

```
1 import nltk
2 from nltk.corpus import stopwords
3 from nltk.tokenize import word_tokenize
```

```
1 # Download necessary data
2 nltk.download('punkt')
3 nltk.download('stopwords')
```

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

```
1 # Input text
2 text = "Hey devs, hope your code is bug-free and your models are learning fast!"
3 print(f"Input text: \n{text}")
4 print(f"Length of text: {len(text)}")
```

```
↗ Input text:
Hey devs, hope your code is bug-free and your models are learning fast!
Length of text: 71
```

```
1 # Tokenization
2 tokens = word_tokenize(text)
3 print(f"Tokens: \n{tokens}")
4 print(f"Length of tokens: {len(tokens)}")
```

```
↗ Tokens:
['Hey', 'devs', ',', 'hope', 'your', 'code', 'is', 'bug-free', 'and', 'your', 'models', 'are', 'learning', 'fast', '!']
Length of tokens: 15
```

```
1 # Stopword removal
2 filtered_tokens = [word for word in tokens if word.lower() not in stopwords.words('english')]
3 print(f"Filtered tokens: \n{filtered_tokens}")
4 print(f"Length of filtered tokens: {len(filtered_tokens)}")
```

```
↗ Filtered tokens:
['Hey', 'devs', ',', 'hope', 'code', 'bug-free', 'models', 'learning', 'fast', '!']
Length of filtered tokens: 10
```