Brianna Murray bnm180003
CS 4395.001

MoodyBot

System Description

My chatbot is called MoodyBot. MoodyBot gets its name from the sentiment analysis feature that comes into play as the user converses with the bot. It has a static knowledge base that is created upon new user integration from a topic prompt provided by the user. When the user gives a prompt, MoodyBot sends that query to ChatGPT to retrieve 10 wikipedia links on the topic. After the links are retrieved, the bot webscrapes each link. During the **web scraping** a .txt file for each link is created and named accordingly using **Regex** while also being filled with the information from the given wikipedia page via **BeautifulSoup**. Once all .txt files are filled with information, stopwords are removed and the bot tokenizes the documents. Then the bot will find the most popular terms from each document via **term frequency** and also **inverse document frequency**. Those gathered words are then saved to a list of dictionaries all containing the frequency variable. From that list, the top 5 words of each document are selected and saved as the "patterns" within a .json file for the current user. The .json "responses" are all sentences containing the top words from each file. So each document has 5 top words being saved to the .json file with every sentence containing those words as its response. These are all list values. Also, default intents are also added to the .json file before official creation. Once that is complete, a system command line executes and calls the machine learning part of the bot. The bot uses **TensorFlow** to create a model based on the .json file in order to retrieve the most accurate responses to a user based on the knowledge base. The input layer is the .json file while the output layer is the created model. TensorFlow creates a bag of words by **stemming** from the provided strings and then converts those into a numeric value via **numpy**. When training TensorFlow creates accurate connections between the created bag of words and the actual .json tags/patterns. The model also makes a **pickle** file containing all of the data as arrays so that it is easy for the model to train on. Once the model is created, the chat conversation can begin.

When the user model has been created MoodyBot will greet the user by name and pull up their knowledge base file on the user containing their given topic of choice. Whenever the user asks for a fact the chatbot will use **cosine similarity** from **sklearn** to find the best fitting fact to present to the user from the knowledge base. First the model will find the appropriate tag from the user's prompt, then from cosine similarity the chatbot will find the best response fact from the tagged category to give back to the user. While conversing with the user each time a fact from the knowledge base is provided, the chatbot will also tell the user if the fact being shared is a simple or complex sentence based on the number of clauses within the sentence. This is done by

using **parts of speech tagging** with **NLTK** on the sentence and then splitting the sentence whenever a verb phrase is found. If more than one verb phrase exists in the sentence the chatbot will display all of the clauses within the sentence and inform the user of its findings.

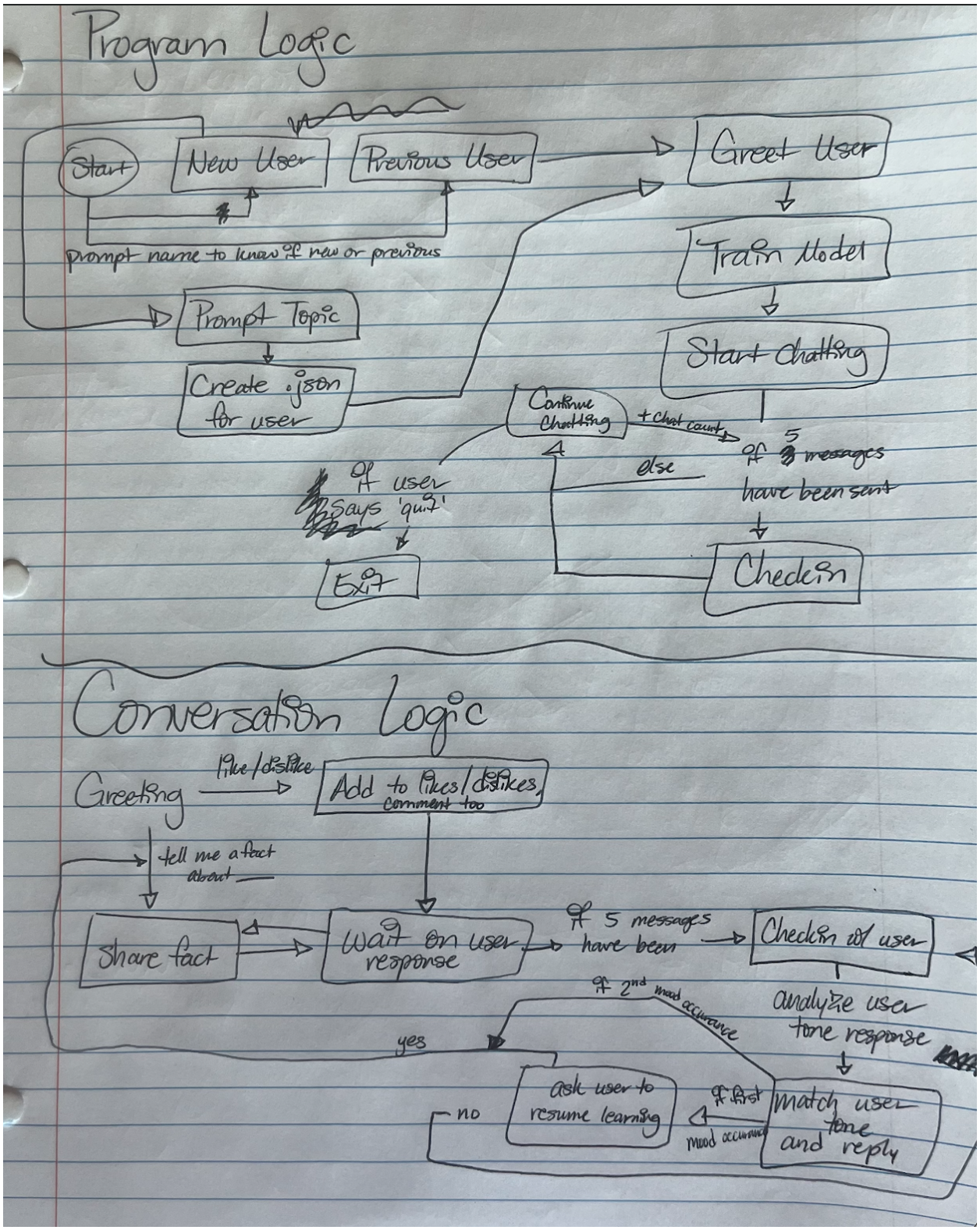
In addition to sharing facts the chatbot will also remember the user's likes and dislikes. Whenever the user says they like or dislike something that will prompt the bot to append to the .json knowledge base on the user and remember that information for the next time the user and bot chat (after the current conversion ends and a new one begins). When the chatbot is started back up again the bot will greet the user by mentioning their likes and possibly a dislike that was saved and randomly selected from the previous conversation.

Every 5 messages between the user and chatbot will be interrupted by the chatbot's check in feature. Here the chatbot will ask the user if they are enjoying their conversation with the bot. When the user responds the bot will do a **sentiment analysis** on their reply using **VADER** to see if the user is speaking positively, negatively, or neutral. If the user is positive or neutral, the bot will respond in a similar fashion. However, if the user is negative the bot will either reply mad, sad, or mockingly. The bot's mood, which is based on the user's reply, is then passed to ChatGPT to get a variety of phrases/expressions as a response. One of the responses is randomly selected and then sent back to the user. After the first moody interfaction the bot will ask the user if they are ready to "get back to learning." Here the user has the choice of agreeing or disagreeing. If the user agrees, the conversation will resume back to normal until the next check in. If the user diagrees, the bot will try to reason with them by asking what they would rather do instead. That response is again passed to the sentiment analysis function to be processed. And the mood analyzer function runs one more time with a new response, and possibly different mood. After the second moody interaction the bot will resume back to normal and the user can prompt it once again for a fact.

Brianna Murray bnm180003
CS 4395.001

Logic Diagram

Visual summary of the above description.



Brianna Murray bnm180003
CS 4395.001

Sample Dialog

MoodyBot can have a very basic conversation with the user. Things get fun when MoodyBot deals with sentiment analysis. MoodyBot works best when you tell it your likes and dislikes within your first conversation as a new user with the bot.

Brianna Murray bnm180003
CS 4395.001

```
/Users/brimurray/Desktop/ChatBot/venv/bin/python /Users/brimurray/Desktop/ChatBot/main.py
Please enter your name...
You: bri
What is your favorite hobby?
You: houseplants
Ok, I will learn about this...please wait
[nltk_data] Error loading punkt: <urlopen error [SSL:
[nltk_data]     CERTIFICATE_VERIFY_FAILED] certificate verify failed:
[nltk_data]     unable to get local issuer certificate (_ssl.c:1129)>
2023-04-15 10:17:17.193235: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is optimized to use available CPU instructions in performance-
To enable the following instructions: AVX2 AVX512F AVX512_VNNI FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.
WARNING:tensorflow:From /Users/brimurray/Desktop/ChatBot/venv/lib/python3.9/site-packages/tensorflow/python/compat/v2_compat.py:107: disable_resource_variables (from
Instructions for updating:
non-resource variables are not supported in the long term
bri_data.json
Train the model type in (train): train
WARNING:tensorflow:From /Users/brimurray/Desktop/ChatBot/venv/lib/python3.9/site-packages/tflearn/initializations.py:164: calling TruncatedNormal.__init__ (from tenso
Instructions for updating:
Call initializer instance with the dtype argument instead of passing it to the constructor
---------------------------------
Run id: PIL99B
Log directory: /tmp/tflearn_logs/
---------------------------------
Training samples: 55
Validation samples: 0
--
Training Step: 1  | time: 0.280s
| Adam | epoch: 001 | loss: 0.00000 - acc: 0.0000 -- iter: 08/55
Training Step: 2  | total loss: 2.30845 | time: 0.283s
| Adam | epoch: 001 | loss: 2.30845 - acc: 0.0000 -- iter: 16/55
Training Step: 3  | total loss: 2.51772 | time: 0.285s
| Adam | epoch: 001 | loss: 2.51772 - acc: 0.0000 -- iter: 24/55
Training Step: 4  | total loss: 2.55358 | time: 0.287s
```

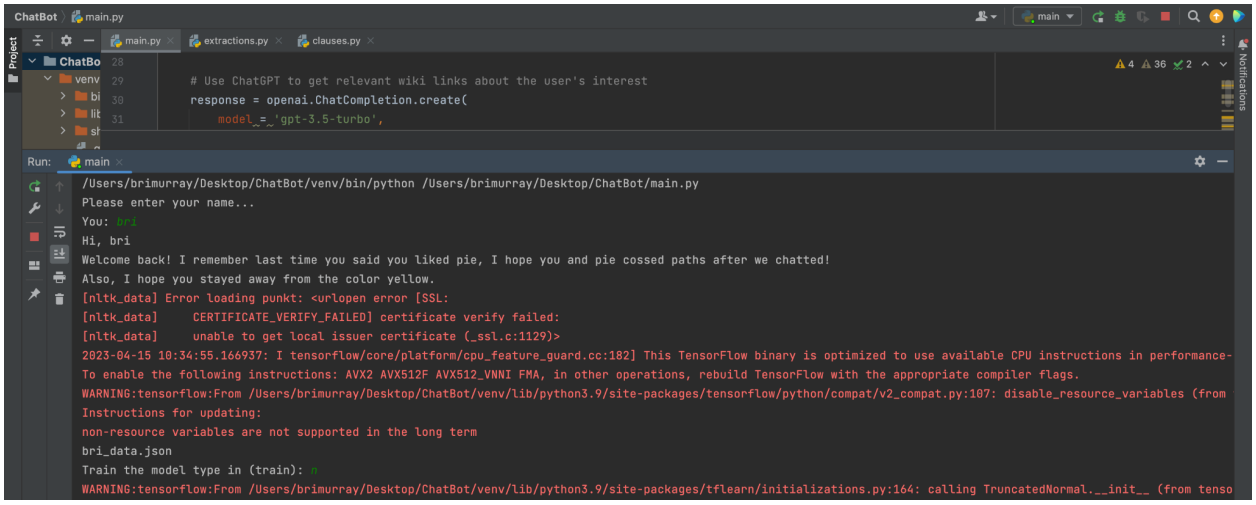
```
Start talking with bot!(type 'quit' to stop)
You: Hi
Hi, bri!
You: I like coloring
I enjoy coloring too
You: I dislike the color yellow
I see
You: I like pie
I enjoy pie too
You: Tell me a fact about houseplants
In fact, many succulents are able to thrive in dry conditions, and some are able to last up to two years without water depending on their surroundings and adaptations
That fact was a complex sentence!
1 In fact , many succulentsare able
2 to thrivein dry conditions , and some are able
3 to last upto two years
4 without waterdepending
You: tell me about lilies
Are you having fun?
You: I was until you cut me off. That was rude
["I can't believe this is happening! I've put in so many hours of work for this project and now it's all falling apart!", "Are you kidding me? I specifically asked fo
Do you have any idea how much time and effort I've put into this? And now it's all for nothing because of one mistake!
Are you ready to get back to learning? (y/n)
You: q
What would you rather do?
You: I want to ride a bmx bike
['I am feeling on top of the world today!', 'I am so grateful for this opportunity to work and grow', 'I simply love what I do for a living', 'I am so happy to be abl
I am proud of the work I have accomplished and the progress I have made
Let's get back to learning.
You: wit
In South Africa, several species have been threatened with extinction due to poaching from the wild for the black market and mining related activities
quit
That fact was a complex sentence! Here is a list of the clauses from that fact:
1 In South Africa , several specieshave
2 beenthreatened
3 with extinction dueto poaching
```

Brianna Murray bnm180003
CS 4395.001

```
I enjoy coloring too
You: I dislike the color yellow
I see
You: I like pie
I enjoy pie too
You: Tell me a fact about houseplants
In fact, many succulents are able to thrive in dry conditions, and some are able to last up to two years without water depending on their surroundings and adaptations
That fact was a complex sentence!
1 In fact , many succulentsare able
2 to thrivein dry conditions , and some are able
3 to last upto two years
4 without waterdepending
You: tell me about lilies
Are you having fun?
You: I was until you cut me off. That was rude
["I can't believe this is happening! I've put in so many hours of work for this project and now it's all falling apart!", "Are you kidding me? I specifically asked fo
Do you have any idea how much time and effort I've put into this? And now it's all for nothing because of one mistake!
Are you ready to get back to learning? (y/n)
You: q
What would you rather do?
You: I want to ride a bmx bike
['I am feeling on top of the world today!', 'I am so grateful for this opportunity to work and grow', 'I simply love what I do for a living', 'I am so happy to be abl
I am proud of the work I have accomplished and the progress I have made
Let's get back to learning.
You: quit
In South Africa, several species have been threatened with extinction due to poaching from the wild for the black market and mining related activities
quit
That fact was a complex sentence! Here is a list of the clauses from that fact:
1 In South Africa , several specieshave
2 beenthreatened
3 with extinction dueto poaching
4 from the wildfor the black market and mining related activities
You:
Process finished with exit code 0
```

```
You: hi
Hi, bri!
You: tell me about plants
Plants form about 80% of the world biomass at about 450 gigatonnes (4
You: I like pie
I enjoy pie too
You: I dislike bugs
I see
You: Tell me about how to water plants
This is how early plants reproduced
You: ok
Are you having fun?
You: Not really, you are dumb
Overall sentiment dictionary is :  {'neg': 0.471, 'neu': 0.529, 'pos': 0.0, 'compound': -0.552}
sentence was rated as  47.099999999999994 % Negative
sentence was rated as  52.900000000000006 % Neutral
sentence was rated as  0.0 % Positive
Sentence Overall Rated As Negative
["I just can't seem to shake this feeling of sadness.", "I'm trying my best to focus on work, but everything just feels so overwhelming right nou
I'm trying my best to focus on work, but everything just feels so overwhelming right now.
Are you ready to get back to learning? (y/n)
You: n
What would you rather do?
You: Talk to you about my dog
Overall sentiment dictionary is :  {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
sentence was rated as  0.0 % Negative
sentence was rated as  100.0 % Neutral
sentence was rated as  0.0 % Positive
Sentence Overall Rated As Neutral
["I'm here to do my job and get it done to the best of my abilities.", "I don't have any personal feelings about this task, I just want to compl
As a hard worker, I strive to remain unbiased and approach each task with an open mind.
Let's get back to learning.
```

Brianna Murray bnm180003
CS 4395.001



Brianna Murray bnm180003
CS 4395.001

Knowledge Base Appendix

The knowledge base is live upon creation because it is based on the user's preferred topic of discussion. However, after the knowledge base is created it is then static. In this example you can see the user selected the knowledge base to be about plants. The patterns are the top 5 words from the document and the responses are all the sentences that contain those top words. This file is created as a .json file for loop up retrieval by the TensorFlow model.

main.py | bri_data.json | extractions.py | clauses.py


```
{
    "intents": [
        {
            "tag": "areas",
            "patterns": [
                "plants",
                "houseplants",
                "plant",
                "grown",
                "indoor"
            ],
            "responses": [
                "Common houseplants are usually tropical or semi-tropical, and are often epiphytes, succulents or cacti",
                "The history of houseplants is intertwined with the history of container gardening in general",
                "Ancient Egyptians and Sumerians grew ornamental and fruiting plants in decorative containers",
                "In ancient China, potted plants were shown at garden exhibitions over 2,500 years ago",
                "In the Renaissance, plant collectors and affluent merchants from Italy, the Netherlands and Belgium importe
                "In the 17th century, fascination in exotic plants grew among the aristocracy of France and England, with in
                "Less wealthy people would also grow plants, especially flowers, indoors",
                "In the early 18th century, windows in London became wider and brighter, expanding the opportunities for the
                "The expansion of European colonialism brought Europeans into contact with a wide variety of new plants, esp
                "Forcing plants to bloom out of season, especially bulbs, was very popular in this era",
                "The decorative pot or cachepot specifically for growing houseplants (as opposed to a simple terracotta pot,
                "Furthermore, nurseries were flourishing in the 18th century, which stocked thousands of plants, including
                "The Victorian era saw the first use of houseplants by the middle class, which were perceived as a symbol o
                "Some foliage plants which tolerated the typical gloomy and snug environment inside a Victorian house became
                "The quintessential Victorian plants were palms (such as kentia palms and parlour palms), the cast iron pla
```


bri_data.json | extractions.py | clauses.py

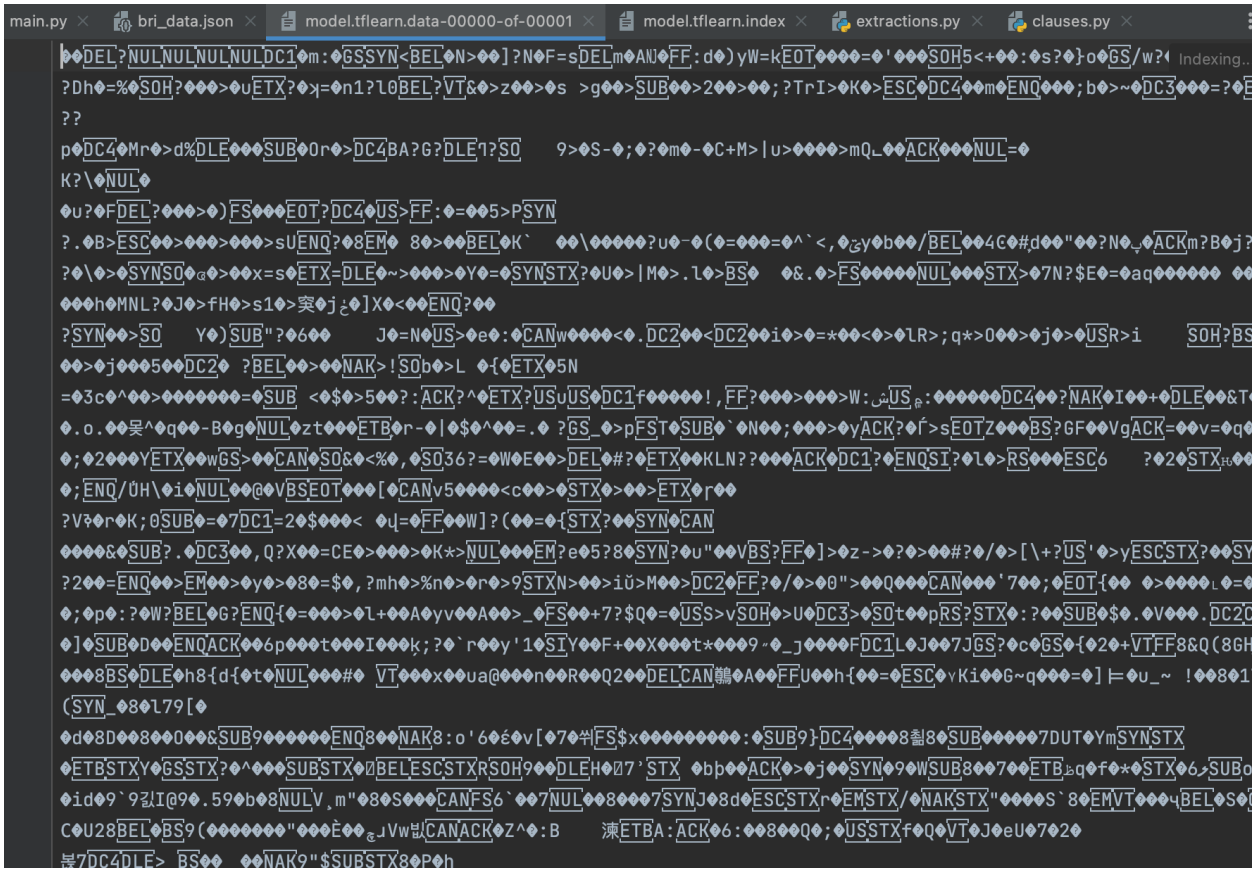

```
        "patterns": [
            "lily",
            "peace",
            "spathiphyllum",
            "cm",
            "spathe"
        ],
        "responses": [
            "The more common name of peace lily derives from the plants symbolism for peace, purity, and he
            "Although it is called a \"lily\", the peace lily is not a true lily from the family Liliaceae"
            "True lilies are highly toxic (poisonous) to cats and dogs, but the peace lily, spathiphyllum i
            "Certain species of Spathiphyllum are commonly known as spath or peace lilies",
            "The more common name of peace lily derives from the plants symbolism for peace, purity, and he
            "Although it is called a \"lily\", the peace lily is not a true lily from the family Liliaceae"
            "True lilies are highly toxic (poisonous) to cats and dogs, but the peace lily, spathiphyllum i
            "Explorers and botanists brought over 5,000 species to Europe from South America, Africa, Asia
            "In the 1920s, commercial houseplant production began in California, focused on the Kentia palm
            "Each plant species has different care requirements, and care requirements can vary based on th
            "Most houseplants are species that have adapted to survive in a temperature range between 15 \u
            "Wild collection of plants for the houseplant trade is a major threat to many species, especial
            "In the gradual process of families improving their immediate environment, useful tree and vine
            "Eventually foreign species were also selected and incorporated into the gardens",
            "Among the most popular species planted were date palms, sycamores, fig trees, nut trees, and w
            "In a famous Canadian Charter of Rights case, \"Sandra Bell vs",
            " City of Toronto\", 1997, the right to cultivate all native species, even most varieties deeme
            "Pesticides may affect the ecology of a garden due to their effects on the populations of both
            "Succulents are not a taxonomic category, since the term describes only the attributes of a par
```

Brianna Murray bnm180003
CS 4395.001

Sample User Appendix

For simplicity the user model starting data and knowledge base have been merged. The user's name is saved and known by the filename of the knowledge base and the likes/dislikes of the user are saved at the bottom of the knowledge base file. The user model that TensorFlow creates is not readable and I have attached a picture to prove this. The best way to understand the user model is to look at the output provided when the model is being trained. The .json file contains X amount of terms when all pattern values are summed together. The model will train itself X times for each term within the pattern values in the .json file. You can see in the provided images below that when the model is finished training it has an accuracy of 98% with a loss around 1%. This value fluctuates however the numbers I typically see are between 96%-100% varying based on what data is pulled for each given user.

Brianna Murray bnm180003
CS 4395.001



Brianna Murray bnm180003
CS 4395.001

bri_data.json extractions.py clauses.p

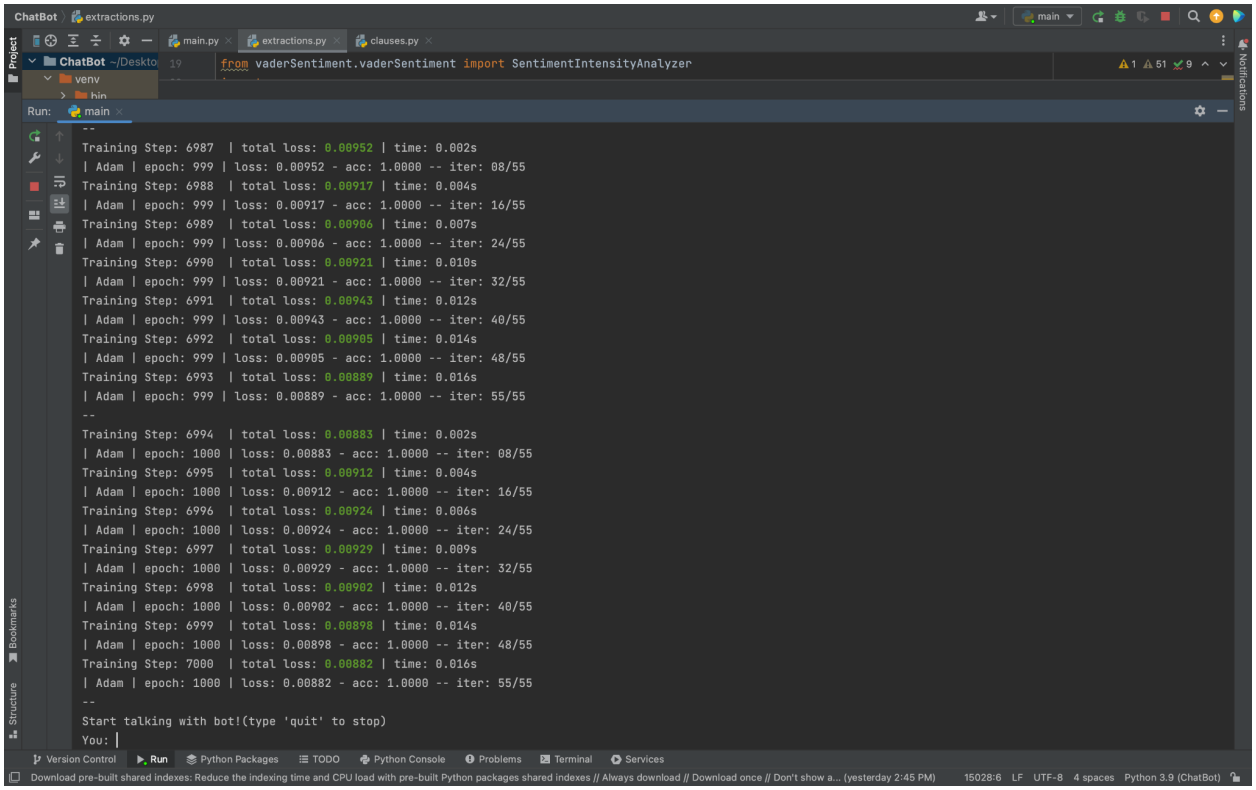
```
{
    "tag": "likes",
    "patterns": [
        "I like",
        "I love",
        "I enjoy",
        "favorite"
    ],
    "responses": [
        "coloring",
        "pie"
    ]
},
{
    "tag": "dislikes",
    "patterns": [
        "I don't like",
        "I do not like",
        "I hate",
        "I don't enjoy",
        "I do not enjoy",
        "least favorite"
    ],
    "responses": [
        "the color yellow"
    ]
```

Brianna Murray bnm180003
CS 4395.001

```
/Users/brimurray/Desktop/ChatBot/venv/bin/python /Users/brimurray/Desktop/ChatBot/main.py
Please enter your name...
You: bri
Hi, bri
[nltk_data] Error loading punkt: <urlopen error [SSL:
[nltk_data]     CERTIFICATE_VERIFY_FAILED] certificate verify failed:
[nltk_data]     unable to get local issuer certificate (_ssl.c:1129)>
2023-04-15 09:22:58.623264: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is optimized to use available CPU instructions in performance-
To enable the following instructions: AVX2 AVX512F AVX512_VNNI FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.
WARNING:tensorflow:From /Users/brimurray/Desktop/ChatBot/venv/lib/python3.9/site-packages/tensorflow/python/compat/v2_compat.py:107: disable_resource_variables (from
Instructions for updating:
non-resource variables are not supported in the long term
bri_data.json
Train the model type in (train): train
WARNING:tensorflow:From /Users/brimurray/Desktop/ChatBot/venv/lib/python3.9/site-packages/tflearn/initializations.py:164: calling TruncatedNormal.__init__ (from tenso
Instructions for updating:
Call initializer instance with the dtype argument instead of passing it to the constructor
---------------------------------
Run id: VE55CJ
Log directory: /tmp/tflearn_logs/
---------------------------------
Training samples: 51
Validation samples: 0
--
Training Step: 1  | time: 0.276s
| Adam | epoch: 001 | loss: 0.00000 - acc: 0.0000 -- iter: 08/51
Training Step: 2  | total loss: 2.37516 | time: 0.279s
| Adam | epoch: 001 | loss: 2.37516 - acc: 0.0000 -- iter: 16/51
Training Step: 3  | total loss: 2.59018 | time: 0.280s
| Adam | epoch: 001 | loss: 2.59018 - acc: 0.2045 -- iter: 24/51
Training Step: 4  | total loss: 2.62650 | time: 0.282s
| Adam | epoch: 001 | loss: 2.62650 - acc: 0.0511 -- iter: 32/51
Training Step: 5  | total loss: 2.63559 | time: 0.284s
```

```
--
Training Step: 6987  | total loss: 0.62407 | time: 0.003s
| Adam | epoch: 999 | loss: 0.62407 - acc: 0.9551 -- iter: 08/51
Training Step: 6988  | total loss: 0.56212 | time: 0.006s
| Adam | epoch: 999 | loss: 0.56212 - acc: 0.9596 -- iter: 16/51
Training Step: 6989  | total loss: 0.50633 | time: 0.008s
| Adam | epoch: 999 | loss: 0.50633 - acc: 0.9636 -- iter: 24/51
Training Step: 6990  | total loss: 0.45609 | time: 0.011s
| Adam | epoch: 999 | loss: 0.45609 - acc: 0.9673 -- iter: 32/51
Training Step: 6991  | total loss: 0.41086 | time: 0.013s
| Adam | epoch: 999 | loss: 0.41086 - acc: 0.9705 -- iter: 40/51
Training Step: 6992  | total loss: 0.37000 | time: 0.015s
| Adam | epoch: 999 | loss: 0.37000 - acc: 0.9735 -- iter: 48/51
Training Step: 6993  | total loss: 0.33344 | time: 0.017s
| Adam | epoch: 999 | loss: 0.33344 - acc: 0.9761 -- iter: 51/51
--
Training Step: 6994  | total loss: 0.30059 | time: 0.002s
| Adam | epoch: 1000 | loss: 0.30059 - acc: 0.9785 -- iter: 08/51
Training Step: 6995  | total loss: 0.27085 | time: 0.005s
| Adam | epoch: 1000 | loss: 0.27085 - acc: 0.9807 -- iter: 16/51
Training Step: 6996  | total loss: 0.24458 | time: 0.008s
| Adam | epoch: 1000 | loss: 0.24458 - acc: 0.9826 -- iter: 24/51
Training Step: 6997  | total loss: 0.22046 | time: 0.010s
| Adam | epoch: 1000 | loss: 0.22046 - acc: 0.9843 -- iter: 32/51
Training Step: 6998  | total loss: 0.19906 | time: 0.013s
| Adam | epoch: 1000 | loss: 0.19906 - acc: 0.9859 -- iter: 40/51
Training Step: 6999  | total loss: 0.17958 | time: 0.015s
| Adam | epoch: 1000 | loss: 0.17958 - acc: 0.9873 -- iter: 48/51
Training Step: 7000  | total loss: 0.16208 | time: 0.018s
| Adam | epoch: 1000 | loss: 0.16208 - acc: 0.9886 -- iter: 51/51
--
Start talking with bot!(type 'quit' to stop)
You:
```

Brianna Murray bnm180003
CS 4395.001

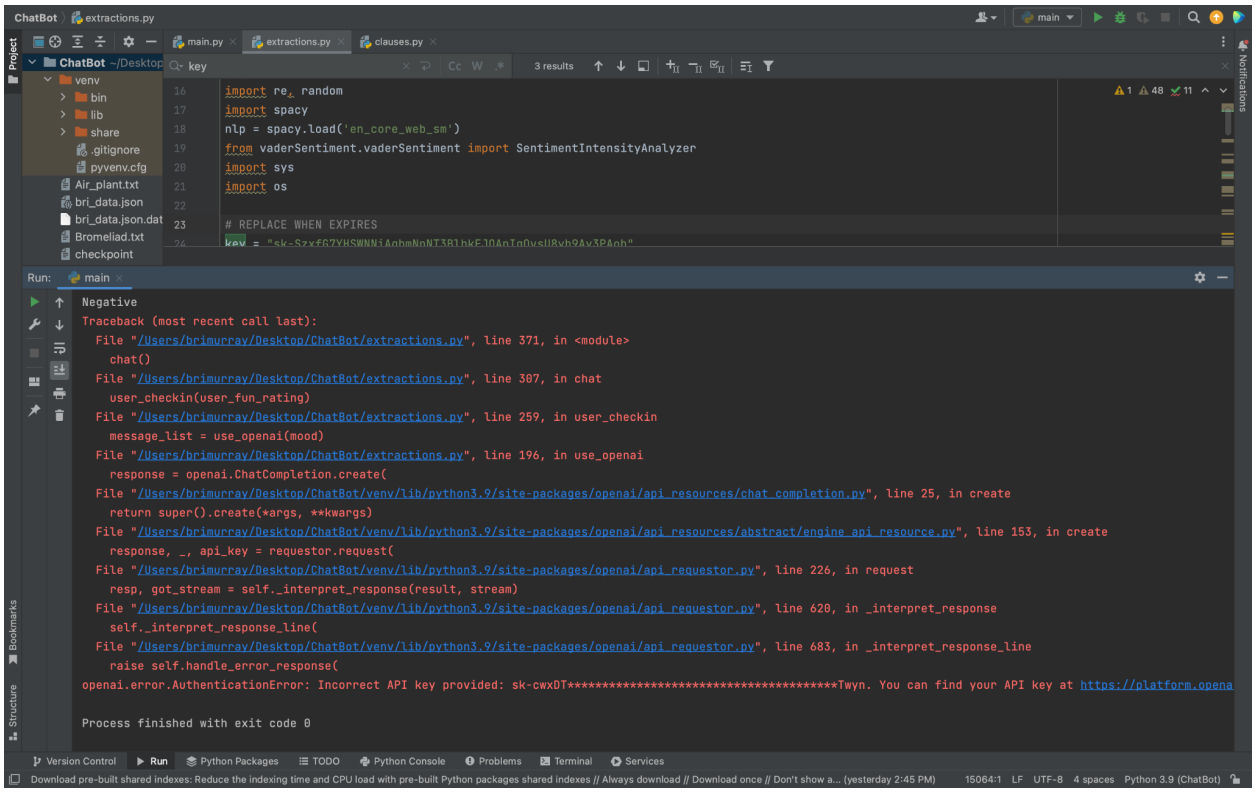


Evaluation

Overall I think the chatbot's strength really shines with the clause recognizer and cosine similarity features that help to further expand the bot's capabilities. TensorFlow works extremely well and matches the user prompt to the correct tag category every time. In addition, the bot is able to accurately recognize the user's tone when checking in every time as well. Also, the clause recognizer is accurate for what it was programmed to do. I think the weakness with the clause recognizer for the sentences it is providing is that it only bases a clause on the amount of verb phrases found within a sentence when there are additional rules that also come into play when looking for a verb phrase. The other shortcoming with the bot is that because the cosine similarity was implemented within the fact retrieval feature, 9/10 the bot will give the same fact to the user repeatedly if the prompt from the user does not change.

Notice: This is a qualm more so with OpenAI rather than the chatbot itself but new API tokens have to be recreated for this project to work due to the fact that they expire frequently. When running this project be sure to create a brand new Secret Key API from OpenAI to properly run the bot. ***(I've attached a screenshot below of what the error looks like if you need to make a new key.)***

Brianna Murray bnm180003
CS 4395.001

Also, the red text shown within the other screenshots are not errors. They are messages / warnings / suggestions of ways to refine the tools being used so that they can be configured to their most efficient capabilities. Those can be overlooked, program will still run.

If first model does not output desired results, delete created files and retrain the model.