Date: 16 oct 2023

Team ID: 329

Project ID: Proj_227277_Team_1

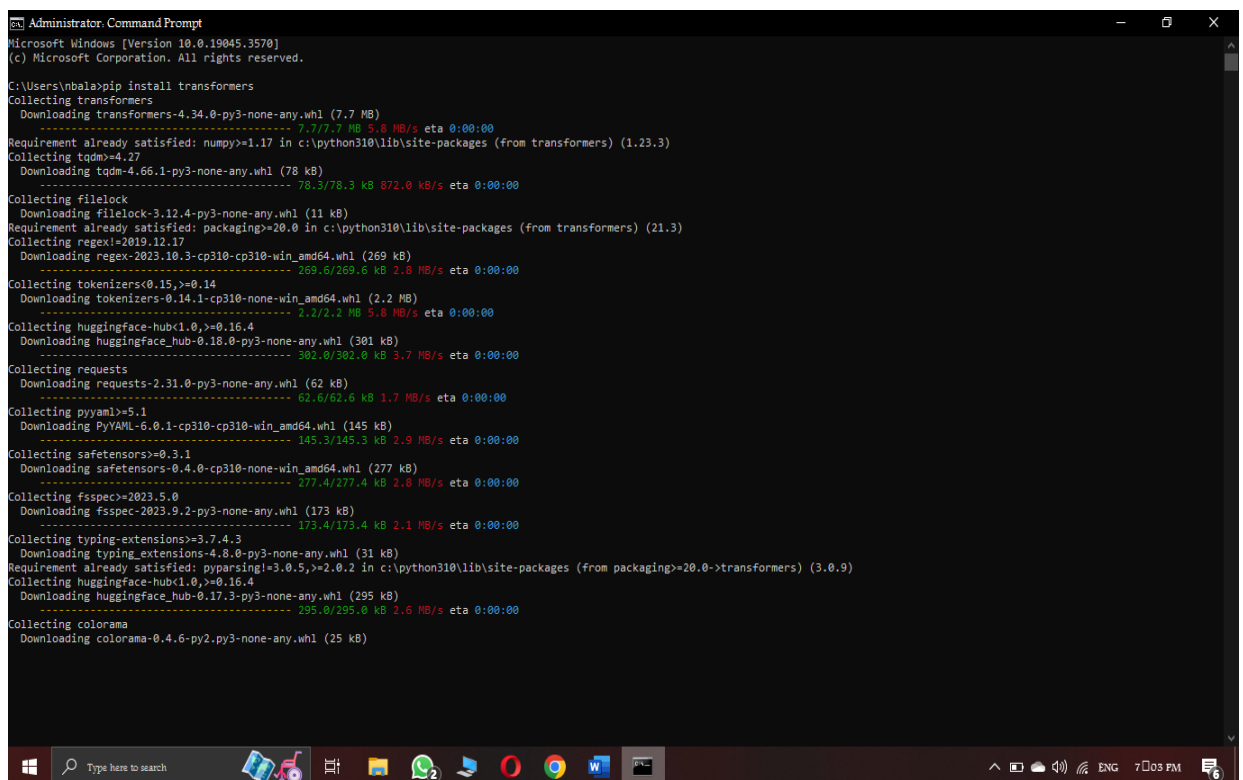**Name: B. Hemanth**

Installation of Required Packages

**1.Package Name: Transformers
Use: For Gpt-3 integration
Command to install: pip install
transformers**

2.Package Name: Flask
Use: for web app development
Command to install: pip install flask

Program for basic chat bot creation

I provided source code file in my git hub repository.

```python
# 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.svm import SVC
from sklearn.feature_extraction.text import
TfidfTransformer,TfidfVectorizer
from sklearn.pipeline import Pipeline

# importing the dataset
df =
pd.read_csv(r"C:\Users\nbala\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?

```python
#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

```python
# Data Preprocessing

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

# Create different types of plots
plt.style.use('seaborn-darkgrid')
fig, ax = plt.subplots(nrows=2, ncols=2, figsize=(15, 10))

# Histograms
sns.histplot(x='question tokens', data=df, kde=True, ax=ax[0, 0], color='skyblue')
sns.histplot(x='answer tokens', data=df, kde=True, ax=ax[0, 1], color='salmon')

# KDE Plots
sns.kdeplot(x='question tokens', data=df, ax=ax[1, 0], color='skyblue')
```

```
sns.kdeplot(x='answer tokens', data=df, ax=ax[1, 1],
color='salmon')

# Titles and labels
ax[0, 0].set_title('Question Tokens Histogram')
ax[0, 1].set_title('Answer Tokens Histogram')
ax[1, 0].set_title('Question Tokens KDE')
ax[1, 1].set_title('Answer Tokens KDE')
# Adding labels
ax[0, 0].set_xlabel('Tokens')
ax[0, 1].set_xlabel('Tokens')
ax[1, 0].set_xlabel('Tokens')
ax[1, 1].set_xlabel('Tokens')

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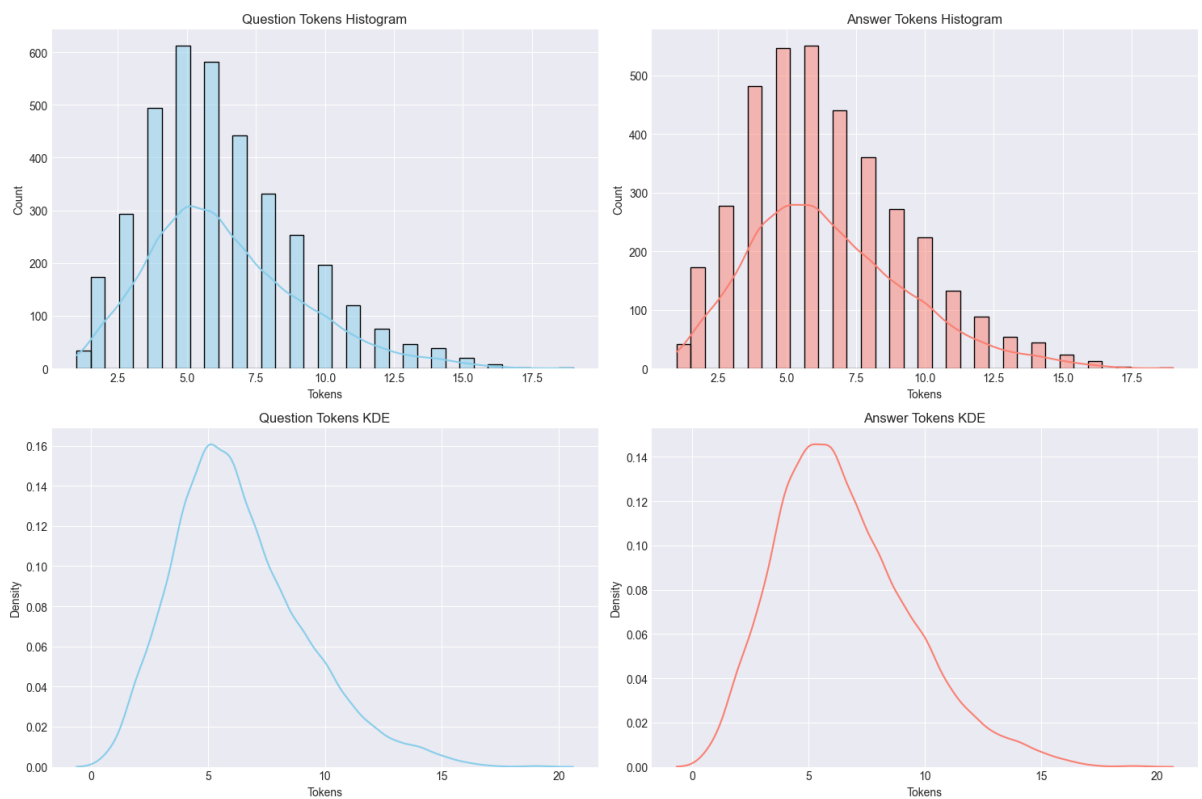


```
#Function for converting upper to lower case
def cleaner(x):
```

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

Pipe = Pipeline([
    ('bow',CountVectorizer(analyzer=cleaner)),
    ('tfidf',TfidfTransformer()),
    ('classifier',SVC(kernel='linear'))
])

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

```
▶           Pipeline

  ▶ CountVectorizer

  ▶ TfidfTransformer

      ▶ SVC
```

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

```
▶           Pipeline

  ▶ CountVectorizer

  ▶ TfidfTransformer

      ▶ SVC
```
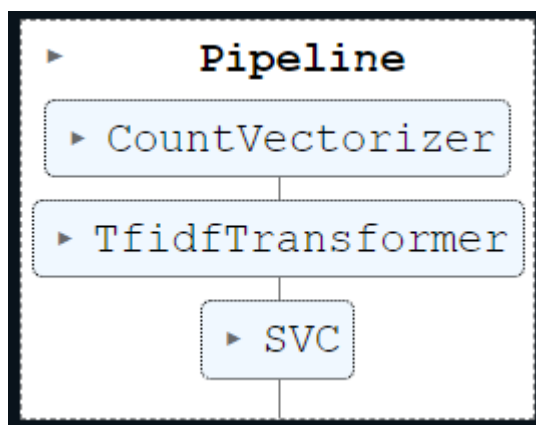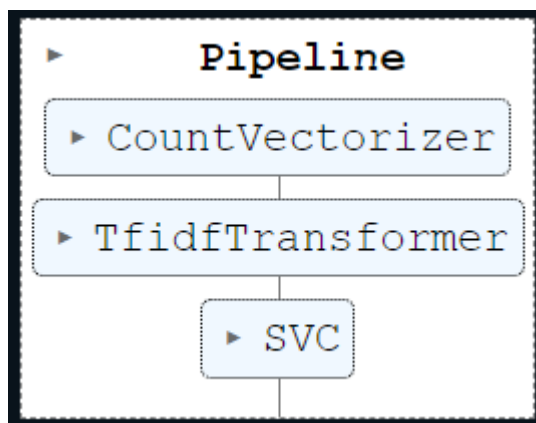
```python
Pipe.predict(["how's it going?"])[0]
```

```
"i'm doing well. how about you?"
```

Pipe.predict(["i'm fine. how about yourself?"])[0]

```
"i'm pretty good. thanks for asking."
```

Pipe.predict(["i'm actually in school right now."])[0]

```
'what school do you go to?'
```

Pipe.predict(["which school do you attend?"])[0]

```
"i'm attending pcc right now."
```

Pipe.predict(["great"])[0]

```
'i appreciate that.'
```

Pipe.predict(["it wouldn't rain in the middle of the summer."])[0]

```
"it wouldn't seem right if it started raining right now."
```