```
!pip install spacy
!python -m spacy download en_core_web_sm
!pip install beautifulsoup4
!pip install textblobm
```

```
Requirement already satisfied: spacy in /usr/local/lib/python3.10/dist-packages (3.7.4)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.0.10)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.0.9)
Requirement already satisfied: thinc<8.3.0,>=8.2.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (8.2.3)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.1.2)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.4.8)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.0.10)
Requirement already satisfied: weasel<0.4.0,>=0.1.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (0.3.4)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (0.9.4)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /usr/local/lib/python3.10/dist-packages (from spacy) (6.4.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (4.66.4)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.31.0)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.7.
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.1.4)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from spacy) (67.7.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (24.0)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.4.0)
Requirement already satisfied: numpy>=1.19.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.25.2)
Requirement already satisfied: language-data>=1.2 in /usr/local/lib/python3.10/dist-packages (from langcodes<4.0.0,>=3.2.0->spacy
Requirement already satisfied: annotated-types>=0.4.0 in /usr/local/lib/python3.10/dist-packages (from pydantic!=1.8,!=1.8.1,<3.0
Requirement already satisfied: pydantic-core==2.18.2 in /usr/local/lib/python3.10/dist-packages (from pydantic!=1.8,!=1.8.1,<3.0.
Requirement already satisfied: typing-extensions>=4.6.1 in /usr/local/lib/python3.10/dist-packages (from pydantic!=1.8,!=1.8.1,<3
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0-
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.7
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /usr/local/lib/python3.10/dist-packages (from thinc<8.3.0,>=8.2.2->spacy) (0
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.10/dist-packages (from thinc<8.3.0,>=8.2.2->spa
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.10/dist-packages (from typer<0.10.0,>=0.3.0->spacy)
Requirement already satisfied: cloudpathlib<0.17.0,>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from weasel<0.4.0,>=0.1.0-
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from jinja2->spacy) (2.1.5)
Requirement already satisfied: marisa-trie>=0.7.7 in /usr/local/lib/python3.10/dist-packages (from language-data>=1.2->langcodes<
Collecting en-core-web-sm==3.7.1
  Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.7.1/en_core_web_sm-3.7.1-py3-none-any.
  ────────────────────────────────────────── 12.8/12.8 MB 28.8 MB/s eta 0:00:00
Requirement already satisfied: spacy<3.8.0,>=3.7.2 in /usr/local/lib/python3.10/dist-packages (from en-core-web-sm==3.7.1) (3.7.4
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-cor
Requirement already satisfied: thinc<8.3.0,>=8.2.2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-core
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-c
Requirement already satisfied: weasel<0.4.0,>=0.1.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-core
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-core
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-c
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-
```

```
#load the datset -> feature extration -> data visulization
# data cleaning-> train_test_split->model building
#->model training-> model evaluation->model saving->stramilit application deploy


import warnings
warnings.filterwarnings('ignore')

import pandas as pd


df=pd.read_csv('https://raw.githubusercontent.com/laxmimerit/All-CSV-ML-Data-Files-Download/master/twitter_sentiment.csv', header=None,
df.head()
```

		1	2		3
	0				
2401	Borderlands	Positive		im getting on borderlands and i will murder yo...
2401	Borderlands	Positive		I am coming to the borders and I will kill you...
2401	Borderlands	Positive		im getting on borderlands and i will kill you ...
2401	Borderlands	Positive		im coming on borderlands and i will murder you...
2401	Borderlands	Positive		im getting on borderlands 2 and i will murder ...

```python
df=df[[2,3]].reset_index(drop=True)
df.head()
```

	2	**3**
0	Positive	im getting on borderlands and i will murder yo...
1	Positive	I am coming to the borders and I will kill you...
2	Positive	im getting on borderlands and i will kill you ...
3	Positive	im coming on borderlands and i will murder you...
4	Positive	im getting on borderlands 2 and i will murder ...

```python
df.columns=['sentiment','text']
df.head()
```

	sentiment	**text**
0	Positive	im getting on borderlands and i will murder yo...
1	Positive	I am coming to the borders and I will kill you...
2	Positive	im getting on borderlands and i will kill you ...
3	Positive	im coming on borderlands and i will murder you...
4	Positive	im getting on borderlands 2 and i will murder ...

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 75682 entries, 0 to 75681
Data columns (total 2 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   sentiment  75682 non-null  object
 1   text       74996 non-null  object
dtypes: object(2)
memory usage: 1.2+ MB
```

```python
df.isnull().sum()
df.dropna(inplace=True)
```

```python
df=df[df['text'].apply(len)>5]
```

```python
df.shape
```

```
(72883, 2)
```

```python
df['sentiment'].value_counts()
```

```
sentiment
Negative      22020
Positive      20186
Neutral       17899
Irrelevant    12778
Name: count, dtype: int64
```

```python
pip install git+https://github.com/laxmimerit/preprocess_kgptalkie.git --upgrade --force-reinstall
```

```
Collecting git+https://github.com/laxmimerit/preprocess_kgptalkie.git
  Cloning https://github.com/laxmimerit/preprocess_kgptalkie.git to /tmp/pip-req-build-fkdfhb2r
  Running command git clone --filter=blob:none --quiet https://github.com/laxmimerit/preprocess_kgptalkie.git /tmp/pip-req-build-fk
  Resolved https://github.com/laxmimerit/preprocess_kgptalkie.git to commit 96bf02872d9756f29d6cddb8aafaedcd2a39bbb4
```

```
      Preparing metadata (setup.py) ... done
  Building wheels for collected packages: preprocess-kgptalkie
    Building wheel for preprocess-kgptalkie (setup.py) ... done
    Created wheel for preprocess-kgptalkie: filename=preprocess_kgptalkie-0.1.3-py3-none-any.whl size=7603 sha256=2bea78e17aaef19f99b4
    Stored in directory: /tmp/pip-ephem-wheel-cache-2s0nd5ls/wheels/5c/94/34/99d5ff65e88b8d9a6c5e8d8652f2311d87790a61a1b7466e21
  Successfully built preprocess-kgptalkie
  Installing collected packages: preprocess-kgptalkie
  Successfully installed preprocess-kgptalkie-0.1.3
```

```python
import preprocess_kgptalkie as kgp
```

```python
df.columns
```

```
Index(['sentiment', 'text'], dtype='object')
```

```python
df=kgp.get_basic_features(df)
```

```python
df.head()
```

	sentiment	text	char_counts	word_counts	avg_wordlength	stopwords_counts
0	Positive	im getting on borderlands and i will murder yo...	43	11	3.909091	6
		I am				

```python
import matplotlib.pyplot as plt
import seaborn as sns
```

```python
plt.figure(figsize=(30,20))
num_cols=df.select_dtypes(include='number').columns
for index,col in enumerate(num_cols):
  plt.subplot(2,4,index+1)
  sns.kdeplot(data=df, x=col,hue='sentiment',fill=True)
```
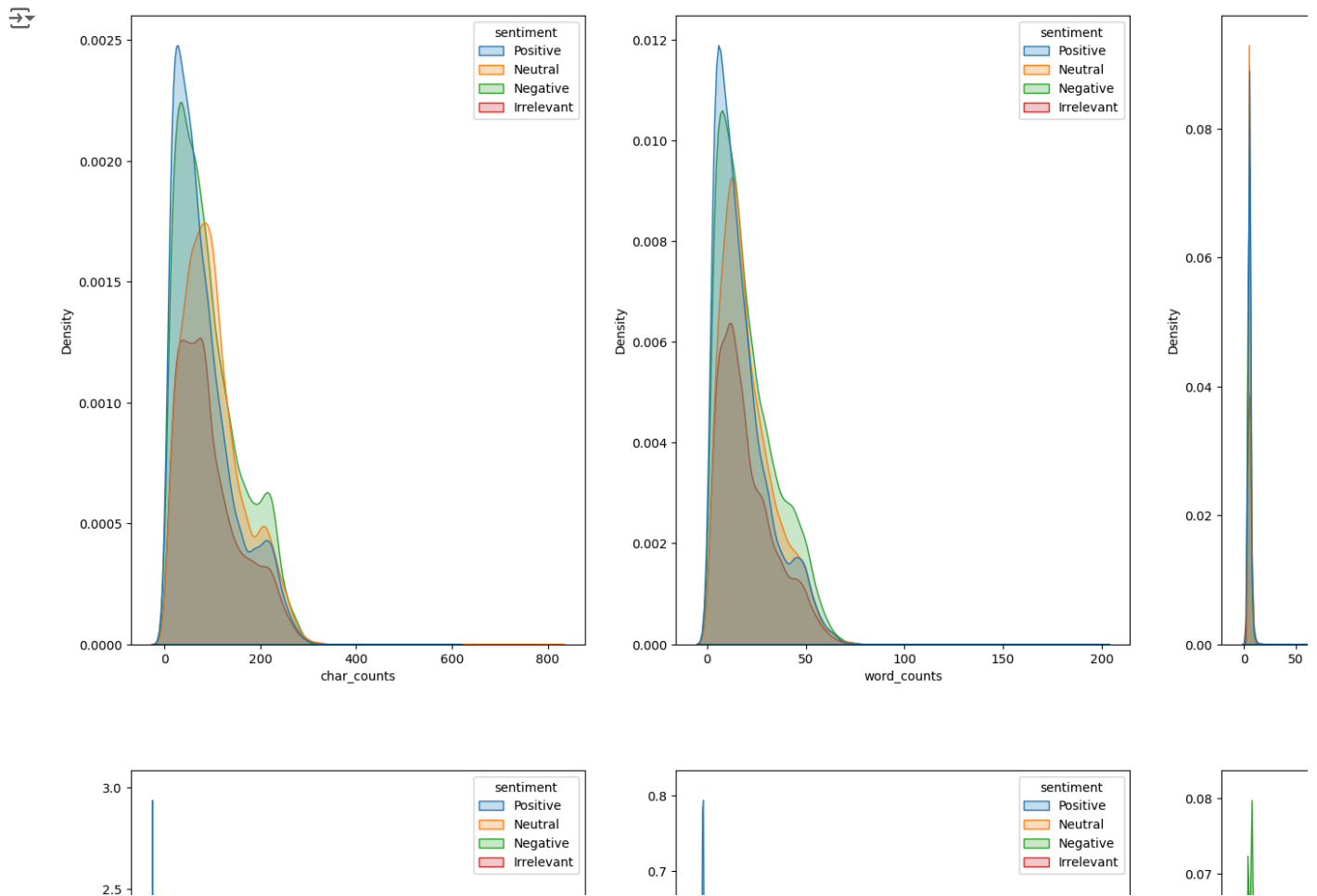
```
df['sentiment'].value_counts().plot(kind='pie',autopct="%1.0f%%")
```

```
<Axes: ylabel='count'>
```
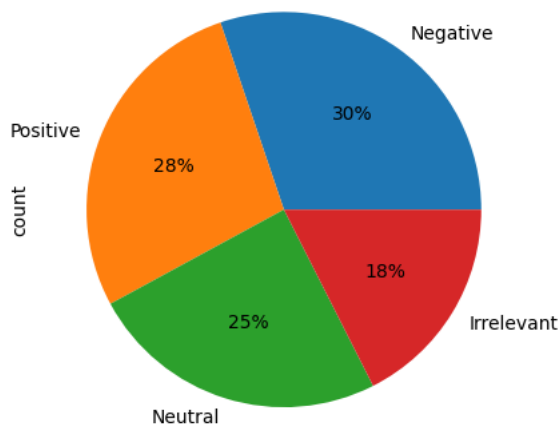


```
#worldcloud
!pip install wordcloud
```

```
Requirement already satisfied: wordcloud in /usr/local/lib/python3.10/dist-packages (1.9.3)
Requirement already satisfied: numpy>=1.6.1 in /usr/local/lib/python3.10/dist-packages (from wordcloud) (1.25.2)
Requirement already satisfied: pillow in /usr/local/lib/python3.10/dist-packages (from wordcloud) (9.4.0)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-packages (from wordcloud) (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (1.2.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (4.51.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (1.4.5)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (24.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplotlib->wordclou
```

```
from wordcloud import WordCloud, STOPWORDS
```

```
stopwords=set(STOPWORDS)
```

```
wordcloud=WordCloud(background_color='white',stopwords=stopwords,max_words=300,max_font_size=40,scale=5).generate(str(df['text']))
plt.imshow(wordcloud)
```

<matplotlib.image.AxesImage at 0x786538b7f6a0>


```
plt.figure(figsize=(40,20))
```

```
for index, sent in enumerate(df['sentiment'].unique()):
  plt.subplot(2,2,index+1)
  data=df[df['sentiment']==sent]['text']
  wordcloud=WordCloud(background_color='white',stopwords=stopwords,max_words=300,max_font_size=40,scale=5).generate(str(df['text']))
  plt.imshow(wordcloud)
  plt.xticks([]),plt.yticks([])
  plt.title(sent,fontsize=40)
plt.show
```

matplotlib.pyplot.show
```
def show(*args, **kwargs)
```

Display all open figures.

Parameters
----
block : bool, optional
    Whether to wait for all figures to be closed before returning.

Positive

Negative

˅ data cleaning

```python
df['text']=df['text'].apply(lambda x:x.lower())
df['text']=df['text'].apply(lambda x:kgp.remove_urls(x))
df['text']=df['text'].apply(lambda x:kgp.remove_html_tags(x))
df['text']=df['text'].apply(lambda x:kgp.remove_rt(x))
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(df['text'],df['sentiment'],test_size=0.2,random_state=0)
```

```python
X_train
```

```
70828                             st ghostrecon
44616    okay i was scared and didnt know what was goin...
64280    this bull helvick is a problem eamaddennfl eas...
65559    johnson johnson to stop selling whitening powd...
50491             piss people off at fifa and on twitter
                        ...
22067    the construction of this building has failed d...
47727    just spend a hour in home depot home renovatio...
44342                       hey vzwsupport no response
45322    team nice try rhandlerr but your photoshopped ...
70925    was good fun watch wolves get torn apart by th...
Name: text, Length: 58306, dtype: object
```

```python
X_train.shape, X_test.shape
```
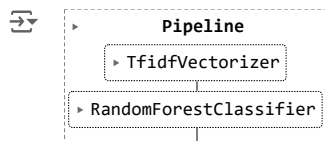
```
((58306,), (14577,))
```

model building

```python
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.ensemble import RandomForestClassifier
from sklearn.pipeline import Pipeline
```

```python
clf=Pipeline([('tfid',TfidfVectorizer()),('rfc',RandomForestClassifier(n_jobs=-1))])
```

```python
clf.fit(X_train,y_train)
```

```
        Pipeline
    ▸ TfidfVectorizer
 ▸ RandomForestClassifier
```

```python
from sklearn.metrics import classification_report
```

```python
y_pred=clf.predict(X_test)
print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

  Irrelevant       0.98      0.86      0.91      2561
    Negative       0.92      0.96      0.94      4481
     Neutral       0.93      0.92      0.92      3553
    Positive       0.91      0.95      0.93      3982

    accuracy                           0.93     14577
   macro avg       0.93      0.92      0.92     14577
weighted avg       0.93      0.93      0.93     14577
```

```python
import pickle
pickle.dump(clf,open('twitter_sentiment.pkl','wb'))
```

```python
clf.predict(['let met not upset you'])
```

```
array(['Negative'], dtype=object)
```

```python
clf.predict(['you are looking beautiful'])
```