

Date: 16 October 2023

Team ID: 329 PROJECT

ID: Proj_227277_Team_1

NAME: Gunasekar B

Installing Packages

1. Package name: transformers

Use: For GPT-3 integration

Command to install: pip install transformers

```
C:\Windows\system32\cmd.exe - pip install transformers
Microsoft Windows [Version 10.0.18363.1556]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\COMPAQ>pip install transformers
Defaulting to user installation because normal site-packages is not writeable
Collecting transformers
  Downloading transformers-4.34.0-py3-none-any.whl (7.7 MB)
----- 7.7/7.7 MB 3.6 MB/s eta 0:00:00
Collecting safetensors<0.3.1
  Downloading safetensors-0.4.0-cp310-none-win_amd64.whl (277 kB)
----- 277.4/277.4 kB 2.1 MB/s eta 0:00:00
Collecting tokenizers<0.15,>=0.14
  Downloading tokenizers-0.14.1-cp310-none-win_amd64.whl (2.2 MB)
----- 2.2/2.2 MB 6.4 MB/s eta 0:00:00
Collecting huggingface-hub<1.0,>=0.16.4
  Downloading huggingface_hub-0.18.0-py3-none-any.whl (301 kB)
----- 302.0/302.0 kB 3.1 MB/s eta 0:00:00
Collecting numpy>=1.17
  Downloading numpy-1.26.1-cp310-cp310-win_amd64.whl (15.8 MB)
----- 15.8/15.8 MB 2.9 MB/s eta 0:00:00
Collecting requests
  Downloading requests-2.31.0-py3-none-any.whl (62 kB)
----- 62.6/62.6 kB 844.1 kB/s eta 0:00:00
Collecting packaging>=20.0
  Downloading packaging-23.2-py3-none-any.whl (53 kB)
----- 53.0/53.0 kB 910.0 kB/s eta 0:00:00
Collecting filelock
  Downloading filelock-3.12.4-py3-none-any.whl (11 kB)
----- 11.0/11.0 kB 1.4 MB/s eta 0:00:00
Collecting tqdm>=4.27
  Downloading tqdm-4.66.1-py3-none-any.whl (78 kB)
----- 78.3/78.3 kB 1.4 MB/s eta 0:00:00
Collecting regex!=2019.12.17
  Downloading regex-2023.10.3-cp310-cp310-win_amd64.whl (269 kB)
----- 269.6/269.6 kB 2.1 MB/s eta 0:00:00
Collecting pyyaml>=5.1
  Downloading PyYAML-6.0.1-cp310-cp310-win_amd64.whl (145 kB)
----- 145.3/145.3 kB 958.1 kB/s eta 0:00:00
Collecting fsspec>=2023.5.0
  Downloading fsspec-2023.9.2-py3-none-any.whl (173 kB)
----- 173.4/173.4 kB 2.1 MB/s eta 0:00:00
Collecting typing-extensions>=3.7.4.3
  Downloading typing_extensions-4.8.0-py3-none-any.whl (31 kB)
----- 31.0/31.0 kB 1.4 MB/s eta 0:00:00
Collecting huggingface-hub<1.0,>=0.16.4
  Downloading huggingface_hub-0.17.3-py3-none-any.whl (295 kB)
```

2. Package name: Flask

Use: For web app development

Command to install: pip install Flask

```
Microsoft Windows [Version 10.0.18363.1556]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\COMPAG>pip install Flask
Defaulting to user installation because normal site-packages is not writeable
Collecting Flask
  Downloading Flask-3.0.0-py3-none-any.whl (99 kB)
----- 99.7/99.7 kB 440.8 kB/s eta 0:00:00
Collecting blinker>=1.6.2
  Downloading blinker-1.6.3-py3-none-any.whl (13 kB)
Collecting click>=8.1.3
  Downloading click-8.1.7-py3-none-any.whl (97 kB)
----- 97.9/97.9 kB 622.0 kB/s eta 0:00:00
Collecting itsdangerous>=2.1.2
  Downloading itsdangerous-2.1.2-py3-none-any.whl (15 kB)
Collecting Jinja2>=3.1.2
  Downloading Jinja2-3.1.2-py3-none-any.whl (133 kB)
----- 133.1/133.1 kB 1.1 MB/s eta 0:00:00
Collecting Werkzeug>=3.0.0
  Downloading Werkzeug-3.0.0-py3-none-any.whl (226 kB)
----- 226.6/226.6 kB 1.3 MB/s eta 0:00:00
Requirement already satisfied: colorama in c:\users\compag\appdata\roaming\python\python310\site-packages (from click>=8.1.3->Flask) (0.4.6)
Collecting MarkupSafe>=2.0
  Downloading MarkupSafe-2.1.3-cp310-cp310-win_amd64.whl (17 kB)
Installing collected packages: MarkupSafe, itsdangerous, click, blinker, Werkzeug, Jinja2, Flask
WARNING: The script flask.exe is installed in 'C:\Users\COMPAG\AppData\Roaming\Python\Python310\Scripts' which is not on PATH.
Consider adding this directory to PATH on, if you prefer to suppress this warning, use --no-warn-script-location.
```

Program for basic chat bot conversation

I provided source code file called “AI_Phase3_source_code.ipynb” in my git hub repository

```
# import all required libraries
import numpy as np
import string
from nltk.corpus import stopwords
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.feature_extraction.text import CountVectorizer
```

```

from sklearn.neural_network import MLPClassifier
from sklearn.feature_extraction.text import TfidfTransformer,TfidfVectorizer
from sklearn.pipeline import Pipeline

```

```

# importing the dataset
df = pd.read_csv(r"C:\Users\COMPAQ\Desktop\IBM\dataset\dialogs.txt", sep='\t')

df.head()

```

	hi, how are you doing?	i'm fine. how about yourself?
0	i'm fine. how about yourself?	i'm pretty good. thanks for asking.
1	i'm pretty good. thanks for asking.	no problem. so how have you been?
2	no problem. so how have you been?	i've been great. what about you?
3	i've been great. what about you?	i've been good. i'm in school right now.
4	i've been good. i'm in school right now.	what school do you go to?

#add column names

```

df.columns=['Questions','Answers']
df

```

	Questions	Answers
0	i'm fine. how about yourself?	i'm pretty good. thanks for asking.
1	i'm pretty good. thanks for asking.	no problem. so how have you been?
2	no problem. so how have you been?	i've been great. what about you?
3	i've been great. what about you?	i've been good. i'm in school right now.
4	i've been good. i'm in school right now.	what school do you go to?
...
3719	that's a good question. maybe it's not old age.	are you right-handed?
3720	are you right-handed?	yes. all my life.
3721	yes. all my life.	you're wearing out your right hand. stop using...
3722	you're wearing out your right hand. stop using...	but i do all my writing with my right hand.
3723	but i do all my writing with my right hand.	start typing instead. that way your left hand ...
3724 rows × 2 columns		

```
# Data Preprocessing
# Assuming you have a DataFrame 'df' with 'Questions' and 'Answers' columns

df['question tokens'] = df['Questions'].apply(lambda x: len(x.split()))
df['answer tokens'] = df['Answers'].apply(lambda x: len(x.split()))

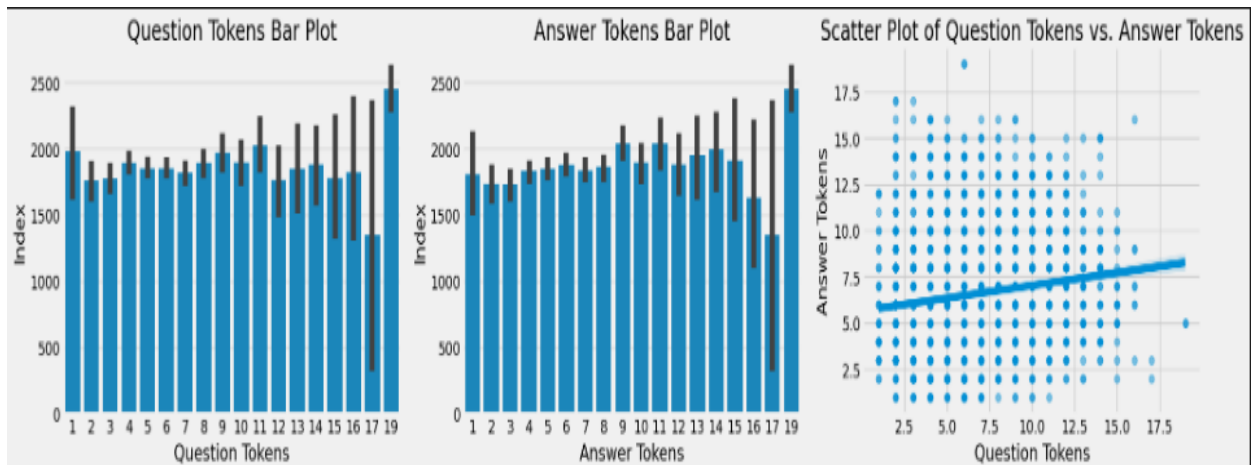
plt.style.use('fivethirtyeight')
fig, ax = plt.subplots(nrows=1, ncols=3, figsize=(20, 5))
sns.set_palette('Set2')

# Create bar plots for question tokens and answer tokens
sns.barplot(x='question tokens', y=df.index, data=df, ax=ax[0])
ax[0].set_xlabel('Question Tokens')
ax[0].set_ylabel('Index')
ax[0].set_title('Question Tokens Bar Plot')

sns.barplot(x='answer tokens', y=df.index, data=df, ax=ax[1])
ax[1].set_xlabel('Answer Tokens')
ax[1].set_ylabel('Index')
ax[1].set_title('Answer Tokens Bar Plot')

# Create a scatter plot with a regression line for the relationship between question and
answer tokens
sns.regplot(x='question tokens', y='answer tokens', data=df, ax=ax[2],
scatter_kws={'alpha':0.5})
ax[2].set_xlabel('Question Tokens')
ax[2].set_ylabel('Answer Tokens')
ax[2].set_title('Scatter Plot of Question Tokens vs. Answer Tokens')

plt.tight_layout()
plt.show()
```



#Function for converting upper to lower case

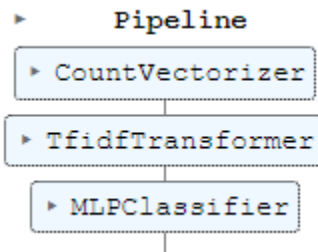
def cleaner(x):

return [a for a in (''.join([a for a in x if a not in string.punctuation])).lower().split()]

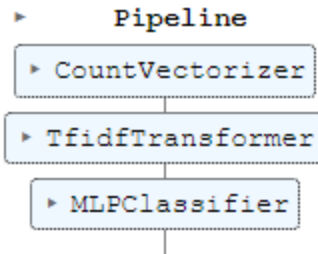
#Model

```
Pipe = Pipeline([
    ('bow',CountVectorizer(analyzer=cleaner)),
    ('tfidf',TfidfTransformer()),
    ('classifier',MLPClassifier())
])
```

Pipe.fit(df['Questions'],df['Answers'])



Pipe.fit(df['Questions'],df['Answers'])



#Text

```
Pipe.predict(['like how clear the sky gets after it rains.'])[0]
```

```
'i feel the same way. it smells so good after it rains.'
```

```
Pipe.predict(['i want this trip to be perfect, i hope it stays warm.'])[0]
```

```
"this california weather is so uncertain, it's impossible to know what'll happen."
```

```
Pipe.predict(['it would not be good if it got cold this weekend.'])[0]
```

```
'i want this trip to be perfect, i hope it stays warm.'
```

```
Pipe.predict(['it would be nice if the weather would never change.'])[0]
```

```
'that would be great, then we could plan things sooner.'
```

```
Pipe.predict(['why is that?'])[0]
```

```
'because i love the snow.'
```

```
Pipe.predict(['What are you doing'])[0]
```

```
"i'm going to change the light bulb. it burnt out."
```