Name: Akhil

Hall_ticket: 2403A52344

Lab: 3

Aim : Understand POS tagging challenges in informal, noisy text.

step 1:install libraties and import

```
!pip install nltk
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.12/dist-package
Requirement already satisfied: click in /usr/local/lib/python3.12/dist-packag
Requirement already satisfied: joblib in /usr/local/lib/python3.12/dist-packa
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.12/d
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-package
```

```
import nltk
from nltk.tokenize import TweetTokenizer
from nltk.corpus import twitter_samples
```

step 2:Download required nltk resources

```
nltk.download('twitter_samples')
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger_eng')
```

```
[nltk_data] Downloading package twitter_samples to /root/nltk_data...
[nltk_data]   Package twitter_samples is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger_eng to
[nltk_data]     /root/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger_eng is already up-to-
[nltk_data]       date!
True
```

step 3: Load Tweets Dataset

```
tweet=twitter_samples.strings('positive_tweets.json')
for i in range(3):
  print("Tweet",i+1)
  print(tweet[i])
  print()
```

```
Tweet 1
#FollowFriday @France_Inte @PKuchly57 @Milipol_Paris for being top engaged me
```

```
Tweet 2
@Lamb2ja Hey James! How odd :/ Please call our Contact Centre on 02392441234

Tweet 3
@DespiteOfficial we had a listen last night :) As You Bleed is an amazing tra
```

step 4: Totenization

```
tokenizer=TweetTokenizer()
preserve_case=False,
strip_handes=True,
reduce_len=True

tokenized_tweets =[tokenizer.tokenize(t) for t in tweet[:5]]

for i,token in enumerate(tokenized_tweets):
  print("Tweet",i+1, "Tokens:")
  print(token)
  print()
```

```
Tweet 1 Tokens:
['#FollowFriday', '@France_Inte', '@PKuchly57', '@Milipol_Paris', 'for', 'bei

Tweet 2 Tokens:
['@Lamb2ja', 'Hey', 'James', '!', 'How', 'odd', ':/', 'Please', 'call', 'our'

Tweet 3 Tokens:
['@DespiteOfficial', 'we', 'had', 'a', 'listen', 'last', 'night', ':)', 'As',

Tweet 4 Tokens:
['@97sides', 'CONGRATS', ':)']

Tweet 5 Tokens:
['yeaaaah', 'yippppy', '!', '!', '!', 'my', 'accnt', 'verified', 'rqst', 'has
```

step 5: pos tagging

```
rext="i love india ❤️"
token=tokenizer.tokenize(rext)
tags=nltk.pos_tag(token)
print("orginal",rext)
print("token",token)
print("pos tags",tags)
```

```
orginal i love india ❤️
token ['i', 'love', 'india', '❤', '']
pos tags [('i', 'JJ'), ('love', 'VBP'), ('india', 'NN'), ('❤', 'NN'), ('', '
```

Step 6: Extracts noven and Verd

```
Nouns=[]
VERB=[]
for word in tags:
  if word[1].startswith('NN'):
    Nouns.append(word)
  elif word[1].startswith('VB'):
    VERB.append(word)
print("Nouns",Nouns)
print("Verbs",VERB)
```

```
Nouns [('india', 'NN'), ('❤', 'NN'), ('', 'NN')]
Verbs [('love', 'VBP')]
```