**Create a Chatbot in Python**

**Phase 3: Development Part 1**

BALAMURUGAN G

953621104009



pip install pandas scikit-learn numpy

Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (1.5.3) Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (1.2.2) Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (1.23.5)

Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2) Requirement already satisfied: pytz>=2020.I in /usr/local/lib/python3.10/dist-packages (from pandas) (2023.3.post1) Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.11.3) Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.3.2) Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (3.2.0) Requirement already satisfied: six›=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.1-›pandas) (

|  |  |  |
| --- | --- | --- |
|  |  |  |
| import pandas as pd  from sklearn.model\_selection import train\_test\_split  from sklearn.feature\_extraction.text import CountVectorizer from sklearn.naive\_bayes import MultinomialNB  from sklearn.pipeline import make\_pipeline from sklearn import metrics | | |
| dataset\_link = ["https://www.kaggle.com/datasets/grafstor/simple-dialogs-for-chatbot"](http://www.kaggle.com/datasets/grafstor/simple-dialogs-for-chatbot) | | |
| file\_path = "dialogs.txt" | | |
| columns = ['user', 'bot']  df = pd.read\_csv(file\_path, sep='\t', header=None, names=columns) | | |
| print(df.head()) | | |

user \

1. hi, how are you doing?
2. i'm fine. how about yourself?
3. i'm pretty good. thanks for asking.
4. no problem. so how have you been?
5. i've been great. what about you?

bot

1. i'm fine. how about yourself7
2. i'm pretty good. thanks for asking.
3. no problem. so how have you been7
4. i've been great. what about you?
5. i've been good. i'm in school right now.

print(df.tail())

|  |  |  |
| --- | --- | --- |
| 3720 | user  that's a good question. maybe it's not old age. | \ |
| 3721 | are you right-handed7 |  |
| 3722 | yes. all my life. |  |
| 3723 | you're wearing out your right hand. stop using... |  |
| 3724 | but i do all my writing with my right hand. |  |
|  | bot |  |
| 3720 | are you right-handed? |  |
| 3721 | yes. all my life. |  |
| 3722 | you're wearing out your right hand. stop using... |  |
| 3723 | but i do all my writing with my right hand. |  |
| 3724 | start typing instead. that way your left hand ... |  |

pnint(df.info())

<class 'pandas.core.frame.DataFrame'› Rangelndex: 3725 entries, 0 to 3724 Data columns (total 2 columns):

# Column Non-Null Count Dtype

* 1. user 3725 non-null object
  2. bot 3725 non-null object dtypes: object(2)

memory usage: 58.3+ KB None

print(df.describe())

|  |  |  |
| --- | --- | --- |
| count | user  3725 | bot  3725 |
| un1que | 3510 | 3512 |
| top | what do you mean? | what do you mean? |
| freq | 22 | 22 |

print(type(df))

<class 'pandas.core.frame.DataFrame'›

print(type(file\_path))

<class 'str'>

print(df.shape)

(3725, 2)

pnint(df.size)

7450

df[”user“].value\_counts()

what do you mean ? 22

why not? 16

what happened? 10

why is that? 9

me too. 7

you can yell at them. 1

and they will yell back at you. 1

yelling doesn't do any good. 1

will you look at this form? 1

but i do all my writing with my right hand. 1

Name: user, Length: 3510, dtype: int64

df[“bot”].value\_counts()

what do you mean? 22

why not7 16

what happened? 10

why is that? 9

of course. 8

i don't understand some things. 1

let me help you. 1

what does “mi” mean? 1

mi stands for middle initial. 1

start typing instead. that way your left hand will do half the work. 1

Name: bot, Length: 3512, dtype: int64

df.index

Rangelndex(start=0, stop=3725, step=1)

df.values

array([['hi, how are you doing?', “i'm fine. how about yourself?”], [“i'm fine. how about yourself?”,

“i'm pretty good. thanks for asking.”], [“i'm pretty good. thanks for asking.”, 'no problem. so how have you been7'],

['yes. all my life.',

“you're wearing out your right hand. stop using it so much.“], [“you're wearing out your right hand. stop using it so much.”,

'but i do all my writing with my right hand.'], ['but i do all my writing with my right hand.',

'start typing instead. that way your left hand will do half the work.']], dtype=object)

df.axes

[Rangelndex(start=0, stop=3725, step=1), Index(['user', 'bot'], dtype='object')]

df.axes[0]

Rangelndex(start=0, stop=3725, step=1)

df.axes[1]

Index(['user', 'bot'], dtype='object')

df.keys()

Index(['user', 'bot'], dtype='object')

df.sum()

user hi, how are you doing?i'm fine. how about your... bot i'm fine. how about yourself?i'm pretty good. dtype: object

df.max()

user

and you know you should be glad! bot

there will be an answer, let it be! dtype: object

df.min()

user $1,000.

bot it doesn't matter. here he comes now with the... dtype: object

df.count()

user 3725

bot 3725

dtype: 1nt64

[('user', 0

list(df.items())

1

2

3

4

hi, how are you doing7 i'm fine. how about yourself?

i'm pretty good. thanks for asking. no problem. so how have you been7 i've been great. what about you?

3720 that's a good question. maybe it's not old age. 3721 are you right-handed?

3722 yes. all my life.

3723 you're wearing out your right hand. stop using... 3724 but i do all my writing with my right hand. Name: user, Length: 3725, dtype: object),

('bot',

* + 1. i'm fine. how about yourself?
    2. i'm pretty good. thanks for asking.
    3. no problem. so how have you been?
    4. i've been great. what about you?
    5. i've been good. i'm in school right now.

3720 are you right-handed?

3721 yes. all my life.

3722 you're wearing out your right hand. stop using... 3723 but i do all my writing with my right hand. 3724 start typing instead. that way your left hand ... Name: bot, Length: 3725, dtype: object)]

df.T

0 3 4

i'm no i've

6 7

i've

8

it's

3715 3716 3717 3718

don't if

fine. pretty problem. been

been what

okay.

what's

it what ' know. ag

hi, how '

user are you

how good. so how great. about thanks have what

good. school i'm in do

i go to pcc.

do you like it

it's a

really

wrong

with

aches do

most you

i think

it's

doing?

yourself? for you about asking. been? you?

school you

right go to? now.

there?

big

campus.

it?

of the think

time. it is?

old

age. ha

bot

i'm fine. i'”

how

pretty good.

no i've

problem. been

so how great.

i've

been what good. school

it's

do okay. good

i go to you it's a luck

it what

aches do

if it's

i old

don't age,

know. why

i don't quo

about yourself?

thanks

for asking.

have what

you about

been? you?

i'm in do

school you p CC

right go to?

now.

like it really with

there? big school. campus.

most you

of the think

time. it is?

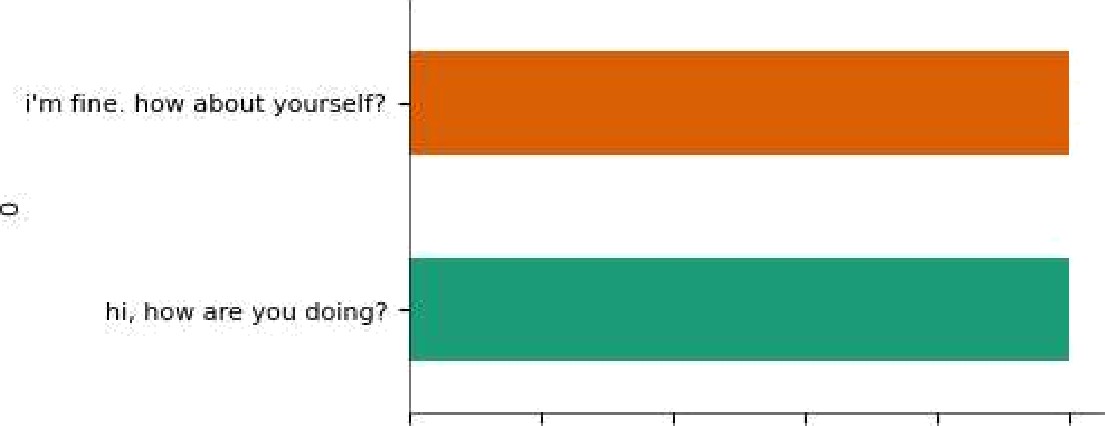
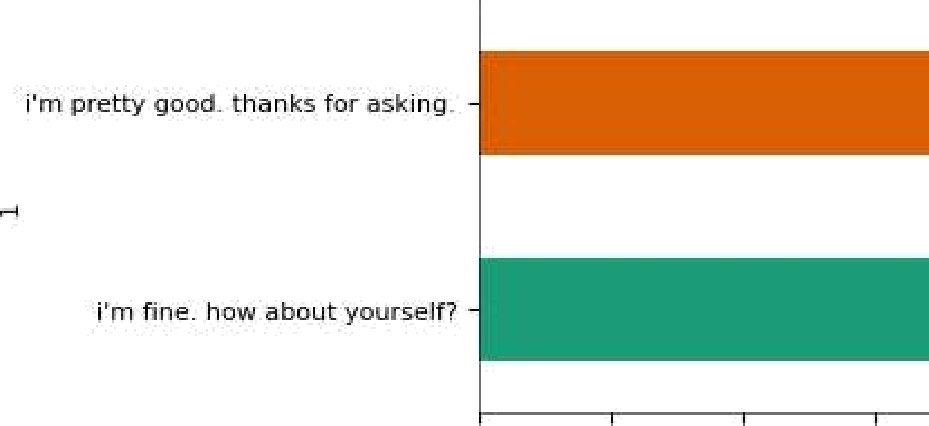
think both i

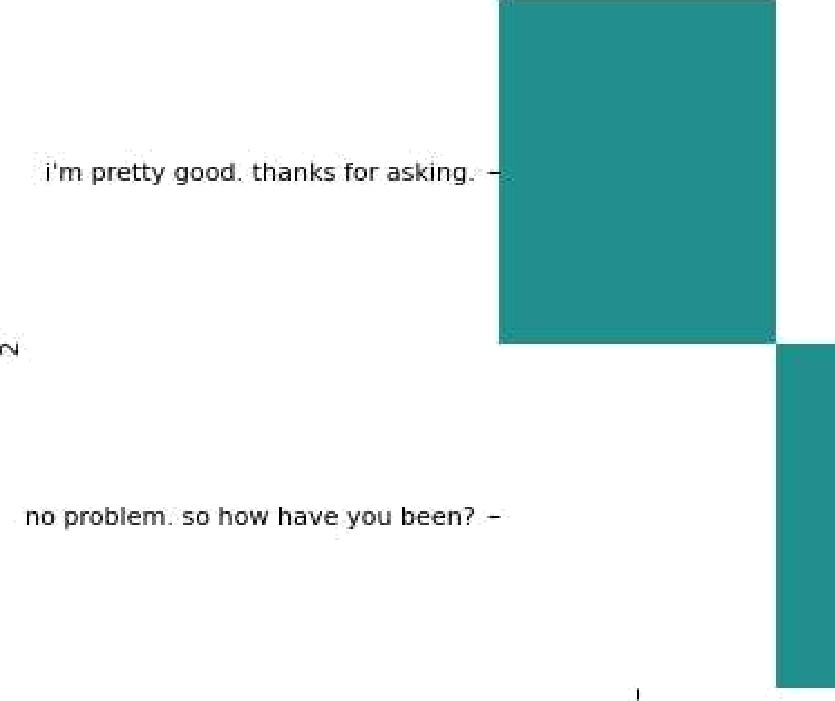
it's of

old your ol age. hands

* 1. rows > 3725 columns

Categorical distributions

2-d categorical distributions

0.0

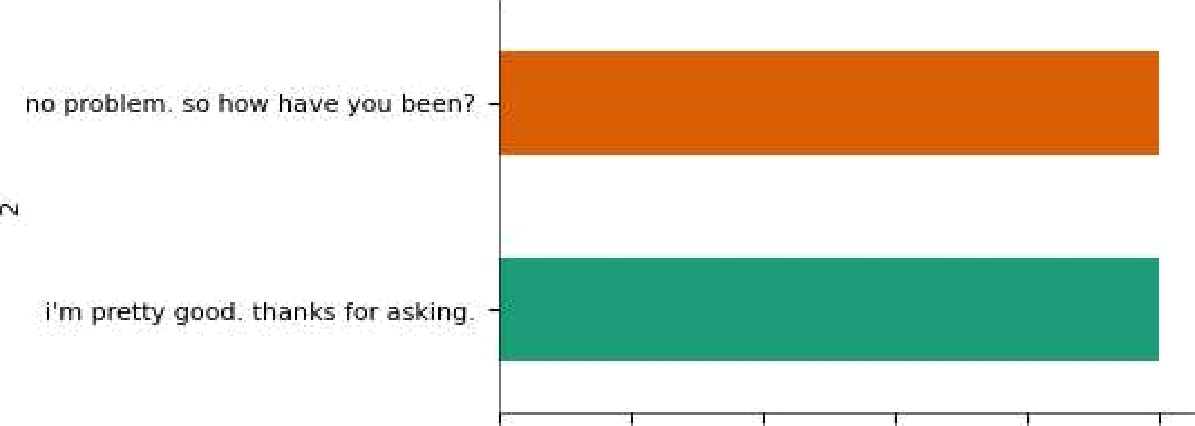
0.0

0.2

0.2

0.4

0.6

0.4 06

E

0

0.8

0.8

1.0

1.0

-1.100



’’-1.050

-1.025

-1.000

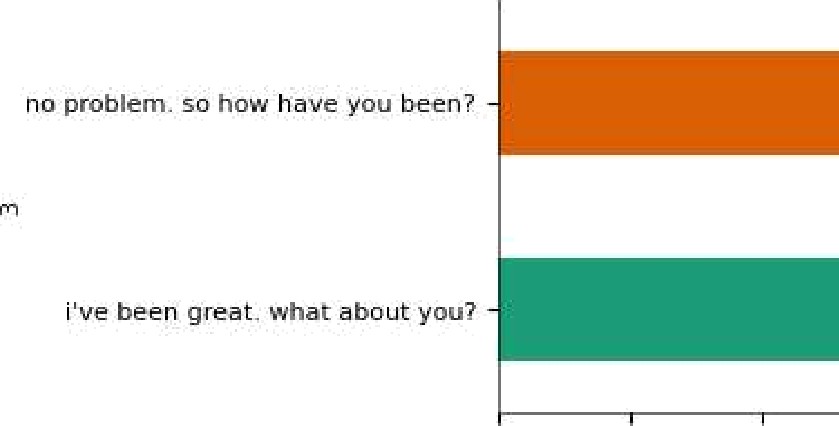
— 0.975

-0.950

- 0.925

— 0.900

0.Q 0.2

0.0

0.4

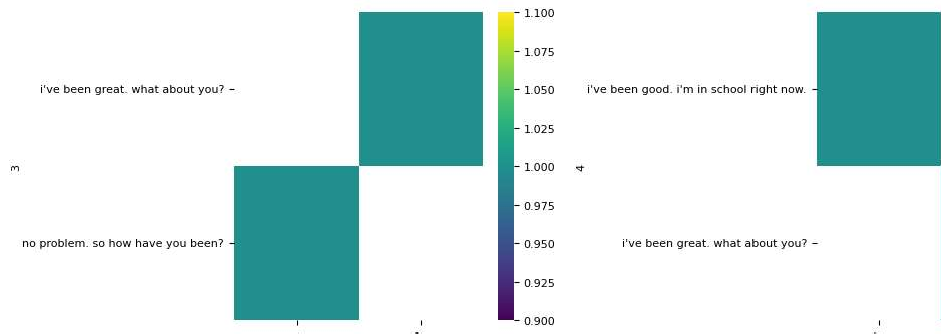
0.2



0.6

0.Q





|  |
| --- |
| df['user'] = df['user'].str.lower()  df['bot'] = df['bot'].str.lower() |
| X\_train, X\_test, y\_train, y\_test = train\_test\_split(df['user'], df['bot'], test\_size=0.2, random\_state=42) |
| model = make\_pipeline(CountVectorizer(), MultinomialNB()) |
| model.fit(X\_train, y\_train) |

,• Pipeline



CountVectorizer '

•Multi omialNB

y\_pred - model.predict(X test)

accuracy = metrics.accuracy\_score(y\_test, y\_pred) print(f“Accuracy: {accuracy}”)

Accuracy: 0.006711409395973154

accuracy = metrics.accuracy\_score(y\_test, y\_pred) accuracy\_percentage = accuracy \* 100

print(f“Accuracy: {accuracy\_percentage:.2f}%“)

Accuracy: 0.67%

precision = metrics.precision\_score(y\_test, y\_pred, average='weighted') recall = metrics.recall\_score(y\_test, y\_pred, average='weighted')

f1 = metrics.f1\_score(y\_test, y\_pred, average='weighted')

print(f“Precision: {precision}”) print(f“Recall: {recall}") print(f“F1 Score: {f1}”)

Precision: 0.00014766274113733603

Recall: 0.006711409395973154

F1 Score: 0.0002863846471638254

/usr/local/lib/python3.10/dist-packages/sklearn/metrics/\_classification.py:1344: UndefinedMetricWarning: Precision is ill-

\_warn\_prf(average, modifier, msg\_start, len(result))

/usr/local/lib/python3.10/dist-packages/sklearn/metrics/\_classification.py:1344: UndefinedMetricWarning: Recall is ill-def

\_warn\_prf(average, modifier, msg\_start, len(result))

class\_report = metrics.classification\_report(y\_test, y\_pred)

print(”Classification Report:”)

print(class\_report)

/usr/local/lib/python3.10/dist packages/sklearn/metrics/\_classification.py:1344: UndefinedMetricWarning: Precision and F’ ' \_warn\_prf(average, modifier, msg\_start, len(result))

/usr/local/lib/python3.10/dist-packages/sklearn/metrics/\_classification.py:1344: UndefinedMetricWarning: Recall and F-sc



\_warn\_prf(average, modifier, msg\_start, len(result)) Classification Report:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| precision | | recall | fl-score | suppo | |
| a blind date doesn't mean that she is blind! | 0.00 | 0.00 | 0.00 | |  |
| a girl i met at the market. | 0.00 | 0.00 | 0.00 | |  |
| a lot, like the party i'm having on friday. | 0.00 | 0.00 | 0.00 | |  |
| a team of doctors gave her a whole new face. | 0.00 | 0.00 | 0.00 | |  |
| a third guy punched one of the two guys. | 0.00 | 0.00 | 0.00 | |  |
| about 50 percent of the time he's traveling. | 0.00 | 0.00 | 0.00 | |  |
| after that, we can dry it with a towel. | 0.00 | 0.00 | 0.00 | |  |
| all right, see you. | 0.00 | 0.00 | 0.00 | |  |
| all right. i'll see you next time. | 0.00 | 0.00 | 0.00 | |  |
| also, i bought a can of cheap coffee and a bag of good coffee. | 0.00 | 0.00 | 0.00 | |  |
| and 4,000 people showed up. | 0.00 | 0.00 | 0.00 | |  |
| and i don't have either. | 0.00 | 0.00 | 0.00 | |  |
| and it's just as easy to prepare. | 0.00 | 0.00 | 0.00 | |  |
| and it's not too small. | 0.00 | 0.00 | 0.00 | |  |
| and now i'm glad that i didn't. | 0.00 | 0.00 | 0.00 | |  |
| and then we can tunn right at the light. | 0.00 | 0.00 | 0.00 | |  |
| and then you chase it. | 0.00 | 0.00 | 0.00 | |  |
| and we can afford it! | 0.00 | 0.00 | 0.00 | |  |
| and you're mine forever. | 0.00 | 0.00 | 0.00 | |  |
| any rain right now would be pointless. | 0.00 | 0.00 | 0.00 | |  |
| anything else? | 0.00 | 0.00 | 0.00 | |  |
| are you going to be buried next to each other? | 0.00 | 0.00 | 0.00 | |  |
| are you going to buy a desktop or laptop? | 0.00 | 0.00 | 0.00 | |  |
| are you going to go and visit her and the baby? | 0.00 | 0.00 | 0.00 | |  |
| are you going to start a family? | 0.00 | 0.00 | 0.00 | |  |
| are you going to wash it yourself? | 0.00 | 0.00 | 0.00 | |  |
| artists see things differently. because i want to make a lot of money. because i was watching you. | 0.00  *.*  0.00 | 0.00  *.*  0.00 | 0.00  *.*  0.00 | |  |
| because of all the security? | 0.00 | 0.00 | 0.00 | |  |
| because parents don't teach their kids right from wrong. | 0.00 | 0.00 | 0.00 | |  |
| because tax cuts will save them money. | 0.00 | 0.00 | 0.00 | |  |
| because you get to meet a lot of people. | 0.00 | 0.00 | 0.00 | |  |
| because you helped make it dirty. | 0.00 | 0.00 | 0.00 | |  |
| because you're thinking too much. | 0.00 | 0.00 | 0.00 | |  |
| but a house is so expensive. | 0.00 | 0.00 | 0.00 | |  |
| but big earthquakes happen only once in a while. | 0.00 | 0.00 | 0.00 | |  |
| but bush visited some of them in the hospital once. | 0.00 | 0.00 | 0.00 | |  |
| but i'm listening to it. | 0.00 | 0.00 | 0.00 | |  |
| but now work is hard, too. | 0.00 | 0.00 | 0.00 | |  |
| but someone's going to break it, so what good is it? | 0.00 | 0.00 | 0.00 | |  |
| but the same thing day after day gets old. | 0.00 | 0.00 | 0.00 | |  |
| but the smell is different. | 0.00 | 0.00 | 0.00 | |  |
| but there are a lot of police. | 0.00 | 0.00 | 0.00 | |  |
| but you can't win if you don't play. | 0.00 | 0.00 | 0.00 | |  |
| but you eat fruits. | 0.00 | 0.00 | 0.00 | |  |
| but you know that california weather is really unpredictable. | 0.00 | 0.00 | 0.00 | |  |
| but you're right. there are a lot of fires. | 0.00 | 0.00 | 0.00 | |  |
| buy a new one. | 0.00 | 0.00 | 0.00 | |  |
| can i have all those cans of tomato sauce you bought for your pasta? | 0.00 | 0.00 | 0.00 | | \* |
|  |  |  |  | |  |

**PROJECT SETUP**

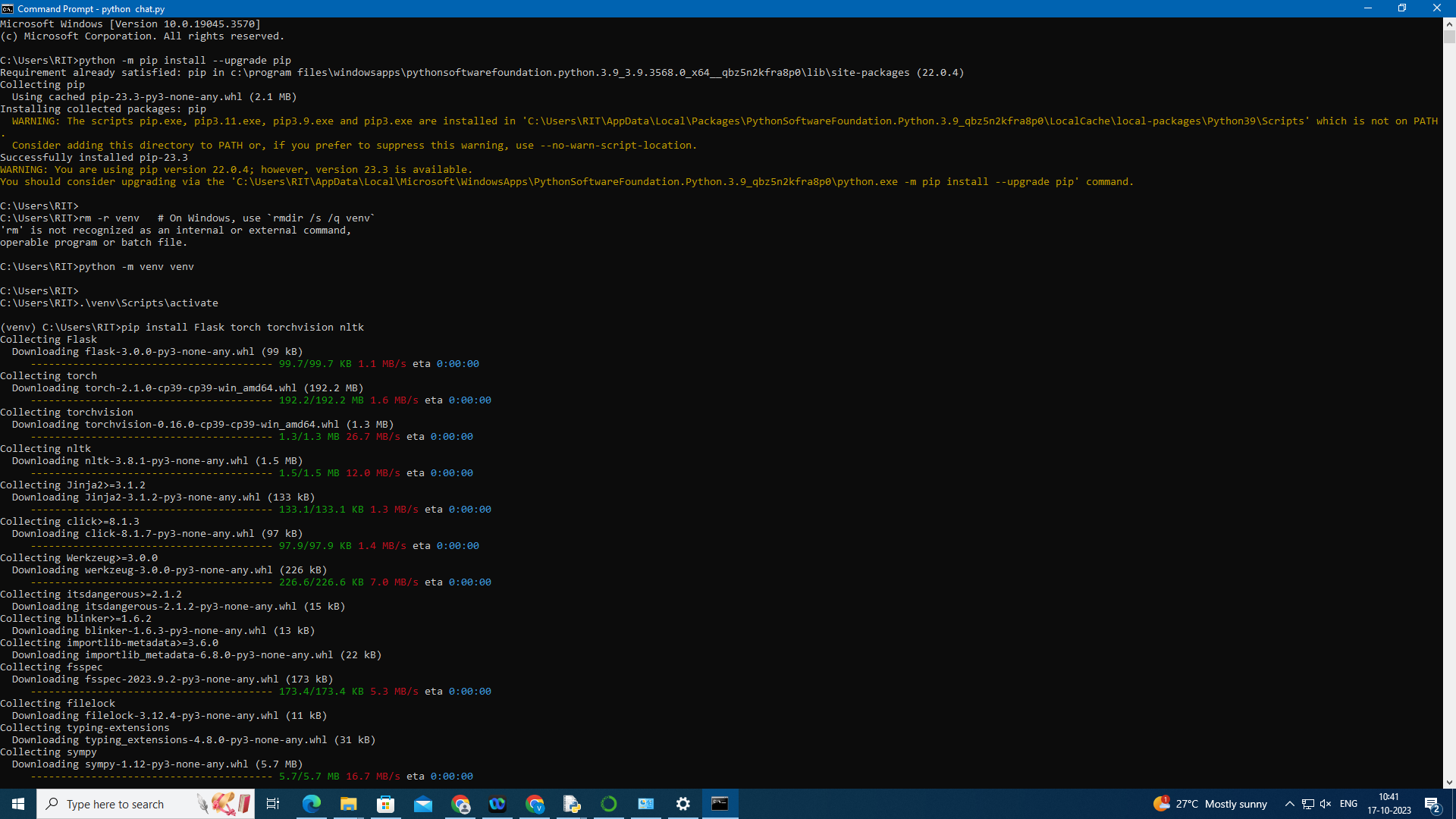
To build a chatbot and start with the development phase. Here are the steps we can follow:

**Step1:** Create a Virtual Environment

cd chatbot

python3 -m venv venv

.\venv\Scripts\activate



**Step2:** Install Required Libraries

pip install transformers

pip install flask

