```
import nltk
nltk.download('punkt')
nltk.download('wordnet')
nltk.download('averaged_perceptron_tagger')
nltk.download('stopwords')
from nltk import sent_tokenize
from nltk.corpus import stopwords
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /root/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Tokenization

+ Code + Text

```
text = "Hello, Friends my name Aditya Shinde. Today I am going to perform tokenization assignment"


tokens_sents = nltk.sent_tokenize(text)


print(tokens_sents)
```

```
['Hello, Friends my name Aditya Shinde.', 'Today I am going to perform tokenization assignment']
```

```
tokens_words = nltk.word_tokenize(text)
print(tokens_words)
```

```
['Hello', ',', 'Friends', 'my', 'name', 'Aditya', 'Shinde', '.', 'Today', 'I', 'am', 'going', 'to', 'perform', 'tokenization', 'assi
```

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

```
['hello', ',', 'friend', 'my', 'name', 'aditya', 'shind', '.', 'today', 'i', 'am', 'go', 'to', 'perform', 'token', 'assign']
```

Lemmatization

```
import nltk
from nltk.stem import WordNetLemmatizer
lemmatizer = WordNetLemmatizer()


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

```
hello , friend my name aditya shind . today i am go to perform token assign
```

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

```
['hello', ',', 'friend', 'my', 'name', 'aditya', 'shind', '.', 'today', 'i', 'am', 'go', 'to', 'perform', 'token', 'assign']
```

Part of Speech Tagging

```
print("Parts of Speech: ",nltk.pos_tag(leme))
```

```
Parts of Speech:  [('hello', 'NN'), (',', ','), ('friend', 'VB'), ('my', 'PRP$'), ('name', 'NN'), ('aditya', 'RB'), ('shind', 'NN'),
```

Stop Word

```
sw_nltk = stopwords.words('english')
print(sw_nltk)
```

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

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

```
Hello, Friends name Aditya Shinde. Today going perform tokenization assignment
```

Start coding or generate with AI.