

# Text Summarization



*Mentor: Mr. Narendra Kumar*



*Group: 4*

# Introduction

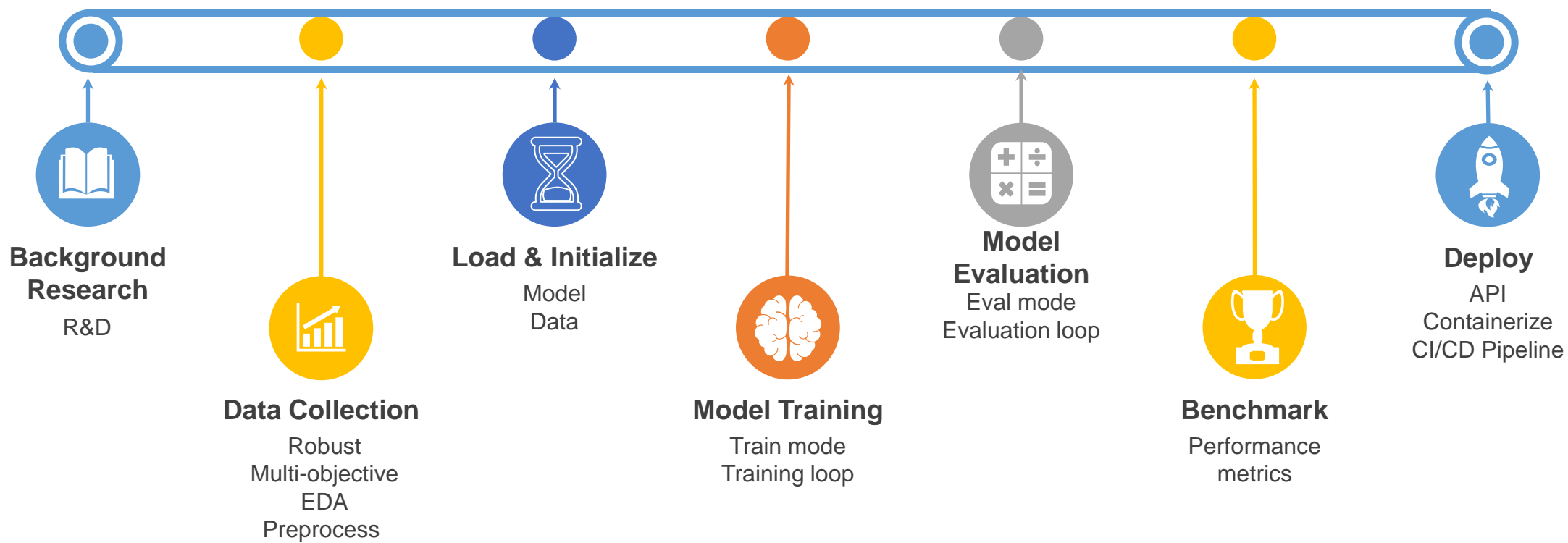
## Problem Statement & Planning

# Introduction

## 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.

# INTENDED PLAN



# Background Research

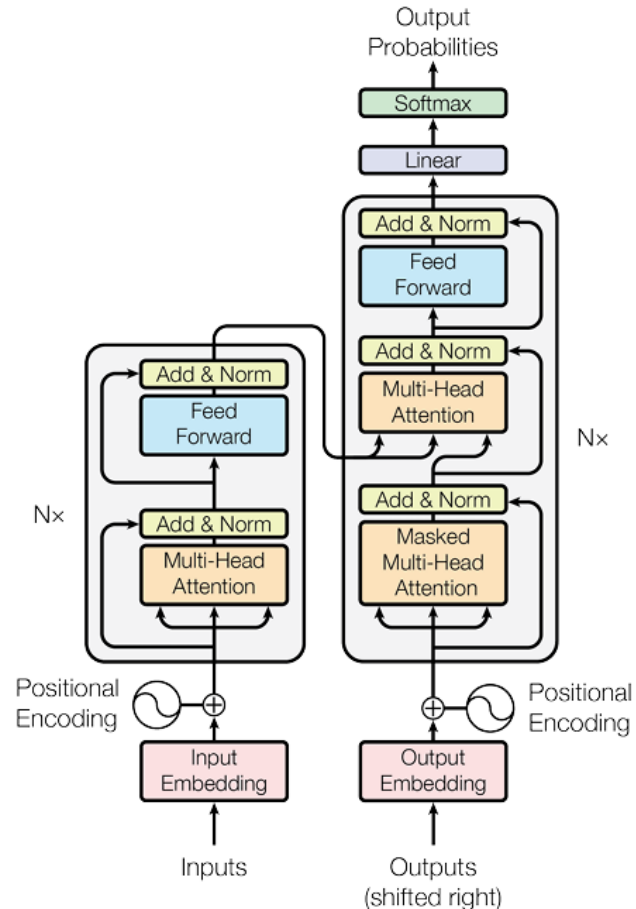
## Literature Review & Findings

# Background Research Literature Review

S. No	Use-Case	Paper Title	Year	Method	Dataset	Results	Limitations
1	General text summarization	Text Summarization Using Deep Learning Techniques: A Review	2023	Deep Learning (Seq2Seq, Attention, Transformers)	CNN/Daily Mail, XSum	Improved performance in capturing semantic relationships, better coherence	Computationally expensive, requires large datasets
2	Implementation of the Transformer architecture	Attention is all you need	2023	Transformer	WMT 2014 English-German, WMT 2014 English-French	Introduced the Transformer architecture, significantly improving the performance of text summarization tasks.	Requires large datasets and computational resources for training.
3	Multi-document summarization	Surveying the Landscape of Text Summarization with Deep Learning	2023	Deep learning methods. Various techniques like RBMs and fuzzy logic employed for summarization.	CNN/DailyMail	Incorporating transfer learning enhances summary quality and reduces data demand.	Complex models, high computational resources
4	Abstractive summarization	Pegasus: Pre-training with gap-sentences for abstractive summarization	2020	Transformer (Pegasus)	XSum, CNN/DailyMail, and Reddit TIFU	Significant improvements in abstractive summarization quality	Resource-intensive
5	Extractive summarization	Text Summarization with Pretrained Encoders	2019	Intersentence Transformer layers for summarization	CNN/Daily Mail, NYT, Xsum, DailyMail	BERT-based models outperformed other approaches in abstractive summarization.	High computational resources required

# Research

## Selected Architecture



[2] Fig. :Transformer architecture:

- **Implementation methods:**

- *From Scratch*

- Build Model
  - NN
- Initialize normalized W&B
- Train model with extensive data
- Hence,
  - Computationally Intensive
  - Sub-Optimal usage of resources
  - Out-of-scope

- *Using Pre-trained model*

- Load Model & its parameters
- Re-Train with specific dataset
- Evaluate
- Hence,
  - Innovation can be done at intended tasks
  - Optimal utilization of resources

# Proposals



# Proposal Workflow

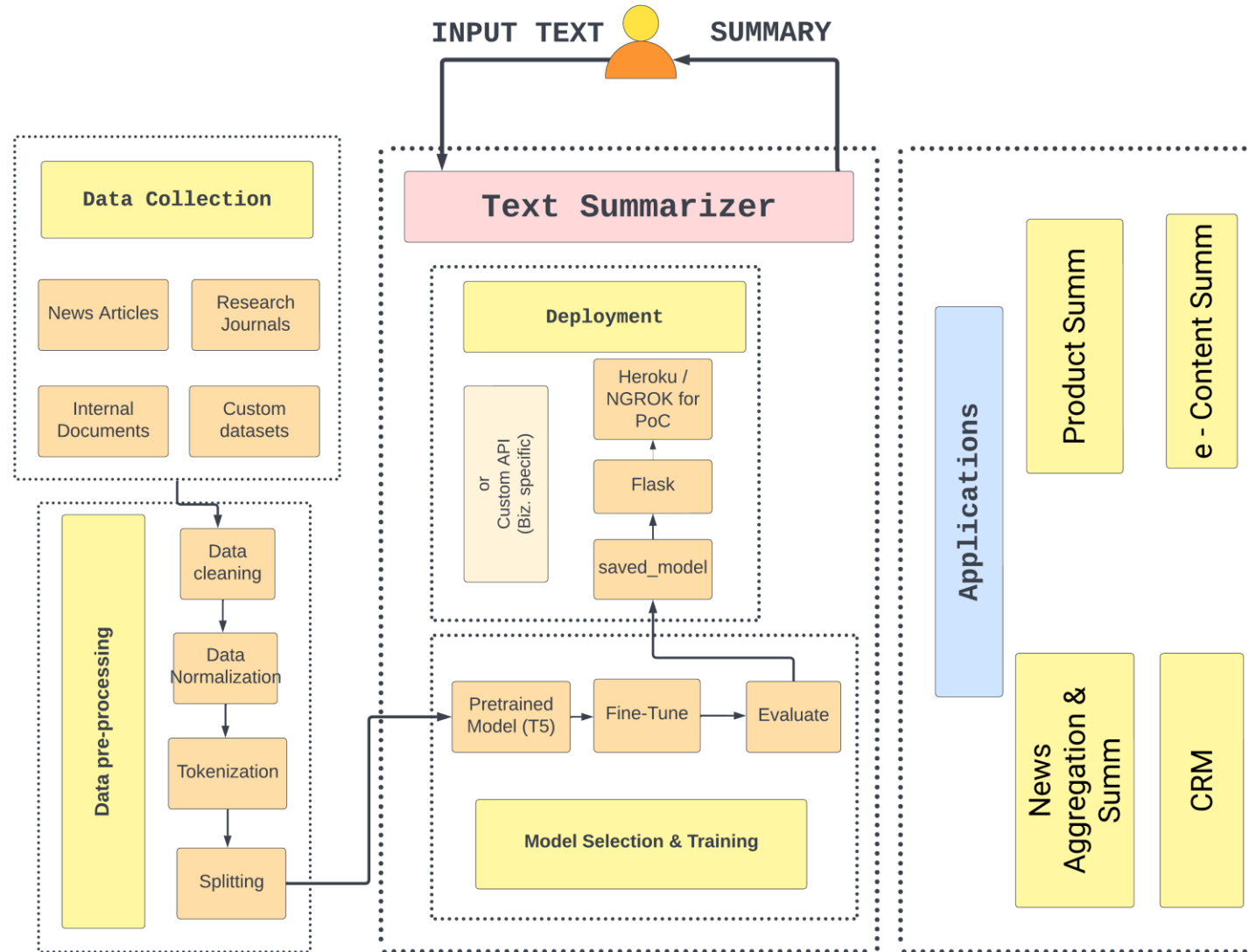


Fig. : Proposed Workflow for Abstractive Text Summarization

# Proposal Workflow

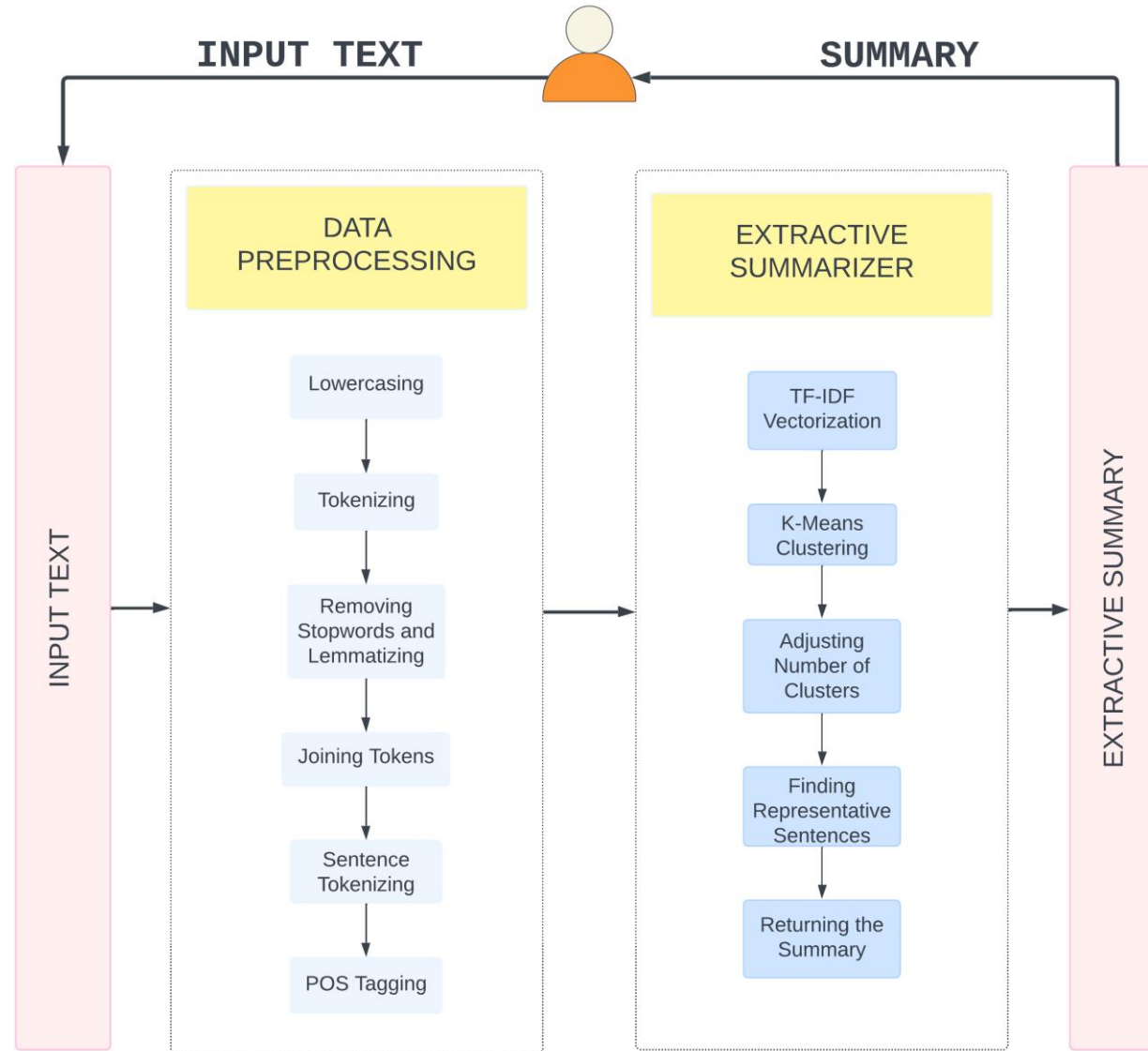


Fig. : Proposed Workflow for Extractive Text Summarization

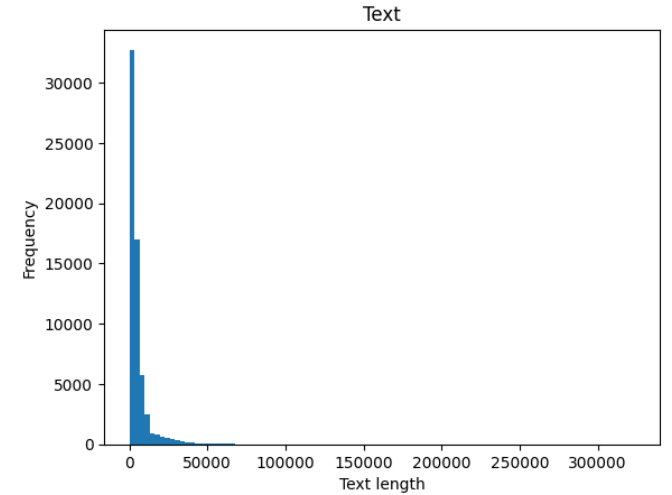
# Proposal Dataset

- Merged selective dataset from
  - CNN, Daily Mail : News,
  - BillSum: Legal,
  - ArXiv : Scientific
  - Dialoguesum : Conversations.
- Completed - data preprocessing
  - Removed
    - NULL records, punctation, stop-words
  - Lowercasing, lemmatization.

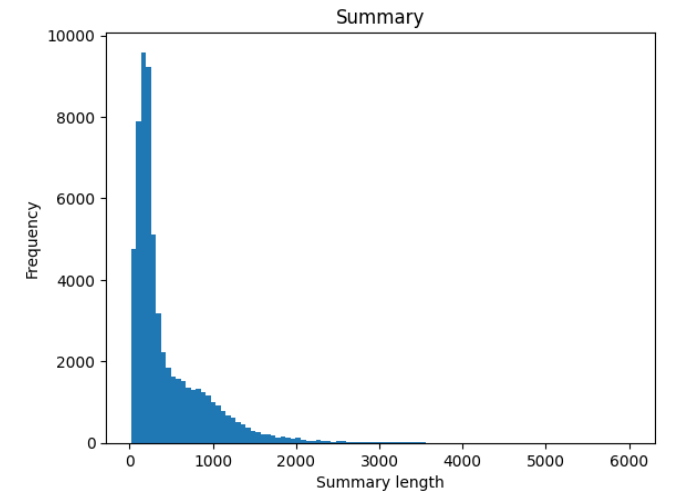
	text	summary
0	section 1 liability business entity providing ...	shield business entity civil liability relatin...
1	section 1 short title act may cited human righ...	human right information act requires certain f...
2	section 1 short title act may cited jackie rob...	jackie robinson commemorative coin act directs...
3	section 1 nonrecognition gain rollover small b...	amends internal revenue code provide temporari...
4	section 1 short title act may cited native ame...	native american energy act sec 3 amends energy...
...	...	...
62702	person1 excuse mr green manchester arent perso...	tan ling pick mr green easily recognized white...
62703	person1 mister ewing said show conference cent...	person1 person2 plan take underground together...
62704	person1 help today person2 would like rent car...	person2 rent small car 5 day help person1
62705	person1 look bit unhappy today whats person2 w...	person2s mom lost job person2 hope mom wont fe...
62706	person1 mom im flying visit uncle lee family n...	person1 asks person2s idea packing bag visitin...

62707 rows × 2 columns

```
count    62707.000000
mean      5211.270975
std       7794.860686
min        83.000000
25%      1275.000000
50%      3176.000000
75%      5684.500000
max     323742.000000
Name: text, dtype: float64
```



```
count    62707.000000
mean      448.081937
std       459.087443
min        16.000000
25%       154.000000
50%       255.000000
75%       618.000000
max     6014.000000
Name: summary, dtype: float64
```



\* In characters.

[https://drive.google.com/drive/folders/1yH89iZmARdc-R7QY6pwfE8tbOJI\\_n9K8?usp=sharing](https://drive.google.com/drive/folders/1yH89iZmARdc-R7QY6pwfE8tbOJI_n9K8?usp=sharing)

[Infosys Springboard Text-Summarization/src/data\\_preprocessing.ipynb](https://github.com/MohanKrishnaGR/Infosys-Springboard-Text-Summarization) at main · MohanKrishnaGR/Infosys Springboard Text-Summarization (github.com)

# Proposal

## Model Training



Fig. : Fine-Tuning Overview

- Proposed implementation – Two – 2 Methods
  - Method 1 – Native PyTorch Method
  - Method 2 – Trainer Class Method

# Proposal

## Model Training (Method 1)



Hugging Face

PyTorch

- Load pre-trained transformer
  - Facebook's Bart Large
- OOP implementation of Dataset
  - Feature, Target
  - Tokenize
  - Padding, Truncate
  - Convert to Tensor
  - Pass to: DataLoader – with batch size
- Training Loop
  - Adam optimizer
  - Forward pass & compute loss
  - Backward pass
  - Update params – compute gradient
  - Update LR
  - Zero the gradients
  - Update total loss
- Only minimal train loss of 1.3280.
  - But, produced inconsistent results.
  - Cannot be pushed into production.
- Raises the need for optimized training and eval loop for Transformer.

Fig. : Screenshot

# Proposal

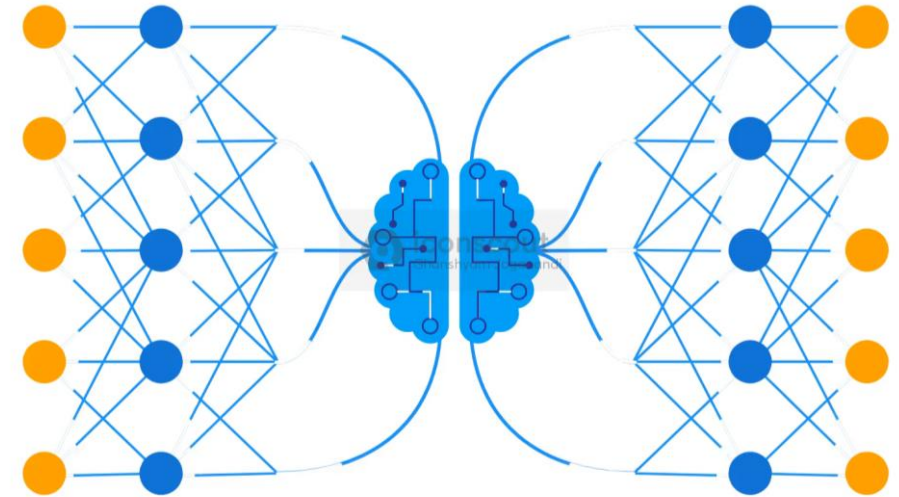
## Model Training (Method 2)

- Trainer Method
- Implemented in src/bart.ipynb.
- A function was implemented for the dataset, to convert text data into model inputs and targets.
- Trainer class from transformer package was utilized for training and evaluation. Tainer is a simple but feature-complete training and eval loop for PyTorch, optimized for transformers.
- The model was trained with whole dataset for 10 epochs for 26:24:22 (HH:MM:SS) in 125420 steps.
- Training Loss = 17.4700
- Considered the performance metrics of the models trained by the forementioned methods. After the due analysis, the model trained using 'Method 2' was selected.



Hugging Face

PyTorch



[Infosys\\_Springboard\\_Text-Summarization/src/bart.ipynb at main · MohanKrishnaGR/Infosys\\_Springboard\\_Text-Summarization \(github.com\)](https://github.com/MohanKrishnaGR/Infosys_Springboard_Text-Summarization)

[https://drive.google.com/drive/folders/1tNdLI67UTc5es6VB\\_dml8b5gkRUWzupl?usp=drive\\_link](https://drive.google.com/drive/folders/1tNdLI67UTc5es6VB_dml8b5gkRUWzupl?usp=drive_link)

# Proposal

## Model Validation

- Performance metrics – **ROUGE** (Recall-Oriented Understudy for Gisting Evaluation)
  - Overlap between generated summary and reference summary.
  - Best suited : evaluating 'Text Summarization' tasks.
  - Other options : **BLEU**.
- ROUGE-N: Measures the overlap of n-grams (contiguous sequences of n items) between the candidate summary and the reference summaries.
  - **ROUGE-1:**
    - Overlap of unigrams (single words).
  - **ROUGE-2:**
    - Overlap of bigrams (two-word sequences).
  - **ROUGE-L:**
    - Measures the longest common subsequence (LCS) between the candidate and reference summaries.
  - **ROUGE-LSUM**
    - (LCS Summary) - variant of the ROUGE-L metric, specifically designed to evaluate the quality of summaries.
- Aimed to: implement custom evaluation function, using ROUGE based on model's inference.

[Infosys\\_Springboard\\_Text-Summarization/src/evaluation.ipynb at main · MohanKrishnaGR/Infosys\\_Springboard\\_Text-Summarization \(github.com\)](#)

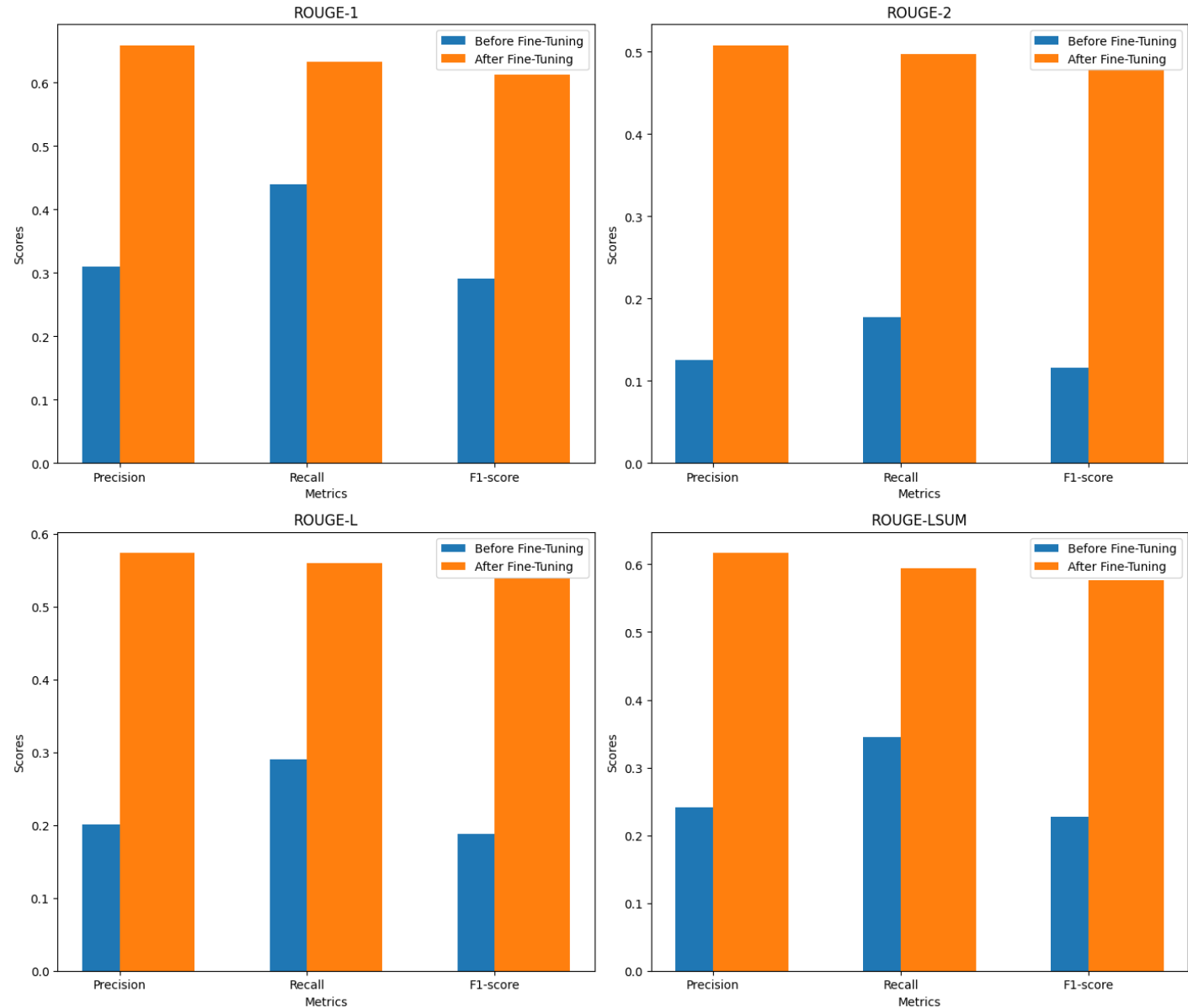
[Infosys\\_Springboard\\_Text-Summarization/src/roque.ipynb at main · MohanKrishnaGR/Infosys\\_Springboard\\_Text-Summarization \(github.com\)](#)



# Proposal

## Comparative Analysis

- Analysis of the transformer's performance metrics before and after Fine-Tuning.
- The transformer model shows significant improvements across all ROUGE metrics after fine-tuning.
- The most substantial gains observed in ROUGE-2 scores. (F1-score=61.32)
- This indicates that the fine-tuning process has notably enhanced the model's ability to generate more accurate and relevant summaries.
- The model is now more proficient at generating summaries that are precise, comprehensive, and contextually accurate.
- Will act as a powerful tool for a variety of Business applications that require efficient and effective text summarization.





# Proposal Testing



**Infosys Springboard - Text Summarizer**

A simple and efficient text summarizer. Enter your text in the box below and get a concise summary.

Input text

At least 49 migrant workers, including around 40 Indian citizens, have died in a deadly fire that devastated a building in Kuwait's southern district of Al-Mangaf. The fire that broke out in the apartment building located in Kuwait's Al Ahmadi Governorate early on Wednesday also left more than a dozen injured, who were admitted to nearby hospitals, reported the Kuwait News Agency (KUNA). Prime Minister Narendra Modi and External Affairs Minister S. Jaishankar expressed shock over the incident, and Congress leader Rahul Gandhi expressed 'serious concern' about the condition of Indians in the Gulf region.

"My thoughts are with all those who have lost their near and dear ones. I pray that the injured recover at the earliest. The Indian Embassy in Kuwait is closely monitoring the situation and working with the authorities there to assist the affected," said Mr. Modi in a message. Mr. Modi held a review meeting on Wednesday evening about the condition of the affected Indians in Kuwait. He deputed Minister of State for External Affairs Kirti Vardhan Singh to oversee the help being rendered to the injured and to bring back the remains of the Indians who perished in the incident.

Indian ambassador to Kuwait Adarsh Swaika visited the Mubarak Al-Kabeer Hospital where 11 injured workers were admitted. "Ten of them are expected to be released today and one in hospital is reportedly stable," the Indian embassy said in a statement. The Government of Kuwait has not made any statement officially so far, but Interior Minister Sheikh Fahad Al-Yousuf Al-Sabah has ordered the police to arrest the owner of the building located in Al-Mangaf.

The incident has highlighted the poor living conditions of blue-collar Indian workers in the region.

Clear Submit

Summarized Text

Fire broke out in apartment building located in Kuwait's Al Ahmadi Governorate early on Wednesday. At least 49 migrant workers, including around 40 Indian citizens, have died in the blaze. The Indian embassy in Kuwait is monitoring the situation and working with the authorities there to assist the affected people.

Flag

Powered by Infosys Springboard Intern. Let's connect, [LinkedIn](#).

- Simple interface for the Deep Learning model, developed using Gradio.
- Gradio is an open-source Python package that allows us to quickly build a demo - web-application for the trained models.
- Enables us to test and even deploy the trained model.

[Infosys\\_Springboard\\_Text-Summarization/src/interface.ipynb](#) at main · MohanKrishnaGR/Infosys\_Springboard\_Text-Summarization ([github.com](#))

# Proposal

## Extractive Text Summarization

- Rather than choosing computationally intensive deep-learning models, utilizing a rule based approach will result in optimal solution. Utilized a new-and-novel approach of combining the matrix obtained from TF-IDF and KMeans Clustering methodology.
- It is the expanded topic modeling specifically to be applied to multiple lower-level specialized entities (i.e., groups) embedded in a single document. It operates at the individual document and cluster level.
- The sentence closest to the centroid (based on Euclidean distance) is selected as the representative sentence for that cluster.
- Implementation: Preprocess text, extract features using TF-IDF, and summarize by selecting representative sentences.
- Source code for implentation & evaluation: `src/Extractive_Summarization.ipynb`
- ROUGE1 (F-Measure) = 24.71

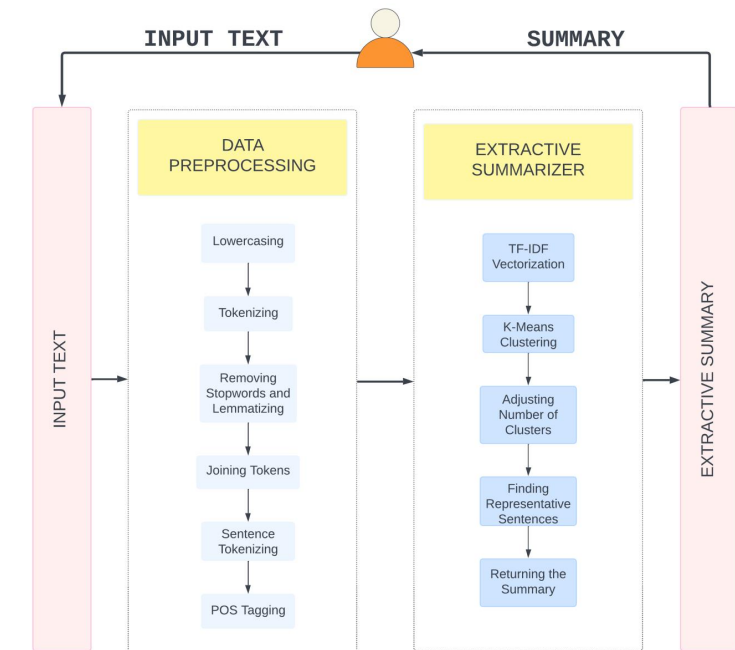
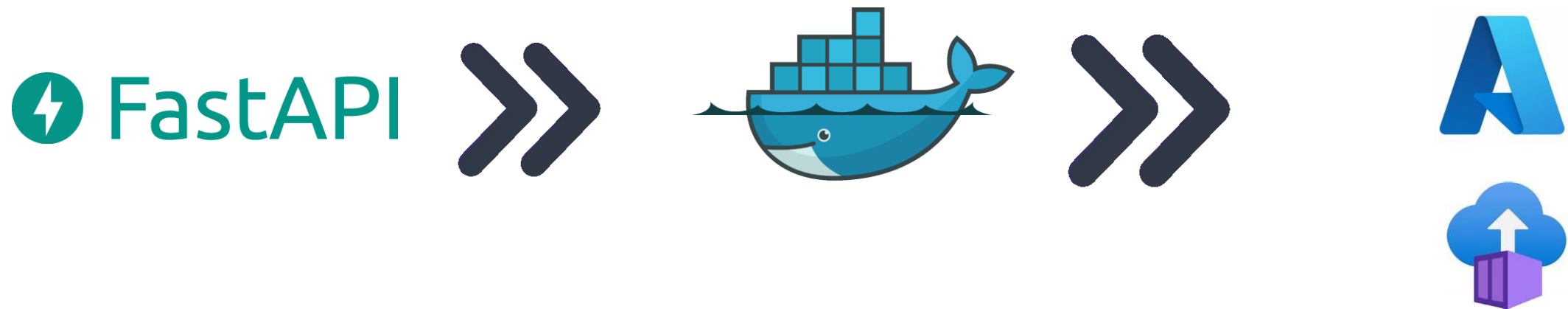


Fig. : Proposed Workflow for Extractive Text Summarization

# Proposal

## Deployment



[Infosys\\_Springboard\\_Text-Summarization/summarizer at main · MohanKrishnaGR/Infosys\\_Springboard\\_Text-Summarization \(github.com\)](#)

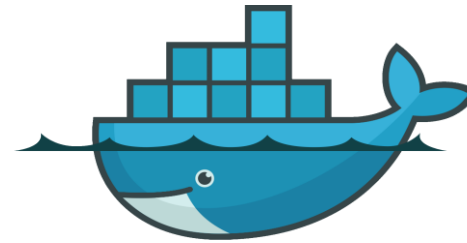
# Proposal

## Deployment



- Implemented extractor modules for text extraction from URL, PDF, docx.
- Defined the API endpoints. (FastAPI)
  - Accepts: Text, URL, Files (PDF, docx)
  - Returns:
    - Abstractive & Extractive Summary
- Utilized 'jQuery' for a dynamic webpage.

- Containerized the entire application along with the deep-learning models.
  - Built the image & Pushed into docker hub.



- Drawback = Less computation for free-tier plan (t2.micro)

- Deployed the docker image using Azure Container Instance
- Integrated with GitHub actions – CI/CD pipeline
- Advantage = 4 CPU cores for Free Trail



# Results

# Results

## GitHub Repository

The screenshot shows a GitHub repository page for 'Infosys Springboard Text Summarization'. The repository is owned by 'MohanKrishnaGR' and has 27 commits. The main branch is 'main'. The repository contains several files and folders: '.github/workflows', 'models', 'src', 'summarizer', 'Group Report.pdf', and 'README.md'. The 'README.md' file is selected and displays the 'Infosys Springboard' logo. Below the logo, there is a section titled 'Text Summarization' which states: 'A project by AI/ML Interns (Group 4) @ Infosys Springboard, Summer 2024.' The 'Mentor' is listed as 'Mr. Narendra Kumar'. The repository also features a CI/CD Pipeline status showing 'passing'. On the right side, the 'About' section describes the repository as containing the implementation of a Transformer-based model for abstractive text summarization and a rule-based approach for extractive text summarization. The 'Releases' and 'Packages' sections indicate no releases or packages have been published. The 'Languages' section shows that the repository is primarily composed of Jupyter Notebook files (99.8%) and other files (0.2%).

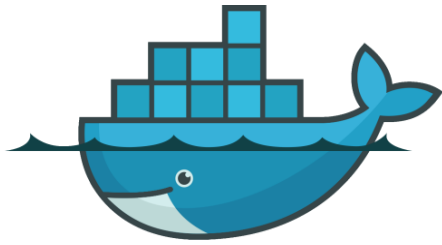
Fig. : Screenshot of the GitHub repository.

[Infosys Springboard Text-Summarization](#)

GROUP 4

# Results

## Deployment - Ref. Links



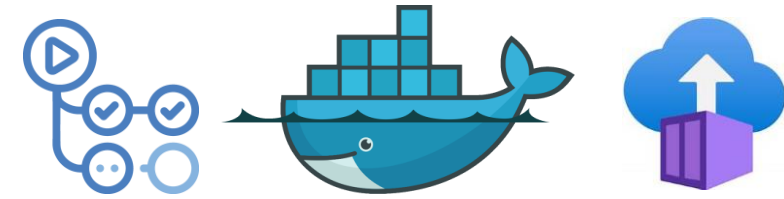
- [mohankrishnagr/infosys\\_text-summarization - Docker Image | Docker Hub](https://hub.docker.com/r/mohankrishnagr/infosys_text-summarization)



- [Text-Summarizer](#)
  - <http://text-summarizer.bqegenbyedfzhpa3.centralindia.azurecontainer.io:8000/>
  - <http://20.235.235.107:8000/>

# Results

## CI/CD pipeline



Screenshot of the GitHub Actions CI/CD pipeline interface for the repository `MohanKrishnaGR / Infosys_Springboard_Text-Summarization`.

The interface shows the **Actions** tab selected, displaying the **CI/CD Pipeline** workflow. The workflow status is **Updated Group Report With CI/CD #10**, which has succeeded 3 days ago in 9m 4s.

The **build-and-deploy** job is selected, showing a list of steps that all succeeded:

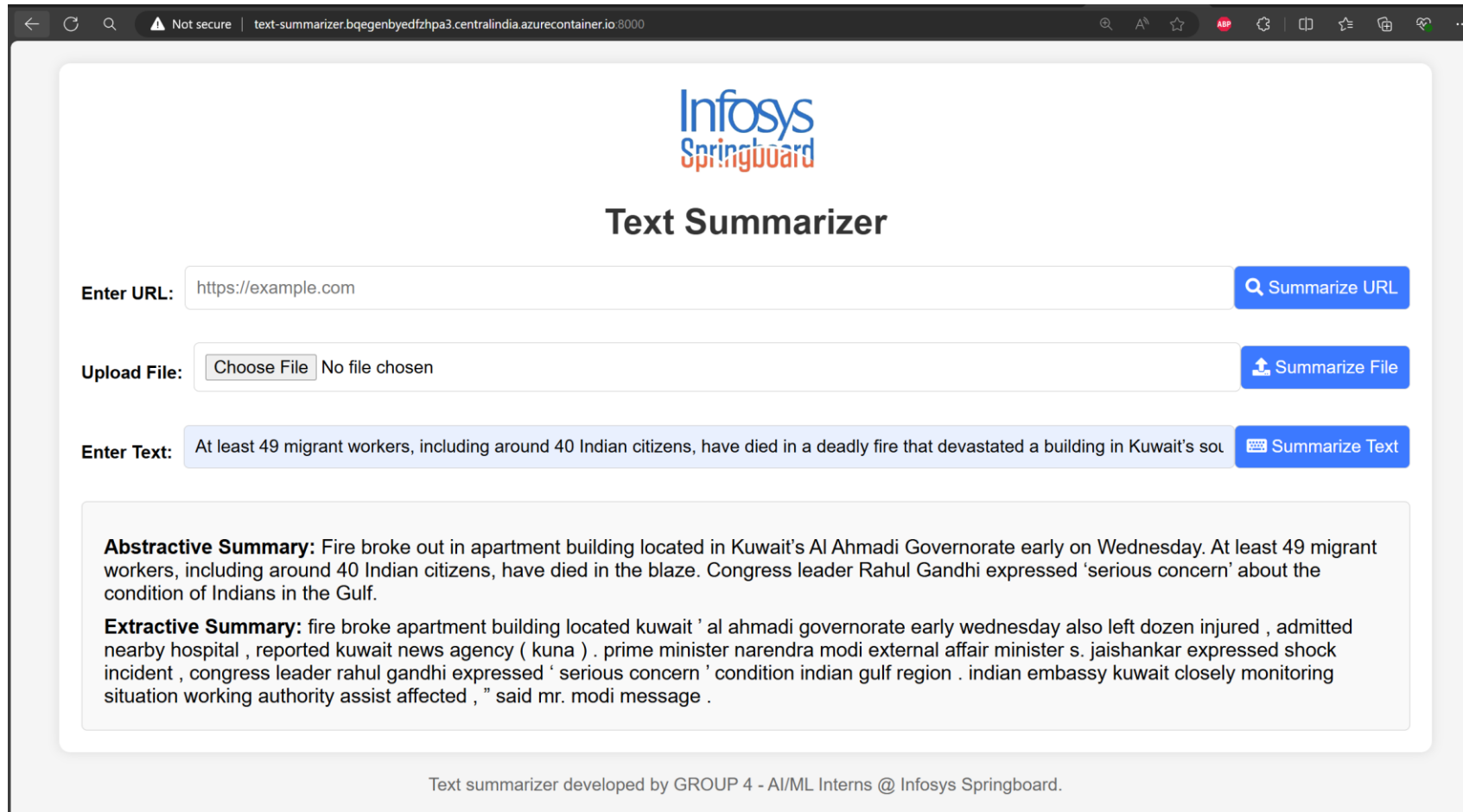
Step	Duration
Set up job	2s
Pre Login via Azure CLI	15s
Checkout code	1s
Log in to Docker Hub	0s
Build and push Docker image	5m 18s
Login via Azure CLI	2s
Install Azure CLI	13s
Restart Azure Container Instance	3m 10s
Post Checkout code	0s
Post Login via Azure CLI	0s
Complete job	0s

Fig. : Screenshot of the CI/CD pipeline.



# Results

## Deployed Application



The screenshot shows a web browser window with the URL `text-summarizer.bqegenbyedfzhp3.centralindia.azurecontainer.io:8000`. The page features the Infosys Springboard logo and the title "Text Summarizer". It has three input methods: "Enter URL:" with a text box containing `https://example.com` and a "Summarize URL" button; "Upload File:" with a "Choose File" button and a "Summarize File" button; and "Enter Text:" with a text box containing a news snippet about a fire in Kuwait and a "Summarize Text" button. Below the inputs, the application displays two types of summaries for the provided text: an "Abstractive Summary" which is a concise, rephrased version of the text, and an "Extractive Summary" which is a direct extraction of key sentences from the original text. At the bottom, a footer states "Text summarizer developed by GROUP 4 - AI/ML Interns @ Infosys Springboard."

Infosys  
Springboard

### Text Summarizer

Enter URL:  Summarize URL

Upload File:  No file chosen Summarize File

Enter Text:  Summarize Text

**Abstractive Summary:** Fire broke out in apartment building located in Kuwait's Al Ahmadi Governorate early on Wednesday. At least 49 migrant workers, including around 40 Indian citizens, have died in the blaze. Congress leader Rahul Gandhi expressed 'serious concern' about the condition of Indians in the Gulf.

**Extractive Summary:** fire broke apartment building located kuwait ' al ahmadi governorate early wednesday also left dozen injured , admitted nearby hospital , reported kuwait news agency ( kuna ) . prime minister narendra modi external affair minister s. jaishankar expressed shock incident , congress leader rahul gandhi expressed ' serious concern ' condition indian gulf region . indian embassy kuwait closely monitoring situation working authority assist affected , " said mr. modi message .

Text summarizer developed by GROUP 4 - AI/ML Interns @ Infosys Springboard.

Fig. : Screenshot of the Deployed application – Text input.

# Results

## Deployed Application

The screenshot displays the 'Infosys Springboard Text Summarizer' web application. It features three input methods: URL, File Upload, and Text Entry. The 'Upload File' section is active, showing a file named 'optimizing transformer inference.pdf'. Below the inputs, the 'Abstractive Summary' of the PDF is displayed, detailing a survey of transformer optimization techniques. The footer indicates the application was developed by GROUP 4 - AI/ML Interns @ Infosys Springboard.

**Infosys Springboard**

### Text Summarizer

Enter URL:  [Summarize URL](#)

Upload File:   [Summarize File](#)

Enter Text:  [Summarize Text](#)

**Abstractive Summary:** This paper presents a comprehensive survey of techniques for optimizing the inference phase of transform networks. Researchers have proposed techniques to optimize transformer inference at all levels of abstraction. The survey includes: (1) knowledge distillation, pruning, quantization, neural architecture search and lightweight network design at the algorithmic level; (2) hardware-level optimization techniques and the design of novel hardware accelerators for transformers; (3) quantitative results on the number of parameters/FLOPs and the accuracy of several models/techniques to showcase the tradeoff exercised by them; and (4) future directions in this rapidly evolving field of research.

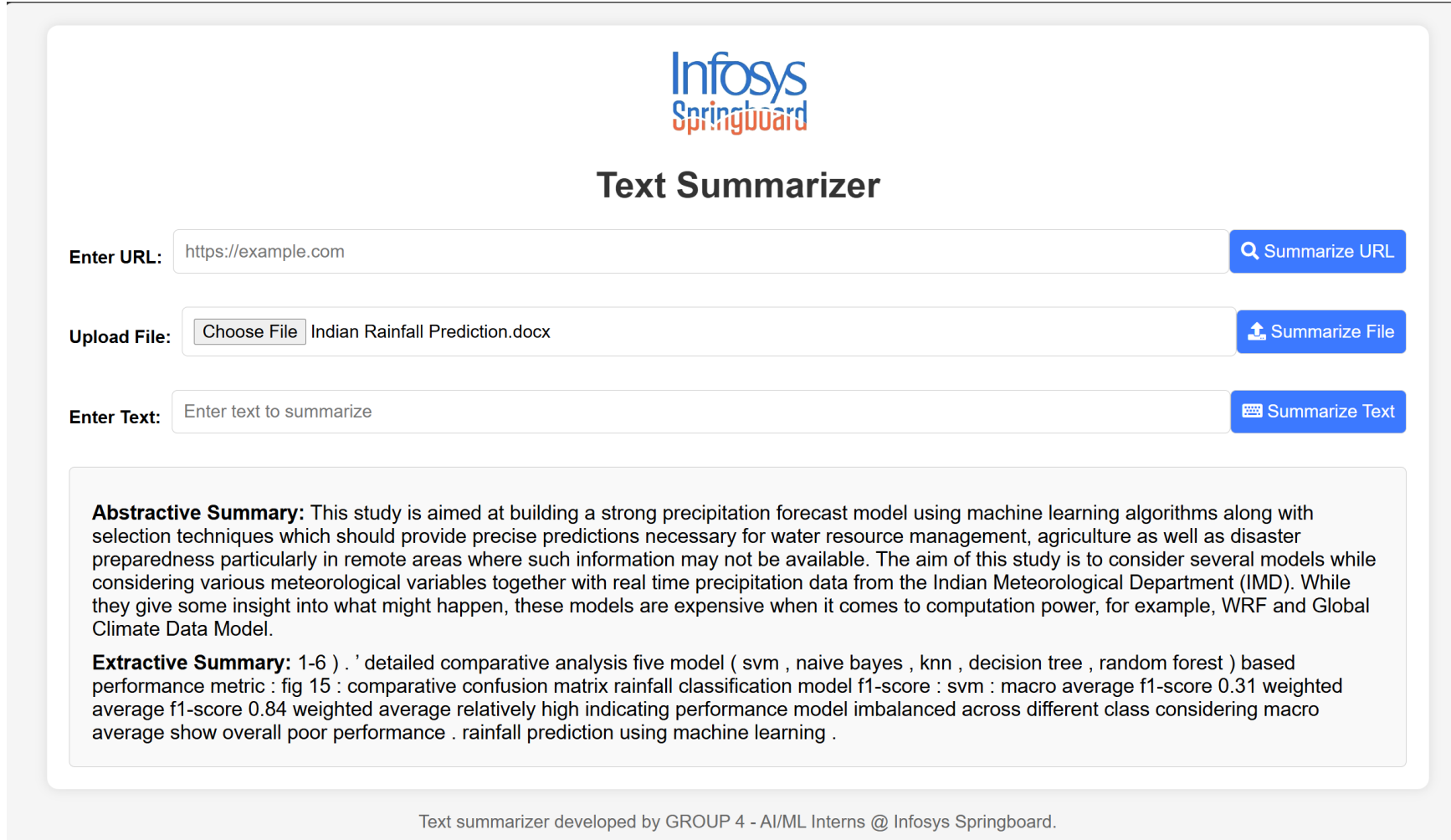
**Extractive Summary:** , keep comparable non-zero element sub-row , non-zero element pe array row-width split multiple row . , training compute-optimal large language model , 2022 , arxiv preprint arxiv:2203.15556 . 1 % accuracy loss , proposed accelerator provides high speedup energy efficiency gain compared gpu . fig . 6.2.2. matching full precision model subsection , discuss quantization method quantize model achieving accuracy original model . output attention value column-wise manner . [ 6 ] . non-uniform token pruning technique adapt pruning percentage based characteristic input sequence . 4.3.4. embedding-layer distillation addition model-level , attention-level hidden state , knowledge teacher embedding layer transferred student ' equivalent layer learn embedding layer . : proceeding ieee/cvf conference computer vision pattern recognition , 2022 , pp .

Text summarizer developed by GROUP 4 - AI/ML Interns @ Infosys Springboard.

Fig. : Screenshot of the Deployed application – PDF input.

# Results

## Deployed Application



The screenshot displays the 'Text Summarizer' web application. At the top, the 'Infosys Springboard' logo is centered. Below it, the title 'Text Summarizer' is prominently displayed. The interface features three input methods: a URL input field with the placeholder 'https://example.com' and a 'Summarize URL' button; a file upload section with a 'Choose File' button, the filename 'Indian Rainfall Prediction.docx', and a 'Summarize File' button; and a text input field with the placeholder 'Enter text to summarize' and a 'Summarize Text' button. The output area, enclosed in a light gray box, contains two paragraphs of summarized text. The first paragraph is an 'Abstractive Summary' describing a study on precipitation forecast models. The second paragraph is an 'Extractive Summary' providing a detailed comparative analysis of five machine learning models (svm, naive bayes, knn, decision tree, random forest) based on performance metrics like f1-score and macro average. At the bottom of the application, a footer note states: 'Text summarizer developed by GROUP 4 - AI/ML Interns @ Infosys Springboard.'

Infosys  
Springboard

### Text Summarizer

Enter URL:  Summarize URL

Upload File:   Summarize File

Enter Text:  Summarize Text

**Abstractive Summary:** This study is aimed at building a strong precipitation forecast model using machine learning algorithms along with selection techniques which should provide precise predictions necessary for water resource management, agriculture as well as disaster preparedness particularly in remote areas where such information may not be available. The aim of this study is to consider several models while considering various meteorological variables together with real time precipitation data from the Indian Meteorological Department (IMD). While they give some insight into what might happen, these models are expensive when it comes to computation power, for example, WRF and Global Climate Data Model.

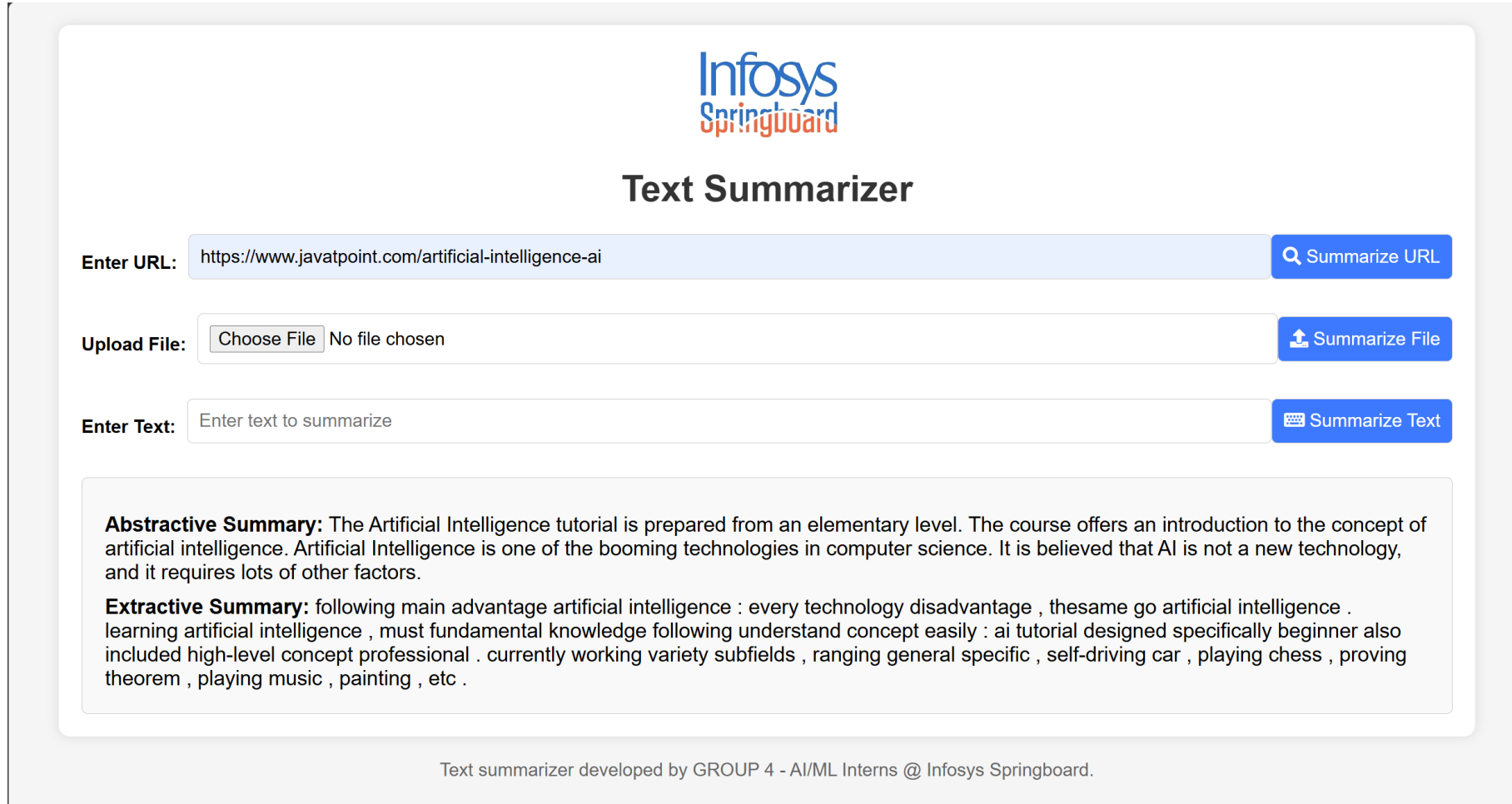
**Extractive Summary:** 1-6 ) . ' detailed comparative analysis five model ( svm , naive bayes , knn , decision tree , random forest ) based performance metric : fig 15 : comparative confusion matrix rainfall classification model f1-score : svm : macro average f1-score 0.31 weighted average f1-score 0.84 weighted average relatively high indicating performance model imbalanced across different class considering macro average show overall poor performance . rainfall prediction using machine learning .

Text summarizer developed by GROUP 4 - AI/ML Interns @ Infosys Springboard.

Fig. : Screenshot of the Deployed application – DOCX input.

# Results

## Deployed Application



The screenshot displays the 'Infosys Springboard Text Summarizer' web application. At the top, the Infosys Springboard logo is centered. Below it, the title 'Text Summarizer' is prominently displayed. The interface features three input methods: a URL input field with the value 'https://www.javatpoint.com/artificial-intelligence-ai' and a 'Summarize URL' button; a file upload section with a 'Choose File' button, the text 'No file chosen', and a 'Summarize File' button; and a text input field with the placeholder 'Enter text to summarize' and a 'Summarize Text' button. The results are shown in a light gray box at the bottom, containing an 'Abstractive Summary' and an 'Extractive Summary'. The footer text reads: 'Text summarizer developed by GROUP 4 - AI/ML Interns @ Infosys Springboard.'

**Infosys Springboard**

### Text Summarizer

Enter URL:  Summarize URL

Upload File:  No file chosen Summarize File

Enter Text:  Summarize Text

**Abstractive Summary:** The Artificial Intelligence tutorial is prepared from an elementary level. The course offers an introduction to the concept of artificial intelligence. Artificial Intelligence is one of the booming technologies in computer science. It is believed that AI is not a new technology, and it requires lots of other factors.

**Extractive Summary:** following main advantage artificial intelligence : every technology disadvantage , thesame go artificial intelligence . learning artificial intelligence , must fundamental knowledge following understand concept easily : ai tutorial designed specifically beginner also included high-level concept professional . currently working variety subfields , ranging general specific , self-driving car , playing chess , proving theorem , playing music , painting , etc .

Text summarizer developed by GROUP 4 - AI/ML Interns @ Infosys Springboard.

Fig. : Screenshot of the Deployed application – URL input.



THANK YOU !!!

