In google colab (or) Jupyter Notebook

1. import nltk
   from nltk.tokenize import word-tokenize
   from nltk . import pos-tag
   from nltk .stem import PorterStemmer
   from nltk . corpus import StopWords
   from nltk .stem import WordNetLemmatizer

2. x = open (" /content/test.dock-txt "). read ( )

3. x
   → ' Artificial Intelligence (AI) & Machine Leary (ML) : \n \n---

   ---

4. import nltk
   nltk. download ('punkt=lab')

   → [nltk_date] (Download --
   [nltk _data]   Unzipping tokenize
   true

5. # Tokenization
   tokens = word--tokenize (a)
   print (tokens).

   → [' Artificial '; ' Intelligence ', '(', 'AI ', ')', '&'
      ' Machine', 'learning' . -------]

6. import nltk
   nltk.download ('averaged-perception_tagger-eng

   ↳ [nltk_data] Downloading packy aaye..
   [nltk_data]   root [nltk_dat
   [nltk_data]   unzipping tagger /averaged-perg_h

   True

7. # POS TAGGING
   postags = pos_tag (token)
   print (postags)

   ↳ [('Artificial', "JJ'), ('Intelligence', 'NNP'), ('C', 'L
        _ _ _]

8. import nltk
   nltk.download ('stopwords').

   ↳ [nltk]_dat] Downloady
      [nltk_date] Unzipping

   True)

9. # Removing Stop Words
   stop_words = set (stopwords.words ('english'))
   print (stopwords).

   ↳ { "we've", 'by', 'doy', 'about', 've', 's',
        _ _ _ }

```
10.  li = [ ]
     for words in token :
         if words not in stop words :
             li . append (words)
     print (li)
```
↳ ['Artificial', 'Intelligence', '(', 'AI', ',', ')', ...]

```
11.  import nltk
     nltk. download ('wordnet')
```
↳ [nltk_data] Downloady package wordnet to /root/nltk
   true

```
12.  # lemmatization
     wl = WordNet lemmatizer ()
     lemilist = []
     for words in li :
         lemilist . append ([words, wl. lemmatize (words)])
     print (lemilist)
```
↳ [['Artificial', 'Artificial'], ['Intelligence', 'Intelligence'],
   ['(', '('], ['AI', 'AI'], [')', ')'],
   ['&', '&'], ['Machine', 'Machine'], ['Leary',
   'Learning'], ['(', '('], ['ML', 'ML'],
   ......]