

# chatbot

March 4, 2024

## 1 Question & Answering Chatbot

Python 3.7.13

[ ]:

[ ]:

[ ]: `!python3.7 -m pip install chatterbot==1.0.4 pytz`

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: chatterbot==1.0.4 in
/home/rohi/.local/lib/python3.7/site-packages (1.0.4)
Requirement already satisfied: pytz in /home/rohi/.local/lib/python3.7/site-
packages (2024.1)
Requirement already satisfied: chatterbot-corpus<1.3,>=1.2 in
/home/rohi/.local/lib/python3.7/site-packages (from chatterbot==1.0.4) (1.2.0)
Requirement already satisfied: mathparse<0.2,>=0.1 in
/home/rohi/.local/lib/python3.7/site-packages (from chatterbot==1.0.4) (0.1.2)
Requirement already satisfied: nltk<4.0,>=3.2 in
/home/rohi/.local/lib/python3.7/site-packages (from chatterbot==1.0.4) (3.8.1)
Requirement already satisfied: pymongo<4.0,>=3.3 in
/home/rohi/.local/lib/python3.7/site-packages (from chatterbot==1.0.4) (3.13.0)
Requirement already satisfied: python-dateutil<2.8,>=2.7 in
/home/rohi/.local/lib/python3.7/site-packages (from chatterbot==1.0.4) (2.7.5)
Requirement already satisfied: sqlalchemy<1.3,>=1.2 in
/home/rohi/.local/lib/python3.7/site-packages (from chatterbot==1.0.4) (1.2.19)
Requirement already satisfied: pint>=0.8.1 in
/home/rohi/.local/lib/python3.7/site-packages (from chatterbot==1.0.4) (0.18)
Requirement already satisfied: PyYAML<4.0,>=3.12 in
/home/rohi/.local/lib/python3.7/site-packages (from chatterbot-
corpus<1.3,>=1.2->chatterbot==1.0.4) (3.13)
Requirement already satisfied: click in /home/rohi/.local/lib/python3.7/site-
packages (from nltk<4.0,>=3.2->chatterbot==1.0.4) (8.1.7)
Requirement already satisfied: joblib in /home/rohi/.local/lib/python3.7/site-
packages (from nltk<4.0,>=3.2->chatterbot==1.0.4) (1.3.2)
Requirement already satisfied: regex>=2021.8.3 in
/home/rohi/.local/lib/python3.7/site-packages (from
nltk<4.0,>=3.2->chatterbot==1.0.4) (2023.12.25)
```

Requirement already satisfied: tqdm in /home/rohi/.local/lib/python3.7/site-packages (from nltk<4.0,>=3.2->chatterbot==1.0.4) (4.66.2)

Requirement already satisfied: packaging in /home/rohi/.local/lib/python3.7/site-packages (from pint>=0.8.1->chatterbot==1.0.4) (23.2)

Requirement already satisfied: importlib-metadata in /home/rohi/.local/lib/python3.7/site-packages (from pint>=0.8.1->chatterbot==1.0.4) (6.7.0)

Requirement already satisfied: six>=1.5 in /home/rohi/.local/lib/python3.7/site-packages (from python-dateutil<2.8,>=2.7->chatterbot==1.0.4) (1.16.0)

Requirement already satisfied: zipp>=0.5 in /home/rohi/.local/lib/python3.7/site-packages (from importlib-metadata->pint>=0.8.1->chatterbot==1.0.4) (3.15.0)

Requirement already satisfied: typing-extensions>=3.6.4 in /home/rohi/.local/lib/python3.7/site-packages (from importlib-metadata->pint>=0.8.1->chatterbot==1.0.4) (4.7.1)

**Simple Chatterbot Example** We first import modules and libraries we need, such as ChatBot and ListTrainer from chatterbot. Then we create an instance of the ChatBot class and store it into a variable (chatbot here). To train our chatbot to give appropriate response, we start with small amount of data. We simply write a conditional statement using while loop that the loop will keep going as long as user have not entered special character or string via input, in the loop, user type their message, we store it in a variable named 'query' check if the user did or did not use the special character we defined to exit the loop, if they did loop will break and our program ends, if not, we call 'get\_response' function from our chatbot instance earlier and pass our 'query' which is the user message to the function and what it will return is the response to the message, the accuracy and quality of the responses depends on the quality and the amount of the data our chatbot got trained with.

By running the next block, ChatterBot might download some data and language models associated with NLTK.

```
[ ]: from chatterbot import ChatBot
from chatterbot.trainers import ListTrainer

chatbot = ChatBot("Rohi&Bahadir")

trainer = ListTrainer(chatbot)

trainer.train([
    "Hi, can I help you?",
    "Sure, I'd like to book a flight to Iceland.",
    "Your flight has been booked."
])

exit_conditions = (":q", "quit", "exit")
while True:
    query = input(">>> ")
```

```

print(f"You: {query}")
if query in exit_conditions:
    break
else:
    print(f"Bot: {chatbot.get_response(query)}")

```

List Trainer: [#####] 100%

```

[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]      /home/rohi/nltk_data...
[nltk_data] Package averaged_perceptron_tagger is already up-to-
[nltk_data]      date!
[nltk_data] Downloading package punkt to /home/rohi/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /home/rohi/nltk_data...
[nltk_data] Package stopwords is already up-to-date!

```

```

You: Hi
Bot: hi
You: How are you doing?
Bot: q:
You: I need to book flight
Bot: Your flight has been booked.
You: :q

```

As we said earlier, the chatbot did not trained well yet, has very limited responses, but, by keep running it and have conversation with it, it do remember your previous conversation.

Chatterbot require database to store all inputs and connect them with possible responses, to do that, by default, it uses SQLite database file.

Chatterbot will create a file named “db.sqlite3” to mainly store all inputs and possible corresponding answers, and two other with same name expect they ending with “wal” and “-shm” which they are temporary support files.

## 1.1 Train chatterbot

### 1.1.1 Preparing Data for Training

in this part, we want to read out data set file using pandas, to do that we first need to install pandas module in our environment, and that import it.

```
[ ]: import pandas as pd
```

in this case, we used a Mental Health FAQ dataset, containing 98 questions and their corresponding answers.

```
[ ]: df = pd.read_csv('data/Mental_Health_FAQ.csv') #from kaggle
df
```

```
[ ]:      Question_ID      Questions \
0      1590140      What does it mean to have a mental illness?
1      2110618      Who does mental illness affect?
2      6361820      What causes mental illness?
3      9434130      What are some of the warning signs of mental i...
4      7657263      Can people with mental illness recover?
..      ...      ...
93     4373204      How do I know if I'm drinking too much?
94     7807643      If cannabis is dangerous, why are we legalizin...
95     4352464      How can I convince my kids not to use drugs?
96     6521784      What is the legal status (and evidence) of CBD...
97     3221856      What is the evidence on vaping?

                                Answers
0      Mental illnesses are health conditions that di...
1      It is estimated that mental illness affects 1 ...
2      It is estimated that mental illness affects 1 ...
3      Symptoms of mental health disorders vary depen...
4      When healing from mental illness, early identi...
..      ...
93     Sorting out if you are drinking too much can b...
94     Cannabis smoke, for example, contains cancer-c...
95     You can't. But you can influence their capacit...
96     Cannabidiol or CBD is a naturally occurring co...
97     "Vaping" is the term for using a device where ...

[98 rows x 3 columns]
```

**Text cleaning and Normalization** Python is case sensitive programming language, which means the string 'Hello' and 'helLO' are two different words, so to solve this problem we make all the strings to lower case, and remove all punctuation (using a string substitution function from "re" module, this function searches for all occurrences of a pattern in a string and replaces them with a specified replacement string. If the pattern is not found, the string is returned unchanged) by writing the function 'normalize\_text' (the name of the function could be anything) and then we store it in a variable and return it from the function.

In the next step, we use our function to apply those changes in our pandas DataFrame using ".apply()" function from pandas library, it allows you to apply the function along an axis of the DataFrame or to all elements of a Series

```
[ ]: import re

# Function to normalize text
def normalize_text(text):
    # Convert text to lowercase
    text = text.lower()
    # Remove punctuation
    text = re.sub(r'[\w\s]', '', text)
```

```

    return text

# Normalize the 'Questions' and 'Answers' columns
df['Questions'] = df['Questions'].apply(normalize_text)
df['Answers'] = df['Answers'].apply(normalize_text)

# Display the first few rows to verify changes
df.head()

```

```

[ ]:   Question_ID      Questions \
0      1590140      what does it mean to have a mental illness
1      2110618      who does mental illness affect
2      6361820      what causes mental illness
3      9434130  what are some of the warning signs of mental i...
4      7657263      can people with mental illness recover

      Answers
0  mental illnesses are health conditions that di...
1  it is estimated that mental illness affects 1 ...
2  it is estimated that mental illness affects 1 ...
3  symptoms of mental health disorders vary depen...
4  when healing from mental illness early identif...

```

### 1.1.2 Removing Duplicates and Missing Values

To check if our DataFrame has any missing or duplicated value, we can use “`isnull()`” function to count any missing value, and “`duplicated()`” for any duplicate data in DataFrame

```

[ ]: missing_values = df.isnull().sum()
      duplicate_questions = df['Questions'].duplicated().sum()

```

As we can see, our DataFrame has no missing or duplicated value

```

[ ]: missing_values, duplicate_questions

```

```

[ ]: (Question_ID      0
      Questions      0
      Answers        0
      dtype: int64,
      1)

```

```

[ ]: from chatterbot import ChatBot
      chatbot = ChatBot("rohi&bahadir")

```

```

[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data] /home/rohi/nltk_data...
[nltk_data] Package averaged_perceptron_tagger is already up-to-
[nltk_data] date!
[nltk_data] Downloading package punkt to /home/rohi/nltk_data...

```

```
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /home/rohi/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

```
[ ]: from chatterbot.trainers import ChatterBotCorpusTrainer
```

```
trainer = ChatterBotCorpusTrainer(chatbot)
trainer.train("chatterbot.corpus.english")
```

```
Training ai.yml: [#####] 30%
Training ai.yml: [#####] 100%
Training botprofile.yml: [#####] 100%
Training computers.yml: [#####] 100%
Training conversations.yml: [#####] 100%
Training emotion.yml: [#####] 100%
Training food.yml: [#####] 100%
Training gossip.yml: [#####] 100%
Training greetings.yml: [#####] 100%
Training health.yml: [#####] 100%
Training history.yml: [#####] 100%
Training humor.yml: [#####] 100%
Training literature.yml: [#####] 100%
Training money.yml: [#####] 100%
Training movies.yml: [#####] 100%
Training politics.yml: [#####] 100%
Training psychology.yml: [#####] 100%
Training science.yml: [#####] 100%
Training sports.yml: [#####] 100%
Training trivia.yml: [#####] 100%
```

```
[ ]: from chatterbot.trainers import ListTrainer
```

```
training_data = df.apply(lambda row: [row['Questions'], row['Answers']],
    ↪axis=1).tolist()

flattened_training_data = [item for pair in training_data for item in pair]

trainer = ListTrainer(chatbot)

trainer.train(flattened_training_data)
```

```
List Trainer: [#####] 100%
```

```
[ ]: exit_conditions = (":q", "quit", "exit")
while True:
    query = input(">>> ")
    print(f"You: {query}")
    if query in exit_conditions:
```

```
        break
    else:
        print(f"Bot: {chatbot.get_response(query)}")
```

You: what are you?

Bot: I am just an artificial intelligence.

You: What's the difference between anxiety and stress?

Bot: stress and anxiety are often used interchangeably and there is overlap between stress and anxiety stress is related to the same fight flight or freeze response as anxiety and the physical sensations of anxiety and stress may be very similar the cause of stress and anxiety are usually different however stress focuses on mainly external pressures on us that were finding hard to cope with when we are stressed we usually know what were stressed about and the symptoms of stress typically disappear after the stressful situation is over anxiety on the other hand isnt always as easy to figure out anxiety focuses on worries or fears about things that could threaten us as well as anxiety about the anxiety itself stress and anxiety are both part of being human but both can be problems if they last for a long time or have an impact on our wellbeing or daily life

You: What does rapid cycling mean?

Bot: rapid cycling means that someone diagnosed with bipolar disorder or depression experiences four or more episodes of depression and/or mania in one year rapid cycling can happen any time someone experiences bipolar disorder about of people diagnosed with bipolar disorder experience rapid cycling at some point in many cases rapid cycling eventually goes away on its own and people return to a pattern of longer and less frequent episodes

You: What is cyclothymic disorder?

Bot: cyclothymic disorder is a subtype of bipolar disorder much like bipolar disorder the symptoms of cyclothymia include three or more symptoms of hypomania and five or more symptoms of depression like bipolar disorder people may experience wellness between episodes of hypomania and depression symptoms of hypomania include symptoms of depression include talk to a doctor or mental health professional if you think you might have cyclothymic disorder

You: :q

[ ]: