

```
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer, WordNetLemmatizer
```

```
nltk.download('stopwords')
nltk.download('punkt')
nltk.download('wordnet')
```

```
↳ [nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[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...
True
```

```
text_document = 'Text Preprocessing is an essential step in Natural Language Processing. It includes tasks like stopword removal and stem
```

```
words = word_tokenize(text_document)
words
```

```
↳ ['Text',
    'Preprocessing',
    'is',
    'an',
    'essential',
    'step',
    'in',
    'Natural',
    'Language',
    'Processing',
    '.',
    'It',
    'includes',
    'tasks',
    'like',
    'stopword',
    'removal',
    'and',
    'stemming',
    '.']
```

```
stop_word = set(stopwords.words('english'))
filtered_words = [word for word in words if word.lower() not in stop_word]
filtered_words
```

```
↳ ['Text',
    'Preprocessing',
    'essential',
    'step',
    'Natural',
    'Language',
    'Processing',
    '.',
    'includes',
    'tasks',
    'like',
    'stopword',
    'removal',
    'stemming',
    '.']
```

```
stemmer = PorterStemmer()
```

```
stemmed_words = [stemmer.stem(word) for word in filtered_words]
stemmed_words
```

```
↳ ['text',
    'preprocess',
    'essenti',
    'step',
    'natur',
    'languag',
    'process',
    '.',
    'includ',
    'task',
    'like',
    'stopword',
    'remov',
    'stem',
    '.']
```

```
lemmatizer = WordNetLemmatizer()
```

```
lemmatized_words = [lemmatizer.lemmatize(word) for word in filtered_words]  
lemmatized_words
```

```
↔ ['Text',  
  'Preprocessing',  
  'essential',  
  'step',  
  'Natural',  
  'Language',  
  'Processing',  
  '.',  
  'includes',  
  'task',  
  'like',  
  'stopword',  
  'removal',  
  'stemming',  
  '.']
```

Start coding or [generate](#) with AI.