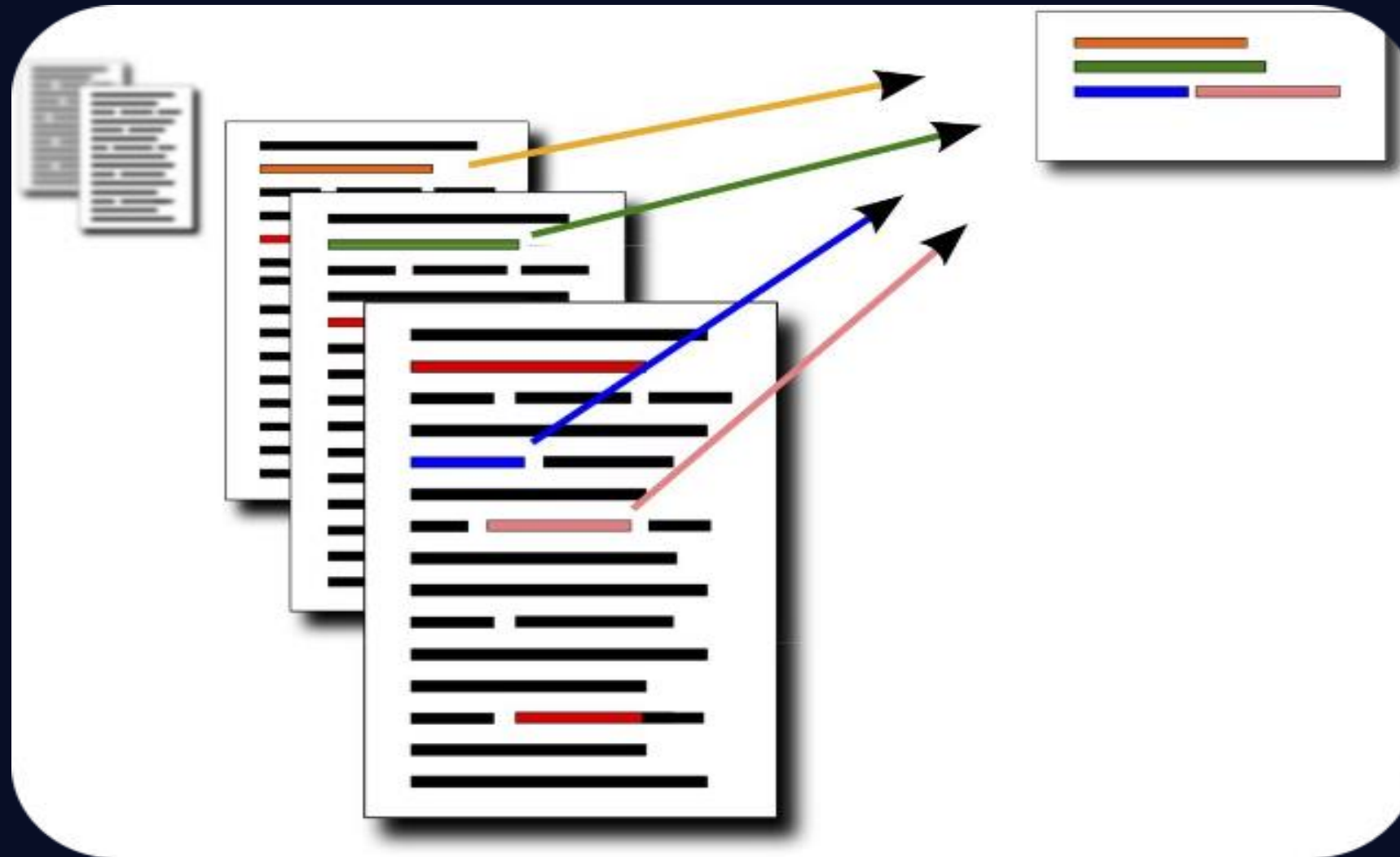


Text Summarizer



**Team-1 : Text Summarizer under the guidance of
NARENDRA KUMAR SIR**

key points:

1. Team
2. Introduction
3. Dataset
4. Model
5. Approach
6. Preprocessing
7. Extractive Model
8. Rough Scores-Extractive
9. Abstractive Model
10. Rough Scores-Abstractive
11. Interface-Demo
12. Github Repo

Text Summarization Project Ideas

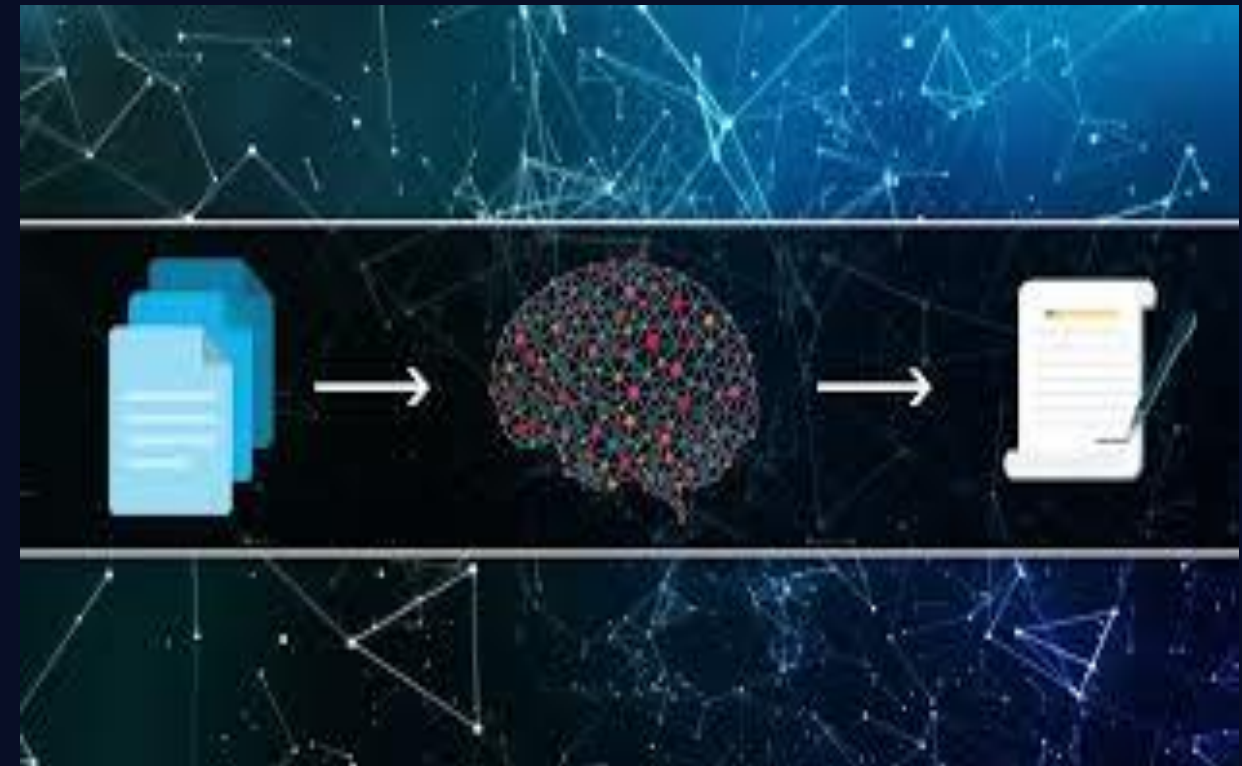


Team-1

- DHARMAVARAPU SANTHI KUMARI
- SAI TEJA ANKAM
- HETU PATEL
- AVANIGADDA SIVAYYA
- VIVEK SADASIVAN

INTRODUCTION

Text summarization is the process of condensing a lengthy text into a concise and informative summary, capturing the key points and main ideas. It is a crucial task in natural language processing with applications in various domains, from news articles to research papers. Text summarization is the creation of a short, accurate, and fluent summary of a longer text document. Automatic text summarization methods are greatly needed to address the ever-growing amount of text data available online. This could help to discover relevant information and to consume relevant information faster.



Dataset

Research:

Considering many datasets and working on them for better understanding we consider two datasets for model.

Xsum.csv

For Abstractive model we choose Xsum.csv dataset from Hugging face

train.csv

For Extractive model we choose train.csv dataset from Cnn/Daily mails

Overview of T5 Model

1. Transformers

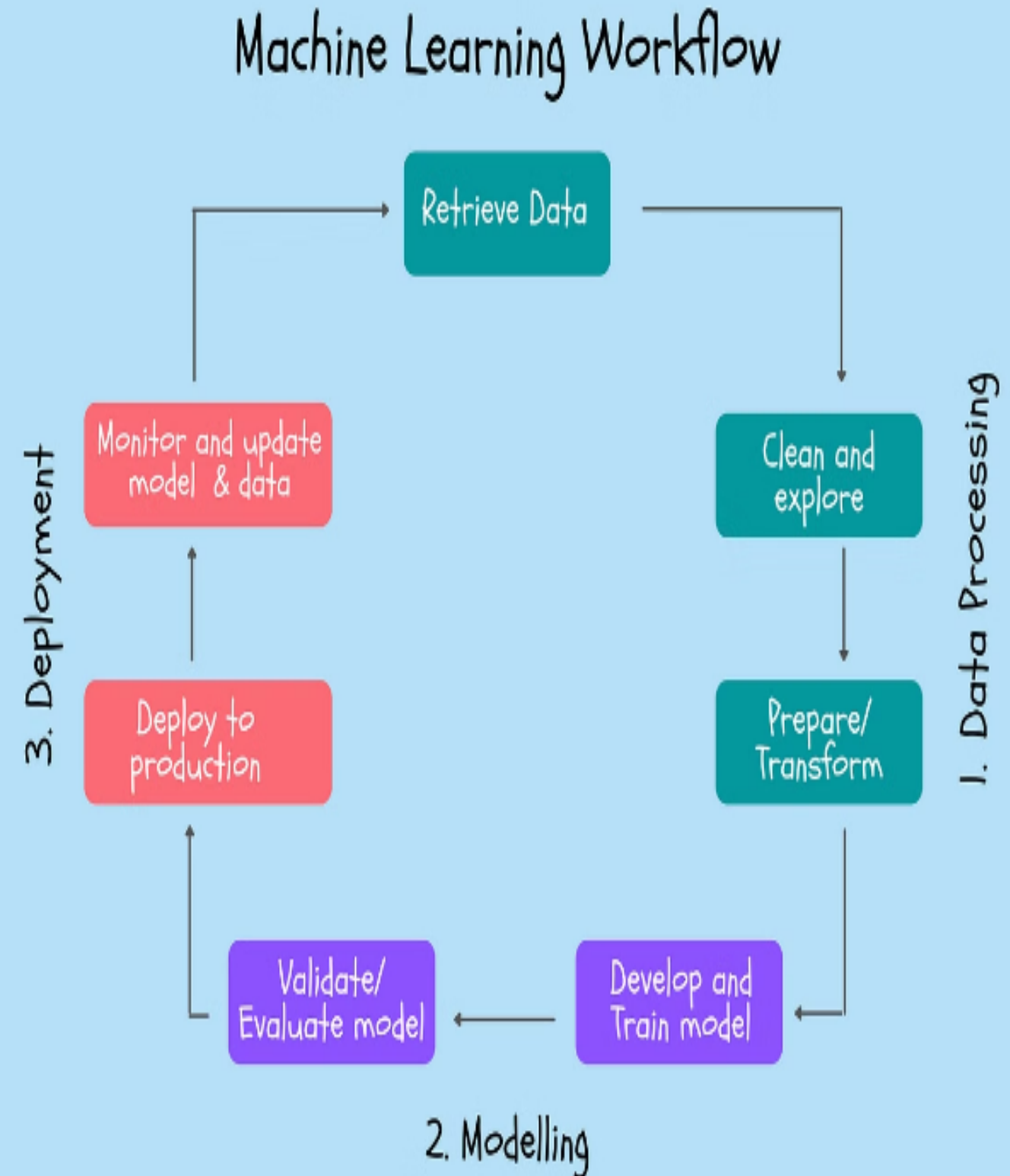
The T5 model is built upon the Transformer architecture, which has revolutionized language understanding and generating tasks.

2. Text-to-Text

T5 is a unified text-to-text model, capable of handling a wide range of natural language processing tasks, including summarization.

3. Pre-training

T5 is pre-trained on a massive corpus of text data, providing it with a strong foundation for fine-tuning on specific tasks.



Approach

1 Extract

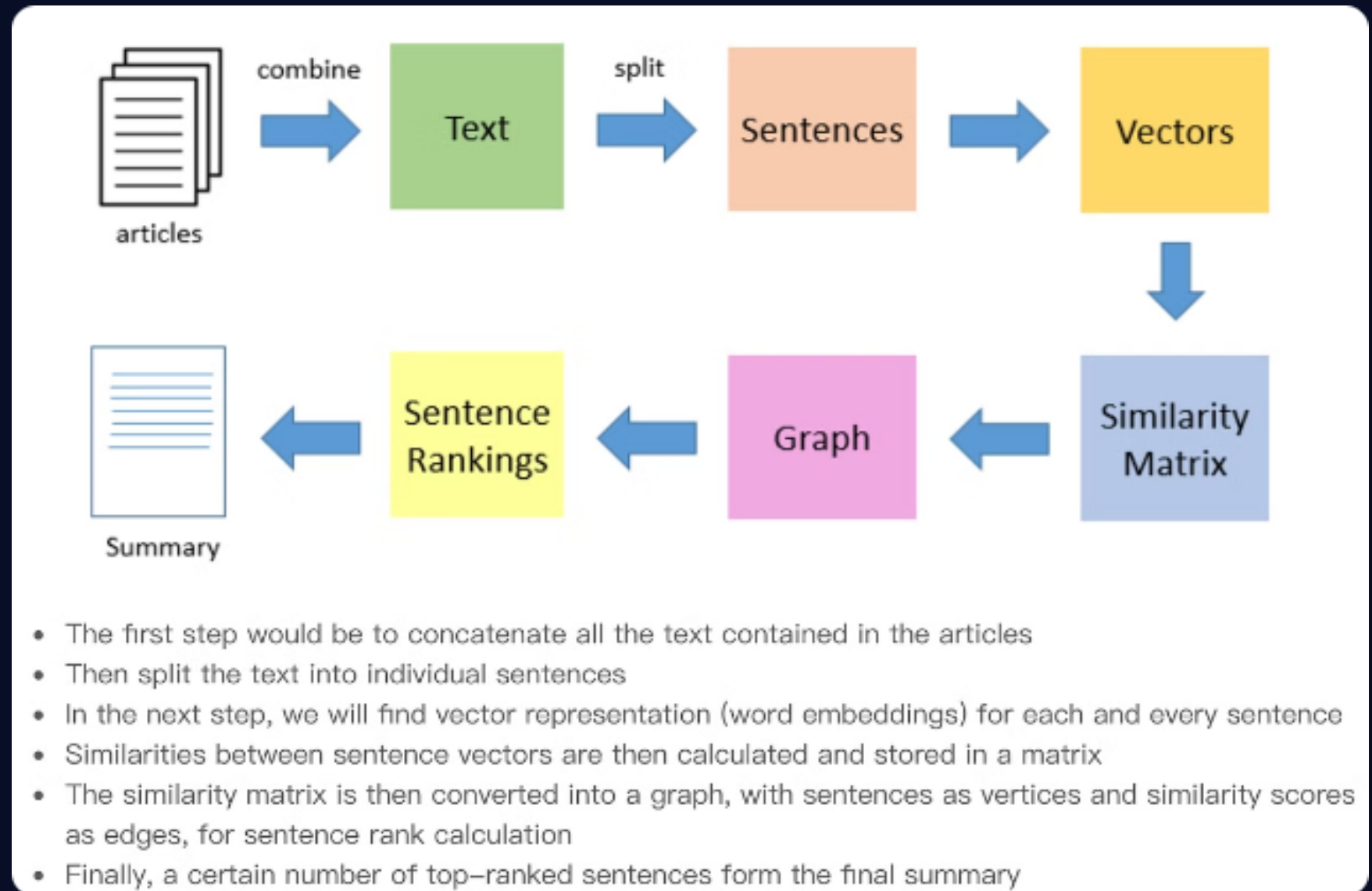
Our model identifies the most essential information in the source text.

2 Compress

The extracted content is intelligently compressed to create a concise summary.

3 Refine

The summary is polished to ensure coherence, clarity, and readability.



Preprocessing



Cleaning

Removing noise, formatting, and irrelevant content from the source text.



Tokenization

Breaking down the text into individual words, sentences, and paragraphs.



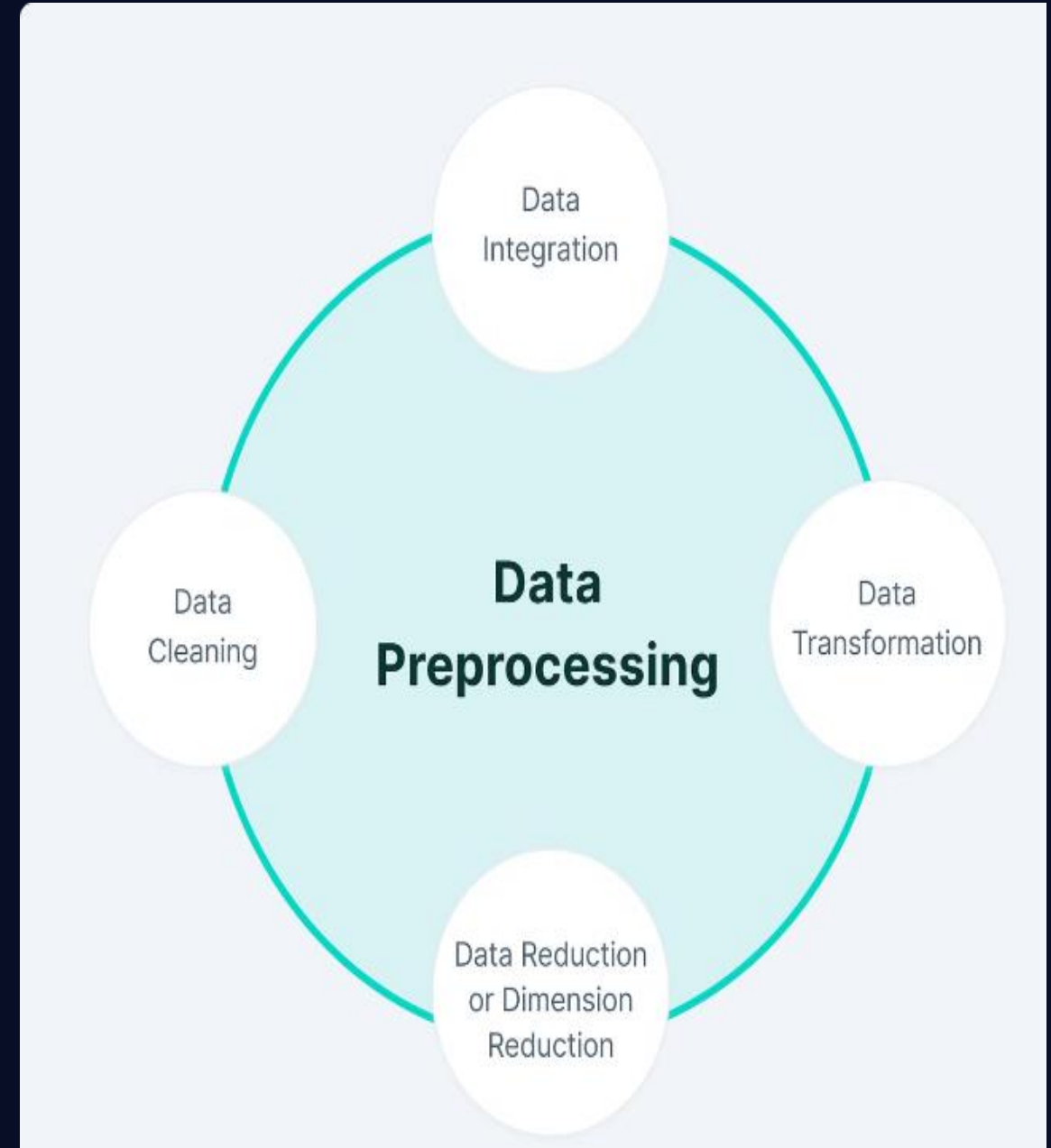
Analysis

Applying natural language processing techniques to understand the text structure and meaning.



Encoding

Converting the preprocessed text into a format suitable for the summarization model.



Extractive Summarization

Key Idea:

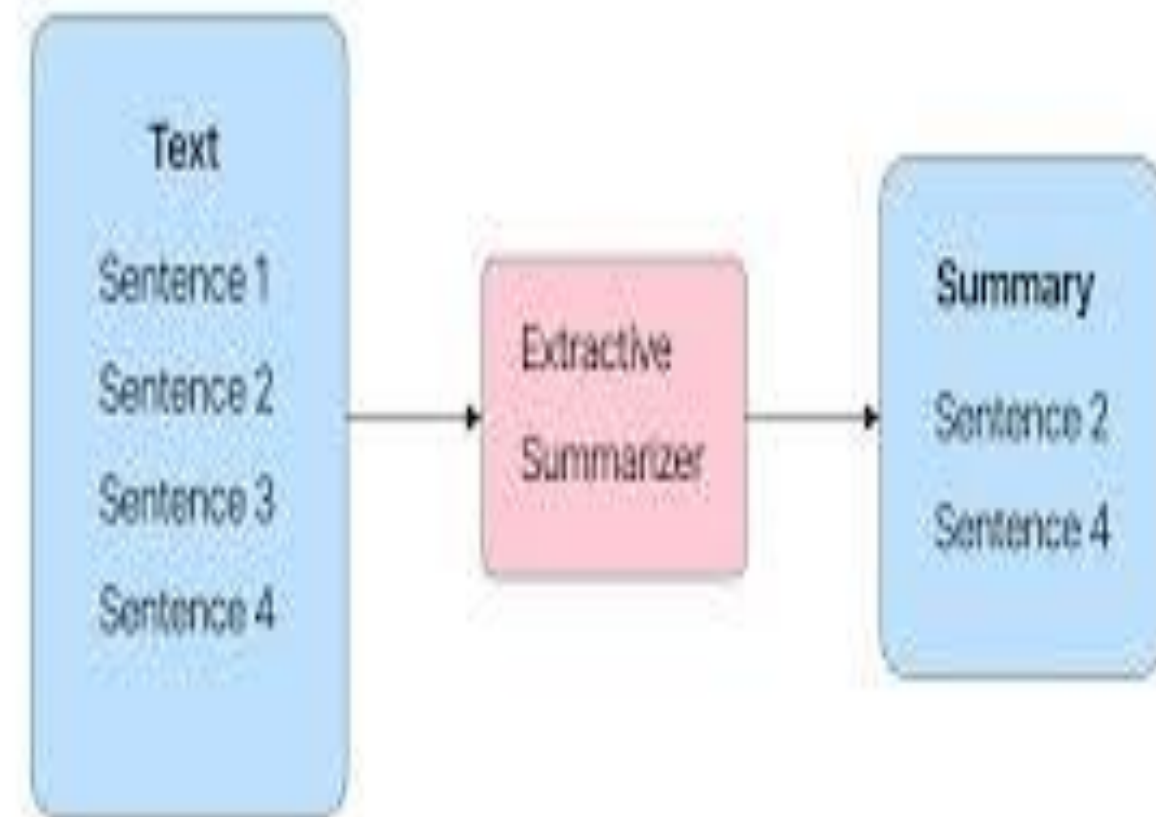
Extractive summarization selects the most important sentences or phrases from the original text and concatenates them to form the summary.

Advantages:

Extractive summaries are typically more faithful to the original text and easier to generate.

Limitations:

Extractive methods can struggle to capture the overall meaning and coherence of the text, leading to less informative summaries



Rough Scores

- **ROUGE-1:** 0.5435 (measures unigram overlap)
- **ROUGE-2:** 0.2868 (measures bigram overlap)
- **ROUGE-L:** 0.4135 (measures the longest common subsequence)
- **ROUGE-Lsum:** 0.4918 (measures the longest common subsequence for summaries)

Rough scores for extractive model

Abstractive Summarization

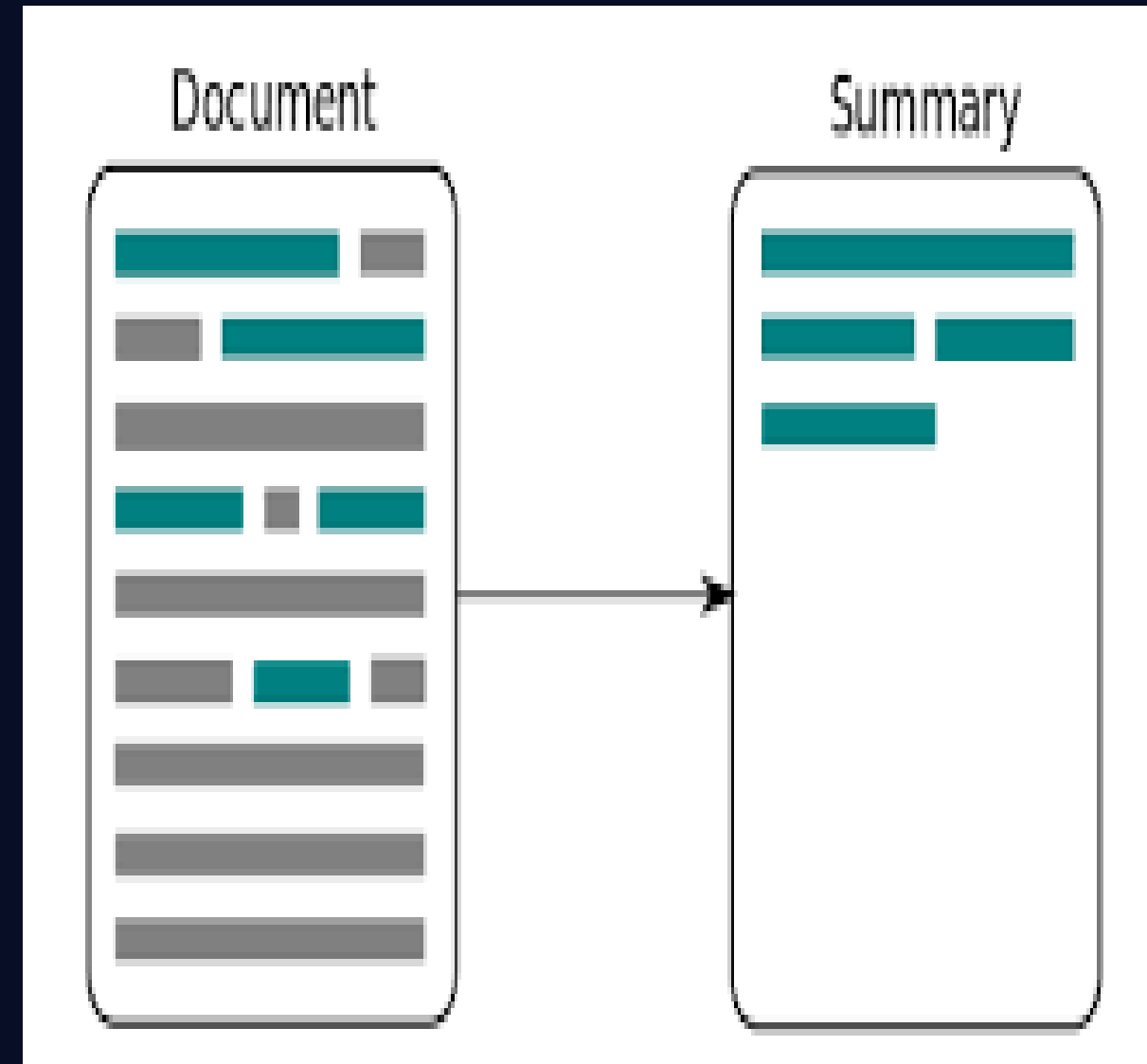
Generating New Text: Abstractive Summarization generates new text that captures the main ideas and essence of the original document, rather than just extracting existing sentences

Deeper Understanding: Abstractive models require a deeper understanding of the text to rephrase the content in a concise and coherent manner.

Challenges: Abstractive summarization is more complex and prone to generating inaccurate or irrelevant information if not properly trained.

Potential: Abstractive summarization has the potential to produce more informative and human-like summaries that better capture the essence of the original text.

Abstractive Summarizer



Rough scores

- **ROUGE-1:** Precision=0.4983, Recall=0.2490, F1=0.4508
- **ROUGE-2:** Precision=0.2162, Recall=0.2734, F1=0.2289
- **ROUGE-L:** Precision=0.5596, Recall=0.6520, F1=0.6181
- **ROUGE-Lsum:** Precision=0.4580, Recall=0.4028, F1=0.4185

Rough scores of abstractive model

Interface

User-Friendly Design

Developing an intuitive and visually appealing interface that makes it easy for users to input text and view the generated summaries.

Interactive Features

Incorporating interactive elements, such as sliders or dropdown menus, to allow users to customize the summarization settings and preferences.

Real-Time Summarization

Implementing the text summarization models to provide instant summaries as users type or upload their input text.

Text Summarization

Not secure ec2-54-166-136-40.compute-1.amazonaws.com:8000

☆

Text Summarization

Enter Text:

Extractive ▼

A Triple M Radio producer has been inundated with messages from prospective partners after a workplace ploy. After Tuesday's Grill Team show, hosts Matty Johns, Mark Geyer and Gus Worland uploaded a picture of 26-year-old Nick Slater to Facebook with a mobile number where people could reach him. In less than 24 hours, he had received over 130 messages from a varied range of male and female listeners, reports News.com. Triple M producer Nick Slater, (C), pictured with his Grill Team hosts, was flooded with 130 voicemails in 24 hours . Workmates and Grill Team hosts Matty Johns, Mark Geyer and Gus Worland uploaded a picture of 26-year-old Nick Slater to Facebook with a mobile number where people could reach out . The ploy came about after a waitress handed the audio engineer her number while out at some work drinks. Unconvinced it was a one off, his colleagues decided to put it to the test and see if anyone else was romantically interested in him. 'The Producers had a few drinks on Friday & Handsome Nick got a number off the waitress in the first 10 minutes!' 'We don't believe him that this never happens to him. Here's Nicks number...let's see how many calls he gets!' Slater received a torrent of voicemails, ranging from date proposals to 'heavy panting' 26-year-old Slater, a sound engineer, and his Triple M Grill Team workmates . In the following 24 hours Slater received a

Paste ▼

ploy came about after a waitress handed the audio engineer her number while out at some work drinks . Unconvinced it was a one off, his colleagues decided to put it to the test and see if anyone else was romantically interested in him

Copy

Generate

Text Summarization

ec2-54-166-136-40.compute-1.amazonaws.com:8000

Text Summarization

Enter Text:

A Triple M Radio producer has been inundated with messages from prospective partners after a workplace ploy. After Tuesday's Grill Team show, hosts Matty Johns, Mark Geyer and Gus Worland uploaded a picture of 26-year-old Nick Slater to Facebook with a mobile number where people could reach him. In less than 24 hours, he had received over 130 messages from a varied range of male and female listeners, reports News.com. Triple M producer Nick Slater, (C), pictured with his Grill Team hosts, was flooded with 130 voicemails in 24 hours . Workmates and Grill Team hosts Matty Johns, Mark Geyer and Gus Worland uploaded a picture of 26-year-old Nick Slater to Facebook with a mobile number where people could reach out . The ploy came about after a waitress handed the audio engineer her number while out at some work drinks. Unconvinced it was a one off, his colleagues decided to put it to the test and see if anyone else was romantically interested in him. 'The Producers had a few drinks on Friday & Handsome Nick got a number off the waitress in the first 10 minutes!' 'We don't believe him that this never happens to him. Here's Nicks number...let's see how many calls he gets!' Slater received a torrent of voicemails, ranging from date proposals to 'heavy panting' 26-year-old Slater, a sound engineer, and his Triple M Grill Team workmates . In the following 24 hours Slater received a

Paste

Abstractive

Copy

Generate

Conclusion

In conclusion, this project has explored the potential of the T5 model for both extractive and abstractive text summarization.

The results demonstrate the strengths and limitations of each approach, highlighting the importance of tailoring the summarization method to the specific needs of the task and the target audience.



Github Link

[**https://github.com/saitej-a/Infosys_text_summarizer**](https://github.com/saitej-a/Infosys_text_summarizer)

Thank you