



NLP BASED CHATBOT FOR CS STUDENTS

¹G. Jemilda, ²M. Mahalakshmi, ³M. Maha Shree, ⁴M. Murugeswari, ⁵D. Priyadharshini

¹Faculty, ^{2,3,4,5}UG Scholar

Computer Science and Engineering

JayarajAnnapackiam CSI College of Engineering, Nazareth, India.

Abstract: A chatbot is a computer program designed to simulate conversation with human users, especially over the Internet. Chatbots use Natural Language Processing (NLP) and Machine Learning (ML) algorithms to understand and generate responses to user inputs. It can be integrated into messaging platforms, and mobile apps to provide automated computer support, answering frequently asked questions. User interacts with the system using text and voice messages. It gives the answer for Computer Science (CS) department questions. It gives impressive response from chatbot. It is an open source platform. Some chatbots are rule based, using a set of predetermined responses. While others are more advanced and use machine learning to generate responses based on the context of the conversation.

Index Terms - Chatbot, NLP, ML, CS department Queries.

I. INTRODUCTION

Chatbots, also known as conversational agents, enable the interaction of humans with computers through natural language, by applying the technology of NLP. Due to their ability to emulate human conversations, services automated and efforts reduced. The adaptation of educational chatbots is on the rise due to their ability to provide a cost-effective method to engage students and provide a personalized learning experience. Chatbot adoption is especially crucial in online classes that include many students where individual support from educators to students is challenging. Chatbots can facilitate learning within the educational context, for instance by instantaneously providing students with course. Moreover, chatbots may interact with students individually or support collaborative learning activities by applying text, speech and other modes of communication to assist learners.

II. LITERATUREREVIEW

Theodora A. Maniou et al., described an Android General Knowledge Chatbot using Naive Bayes algorithm, It analyzes user's queries and understand user's message. This system provides answers to the queries of the users. The user can ask questions about the different categories like sports, education, health, etc. The system will compute the question with highly optimize algorithm and respond to the user. Chatbot for answering the user's query in fastest ways save time and requires active internet connection. The system can be used for schools, colleges, or universities as a smart assistance. In this paper, a chatbot application in android is introduced which is able to interact with users. This chatbot can answer the queries in textual form. But system may provide inaccurate results if data not entered properly.

Rohan Tiwari et al., measured human machine conversation using chatbot by implementing ML. It uses K-Nearest Neighbor (KNN) algorithm. It is inspired from the working of human brain, how neurons are inter connected in the brain to those clients will hold discussions, regularly by means of text. The website gives all fields of questions. They help you to know your customers. The chatbot is trained using datasets of various types so that it can answer a broad range of questions. An advantage of human machine conversation using chatbot reduces the operational cost and has disadvantages like less understanding of NLP technique using Parts of speech. Our work overcomes the problem of less understanding of NLP techniques.

Atlanta et al., proposed intelligent chatbot tool to assist high school students. In education field also, features of ML, deep learning, and NLP techniques are most widely used for building Artificial Intelligence (AI) based chatbots. This artificial intelligent chatbot tool is developed to assist high school students for learning their general knowledge subjects. The purpose of a chatbot system is to simulate a human conversation. AI chatbots are used to service more customers, collect and analyze data, reduce human resource costs, lack human empathy, limited decision-making, expensive, etc. It is often impossible to get all the data on a single interface without the complications of going through multiple forms and windows.

AvniPrajapathi et al., proposed android based chatbot for college using the algorithm of Naive Bayes. The main advantage gives timely response. System return answers in few seconds rather than minutes. Chatbots hold the promise of revolutionizing education by engaging learners, personalizing, supporting educators, and developing deep insight into learners' behavior. However,

there is a lack of studies that analyze the recent evidence-based chatbot-learner interaction design techniques applied in education. This system provides answers to the queries of the users by analyzing the keywords.

III. RESEARCH METHODOLOGY

Intellectual chatbot for CS emphasize the significance by putting an algorithm which is relatively easy for implementation and it gives impressive response. In our work, we are having six modules such as Data collection, preprocessing, Feature Extraction, Training, Testing and predicted output.

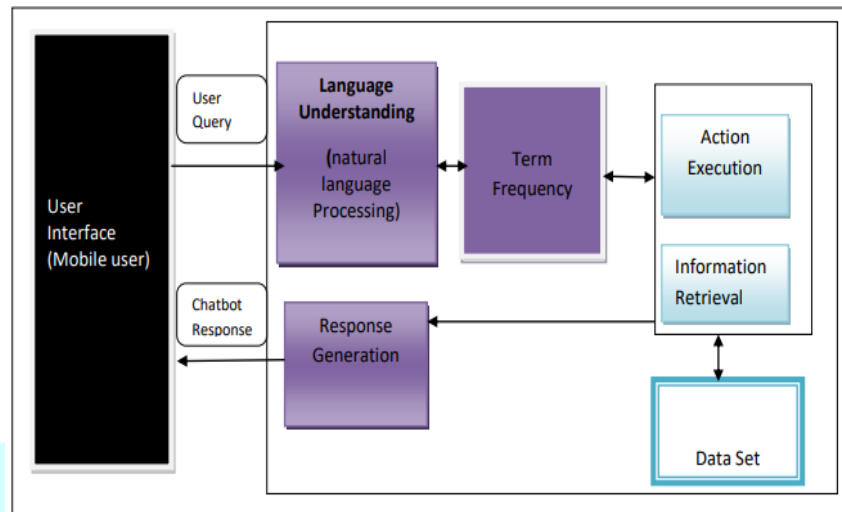


Fig.1 Architecture Diagram

In the above Figure 1, user asks the query using text and voice messages to the chatbot. Chatbot understands the query using NLP. Preprocessing done for user query, then features extracted, and then training of data set done by KNN algorithm. If user query matches the data set then answer retrieved from the database by giving response to the user query.

3.1 Data Collection

It is a process of gathering and measuring the information on variables of interest, in an established fashion that enables to answer stated research questions, test hypothesis and evaluate outcomes. Large data set of conversation between user and chatbot was collected which contains CS department syllabus questions with detailed explanation of these answers which is shown in Figure 2.

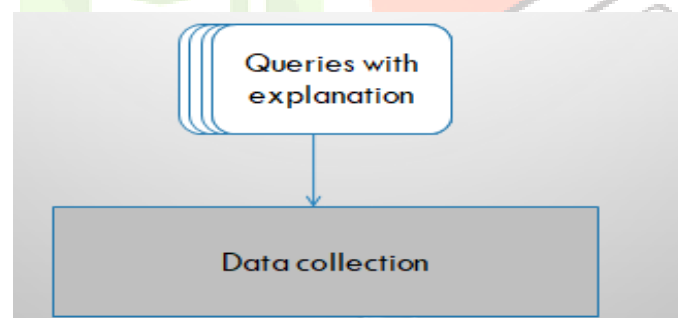


Fig.2Data Collection

3.2 Preprocessing

Data preprocessing, a component of data preparation, describes any type of processing performed on raw data to prepare it for another data processing procedure. Preprocessing means cleaning the collected data to remove irrelevant information and preparing it for training.

3.2.1 Pseudo code

```

words=[lemmatizer.lemmatize(w.lower())forwinwordsifwnotinignore_words]
words = sorted(list(set(words)))
classes = sorted(list(set(classes)))
print(len(documents), "documents")
print(len(classes),"classes",classes)
print(len(words),"uniquelemmatizedwords",words)
pickle.dump(words,open('texts.pkl','wb'))
pickle.dump(classes,open('labels.pkl','wb'))
  
```

3.3 Feature Extraction

Feature Extraction done by term frequency. Term frequency means number of times the term appears in a query. It is compared with the total number of times the term appears in the dataset.

3.3.1 Pseudo code

```
fordocindocuments:
  bag=[]
  for the pattern
    pattern_words=doc[0]
    pattern_words=[lemmatizer.lemmatize(word.lower())forwordinpattern_words]
    forwinwords:
      bag.append(1)ifwinpattern_wordseelsebag.append(0)
    output_row= list(output_empty)output_row[classes.index(doc[1])]=1
```

3.4 Training

The training process is explained in Figure 3. Train a ML model on the preprocessed data and extract features to learn the language. Data set is trained using KNN algorithm. The observations in the training set uses supervised learning algorithms. Each observation consists of an observed output variable and one or more observed input variables. If the test set contain examples from the training set, it will be difficult to access the algorithm. A program that generalizes well will be able to effectively perform a task with new data.

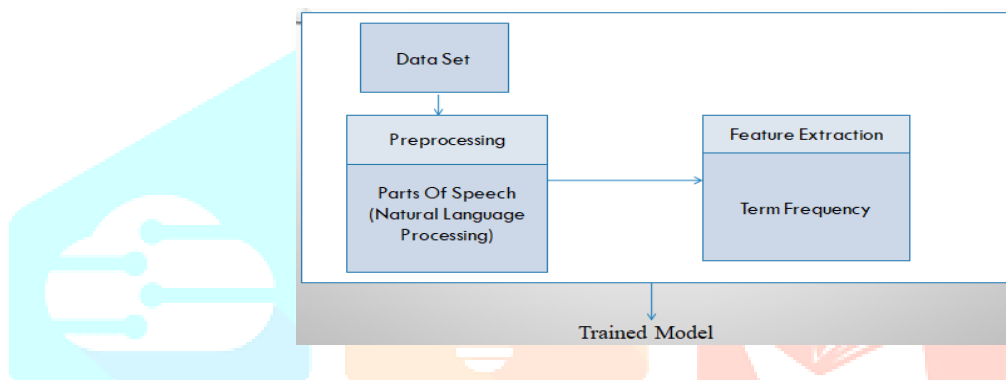


Fig.3 Training

3.5 Testing

The testing process is explained in Figure 4. Evaluate the performance of the trained model on a separate data set to determine its accuracy. The testing data is used to evaluate the performance of the model using some performance metric.

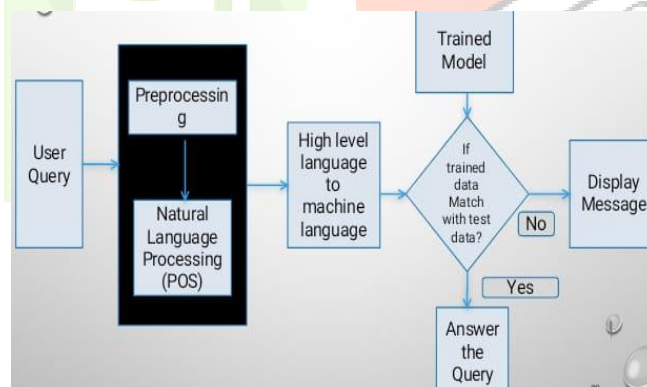


Fig.4 Testing

3.5.1 Steps

- Collect data set with labeled data.
- Preprocess the data.
- Split the data into training and testing sets.
- Choose a value for k.
- For each instance in the training set,
- Calculate the distance between the instance and all instances in the training set.
- Sort the distances in ascending order.
- Select the first k instances with the smallest distances.
- Determine the class of Instances based on the majority class of its k nearest neighbors.

Calculate the accuracy of the model on the training Set.

3.6 Predicted Output

NLP techniques used to understand the user's text inputs and provide personalized responses based on the user's need which is shown in Figure 5.

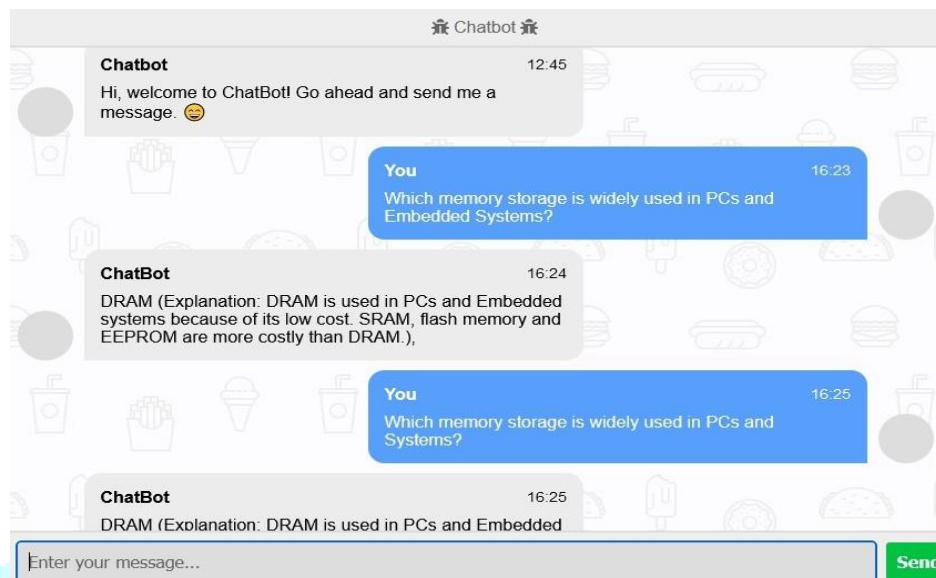


Fig.5 Sample Output

IV. RESULTS AND DISCUSSION

The results obtained by the experimental campaign are satisfying and show the good perspective. Further developments involve the application of the proposed approach in various contexts and an improvement of e-learning platform and these findings provide further evidence to suggest that chatbot programming is still in its early stages. Future research can add two more fields. The first field is to focus on the developers support to create and offer tools that allow user interact with chatbot using voice and text messages. It provides educational chatbot guidelines to support students learning successfully. The second field is a content analysis of the actual conversations with students. Chatbots using AI have revolutionized the businesses to interact with their customers. AI-powered chatbots can understand natural language, analyze data, and provide personalized responses to customers in real-time.

AI chatbots can also be integrated with other technologies such as voice assistants, virtual and augmented reality, and the Internet of Things (IoT), creating an intelligent ecosystem that can provide more engaging and immersive experiences for users. However, it is important to note that chatbots are not perfect and can sometimes fail to understand or respond appropriately to user queries. Therefore, it is crucial to continually monitor and improve the chatbot's performance by analyzing user feedback and making necessary adjustments. Overall, chatbots using artificial intelligence have become an essential tool for businesses looking to improve customer engagement, increase efficiency, and stay competitive in today's digital landscape.

4.1. Sample Screenshots

Figures 6 to 9 represents the sample screenshots.



Fig.6 Home Page



Fig.7Registration



Fig.8Login

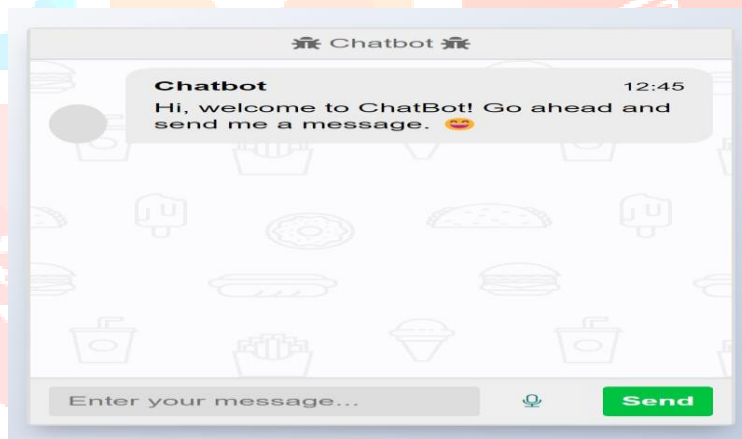


Fig.9Main Page

V. FUTURE ENHANCEMENT

Intellectual chatbot for CS help students to find the right sources of information not only gives answer for queries but also gives explanation. It will provide an instant as well as accurate response. For future enhancement, some more department questions can be included.

VI. ACKNOWLEDGMENT

We would like to acknowledge our sincere thanks to the Management of our College and our family members who have supported and helped us in different stages of this project work.

REFERENCES

- [1] Rohan Tiwari, G.M.S.S. Pranav, R. Prema, 2022. Human Machine Conversation using chatbot. International Research Journal of Engineering and Technology (IRJET), 09 (03).
- [2] Atlanta, Georgia, 2017. Developing an Intelligent Chat-bot Tool to assist high school students. Georgia Institute of Technology, 11.
- [3] AvniPrajapati, PranaliNaik, Sushmita Singh, Anil Kale, 2018. Android based Chatbot for College. International Journal of Scientific & Engineering Research, 9 (04).
- [4] Theodora A. Maniou, Andreas Veglis, 2020. Android General Knowledge Chatbot. Aristotle University of Thessaloniki, 30.

