CREATE A CHATBOT USING PYTHON:

College Code:5113

Team Members:

D.Aakash(511321106001)

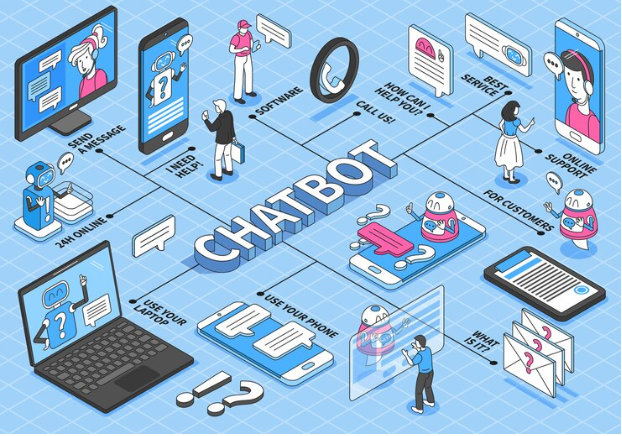
S.Chandrsekran(511321106005)

A.Dinesh(511321106010)

J.Yukendiran(511321106306)

**Phase 3 Project:Devlopment**

**Project Title:**Create A Chatbot using Python



**Introduction:**

* Chatbots are computer programs designed to simulate human conversation by interacting with users via text or voice. They have become increasingly popular in various applications, including customer support, virtual assistants, and information retrieval. Building a chatbot using Python can be a rewarding project, and it often involves various components, such as data preprocessing, natural language processing (NLP), and machine learning.
* Chatbots, also known as conversational agents, are designed with the help of AI (Artificial Intelligence) software. They simulate a conversation (or a chat) with users in a natural language via messaging applications, websites, mobile apps, or phone.

**Purpose of Using Chatbot:**

Chatbots allow businesses to connect with customers in a personal way without the expense of human representatives. For example, many of the questions or issues customers have are common and easily answered. That's why companies create FAQs and troubleshooting guides.



**Preprocessing the Dataset:**

The quality of a chatbot heavily relies on the dataset it's trained on. Preprocessing the dataset is a critical step to ensure the chatbot can understand and respond appropriately to user inputs. Here's an overview of dataset preprocessing in chatbot development:

1. **Data Collection**: Begin by collecting a dataset that consists of questions or user inputs and corresponding responses. This dataset can be gathered from various sources, such as FAQs, chat logs, or custom-created dialogues.
2. **Data Cleaning**: Raw text data often contains noise, special characters, or unnecessary formatting. Clean the data by removing unwanted elements, such as HTML tags, non-alphanumeric characters, and irrelevant metadata.
3. **Text Normalization**: Normalize the text by converting it to lowercase. This ensures that the chatbot's responses are consistent and independent of the user's input casing.
4. **Tokenization**: Tokenization is the process of splitting text into individual words or tokens. This step allows the chatbot to analyze and understand the text on a word-by-word basis.
5. **Stopword Removal**: Stopwords are common words like "and," "the," "is," etc., that often don't carry significant meaning. Removing stopwords can help improve the efficiency and accuracy of the chatbot's responses.
6. **Lemmatization or Stemming**: Reducing words to their root form helps the chatbot recognize different word forms as equivalent. You can use lemmatization (reducing words to their base form) or stemming (removing prefixes and suffixes) for this purpose.
7. **Handling Synonyms and Variations**: Addressing synonyms and variations in the dataset is crucial. You can create mappings or equivalence tables to link synonyms or different ways of phrasing the same question to a common response.
8. **Data Structure**: Organize the preprocessed data into a suitable format, such as a JSON file, CSV, or a database, making it accessible for training and future use.

**Requirment Packages:**

* Numpy
* Pandas
* Keras
* Json
* Pickels
* Warnings
* Tensorflow
* Keras
* Nltk
* Matplotlib.pyplot

Necessary step to follow:

**Step 1: Define Your Objective and Collect Data**: Before you start coding, define the purpose and objective of your chatbot. Then, gather a dataset containing questions and their corresponding responses. For simplicity, let's use a JSON file for the dataset.

import json

# Load your dataset from a JSON file (questions and responses).

with open("dataset.json", "r") as file:

data = json.load(file)

**Step 2: Preprocess the Data**

Preprocessing the data is crucial to make it suitable for training and usage by the chatbot. You can tokenize and clean the text, remove punctuation, and convert it to lowercase.

import string

import nltk

nltk.download('punkt')

# Tokenize and clean the text

def preprocess\_text(text):

text = text.lower()

text = text.translate(str.maketrans("", "", string.punctuation))

return text

# Preprocess the dataset

for pair in data:

pair['question'] = preprocess\_text(pair['question'])

pair['response'] = preprocess\_text(pair['response'])

**Step 3: Build a Simple Rule-Based Chatbot**

For simplicity, let's create a basic rule-based chatbot that looks for exact matches in the dataset and responds accordingly.

def simple\_chatbot(input\_text):

input\_text = preprocess\_text(input\_text)

for pair in data:

if pair['question'] == input\_text:

return pair['response']

return "I'm sorry, I don't understand your question."

# Testing the chatbot

while True:

user\_input = input("You: ")

if user\_input.lower() == 'exit':

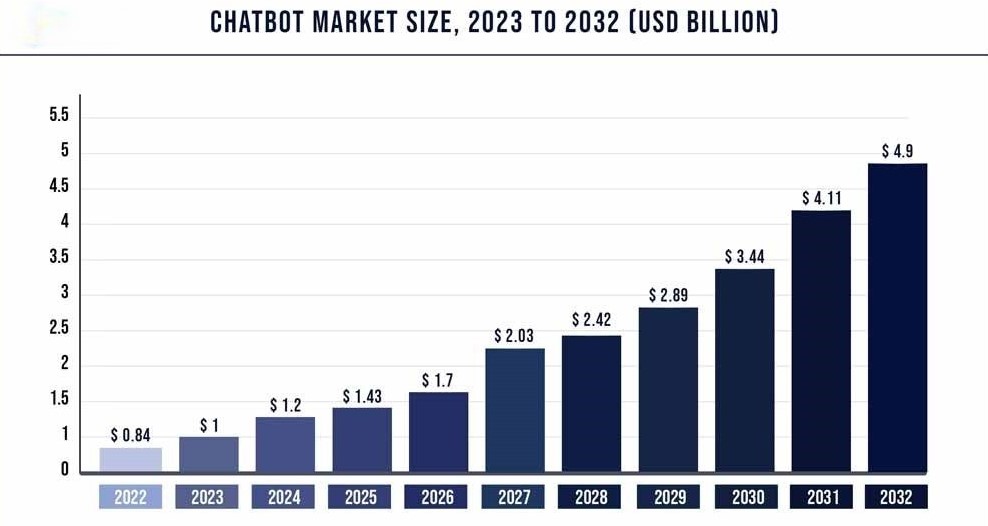
break

response = simple\_chatbot(user\_input)

print("Chatbot:", response

**Devlopment of Chatbot:**

The chatbot market is estimated to grow from USD 5.27 billion in 2023 and is likely to grow at a CAGR of 23.28% during 2023-2028 to reach USD 14.95 billion by 2028.

**Program:**

In [1]:

import nltk

nltk.download('punkt')*#Sentence tokenizer*

[nltk\_data] Downloading package punkt to /usr/share/nltk\_data...

[nltk\_data] Package punkt is already up-to-date!

Out[1]:

True

In [2]:

import nltk

from nltk.stem import WordNetLemmatizer

lemmatizer = WordNetLemmatizer()

import json

import pickle

import warnings

warnings.filterwarnings('ignore')

Out[2]:True

In [3]:

import numpy as np

import tensorflow as tf

from keras.models import Sequential

from keras.layers import Dense, Activation, Dropout

from tensorflow.keras.optimizers import SGD

import random

Out[3]:True

**Preprocessing:**

In [4]:

words=[]

classes = []

documents = []

ignore\_words = ['?', '!']

data\_file = open('/kaggle/input/chatbot-dataset/intents.json').read() *# read json file*

intents = json.loads(data\_file) *# load json file*

When working with text data, we need to perform various preprocessing on the data before we make a machine learning or a deep learning model. Based on the requirements we need to apply various operations to preprocess the data.

* Tokenizing is the most basic and first thing you can do on text data.
* Tokenizing is the process of breaking the whole text into small parts like words.
* Here we iterate through the patterns and tokenize the sentence using nltk.word\_tokenize() function and append each word in the words list. We also create a list of classes for our tags.

In [5]:

for intent in intents['intents']:

for pattern in intent['patterns']:

*#tokenize each word*

w = nltk.word\_tokenize(pattern)

words.extend(w)*# add each elements into list*

*#combination between patterns and intents*

documents.append((w, intent['tag']))*#add single element into end of list*

*# add to tag in our classes list*

if intent['tag'] not in classes:

classes.append(intent['tag'])

In [6]:

nltk.download('wordnet') *#lexical database for the English language*

[nltk\_data] Downloading package wordnet to /usr/share/nltk\_data...

[nltk\_data] Package wordnet is already up-to-date!

Out[6]:

True

In [7]:

nltk.download('omw-1.4')

[nltk\_data] Downloading package omw-1.4 to /usr/share/nltk\_data...

Out[7]:

True

Now we will lemmatize each word and remove duplicate words from the list.

* Lemmatizing is the process of converting a word into its lemma form and then creating a pickle file to store the Python objects which we will use while predicting.

In [8]:

*# lemmatize, lower each word and remove duplicates*

words = [lemmatizer.lemmatize(w.lower()) for w in words if w not in ignore\_words]

words = sorted(list(set(words)))

*# sort classes*

classes = sorted(list(set(classes)))

*# documents = combination between patterns and intents*

print (len(documents), "documents\n", documents, "\n")

*# classes = intents[tag]*

print (len(classes), "classes\n", classes, "\n")

*# words = all words, vocabulary*

print (len(words), "unique lemmatized words\n", words, "\n")

pickle.dump(words,open('words.pkl','wb'))

pickle.dump(classes,open('classes.pkl','wb'))

405 documents

[(['Hi'], 'greeting'), (['How', 'are', 'you', '?'], 'greeting'), (['Is', 'anyone', 'there', '?'], 'greeting'), (['Hello'], 'greeting'), (['Good', 'day'], 'greeting'), (['What', "'s", 'up'], 'greeting'), (['how', 'are', 'ya'], 'greeting'), (['heyy'], 'greeting'), (['whatsup'], 'greeting'), (['?', '?', '?', '?', '?', '?', '?', '?'], 'greeting'), (['cya'], 'goodbye'), (['see', 'you'], 'goodbye'), (['bye', 'bye'], 'goodbye'), (['See', 'you', 'later'], 'goodbye'), (['Goodbye'], 'goodbye'), (['I', 'am', 'Leaving'], 'goodbye'), (['Bye'], 'goodbye'), (['Have', 'a', 'Good', 'day'], 'goodbye'), (['talk', 'to', 'you', 'later'], 'goodbye'), (['ttyl'], 'goodbye'), (['i', 'got', 'to', 'go'], 'goodbye'), (['gtg'], 'goodbye'), (['what', 'is', 'the', 'name', 'of', 'your', 'developers'], 'creator'), (['what', 'is', 'the', 'name', 'of', 'your', 'creators'], 'creator'), (['what', 'is', 'the', 'name', 'of', 'the', 'developers'], 'creator'), (['what', 'is', 'the', 'name', 'of', 'the', 'creators'], 'creator'), (['who', 'created', 'you'], 'creator'), (['your', 'developers'], 'creator'), (['your', 'creators'], 'creator'), (['who', 'are', 'your', 'developers'], 'creator'), (['developers'], 'creator'), (['you', 'are', 'made', 'by'], 'creator'), (['you', 'are', 'made', 'by', 'whom'], 'creator'), (['who', 'created', 'you'], 'creator'), (['who', 'create', 'you'], 'creator'), (['creators'], 'creator'), (['who', 'made', 'you'], 'creator'), (['who', 'designed', 'you'], 'creator'), (['name'], 'name'), (['your', 'name'], 'name'), (['do', 'you', 'have', 'a', 'name'], 'name'), (['what', 'are', 'you', 'called'], 'name'), (['what', 'is', 'your', 'name'], 'name'), (['what', 'should', 'I', 'call', 'you'], 'name'), (['whats', 'your', 'name', '?'], 'name'), (['what', 'are', 'you'], 'name'), (['who', 'are', 'you'], 'name'), (['who', 'is', 'this'], 'name'), (['what', 'am', 'i', 'chatting', 'to'], 'name'), (['who', 'am', 'i', 'taking', 'to'], 'name'), (['what', 'are', 'you'], 'name'), (['timing', 'of', 'college'], 'hours'), (['what', 'is', 'college', 'timing'], 'hours'), (['working', 'days'], 'hours'), (['when', 'are', 'you', 'guys', 'open'], 'hours'), (['what', 'are', 'your', 'hours'], 'hours'), (['hours', 'of', 'operation'], 'hours'), (['when', 'is', 'the', 'college', 'open'], 'hours'), (['college', 'timing'], 'hours'), (['what', 'about', 'college', 'timing'], 'hours'), (['is', 'college', 'open', 'on', 'saturday'], 'hours'), (['tell', 'something', 'about', 'college', 'timing'], 'hours'), (['what', 'is', 'the', 'college', 'hours'], 'hours'), (['when', 'should', 'i', 'come', 'to', 'college'], 'hours'), (['when', 'should', 'i', 'attend', 'college'], 'hours'), (['what', 'is', 'my', 'college', 'time'], 'hours'), (['college', 'timing'], 'hours'), (['timing', 'college'], 'hours'), (['more', 'info'], 'number'), (['contact', 'info'], 'number'), (['how', 'to', 'contact', 'college'], 'number'), (['college', 'telephone', 'number'], 'number'), (['college', 'number'], 'number'), (['What', 'is', 'your', 'contact', 'no'], 'number'), (['Contact', 'number', '?'], 'number'), (['how', 'to', 'call', 'you'], 'number'), (['College', 'phone', 'no', '?'], 'number'), (['how', 'can', 'i', 'contact', 'you'], 'number'), (['Can', 'i', 'get', 'your', 'phone', 'number'], 'number'), (['how', 'can', 'i', 'call', 'you'], 'number'), (['phone', 'number'], 'number'), (['phone', 'no'], 'number'), (['call'], 'number'), (['list', 'of', 'courses'], 'course'), (['list', 'of', 'courses', 'offered'], 'course'), (['list', 'of', 'courses', 'offered', 'in'], 'course'), (['what', 'are', 'the', 'courses', 'offered', 'in', 'your', 'college', '?'], 'course'), (['courses', '?'], 'course'), (['courses', 'offered'], 'course'), (['courses', 'offered', 'in', '(', 'your', 'univrsity', '(', 'UNI', ')', 'name', ')'], 'course'), (['courses', 'you', 'offer'], 'course'), (['branches', '?'], 'course'), (['courses', 'available', 'at', 'UNI', '?'], 'course'), (['branches', 'available', 'at', 'your', 'college', '?'], 'course'), (['what', 'are', 'the', 'courses', 'in', 'UNI', '?'], 'course'), (['what', 'are', 'branches', 'in', 'UNI', '?'], 'course'), (['what', 'are', 'courses', 'in', 'UNI', '?'], 'course'), (['branches', 'available', 'in', 'UNI', '?'], 'course'), (['can', 'you', 'tell', 'me', 'the', 'courses', 'available', 'in', 'UNI', '?'], 'course'), (['can', 'you', 'tell', 'me', 'the', 'branches', 'available', 'in', 'UNI', '?'], 'course'), (['computer', 'engineering', '?'], 'course'), (['computer'], 'course'), (['Computer', 'engineering', '?'], 'course'), (['it'], 'course'), (['IT'], 'course'), (['Information', 'Technology'], 'course'), (['AI/Ml'], 'course'), (['Mechanical', 'engineering'], 'course'), (['Chemical', 'engineering'], 'course'), (['Civil', 'engineering'], 'course'), (['information', 'about', 'fee'], 'fees'), (['information', 'on', 'fee'], 'fees'), (['tell', 'me', 'the', 'fee'], 'fees'), (['college', 'fee'], 'fees'), (['fee', 'per', 'semester'], 'fees'), (['what', 'is', 'the', 'fee', 'of', 'each', 'semester'], 'fees'), (['what', 'is', 'the', 'fees', 'of', 'each', 'year'], 'fees'), (['what', 'is', 'fee'], 'fees'), (['what', 'is', 'the', 'fees'], 'fees'), (['how', 'much', 'is', 'the', 'fees'], 'fees'), (['fees', 'for', 'first', 'year'], 'fees'), (['fees'], 'fees'), (['about', 'the', 'fees'], 'fees'), (['tell', 'me', 'something', 'about', 'the', 'fees'], 'fees'), (['What', 'is', 'the', 'fees', 'of', 'hostel'], 'fees'), (['how', 'much', 'is', 'the', 'fees'], 'fees'), (['hostel', 'fees'], 'fees'), (['fees', 'for', 'AC', 'room'], 'fees'), (['fees', 'for', 'non-AC', 'room'], 'fees'), (['fees', 'for', 'Ac', 'room', 'for', 'girls'], 'fees'), (['fees', 'for', 'non-Ac', 'room', 'for', 'girls'], 'fees'), (['fees', 'for', 'Ac', 'room', 'for', 'boys'], 'fees'), (['fees', 'for', 'non-Ac', 'room', 'for', 'boys'], 'fees'), (['where', 'is', 'the', 'college', 'located'], 'location'), (['college', 'is', 'located', 'at'], 'location'), (['where', 'is', 'college'], 'location'), (['where', 'is', 'college', 'located'], 'location'), (['address', 'of', 'college'], 'location'), (['how', 'to', 'reach', 'college'], 'location'), (['college', 'location'], 'location'), (['college', 'address'], 'location'), (['wheres', 'the', 'college'], 'location'), (['how', 'can', 'I', 'reach', 'college'], 'location'), (['whats', 'is', 'the', 'college', 'address'], 'location'), (['what', 'is', 'the', 'address', 'of', 'college'], 'location'), (['address'], 'location'), (['location'], 'location'), (['hostel', 'facility'], 'hostel'), (['hostel', 'servive'], 'hostel'), (['hostel', 'location'], 'hostel'), (['hostel', 'address'], 'hostel'), (['hostel', 'facilities'], 'hostel'), (['hostel', 'fees'], 'hostel'), (['Does', 'college', 'provide', 'hostel'], 'hostel'), (['Is', 'there', 'any', 'hostel'], 'hostel'), (['Where', 'is', 'hostel'], 'hostel'), (['do', 'you', 'have', 'hostel'], 'hostel'), (['do', 'you', 'guys', 'have', 'hostel'], 'hostel'), (['hostel'], 'hostel'), (['hostel', 'capacity'], 'hostel'), (['what', 'is', 'the', 'hostel', 'fee'], 'hostel'), (['how', 'to', 'get', 'in', 'hostel'], 'hostel'), (['what', 'is', 'the', 'hostel', 'address'], 'hostel'), (['how', 'far', 'is', 'hostel', 'from', 'college'], 'hostel'), (['hostel', 'college', 'distance'], 'hostel'), (['where', 'is', 'the', 'hostel'], 'hostel'), (['how', 'big', 'is', 'the', 'hostel'], 'hostel'), (['distance', 'between', 'college', 'and', 'hostel'], 'hostel'), (['distance', 'between', 'hostel', 'and', 'college'], 'hostel'), (['events', 'organised'], 'event'), (['list', 'of', 'events'], 'event'), (['list', 'of', 'events', 'organised', 'in', 'college'], 'event'), (['list', 'of', 'events', 'conducted', 'in', 'college'], 'event'), (['What', 'events', 'are', 'conducted', 'in', 'college'], 'event'), (['Are', 'there', 'any', 'event', 'held', 'at', 'college'], 'event'), (['Events', '?'], 'event'), (['functions'], 'event'), (['what', 'are', 'the', 'events'], 'event'), (['tell', 'me', 'about', 'events'], 'event'), (['what', 'about', 'events'], 'event'), (['document', 'to', 'bring'], 'document'), (['documents', 'needed', 'for', 'admision'], 'document'), (['documents', 'needed', 'at', 'the', 'time', 'of', 'admission'], 'document'), (['documents', 'needed', 'during', 'admission'], 'document'), (['documents', 'required', 'for', 'admision'], 'document'), (['documents', 'required', 'at', 'the', 'time', 'of', 'admission'], 'document'), (['documents', 'required', 'during', 'admission'], 'document'), (['What', 'document', 'are', 'required', 'for', 'admission'], 'document'), (['Which', 'document', 'to', 'bring', 'for', 'admission'], 'document'), (['documents'], 'document'), (['what', 'documents', 'do', 'i', 'need'], 'document'), (['what', 'documents', 'do', 'I', 'need', 'for', 'admission'], 'document'), (['documents', 'needed'], 'document'), (['size', 'of', 'campus'], 'floors'), (['building', 'size'], 'floors'), (['How', 'many', 'floors', 'does', 'college', 'have'], 'floors'), (['floors', 'in', 'college'], 'floors'), (['floors', 'in', 'college'], 'floors'), (['how', 'tall', 'is', 'UNI', "'s", 'College', 'of', 'Engineering', 'college', 'building'], 'floors'), (['floors'], 'floors'), (['Syllabus', 'for', 'IT'], 'syllabus'), (['what', 'is', 'the', 'Information', 'Technology', 'syllabus'], 'syllabus'), (['syllabus'], 'syllabus'), (['timetable'], 'syllabus'), (['what', 'is', 'IT', 'syllabus'], 'syllabus'), (['syllabus'], 'syllabus'), (['What', 'is', 'next', 'lecture'], 'syllabus'), (['is', 'there', 'any', 'library'], 'library'), (['library', 'facility'], 'library'), (['library', 'facilities'], 'library'), (['do', 'you', 'have', 'library'], 'library'), (['does', 'the', 'college', 'have', 'library', 'facility'], 'library'), (['college', 'library'], 'library'), (['where', 'can', 'i', 'get', 'books'], 'library'), (['book', 'facility'], 'library'), (['Where', 'is', 'library'], 'library'), (['Library'], 'library'), (['Library', 'information'], 'library'), (['Library', 'books', 'information'], 'library'), (['Tell', 'me', 'about', 'library'], 'library'), (['how', 'many', 'libraries'], 'library'), (['how', 'is', 'college', 'infrastructure'], 'infrastructure'), (['infrastructure'], 'infrastructure'), (['college', 'infrastructure'], 'infrastructure'), (['food', 'facilities'], 'canteen'), (['canteen', 'facilities'], 'canteen'), (['canteen', 'facility'], 'canteen'), (['is', 'there', 'any', 'canteen'], 'canteen'), (['Is', 'there', 'a', 'cafetaria', 'in', 'college'], 'canteen'), (['Does', 'college', 'have', 'canteen'], 'canteen'), (['Where', 'is', 'canteen'], 'canteen'), (['where', 'is', 'cafetaria'], 'canteen'), (['canteen'], 'canteen'), (['Food'], 'canteen'), (['Cafetaria'], 'canteen'), (['food', 'menu'], 'menu'), (['food', 'in', 'canteen'], 'menu'), (['Whats', 'there', 'on', 'menu'], 'menu'), (['what', 'is', 'available', 'in', 'college', 'canteen'], 'menu'), (['what', 'foods', 'can', 'we', 'get', 'in', 'college', 'canteen'], 'menu'), (['food', 'variety'], 'menu'), (['What', 'is', 'there', 'to', 'eat', '?'], 'menu'), (['What', 'is', 'college', 'placement'], 'placement'), (['Which', 'companies', 'visit', 'in', 'college'], 'placement'), (['What', 'is', 'average', 'package'], 'placement'), (['companies', 'visit'], 'placement'), (['package'], 'placement'), (['About', 'placement'], 'placement'), (['placement'], 'placement'), (['recruitment'], 'placement'), (['companies'], 'placement'), (['Who', 'is', 'HOD'], 'ithod'), (['Where', 'is', 'HOD'], 'ithod'), (['it', 'hod'], 'ithod'), (['name', 'of', 'it', 'hod'], 'ithod'), (['Who', 'is', 'computer', 'HOD'], 'computerhod'), (['Where', 'is', 'computer', 'HOD'], 'computerhod'), (['computer', 'hod'], 'computerhod'), (['name', 'of', 'computer', 'hod'], 'computerhod'), (['Who', 'is', 'extc', 'HOD'], 'extchod'), (['Where', 'is', 'extc', 'HOD'], 'extchod'), (['extc', 'hod'], 'extchod'), (['name', 'of', 'extc', 'hod'], 'extchod'), (['what', 'is', 'the', 'name', 'of', 'principal'], 'principal'), (['whatv', 'is', 'the', 'principal', 'name'], 'principal'), (['principal', 'name'], 'principal'), (['Who', 'is', 'college', 'principal'], 'principal'), (['Where', 'is', 'principal', "'s", 'office'], 'principal'), (['principal'], 'principal'), (['name', 'of', 'principal'], 'principal'), (['exam', 'dates'], 'sem'), (['exam', 'schedule'], 'sem'), (['When', 'is', 'semester', 'exam'], 'sem'), (['Semester', 'exam', 'timetable'], 'sem'), (['sem'], 'sem'), (['semester'], 'sem'), (['exam'], 'sem'), (['when', 'is', 'exam'], 'sem'), (['exam', 'timetable'], 'sem'), (['exam', 'dates'], 'sem'), (['when', 'is', 'semester'], 'sem'), (['what', 'is', 'the', 'process', 'of', 'admission'], 'admission'), (['what', 'is', 'the', 'admission', 'process'], 'admission'), (['How', 'to', 'take', 'admission', 'in', 'your', 'college'], 'admission'), (['What', 'is', 'the', 'process', 'for', 'admission'], 'admission'), (['admission'], 'admission'), (['admission', 'process'], 'admission'), (['scholarship'], 'scholarship'), (['Is', 'scholarship', 'available'], 'scholarship'), (['scholarship', 'engineering'], 'scholarship'), (['scholarship', 'it'], 'scholarship'), (['scholarship', 'ce'], 'scholarship'), (['scholarship', 'mechanical'], 'scholarship'), (['scholarship', 'civil'], 'scholarship'), (['scholarship', 'chemical'], 'scholarship'), (['scholarship', 'for', 'AI/ML'], 'scholarship'), (['available', 'scholarships'], 'scholarship'), (['scholarship', 'for', 'computer', 'engineering'], 'scholarship'), (['scholarship', 'for', 'IT', 'engineering'], 'scholarship'), (['scholarship', 'for', 'mechanical', 'engineering'], 'scholarship'), (['scholarship', 'for', 'civil', 'engineering'], 'scholarship'), (['scholarship', 'for', 'chemical', 'engineering'], 'scholarship'), (['list', 'of', 'scholarship'], 'scholarship'), (['comps', 'scholarship'], 'scholarship'), (['IT', 'scholarship'], 'scholarship'), (['mechanical', 'scholarship'], 'scholarship'), (['civil', 'scholarship'], 'scholarship'), (['chemical', 'scholarship'], 'scholarship'), (['automobile', 'scholarship'], 'scholarship'), (['first', 'year', 'scholarship'], 'scholarship'), (['second', 'year', 'scholarship'], 'scholarship'), (['third', 'year', 'scholarship'], 'scholarship'), (['fourth', 'year', 'scholarship'], 'scholarship'), (['What', 'facilities', 'college', 'provide'], 'facilities'), (['College', 'facility'], 'facilities'), (['What', 'are', 'college', 'facilities'], 'facilities'), (['facilities'], 'facilities'), (['facilities', 'provided'], 'facilities'), (['max', 'number', 'of', 'students'], 'college intake'), (['number', 'of', 'seats', 'per', 'branch'], 'college intake'), (['number', 'of', 'seats', 'in', 'each', 'branch'], 'college intake'), (['maximum', 'number', 'of', 'seats'], 'college intake'), (['maximum', 'students', 'intake'], 'college intake'), (['What', 'is', 'college', 'intake'], 'college intake'), (['how', 'many', 'stundent', 'are', 'taken', 'in', 'each', 'branch'], 'college intake'), (['seat', 'allotment'], 'college intake'), (['seats'], 'college intake'), (['college', 'dress', 'code'], 'uniform'), (['college', 'dresscode'], 'uniform'), (['what', 'is', 'the', 'uniform'], 'uniform'), (['can', 'we', 'wear', 'casuals'], 'uniform'), (['Does', 'college', 'have', 'an', 'uniform'], 'uniform'), (['Is', 'there', 'any', 'uniform'], 'uniform'), (['uniform'], 'uniform'), (['what', 'about', 'uniform'], 'uniform'), (['do', 'we', 'have', 'to', 'wear', 'uniform'], 'uniform'), (['what', 'are', 'the', 'different', 'committe', 'in', 'college'], 'committee'), (['different', 'committee', 'in', 'college'], 'committee'), (['Are', 'there', 'any', 'committee', 'in', 'college'], 'committee'), (['Give', 'me', 'committee', 'details'], 'committee'), (['committee'], 'committee'), (['how', 'many', 'committee', 'are', 'there', 'in', 'college'], 'committee'), (['I', 'love', 'you'], 'random'), (['Will', 'you', 'marry', 'me'], 'random'), (['Do', 'you', 'love', 'me'], 'random'), (['fuck'], 'swear'), (['bitch'], 'swear'), (['shut', 'up'], 'swear'), (['hell'], 'swear'), (['stupid'], 'swear'), (['idiot'], 'swear'), (['dumb', 'ass'], 'swear'), (['asshole'], 'swear'), (['fucker'], 'swear'), (['holidays'], 'vacation'), (['when', 'will', 'semester', 'starts'], 'vacation'), (['when', 'will', 'semester', 'end'], 'vacation'), (['when', 'is', 'the', 'holidays'], 'vacation'), (['list', 'of', 'holidays'], 'vacation'), (['Holiday', 'in', 'these', 'year'], 'vacation'), (['holiday', 'list'], 'vacation'), (['about', 'vacations'], 'vacation'), (['about', 'holidays'], 'vacation'), (['When', 'is', 'vacation'], 'vacation'), (['When', 'is', 'holidays'], 'vacation'), (['how', 'long', 'will', 'be', 'the', 'vacation'], 'vacation'), (['sports', 'and', 'games'], 'sports'), (['give', 'sports', 'details'], 'sports'), (['sports', 'infrastructure'], 'sports'), (['sports', 'facilities'], 'sports'), (['information', 'about', 'sports'], 'sports'), (['Sports', 'activities'], 'sports'), (['please', 'provide', 'sports', 'and', 'games', 'information'], 'sports'), (['okk'], 'salutaion'), (['okie'], 'salutaion'), (['nice', 'work'], 'salutaion'), (['well', 'done'], 'salutaion'), (['good', 'job'], 'salutaion'), (['thanks', 'for', 'the', 'help'], 'salutaion'), (['Thank', 'You'], 'salutaion'), (['its', 'ok'], 'salutaion'), (['Thanks'], 'salutaion'), (['Good', 'work'], 'salutaion'), (['k'], 'salutaion'), (['ok'], 'salutaion'), (['okay'], 'salutaion'), (['what', 'can', 'you', 'do'], 'task'), (['what', 'are', 'the', 'thing', 'you', 'can', 'do'], 'task'), (['things', 'you', 'can', 'do'], 'task'), (['what', 'can', 'u', 'do', 'for', 'me'], 'task'), (['how', 'u', 'can', 'help', 'me'], 'task'), (['why', 'i', 'should', 'use', 'you'], 'task'), (['ragging'], 'ragging'), (['is', 'ragging', 'practice', 'active', 'in', 'college'], 'ragging'), (['does', 'college', 'have', 'any', 'antiragging', 'facility'], 'ragging'), (['is', 'there', 'any', 'ragging', 'cases'], 'ragging'), (['is', 'ragging', 'done', 'here'], 'ragging'), (['ragging', 'against'], 'ragging'), (['antiragging', 'facility'], 'ragging'), (['ragging', 'juniors'], 'ragging'), (['ragging', 'history'], 'ragging'), (['ragging', 'incidents'], 'ragging'), (['hod'], 'hod'), (['hod', 'name'], 'hod'), (['who', 'is', 'the', 'hod'], 'hod')]

38 classes

['admission', 'canteen', 'college intake', 'committee', 'computerhod', 'course', 'creator', 'document', 'event', 'extchod', 'facilities', 'fees', 'floors', 'goodbye', 'greeting', 'hod', 'hostel', 'hours', 'infrastructure', 'ithod', 'library', 'location', 'menu', 'name', 'number', 'placement', 'principal', 'ragging', 'random', 'salutaion', 'scholarship', 'sem', 'sports', 'swear', 'syllabus', 'task', 'uniform', 'vacation']

263 unique lemmatized words

["'s", '(', ')', 'a', 'about', 'ac', 'active', 'activity', 'address', 'admision', 'admission', 'against', 'ai/ml', 'allotment', 'am', 'an', 'and', 'antiragging', 'any', 'anyone', 'are', 'as', 'asshole', 'at', 'attend', 'automobile', 'available', 'average', 'be', 'between', 'big', 'bitch', 'book', 'boy', 'branch', 'bring', 'building', 'by', 'bye', 'cafetaria', 'call', 'called', 'campus', 'can', 'canteen', 'capacity', 'case', 'casuals', 'ce', 'chatting', 'chemical', 'civil', 'code', 'college', 'come', 'committe', 'committee', 'comp', 'company', 'computer', 'conducted', 'contact', 'course', 'create', 'created', 'creator', 'cya', 'date', 'day', 'designed', 'detail', 'developer', 'different', 'distance', 'do', 'document', 'doe', 'done', 'dress', 'dresscode', 'dumb', 'during', 'each', 'eat', 'end', 'engineering', 'event', 'exam', 'extc', 'facility', 'far', 'fee', 'first', 'floor', 'food', 'for', 'fourth', 'from', 'fuck', 'fucker', 'function', 'game', 'get', 'girl', 'give', 'go', 'good', 'goodbye', 'got', 'gtg', 'guy', 'have', 'held', 'hell', 'hello', 'help', 'here', 'heyy', 'hi', 'history', 'hod', 'holiday', 'hostel', 'hour', 'how', 'i', 'idiot', 'in', 'incident', 'info', 'information', 'infrastructure', 'intake', 'is', 'it', 'job', 'junior', 'k', 'later', 'leaving', 'lecture', 'library', 'list', 'located', 'location', 'long', 'love', 'made', 'many', 'marry', 'max', 'maximum', 'me', 'mechanical', 'menu', 'more', 'much', 'my', 'name', 'need', 'needed', 'next', 'nice', 'no', 'non-ac', 'number', 'of', 'offer', 'offered', 'office', 'ok', 'okay', 'okie', 'okk', 'on', 'open', 'operation', 'organised', 'package', 'per', 'phone', 'placement', 'please', 'practice', 'principal', 'process', 'provide', 'provided', 'ragging', 'reach', 'recruitment', 'required', 'room', 'saturday', 'schedule', 'scholarship', 'seat', 'second', 'see', 'sem', 'semester', 'servive', 'should', 'shut', 'size', 'something', 'sport', 'start', 'student', 'stundent', 'stupid', 'syllabus', 'take', 'taken', 'taking', 'talk', 'tall', 'technology', 'telephone', 'tell', 'thank', 'thanks', 'the', 'there', 'these', 'thing', 'third', 'this', 'time', 'timetable', 'timing', 'to', 'ttyl', 'u', 'uni', 'uniform', 'univrsity', 'up', 'use', 'vacation', 'variety', 'visit', 'we', 'wear', 'well', 'what', 'whats', 'whatsup', 'whatv', 'when', 'where', 'wheres', 'which', 'who', 'whom', 'why', 'will', 'work', 'working', 'ya', 'year', 'you', 'your']

**Phase 3 Conclusion:**

In this phase we started to build a chatbot and summarized and discussed about it’s  preprossing and dataset chatbot not only used for texts it also used in various purspose like bussines applications,websites, applications, websites, mobile apps, or phone and upcoming phases we are going to build it advance level like featuring and evaluation,modeling.