```
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords


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

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

```
text = """
The playful kitten chased the butterfly through the garden, its tiny paws barely touching the ground. The sun shone brightly, casting dancin
"""


# 1. Tokenization


tokens = word_tokenize(text)


print(tokens)
```

```
['The', 'playful', 'kitten', 'chased', 'the', 'butterfly', 'through', 'the', 'garden', ',', 'its', 'tiny', 'paws', 'barely', 'touching',
```

```
# 2. Normalization


norm_tokens = [word.lower() for word in tokens]


print(norm_tokens)
```

```
['the', 'playful', 'kitten', 'chased', 'the', 'butterfly', 'through', 'the', 'garden', ',', 'its', 'tiny', 'paws', 'barely', 'touching',
```

```
# 3. Stemming


from nltk.stem import PorterStemmer


ps = PorterStemmer()


stem_token = [ps.stem(word) for word in tokens]


print(stem_token)
```

```
['the', 'play', 'kitten', 'chase', 'the', 'butterfli', 'through', 'the', 'garden', ',', 'it', 'tini', 'paw', 'bare', 'touch', 'the', 'gr
```

```
# 4. Removal of Stopwords


stop_word = set(stopwords.words('english'))


stop_word_tokens = [word for word in stem_token if word not in stop_word]


print(stop_word_tokens)
```

```
['play', 'kitten', 'chase', 'butterfli', 'garden', ',', 'tini', 'paw', 'bare', 'touch', 'ground', '.', 'sun', 'shone', 'brightli', ',',
```

Start coding or generate with AI.