

# **TEXT SUMMARIZER USING LSTM WITH ATTENTION LAYER**

**PROJECT SYNOPSIS**

## **Machine Intelligence**

**BACHELOR OF TECHNOLOGY- V Sem CSE**  
**Department of Computer Science & Engineering**

**SUBMITTED BY**

**Batch No: - 3**

Student name 1: Rakshika S  
Srn: PES2UG20CS264

Student name 2: Shravya U  
Srn: PES2UG20CS326

# **PES UNIVERSITY**

**(Established under Karnataka Act No. 16 of 2013)**

**100 Feet Ring Road, BSK III Stage, Bengaluru-560085**

## **Abstract and Scope:**

Text summarization is a technique that extracts text from a large amount of data. It helps in creating a shorter version of the large text available. It is important because it reduces reading time, helps in better research work, increases the amount of information that can fit in an area.

The steps to generate the summarization of the paragraph are:

- Obtain Data
- Text Pre-processing
- Convert paragraphs to sentences
- Tokenizing the sentences
- Find weighted frequency of occurrence
- Replace words by weighted frequency in sentences
- Sort sentences in descending order of weights
- Summarizing the Article

The sections of this project contain:

- Data acquisition
- Preprocessing of Data
- Analyzing and Preparing the Data
- Model Building
- Model Training
- Inference

The text summarizer will be done using the natural language processing method.

LSTM is a special kind of RNN. Which is capable of learning long term dependencies thus treating the problem of short-term dependencies of a simple RNN. It is not possible for a RNN to remember and to understand the context behind the input while it is achieved using a LSTM.

It is difficult for the encoder to memorize long sequences into a fixed-length vector - this is the major drawback of the encoder - decoder model. Hence, an additional layer called the Attention layer is used.

### Feasibility study:

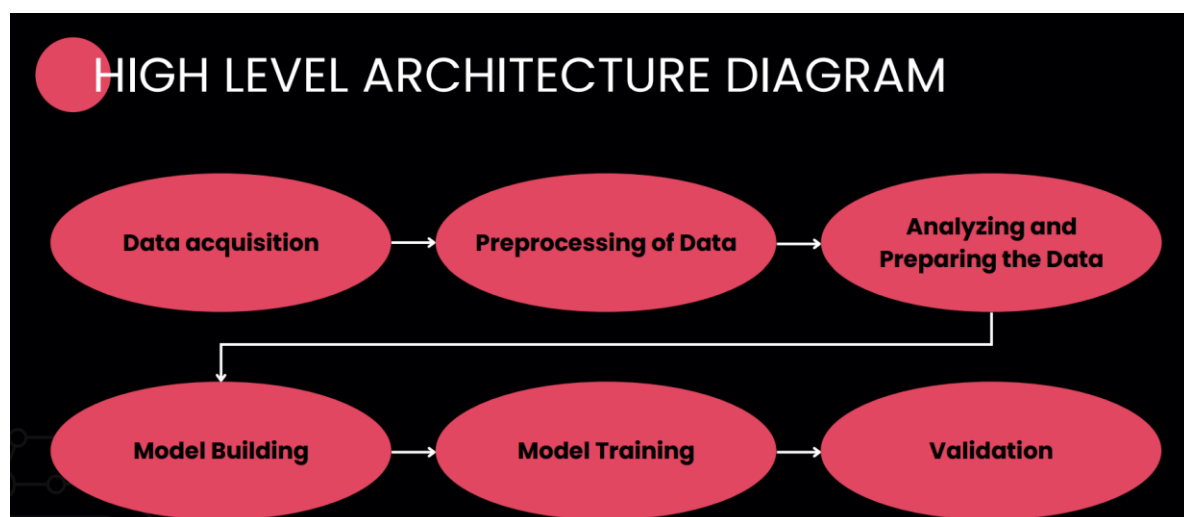
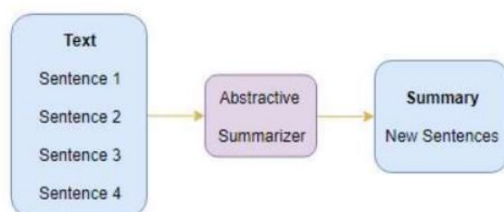
Text summarizer is a very important tool. It is the process of condensing large amounts of information into concise, consumable texts. It helps readers understand complex topics quickly by reducing the topics down to their main points. Companies often ask writers to summarize text because they want to ensure that the reader understands what they've written. They also want to ensure that the content is easy to read. The main advantages of it are that it helps keep up-to-date and increases user engagement.

The purpose of a title is to: Grab Attention – Compel the reader to read the lead-in. Predict Content – Set an expectation of what awaits them. Establish Tone – Create an atmosphere appropriate to the subject matter. So, it is important to have a title for a paragraph.

### Design Approach of the Project:

Text summarizer:

This is an abstractive text summarization tool. The text summarization will be done using the NLP method. We will be separating the stop words from the paragraph that is given. Stop words are the unnecessary words in a paragraph. We will also be removing the punctuations and giving a summary only with text. The whole paragraph will be tokenized. The frequency of the words will be tracked with a counter. Then the word frequencies will be normalized. A threshold will be set which will specify what percentage of the paragraph given will consist the summary. If 30% of a paragraph containing 10 lines should be the summary, then the tool will output a summary of 3 lines about the paragraph that is fed.



## **References:**

[1] Rahul, S. Adhikari and Monika, "NLP based Machine Learning Approaches for Text Summarization," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 535-538, doi: 10.1109/ICCMC48092.2020.ICCMC 00099.

[2] Abstractive Document Summarization without Parallel Data

[3] A survey automatic text summarization

[4] Diverse Decoding for Abstractive Document Summarization