Project: Text Summarization

Introduction:

Today we know that machines have become smarter than us and can help us with every aspect of life, the technologies have reached to an extent where they can do all. The field which makes these things happen is Machine Learning. Machine Learning train the machines with some data which makes it capable of acting when tested by the similar type of data. The machines have become capable of understanding human languages using Natural Language Processing.

As the project title suggests, Text Summarizer is a web-based application which helps in summarizing the text. We can upload our data and this application gives us the summary of that data in as many numbers of lines as we want. The product is mainly a text summarizing using Deep Learning concepts.

Language Used: Python.

Libraries Used: Nltk,heap,pandas,re.

Project Components:

1. First of all , we have to load our text(.txt) file for data processing.

2.Now we have to do some data cleaning using re(regex) library for removing some unwanted data(like: [],””...) .

3. We have to tokenize our file as sentence Tokenize.

4. Also we have to tokenize our file as word Tokenize .

5. Set stopwords as english using Nltk library.

6. Removing stopwords from our word\_Tokenize File .

7. Then we have to calculate weighted frequecies of each word using python Loops.

8.Now In Sentence\_Tokenize we have to calculate the weight Frequency of each sentence using word weighted Frequency.

9.Join the words using .split function and summarize the .txt file into desired no of lines.

Submitted By:

Gagan Varshney