

```
In [1]: import nltk
        from nltk.tokenize import word_tokenize, TweetTokenizer, MWETokenizer, TreebankWordTokenizer
        from nltk.stem import PorterStemmer, SnowballStemmer, WordNetLemmatizer
        import string

        nltk.download('punkt')
        nltk.download('wordnet')
        nltk.download('omw-1.4')
```

```
[nltk_data] Downloading package punkt to C:\Users\Lalit
[nltk_data]   H\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping tokenizers\punkt.zip.
[nltk_data] Downloading package wordnet to C:\Users\Lalit
[nltk_data]   H\AppData\Roaming\nltk_data...
[nltk_data] Downloading package omw-1.4 to C:\Users\Lalit
[nltk_data]   H\AppData\Roaming\nltk_data...
```

Out[1]: True

```
In [2]: text = "Hello there! I'm using NLTK. It's amazing . #AI #MachineLearning"
        print("Original Text:\n", text)
```

Original Text:

Hello there! I'm using NLTK. It's amazing . #AI #MachineLearning

```
In [3]: whitespace_tokens = text.split()
        print("Whitespace Tokenization:\n", whitespace_tokens)
```

Whitespace Tokenization:

['Hello', 'there!', "I'm", 'using', 'NLTK.', "It's", 'amazing', '.', '#AI', '#MachineLearning']

```
In [4]: punct_tokens = []
        for word in text.split():
            word = word.strip(string.punctuation)
            if word:
                punct_tokens.append(word)

        print("Punctuation-based Tokenization:\n", punct_tokens)
```

Punctuation-based Tokenization:

['Hello', 'there', "I'm", 'using', 'NLTK', "It's", 'amazing', 'AI', 'MachineLearning']

```
In [5]: treebank_tokenizer = TreebankWordTokenizer()
        treebank_tokens = treebank_tokenizer.tokenize(text)

        print("Treebank Tokenization:\n", treebank_tokens)
```

Treebank Tokenization:

['Hello', 'there', '!', 'I', "'m", 'using', 'NLTK.', 'It', "'s", 'amazing', '.', '#', 'AI', '#', 'MachineLearning']

```
In [6]: tweet_tokenizer = TweetTokenizer()
        tweet_tokens = tweet_tokenizer.tokenize(text)
```

```
print("Tweet Tokenization:\n", tweet_tokens)
```

Tweet Tokenization:

```
['Hello', 'there', '!', "I'm", 'using', 'NLTK', '.', "It's", 'amazing', '.', '#AI', '#MachineLearning']
```

```
In [7]: mwe_tokenizer = MWETokenizer([('Machine', 'Learning')])
mwe_tokens = mwe_tokenizer.tokenize(text.split())

print("MWE Tokenization:\n", mwe_tokens)
```

MWE Tokenization:

```
['Hello', 'there!', "I'm", 'using', 'NLTK.', "It's", 'amazing', '.', '#AI', '#MachineLearning']
```

```
In [10]: nltk.download("punkt_tab")
porter = PorterStemmer()
porter_stems = [porter.stem(word) for word in word_tokenize(text)]

print("Porter Stemming:\n", porter_stems)
```

```
[nltk_data] Downloading package punkt_tab to C:\Users\Lalit
[nltk_data]      H\AppData\Roaming\nltk_data...
[nltk_data]      Unzipping tokenizers\punkt_tab.zip.
```

Porter Stemming:

```
['hello', 'there', '!', 'i', "'m", 'use', 'nltk', '.', 'it', "'s", 'amaz', '.', '#', 'ai', '#', 'machinelearn']
```

```
In [11]: snowball = SnowballStemmer("english")
snowball_stems = [snowball.stem(word) for word in word_tokenize(text)]

print("Snowball Stemming:\n", snowball_stems)
```

Snowball Stemming:

```
['hello', 'there', '!', 'i', "'m", 'use', 'nltk', '.', 'it', "'s", 'amaz', '.', '#', 'ai', '#', 'machinelearn']
```

```
In [12]: lemmatizer = WordNetLemmatizer()
lemmatized_words = [lemmatizer.lemmatize(word) for word in word_tokenize(text)]

print("Lemmatization:\n", lemmatized_words)
```

Lemmatization:

```
['Hello', 'there', '!', 'I', "'m", 'using', 'NLTK', '.', 'It', "'s", 'amazing', '.', '#', 'AI', '#', 'MachineLearning']
```

```
In [ ]:
```