PROG 7   ANALYSIS AND INTERPRETATION OF DATA

AIM

To manipulate the Twitter Data Set by removing the Punctuation, Numbers, Special Characters and word length<=3, tokenize the Words and Stem and generate a word cloud for the Twitter dataset and retrieve the top 15 positive and negative tags.

SOURCE CODE

```
import numpy as np

import pandas as pd

from nltk.corpus import stopwords

from nltk.tokenize import word_tokenize

import nltk

from wordcloud import WordCloud, STOPWORDS

import matplotlib.pyplot as plt

data=pd.read_csv('Twitter_Data.csv')

data=data.iloc[:50,:]

nltk.download('punkt')

nltk.download('stopwords')


print(data.head())


stop_words = set(stopwords.words('english'))


data['tokenized_sents'] = data.apply(lambda row:
nltk.word_tokenize(str(row['clean_text'])), axis=1)
for i in range(50):

    words = data.loc[i,'tokenized_sents']

    wordsFiltered = []

    for w in words:

        if w not in stop_words:

            wordsFiltered.append(w)
```

```
    data.at[i,'tokenized_sents']=wordsFiltered
```

```
print(data.head())
```

```
comment_words=''
for i in range(50):
    tokens=data.iloc[i,2]
    comment_words += " ".join(tokens)+" "
```

```
wordcloud = WordCloud(collocations = False, background_color =
'white').generate(comment_words)
```

```
# plot the WordCloud image
plt.figure(figsize = (8, 8), facecolor = None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad = 0)
```

```
plt.show()
```

OUTPUT

```
                                  clean_text  category  \
0  when modi promised "minimum government maximum...    -1.0
1  talk all the nonsense and continue all the dra...     0.0
2  what did just say vote for modi  welcome bjp t...     1.0
3  asking his supporters prefix chowkidar their n...     1.0
4  answer who among these the most powerful world...     1.0
```

```
                                  clean_text  category  \
0  when modi promised "minimum government maximum...    -1.0
1  talk all the nonsense and continue all the dra...     0.0
2  what did just say vote for modi  welcome bjp t...     1.0
3  asking his supporters prefix chowkidar their n...     1.0
4  answer who among these the most powerful world...     1.0

                                  tokenized_sents
0  [modi, promised, ", minimum, government, maxim...
1       [talk, nonsense, continue, drama, vote, modi]
2  [say, vote, modi, welcome, bjp, told, rahul, m...
3  [asking, supporters, prefix, chowkidar, names,...
4  [answer, among, powerful, world, leader, today...
```

RESULT:

Thus the program is executed and output is verified successfully.