

**Date: 16 oct 2023**

**Team ID: 329**

**Project ID: Proj\_227277\_Team\_1**

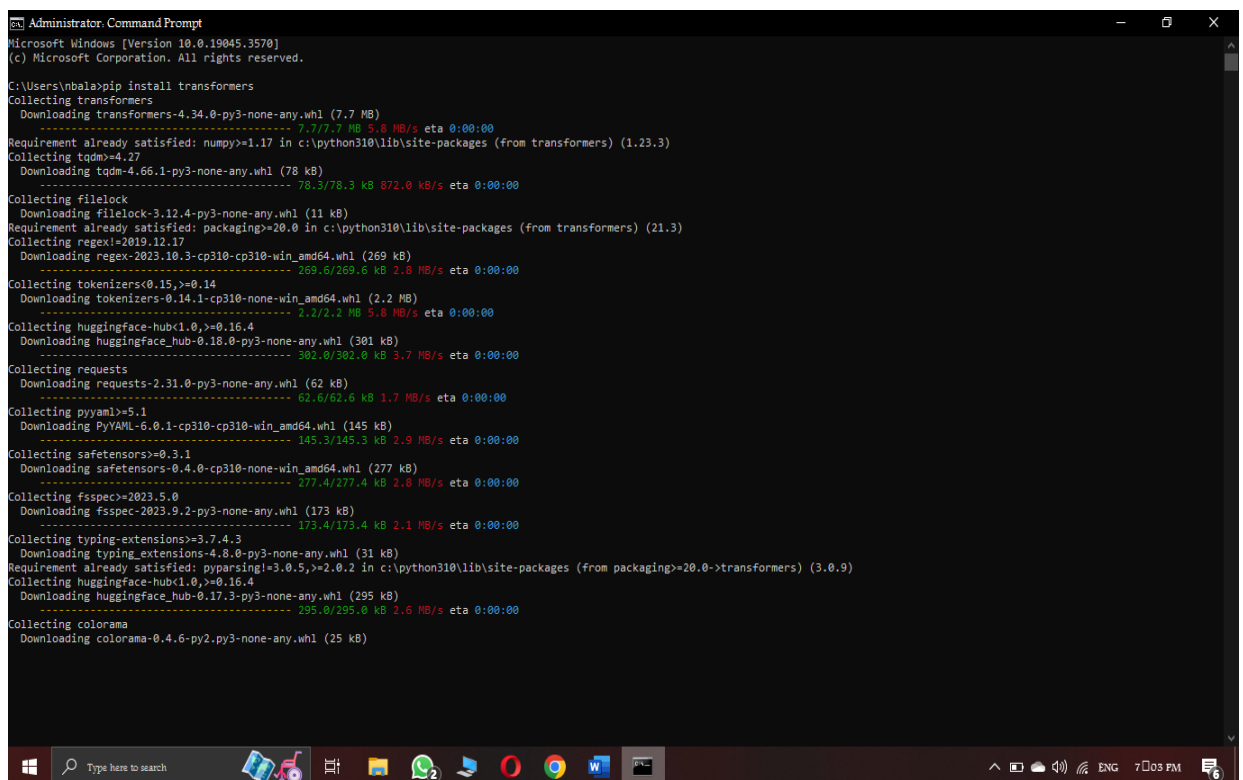
**Name: B. Hemanth**

# Installation of Required Packages

## 1. Package Name: Transformers

**Use: For Gpt-3 integration**

**Command to install: pip install transformers**



```
Administrator: Command Prompt
Microsoft Windows [Version 10.0.19045.3570]
(c) Microsoft Corporation. All rights reserved.

C:\Users\nbalax>pip install transformers
Collecting transformers
  Downloading transformers-4.34.0-py3-none-any.whl (7.7 MB)
----- 7.7/7.7 MB 5.8 MB/s eta 0:00:00
Requirement already satisfied: numpy>=1.17 in c:\python310\lib\site-packages (from transformers) (1.23.3)
Collecting tqdm>=4.27
  Downloading tqdm-4.66.1-py3-none-any.whl (78 kB)
----- 78.3/78.3 kB 872.0 kB/s eta 0:00:00
Collecting filelock
  Downloading filelock-3.12.4-py3-none-any.whl (11 kB)
Requirement already satisfied: packaging>=20.0 in c:\python310\lib\site-packages (from transformers) (21.3)
Collecting regex!=2019.12.17
  Downloading regex-2023.10.3-cp310-cp310-win_amd64.whl (269 kB)
----- 269.6/269.6 kB 2.8 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 5.8 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 3.7 MB/s eta 0:00:00
Collecting requests
  Downloading requests-2.31.0-py3-none-any.whl (62 kB)
----- 62.6/62.6 kB 1.7 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 2.9 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 2.8 MB/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)
Requirement already satisfied: pyarsing!=3.0.5,>=2.0.2 in c:\python310\lib\site-packages (from packaging>=20.0->transformers) (3.0.9)
Collecting huggingface-hub<1.0,>=0.16.4
  Downloading huggingface_hub-0.17.3-py3-none-any.whl (295 kB)
----- 295.0/295.0 kB 2.6 MB/s eta 0:00:00
Collecting colorama
  Downloading colorama-0.4.6-py2.py3-none-any.whl (25 kB)
```

## 2.Package Name: Flask

Use: for web app development

Command to install: `pip install flask`

```
Administrator: Command Prompt
Collecting Jinja2>=3.1.2 (from flask)
  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 itsdangerous>=2.1.2 (from flask)
  Downloading itsdangerous-2.1.2-py3-none-any.whl (15 kB)
Collecting click>=8.1.3 (from flask)
  Obtaining dependency information for click>=8.1.3 from https://files.pythonhosted.org/packages/00/2e/d53fa4befbf2cfa713304affc7ca780ce4fc1fd8710527771b58311a3229/click-8.1.7-py3-none-any.whl.metadata
  Downloading click-8.1.7-py3-none-any.whl.metadata (3.0 kB)
Collecting blinker>=1.6.2 (from flask)
  Obtaining dependency information for blinker>=1.6.2 from https://files.pythonhosted.org/packages/bf/2b/11bcedb7dee4923253a4a21bae3be854bcc4f06295bd827756352016d97c/blinker-1.6.3-py3-none-any.whl.metadata
  Downloading blinker-1.6.3-py3-none-any.whl.metadata (1.9 kB)
Requirement already satisfied: colorama in c:\python310\lib\site-packages (from click>=8.1.3->flask) (0.4.6)
Collecting MarkupSafe>=2.0 (from Jinja2>=3.1.2->flask)
  Obtaining dependency information for MarkupSafe>=2.0 from https://files.pythonhosted.org/packages/84/a8/c4aebb8a14a1d39d5135eb8233a0b95831cdc42c4088358449c3ed657044/MarkupSafe-2.1.3-cp310-cp310-win_amd64.whl.metadata
  Downloading MarkupSafe-2.1.3-cp310-cp310-win_amd64.whl.metadata (3.1 kB)
  Downloading flask-3.0.0-py3-none-any.whl (99 kB)
----- 99.7/99.7 kB 715.9 kB/s eta 0:00:00
Download blinker-1.6.3-py3-none-any.whl (13 kB)
Download click-8.1.7-py3-none-any.whl (97 kB)
----- 97.9/97.9 kB 1.4 MB/s eta 0:00:00
Download werkzeug-3.0.0-py3-none-any.whl (226 kB)
----- 226.6/226.6 kB 2.3 MB/s eta 0:00:00
Download MarkupSafe-2.1.3-cp310-cp310-win_amd64.whl (17 kB)
Installing collected packages: MarkupSafe, itsdangerous, click, blinker, Werkzeug, Jinja2, flask
Successfully installed Jinja2-3.1.2 MarkupSafe-2.1.3 Werkzeug-3.0.0 blinker-1.6.3 click-8.1.7 flask-3.0.0 itsdangerous-2.1.2
C:\Users\nbala>
```

# Program for basic chat bot creation

I provided source code file 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.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?

```
#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

```
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')
```

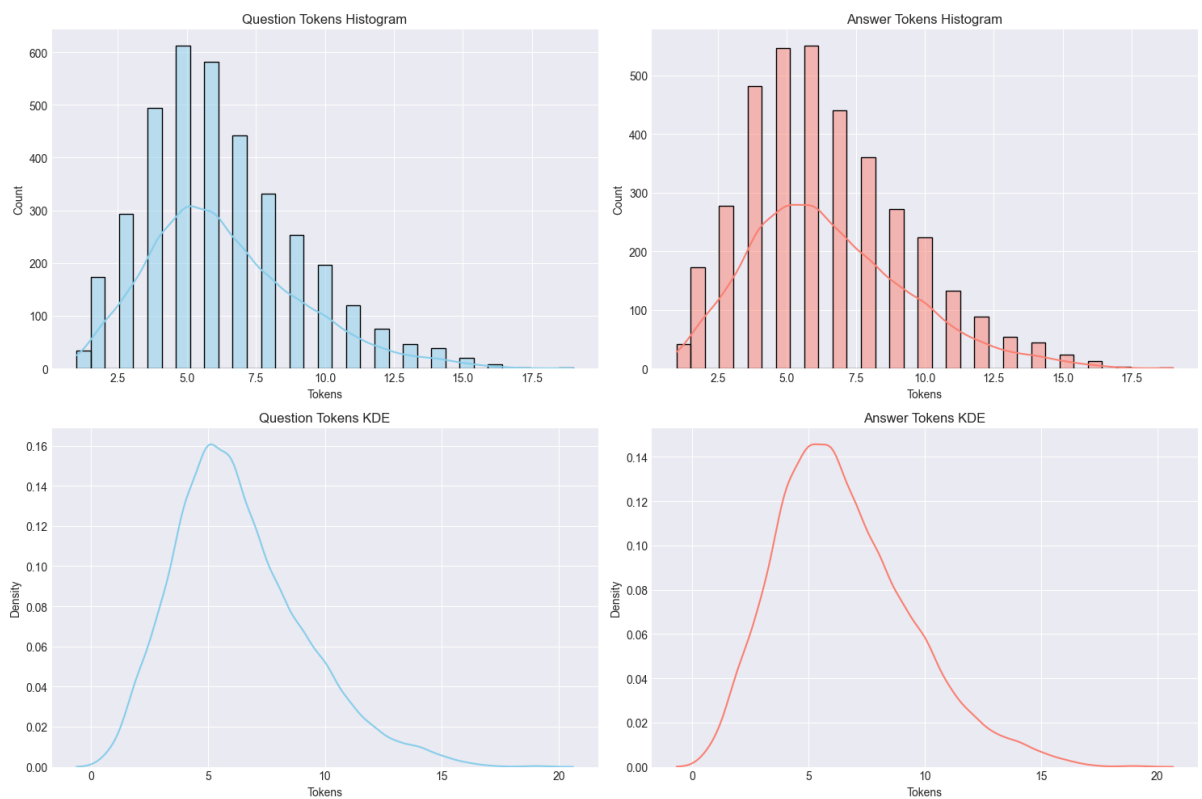
```
ax[0, 1].set_xlabel('Tokens')
```

```
ax[1, 0].set_xlabel('Tokens')
```

```
ax[1, 1].set_xlabel('Tokens')
```

```
plt.tight_layout()
```

```
plt.show()
```



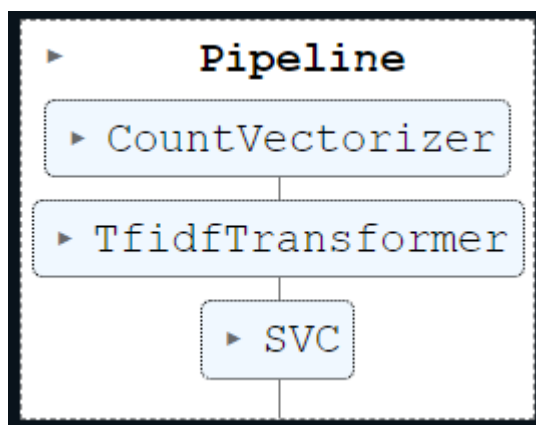
```
#Function for converting upper to lower case
```

```
def cleaner(x):
```

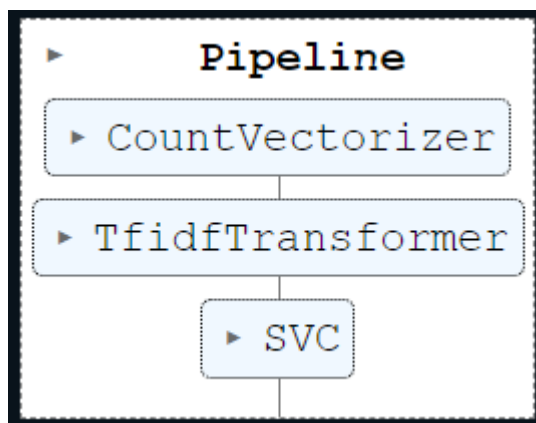
```
return [a for a in ("join([a for a in x if a not in
string.punctuation])).lower().split()]
```

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

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



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



```
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."
```