

1.OUTPUT:-

This is a sample text with , URLs () and mentions to remove.

2.

output:

Lemmatized tokens: ['The', 'boy', 'were', 'running', 'and', 'jumping', 'in', 'the', 'park']

Stemmed tokens: ['the', 'boy', 'were', 'run', 'and', 'jump', 'in', 'the', 'park']

3.

output:

Standardized text: Oh My God! that was so cool yesterday ☐☐

4.

output:

[('This', 'DT'), ('is', 'VBZ'), ('a', 'DT'), ('sample', 'NN'), ('text', 'NN'), ('for', 'IN'), ('part-of-speech', 'JJ'), ('tagging', 'NN'), (',', ',')]

5.

Topic #0:

edu com article university writes posting host nntp cs like

Topic #1:

god people don think just say does know sayings christian

Topic #2:

mac apple use software thanks problem help card pc know

Topic #3:

drive scsi ide drives controller hard floppy disk bios meg

Topic #4:

book read books reading edu page university course ftp software

Topic #5:

armenian turkish armenians turkey people government war genocide russian soviet

Topic #6:

jpeg image file format gif images files color quality version

Topic #7:

space nasa launch orbit shuttle moon mission gov earth satellite

Topic #8:

gun people government law state guns right control states rights

Topic #9:

windows dos file use files problem program version window thanks

6.

```
>
> from sklearn.feature_extraction.text import TfidfVectorizer
> import pandas as pd
>
> documents = [    "The sky is blue.",    "The sun is bright.",    "The sun in the sky is bright."]
>
> tfidf_vectorizer = TfidfVectorizer()
> tfidf_matrix = tfidf_vectorizer.fit_transform(documents)
>
> df = pd.DataFrame(tfidf_matrix.toarray(), columns=tfidf_vectorizer.get_feature_names())
> print(df)
   blue  bright      in      is      sky      sun      the
0.66284  0.000000  0.000000  0.391484  0.504107  0.000000  0.391484
0.00000  0.558478  0.000000  0.433708  0.000000  0.558478  0.433708
0.00000  0.359341  0.472491  0.279061  0.359341  0.359341  0.558121
>
```

7.

output:

```
[-0.00461729 -0.0047736 -0.00135689 -0.00225935 0.0024054 0.00194681
0.00313015 -0.00308307 -0.00218958 -0.00222326 -0.00341184 -0.00312869 -
0.00452644 -0.00394727 0.00379233 0.00394584 -0.00373821 -0.00323984 -
0.0030908 -0.00381518 0.00313715 0.00079613 0.00310879 -0.00249962 -
0.00058269 -0.00131929 -0.00215552 0.00022552 0.00269889 0.00144772 -
0.00111883 0.00089038 0.00263875 -0.00132836 -0.00278709 -0.00071284 -
0.00086249 -0.00379494 -0.00324767 0.00456697 0.00156754 0.00403906 -
0.00131606 -0.00107949 -0.00023786 -0.00252722 0.00246128 -0.00450067
0.00406998 -0.0010762 -0.00010749 0.00163175 0.00059829 -0.00188124
0.00067072 0.00352225 0.00397871 -0.00195305 -0.00164344 -0.0019277 -
0.00426968 -0.00122883 -0.00474226 0.00319609 0.0004977 -0.0003451 -
0.00364027 0.00054388 0.00392523 0.00332062 0.00393645 0.00378703 -
0.0038728 0.00279362 -0.00293605 -0.00204456 -0.00272693 -0.00180922 -
0.00094033 -0.00030417 0.00059067 0.00223104 -0.00420595 -0.00174431
0.00479956 -0.0040663 0.00264079 0.00405627 0.00017403 0.00337022 -
0.00317048 -0.00373291 -0.00242708 0.00121094 -0.00041898 -0.00056284
0.00468122 0.00031098 0.0041314 -0.00289494 0.00461968 -0.00227856 -
0.00216032 -0.00160402 0.00032992 -0.00018028]
```

0.06288836

8.

POS

9.

POS

10.

output:

Cosine similarity: 0.5

11.

output :

Positive: I love my new phone, it's amazing!

Negative: This laptop is the worst, I regret buying it.

Positive: The camera on this tablet is fantastic!

Negative: I hate this new gaming console, it's so slow.

Positive: The battery life on this e-reader is amazing, I love it.

12

output:

Accuracy: 0.75