitive exchanges. Human agents can now concentrate on instances that require their knowledge because they are more complex. NLP-based Chatbot are intelligent in that they can understand speech patterns, text structures, and language semantics. Because of this, they can analyze and derive meaning from massive amounts of unstructured data. A Chatbot has the capacity to perceive small differences in different languages which is improved by cross-linguistic comprehension of morphemes of Natural Language Processing(NLP). In addition, NLP gives Chatbot the capacity to comprehend unused words, adjust to changing abbreviations, and identify emotions through sentiment analysis, simulating human-like comprehension. This model depicts a Chatbot that responds to user inquiries in a brief and straightforward manner. Although conversational AI is frequently associated with Chatbot, not all Chatbot use AI. In general, rules-based Chatbot are used to describe Chatbot that do not use AI. These Chatbot direct users towards particular behaviours using established rules and decision trees. These scripts and rules are predetermined, and any changes call for manual action from the organization. One of the early alternative therapies explored was cognitive behavioural therapy. The user must, however, attend face-to-face counseling, which could last nine to twelve years. The first virtual assistant designed in 1950 to read a person's mental state was the Turing test. The main goal of this proposed system is to overcome the difficulties in determining the appropriate responses to the user's inquiries. Performance metrics such as Similarity Bow and Term frequency-inverse document frequency (TF-IDF) score is used to evaluate the similarity between the exact answer and predicted answer.

Keywords*: *Chatbot, Machine Learning (ML), Mental Health, Natural Language Processing (NLP), Therapy

1. Introduction

The creation of conversational Chatbot, which have already replaced classic Chatbot, has become more popular in recent years. A Chatbot is a computer programme that communicates with people and attends to their needs.A Chatbot responds to a user's query and occasionally has the ability to do other tasks as well. The Chatbot is a piece of software used to have natural language discussions with drug users. Originally, Chatbot evolved for recreation by emulating mortal communication. It embodies knowledge by employing AI. People are less fearful of being judged and less concerned with creating a good impression when speaking with a machine than they are when speaking with a human.

**1,2,3,4,Department of Computer Science & Information Technology, Institute of Technical Education & Research*
Siksha 'O' Anusandhan Deemed to Be University, Bhubaneswar, Odisha-751030, India
ORCID ID : 0000-0003-0547-6609
** Corresponding Author Email: smitarath@soa.ac.in*

Chatbot can replace humans in routine tasks that include answering queries and delivering effective responses, e.g., E-Help divisions, client care, and virtual sidekicks like Cortana in Windows and Siri in macOS. The educational system, family and friend expectations, and a lack of communication about stress among peers or superiors all contribute to the high levels of stress or poor mental health experienced by students aged 15 to 25 studying in Indian institutions. It requires a cooperative effort involving both manual interpretation and artificial intelligence methods such as NLP, NLU, machine literacy, deep literacy, and more. There are many challenges to developing a Chatbot that cannot be ignored.

- **Cold User Interaction:** Occasionally, the Chatbot exchanges may feel a bit scripted or robotic, which is a result of a lack of personalization. This can be answered by using a mortal voice rather than a voice of a robot, using emojis while responding, and using non-professional language to make it easy to understand.

- **Lack of sensitivity:** Due to AI limitations, the lack of feelings in Chatbot is an added problem. Chatbot are programmed to respond to certain keywords or expressions, but they are not always able to comprehend the complications of mortal feelings. They warrant empathy and can respond in a robotic or indifferent manner. Chatbot inventors must incorporate empathy into their Chatbot to overcome this challenge.
- **Lesser responses:** This challenge can lead to the wrong user experience as the inquiry is not properly answered. This can be answered by focusing on perfecting the Chatbot's NLP capabilities.
- **Absence of involvement:** Bots are designed in such a way that they cannot respond to commands outside of the sequence programmed. This results in boring and repetitious exchanges. This can be answered by programming the bots in such a way that they learn from the continued discussion with Stoner.

A functional Chatbot was created via a variety of parameters, some of which are described here.

- **Scalability:** A Chatbot is scalable if it receives a large number of customer inquiries and responds efficiently. A scalable Chatbot will work in any setting.
- **Turing Test:** The Turing test was devised in 1950 by Alan Turing to find out if a machine could display intellectual gestures like those of a person. A computer satisfies the Turing test if the tester fails to distinguish between the responses of humans and those of machines.
- **Interoperability:** The ability of a system to consume and exchange information is known as interoperability. Drug users should be able to switch between various channels easily owing to an interoperable Chatbot.
- **Speed:** The amount of time it takes a Chatbot to produce a helpful answer to a stoner's question is its speed. Smart Chatbot should be able to deliver timely interventions.

2. Literature Review

Although the number of human-to-machine service interactions is growing quickly, organizations must take into account the dynamics of consumer acceptance of new technologies to ensure their long-term viability and development. The majority of Chatbot advantages are efficiency-related, such as cheaper expenses for the business, shorter customer wait times, and alignment with customer preferences for digital rather than voice-based communications.

Chatbot user satisfaction is uneven despite this. Due to poorly interpreted queries, irrelevant answers, and inadequate interaction with human service agents, customers may not be able to get the required responses. A Chatbot (also known as a chatterbot) is software that converses with human users; it serves as a virtual assistant and may respond to a limited number of user questions by giving the right responses. Some of the best-known Chatbot are those created by large corporations, such as Apple Siri, Microsoft Cortana, and IBM Watson. These are but a few of the most often used systems. Numerous lesser-known Chatbot exist that are more pertinent to research and their applications; some of these will be covered in the following chapter. Several researchers have designed their Chatbot to deal with various scenarios, as discussed below.

Dokukinaa and Gumanova provide comprehensive details about the rise of Chatbot in today's era. Lives, allowing multitasking and greatly facilitating and improving the work of ESL educators. On a learner's mobile device, Chatbot are accessible every hour of the day, every day of the week. They offer instruction at convenient times and locations in manageable portions or short pieces, which fits the busy lives of contemporary students perfectly. They also patiently repeat the content numerous times until it is comprehended, retained, and put into practise, check the student's understanding, and offer immediate feedback [1]. In order to improve healthcare communication and support patient interactions, AI-powered health Chatbot are being developed. Khadijaa, Zahraa, and Naceur give us an overview of their general architecture. By utilizing artificial intelligence to enhance patient support, streamline communication, and increase accessibility to healthcare, AI-powered health Chatbot provide a novel solution [2].

An intelligent knowledge-based Chatbot was created by Chan and Liang to understand and effectively respond to user inquiries while providing a natural conversational experience. It does this by utilising natural language processing and machine learning techniques. The proposed system makes use of a number of cutting-edge technologies, such as web crawling, NLP, knowledge bases, and AI. A prototype system is constructed in the context of this study. The evaluation of the system prototype produced positive results that confirm the system's efficacy [3]. As given in the study by Shumanov and Johnson[4], which looks at how businesses and government organisations use intelligent Chatbot while noting varying levels of customer satisfaction, It explores the effects of using language to match the Chatbot's personality with the consumer's personality and shows how personalised interactions increase consumer engagement and purchase outcomes in situations involving social gain. The study emphasises how personality-based customization of Chatbot responses has the potential to improve human-computer interactions [4]. The study by

Ogawa, Oyama, Morito, and Kobayashi examines the use of a Chatbot powered by artificial intelligence to enhance smiles and speech in people with Parkinson's disease (PD) during telemedicine consultations. It looks at the viability and effectiveness of the Chatbot intervention and explores the possibility that motor symptoms, cognition, and mood can all be predicted using only objective facial and speech characteristics. The results suggest that Chatbot conversations significantly affect speech and smile characteristics, which are linked to motor symptoms, cognition, and mood in PD.

The study emphasises how AI-based Chatbot could improve PD telemedicine evaluation and monitoring [5]. Even though Chatbot have the potential to provide organisational support for studies and exams, Zahour, Benlahmar, and Eddaoui help us highlight the scant attention that has been paid to them in the educational sector. In order to identify the dominant personality type of students entering the job market, the article introduces a new Chatbot focused on educational and professional guidance, incorporating John Holland's theory and the RIASEC questionnaire. To close the gap between Chatbot technology and academia, this new research field was established [6]. Sperl in his paper[7] outlines a novel framework for integrating both tangible and intangible cultural objects into a single data model to assist travellers, is available online.

The framework makes use of a conversational agent based on the Seq2Seq model and a microservice architecture. The popularity of Chatbot development and the shift to conversational Chatbot with the assistance of Jwala, Sirisha, and Padma Computer programmes called Chatbot are made to communicate with and assist people. Users can ask questions and get answers from them, and they can even do things. Developing Chatbot was difficult in the beginning, but with the availability of development platforms and source code, it has gotten simpler. Natural Language Processing (NLP) or deep learning techniques can be used to create Chatbot, with deep learning Chatbot needing a lot of training data. The goal of the paper is to examine various methods for creating Chatbot, classify them, and talk about metrics for measuring their effectiveness [8]. Alekseev, Hagalova, and Sokolova conducted research on the effectiveness of Chatbot for individualised student communication in the classroom. Artificial intelligence-powered Chatbot were created to automate educational procedures, interacting with users through the cross-platform Telegram messenger. A graphical application interface for Python was created using Qt5 libraries to make the dataset creation process easier. A classification accuracy of 97 percent was attained using a variety of vectorizers and machine learning models to ascertain user intentions [9]. In order to provide healthcare services to underserved and rural populations in India, Gadge, Deokar, Kachare, and Madane set out to develop a conversational AI-powered Chatbot for medical diagnostics. The Chatbot communicates with the patient via a web-based user interface and uses natural language processing to comprehend their symptoms. The system aims to deliver precise diagnoses, suggest suitable medications, and offer necessary precautions by analysing the input and producing human-like responses.

The project uses the Python NLTK (Natural Language Toolkit) module, which enables symbolic and statistical natural language processing and enhances the Chatbot's conversational capabilities [10]. Konar introduced us to the field of sentiment analysis, which uses computer software to decipher human emotions expressed in writing. Opinion mining is the term for this process of classifying sentiments as either positive, negative, or neutral. Together, natural language processing (NLP) and machine learning (ML) solve data-related issues and are regarded as essential components of artificial intelligence (AI). While ML assists in the analysis of vast amounts of data, NLP focuses on how computers and human language interact. This study compares how well lexicon-based and machine-learning approaches perform sentiment analysis when applied to Chatbot. Python is used to create the Chatbot, including one that categorises movie reviews and another one called "DocBot" that informs users about kidney disease [11].

A Chatbot using machine learning is being created as part of the Palandurkar, Shaikh, Shewale, and Raut project to help students with questions about their classes and their schedules. To improve its precision and accuracy over time, the Chatbot makes use of chat history. Information on courses, professors, admissions, college facilities, and frequently asked questions is available under the College Enquiry option. The e-learning option focuses on providing a single correct response to questions related to the IT branch syllabus in order to prevent confusion brought on by search engine results. Future work entails expanding the Chatbot's training data to include all courses in all college departments [12]. Software provided by Reddy Karri and Kumar enables communication with clients in natural language. The objective has been to develop Chatbot that can pass the Turing test, which involves tricking users into believing they are speaking with real people. The ELIZA Chatbot was first introduced in 1966, and since then, there have been efforts to develop other Chatbot using a variety of methodologies and technologies. In order to analyse speech input and produce responses that resemble human speech, the Python NLTK module is frequently used. The demand for Chatbot has increased as a result of speech-based search engines and virtual assistants like Siri, Cortana, Alexa, Google Assistant, and Microsoft Cortana. They are frequently used in the business sector to automate client services and

minimize human effort. With the goal of giving accurate and pertinent answers, Chatbot play a part in information gathering within dialogue systems. This study compares various Chatbot technologies and covers Chatbot system design and implementation [13].

According to Kandpal, Jasnan, Raut, and Bhorge, new technologies like Chatbot or virtual assistants are emerging as machine learning and AI demand rise. Chatbot have developed from menu/button-based to keyword-based to contextual-based approaches, utilising machine learning and AI techniques to provide more accurate answers to user queries. A contextual-based Chatbot model's operation is covered in this paper, along with applications, related research in the area, difficulties, and potential applications. The model uses neural networks and pertinent software packages to combine deep learning and natural language processing techniques for better outcomes. By assisting with appointment scheduling and providing predictive diagnosis, among other features, Chatbot in the healthcare sector have the potential to completely change how patients interact with medical professionals [14].

The use of natural language in information systems, according to Reuterfelt, Carrera, Iglesias, Araque, Rada, and Muoz, can improve student interaction and the learning process. This investigation focuses on how cognitive computing is used in blended learning settings. It is suggested that an architecture be developed for a modular cognitive agent that is tailored for social interaction and answering pedagogical questions. For students learning data science and machine learning, the system acts as a personal assistant. Natural language processing algorithms and machine learning models are combined in a user interface that resembles a human. An experiment has been used to verify the system's efficacy [15].

3. Proposed Rule-Based Chatbot Approach (RCA)

In recent years, various virtual assistants have come into existence to handle the mental health of a person. This paper gives a brief description of the mental state of a person with the help of a rule-based Chatbot. This Chatbot is implemented using the natural language toolkit. This model represents a quick and simple conversation Chatbot that provides answers to the user's queries. The use of cognitive behavioural therapy was one of the alternative therapies used earlier. But the limitation was that the user has to attend face-to-face therapy, which may last for nine to twelve years. The Turing test was the first virtual assistant developed in 1950 to read the mental health of a person. In this proposed system, the primary focus lies on addressing the challenges associated with identifying the correct answers to the user query.

The core approach for building this Chatbot involves the utilisation and importation of essential libraries. The Natural Language Toolkit (NLTK) is employed, offering a comprehensive set of text processing libraries. Additionally, SKlearn 0.0, a Python library with a variety of unsupervised and supervised learning algorithms, is leveraged. NumPy, Pandas, and regular expression libraries are also utilised, providing efficient tools for machine learning and statistical modelling tasks, including classification, regression, clustering, and feature reduction. This combination of resources aims to develop a robust Chatbot solution capable of effectively analysing text message sentiment and identifying any lewd or vulgar content. Users will be promptly warned if their message contains inappropriate language, enabling them to make the necessary revisions before sending it.

3.1 Data Acquisition

In this project, questionnaires were created using Google Forms, a web-based application for online surveys, and shared with friends and relatives, who provided responses. This approach generated a dataset with questions and answers from a diverse group. The collected data can be exported in different formats for further analysis and examination, enabling the use of statistical tools or machine learning algorithms to gain insights and extract valuable information. Google Forms proved to be a convenient and efficient method for collecting valuable data that aligned with the project's objectives. The whole set of questions and answers is collected in the above-mentioned process and stored in a CSV file. This file is then split into training and testing in the ratio of 8:2.

3.2 Data Pre-processing

The objective of this project is to process a dataset stored in a CSV file by identifying and removing missing values. To achieve this, the project utilises the isnull() function, commonly available in data analysis libraries like Pandas. This function allows for the detection of null or missing values within the dataset. Missing values can occur due to various reasons, such as incomplete data collection or data entry errors. They can hinder accurate analysis and modelling if not handled appropriately. Therefore, it is crucial to identify and address these missing values. Once the missing values are identified using the isnull() function, the project proceeds to remove them from the dataset. This can involve either removing entire rows or columns containing missing values or employing strategies like imputation to fill in the missing values with appropriate estimates.

After the missing values have been handled, the dataset undergoes a process of word splitting. This step suggests that the dataset likely contains textual data. By splitting the dataset into individual words, it becomes easier to perform word-level analysis, enabling tasks such as sentiment

analysis, text classification, or information extraction. Additionally, the project aims to recognise similar words within the dataset. Techniques like stemming, lemmatization, or word embeddings can be applied to identify words with similar semantic meanings or contextual relationships. This recognition of similar words can help in various applications, including data grouping, pattern identification, or clustering of similar data points. Overall, this project employs the isnull() function to identify and remove missing values from a CSV dataset. It then splits the dataset into words and recognises similar words to enable subsequent analysis tasks. The specific implementation details, programming language, and libraries used may vary depending on the project's requirements and the characteristics of the dataset.

3.3 Model Training

NLTK also offers access to diverse corpora and lexical resources, allowing you to train and evaluate NLP models. It includes pre-trained models for part-of-speech tagging and named entity recognition, making it easier to get started with NLP tasks[1]. One of the essential features of NLTK is its ease of use. It provides intuitive interfaces and well-documented APIs, enabling developers to quickly implement NLP functionality in their applications. NLTK has gained popularity in both academia and industry due to its extensive capabilities and active community support. It serves as a valuable tool for researchers, educators, and developers working with text data and NLP applications. Overall, NLTK is a powerful and versatile library for NLP tasks in Python. Its wide range of functionalities, access to corpora, and user-friendly design make it a go-to choice for anyone working with text analysis and NLP. The flowchart in Fig. 1 provides a step by step procedure to train the model.

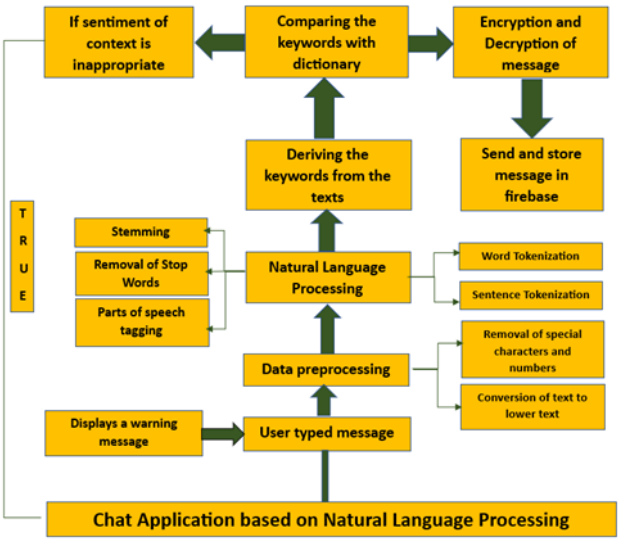

Fig 1. Flowchart of the Proposed Rule-based Chatbot Approach (RCA)to deal with the mental health of a person

3.4 Model Evaluation

3.4.1 . Tokenization

It is the process of splitting a large amount of text into smaller parts called tokens. Natural Language Processing (NLP) is used in applications like text classification, intelligent Chatbot, sentiment analysis, language translation, and more. Understanding the patterns within the text becomes crucial for achieving these goals. Tokens play a vital role in discovering such patterns and are considered a foundational step for tasks like stemming and lemmatization[16]. For now, it is important to focus on the concept of tokenization and its significance in textual data cleaning using NLP.

3.4.2. Stemming

It is the process of generating morphological variations of a root or base word. Stemming programmes are often referred to as stemming algorithms or stemmers. A stemmer reduces words like "chocolates," "chocolatey," and "choco" to the root word "chocolate," while words like "retrieval," "retrieved," and "retrieves" are reduced to the stem "retrieve." Stemming helps reduce words to their basic form, simplify text analysis, and improve computational efficiency in various NLP tasks.

3.4.3. Lemmatization

Lemmatization is the process of grouping together different inflected forms of a word so that they can be analysed as a single item. Unlike stemming, lemmatization considers the context of words, linking them to their base or dictionary form based on their meaning. Text pre-processing encompasses both stemming and lemmatization, but these terms are often confused. However, lemmatization is preferred over stemming because it performs a more sophisticated morphological analysis of words, resulting in better accuracy and context preservation. Lemmatization aims to bring words with similar meanings together, facilitating more accurate analysis and language understanding. Here is the output of the programme Lemmatization Text. The process of lemmatized text is as illustrated in Table 1.

Table 1. Process of lemmatized text

Question ID	Questions	Answers
1590140	What does it mean to have a mental illness?	\n Mental illness refers to conditions that aff….
2110618	Who does mental illness affect?	Mental illness can impact individuals of any a…
9434130	What are some warning signs of mental i….	Some warning signs of mental illness include…
7657263	Can people with mental illness recover?	yes, people with mental illness can recover
1619387	What should I do if I know who appears…	Encourage them to seek professional help and...

3.4.4. Punctuation

Removing punctuation from raw text is a pre-processing technique used to clean and prepare the text for machine learning modeling. Once the text is cleaned, each character and word are compared to the stored words in the database. If there is a match, the Chatbot considers it a meaningful query and generates a response based on it. This process helps improve the accuracy and relevance of the Chatbot's replies by ensuring that the input is processed without the interference of punctuation marks.

4. Performance Analysis

The model is evaluated based on the following performance matrix as stated below.

4.1. Similarity of Bag of Words (BOW)

The Bag of Words (BOW) technique extracts characteristics for machine learning from text. The training set is used to generate a vocabulary of original terms, and each document is represented as a numerical vector based on word frequencies. BOW is widely used but does not capture word context or meaning. It is effective for tasks like sentiment analysis and text classification. BOW is a commonly employed approach that is extensively used in conjunction with natural language processing, information retrieval from documents, and document classifications. The output of program similarity bow is as shown in Table 2.

Table 2. Output of Similarity BOW

Rows of the Text	Answers	Similarity BOW
6	Treatment options for mental illness include t..	1.000000
17	Different types of mental health professionals...	0.707107
8	Different types of mental health professionals...	0.707107
2	Some warning signs of mental illness include...	0.500000
12	Just be tension free and relax	0.408248

Implementing the Bag of Words algorithm with Python in the code is as illustrated in Fig. 2.

```
def text_normalization(text):
    text = str(text).lower()
    spl_char_text = re.sub(r'[^ a-z]','',text)
    tokens = nltk.word_tokenize(spl_char_text)
    lema = wordnet.WordNetLemmatizer()
    tags_list = pos_tag(tokens,tagset=None)
    lema_words = []
    for token,pos_token in tags_list:
        if pos_token.startswith('V'):
            pos_val = 'v'
        elif pos_token.startswith('J'):
            pos_val = 'a'
        elif pos_token.startswith('R'):
            pos_val = 'r'
        else:
            pos_val = 'n'
        lema_token = lema.lemmatize(token,pos_val)

        if lema_token in stop:
          lema_words.append(lema_token)

    return " ".join(lema_words)
```

Fig 2. Pseudo code for Similarity Bow

4.2. Term Frequency-inverse Document Frequency

The TF-IDF score for a term in a text is calculated by combining two distinct metrics through multiplication. There are several methods for calculating the term frequency (TF) of a word in a text, including a simple count of the phrase's occurrences. There are other ways to change the frequency, though. One method is to divide the word's raw count by the length of the document or the word's raw frequency, whichever is higher.

The following formula in Equation (1) is used to determine phrase frequency:

$$Tf(i,j) = \frac{n(i,j)}{\sum n(i,j)} \qquad (1)$$

Where,

$n(i,j)$= number of times nth word occurred in a document

$\sum n(i,j)$ = total number of words in a document.

The inverse document frequency (IDF) of a word as in Equation(2) across a document set determines its commonness or rarity. A lower IDF value indicates a more common word. IDF may be computed by taking the logarithm of the number of documents divided by the number of documents containing the word. So, if a term is widely used and appears in many papers, the IDF value will probably be close to 0. On the other hand, if a term is uncommon and only appears in a few publications, the IDF value will be close to Expressed mathematically, the TF-IDF score is computed using the following formula:

$$IDF = 1 + \log \frac{N}{dN} \qquad (2)$$

Where,

N=Total number of documents in the dataset

dN=total number of documents in which nth word occur

The TF-IDF in Equation (3) is acquired by performing the following calculation:

$$TF\text{-}IDF = TF*IDF \qquad (3)$$

The code of execution process to find TF-IDF is as illustrated in Fig. 3.

```
tfidf = TfidfVectorizer()
x_tfidf = tfidf.fit_transform(df['lemmatized_text']).toarray()
Question_lemma1 = text_normalization(Question1)
Question_tfidf = tfidf.transform([Question_lemma1]).toarray()
df_tfidf = pd.DataFrame(x_tfidf,columns = tfidf.get_feature_names())
df_tfidf.head()
cos = 1-pairwise_distances(df_tfidf,Question_tfidf,metric='cosine')
cos
df['similarity_tfidf'] = cos
df_simi_tfidf = pd.DataFrame(df, columns=['Answers','similarity_tfidf'])
df_simi_tfidf
df_simi_tfidf_sort = df_simi_tfidf.sort_values(by='similarity_tfidf', ascending=False)
df_simi_tfidf_sort.head(10)
threshold = 0.1
df_threshold = df_simi_tfidf_sort[df_simi_tfidf_sort['similarity_tfidf'] > threshold]
df_threshold
```

Fig 3. Pseudo Code of Execution Process to Find the (TF-IDF)

The figure Fig.4. below gives a snapshot of a output obtained during conversation with the Chatbot. A set of questions are asked to the Chatbot and accordingly the Chatbot provides an answer to the user.

Fig 4. A Snapshot of the Test Data of Chatbot produced during Testing

5. Conclusion

The objective of this project is to offer a viable solution to address the escalating mental health challenges encountered by students. To effectively confront this issue, it was imperative to gather direct input from the students themselves via a survey form. Examining the data yielded significant insights into the diverse terminologies associated with mental health issues among students and the correlations between these terms. Subsequent investigation into these methodologies revealed that they encompassed most parameters requiring testing, leading to the selection of medically approved questionnaires for the initial analysis of the users. The training takes a simplified version of the dataset and NLTK, which provides a better model to find queries and answers related to the mental health of a person. The Bag of Words (BOW) approach uses texts to extract features for finding accurate answers. The similarity score (similarity bow) ranges from 0 to 1, where 1 represents the most similar and 0 represents the least similar. It is seen that in most of the cases, it is nearly 0.7 or 1.0. Thus, this Chatbot is built on rules such that it is familiar with the problems and can deliver solutions. The advantages of using a rule-based Chatbot lie in its reduced training time, low cost, and security.

Acknowledgements

The authors are highly grateful to Department of Computer Science & Information Technology, Siksha 'O' Anusandhan Deemed to be University for making this investigation successful.

Author contributions

Smita Rath, Adyasha Pattanayak, Sashikanta Tripathy: Conceptualization, Methodology, Software, Field study,Writing-Reviewing, Writing-Original draft

preparation **Sushree Bibhuprada B. Priyadarshini , Anjela Tripathy:** Data curation, Software, Validation., Field study **Shipra Tanvi:** Visualization, Investigation, and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Dokukina, J. Gumanova, "The rise of chatbots–new personal assistants in foreign language learning",*Procedia Computer Science*, vol.169,pp.542-6,2020 Jan 1.

[2] W.-K. Chen, A. Khadija,F.F. Zahra, & A. Naceur, "AI-Powered Health Chatbots: Toward a general architecture", *Procedia Computer Science*, vol. *191*,pp. 355-360,2021.

[3] E.W. Ngai, M.C. Lee, M. Luo, P.S.Chan, and T.Liang, "An intelligent knowledge-based chatbot for customer service", *Electronic Commerce Research and Applications*, vol. *50*, pp.101098, 2021.

[4] J. U. Duncombe,M. Shumanov,and L. Johnson, "Making conversations with chatbots more personalized", *Computers in Human Behavior*, vol.*117*, p.106627,2021.

[5] *M.* Ogawa,G. Oyama, K. Morito, M. Kobayashi, Y. Yamada, K. Shinkawa, H. Kamo, T. Hatano, and N.Hattori, "Can AI make people happy? The effect of AI-based chatbot on smile and speech in Parkinson's disease", *Parkinsonism & Related Disorders*, vol. *99*, pp.43-46,2022.

[6] O. Zahour, A. Eddaoui, H. Ouchra, and O. Hourrane, "A system for educational and vocational guidance in Morocco: Chatbot E-Orientation", Procedia Computer Science, vol.175, pp.554-559,2020.

[7] G. Sperlí, " A cultural heritage framework using a Deep Learning based Chatbot for supporting tourist journey", *Expert Systems with Applications*, vol.*183*, p.115277,2021.

[8] K. Jwala, G.N.V.G. Sirisha, and G.P. Raju, "Developing a chatbot using machine learning", *International Journal of Recent Technology and Engineering (IJRTE)*, vol.*8*(1S3), pp.89-92,2019.

[9] D.Alekseev, P. Shagalova, and E. Sokolova, "Development of a Chatbot Using Machine Learning Algorithms to Automate Educational Processes", In *Графикон-конференции по компьютерной графике и зрению*",vol. 31, pp. 1104-1113,2021.

[10] S. Ayanouz, B.A. Abdelhakim, and M.Benhmed, "A smart chatbot architecture based NLP and machine learning for health care assistance", In *Proceedings of the 3rd international conference on networking, information systems & security* (pp. 1-6),2020.

[11] K. Konar, "A Comparative Study on Chatbot Based on Machine Learning and Lexicon Based Technique", *International Journal of Innovative Science and Research Technology*, vol.5(5), pp.1534-1542,2020.

[12] F. Colace, F., M. De Santo, M. Lombardi, F. Pascale, A. Pietrosanto, A,and S.Lemma, "Chatbot for e-learning: A case of study", *International Journal of Mechanical Engineering and Robotics Research*, vol.*7*(5), pp.528-533,2018.

[13] S.P.R. Karri,and B.S. Kumar, "Deep learning techniques for implementation of chatbots", In *2020 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-5). IEEE,2020 January.

[14] P. Kandpal, K. Jasnani, R. Raut, and S. Bhorge, " Contextual Chatbot for healthcare purposes (using deep learning)", *In 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, (pp. 625-634). IEEE,2020,July.

[15] D. Carlander-Reuterfelt, Á. Carrera, C.A.Iglesias, O. Araque, J.F.S. Rada, and S. Muñoz, "JAICOB: A data science chatbot", *IEEE Access*, vol. *8*, pp.180672-180680.19,2020.

[16] Boban, A. Doko ,and S. Gotovac, "Sentence retrieval using stemming and lemmatization with different length of the queries", Advances in Science, Technology and Engineering Systems, vol.5(3), pp.349-354,2020.

[17] Motghare, S. M. ., & Nair, P. S. . (2023). Empirical Analysis of Privacy Preservation Models for Cyber Physical Deployments from a Pragmatic Perspective. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3s), 19–29. https://doi.org/10.17762/ijritcc.v11i3s.6151

[18] Mr. Kaustubh Patil. (2013). Optimization of Classified Satellite Images using DWT and Fuzzy Logic. International Journal of New Practices in Management and Engineering, 2(02), 08 - 12. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/15