

# **SUMMAI: AUTOMATED TEXT SUMMARIZATION FOR BUSINESS INSIGHTS**

A Project on NLP-based Extractive Text Summarization

# **GROUP MEMBERS**

---

**AP23110010866 - Madhav**

**AP23110010494 - Dhanush**

**AP23110011073 - Manish**

**AP23110010735 - Snehith**

**AP23110010301 - Shashank**

# PROBLEM STATEMENT

---

Janani, a business analyst, reads many long reports and documents to understand market trends and competitors. Reading all the information takes a lot of time and effort. She needs a tool that can quickly highlight the important points for her. SummAI solves this problem by automatically creating summaries, helping her save time and work more efficiently.

# OBJECTIVE

---

The main objective of SummAI is to automatically summarize large text into a shorter and easier form. It helps people understand the main ideas faster and supports better decision-making. The project uses Python and simple Natural Language Processing (NLP) techniques to clean the text, find important parts, and create useful summaries.

# PROPOSED SOLUTION & WORKFLOW

---

## Proposed Solution:

We built SummAI, an extractive summarization tool that identifies important sentences and presents them as a short summary.

## Workflow:

1. Import libraries
2. Clean the text
3. Remove stopwords
4. Tokenize words and sentences
5. Create a frequency table
6. Score each sentence
7. Select top sentences → Generate summary

This simple workflow helps in extracting useful information from large text.

# **DATASET & METHODOLOGY**

---

## **Dataset:**

We used kaggle website's dataset for testing SummAI.  
Any long paragraph or document can be used as input.

## **Methodology:**

- Preprocessing: Cleaning text, removing unwanted words
  - Tokenization: Breaking text into words & sentences
  - Frequency Analysis: Counting important words
  - Sentence Scoring: Finding sentences with strong meaning
  - Summary Creation: Picking sentences with high scores
- Everything is done using Python and NLTK.

# TECHNOLOGIES USED

---

## Programming Language:

- Python,Html,Css

## Libraries:

- NLTK - tokenization, stopwords
- chardet - file encoding
- pandas - data handling
- re - text cleaning

## Tools:

- Jupyter Notebook / Any Python IDE

These tools make text processing easy and efficient.

# RESULTS & FINDINGS

---

- SummAI successfully reduced long paragraphs into short summaries.
- Important sentences were selected automatically.
- Summaries were clear, meaningful, and easy to read.
- The system helped save time by highlighting key information.
- The method works well for business reports, articles, and general text.

# CONCLUSION & FUTURE SCOPE

---

## Conclusion:

SummAI is a simple but effective tool for quick text summarization. It helps users like Janani understand long content faster and make better decisions.

## Future Scope:

- Add advanced NLP models (BERT, T5) for smarter summaries
- Create a user-friendly app or web interface
- Support multi-language summarization
- Improve accuracy using machine learning models