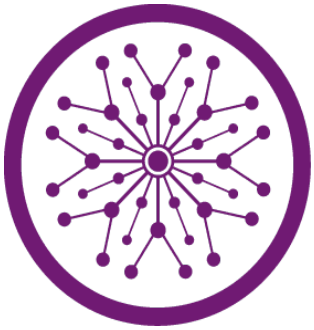

SUPERIOR UNIVERSITY

Name : Minhaj Asghar

Roll No : 108

Lab Report : 09

Project Report: Text Summarization Web App using T5 Transformer

1. Project Title

Text Summarization with T5 and Gradio

2. Objective

The main aim of this project is to build an interactive web application that can summarize long pieces of text into concise summaries using the T5 Transformer model. This helps users quickly extract key information from larger documents or paragraphs.

3. Tools and Technologies Used

- Python
- Transformers Library (Hugging Face)
- T5-small Pre-trained Model
- Gradio (for creating web interface)
- PyTorch (for model execution)
- Google Colab or Local Python Environment

---

4. System Overview

a. Model Overview

- The T5 (Text-to-Text Transfer Transformer) model by Google is used.
- It is a unified framework that treats every NLP problem as a text generation task.
- The pre-trained T5-small variant is used to balance performance and speed.

b. Functionality

- The application takes user-inputted text as input.
- It applies text summarization by prepending "summarize: " to the input.
- The T5 model generates a summarized output using beam search for quality results.
- The summarized result is displayed in a user-friendly interface powered by Gradio.

5. Key Features

- ✅ Pre-trained Model: Utilizes Hugging Face's T5-small model for accurate summarization.
- ✅ Web Interface: Powered by Gradio, making the tool easily usable without needing any coding knowledge.
- ✅ Interactive & Real-time: Users can input any paragraph or article and get an instant summary.
- ✅ Length Control: Uses beam search and length penalty to ensure summaries are concise but meaningful.
- ✅ Truncation Handling: Automatically manages long input texts by truncating appropriately.

6. Use Cases

- Quickly summarizing articles, news, or research papers
- Assisting students and professionals in reviewing large documents
- Helping content creators draft condensed versions of their content
- Improving reading efficiency and information retrieval

7. Deployment and Usage

- The interface is deployed using Gradio.
- It can be run locally, on a server, or in environments like Google Colab.

- After launching, users are presented with a textbox to input text, and the summarized output is displayed below.

8. Challenges Faced

- Managing input length due to model constraints (max 1024 tokens).
- Ensuring quality summaries while keeping output within desired length.
- Optional use of external datasets or evaluation metrics (removed in this case).

9. Conclusion

This project successfully demonstrates how transformer models like T5 can be integrated into web applications for practical NLP tasks like text summarization. The combination of Hugging Face and Gradio provides a powerful yet simple way to deploy machine learning solutions.