

```
In [1]: import nltk
nltk.download('punkt')
nltk.download('wordnet')
nltk.download('averaged_perceptron_tagger')
nltk.download('stopwords')
from nltk import sent_tokenize
from nltk import word_tokenize
from nltk.corpus import stopwords
```

```
[nltk_data] Downloading package punkt to C:\Users\ONKAR
[nltk_data]     MANE\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to C:\Users\ONKAR
[nltk_data]     MANE\AppData\Roaming\nltk_data...
[nltk_data] Package wordnet is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     C:\Users\ONKAR MANE\AppData\Roaming\nltk_data...
[nltk_data] Package averaged_perceptron_tagger is already up-to-
[nltk_data]     date!
[nltk_data] Downloading package stopwords to C:\Users\ONKAR
[nltk_data]     MANE\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

```
In [2]: text='Albert Einstein was born at Ulm, in Württemberg, Germany, on March 14, 1879.'
```

```
In [3]: tokens_sents = nltk.sent_tokenize(text)
print(tokens_sents)
```

```
['Albert Einstein was born at Ulm, in Württemberg, Germany, on March 14, 1879.', 'Six weeks later the family moved to Munich, where he later on began his schooling at the Luitpold Gymnasium.']
```

```
In [4]: tokens_words = nltk.word_tokenize(text)
print(tokens_words)
```

```
['Albert', 'Einstein', 'was', 'born', 'at', 'Ulm', ',', 'in', 'Württemberg', ',', 'Germany', ',', 'on', 'March', '14', ',', '1879', '.', 'Six', 'weeks', 'later', 'the', 'family', 'moved', 'to', 'Munich', ',', 'where', 'he', 'later', 'on', 'began', 'his', 'schooling', 'at', 'the', 'Luitpold', 'Gymnasium', '.']
```

```
In [5]: from nltk.stem import PorterStemmer
from nltk.stem.snowball import SnowballStemmer
from nltk.stem import LancasterStemmer
```

```
In [6]: stem=[]
for i in tokens_words:
    ps = PorterStemmer()
    stem_word= ps.stem(i)
    stem.append(stem_word)
print(stem)
```

```
['albert', 'einstein', 'wa', 'born', 'at', 'ulm', ',', 'in', 'würtemberg', ',', 'germani', ',', 'on', 'march', '14', ',', '1879', '.', 'six', 'week', 'later', 'the', 'famili', 'move', 'to', 'munich', ',', 'where', 'he', 'later', 'on', 'began', 'hi', 'school', 'at', 'the', 'luitpold', 'gymnasium', '.']
```

```
In [7]: import nltk
        from nltk.stem import WordNetLemmatizer
        lemmatizer = WordNetLemmatizer()
```

```
In [8]: lemmatized_output = ' '.join([lemmatizer.lemmatize(w) for w in stem])
        print(lemmatized_output)
```

albert einstein wa born at ulm , in württemberg , germani , on march 14 , 1879 . six week later the famili move to munich , where he later on began hi school at the luitpold gymnasium .

```
In [9]: leme=[]
        for i in stem:
            lemetized_word=lemmatizer.lemmatize(i)
            leme.append(lemetized_word)
        print(leme)
```

['albert', 'einstein', 'wa', 'born', 'at', 'ulm', ',', 'in', 'würtemberg', ',', 'germani', ',', 'on', 'march', '14', ',', '1879', '.', 'six', 'week', 'later', 'the', 'famili', 'move', 'to', 'munich', ',', 'where', 'he', 'later', 'on', 'began', 'hi', 'school', 'at', 'the', 'luitpold', 'gymnasium', '.']

```
In [10]: print("Parts of Speech: ",nltk.pos_tag(leme))
```

Parts of Speech: [('albert', 'JJ'), ('einstein', 'NN'), ('wa', 'NN'), ('born', 'VB N'), ('at', 'IN'), ('ulm', 'JJ'), (',', ','), ('in', 'IN'), ('würtemberg', 'NN'), (',', ','), ('germani', 'NN'), (',', ','), ('on', 'IN'), ('march', 'NN'), ('14', 'CD'), (',', ','), ('1879', 'CD'), ('.', '.'), ('six', 'CD'), ('week', 'NN'), ('late r', 'RB'), ('the', 'DT'), ('famili', 'NN'), ('move', 'NN'), ('to', 'TO'), ('munich', 'VB'), (',', ','), ('where', 'WRB'), ('he', 'PRP'), ('later', 'RB'), ('on', 'IN'), ('began', 'VBD'), ('hi', 'JJ'), ('school', 'NN'), ('at', 'IN'), ('the', 'DT'), ('luitpold', 'JJ'), ('gymnasium', 'NN'), ('.', '.')]

```
In [11]: sw_nltk = stopwords.words('english')
        print(sw_nltk)
```

['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "yo u've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'h is', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itse lf', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'who m', 'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'durin g', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'o n', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there', 'wh en', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'ot her', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'to o', 'very', 's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn', "cou ldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't", 'h aven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'n eedn', "needn't", 'shan', "shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'were n', "weren't", 'won', "won't", 'wouldn', "wouldn't"]

```
In [12]: words = [word for word in text.split() if word.lower() not in sw_nltk]
        new_text = " ".join(words)
```

```
print(new_text)
```

Albert Einstein born Ulm, Württemberg, Germany, March 14, 1879. Six weeks later family moved Munich, later began schooling Luitpold Gymnasium.

In []: