PHASE-3: DEVELOPMENT PART 1

Project-3 Create a chatbot using Python

In this phase of project submission, we are all set to develop a program using python to setup a chatbot. In the previous phases we have acquired enough knowledge about the basics of a chatbot and about some innovations that could make it even better.

To start with we are provided a dataset from Kaggle which is a text file comprising set of dialogues. We are about to process the dataset and build a simple chatbot that would reply to your prompt by comparing it to the existing dataset.

DATA CLEANING:

The dataset provided must be processed prior feeding it to the chatbot program. This essential step is known as 'Data Cleaning'. Data can be sorted using the pandas library in python.

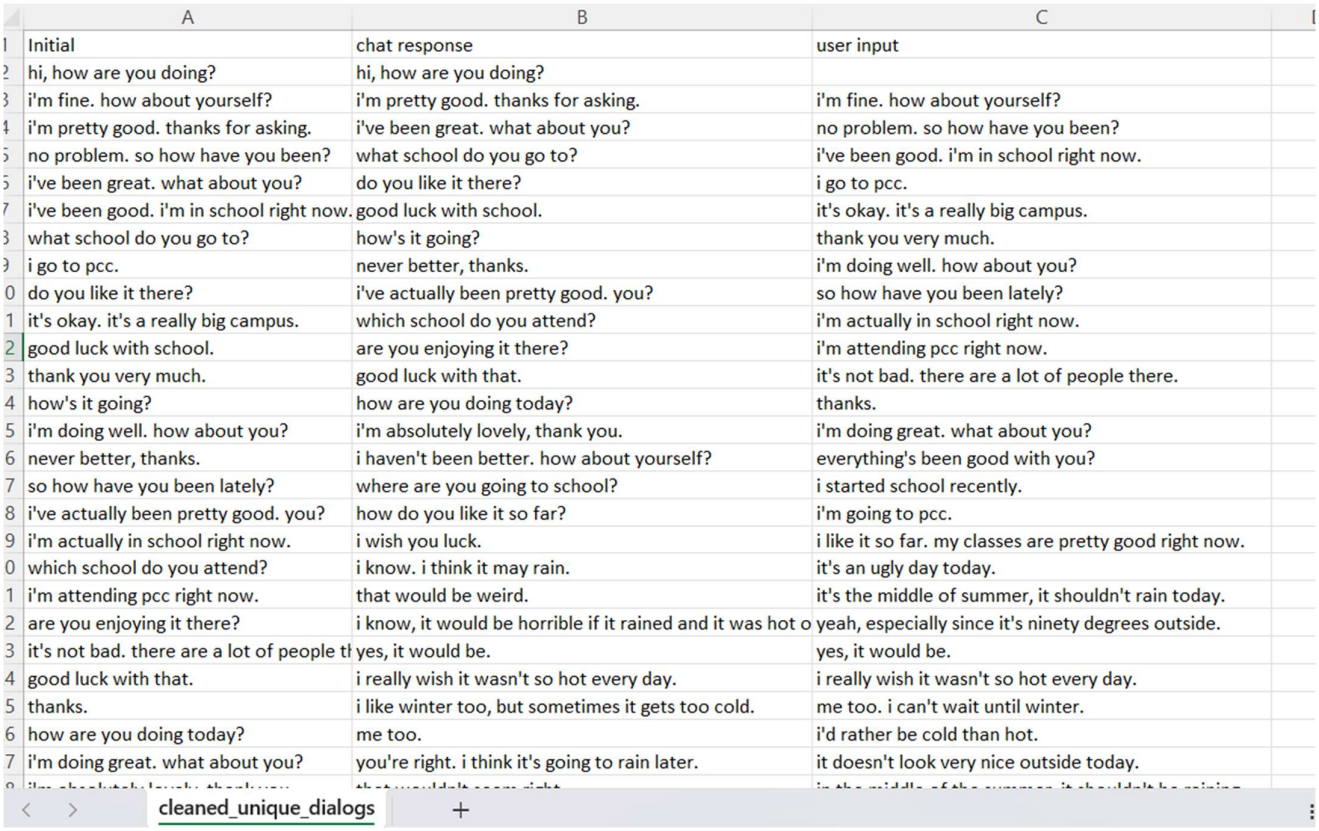
The steps are as follows:

1. Read the text file using pandas library.
2. Store the dialogues as a string in a new csv file under the column ' initial' .
3. Use the unique() function to filter out the repeating values.
4. Iterate through the values of the 'initial' column and store them under 2 new columns- chatbot_response and user_input.
5. Open the CSV file in excel to check the format of the data stored and insert an empty cell in the first row of user_input. This helps in initialization of chatbot's first response later while running the code.

CODE SNIPPET:

```
import pandas as pd


# The path to your text-based dataset
dataset_path = "C:\\Users\\shiva\\Downloads\\archive\\dialogs.txt"


with open(dataset_path, 'r', encoding='utf-8') as file:
    lines = file.readlines()
cleaned_lines = []
for line in lines:
    parts = line.strip().split('\t')  # Split using the tab character
    if len(parts) == 2:
        cleaned_lines.append(parts[0])
        cleaned_lines.append(parts[1])
data = {'Initial': cleaned_lines}
df = pd.DataFrame(data)
df.drop_duplicates(subset='Initial', inplace=True)


# The path to save the cleaned DataFrame as a CSV file
cleaned_csv_path = "C:\\Users\\shiva\\Downloads\\archive\\cleaned_unique_dialogs.csv"


df.to_csv(cleaned_csv_path, index=False)
df = pd.read_csv(cleaned_csv_path)
df['chat_response'] = df.iloc[::2, 0].reset_index(drop=True)
df['user_input'] = df.iloc[1::2, 0].reset_index(drop=True)
df.to_csv(csv_path, index=False)
```

OUTPUT:



	A	B	C
1	Initial	chat response	user input
2	hi, how are you doing?	hi, how are you doing?	
3	i'm fine. how about yourself?	i'm pretty good. thanks for asking.	i'm fine. how about yourself?
4	i'm pretty good. thanks for asking.	i've been great. what about you?	no problem. so how have you been?
5	no problem. so how have you been?	what school do you go to?	i've been good. i'm in school right now.
6	i've been great. what about you?	do you like it there?	i go to pcc.
7	i've been good. i'm in school right now.	good luck with school.	it's okay. it's a really big campus.
8	what school do you go to?	how's it going?	thank you very much.
9	i go to pcc.	never better, thanks.	i'm doing well. how about you?
10	do you like it there?	i've actually been pretty good. you?	so how have you been lately?
11	it's okay. it's a really big campus.	which school do you attend?	i'm actually in school right now.
12	good luck with school.	are you enjoying it there?	i'm attending pcc right now.
13	thank you very much.	good luck with that.	it's not bad. there are a lot of people there.
14	how's it going?	how are you doing today?	thanks.
15	i'm doing well. how about you?	i'm absolutely lovely, thank you.	i'm doing great. what about you?
16	never better, thanks.	i haven't been better. how about yourself?	everything's been good with you?
17	so how have you been lately?	where are you going to school?	i started school recently.
18	i've actually been pretty good. you?	how do you like it so far?	i'm going to pcc.
19	i'm actually in school right now.	i wish you luck.	i like it so far. my classes are pretty good right now.
20	which school do you attend?	i know. i think it may rain.	it's an ugly day today.
21	i'm attending pcc right now.	that would be weird.	it's the middle of summer, it shouldn't rain today.
22	are you enjoying it there?	i know, it would be horrible if it rained and it was hot o	yeah, especially since it's ninety degrees outside.
23	it's not bad. there are a lot of people th	yes, it would be.	yes, it would be.
24	good luck with that.	i really wish it wasn't so hot every day.	i really wish it wasn't so hot every day.
25	thanks.	i like winter too, but sometimes it gets too cold.	me too. i can't wait until winter.
26	how are you doing today?	me too.	i'd rather be cold than hot.
27	i'm doing great. what about you?	you're right. i think it's going to rain later.	it doesn't look very nice outside today.

cleaned_unique_dialogs

CHATBOT DESIGNING:

A simple chatbot can be programmed using the 'dictionary' approach. Wherein the values from the dataset are initialized as a dictionary datatype. The user input are stored as keys while the corresponding chat response is stored as a value in the dictionary.

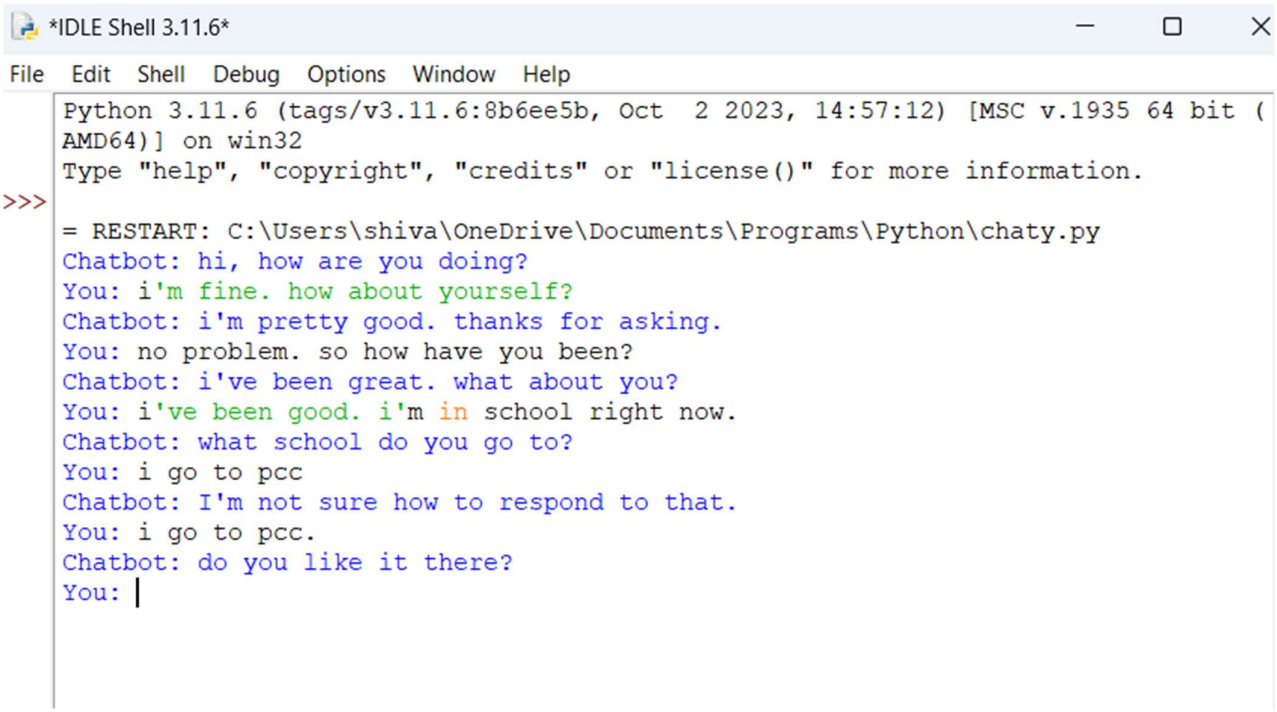
So when a user prompts a dialog similar to the dialogues in the dataset, the chatbot responds by returning the value for the specific key.

CODE SNIPPET:

```
import pandas as pd

dataset =
pd.read_csv("C:\\Users\\shiva\\Downloads\\archive\\cleaned_unique_dialogs.csv")


responses = dict(zip(dataset['user input'], dataset['chat response']))
```

```
def get_chatbot_response(user_input):
    user_input = user_input.lower()
    return responses.get(user_input, "I'm not sure how to respond to that.")
first_dialogue = dataset['user input'].iloc[0]
chatbot_first_response = dataset['chat response'].iloc[0]
print("Chatbot:", chatbot_first_response)
while True:
    user_input = input("You: ")
    chatbot_response = get_chatbot_response(user_input)
    print("Chatbot:", chatbot_response)
```

OUTPUT:



NOTE:

This is a very simple approach to build a chatbot, hence the concept of tokenization in NLU is not applied here. The chat bot's capabilities are very limited and can only reply with the basic templates provided by the dataset. The chatbot is incapable of understanding user intentions as it is not equipped with NLU libraries.

We intend to develop the chatbot capabilities by integrating NLU libraries and will give a try to integrate the chatbot with a web application in subsequent phases of project submission.