CSE413- SNLP- Minor project Report

Title of the Project : Automated Text Summarization Using NLP

Group ID : 20 (Sai Charan-BCD29)

Group Members: 1

NAME: SAI CHARAN

ROLL NUMBER: 2022BCD0029

Link to code/Git if any : https://github.com/SaiCharanBCD29/Automated-Text-Summarizer-NLP.git

1. Problem Description:

Nowadays, people deal with a large amount of text every day — from news articles and research papers to blogs and social media posts. Reading through all this content takes a lot of time and effort. The main problem is how to quickly understand the key points from long pieces of text without missing important information.

This project focuses on solving that issue by building an automated text summarization system using Natural Language Processing (NLP). The goal is to make it easier to get short and meaningful summaries of long text documents, helping users save time and grasp the main idea instantly.

2. Objective of the Project

The main objective of this project is to create a system that can automatically summarize long text into shorter, easy-to-read summaries. It aims to help users quickly understand the main points of any text without reading the entire content.

Specifically, the project focuses on:

- Reducing the time and effort needed to read long documents.

- Highlighting the most important information from a text.
- Demonstrating the use of NLP and pre-trained language models for practical applications.

3. Dataset details

For this project, we used publicly available text data that is suitable for summarization tasks. The dataset contains long paragraphs and articles in English. These texts are used as input for the model, which then generates short, meaningful summaries.

In our implementation, we also tested the system with sample text files created manually to demonstrate how the summarizer works on different types of content, like news articles, blogs, and general information paragraphs.

No special or private datasets were required, making it easy to run and reproduce the project on any computer.

4. Methodology and Models

In this project, we built an automated text summarization system using Natural Language Processing (NLP) and a pre-trained transformer model.

The system works in the following steps:

- Text Input: The user provides a long paragraph or article that needs to be summarized.
- Preprocessing: The input text is cleaned and prepared for the model by removing unnecessary spaces or special characters.
- Model Selection: We used a pre-trained Hugging Face transformer model called BART Large CNN, which is designed for text summarization. This model understands the meaning of the text and can generate concise summaries.
- Summarization: The model processes the input text and produces a shorter version that retains the main ideas.
- Output: The generated summary is displayed, showing the most important points of the original text.

This approach allows us to summarize any text efficiently without training a model from scratch, saving both time and computational resources.

5. Implementation / Steps to Run the Project

The project is implemented using Python and the Hugging Face Transformers library. Here are the steps to run the project:

- Set up the environment:
  - Install Python (version 3.7 or above).
  - Install required libraries using the command:

```
pip install -r requirements.txt
```

- Prepare the input text:
  - You can use the provided sample_input.txt file or create your own text file with a long paragraph for summarization.
- Run the summarization script:
  - Execute the Python script using:

```
python summarization.py
```

- View the output:
  - The system will display the original text and the generated summary.
  - You can also save the summary to a file if needed.
- Optional:
  - You can test the model on different text files or adjust the minimum and maximum summary length in the code for better results.

This step-by-step approach makes it easy to use the system for summarizing any text without needing deep knowledge of NLP.

Results and Discussion

The automated text summarization system was tested on several paragraphs, including news articles, blog posts, and sample text files. The system successfully generated short, meaningful summaries that captured the main points of the original text.

Observations:

- The model effectively reduced the length of long paragraphs while preserving important information.
- The summaries were easy to read and understood by humans.

- For very short texts, the summarizer sometimes returned text similar to the input, which is expected since there is little to compress.
- The system works best with English text and performs well on formal content like news or articles.

Example Output:

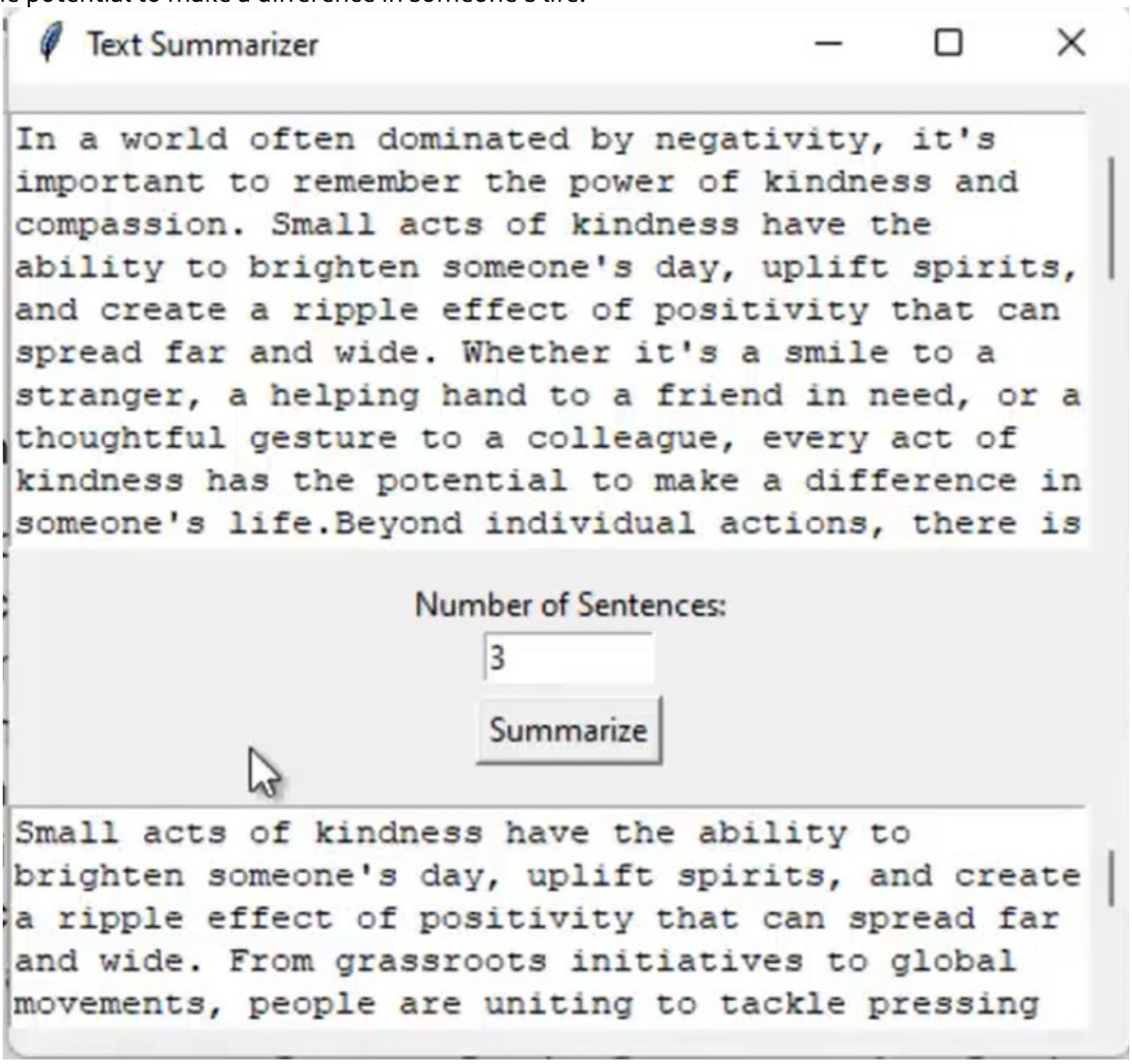
Original Text:

In a world often dominated by negativity, it's important to remember the power of kindness and compassion.", "Small acts of kindness have the ability to brighten someone's day, uplift spirits, and create a ripple effect of positivity that can spread far and wide.", "Whether it's a smile to a stranger, a helping hand to a friend in need, or a thoughtful gesture to a colleague, every act of kindness has the potential to make a difference in someone's life.", 'Beyond individual actions, there is also immense power in collective efforts to create positive change.', 'When communities come together to support one another, incredible things can happen.', 'From grassroots initiatives to global movements, people are uniting to tackle pressing social and environmental issues, driving meaningful progress and inspiring hope for a better future.', "It's also important to recognize the strength that lies within each and every one of us.", 'We all have the ability to make a positive impact, no matter how small our actions may seem.', 'By tapping into our innate compassion and empathy, we can cultivate a culture of kindness and empathy that enriches our lives and those around us.', "So let's embrace the power of kindness, and strive to make the world a better place one small act at a time.", 'Together, we can create a brighter, more compassionate future for all.

Generated Summary:

Small acts of kindness have the ability to brighten someone's day, uplift spirits, and create a ripple effect of positivity that can spread far and wide. From grassroots initiatives to global movements, people are uniting to tackle pressing social and environmental issues, driving meaningful progress and inspiring hope for a better future. Whether it's a smile to a stranger, a helping hand to a friend in need, or a thoughtful gesture to a colleague, every act of kindness has

the potential to make a difference in someone's life.



Discussion:

The project demonstrates how pre-trained models can save time and effort in understanding large texts.

It can be extended to other languages or integrated into applications like chatbots, news summarizers, or content review systems.

Conclusion:

This project successfully demonstrates an Automated Text Summarization system using NLP and a pre-trained transformer model. The system can take long paragraphs or articles and generate concise summaries that retain the main ideas. It reduces the time and effort required to read and understand large amounts of text.

Using the Hugging Face Transformers library and the BART Large CNN model, the project shows how modern NLP models can be applied for real-world text summarization tasks without needing to train models from scratch.

The project can be further improved by adding support for other languages, handling very large documents, and integrating the summarizer into web or mobile applications.

Reference:

- Hugging Face Transformers Documentation: https://huggingface.co/transformers/
- Facebook BART Model Paper: Lewis, Mike, et al. "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension." arXiv preprint arXiv:1910.13461 (2019).
- Python Official Documentation: https://www.python.org/doc/