

## **A Smart Chatbot Using NLP And Machine Learning Technique**

**BHAVANA H V**

M.Tech 4<sup>th</sup> SEM(CSE)

Department Of Computer Science And Engineering

University B D T College Of Engineering ,Davanagere-577004,Karnataka ,India

(A Constitute College Of Visvesvaraya Technological University, Belagavi)

**DR. MOHAMED RAFI**

Department Of Computer Science And Engineering

University B D T College Of Engineering ,Davanagere-577004,Karnataka ,India

(A Constitute College Of Visvesvaraya Technological University, Belagavi)

### **ABSTRACT**

In the past few years, machine learning, deep learning and finally artificial intelligence has grown by developing their features and leaps. This tool is playing the main role and reduces the complex issues that we are tackle today. This machine learning tool will build huge connection between people and machine, so here we are collaborating with natural ` processing also as particular meaning of chatterbot or chatbot is a customer services and communication tool, this chatbot mainly includes communicating through different media like phones, social media such as youtube , facebook ,twitter and etc Chatbot is like a virtual assistant will reminded everything and also perform question and answering. Based on the project it will focus on mapping a sequence of words that constitute the response. This tool emphasis on grammatical mistake sentimental words and more things. Chatbot is kind of chatbox which accept both text and voice based input

---

Keywords: machine learning, artificial intelligence , python ,request ,response.

## I. INTRUCTION

Over many years evolution is taken place in many fields there is drastic changes in area related to machines learning and robotics . we taken to consider that machine learning and artificial intelligence are broadly distributed in the form chatbot. Chatbot is a small example of progression in ML and AI which is user friendly and less inefficient in mode the first word that was experienced in chatbot is ELIZA.it is virtual piece of software where it can mimic to human interaction and also it convers with human artificial intelligence support and process human natural language using many interfaces. Among many intelligent personal assistants currently available, they are structured on rule-based techniques or retrieval-based techniques that generate decent results. from this paper we should overcome limitation of only text based chatbot by introducing voice -based and voice-trained chatbot there are steps in voice based chatbot first step is analysis and second is process. When it is deals with online chat system it would follow the process of client server ,input from client and respond from server this action requires large processing time when large number clients acquires the same server at same time this may lead to limitation because server cannot respond simultaneously AI and ML algorithms are implemented in development of conversations. R&D (Research and Development) are still under progress and experimentation in these fields. conversational agents are predominantly used in many sectors like government administration, banking ,startup companies, online stores, social media. These are largely implemented in corporation and startup companies. However chat are not proper installed in medical fields because One can find many virtual bot development structures in the market, both interface-based and code-based. There are more well known conversational agents are alexa from amazon, google assistant from google and cortana from Microsoft also have limitation in functionality.

## II. RELATED WORKS

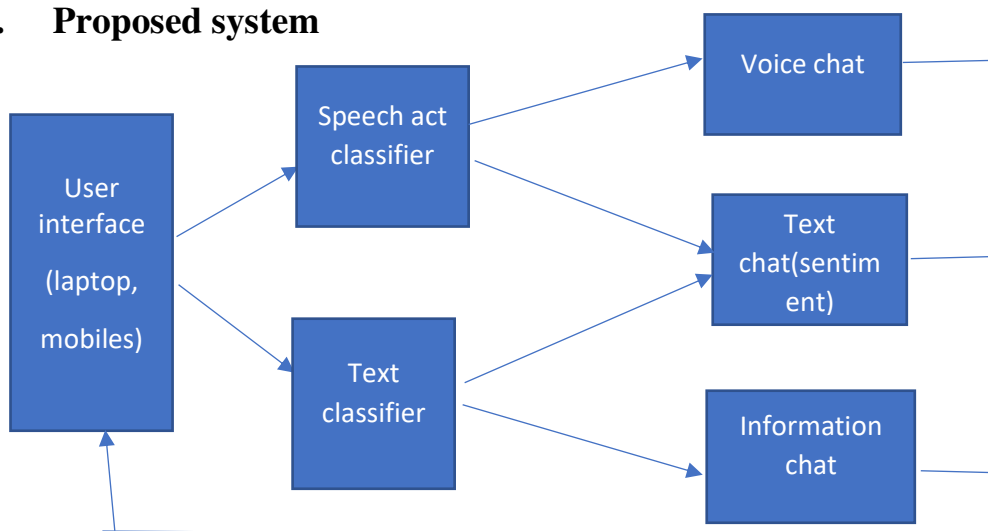
**Wu and. Al. (2017)**analyzed the problem of answer selection for long conversations in retrieval-based chatbots. The goal here is to match between a response candidate and the context in a given conversation, the challenge is to find significant pieces of the context, and know how to implement the relationships between speeches in this context. The matching methods that already exist could lose important information in contexts. The authors proposed a unified framework in which the context is transformed into a fixed-length vector, with no interaction with the answer before matching. This new framework is known as a sequential matching framework (SMF); it can adequately take significant information from the contexts to match the relations between speeches. As the first step, SMF matches a response and converts the pair into a matching vector. Then, the matching vectors are gathered with the help of an RNN. The final

step is the calculation of the context-response matching. The tests of the model's performance were made on two public datasets. The results reveal that both models can surpass the state-of-art matching methods **Saurav Kumar Mishra [1]** describes that medical chatbot will behave as a virtual doctor, which will be allowed to interact with the patients. This chatbot is developed in python language using pattern matching algorithms and Natural Language Processing techniques. According to the survey, which was conducted to check the performance of this chatbot, 80% is the correct answers given by chatbot, while ambiguous/incorrect answers given by this chatbot are 20%. According to the results of this survey, this chatbot can be primarily used as a virtual doctor for care and awareness as well as for teaching medical students.

**Divya Madhu [2]** says that Artificial Intelligence can be used to predict any disease and to provide the list of possible treatments based on given symptoms. Moreover, if periodically analysis is performed on a person's body, AI can help to predict any possible disease even before any damage occurs to the body. The significant challenges for this study are research and development cost and support from the government for the proper successful implementation of all medicines, which are not described in this research paper.

**Hameedullah Kazi [3]** proposed an idea of developing a medical chatbot for medical students. This chatbot uses an open-source AIML based Chatter bean. This AIMML based chatbot can accurately convert natural human language queries into the relevant queries of SQL. Ninety-seven sample questions were collected, and after that, these questions were divided into different groups based upon their types. Depending upon the total number of issues present in each group, the resultant groups were ranked accordingly. According to the queries, questions were made, where 47% of questions are posed questions while other groups have less than 7 percent questions. This system is not specially developed to respond to student queries or, to support natural dialogue in chatbots.

### III. Proposed system



## Input

chatbots solely supported a single adjacency pair, also known as a one-shot conversation. However, modern chatbots can sustain multiple adjacency pairs, remembering states and contexts between conversations and have the capability to associate data in different adjacency pairs which is related. This is the bots ability to preserve the conversation. A chatbot consists of four main parts: frontend, knowledge-base, back-end and corpus which is the training data. The front end is accountable for enabling communication between the bot and the user. The NLU utilises Artificial intelligence methods to identify the intent and context of the user input. An appropriate response is generated from the users' intent. The knowledge base defines the chatbots knowledge, which is created within the NLU and supported by the back-end, the back-end applies the domains corpus to produce the knowledge base Input can be supplied to the chatbot in the form of text or speech. The Input is sent to the dialog management system which is the NLU in this case, which determines an appropriate response and amends the chatbots state accordingly to carry out the required action. The chatbot will produce text and speech responses in the form of both text and speech.

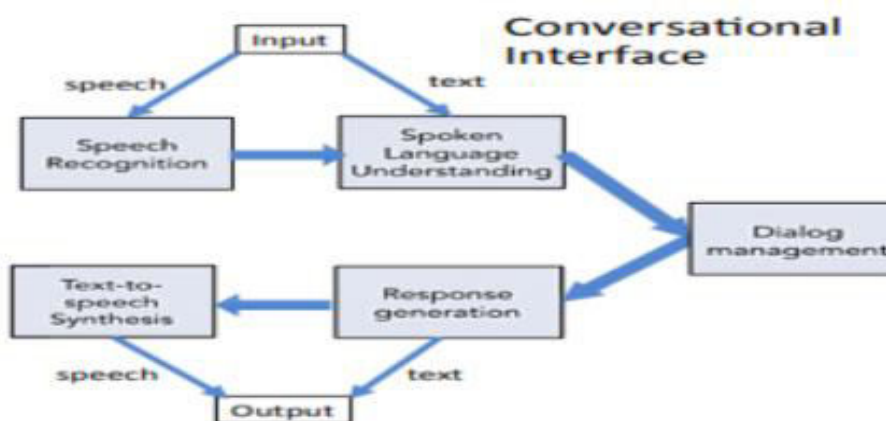


Figure 2: conversational interface

“Natural Language Processing (NLP) is a theory motivated range of computational techniques, for the automatic analysis and representation of human language” Natural Language Processing technology has made great advancements in machine learning based systems to be able to extract meaning from natural language utterances also known as sentiment analysis. Sentiment analysis incorporates multiple natural language processing techniques in order to extract meaning and polarity from text. . The NLU holds liability for the translation of conversational dialogs to actions which are understood by the machine. NLU engines use a variety of artificial intelligence methods to understand the natural language used in conversational interfaces such as chatbots

**Tokenization:** Tokenization is essentially splitting a phrase, sentence, paragraph, or an entire text document into smaller units, such as individual words or terms. Each of these smaller units are called

**tokens** The tokens could be words, numbers or punctuation marks. In tokenization, smaller units are created by locating word boundaries

**NLP:** NLP is a way for computers to analyze, understand, and derive meaning from human language in a smart and useful way. By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation

**NLTK:** The Natural Language Toolkit is an open source library for the Python programming language. It comes with a hands-on guide that introduces topics in computational linguistics as well as programming fundamentals for Python which makes it suitable for linguists who have no deep knowledge in programming, engineers and researchers that need to delve into computational linguistics, students and educators. Natural Language Processing with Python provides a practical introduction to programming for language processing. Written by the creators of NLTK, it guides the reader through the fundamentals of writing Python programs

**Keyword Matching:** Here we define a function for a greeting by the bot that is if a user's input is a greeting, the bot shall return a greeting response.

**Speech Recognition:** Speech recognition helps us to save time by speaking instead of typing. It also gives us the power to communicate with our devices without even writing one line of code. This makes technological devices more accessible and easier to use. Speech recognition is a great example of using machine learning in real life. Speech Recognition library has many classes but we will be focusing on a class called Recognizer. This is the class that will help us to convert audio files into text

- **Duration:** let's say we want only the first 7 seconds of the whole audio file, we have to set the duration parameter to 7.0
- **Offset:** it is used to cut off or skip over a specified amount of second at the start of an audio file. Let's say we don't want the first 1second of the audio files, we have to set the offset parameter to 1.0

Generating Response: To generate a response from our bot for input questions, the concept of document similarity will be used

#### IV. System specification

there are some related to this paper based on there how it works

- Unlike humans who can only communicate with one human at a time, chat bots can simultaneously have conversations with thousands of people.
- are info chat it mainly focus on the knowledge base where it retrieve from the system is voice chat is gives in response in the form of speech to text and text to speech and third is tchat is also called as sentiment analysis
- chatbot gives response as per our mood if are in mood of happy its replay also happy if we chat with bad mood chat also response in sad mood
- A chatbot consists of four main parts: frontend, knowledge-base, back-end and corpus which is the training data

#### V. System implementation

**AIML:** To create our knowledge base for normal conversation, we have used AIML files to store the question and answers pair. When user converses with our chat bot, the input is matched to patterns listed in AIML files and corresponding answer is returned as response. The sample AIML file structure is as:

HELLO USERNAME Hello User! B.

**Lemmatization and POS Tagging Using WordNet:** Information extraction from the input text was done by extracting keywords. For example, “What is the current placement scenario?” contain “current”, “placement” and “scenario” as the keywords. Appropriate Lemmas of the keywords were found using Lemmatization and POS tagging, to group together the different inflected form of the words. For example, requiring, require and required should map to require. WordNet from Python’s “nltk” package was used for this purpose.

**C. Semantic Sentence Similarity:** There are various combinations in which user can input the same query. For example,

Q1: hi how are you?

Q2: how was u your day?

Q1 and Q2 both mean the same thing (same sense). Also, there will be many more combinations for this same query and finding all such combinations will not be feasible. The scalability and performance of the system will also get affected. To overcome this problem, similarity is found out between the user input and the queries present in the available question set (whose answers are available with the system). The query which matches with the input with maximum score Implementation of a Chat Bot System using AI and NLP gets selected (if greater than threshold) and the appropriate response is returned. Similarity score for two sentences is calculated by averaging the similarity of the individual keywords of those sentences. Each keyword of the first sentence is matched with every keyword of the second sentence to find the word with maximum similarity to it. Then the similarity score of individual words are averaged to represent the sentence similarity. To find the word similarity- Path Similarity and Wu-Palmer (WUP) Similarity is used. Path similarity computes shortest number of edges from one word sense to another word sense, assuming a hierarchical structure like WordNet. In general, word senses which have a longer path distance are less similar than those with a very short path distance, e.g. man, dog versus man, tree (expectation is that man is more similar to dog than it is to tree). The Wu-Palmer metric weights the edges based on distance in the hierarchy.

## VI. SIMILAR CHATBOTS

### Medical chatbot-behave like virtual doctor

This is medical chat bot gives instruction to patients based on their health developed by the Saurav Kumar Mishra[2] . According to the survey, which was conducted to check the performance of this chatbot, 80% is the correct answers given by chatbot, while ambiguous/incorrect answers given by this chatbot are 20%.

### Medical chat-for purpose of student

This system was proposed by Hameedullah Kazi [2] where it is used by student for queries Ninety-seven question were collected from that this chatbot possible answer to 47% of questions are posed questions while other groups have less than 7 percent questions.

### One Remission chatbot

The chatbot act like a mental and physical health assistant, it gives the patients the ability to share any positive or negative thoughts; they can communicate either verbally or by text messages, and in return,



they get an accurate explanation of their questions. If they need advice on diets, exercise or sleep, One Remission is here to provide them with the right advice possible.[4]

### Proposed chatbot

Above chatbot some of based on android and there no text based chatbot is merged with voicechatbot but in proposed system we have merged both voice chatbot and text chatbot in single tool when it comes to text chatbot it can give answers to all the which were in database or computer program voice based it can answer to 17 queries 20, and it also depend on database and one more feature is added that is sentiment analysis chatbot also can give answer based on their mode

Results :

|  | Input questions | Output answer |
|--|-----------------|---------------|
| Medical chatbot-behave like virtual doctor | In percent 100% | 80% output    |
| Medical chat-for purpose of student        | 97 questions    | 47% output    |
| Proposed system                            | 20 queries      | 17 output     |

## VII. APPLICATIONS

- It enables the students to be updated with college activities.
- It saves time for the students as well as teaching and non-teaching staffs.
- It is providing us a readily available information source without taking any physical efforts.
- It is easily accessible and saving time and money also.

## VIII. CONCLUSIONS

It is often impossible to get all the data on a single interface without the complications of going through multiple forms and windows. The college chat bot aims to remove this difficulty by providing a



common and user-friendly interface to solve queries of college students and teachers. The purpose of a chat bot system is to simulate a human conversation. Its architecture integrates a language model and computational algorithm to emulate information online communication between a human and a computer using natural language. The college student and employees can freely upload their queries. The chat bot provides fast and efficient search for answers to the queries and gets the relevant links to their question. A background research took place, which included an overview of the conversation procedure and tries to find out the relevant keywords related to that query to provide the proper link. The database storage includes information about questions, answers, keywords, and logs. We have also developed an interface. The interface developed will have two parts, one for users and the other for the administrator.

- IX. **FUTURE ENHANCEMENTS:** Instead of AIML based bot, other algorithms can be implemented. We can include voice-based queries. The users will have to give voice input and the system will give the text output. Also, after successful execution of chat bot in college domain, we can implement it in other domains like medical, forensic, sports, etc. It will be beneficial in all the fields as without spending much time, we are accessing the relevant information and that too without any sorting.

## X. REFERENCE

- [1] C. J. N. J. S. S. Divya Madhu, "A novel approach for medical assistance using trained chatbot," in International Conference on Inventive Communication and Computational Technologies (ICICCT), 2017.
- [2] Saurav Kumar Mishra, DharendraBharti, Nidhi Mishra, " Dr.Vdoc: A Medical Chatbot that Acts as a virtual Doctor", Journal of Medical Science and Technology, Volume: 6, Issue 3,2017
- [3] S. A. N. H. Hameedullah Kazi, "Effect of Chatbot Systems on Student's Learning Outcomes," SYLWAN, 2019.
- [4] D. B. Mesko, "The Medical Futurist," The Medical Futurist Institute, 2020. [Online]. Available: <https://medicalfuturist.com/magazine/>
- [5] Soufyane Ayanouz, Boudhir Anouar Abdelhakim, and Mohammed Benhmed," A Smart Chatbot Architecture based NLP and Machine learning for health care assistance." t: <https://www.researchgate.net/publication/340678278>