**Extractive Text Summarization**

* **Extractive text summarization** is a technique in natural language processing (NLP) where the most important sentences or phrases are selected from the original text to create a summary.
* **Aim: Our aim to convert large paragraph into smaller paragraph without changes its meaning.**

1. **Importing Library:**

* import nltk
* import numpy as np
* import re
* import heapq
* nltk.download('punkt')

**NLTK**: For text tokenization, stopword removal, and word frequency analysis.

**Numpy**: For numerical operations and arrays.

**Heapq**: To retrieve the top *n* most important sentences for summarization.

1. **Loading the data**
2. **Tokenization**

* Tokenization is the process of splitting the text into smaller units like words or sentences.

1. **Sentence Tokenization**: Split the text into sentences.
2. **Word Tokenization**: Split each sentence into individual words.
3. **Calculate Word Frequency**

* Count the frequency of each word in the tokenized text to understand which words are more important. Words that appear more frequently are likely to be important.

1. **Score the sentences**

* Now, score each sentence based on the word frequencies. A sentence score is calculated as the sum of the frequencies of the words in that sentence.

1. **Select the top sentences**

* Use the heapq library to select the n highest-scoring sentences. These will form the summary.