

Department of CSE (AI & AIML)

NewsSnips: GLOBAL NEWS ARTICLE SUMMARIZATION AND CLASSIFICATION

Guide:

Asst.prof : Swati Barik

Name of Projectees :

Mohd Talha Ansari

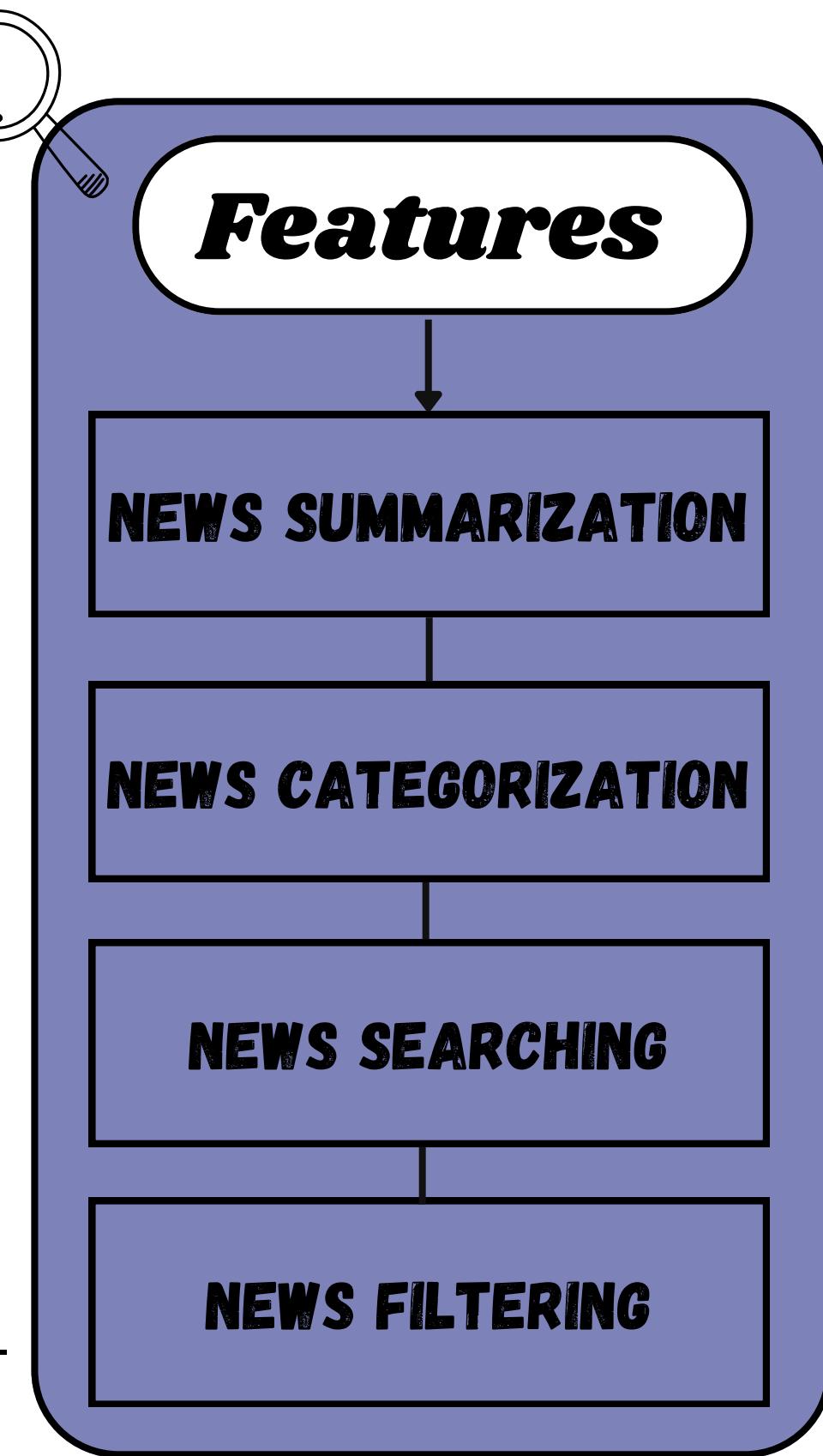
Kajal Metkar

1. Introduction
2. Justifications for Selecting the Title
3. Problem Statement
4. Prototype
5. Literature Survey
6. Block Diagram
7. Expected Result
8. Work plan
9. References

The "Global News Article Summarization and Classification" project is aimed at simplifying the overwhelming amount of information available in today's news landscape. With the vast number of news articles published daily across various platforms, there is a growing need for a system that can filter, categorize, and summarize content in a meaningful and efficient manner. This project leverages Natural Language Processing (NLP) techniques to build a robust system that gathers news from multiple sources, processes the data, and provides users with concise summaries and categorized articles tailored to their interests.

The system implements an array of powerful algorithms such as tokenization, text classification, keyword extraction, and summarization using methods like TF-IDF and Cosine Similarity. By integrating these NLP models, the project enables users to quickly access the most relevant and important content without sifting through lengthy articles.

The interface is designed to enhance user experience through intuitive UI/UX components, ensuring seamless interaction and easy navigation. Additionally, the backend incorporates Flask APIs and JavaScript for efficient data handling and real-time updates, ensuring that users receive the most up-to-date and filtered content.





USE CASES

1. Tailored Articles:

Personalized recommendations for users.

2. Efficient Browsing:

Quick access to relevant content.

3. Summarized Insights:

Brief summaries of extensive articles.

4. Interactive Discussions:

Engage in user conversations.

5. Seamless Sharing:

Easy distribution of articles.

- **Information Overload:** With the exponential growth of online news content, users are often overwhelmed by the sheer volume of information. This makes it difficult for them to find relevant news quickly. By addressing this issue, the project aims to streamline the news consumption process.
- **Time Efficiency:** In today's fast-paced world, users often do not have the time to read lengthy articles. Providing summaries allows users to quickly grasp the key points, saving them time while keeping them informed.
- **Relevance Filtering:** Many news platforms bombard users with a mix of relevant and irrelevant content. By implementing advanced classification and categorization, the project ensures that users can easily access news that is most relevant to their interests or needs.
- **User Experience Enhancement:** A well-organized news platform that categorizes and summarizes content enhances the overall user experience, making it more intuitive and user-friendly. This can lead to increased user engagement and retention.

Problem Statement



Problem

In today's digital age, the overwhelming volume of news from various platforms poses significant challenges for users in accessing relevant information efficiently. Users often face information overload, struggling to find articles that match their interests while sifting through lengthy, irrelevant content. The lack of personalized filtering, accurate categorization, and concise summarization tools further complicates news consumption, making it difficult to stay informed in a time-efficient manner. Existing news aggregation systems often fail to

STRATEGIC INSIGHTS: OUR APPROACH

Goals :

- Enhance User Experience: Create a seamless UI/UX interactive news website allowing customizable news categorization, trending article display, and efficient NLP-driven search.
- Improve Information Retrieval: Utilize advanced NLP and ML-based data preprocessing to summarize large amount of data and streamline news search across large datasets and enable multi-category filtering



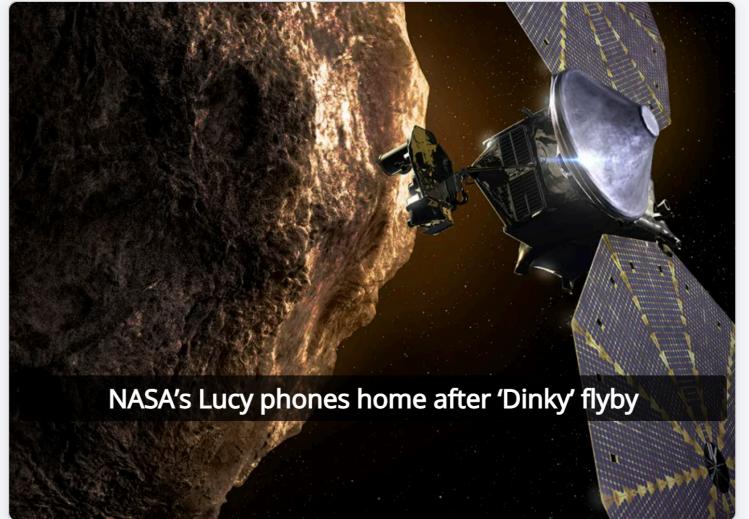
Objectives :

- Develop Advanced NLP: Implement cutting-edge techniques for precise news results.
- Optimize Data Preprocessing: Enhance data quality through data cleaning.
- Enhance Recommendation System: Personalize recommendations based on user preferences.
- Provide Seamless Interaction: Design an intuitive UI for easy navigation and personalized content access.

NewsSnips

Search News

Select Categories
Search by Category



NASA's Lucy phones home after 'Dinky' flyby

Previous
Next



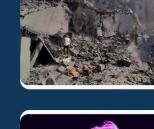
Russia's gold reserves hit record high - RIA



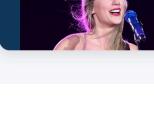
Max is taking 4K away from its legacy ad-free subscribers



Ghostbusters: Rise of the Ghost Lord is light on lore but stuffed with fun



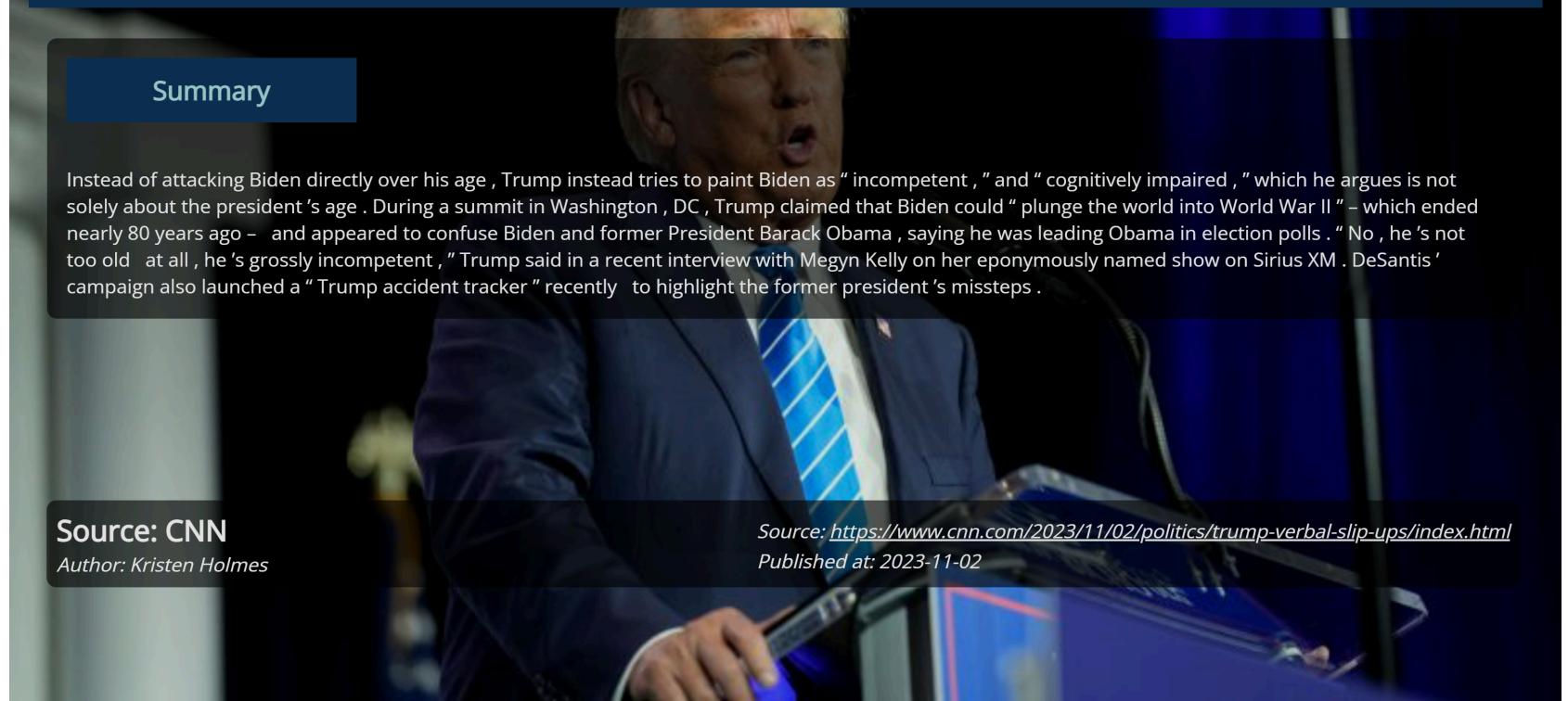
Singapore warns foreigners against using country to further political causes amid Israel-Hamas war



Taylor Swift Adds 2024 Vancouver Dates To Eras Tour

[Trump's verbal slip-ups threaten his argument about Biden | CNN Politics](#)

Summary



Instead of attacking Biden directly over his age , Trump instead tries to paint Biden as " incompetent , " and " cognitively impaired , " which he argues is not solely about the president 's age . During a summit in Washington , DC , Trump claimed that Biden could " plunge the world into World War II " – which ended nearly 80 years ago – and appeared to confuse Biden and former President Barack Obama , saying he was leading Obama in election polls . " No , he 's not too old at all , he 's grossly incompetent , " Trump said in a recent interview with Megyn Kelly on her eponymously named show on Sirius XM . DeSantis ' campaign also launched a " Trump accident tracker " recently to highlight the former president 's missteps .

Source: CNN
Author: Kristen Holmes

Source: <https://www.cnn.com/2023/11/02/politics/trump-verbal-slip-ups/index.html>
Published at: 2023-11-02

1. Main Page

NewsSnips

Search News
Select Categories
Search by Category

Search Results



How a landmark caste census in India threatens Modi's grip on power



Modi's government and some of India's elites are battling for control of one of the country's most elite private clubs



Today in Politics: Modi focus on infra stimulus in poll-bound states, TMC protest set to begin in Delhi



PM Narendra Modi extends Garib Kalyan Anna for 5 years, aims to benefit 80 Crore poor families, slams Congress for caste politics



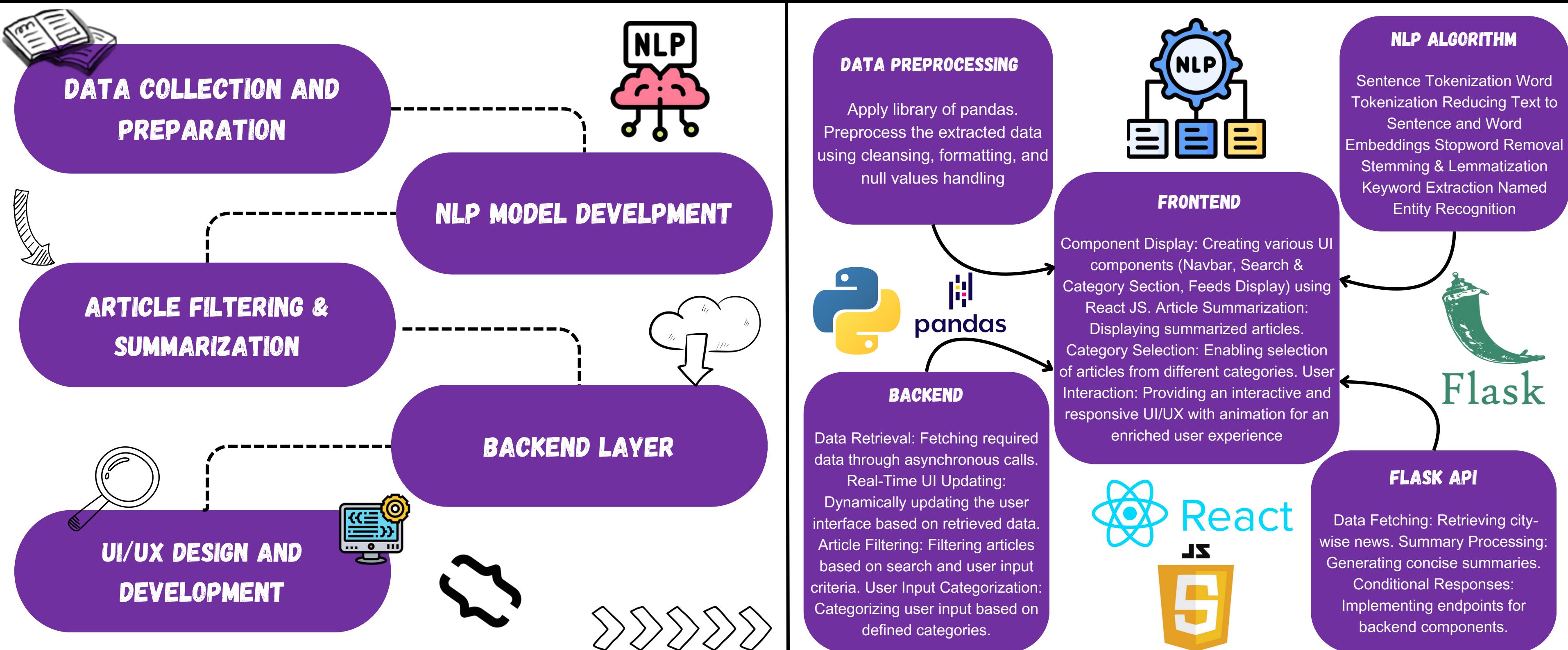
Congress leaders, their statements, announcements are 'filmy', says PM Modi in poll-bound MP

2. Summary Page

3. Search Result

Sr. No.	Paper Title and Year	Details of Publication	Findings
1	Title: "Text Summarization Techniques: A Review" Authors: Priyanka Dey, M. A. Awan -- 2020	Published on: International Journal of Computer Applications link: https://www.researchgate.net/publication/339146200_A_Review_on_Text_Summarization_Techniques	This review summarizes various text summarization methods, including TF-IDF and machine learning approaches, highlighting their application in news summarization.
2	Title: “Analysis of Component Libraries for React JS” Authors: Mukthapuram Praneeth Reddy -- 2021	Published on: IARJSET link: https://www.researchgate.net/publication/353173122_Analysis_of_Component_Libraries_forReact_JS	This paper analyzes various component libraries available for React.js, discussing their features, usability, and performance. It aims to guide developers in choosing the right library based on specific project needs and requirements.
3	Title: “Implementation of TF-IDF and Cosine Similarity Algorithms for Classification of Documents Based on Abstract Scientific Journals”. -- 2021 Authors: Paska Marto Hasugian, Jonson Manurung Logaraz, Uzitha Ram	Published on: JURNAL INFOKUM link : https://sigmodrecord.org/?smd_process_download=1&download_id=4551	This paper discusses the implementation of TF-IDF and cosine similarity algorithms for classifying documents, specifically focusing on abstract scientific journals. The authors present methodologies for effectively utilizing these algorithms to enhance document classification accuracy.
4	Title: “Naïve Bayes Classifier for Classification” -- 2019 Authors: Aji Wibawa, Ahmad Chandra Kurniawan, Della Murbarani, Prawidya Murti, Risky Perdana Adiperkasa	Published on: International Journal of Recent Contributions from Engineering Science & IT (iJES) link: https://www.researchgate.net/publication/33937653_Naive_Bayes_Classifier_for_Journal_Quartile_Classification	This paper explores the application of the Naïve Bayes classifier for the classification of journals based on their quartile rankings. The authors detail the methodology employed to enhance the accuracy of journal classification, providing insights into its effectiveness in academic publishing.

Outlining Our Methodology





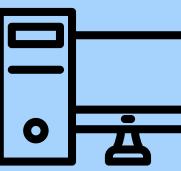
Personalized News Content: Users should receive news articles that are tailored to their interests and preferences. The system should effectively filter and recommend articles that align with the user's specific queries and categories.



Quick and Easy Access to Information: The platform should enable users to quickly find and access relevant news articles without the need to sift through irrelevant content. The search functionality should be efficient, allowing users to retrieve information with minimal effort and a seamless interaction.



Concise Summaries : Users should benefit from concise and accurate summaries of long articles, allowing them to grasp the main points quickly. The summarization feature should save users time while still providing them with the essential information with full information about topic



Intuitive and Engaging User Interface: The interface should be user-friendly, with a clean and organized layout that makes it easy to navigate through the content. Users should be able to filter news, select categories, and adjust summary lengths with ease. The experience should be enhanced with animations and real-time updates, ensuring that the interface is responsive and engaging



Seamless Interaction and Feedback : Users should be able to interact with the system effortlessly, with clear options for filtering, categorizing, and summarizing news articles. The system should respond quickly to user inputs, providing immediate feedback .

Months Activities	August'23	Sept'23	Oct'23	Nov'23
Literature Reviews	✓			
Requirement Analysis	✓			
Designing	✓			
Experimental Analysis		✓		
Module wise Implementation		✓		
Testing and Debugging			✓	
Preparation of Project Report				✓

1.Naïve Bayes Classifier for Classification -- 2019

This paper explores the application of the Naïve Bayes classifier for the classification of journals based on their quartile rankings. The authors detail the methodology employed to enhance the accuracy of journal classification, providing insights into its effectiveness in academic publishing.

2.Application Programming Interface (API) Research: A Review of the Past to Inform the Future 2021

We have learned how to develop robust APIs using Flask, which is crucial for enabling backend communication in a web application. The API endpoints are designed to handle specific tasks like data fetching, summary processing, and conditional responses, ensuring that each function operates efficiently.

3.Article Classification using Natural Language Processing and Machine Learning -- 2019

We have learned that text classification is a critical task in reducing human effort by automatically categorizing large volumes of text data, such as articles, into predefined categories. This automation significantly improves efficiency and accuracy in managing and analyzing textual information.

Thank you

!