

```

!pip install nltk

Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk) (1.3.2)
Requirement already satisfied: regex<=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk) (2023.12.25)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk) (4.66.2)

import nltk

nltk.download('punkt')

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
True

nltk.download("stopwords")

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.
True

from nltk import sent_tokenize, word_tokenize

from nltk.corpus import stopwords

stopwords=stopwords.words("english")

sent="I will walk 500 miles and I would walk 500 more,just to be the man who walks a thousand miles to fall down at your door".lower()
sent

'i will walk 500 miles and i would walk 500 more,just to be the man who walks a thou
sand miles to fall down at your door'

words=word_tokenize(sent)
print(words)

['i', 'will', 'walk', '500', 'miles', 'and', 'i', 'would', 'walk', '500', 'more', ',', 'just', 'to', 'be', 'the', 'man', 'who', 'wal
<
>

sentences=sent_tokenize(sent)
print(sentences)

['i will walk 500 miles and i would walk 500 more,just to be the man who walks a thousand miles to fall down at your door']

print(stopwords)

['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourse
<
>

for word in words:
    if word in stopwords:
        words.remove(word)
print(words)

['will', 'walk', '500', 'miles', 'i', 'would', 'walk', '500', ',', 'the', 'man', 'walks', 'thousand', 'miles', 'to', 'fall', 'at',
<
>

from nltk import PorterStemmer

stemmer=PorterStemmer()

print(words)

['will', 'walk', '500', 'miles', 'i', 'would', 'walk', '500', ',', 'the', 'man', 'walks', 'thousand', 'miles', 'to', 'fall', 'at',
<
>

stemmed_words=[stemmer.stem(word) for word in words]
print(stemmed_words)

['will', 'walk', '500', 'mile', 'i', 'would', 'walk', '500', ',', 'the', 'man', 'walk', 'thousand', 'mile', 'to', 'fall', 'at', 'doc
<
>

```

```
➡ Counter({'NN': 6, 'VB': 3, 'MD': 2, 'CD': 2, ',': 1, 'DT': 1, 'VBP': 1, 'TO': 1, 'IN': 1})
```