Text Summarization with SpaCy

- Text summarization is the process of distilling the most important information from a source (or sources) to produce an abridged version for a particular user (or users) and task (or tasks).
- idea of summarization is to find a subset of data which contains the “information” of the entire set
- Main Idea
  - Text Preprocessing(remove stopwords,punctuations).
  - Frequency table of words/Word Frequency Distribution - how many times each word appears in the document
  - Score each sentence depending on the words it contains and the frequency table
  - Build summary by joining every sentence above a certain score limit

Word Frequency Table

- dictionary of words and their counts
- How many times each word appears in the document
- Using non-stopwords

Maximum Word Frequency

- find the weighted frequency
- Each word over most occurring word
- Long sentence over short sentence

Sentence Score and Ranking of Words in Each Sentence

- Sentence Tokens
- scoring every sentence based on number of words
- non stopwords in our word frequency table

To GET THE OUTPUT:

step 1: open the jupyter notebook or lab

step 2: make sure all the required packages are pre installed

step 3: import all the required packages and library

step 4: assign the values and text to the parameters

step 5: run the `text_summarizer` with the requied parameters

step 6: done!! The output is obtained.

OUTPUT(obtained and verified)

In [4]:
```
1 text_summarizer(text,1)
```

```
Original Document

"Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to progressively
improve their performance on a specific task. Machine learning algorithms build a mathematical model of sample data, known as
"training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine le
arning algorithms are used in the applications of email filtering, detection of network intruders, and computer vision, where
it is infeasible to develop an algorithm of specific instructions for performing the task. Machine learning is closely related
to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization deliv
ers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine l
earning, and focuses on exploratory data analysis through unsupervised learning.In its application across business problems, m
achine learning is also referred to as predictive analytics."

Total Length: 1071


Using the defined function:
Summarized Document

Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions
or decisions without being explicitly programmed to perform the task.

Total Length: 195

 Using the Inbuilt lib:
Summarized Document (verification):

Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions
or decisions without being explicitly programmed to perform the task.

Total Length: 195
```