

PROGRAM-11

Aim: Write a program in python for text classification for the given sentence using NLTK

Logic: Text classification also known as text categorization is the process of sorting text into categories. For example, classification of customer feedback by topic, urgency, sentiment etc. NLTK is python library which can be used to perform text classification.

Algorithm:

1. Import NLTK
2. Import random library
3. Store all sentences
4. Store movie reviews in variable
5. Use nltk.FreqDist and most_common to print result

Implementation:

```
import nltk
import random
from nltk.corpus import movie_reviews

documents = [(list(movie_reviews.words(fileid)), category)
              for category in movie_reviews.categories()
              for fileid in movie_reviews.fileids(category)]

random.shuffle(documents)
print(documents[1])

all_words = []
for w in movie_reviews.words():
    all_words.append(w.lower())

all_words = nltk.FreqDist(all_words)
print(all_words.most_common(15))
print(all_words["stupid"])
```

Output:



```
C:\Windows\System32\Windoc...
[('synopsis', 1), ('in', 1), ('this', 1), ('movie', 1), ('steven', 1), ('spielberg', 1), ('one', 1), ('of', 1), ('today', 1), ('s', 1), ('finest', 1), ('direc
tors', 1), ('attempts', 1), ('to', 1), ('spice', 1), ('up', 1), ('the', 1), ('1800s', 1), ('story', 1), ('of', 1), ('a', 1), ('long', 1), ('courtroom', 1), ('battle', 1), ('over', 1), ('th
e', 1), ('fate', 1), ('of', 1), ('prisoner', 1), ('cinque', 1), ('djimon', 1), ('hounsou', 1), ('a', 1), ('young', 1), ('angry', 1), ('man', 1), ('from', 1), ('sie
rra', 1), ('leone', 1), ('who', 1), ('was', 1), ('kidnapped', 1), ('into', 1), ('slavery', 1), ('and', 1), ('his', 1), ('fellow', 1), ('prisoners', 1), ('cinque', 1), ('s
', 1), ('and', 1), ('friends', 1), ('have', 1), ('landed', 1), ('a', 1), ('ship', 1), ('on', 1), ('the', 1), ('shores', 1), ('of', 1), ('america', 1), ('after', 1), ('escaping', 1), ('spanish', 1), ('s
lave', 1), ('traders', 1), ('but', 1), ('since', 1), ('the', 1), ('americans', 1), ('don', 1), ('t', 1), ('speak', 1), ('cinque', 1), ('s', 1), ('language', 1), ('the', 1), ('black', 1), ('men', 1), ('are', 1), ('hauled', 1), ('into', 1), ('court', 1), ('to', 1), ('determine', 1), ('whether', 1), ('or', 1), ('not', 1), ('they', 1), ('are', 1), ('legally', 1), ('slaves', 1), ('technically', 1), ('since', 1), ('the', 1), ('international', 1), ('slave', 1), ('trade', 1), ('was', 1), ('outlawed', 1), ('at', 1), ('that', 1), ('t
ime', 1), ('people', 1), ('like', 1), ('cinque', 1), ('couldn', 1), ('t', 1), ('be', 1), ('kidnapped', 1), ('into', 1), ('slavery', 1), ('one', 1), ('had', 1), ('to', 1), ('be', 1), ('born', 1), ('a', 1), ('slave', 1), ('to', 1), ('be', 1), ('legally', 1), ('considered', 1), ('a', 1), ('slave', 1), ('lawyers', 1), ('baldwin', 1), ('and', 1), ('adams', 1), ('c
', 1), ('matthew', 1), ('mcconaughey', 1), ('and', 1), ('anthony', 1), ('hopkins', 1), ('must', 1), ('prove', 1), ('cinque', 1), ('and', 1), ('the', 1), ('others', 1), ('were', 1), ('captured', 1), ('into', 1), ('slavery', 1), ('rather', 1), ('than', 1), ('born', 1), ('slaves', 1), ('in', 1), ('order', 1), ('to', 1), ('get', 1), ('them', 1), ('out', 1), ('of', 1), ('prison', 1), ('as', 1), ('free', 1), ('men', 1), ('three', 1), ('lengthy', 1), ('court', 1), ('cases', 1), ('are', 1), ('portrayed', 1), ('with', 1), ('spielberg', 1), ('s', 1), ('trademark', 1), ('panache', 1), ('flashy', 1), ('beginning', 1), ('lots', 1), ('of', 1), ('facial', 1), ('close', 1), ('ups', 1), ('big', 1), ('anti
climax', 1), ('where', 1), ('subtitles', 1), ('show', 1), ('what', 1), ('eventually', 1), ('happened', 1), ('to', 1), ('the', 1), ('various', 1), ('characters', 1), ('opinio
n', 1), ('making', 1), ('a', 1), ('fictional', 1), ('movie', 1), ('is', 1), ('easier', 1), ('than', 1), ('making', 1), ('one', 1), ('about', 1), ('real', 1), ('life', 1), ('in', 1), ('fiction', 1), ('one', 1), ('invents', 1), ('purposeful', 1), ('clear', 1), ('cut', 1), ('good', 1), ('guys', 1), ('and', 1), ('bad', 1), ('guys', 1), ('puts', 1), ('the', 1), ('fictional', 1), ('characters', 1), ('in', 1), ('conflict', 1), ('and', 1), ('takes', 1), ('the', 1), ('tale', 1), ('to', 1), ('its', 1), ('exciting', 1), ('conclus
ion', 1), ('real', 1), ('life', 1), ('however', 1), ('consists', 1), ('of', 1), ('long', 1), ('stretches', 1), ('of', 1), ('boredom', 1), ('with', 1), ('a', 1), ('few', 1), ('dramatic', 1), ('moments', 1), ('and', 1), ('characters', 1), ('who', 1), ('stand', 1), ('around', 1), ('think', 1), ('thoughts', 1), ('and', 1), ('do', 1), ('nothing', 1), ('or', 1), ('come', 1), ('and', 1), ('go', 1), ('before', 1), ('events', 1), ('are', 1), ('resolved', 1), ('spielberg', 1), ('gives', 1), ('us', 1), ('a', 1), ('visually', 1), ('spi
cy', 1), ('and', 1), ('historically', 1), ('accurate', 1), ('real', 1), ('life', 1), ('story', 1), ('djimon', 1), ('hounsou', 1), ('and', 1), ('anthony', 1), ('hopkins', 1), ('tu
rn', 1), ('in', 1), ('excellent', 1), ('performances', 1), ('pos')]
[('the', 77717), ('a', 65876), ('and', 38106), ('of', 35576), ('to', 34123), ('is', 31937), ('s', 30585), ('in', 25195), ('in', 21822), ('s', 18513), ('it', 17612), ('that', 15924), ('a', 15595)]
253
Press Enter to continue...
```