

AI-Powered Document Summariser

Project Proposal by

Anvit Gupta
22114009

anvit_g@cs.iitr.ac.in

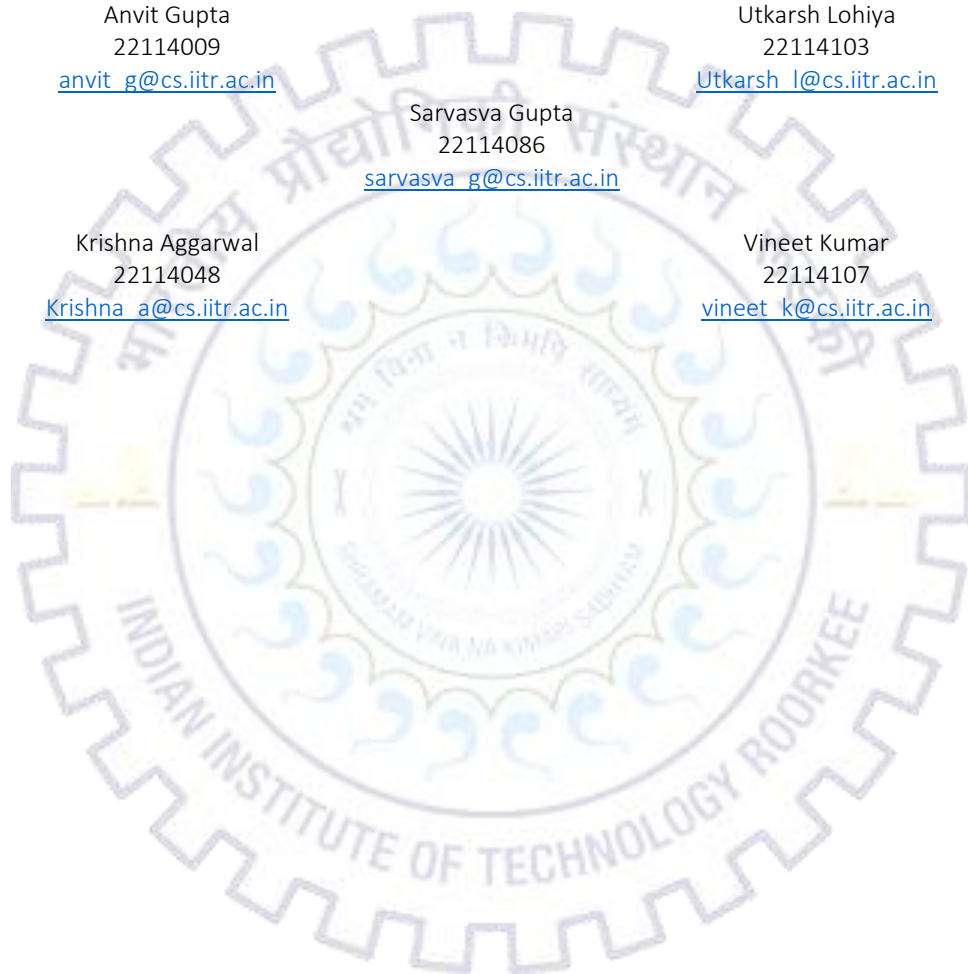
Utkarsh Lohiya
22114103

Utkarsh_l@cs.iitr.ac.in

Sarvasva Gupta
22114086
sarvasva_g@cs.iitr.ac.in

Krishna Aggarwal
22114048
Krishna_a@cs.iitr.ac.in

Vineet Kumar
22114107
vineet_k@cs.iitr.ac.in



Problem Statement

In today's information-rich world, individuals and organizations are inundated with vast amounts of text-based content, ranging from research papers and news articles to legal documents and business reports. This vast amount of information presents a significant challenge when it comes to quickly understanding and extracting key insights from lengthy documents. To address this challenge, there is a growing need for an AI-powered document summarizer that can automatically condense large volumes of text into concise, coherent, and informative summaries.

The problem at hand is the development of an effective and efficient AI-powered document summarisation system. This system must be capable of processing diverse types of textual documents and producing summaries that capture the most important information while maintaining readability, reliability and context.

Objectives and Importance of the Project

1. **Time-saving and Increased productivity:** AI Document Summarizer will automate the process of summarizing documents, reducing the time required to comprehend the content, hence increasing productivity.
2. **Information Extraction:** They effectively identify and extract key information, main ideas, and important details from a document.
3. **Data Analysis:** Summarized data from reports and research can be more readily converted to data analytics processes, allowing organizations to grasp text-based data quickly.
4. **Resource Savings:** The need for man-labour for reading and summarizing the document shall be reduced, allowing them to focus on other essential parts of their works/projects.
5. **Improved Decision-Making:** Summaries provide decision-makers with concise, actionable insights, facilitating better-informed decisions.
6. **Consistency:** AI-based document summarizer would ensure that summaries are consistent in quality and relevance and reduce human bias or errors in the summarization process.

System Functionality

Key Features

1. **Text Extraction:** The system extracts relevant content from diverse textual documents, including research papers, articles, legal documents, and reports.
2. **Customisation:** Allows users to specify the desired length and depth of the summary for tailored results.

Functionality

1. **Document Input:** Users upload or provide text-based documents to the system in various formats, including PDF, DOC, and plain text.
2. **Pre-processing:** The system cleans and pre-processes the text data, removing stop words, punctuation, and formatting.

3. **Summarisation Process:** Utilizes summarisation algorithms to generate a summary that encapsulates the document's main points and critical insights.
4. **Readability Enhancement:** Ensures the summary maintains readability by organising the content logically and coherently.

How These Features Tackle the Problem Statement

The key features and functionality of the AI-powered document summarisation system address the challenges posed by the information-rich world:

- **Efficiency:** By automating the summarisation process, the system significantly reduces the time and effort required to digest large volumes of text, making it efficient for individuals and organisations.
- **Insight Extraction:** The system extracts key insights and essential information from lengthy documents through NLP and summarization algorithms, aiding quick understanding.
- **Maintaining Context:** The system ensures that the generated summaries maintain context and coherence, allowing users to accurately grasp the document's essence.
- **Customisation:** Customization options enable users to tailor the summaries to their needs, enhancing flexibility and usability.
- **Readability:** By enhancing readability, the system delivers informative and easy-to-understand summaries, enhancing overall utility.

The AI-powered document summarization system is a powerful tool for information management, addressing the challenge of extracting crucial knowledge from a sea of text-based content efficiently and effectively.

Project Feasibility

- ❖ **Needed Resources:** For our AI-powered Document Summarizer project, we will leverage Python for backend development, utilizing NLP libraries such as NLTK or Spacy, and Flask for the server setup. On the front-end, we will employ React-JS and CSS to create a user-friendly interface. Additionally, we'll use file handling libraries, Axios for asynchronous HTTP requests, and design tools like Figma or Freeform for UI/UX. With a version control system like Git on GitHub, our team of 3-5 programmers can effectively collaborate.
- ❖ **Scope of the Project:** The scope of the project encompasses document input, real-time NLP summarization, interactive output, error handling, and optional features like multi-document input and feedback systems.
- ❖ **Complexity and Final Verdict:** Implementing extractive and abstractive summarization adds complexity, but our team's technical skills and the availability of open-source NLP libraries make this project achievable within a semester. By breaking down tasks and adhering to a well-defined timeline, our team of 3-5 programmers can smoothly deliver a fully functional AI document summarizer that meets user expectations and academic deadlines.