Most Common Uses of `string`

1. **Predefined character sets**
    - `string.ascii_letters` → `'abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ'`
    - `string.ascii_lowercase` → `'abcdefghijklmnopqrstuvwxyz'`
    - `string.ascii_uppercase` → `'ABCDEFGHIJKLMNOPQRSTUVWXYZ'`
    - `string.digits` → `'0123456789'`
    - `string.punctuation` → `'!"#$%&\'()*+,-./:;<=>?@[\\]^_{|}~'`
    - `string.whitespace` → `' \t\n\r\x0b\x0c'` (space, tab, newline, etc.)

`import nltk`

- `nltk` = **Natural Language Toolkit**
- It's a Python library for working with **human language (text)**.
- It helps with things like tokenization, stemming, stopword removal, etc.

2️⃣ `from nltk.corpus import stopwords`

- **Corpus** means "collection of text."
- `stopwords` are **common words** that usually don't add much meaning in text analysis.
    - Examples: *"is"*, *"the"*, *"an"*, *"at"*, *"in"*
- These are often removed in NLP tasks to focus on the important words.

1️⃣ `import nltk`

- `nltk` = **Natural Language Toolkit**
- It's a Python library for working with **human language (text)**.
- It helps with things like tokenization, stemming, stopword removal, etc.

2️⃣ `from nltk.corpus import stopwords`

- **Corpus** means “collection of text.”
- `stopwords` are **common words** that usually don’t add much meaning in text analysis.
    - Examples: *"is"*, *"the"*, *"an"*, *"at"*, *"in"*
- These are often removed in NLP tasks to focus on the important words.

3️⃣ `from wordcloud import WordCloud`

- `WordCloud` = A tool to make **word cloud images**.
- A word cloud shows words from your text where:
    - **Bigger size** = appears more often
    - **Smaller size** = appears less often

Example word cloud 👇

```
      DATA
   analysis     python
NLP    AI    machine
```

4️⃣ `nltk.download('stopwords')`

- This **downloads the stopwords list** (because NLTK doesn’t have them by default).
- After downloading, you can use it like this:

```
from nltk.corpus import stopwords
print(stopwords.words('english'))
```

Output (shortened):

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'you', 'your',
'yours', ...]
```

WordCloud in ML (Machine Learning / NLP)

A **WordCloud** itself is **not a machine learning algorithm**.
It is a **visualization tool** that helps us:

1. **Understand the dataset (EDA – Exploratory Data Analysis)**
    - Before training ML models, we often explore text data.
    - A word cloud quickly shows **which words are most frequent**.
    - Example: In Twitter sentiment analysis, if negative tweets often contain *"bad"*, *"hate"*, *"worst"*, these will appear big in the word cloud

`Tokenizer` → converts text to numbers

`pad_sequences` → makes sequences the same length

`train_test_split` → splits data into training/testing

`EarlyStopping` → stops training if no progress

`ReduceLROnPlateau` → reduces learning rate when stuck

Sample text with Stop Words	Without Stop Words
GeeksforGeeks – A Computer Science Portal for Geeks	GeeksforGeeks , Computer Science, Portal ,Geeks
Can listening be exhausting?	Listening, Exhausting
I like reading, so I read	Like, Reading, read

Visualization Word Cloud

A word cloud is a text visualization tool that help's us to get insights into the most frequent words present in the corpus of the data.