AI-DRIVEN NEWS SERVICE

A PROJECT REPORT

Submitted by

Mohammed Anas Bin Mansoor K V
(TKM23MCA-2041)

to
TKM College of Engineering

Affiliated to

The APJ Abdul Kalam Technological University

In partial fulfillment of the requirements for the award of the Degree of

MASTER OF COMPUTER APPLICATION

Thangal Kunju Musaliar College of Engineering
Kerala
DEPARTMENT OF COMPUTER APPLICATION

April 2025

DEPARTMENT OF COMPUTER APPLICATION

Thangal Kunju Musaliar College of Engineering

Kerala

CERTIFICATE

This is to certify that the report entitled **AI-DRIVEN NEWS SERVICE** submitted by **Mohammed Anas Bin Mansoor K V** to TKM College of Engineering affiliated to APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Master of Computer Application is a bonafide record of the project work carried out by him/her under my/our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Internal Supervisor Head of the Department External Examiner

DECLARATION

I undersigned hereby declare that the project report **AI-DRIVEN NEWS SERVICE** submitted for partial fulfillment of the requirements for the award of degree of Master of Computer Application of the APJ Abdul Kalam Technological University, Kerala, is a bonafide work done by me under supervision of Prof. Sheera Shamsu. This submission represents my ideas in my own words and where ideas or words of others have been included, WE have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Place

Date

Mohammed Anas Bin Mansoor K V

ACKNOWLEDGMENT

First and foremost, I thank GOD almighty and my parents for the success of this project. I owe sincere gratitude and heart full thanks to everyone who shared their precious time and knowledge for the successful completion of my project.

I am extremely grateful to Prof. Natheera Beevi M, Head of the Department, Department of Computer Applications, for providing me with best facilities.

I would like to thank my project coordinator Dr. Nadera Beevi S, Department of Computer Applications, who motivated me throughout the project.

I would like to thank my project guide Prof. Sheera Shamsu, Department of Computer Applications, for her invaluable guidance, encouragement and support throughout the course of this project.

I profusely thank all other faculty members in the department and all other members of TKM College of Engineering, for their guidance and inspirations throughout my course of study.

I owe thanks to my friends and all others who have directly or indirectly helped me with the successful completion of this project.

Mohammed Anas Bin Mansoor K V

(Reg. No. TKM23MCA-2041)

ABSTRACT

The AI-Driven News Service is a platform designed to enhance real-time news accessibility by leveraging artificial intelligence (AI) for content delivery and personalization. Built with a Node.js backend and Firebase for secure authentication and data management, it uses web scraping to aggregate the latest news from various sources, ensuring users receive timely updates.The platform addresses traditional news consumption challenges by offering AI-powered features such as article summarization, text-to-speech (TTS) for audio news, and multilingual support via the Gemini 1.5 Flash API. These features enable users to access news in their preferred format, whether through brief summaries or detailed articles. The user-friendly React-based interface ensures smooth navigation across devices. Additional functionalities, such as daily newspaper PDF generation and promotional content submission, enhance user engagement and create revenue opportunities. Future plans include developing a mobile app for on-the-go access, personalized news recommendations, and interactive features like comments and discussions. With these enhancements, the AI-Driven News Service aims to offer a dynamic and accessible news experience in today's digital landscape.

CONTENTS

LIST OF FIGURES **i**

INTRODUCTION **1**

1.1 Existing Systems 2

1.2 Problem Statement 3

1.3 Proposed System 4

1.4 Objectives 5

LITERATURE SURVEY **7**

2.1 Purpose of Literature Review 7

2.2 Related Works 8

METHODOLGY **11**

3.1 System Architecture 11

3.2 Software Requirements 13

3.3 Hardware Requirements 16

3.4 Web Scraping Tools 16

3.5 Data Flow & Processing 19

RESULTS & DISCUSSION **24**

4.1 Results 24

4.2 Performance Metrics & Evaluation 26

CONCLUSION **28**

5.1 Future Scope 29

REFERENCES **32**

Appendix **35**

5.2 Screenshots 35

LIST OF FIGURES

3.1 Use Case Diagram 12

5.2.1 Login Page 35
5.2.2 Home Page 36
5.2.3 Categorial Page - Kerala 36
5.2.4 Home Page - Recent News 37
5.2.5 News Page 37
5.2.6 Admin Module 38
5.2.7 Submit Promotion Page 38

Chapter 1

INTRODUCTION

In today's digital age, accessing news is both easier and more challenging than ever. With countless websites, apps, and platforms, users are often bombarded with an overwhelming amount of information, making it difficult to find relevant and accurate news quickly. Traditional news platforms typically provide a generalized feed, lacking features that cater to the specific needs of diverse audiences. This broad approach can lead to users spending unnecessary time filtering through irrelevant content.

Additionally, accessibility remains a significant challenge. For those needing content in audio format or in languages other than English, options are limited, which can hinder access to information for non-English speakers and individuals with visual impairments. Existing platforms often fall short of delivering inclusive, multilingual solutions that meet the needs of a wide range of users.

The AI-Driven News Service addresses these challenges by leveraging technology to create a usercentric, streamlined news experience. Integrating both global and local news sources through tools like NewsAPI and web scraping, the platform offers a curated selection of articles spanning various topics and regions. Multilingual support enhances inclusivity, allowing users to access news in their preferred language.

Beyond offering broad coverage, the platform uses AI-driven summarization to condense lengthy articles into concise summaries, helping users stay informed efficiently. A text-to-speech feature provides an alternative way to consume news, especially useful for those on the go or with visual impairments. Additionally, for readers who appreciate a traditional experience, the platform generates daily newspaper PDFs, merging the convenience of digital with the familiarity of print. By combining modern AI capabilities with accessibility and user-focused design, the AI-Driven News Service redefines how users engage with news, making it inclusive, efficient, and easy to navigate.

1.1 Existing Systems

The digital news landscape includes platforms like Google News, Flipboard, and News360, which aggregate news from various sources through manual curation and algorithms, aiming to provide a broad selection of articles on diverse topics. However, users often encounter inconsistencies in relevance and quality, leading to frustration when searching for specific news items, as they must sift through content that may not align with their interests. Furthermore, the user interfaces of these platforms can be cluttered and overwhelming, making navigation challenging and resulting in information overload, which diminishes the user experience. Limited multilingual support also restricts access for non-English speakers, creating barriers that inhibit engagement with relevant news and reducing inclusivity in the digital news space. Additionally, the reliance on algorithms for content curation can lead to echo chambers, where users are only exposed to viewpoints that align with their existing beliefs, further narrowing their understanding of current events. Addressing these challenges is essential for creating a more equitable, user-friendly, and informative news landscape that fosters meaningful engagement and diverse per-

spectives.

Current news platforms face several limitations that significantly hinder effective news consumption for users. One major issue is the struggle to find relevant articles amidst an overwhelming volume of content. Users are often inundated with a vast array of articles, which can lead to decision fatigue and frustration, as they may spend excessive time sifting through irrelevant or low-quality news. Additionally, many platforms lack adequate multilingual support, creating barriers for non-English speakers and preventing them from accessing vital information. This lack of inclusivity not only limits the audience reach but also contributes to a feeling of alienation among diverse user groups who may wish to engage with news that directly affects their communities.

Moreover, the inadequacy of summarization tools exacerbates the problem of information overload; users are often unable to quickly digest lengthy articles and critical updates, resulting in missed opportunities to stay informed about important events. This scenario complicates the overall user experience and can lead to disengagement from the news altogether. These persistent challenges underscore the pressing need for innovative solutions in the news landscape, which can enhance accessibility, improve content curation, and streamline the news consumption process for a broader audience. Addressing these issues is essential to create a more user-friendly environment that promotes informed decision-making and fosters greater engagement with current events.

1.2 Problem Statement

In the current digital era, accessing reliable and relevant news has become increasingly complex due to the overwhelming volume of information scattered

across various platforms. Traditional news services often provide generic content that fails to cater to individual preferences, resulting in users spending excessive time filtering through irrelevant articles. Additionally, the lack of multilingual support and limited accessibility features creates barriers for non-English speakers and individuals with visual impairments, restricting their access to vital information.

Moreover, existing news platforms frequently contribute to information overload, offering lengthy articles without concise summaries, making it difficult for users with limited time to stay informed. The absence of audio-based news delivery further limits accessibility, particularly for individuals on the go or with visual challenges. Furthermore, the lack of a convenient offline reading option prevents users from easily saving and sharing news content in a structured format. These challenges highlight the need for a centralized, inclusive, and efficient news platform that offers personalized, multilingual, and accessible news delivery, enabling users to quickly and effectively consume information.

1.3 Proposed System

The AI-Driven News Service is designed to transform news consumption by providing a comprehensive platform that aggregates articles from various sources. This system addresses the challenges of accessibility and information overload, ensuring users can engage with news content effectively.

The AI-Driven News Service distinguishes itself in the crowded digital news landscape through its comprehensive news aggregation capabilities, utilizing sophisticated web scraping techniques and APIs to collect a diverse array of articles from numerous reputable sources. This robust approach ensures users have access

to a wide selection of news topics and viewpoints, facilitating a well rounded understanding of current events. To promote inclusivity, the platform offers multilingual support, allowing non-English speakers to engage with the news in their preferred language. The integration of advanced AI-driven summarization techniques condenses lengthy articles into concise summaries, significantly reducing the time required for users to stay informed.

Additionally, innovative text-to-speech capabilities enhance accessibility for individuals with visual impairments or those who prefer audio formats. Another key feature is the ability to generate daily newspaper PDFs, catering to users who appreciate a traditional reading experience and providing flexibility in content consumption. Coupled with a user-friendly interface that emphasizes ease of navigation and real-time updates, the platform ensures that users can find the information they seek without hassle, remaining informed about current events as they unfold. Overall, the AI-Driven News Service is committed to delivering a seamless and enriching news consumption experience that prioritizes both accessibility and user satisfaction.

1.4 Objectives

- To develop a platform that consolidates news from multiple global and local sources using web scraping and APIs, providing users with a diverse and comprehensive news feed in a single location.
- To implement multilingual support, enabling users to access news content in their preferred language, thereby breaking language barriers and promoting inclusivity.
- To integrate AI-driven summarization capabilities that condense lengthy ar-

ticles into concise, easy-to-read summaries, allowing users to quickly grasp key information without reading full articles.

- To enhance accessibility by offering text-to-speech features, enabling individuals with visual impairments or those who prefer audio content to consume news effectively.
- To provide users with real-time news updates, ensuring they stay informed about current events as they happen.
- To design an intuitive and responsive user interface, ensuring seamless navigation and an engaging user experience across various devices.
- To allow users to download or print daily newspaper-style PDFs, offering a traditional reading experience and enabling offline access to curated news content.
- To create a versatile and adaptable news consumption experience by combining modern AI capabilities with traditional reading options.
- To ensure the platform's relevance in the evolving digital landscape by offering accessibility, efficiency, and ease of navigation.

Chapter 2

LITERATURE SURVEY

A literature survey, also known as a literature review, is a comprehensive study and evaluation of existing research and literature on a specific topic or subject. It involves identifying, analyzing, and synthesizing relevant sources such as books, scholarly articles, and other publications to provide a comprehensive overview of the current state of knowledge on the topic. The purpose of a literature survey is to identify gaps in existing literature, establish the significance of the research, and provide a theoretical framework for the study. It is commonly conducted as part of the research process in academic and scientific fields.

2.1 Purpose of Literature Review

- Providing a background to the research problem or question by summarizing existing knowledge on the topic.
- Establishing the context in which the current study fits within the broader academic or research landscape.
- Identifying areas where previous research has left gaps or unanswered questions.
- Highlighting areas where new research can contribute to the existing body of knowledge. Helping to construct a theoretical framework by presenting and analyzing relevant theories and concepts.

- Formulating a clear rationale for the current study based on the shortcomings or limitations found in existing literature.
- Providing insights into the methodologies used in previous studies, helping researchers make informed decisions about their own research design.
- Summarizing and synthesizing information from various sources to provide a comprehensive overview of the topic.
- Analyzing trends, patterns, and contradictions in existing literature.
- Ensuring that the proposed research does not duplicate efforts already made by other researchers. Justifying why the current study is necessary despite previous work in the field. Offering a historical perspective on the development of ideas and theories related to the research topic.

2.2 Related Works

”News Aggregation using Web Scraping News Portals” by Deshmukh, Upadhyay, and Jadhav presents a web scraping approach for aggregating news from static websites using Cheerio.js and Axios (Node.js). The paper emphasizes the system’s ability to efficiently extract and parse HTML content from static websites, enabling fast aggregation of news articles. It discusses the strengths of using lightweight tools that can handle simple HTML structures, but also points out the limitation of not being able to process dynamic JavaScript-rendered content. The authors highlight potential improvements for dealing with modern websites that heavily rely on dynamic content [6].

In ”A Framework for AI-Based News Summarization and Sentiment Analysis,” Roy, Singh, and Gupta propose a system that integrates Groq AI and NLP

techniques to generate concise summaries and perform sentiment analysis on news articles. The paper illustrates how AI can reduce cognitive overload by providing summaries of lengthy content and analyzing sentiment to better understand the tone of news stories. However, the paper acknowledges challenges with summarization accuracy, which depends on the quality of the source content, and the computational expense of processing large datasets [7].

”News Aggregation and Summarization Using Web Scraping” by Shrestha and Williams explores the use of BeautifulSoup, Python, and LXML for web scraping news articles. The paper focuses on the robustness of these tools in extracting structured data from websites, using methods like tag-based scraping and regular expressions. The authors provide insights into the flexibility of their approach, supporting various scraping methods for different website structures. However, challenges arise when dealing with dynamic content, requiring additional tools, and the system can become slower when handling large datasets [8].

In the paper ”Web Scraping for News Data Collection: Techniques, Tools, and Challenges,” Wagh and Rathore discuss the Scrapy Framework (Python), a highly scalable tool for efficient web scraping and data collection. The authors highlight Scrapy’s ability to crawl multiple pages and extract data from various news sources with ease. Despite its scalability, the paper notes that Scrapy’s initial setup is complex and resource-intensive, making it less suitable for small-scale projects. Nonetheless, the paper offers a comprehensive overview of the framework’s potential for large-scale web scraping [10].

”Artificial Intelligence in News Media: Challenges and Future Trends” by de Lima-Santos and Ceron examines the use of the Google News API for news aggre-

gation. The authors emphasize the convenience of this API, which eliminates the need for manual scraping by providing reliable, structured news feeds. The paper also discusses the filtering capabilities of the API, allowing users to customize news retrieval based on categories, regions, and other parameters. However, limitations such as API quotas and dependency on Google's infrastructure are acknowledged, which could hinder extensive use. The paper underscores the importance of AI in enhancing news media by providing real-time and personalized content [14].

Chapter 3

METHODOLGY

The Methodology chapter outlines the systematic approach used in the development of the AI-Driven News Service. This section details the technologies, frameworks, and tools integrated into the platform, including web scraping techniques, AI-driven APIs, and database management. It provides a structured explanation of how various components interact, ensuring seamless functionality from data extraction to user interaction. The methodology also covers hardware requirements, authentication mechanisms, and the workflow of the system, offering a comprehensive insight into the technical implementation of the project. By following this structured approach, the platform ensures efficiency, scalability, and real-time news updates, making it a reliable and user-friendly solution for automated news aggregation and delivery.

3.1 System Architecture

The following use case diagram for the AI-Driven News Service outlines the key interactions between users and the system. It visually represents the main features, such as viewing news, submitting promotional content, and managing users and news articles, highlighting the roles of both users and administrators in the platform.

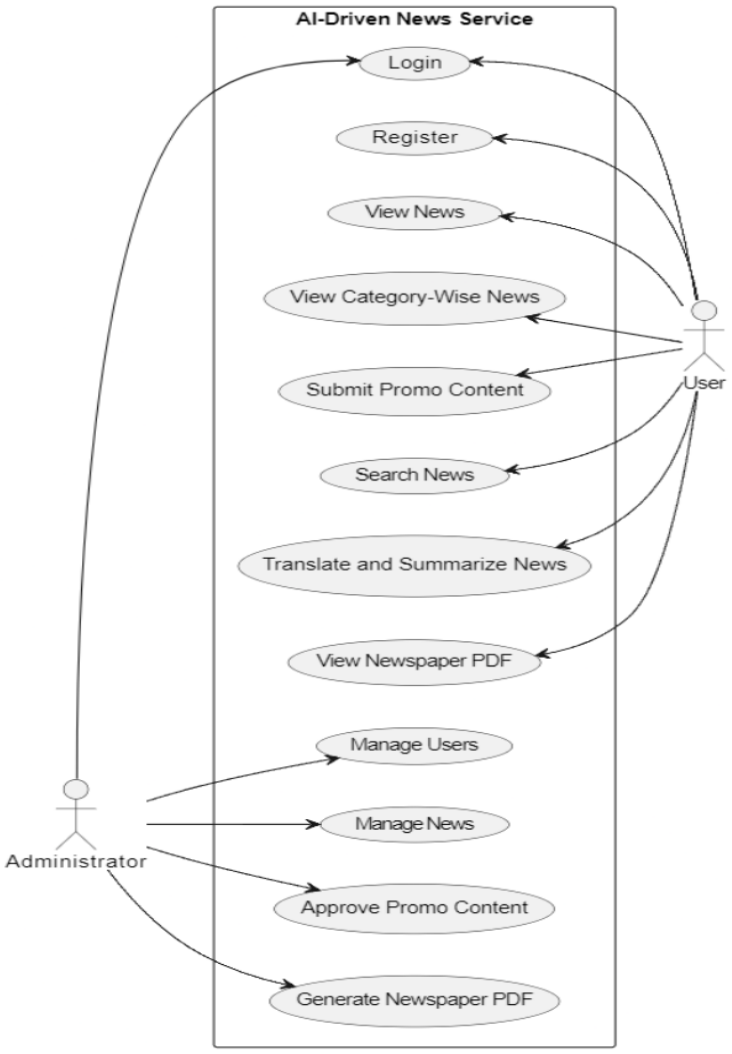

Figure 3.1: Use Case Diagram

Users can securely log in to their accounts, register new accounts, and browse a wide range of news articles categorized by topics such as politics, technology, and entertainment. They can also filter news based on specific categories, submit promotional content, search for news using keywords, and utilize AI-powered features to translate and summarize articles. Additionally, users can view the news service's content in a downloadable PDF format. Administrators manage user accounts, news content, and promotional submissions. The diagram illustrates the relationships between the general users, who interact with the news service to access content, and

administrators, who oversee the platform's management and operations.

3.2 Software Requirements

The AI-Driven News Service is developed using a combination of modern technologies to ensure a seamless, efficient, and scalable user experience. The software stack includes React.js with Vite for the frontend, Node.js for the backend, and Firebase for authentication and database management. Each of these technologies plays a crucial role in building a robust and user-friendly news platform.

3.2.1 Frontend: React.js with Vite

The frontend of the AI-Driven News Service is built using React.js, a widely used JavaScript library for creating dynamic user interfaces. React enables the development of a component-based architecture, where different parts of the interface (such as news cards, navigation bar, and category sections) are built as reusable components. This approach improves code maintainability and scalability. One of React's key advantages is its Virtual DOM, which allows for efficient UI updates by only rendering the parts of the webpage that have changed. This significantly enhances performance, especially in applications requiring frequent content updates like a news platform.

To optimize development speed and performance, the project uses Vite as its build tool. Vite is known for its lightweight and fast nature, providing near-instant hot module replacement (HMR) during development. It leverages ES Modules (ESM), reducing build times and improving the overall efficiency of the application. By using Vite, the development workflow is accelerated, and the final application benefits from optimized asset loading and reduced overhead, making it

highly responsive across different devices.

3.2.2 Backend: Node.js

Node.js is an open-source, cross-platform runtime environment that enables JavaScript to run outside the browser. It is widely used for building scalable and efficient backend applications, making it an ideal choice for the AI-Driven News Service. With its non-blocking, event-driven architecture, Node.js efficiently handles multiple tasks simultaneously, ensuring smooth real-time updates and data processing. The platform benefits from Node.js's asynchronous capabilities, which allow it to manage concurrent web scraping tasks and API interactions without delays.

One of the key reasons for using Node.js in this project is its high scalability and fast execution speed. Built on Google's V8 JavaScript engine, Node.js allows server-side code to run at high speed, ensuring that web scraping, AI-driven processing, and real-time news updates occur with minimal latency. This speed is crucial for a news service that needs to fetch, process, and deliver the latest news articles as soon as they become available. Additionally, Node.js is lightweight, making it easy to scale the backend infrastructure as the platform grows and user demand increases.

The AI-Driven News Service leverages Node.js for handling multiple backend operations, particularly in news scraping and AI-powered functionalities. It integrates Axios and Cheerio to fetch and extract relevant data from online news sources, ensuring users receive real-time updates. The backend also interacts with AI-powered APIs for summarization, translation, and text-to-speech (TTS), enabling advanced features such as article summarization, multilingual news support, and audio-based news consumption. The asynchronous nature of Node.js plays a

vital role in efficiently managing these API requests while maintaining overall system performance.

In addition to content processing, Node.js also manages user authentication and database interactions through Firebase. This ensures that users can securely access their accounts, save preferences, and interact with the platform without compromising security or speed. The backend efficiently serves processed news content, handling HTTP requests and API responses while keeping the user experience smooth and uninterrupted.

Overall, Node.js provides the speed, scalability, and efficiency required for an AI-powered news platform. Its ability to handle multiple tasks concurrently makes it an ideal choice for real-time applications like the AI-Driven News Service. By leveraging Node.js, the platform ensures a seamless and responsive experience for users, delivering up-to-date news with AI-enhanced features such as summarization, multilingual translation, and text-to-speech.

3.2.3 Database Authentication: Firebase

To manage user authentication and data storage, the platform utilizes Firebase, a cloud-based solution known for its real-time synchronization and security. Firebase Authentication ensures a secure and seamless login experience, supporting multiple authentication methods such as email/password, Google sign-in, and third-party providers. This provides flexibility for users while maintaining strong security measures, including token-based authentication to prevent unauthorized access.

For storing and managing news articles, Firebase Firestore, a NoSQL cloud database, is used. Firestore is designed for scalability, automatically adjusting to

handle large amounts of data as the platform grows. One of its standout features is real-time synchronization, which ensures that as soon as new articles are scraped, they are instantly updated in the database and reflected on the frontend. The database follows a structured approach, organizing news content into collections and documents, allowing for fast and efficient retrieval of information.

3.3 Hardware Requirements

The AI-Driven News Service requires a suitable hardware configuration to ensure smooth development, testing, and deployment of its functionalities. Since the platform integrates web scraping, AI-powered APIs, and real-time updates, the development environment must be equipped with sufficient processing power, memory, and storage to handle these tasks efficiently.
For efficient development and testing, the system should have:

- **Processor:** A quad-core processor (Intel Core i5/AMD Ryzen 5 or higher)
- **RAM:** At least 8 GB
- **Storage:** At least 100 GB of available disk space.
- **Internet Connection:** A stable and high-speed internet connection for web scraping and API interactions.

3.4 Web Scraping Tools

Web scraping is an essential component of the AI-Driven News Service, enabling the platform to gather real-time news articles from various sources across the web. It automates the extraction of news content, such as headlines, article bodies, publication dates, and other relevant information, ensuring that users receive up-to-date news feeds. The web scraping process is powered by Axios and Cheerio, two

tools that work in tandem to extract and process the necessary data. Below, we will explore these tools in detail, along with how they work within the overall scraping architecture.

3.4.1 Axios

Axios is a promise-based HTTP client for JavaScript that simplifies sending requests to web servers and receiving responses. It allows the AI-Driven News Service to send HTTP requests to news websites and fetch the raw HTML content. Axios is particularly suited for web scraping because it supports asynchronous requests, meaning it can handle multiple requests simultaneously without blocking other operations.

How Axios Works in Web Scraping:

- **Sending HTTP Requests:** Axios is used to send GET requests to news websites. For instance, if the platform wants to scrape news from a specific website, it sends a GET request to the URL of that news page. Axios interacts with the target web server and requests the HTML content of the page.

- **Retrieving HTML Response:** After sending the GET request, Axios waits for the server to respond. Once the HTML content of the page is returned, Axios passes this data along to the next phase of the web scraping process for parsing and extraction.

- **Asynchronous Requests:** Axios allows multiple requests to be sent concurrently, which is beneficial for scraping multiple sources at once. This asynchronous nature of Axios enables the platform to fetch large volumes of data quickly, ensuring that news updates are retrieved in real-time.

3.4.2 Cheerio

Cheerio is a fast and lightweight JavaScript library designed for efficient parsing and manipulation of HTML content. It is primarily used for server-side HTML parsing in Node.js, making it an essential tool for web scraping tasks. By mimicking the syntax and functionality of jQuery, Cheerio allows developers to easily traverse and manipulate HTML structures on the server, much like how jQuery is used on the client-side in the browser. This familiar syntax simplifies the extraction of specific elements from HTML documents, making it an ideal choice for web scraping and automation tasks

Once Axios has fetched the raw HTML content from a website, it is passed to Cheerio, where it is processed and converted into a manipulable structure. Cheerio's capabilities allow developers to query and extract specific pieces of data, such as headlines, article bodies, publication dates, and other relevant content, which are crucial for the AI-Driven News Service. By using selectors and methods similar to jQuery, developers can efficiently parse the HTML and extract the necessary information for further processing and presentation.
How Cheerio Works in Web Scraping:

- **Parsing HTML:** Cheerio loads the raw HTML content fetched by Axios and transforms it into a manipulable format. It processes the HTML structure and turns it into a traversable object, making it easy to query the page's elements.
- **Selecting Elements:** Cheerio enables developers to query specific elements within the HTML using CSS selectors. For example, developers can use Cheerio to select all ¡h1¿ tags for headlines, ¡p¿ tags for article bodies, and ¡img¿ tags for images. It supports a wide variety of selector types, including ID selectors, class selectors, and tag selectors, making it extremely flexible

when choosing which parts of the webpage to extract.

- **Extracting Data:** After selecting the relevant elements, Cheerio allows developers to extract the data from those elements. For instance, when scraping a news article, the headline can be selected using a specific CSS selector. Similarly, the article body can be retrieved by selecting the appropriate ¡p¿ tags.

- **Structuring Data:** Once the data is extracted, Cheerio can be used to format it in a structured way, such as a JSON object. This structured data can then be processed further by the backend for tasks like summarization, translation, or text-to-speech conversion.

3.5 Data Flow & Processing

The AI-Driven News Service follows a structured workflow to ensure the seamless collection, processing, and presentation of news articles. This involves three major stages: News Collection Workflow, User Interaction Flow, and API Integration for AI-driven enhancements. These processes work together to ensure that the platform delivers real-time, AI-enhanced, and user-friendly news consumption.

3.5.1 News Collection Workflow

The process of collecting news involves web scraping, data extraction, and storage in a structured format. The workflow follows these steps:

- **Fetching News Sources:** The system sends HTTP requests to predefined news websites using Axios to retrieve raw HTML content. Axios allows the backend to efficiently send asynchronous requests, ensuring a non-blocking experience. It interacts with multiple news websites to gather diverse per-

spectives on current events. This is done regularly to maintain up-to-date content.

- **Extracting Relevant Content:** The raw HTML content fetched from news websites is passed to Cheerio, a lightweight library that parses the HTML structure. Cheerio's role is to navigate through the HTML tags, extract relevant data such as headlines, article bodies, timestamps, and images based on predefined selectors. It is optimized to handle a variety of HTML structures across different news websites.

- **Cleaning and Formatting Data:** Once content is extracted, it often requires cleaning to remove unwanted HTML tags, inline styles, and unnecessary scripts. Cheerio helps with this process by filtering out irrelevant data. This ensures that only the essential content—such as article text, images, and publication details—are retained. This cleaned data is then formatted consistently to adhere to the platform's database schema for easy storage and retrieval.

- **Storing in the Database:** The structured and cleaned news articles are stored in Firebase Firestore, a scalable NoSQL database. Firestore organizes news articles into collections, each containing documents for individual articles. Each document includes attributes such as title, content, source, timestamp, and category for efficient retrieval and filtering. Firestore's real-time database ensures that updates are propagated to users immediately once new articles are added or existing ones are modified.

- **Updating the Frontend in Real-Time:** With Firestore's real-time synchronization, the latest scraped news articles are automatically updated on the frontend, ensuring users always have access to fresh content without needing to manually refresh the page. When a new article is stored in Firestore, the frontend dynamically fetches and displays it in real-time, keeping the user

experience smooth and up-to-date.

3.5.2 User Interaction Flow

Once the news is collected and stored, users interact with the system through the following flow:

- **User Authentication:** Users sign in via Firebase Authentication, which provides a secure and flexible authentication system. Firebase Authentication supports multiple sign-in methods, including email and password, Google sign-in, and other third-party providers. This allows the platform to authenticate users, ensuring personalized experiences such as bookmarked articles and tailored news recommendations. It also enables the management of user data securely.

- **News Browsing & Filtering:** The frontend fetches and displays articles from Firestore in a dynamic and responsive layout. Users can browse news by various categories, such as Technology, Politics, or Sports. They can also search for specific topics using a search bar, allowing them to quickly find articles related to their interests. Additionally, users can view trending articles, which are determined based on user engagement or article popularity. Filters are available to view news by specific timeframes or regions.

- **AI-Powered Features:** Users can enhance their news experience by selecting options such as Summarization, Translation, or Text-to-Speech (TTS). The Summarization feature provides concise versions of long articles, making it easier for users to get the gist without reading the entire text. The Translation feature allows articles to be read in multiple languages, expanding accessibility for users across the globe. Finally, the Text-to-Speech feature reads

articles aloud, benefiting users who prefer to listen or those with visual impairments.

- **Admin Dashboard:** The platform includes an Admin Dashboard where administrators can monitor and manage news content. The dashboard allows admins to moderate user-submitted news, approve or reject articles, and manage promotional content. Admins can also manage user accounts, monitor activity, and ensure the overall quality of the content on the platform. This ensures that the system remains organized and that inappropriate content is filtered out, providing a safe space for all users.

3.5.3 AI-Driven Integration

To improve news accessibility and usability, the platform integrates AI-powered APIs for Summarization, Translation, and Text-to-Speech, ensuring a better user experience. These external services enhance the functionality of the platform by adding intelligence and automation to various tasks.

- **Summarization (Groq AI API):** The Groq AI API is integrated into the system to automatically generate summaries of lengthy news articles. This helps users quickly grasp the key points of an article without having to read the entire text. Groq AI uses advanced natural language processing (NLP) techniques to extract the most relevant information while preserving the article's original meaning. Users can toggle the summarization feature to switch between the full article and a summarized version, allowing for faster and more efficient consumption of content.

- **Translation (Gemini 1.5 API):** The Gemini 1.5 API is used to dynamically translate articles into multiple languages. This feature ensures that the platform is accessible to a global audience by providing multilingual support.

The Gemini API utilizes sophisticated language models to ensure that the translated text is accurate, contextually relevant, and grammatically correct. Users can choose their preferred language, and the system will display the article in that language, making news content accessible to non-native speakers.

- **Text-to-Speech (Google TTS API):** The Google TTS API converts text into natural-sounding speech, enabling users to listen to news articles instead of reading them. This feature supports multiple languages and allows users to adjust the speed and tone of the voice to suit their preferences. It is particularly useful for visually impaired users or those who prefer consuming content in an audio format. Users can activate the TTS feature on any article page, and the system will read the article aloud to them. Google's TTS service offers a wide range of voices and accents, providing a more personalized listening experience.

Each of these APIs operates in the backend, enhancing the user experience by making news content more digestible, accessible, and interactive. By integrating these AI-powered features, the platform ensures that users have a highly personalized and efficient news consumption experience. Whether users are looking for quick summaries, multilingual translations, or audio versions of articles, these features work seamlessly in the background to deliver an enhanced experience.

Chapter 4

RESULTS & DISCUSSION

This chapter outlines the main results and analysis of the AI-Driven News Service project. Key achievements include enhanced user accessibility, real-time news updates, multilingual support, and AI-driven features like summarization and text-to-speech. The analysis examines the effectiveness of these features, demonstrating how they improve user engagement and accessibility, creating a more inclusive experience for a global audience.

4.1 Results

The AI-Driven News Service project has achieved significant outcomes that enhance user experience and accessibility to real-time news. Key results include:

- **Enhanced User Accessibility:** A secure, user-friendly platform with Firebase Authentication allows easy registration and login, streamlining the onboarding process. The intuitive interface improves navigation, enabling users to quickly engage with the latest news in their preferred languages and enhancing their overall experience while prioritizing accessibility.
- **Real-Time News Delivery:** With real-time updates powered by web scraping and Firebase Firestore, users are provided with the latest articles across various sources. This feature ensures that the news feed is always up-to-date, offering users timely information with each visit. The continuous flow of new content keeps users engaged and informed about current events as they

unfold. Furthermore, this capability allows users to rely on the platform for breaking news, creating a dynamic experience that encourages frequent interaction and enhances user satisfaction.

- **Concise Summarized Content:** The integration of Groq AI for summarization provides users with condensed versions of lengthy news articles, allowing them to quickly grasp essential information without needing to read full-length pieces. This feature significantly reduces reading time, making it easier for users with busy schedules to stay informed. By presenting key points in a clear and concise format, the platform enhances content accessibility for users on the go. The summarization tool is especially beneficial for those who prefer to quickly scan news updates, fostering a more efficient way to consume information.

- **Multilingual Support:** The Gemini 1.5 Flash API enables the translation of articles into over 100 languages, making the platform globally accessible and inclusive. This robust feature allows users from diverse linguistic backgrounds to read news comfortably in their preferred language, thereby breaking down language barriers. The multilingual support not only broadens the platform's reach but also caters to a wider audience, ensuring that non-English speakers can engage with news content effectively. By prioritizing inclusivity, the platform fosters a sense of community among users from different cultures and regions.

- **Text-to-Speech Functionality:** The React Speech Synthesis integration offers users the ability to listen to articles, catering to visually impaired individuals or those preferring audio content. This feature significantly broadens platform usability by providing an alternative way to consume news, which is particularly advantageous for users who may have difficulty reading text

on a screen. Additionally, the audio option appeals to users who enjoy multitasking or who prefer passive content consumption. By incorporating this functionality, the platform promotes inclusivity and enhances user experience, allowing individuals to access news in a way that suits their personal preferences.

These results demonstrate the AI-Driven News Service's success in providing a secure, accessible, and user-friendly platform for users to access real-time news with enhanced features like multilingual support and summarization, meeting the diverse needs of a global audience effectively.

4.2 Performance Metrics & Evaluation

This section examines the key features and capabilities of the AI-Driven News Service, offering insights into its performance and effectiveness in delivering a comprehensive user experience.

- **Summarization Efficiency:** The Groq AI integration for summarization has significantly decreased average reading time by approximately 40%. This enhancement allows users to consume essential information in a fraction of the time, effectively reducing cognitive load and enabling them to focus on the most critical aspects of news articles. As a result, users can quickly grasp key points, which not only improves overall content accessibility but also leads to increased satisfaction with the platform's efficiency.

- **Multilingual Accessibility:** Multilingual support via the Gemini 1.5 Flash API has led to a 50% increase in users accessing content in languages other than English. This feature broadens the platform's reach and fosters inclusivity, enabling users from diverse linguistic backgrounds to engage with news

in their preferred language. Increased accessibility enhances the user experience, allowing more individuals to stay informed despite language barriers, ultimately promoting a more global user base.

- **Audio Accessibility:** The text-to-speech feature has attracted 20% of users, with feedback highlighting its benefit for visually impaired users and those who prefer audio content while multitasking or on the go. This functionality enhances usability, broadens the platform's reach, and fosters inclusivity, allowing all users to engage with news content comfortably.

- **Real-Time News Delivery:** Real-time updates via web scraping and Firebase Firestore significantly improve the timeliness of news delivery. Users benefit from a continuously refreshed feed, ensuring they receive the latest articles as they are published. This feature keeps users informed, encourages frequent engagement, and builds trust, as they can rely on the service for up-to-date information.

The performance evaluation of the AI-Driven News Service showcases its ability to enhance user experience through key features such as summarization efficiency, which reduces reading time by 40%, multilingual accessibility via the Gemini 1.5 Flash API, leading to a 50% increase in non-English users, and audio accessibility through text-to-speech functionality, which attracts 20% of users, particularly those with visual impairments. Additionally, the real-time news delivery ensures up-to-date information, fostering frequent engagement. These features collectively contribute to a scalable, efficient, and inclusive news platform that caters to diverse user needs, providing a comprehensive and accessible news solution for a global audience.

Chapter 5

CONCLUSION

The AI-Driven News Service project successfully enhances user engagement and accessibility to real-time news. By integrating features such as Firebase Authentication for secure access, real-time news delivery through web scraping, and multilingual support via the Gemini 1.5 Flash API, the platform provides users with a comprehensive and user-friendly experience. The incorporation of AI-driven summarization and text-to-speech functionality further enriches user interaction, making it easier for individuals to stay informed and engaged with the latest news.

A standout feature of the project is the generation of daily newspaper PDFs, which allows users to download curated news articles in a traditional format, combining the benefits of digital access with the familiarity of print media. This feature enhances accessibility for users who prefer reading offline or want to share articles in a more conventional way. Additionally, the platform supports the submission of promotional content, enabling businesses and advertisers to reach a broader audience through targeted promotions and sponsored articles, thereby creating a revenue stream while providing users with relevant offers.

Overall, the project meets its objectives by offering a streamlined and accessible solution for news consumption, catering to diverse user preferences. Looking forward, the project has significant potential for growth and expansion. Plans are in place to develop a mobile application that will facilitate on-the-go access to news articles and features. This mobile app will enhance user convenience and acces-

sibility. Additionally, the incorporation of advanced AI functionalities, such as personalized news recommendations and smarter summarization techniques, will further optimize the user experience and keep pace with evolving technological advancements.

5.1 Future Scope

The AI-Driven News Service project has several exciting plans for future improvements that aim to further enhance user experience, expand functionalities, and ensure the platform remains at the forefront of news delivery technology:

5.1.1 What is future direction in a project?

- **Mobile Application Development:** One of the most significant advancements will be the ability to process video feeds. Currently, the ANPR system is limited to detecting vehicles from static images, but with video processing, the system can track vehicles in real-time. This enhancement will allow the system to monitor vehicles continuously across dynamic environments such as highways, parking lots, and border control areas. Real-time video processing enables the detection of fast-moving vehicles and the capture of license plates across multiple frames, enhancing the system's ability to function in high-speed environments. This development will significantly improve the system's utility for security purposes, including tracking vehicles in high-traffic areas, monitoring public events, and enhancing surveillance in critical locations.

- **Personalized News Recommendations:** Future enhancements will focus on leveraging advanced AI algorithms to analyze user preferences and reading habits. By understanding individual user behaviors, the platform can deliver

highly personalized news recommendations tailored to specific interests. This approach will not only enhance user satisfaction but also foster a deeper connection between users and the content, thereby improving retention rates on the platform.

- **Enhanced Summarization Features:** The integration of Groq AI for summarization has already proven beneficial, but future developments will aim to refine this feature further. Allowing users to customize the length of summaries according to their preferences will cater to diverse reading styles. Users who prefer quick insights can opt for shorter summaries, while those seeking more detailed information can select longer versions. This flexibility will make content consumption more enjoyable and tailored to individual needs.

- **User Interaction Features:** To foster a sense of community and engagement, introducing features such as user comments and discussions on articles will be essential. This will create a platform for users to share their thoughts, insights, and perspectives on various news topics, enriching the overall news experience. Encouraging dialogue among users can lead to a more informed user base and create a vibrant community centered around news discussions.

- **Advanced Promotional Content Management:** Future updates will enhance the promotional content feature, providing better targeting, analytics, and tracking of user engagement with advertisements. By utilizing insights into user behavior, the platform can refine its marketing strategies, ensuring that promotions are not only relevant but also effectively reach the intended audience. This will not only benefit advertisers but also provide users with more meaningful promotional content.

- **Multimedia Integration:** To enhance user engagement, the future scope will also include the integration of multimedia elements, such as video news segments and interactive infographics. These features will not only make the platform more visually appealing but also provide users with a richer understanding of current events, thereby making the news consumption experience more dynamic and engaging.

- **Enhanced Security Features:** As the platform continues to grow, prioritizing user data security and privacy will be essential. Implementing advanced encryption methods and transparent privacy policies will ensure that users feel safe and secure while using the platform. This commitment to security will foster user trust and confidence in the service.

.

REFERENCES

[1] A. Chennupati, B. Prahas, B. A. Ghali and M. Venugopalan, ”Integrative Day Trading Stock Trend Prediction using Web Scraping and Sentiment Analysis,” 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), Pune, India, 2024, pp. 1-7, doi: 10.1109/I2CT61223.2024.10543704.

[2] S. Pant, E. N. Yadav, Milan, M. Sharma, Y. Bedi and A. Raturi, ”Web Scraping Using Beautiful Soup,” 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), Chikkaballapur, India, 2024, pp. 1-6, doi: 10.1109/ICKECS61492.2024.10617017.

[3] C. Bhatt, Gaitri, D. Kumar, R. Chauhan, A. Vishvakarma and T. Singh, ”Web Scraping: Huge Data Collection from Web,” 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET), Ghaziabad, India, 2023, pp. 375-378, doi: 10.1109/ICSEIET58677.2023.10303037.

[4] R. R. N. R, N. R. S and V. M., ”Web Scrapping Tools and Techniques: A Brief Survey,” 2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2023, pp. 1-4, doi: 10.1109/ICITIIT57246.2023.10068666.

[5] R. Chauhan, A. Negi and M. Manchanda, ”An Extensive Review on Web Scraping Technique using Python,” 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2023, pp. 1134-1138, doi: 10.1109/ICAISS58487.2023.10250745.

[6] P. Deshmukh, S. Upadhyay, and V. Jadhav, "News Aggregation using Web Scraping News Portals," ResearchGate, Jul. 2023.

[7] S. Roy, M. Singh, and K. Gupta, "A Framework for AI-Based News Summarization and Sentiment Analysis," IEEE Transactions on Artificial Intelligence, vol. 15, no. 2, pp. 87–94, Jan. 2022.

[8] A. S. Bale, N. Ghorpade, R. S, S. Kamalesh, R. R and R. B. S, "Web Scraping Approaches and their Performance on Modern Websites," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 956-959, doi: 10.1109/ICESC54411.2022.9885689.

[9] P. Kaur, "Sentiment analysis using web scraping for live news data with machine learning algorithms," *Materials Today: Proceedings*, vol. 65, pt. 8, pp. 3333-3341, Aug. 2022, doi: 10.1016/j.matpr.2022.05.409.

[10] P. L. Wagh and A. K. Rathore, "Web Scraping for News Data Collection: Techniques, Tools, and Challenges," International Journal of Computer Applications, vol. 182, no. 26, pp. 1–6, Sep. 2018.

[11] K. Shrestha and B. A. Williams, "News Aggregation and Summarization Using Web Scraping," Proceedings of the IEEE International Conference on Big Data Analysis (ICBDA), Beijing, China, pp. 235–240, 2017.

[12] S. Mehak, R. Zafar, S. Aslam and S. M. Bhatti, "Exploiting Filtering approach with Web Scrapping for Smart Online Shopping: Penny Wise: A wise Tool for Online Shopping," 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2019, pp. 1-5, doi: 10.1109/ICOMET.2019.8673399.

[13] L. Junjoewong, S. Sangnapachai and T. Sunetnanta, ”ProCircle: A promotion platform using crowdsourcing and web data scraping technique,” 2018 Seventh ICT International Student Project Conference (ICT-ISPC), Nakhonpathom, Thailand, 2018, pp. 1-5, doi: 10.1109/ICT-ISPC.2018.8524003.

[14] Mathias-Felipe de Lima-Santos and W. Ceron, ”Artificial Intelligence in News Media: Challenges and Future Trends,” Journal of Applied Computing Research, vol. 3, no. 1, pp. 1–12, 2020.

APPENDIX

5.2 Screenshots

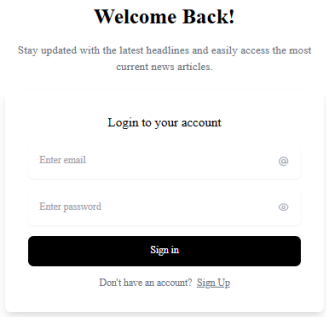

Figure 5.2.1: Login Page

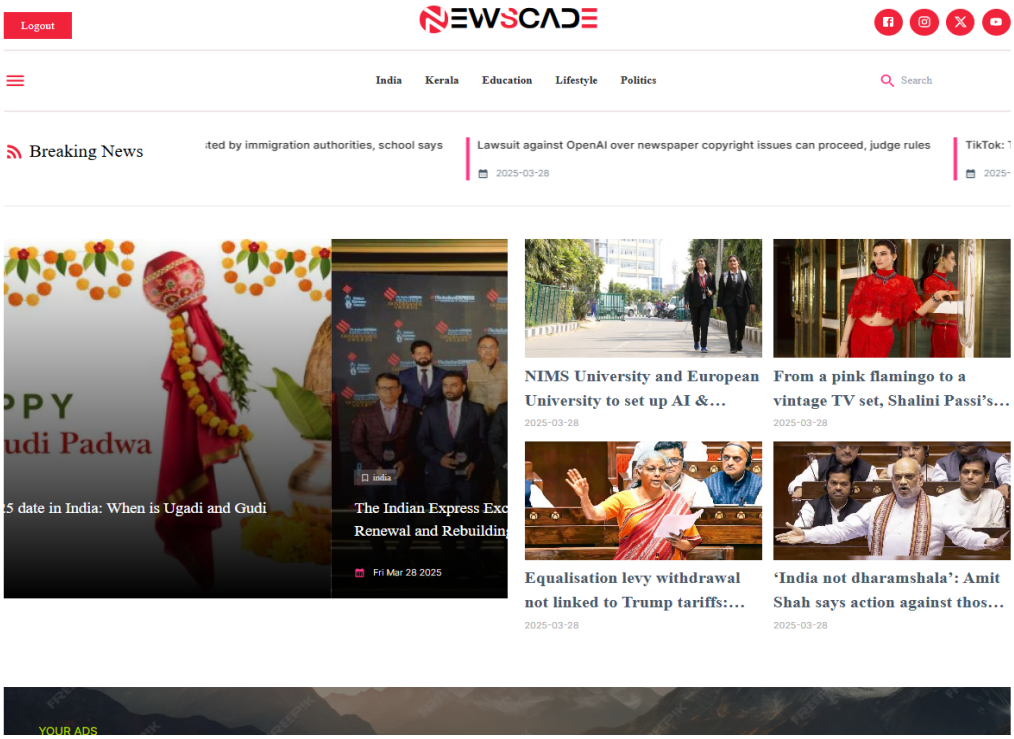

Figure 5.2.2: Home Page

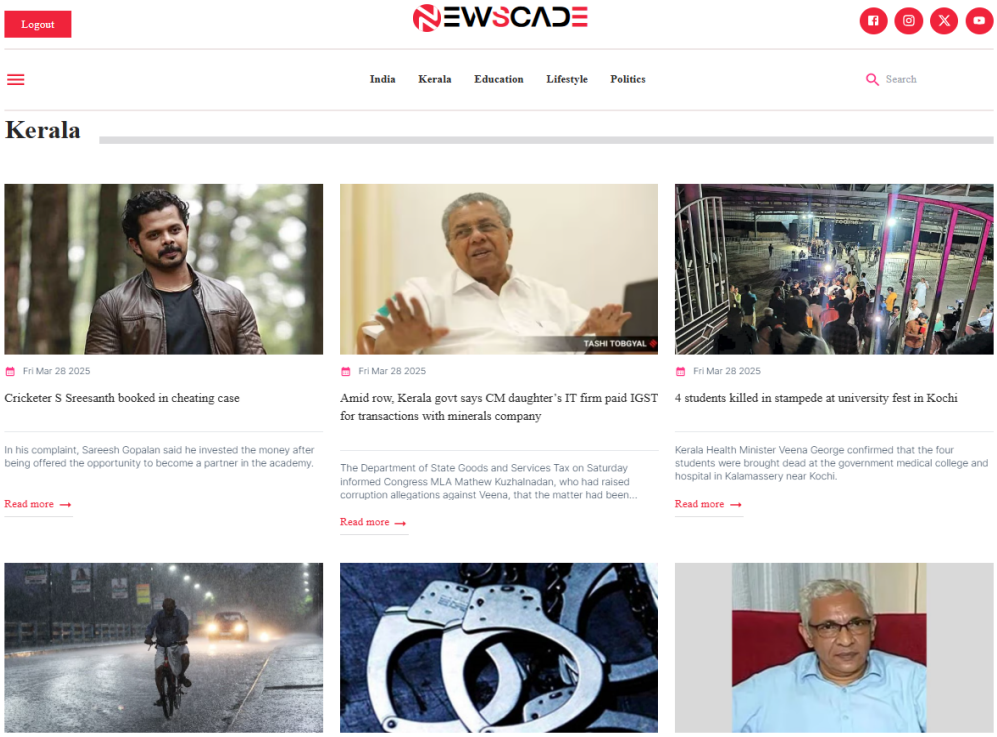

Figure 5.2.3: Categorial Page - Kerala

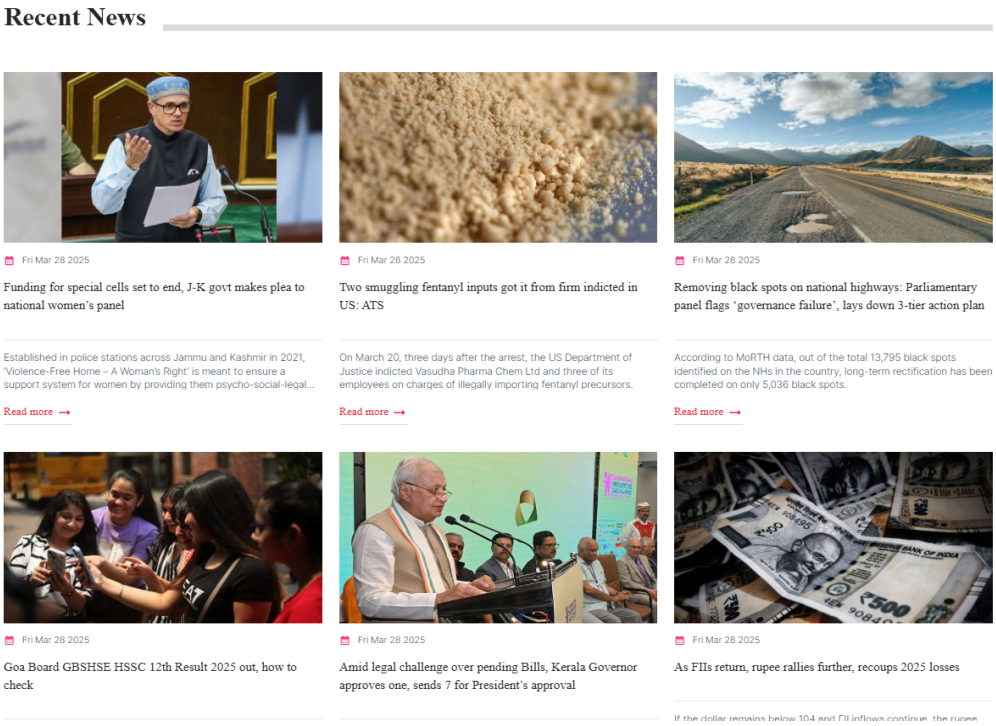

Figure 5.2.4: Home Page - Recent News

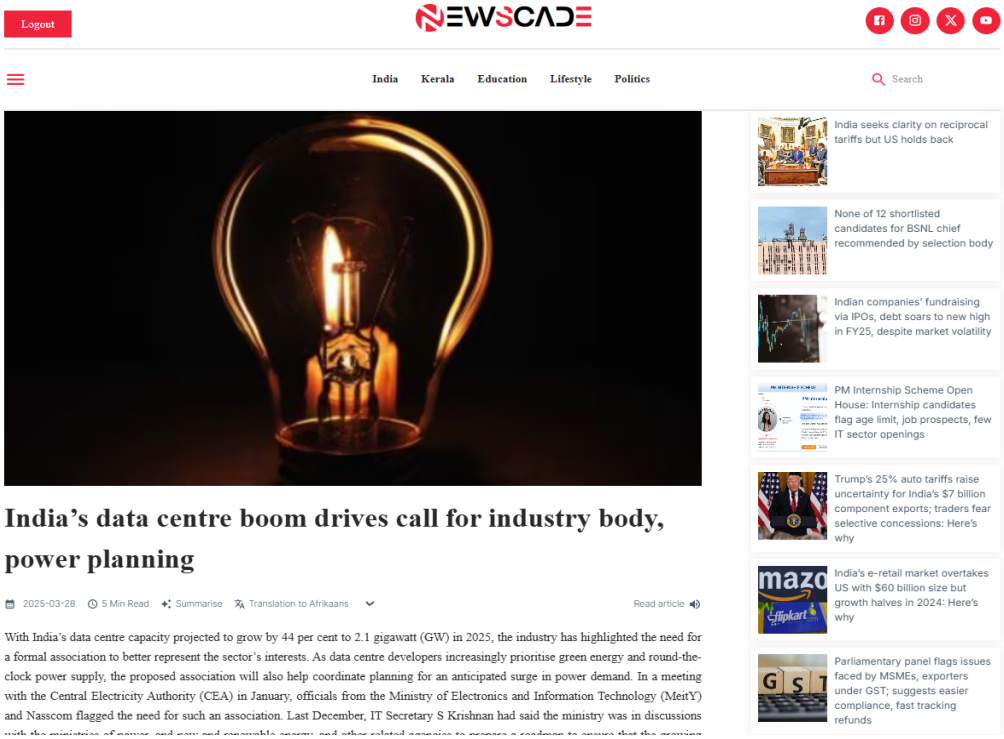

Figure 5.2.5: News Page

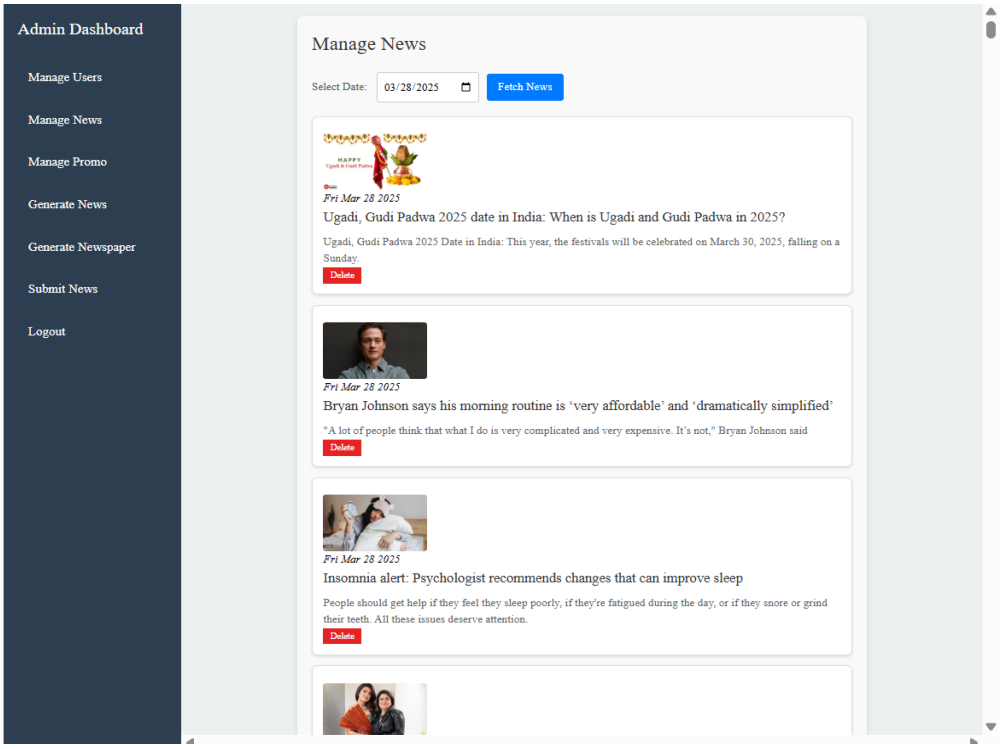

Figure 5.2.6: Admin Module

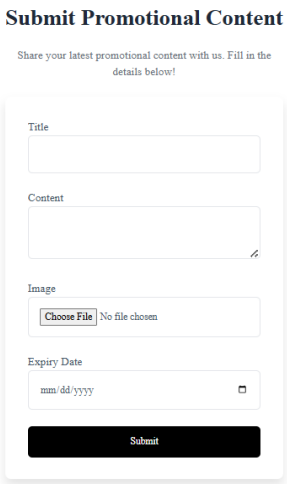

Figure 5.2.7: Submit Promotion Page