



[TEXT SUMMARIZATION]

Mentor: Mr. Narendra Kumar

Intern: Patel Hetu

3rd -B.Tech, ICT,

Pandit Deendayal Energy University (PDEU)

Gandhinagar, Gujarat

Author Note:

E-mail id: hetupatel2405@gmail.com , Hetu.pict22@sot.pdpu.ac.in

GitHub repo: <https://github.com/Patelhlt/text-summarizer>

LinkedIn: <https://www.linkedin.com/in/hetu-patel-04044127b/>

Contact info: +919727731270, +919428220888

ACKNOWLEDGEMENT

I **Patel Hetu**, would like to express my heartfelt gratitude to everyone who has contributed to the successful completion of this project.

First and foremost, I would like to thank my mentor, **Mr. Narendra Kumar**, for his invaluable guidance, support, and encouragement throughout this project. His expertise and insights were instrumental in shaping the direction of my work and ensuring its success.

I am also deeply grateful to **Infosys** and the **Springboard Internship Program** for providing me with this opportunity to explore and work on such an exciting and impactful project. The resources and support provided by the organization have been crucial in the development and execution of my project.

Additionally, I would like to extend my appreciation to my peers and colleagues in Group-1, whose collaboration and feedback have greatly enriched my work. Their diverse perspectives and constructive criticism have been invaluable in refining my ideas and improving the quality of my project.

Lastly, I would like to thank my family and friends for their unwavering support and encouragement throughout my internship. Their belief in my abilities has been a constant source of motivation and inspiration.

This project would not have been possible without the collective efforts and support of all these individuals, and I am deeply thankful for their contributions.

Table of Contents

1. [Acknowledgement](#)
2. [Introduction](#)
3. [Intended Plan](#)
4. [Techniques and Algorithms](#)
5. [Types of Text Summarization](#)
6. [Dataset Proposal](#)
7. [Extractive Summarization](#)
8. [Architecture of the Extractive Text Summarization System](#)
9. [ROUGE Score](#)
10. [Abstractive Summarization](#)
11. [Architecture of the Abstractive Text Summarization System](#)
12. [Comparing Models](#)
13. [Model Training](#)
14. [Evaluate the Model \(Model Validation\)](#)
15. [ROUGE Score Example](#)
16. [Comparative Analysis](#)
17. [Interface Components](#)
18. [Testing Interface Using Gradio](#)
19. [Deployment Part](#)

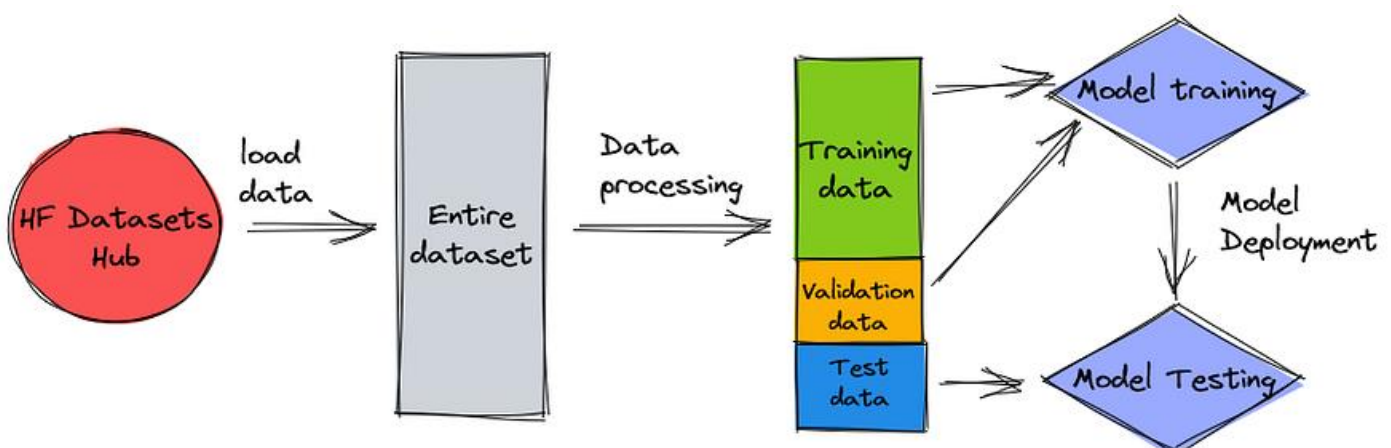
Problem Statement

- Developing an automated text summarization system that can accurately and efficiently condense large bodies of text into concise summaries is essential for enhancing business operations.
- This project aims to deploy NLP techniques to create a robust text summarization tool capable of handling various types of documents across different domains.
- The system should deliver high-quality summaries that retain the core information and contextual meaning of the original text.

Project Statement

- Text Summarization focuses on converting large bodies of text into a few sentences summing up the gist of the larger text.
- There is a wide variety of applications for text summarization including News Summary, Customer Reviews, Research Papers, etc.
- This project aims to understand the importance of text summarization and apply different techniques to fulfill the purpose.

Intended plan:



Introduction

Definition of Text Summarization: Text summarization is the process of creating a concise and coherent version of a longer text document by extracting the most important information. It aims to reduce the length of the original document while preserving its key content and overall meaning.

Importance: Summarization is crucial in today's information-rich world where quick comprehension of large volumes of text is necessary. It helps in:

- Enhancing productivity by providing quick overviews of lengthy documents.
- Improving accessibility by making information easier to understand.
- Supporting decision-making by highlighting critical information.

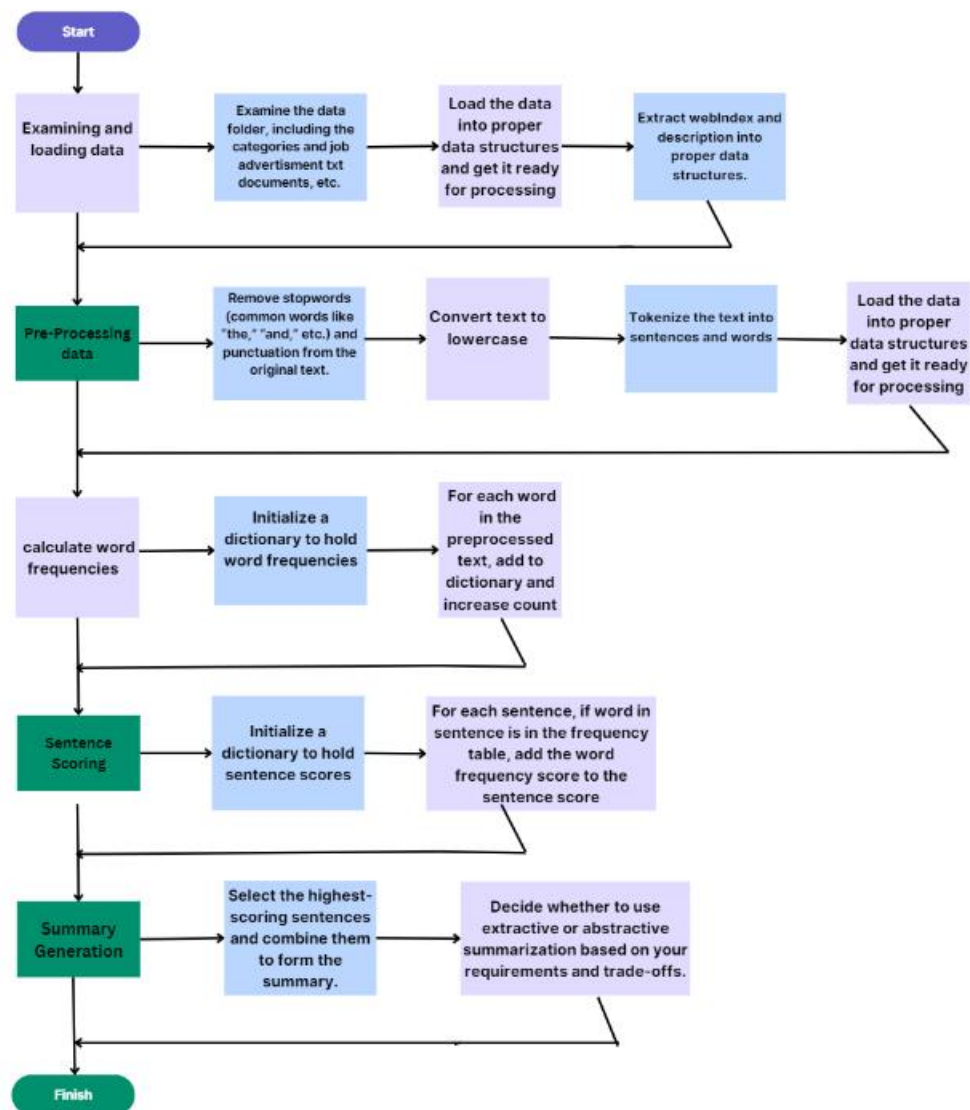
Applications:

- **News Articles:** Summarizing daily news to provide quick updates.
- **Research Papers:** Creating abstracts and summaries for academic papers.
- **Legal Documents:** Simplifying complex legal texts for better understanding.
- **Customer Feedback:** Summarizing reviews and feedback to identify common themes.
- **Social Media:** Summarizing trends and topics from social media posts.

Used Datasets:

- **CNN/Daily Mail:** Used for extractive summarization, providing a large collection of news articles and corresponding summaries.
- **Samsum:** Used for abstractive summarization, containing dialogues and their summaries.

WORK FLOW:



- **Extractive Summarization:**

- Extractive summarization selects key sentences, phrases, or sections directly from the source text.
- It often uses methods like frequency-based selection and TF-IDF.
- Pros: Simple and quick.
- Cons: May lack coherence and context.

ROUGE SCORE:

- 'rouge1': AggregateScore(low=Score(precision=0.75, recall=0.6, fmeasure=0.6666666666666665), mid=Score(precision=0.75, recall=0.6, fmeasure=0.6666666666666665), high=Score(precision=0.75, recall=0.6, fmeasure=0.6666666666666665)), 'rouge2': AggregateScore(low=Score(precision=0.3333333333333333, recall=0.25, fmeasure=0.28571428571428575), mid=Score(precision=0.3333333333333333, recall=0.25, fmeasure=0.28571428571428575), high=Score(precision=0.3333333333333333, recall=0.25, fmeasure=0.28571428571428575)), 'rougeL': AggregateScore(low=Score(precision=0.75, recall=0.6, fmeasure=0.6666666666666665), mid=Score(precision=0.75, recall=0.6, fmeasure=0.6666666666666665), high=Score(precision=0.75, recall=0.6, fmeasure=0.6666666666666665)), 'rougeLsum': AggregateScore(low=Score(precision=0.75, recall=0.6, fmeasure=0.6666666666666665), mid=Score(precision=0.75, recall=0.6, fmeasure=0.6666666666666665), high=Score(precision=0.75, recall=0.6, fmeasure=0.6666666666666665))

- **Abstractive Summarization:**

- Abstractive summarization generates new sentences that convey the main ideas of the source text.
- It involves complex techniques like neural networks and transformer models.
- Pros: Can produce more coherent and human-like summaries.
- Cons: More computationally intensive and may introduce inaccuracies.

model training :

- TrainOutput(global_step=3683, training_loss=0.5255669773256134, metrics={'train_runtime': 814.185, 'train_samples_per_second': 18.094, 'train_steps_per_second': 4.524, 'total_flos': 1993855419285504.0, 'train_loss': 0.5255669773256134, 'epoch': 1.0})

ROUGE SCORE:

- rouge1: Score(precision=0.767627392778128, recall=0.22941014983341268, fmeasure=0.33640983854210316)
- rouge2: Score(precision=0.41012512677602814, recall=0.11619754543090752, fmeasure=0.17460415842663637)
- rougeL: Score(precision=0.5740780603041631, recall=0.19681507337121906, fmeasure=0.28288022661661905)
- rougeLsum: Score(precision=0.7189119882750398, recall=0.21573804703761962, fmeasure=0.3182781398546788)

Interface

Components of a Text Summarization Interface

1. Input Field:
 1. Text Box: A large text box where users can paste or type the text they want to summarize.
 2. File Upload: Option to upload text files (e.g., .txt, .docx, .pdf) for summarization.
2. Summarization Options:
 1. Type of Summarization: Choose between extractive (selects important sentences from the original text) and abstractive (generates new sentences that convey the main ideas) summarization.
 2. Summary Length: Slider or input box to specify the desired length or percentage of the summary.
3. Output Display:
 1. Summary Box: A text box or display area where the summarized text is shown.
 2. Download Option: Button to download the summary as a text file.
4. Additional Features:
 1. Language Selection: Option to select the language for summarization.
 2. Adjustable Parameters: Advanced settings for adjusting parameters like temperature, beam size, or max tokens for abstractive models.
5. User Feedback:
 1. Edit and Improve: Option for users to manually edit the generated summary.
 2. Feedback Form: Collect user feedback to improve the summarization model.

Design Considerations

1. User Experience (UX):
 1. Simplicity: Keep the interface clean and intuitive.
 2. Responsiveness: Ensure the interface works well on various devices, including desktops, tablets, and smartphones.
2. Performance:

1. Speed: The summarization process should be fast to enhance user satisfaction.
2. Scalability: The system should handle multiple requests efficiently.
3. Accuracy and Quality:
 1. Model Selection: Use state-of-the-art models for better summarization quality.
 2. Continuous Improvement: Regularly update the models based on user feedback and advancements in NLP.
4. Security and Privacy:
 1. Data Privacy: Ensure that user data is not stored or misused.
 2. Secure Uploads: Implement secure file handling practices.

deployment part

check on my website : [HETU PATEL: http://patelhetu.andtehdugrfufscr.centralindia.azurecontainer.io:7860/](http://patelhetu.andtehdugrfufscr.centralindia.azurecontainer.io:7860/)

CONCLUSION:

This project successfully explored and implemented both extractive and abstractive summarization models. Despite initial challenges with the `bert-base-uncased` model, transitioning to the `t5-small` model enabled us to build effective summarization systems. The development of a user-friendly interface further enhances the usability of these models, making text summarization accessible to a broader audience. The deployment phase will bring this tool into practical use, providing a valuable resource for users needing efficient text summarization.

By completing this project, we have gained valuable insights into the complexities of text summarization and the practical aspects of model training and deployment. This experience will undoubtedly contribute to future endeavors in the field of natural language processing and machine learning.