

ChatterBot for Farmers

Kareena Arora

School of Computer Science and Engineering

Lovely Professional University, Phagwara

kareenaarora99@gmail.com

Abstract: — Artificial Intelligence and Machine Learning are driving the IT industry to a new scene. Ranchers comprise 54.6% of the Indian populace, yet acquire just 13.9% of the national GDP. This gross jumble can be eased by improving ranchers' entrance to data and master guidance. The chatbot can be portrayed as programming that can talk with individuals utilizing man-made reasoning. This product is utilized to perform errands, for example, rapidly reacting to clients, advising them, and assisting with buying items and offering better support to clients. The expression "ChatterBot" was initially instituted by Michael Mauldin (maker of the first Verbot) in 1994 to depict these conversational projects. Chatbots are right now being embraced in practically all fields. The undertaking is about the connection among ranchers and machines which can be gotten to from anyplace whenever. It gives different data about cultivating, agribusiness, pesticides, composts and a lot more which will assist the ranchers with growing harvests in a right way and to scale up with up and coming business sector patterns and innovations in an easy to use way. A chatbot is built utilizing characteristic language preparing with the assistance of AI calculation for preparing the bot. Python language is utilized for the advancement of the Chabot. The bot is prepared by data gathered from different locales.

Keywords: Artificial Intelligence (AI), Machine Learning (ML), Natural Language Processing (NLP), Chatbot.

1. Introduction

The universe of agribusiness business is getting progressively serious. In the present market, delivering an excellent item is just the principal half of the condition. The second is viable, effective showcasing, a now-essential segment of the formula for progress for most any cultivating business.

One instrument that has been causing a ripple effect in the advertising scene most as of late is man-made reasoning. Man-made intelligence is changing how organizations work over a wide assortment of ventures—and farming is no special case. On the off chance that you've at any point pondered about the potential uses of this kind of innovation for cultivating

organizations, it's time you investigated Facebook Messenger chatbots. We've gathered our best data, tips, and contextual investigations about chatbots for ranchers, so you can utilize this ground-breaking new device to help fabricate your horticulture business.

1.1.Chatbot

A chatbot is an artificial intelligence (AI) programming that can reproduce a discussion (or a visit) with a client in characteristic language through informing applications, sites, and versatile applications or through the phone. Chatbots are ordinarily utilized in exchange frameworks for different commonsense purposes including client care or data procurement. The expression "ChatterBot" was initially begotten by Michael Mauldin (maker of the first Verbot, Julia) in 1994 to depict these conversational projects. A chatbot is frequently depicted as one of the most exceptional and promising articulations of connection among people and machines. In any case, from an innovative perspective, a chatbot just speaks to the characteristic advancement of a Question Answering framework utilizing Natural Language Processing (NLP). Defining reactions to inquiries in normal language is one of the most run of the mill instances of Natural Language Processing applied in different endeavours' end-use applications.

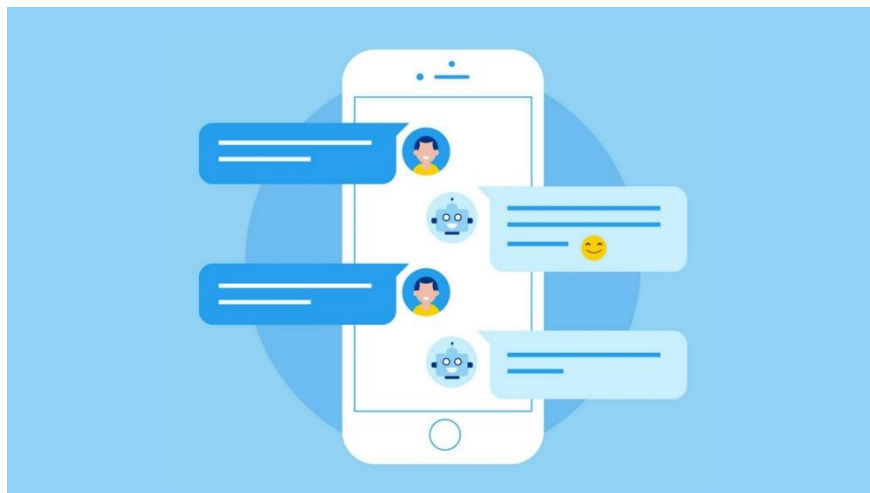


Fig. 1. Chatbot

1.2.Machine Learning

Machine Learning (ML) is the logical investigation of calculations and measurable models that PC frameworks use to play out a particular undertaking without utilizing unequivocal directions, depending on examples and deduction. It is viewed as a subset of man-made consciousness. AI calculations construct a scientific model dependent on test information, known as "preparing information", to settle on forecasts or choices without being unequivocally modified to play out the assignment. AI is firmly identified with computational measurements, which centers around making expectations utilizing PCs. The investigation of numerical streamlining conveys techniques, hypothesis and application areas to the field of AI. Information mining is a field of study inside AI and spotlights on exploratory information examination

through unaided learning. In its application across business issues, AI is additionally alluded to as prescient investigation.



Fig. 2. Machine Learning

2. Literature Review

Albayrak, N., Ozdemir, An., and Zeydan, E. depicted the chatbot and its different application. Ravi, R., in their paper gives us a survey of frameworks which utilizes AIML based chatbots to connect with the client.

Neelkumar P. Patel, Devangi R. Parikh, Prof. Darshan A. Patel, and Prof. Ronak R. Patel depicted about the AI and online human-like intuitive college chatbot (UNIBOT). A chatbot is the best instrument which furnishes a brisk approach to communicate with the clients. In their paper, insights concerning plan, calculation utilized and usage of the Unibot are introduced. The client doesn't have to accumulate data by visiting sites or universities.

Kohli, B., Choudhury, T., Sharma, S., and Kumar, P. have presented the examination of various chatbots and make our own customer utilizing python and electronic applications, up as far as we could possibly know, an activity to watch and clarify chatbot-human associations.

Mrs L. Kannagi, Ramya C., Shreya R., and Sowmiya R. cause ranchers to question about the farming get the reaction in content just as discourse and help in anticipating the future information of cost, with the goal that they can design their exercises.

3. How do Chatbots work?

There are broadly two variants of chatbots: **Rule-Based** and **Self-learning**.

3.1. In a *Rule-based methodology*, a bot responds to questions dependent on certain guidelines on which it is prepared on. The standards characterized can be easy to exceptionally perplexing. The bots can deal with basic inquiries yet neglect to oversee complex ones.

3.2. *Self-learning bots* are the ones that utilization some Machine Learning-based methodologies and are certainly more proficient than rule-based bots. These bots can be of further two sorts: Retrieval Based or Generative

i. In recovery-based models, a chatbot utilizes some heuristic to choose a reaction from a library of predefined reactions. The chatbot utilizes the message and setting of the discussion for choosing the best reaction from a predefined rundown of bot messages. The setting can remember a present situation for the discourse tree, every single past message in the discussion, recently spared factors (for example username). Heuristics for choosing a reaction can be built from multiple points of view, from rule-based if-else restrictive rationale to AI classifiers.

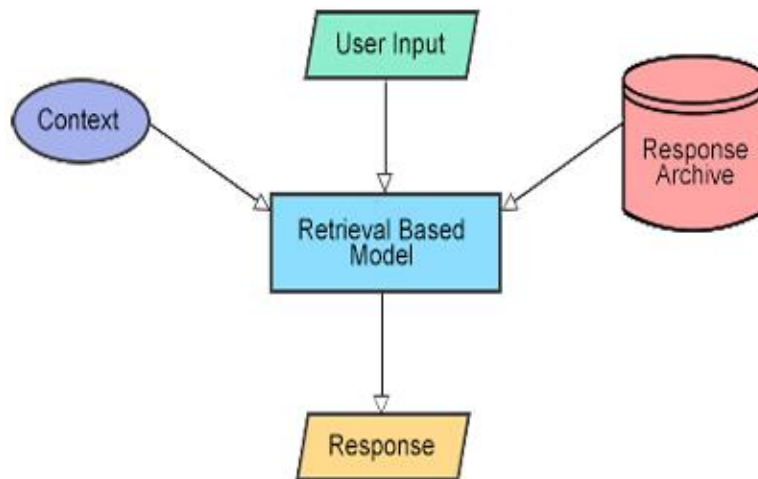


Fig. 3. Retrieval – based Models

ii. Generative bots can produce the appropriate responses and not generally answers with one of the appropriate responses from a lot of answers. This makes them progressively savvy as they take word by word from the inquiry and produces the appropriate responses.

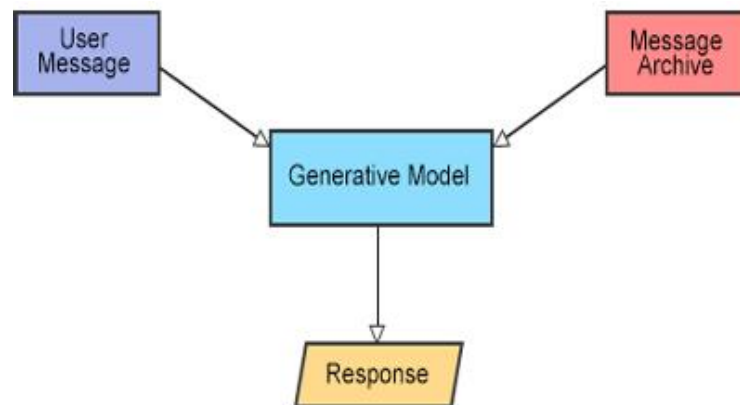


Fig. 4. Generative – based Models

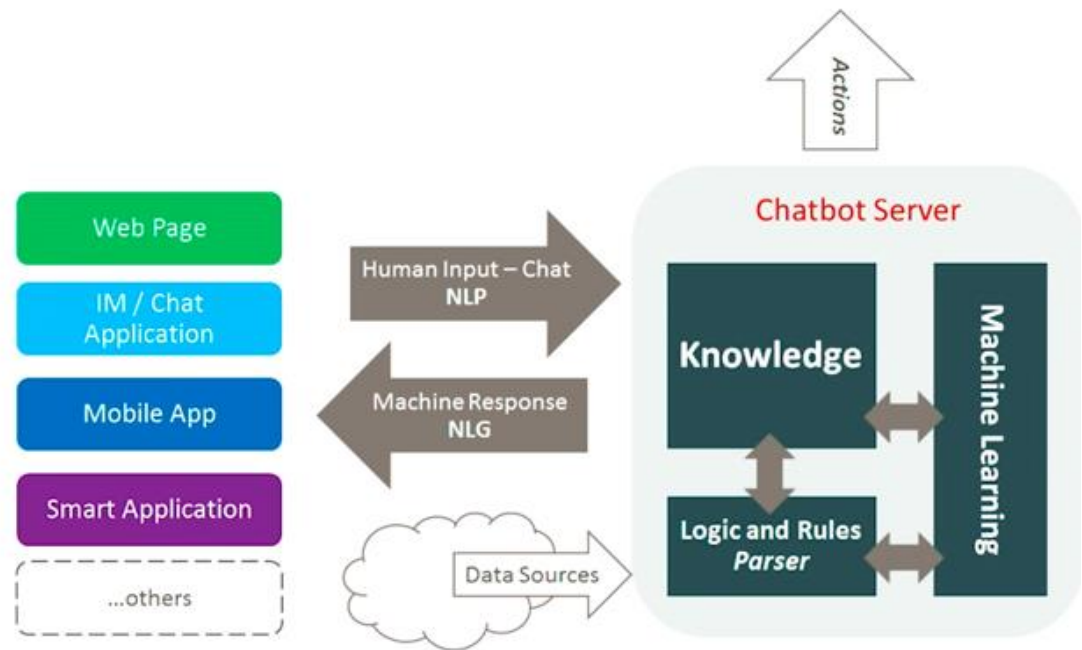


Fig. 5. Anatomy of a Chatbot

4. Proposed Methodology

The underneath figure shows the engineering chart of Chatbot. Here the client is a rancher who needs to think about cultivating, agribusiness and harvests.

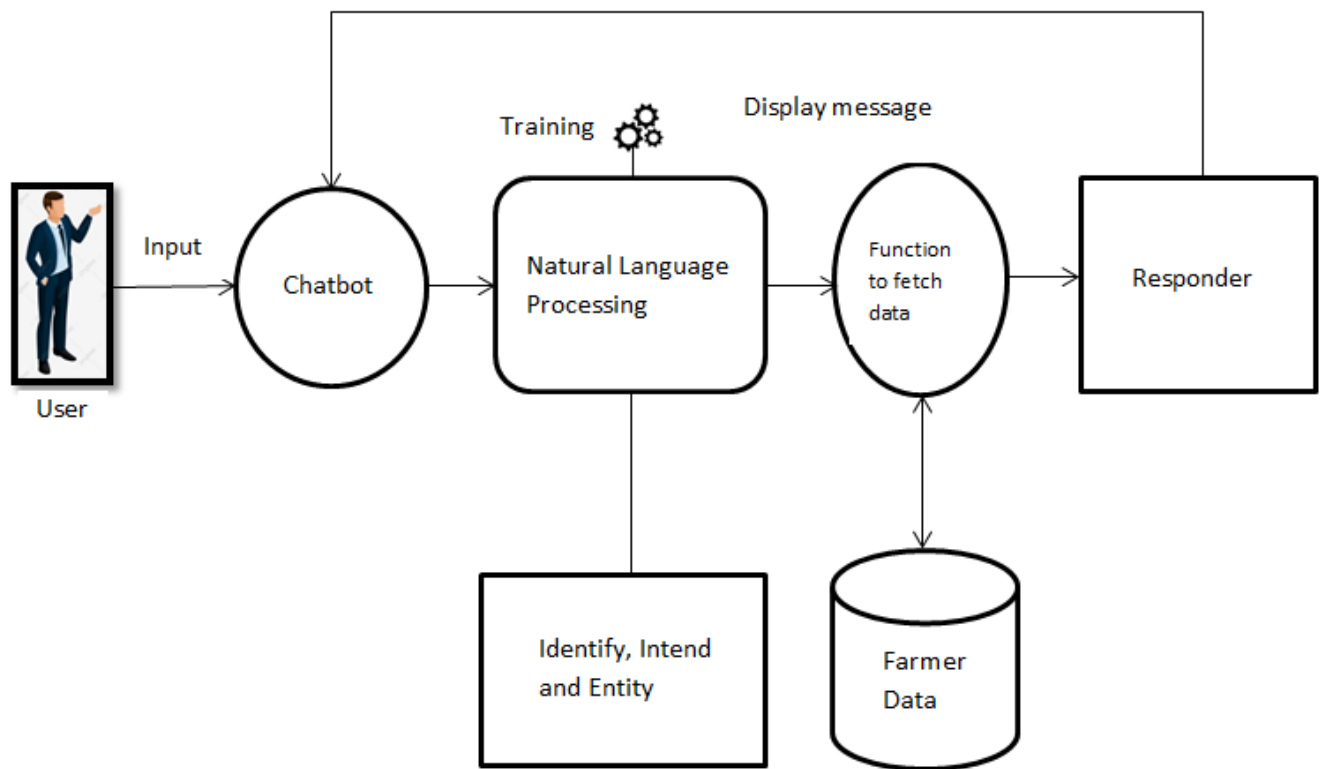


Fig. 6. Architecture Diagram of Chatbot

4.1. Pre-requisites

Hands-On information on scikit library and NLTK is accepted. Be that as it may, on the off chance that you are new to NLP, you can even now peruse the article and afterwards allude back to assets.

4.2. NLP

The field of study that centers around the communications between human language and PCs is called Natural Language Processing, or NLP for short. It sits at the crossing point of software engineering, computerized reasoning, and computational linguistics [Wikipedia]. NLP is a path for PCs to investigate, comprehend, and get significance from human language keenly and helpfully. By using NLP, engineers can sort out and structure information to perform undertakings, for example, programmed outline, interpretation, named substance acknowledgement, relationship extraction, estimation investigation, discourse acknowledgement, and point division.

4.3. NLTK: A Brief Intro

NLTK (Natural Language Toolkit) is the main stage for building Python projects to work with human language information. It gives simple to-utilize interfaces to more than 50 corpora and lexical assets, for example, WordNet, alongside a suite of content preparing libraries for ordering, tokenization, stemming, labelling, parsing, and semantic thinking, and wrappers for modern quality NLP libraries.

NLTK has been classified "a superb apparatus for instructing and working in, computational semantics utilizing Python," and "an astonishing library to play with characteristic language."

Regular Language Processing with Python gives a reasonable prologue to programming for language handling. I enthusiastically prescribe this book to individuals starting in NLP with Python.

4.4. Downloading and introducing NLTK

1. Install NLTK: run `pip install nltk`
2. Test establishment: run `python` at that point type `import nltk`

4.5. Introducing NLTK Packages

Import NLTK and run `nltk.download()`. This will open the NLTK downloader from where you can pick the corpora and models to download. You can likewise download all bundles on the double.

4.6. Text Pre-Processing with NLTK

The fundamental issue with content information is that it is all in the content arrangement (strings). Be that as it may, Machine learning calculations need a type of numerical element vector to

play out the errand. So before we start with any NLP venture we have to pre-process it to make it perfect for work. Essential content pre-preparing incorporates:

- Converting the whole content into capitalized or lowercase, with the goal that the calculation doesn't treat indistinguishable words in various cases from various.
- *Tokenization*: Tokenization is only the term used to depict the way toward changing over the ordinary content strings into a rundown of tokens i.e. words that we really need. Sentence tokenizer can be utilized to discover the rundown of sentences and Word tokenizer can be utilized to discover the rundown of words in strings.

The NLTK information bundle incorporates a pre-prepared Punkt tokenizer for English.

- Removing Noise i.e. everything that isn't in a standard number or letter.
- *Removing Stop words*: Once in a while, some amazingly basic words which would seem, by all accounts, to be of little an incentive in choosing reports coordinating a client need are barred from the jargon completely. These words are called stop words.
- *Stemming*: Stemming is the way toward diminishing curved (or now and again inferred) words to their stem, base or root structure — for the most parts a composed word structure. Model if we somehow happened to stem the accompanying words: "Stems", "Stemming", "Stemmed", "and Stementization", the outcome would be a solitary word "stem".
- *Lemmatization*: A slight variation of stemming is lemmatization. The significant contrast between these is that stemming can frequently make non-existent words, though lemmas are real words. In this way, your root stem, which means the word you end up with, isn't something you can simply gaze upward in a word reference, yet you can look into a lemma. Instances of Lemmatization are that "run" is a base structure for words like "running" or "ran" or that "better" and "great" are in a similar lemma so they are viewed as the equivalent.

4.7. Bag of Words

After the underlying preprocessing stage, we have to change the content into an important vector (or cluster) of numbers. The pack of-words is a portrayal of content that depicts the event of words inside a report. It includes two things:

- A jargon of known words.
- A proportion of the nearness of known words.

For what reason is it known as a "bag" of words? That is because any data about the request or structure of words in the report is disposed of and the model is just worried about whether the realized words happen in the archive, not where they happen in the record.

The instinct behind the Bag of Words is that archives are comparable on the off chance that they have comparable substance. Likewise, we can pick up something about the significance of the archive from its substance alone.

For instance, if our word reference contains the words {Learning, is, the, not, great}, and we need to vectorize the content "Learning is incredible", we would have the accompanying vector: (1, 1, 0, 0, 1).

4.8. TF-IDF Approach

An issue with the Bag of Words approach is that exceptionally visit words begin to command in the report (for example bigger score), yet may not contain as much "instructive substance". Likewise, it will give more weight to longer records than shorter archives.

One methodology is to rescale the recurrence of words by how regularly they show up in all reports with the goal that the scores for visit words like "the" that likewise visit overall records are punished. This way to deal with scoring is called Term Frequency-Inverse Document Frequency or TF-IDF for short, where:

Term Frequency: is a scoring of the recurrence of the word in the present record.

$$TF = (\text{Number of times term } t \text{ shows up in a record}) / (\text{Number of terms in the report})$$

Inverse Document Frequency: is a scoring of how uncommon the word is across reports.

$IDF = 1 + \log (N/n)$, where, N is the number of reports and n is the number of archives a term t has shown up in.

Tf-IDF weight is a weight regularly utilized in data recovery and content mining. This weight is a factual measure used to assess how significant a word is to a report in an assortment or corpus.

4.9. Cosine Similarity

TF-IDF is a change applied to writings to get two genuine esteemed vectors in vector space. We would then be able to acquire the Cosine likeness of any pair of vectors by taking their speck item and partitioning that by the result of their standards. That yields the cosine of the point between the vectors. Cosine closeness is a proportion of comparability between two non-zero vectors. Utilizing this equation we can discover the closeness between any two archives $d1$ and $d2$.

$$\text{Cosine Similarity } (d1, d2) = \text{Dot product } (d1, d2) / ||d1|| * ||d2||$$

where $d1, d2$ are two non-zero vectors.

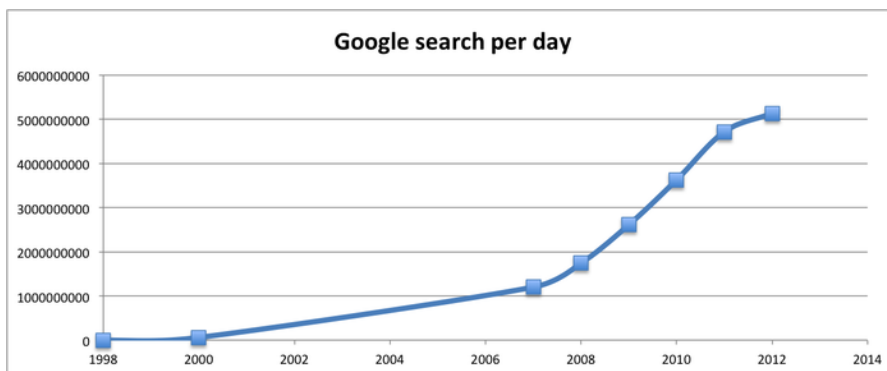


Fig. 7.

4.10. Importing the necessary libraries

```
In [1]: import numpy as np
import random
import string # to process standard python strings
```

```
In [2]: import nltk
```

4.11. Corpus

For our model, we will utilize the Wikipedia page for chatbots as our corpus. Duplicate the substance from the page and spot it in a book document named 'chatbot.txt'. Be that as it may, you can utilize any corpus of your decision.

4.12. Reading in the data

We will peruse in the corpus.txt document and convert the whole corpus into a rundown of sentences and a rundown of words for additional pre-handling.

```
In [5]: f=open('chatbot.txt','r',errors = 'ignore')
raw=f.read()
raw=raw.lower() # converts to lowercase
# nltk.download('punkt') # first-time use only
# nltk.download('wordnet') # first-time use only
nltk.download('popular',quiet=True)
sent_tokens = nltk.sent_tokenize(raw) # converts to list of sentences
word_tokens = nltk.word_tokenize(raw) # converts to list of words
```

Let see an example of the sent_tokens and the word_tokens

```
In [6]: sent_tokens[:2]
```

```
Out[6]: ['chatbot\na chatbot is a piece of software that conducts a conversation via auditory or textual methods.',
'agriculture\nagriculture is the process of producing food, feed, fiber and many other desired products by the cultivation of certain plants and the raising of domesticated animals (livestock).']
```

```
In [7]: word_tokens[:2]
```

```
Out[7]: ['chatbot', 'a']
```

4.13. Pre-processing the raw text

We will currently characterize a capacity called LemTokens which will take as info the tokens and return standardized tokens.

```
In [8]: lemmmer = nltk.stem.WordNetLemmatizer()
#WordNet is a semantically-oriented dictionary of English included in NLTK.
def LemTokens(tokens):
    return [lemmer.lemmatize(token) for token in tokens]
remove_punct_dict = dict((ord(punct), None) for punct in string.punctuation)
def LemNormalize(text):
    return LemTokens(nltk.word_tokenize(text.lower().translate(remove_punct_dict)))
```

4.14. Keyword matching

Next, we will characterize a capacity for a welcome by the bot i.e. if a client's information is welcome, the bot will restore a welcome response. ELIZA utilizes a basic catchphrase coordinating for the welcome. We will use a similar idea here.

```
In [9]: GREETING_INPUTS = ("hello", "hi", "greetings", "sup", "hey")
GREETING_RESPONSES = ["hi", "hey", "*nods*", "hi there", "hello",
                        "I am glad! You are talking to me"]

def greeting(sentence):

    for word in sentence.split():
        if word.lower() in GREETING_INPUTS:
            return random.choice(GREETING_RESPONSES)
```

4.15. Generating Response

To produce a reaction from our bot for input questions, the idea of record similitude will be utilized. So we start by bringing in the essential modules. From sci-kit learn library, import the Tfidf vectorizer to convert a collection of raw documents to a matrix of TF-IDF features.

```
In [10]: from sklearn.feature_extraction.text import TfidfVectorizer
```

Also, import cosine similarity module from sci-kit learn library

```
In [12]: from sklearn.metrics.pairwise import cosine_similarity
```

This will be utilized to discover the similitude between words entered by the client and the words in the corpus. This is the easiest conceivable execution of a chatbot.

We characterize a capacity reaction which scans the client's expression for at least one known watchwords and returns one of a few potential reactions. On the off chance that it doesn't discover the information coordinating any of the catchphrases, it restores a reaction: "I am grieved! I don't get you".

```
In [13]: def response(user_response):
    robo_response=''
    sent_tokens.append(user_response)
    TfidfVec = TfidfVectorizer(tokenizer=LemNormalize, stop_words='english')
    tfidf = TfidfVec.fit_transform(sent_tokens)
    vals = cosine_similarity(tfidf[-1], tfidf)
    idx=vals.argsort()[0][-2]
    flat = vals.flatten()
    flat.sort()
    req_tfidf = flat[-2]
    if(req_tfidf==0):
        robo_response=robo_response+"I am sorry! I don't understand you"
        return robo_response
    else:
        robo_response = robo_response+sent_tokens[idx]
        return robo_response
```

At long last, we will take care of the lines that we need our bot to the state while beginning and closure a discussion relying on the client's information.

```
In [*]: flag=True
print("ROBO: My name is Robo. I will answer your queries about Chatbots. If you want to exit, type Bye!")
while(flag==True):
    user_response = input()
    user_response=user_response.lower()
    if(user_response!='bye'):
        if(user_response=='thanks' or user_response=='thank you' or user_response=='thank u'):
            flag=False
            print("ROBO: You are welcome..")
        else:
            if(greeting(user_response)!=None):
                print("ROBO: "+greeting(user_response))
            else:
                print("ROBO: ",end="")
                print(response(user_response))
                sent_tokens.remove(user_response)
    else:
        flag=False
        print("ROBO: Bye! take care..")
```

ROBO: My name is Robo. I will answer your queries about Chatbots. If you want to exit, type Bye!

5. Result and Discussion

The farmer can effectively pose inquiries to the bot as they ask different people. The bot will answer as indicated by the information accessible in the corpus document. We can without much of a stretch include information in the content document and on the base of that ranchers can extricate the data. This strategy is truly financially savvy.

```
ROBO: My name is Robo. I will answer your queries about Chatbots. If you
want to exit, type Bye!
hi
ROBO: I am glad! You are talking to me
what is chatbot?
ROBO: chatbot
a chatbot is a piece of software that conducts a conversation via audio
ry or textual methods.
```

whata is agriculture?

ROBO: agriculture

agriculture is the process of producing food, feed, fiber and many other desired products by the cultivation of certain plants and the raising of domesticated animals (livestock).

can u tell me about pesticides?

ROBO: pesticide

pesticides are substances that are meant to control pests, including weeds the term pesticide includes all of the following: herbicide, insecticides (which may include insect growth regulators, termiticides, etc) nematocide, molluscicide, piscicide, avicide, rodenticide, bactericide, insect repellent, animal repellent, antimicrobial, and fungicide the most common of these are herbicides which account for approximately 80% of all pesticide use most pesticides are intended to serve as plant protection products (also known as crop protection products), which in general, protect plants from weeds, fungi, or insects in general, a pesticide is a chemical[vague] or biological agent (such as a virus, bacterium, or fungus) that deters, incapacitates, kills, or otherwise discourages pests target pests can include insects, plant pathogens, weeds, molluscs, birds, mammals, fish, nematodes (roundworms), and microbes that destroy property, cause nuisance, or spread disease, or are disease vectors along with these benefits, pesticides also have drawbacks, such as potential toxicity to humans and other species.

Thank you

ROBO: You are welcome..

6. Conclusion

A chatbot is a rising pattern and chatbot expands the viability of the business by giving a superior encounter minimal effort. A straightforward chatbot is not a provoking assignment when contrasted with complex chatbots and engineers ought to comprehend and think about the strength, versatility and adaptability issues alongside an elevated level of aim on human language. To put it plainly, Chatbot is the environment and moving very quickly and with the progression of time, new highlights are included in the current stage.

Chatbot applications streamline connections among individuals and administrations, improving client experience. Simultaneously, they offer organizations new chances to improve the client's commitment process and operational effectiveness by decreasing the regular expense of client care.

Right now, propose a chatbot that would empower bot users to simply type in the inquiry identified with cultivating and will get a reaction right away. It is useful to the clients as it permits to enter inquiries in characteristic language and wanted data is acquired effectively to the client. Tests were led to comprehend the presentation of the apparatus. The apparatus was assessed dependent on the nature of the reaction and it performed well. A rancher needs not to rely upon some other individual to get the hang of something about harvests or the strategies for cultivating. They can undoubtedly contact the bot for any data. By changing the content document based on specialists, it will be progressively useful to ranchers.

References

- 1) Albayrak, N., Ozdemir, A., & Zeydan, E., *An overview of artificial intelligence-based chatbots and an example chatbot application*, 2018.
- 2) Ravi, R., *Intelligent Chatbot for Easy Web-Analytics Insights*, 2018.
- 3) Neelkumar P. Patel, Devangi R. Parikh, Prof. Darshan A. Patel, & Prof. Ronak R. Patel, *AI and Web-Based Human-Like Interactive University Chatbot (UNIBOT)*, 2019.
- 4) Kohli, B., Choudhury, T., Sharma, S., & Kumar, P., *A Platform for Human-Chatbot Interaction Using Python*, 2018.
- 5) V, B. R., J, S. A., Manilal, M., Mathew, A. A., & Babu, B., *A Graph-Based Chatbot for Cancer Patients*, 2019.
- 6) Rahman, A. M., Mamun, A. A., & Islam, A., *Programming challenges of chatbot: Current and future perspective*, 2017.
- 7) N. Harid, A. C. Bogias, H. Griffiths, S. Robson, & A. Haddad, *A Wireless System for Monitoring Leakage Current in Electrical Substation Equipment*, 2016.
- 8) Shanmugam, U., Mani, S., Sivakumar, S., & Rajeswari, P., *Human-Computer text conversation through NLP in Tamil using Internet Recognition*, 2019.
- 9) Mrs V. Selvi, M.E. Ms Saranya S., Ms Chidida K., Ms Abrana R., *Chatbot and bully-free chat*, 2019.
- 10) Bayan Abu Shawar, Eric Atwell, *Chatbots: Are they Really Useful?*, 2017.
- 11) J. Vijayalakshmi, K. PandiMeena, *Agriculture TalkBot Using AI*, 2019.
- 12) Mohit Jain, Pratyush Kumar, Ishita Bhansali, Q. Vera Liao, Khai Truong, & Shwetak Patel, *FarmChat: A Conversational Agent to Answer Farmer Queries*, 2018.
- 13) Mrs L. Kannagi, Ramya C., Shreya R., & Sowmiya R., *Virtual Conversation Assistant- "The FARMBOT"*, 2018.
- 14) Gustavo Marques Mostaco, Leonardo Barreto Campos, & Carlos Eduardo Cugnasca, *AgronomotBot: a smart answering Chatbot applied to agricultural sensor networks*, 2018.
- 15) Madhu, D., Jain, C. J. N., Sebastain, E., Shaji, S., & Ajaykumar, A., *A novel approach for medical assistance using trained chatbot*, 2017.