

ACKNOWLEDGEMENT

I would like to express my heartfelt appreciation to U. V. Patel College of Engineering, Ganpat University, Kherva, for providing me with an opportunity to explore and innovate my ideas by undertaking a project on News Hub in which news articles are ingested and processed with NLP. ML algorithms analyze user behavior and preferences. Articles are summarized using NLP. Users receive personalized news recommendations. The app delivers concise, tailored news articles, enhancing user engagement and satisfaction. This project has helped me to broaden my knowledge and skills in this field, which will be useful for my future endeavours.

I would also like to extend my sincere gratitude to **Prof. Hiten Sadani**, a faculty member at Ganpat University, for his invaluable guidance, support, and motivation throughout the project. He provided me with insightful suggestions and constructive feedback, which helped me to enhance the quality of my work. I am grateful to all the competitive students who were part of this project and were always willing to assist and collaborate with me. Their camaraderie and enthusiasm made this project a remarkable learning experience. Thank you all for your invaluable contributions to this project. Without your support, this project would not have been possible.

With Sincere Regards From,

MR. TIRTH SHAH

MS. KRUTI SHAH

ABSTRACT

In an era dominated by an abundance of online news sources, the challenge of navigating through an information-saturated landscape is ever-present. This report introduces the Duke Dunes News Hub, a sophisticated web application designed to revolutionize the way users consume news. Developed using Streamlit and powered by Firebase, Duke Dunes serves as a centralized platform, offering a streamlined and user-friendly interface for accessing the latest news from a multitude of sources. By prioritizing simplicity and clarity, Duke Dunes eliminates distractions inherent in many news platforms, ensuring an ad-free and privacy-respecting experience without the necessity for user logins.

One of Duke Dunes' key innovations lies in its integration of Natural Language Processing (NLP). This feature allows the application to distil lengthy news articles into concise and easily digestible summaries, addressing the common user pain point of information overload. The NLP-based summarization not only enhances efficiency but also provides users with a quick understanding of the core elements of a news story. The report comprehensively outlines the development process, the integration of NLP technologies, and the utilization of Firebase for efficient data storage.

Duke Dunes stands as a testament to the commitment to user-centric design, offering a solution that values simplicity, relevance, and privacy. This report delves into the intricacies of Duke Dunes, providing insights into its features, functionalities, and the broader implications it holds for the future of online news consumption.

INDEX

1. INTRODUCTION.....	1
1.1 Project Purpose	1
1.2 Project Statement.....	1
1.3 Objective	2
2. PROJECT SCOPE.....	3
2.1 Key Objectives:.....	3
2.2 Out of Scope:	3
2.3 Constraints:	3
3. FEASIBILITY ANALYSIS	5
3.1 Technical Feasibility:	5
3.2 Time Schedule Feasibility:.....	5
3.3 Operational Feasibility:	5
3.4 Implementation Feasibility:.....	5
3.5 Economic Feasibility:.....	5
4. SOFTWARE AND HARDWARE REQUIREMENTS.....	6
4.1 Software Requirements:	6
4.2 Hardware Requirements:.....	6
4.3 Client-Side Requirements:	6
4.4 Functional and Non-Functional Requirements.....	7
4.4.1 Functional Requirements:.....	7
4.4.2 Non-Functional Requirements:	7
5. PROCESS MODEL	8
5.1 Summarizing the articles	8
6. PROJECT PLAN.....	10
6.1 Gantt Chart of entire Project Plan with duration, start date and end date:	10
7. SYSTEM DESIGN.....	11
7.1 Use case diagram.....	11
7.2 Activity Diagram for User.....	12
7.3 Class Diagram	13
7.4 Sequence Diagram.....	14
7.5 Data Flow Diagram	15
8. IMPLEMENTATION DETAILS.....	16
9. TESTING	21
9.1 Test cases and Test results.....	21
9.1.1 Test Cases.....	21
9.1.2 Test Results	21

9.1.3 Test Case for Duke Dunes – News Hub.....	21
10. USER MANUAL	25
10.1 Steps of using entire Duke Dunes – Web App:.....	25
10.2 Database.	30
10.3 Features:	31
10.3.1 Recommended News	31
10.4 News Summarization	33
11. CONCLUSION AND FUTURE WORK.....	35
ANNEXURE.....	36
ABOUT TOOLS AND TECHNOLOGY	38
REFERENCE	40

LIST OF FIGURES

Figure 1 NLP Data Preprocessing	9
Figure 2 Gantt Chart of entire Project Plan	10
Figure 3 Use case diagram	11
Figure 4 Activity Diagram for User	12
Figure 5 Class Diagram.....	13
Figure 6 Sequence Diagram	14
Figure 7 Data Flow Diagram.....	15
Figure 8 QR Code	17
Figure 9 Homepage Navigation	25
Figure 10 Searching for News.....	26
Figure 11 search query	26
Figure 12 Registration details.....	27
Figure 13 Sign Up Authentication.....	28
Figure 14 username and password	28
Figure 15 Login Authentication	29
Figure 16 Recommended Topics.....	29
Figure 17 News Summarizer.....	30
Figure 18 Scroll Bar	30
Figure 19 Database.....	31
Figure 20 Business of Recommended News	31
Figure 21 Entertainment of Recommended News.....	32
Figure 22 Health of Recommended News.....	32
Figure 23 Nation of Recommended News	33

LIST OF TABLES

Table 1 : Test Case 1	22
Table 2 : Test Case 2	22
Table 3 : Test Case 3	23
Table 4 : Test Case 4	24
Table 5 : Test Case 5	24

Vision and Mission of Ganpat University

Vision	It shall be the constant endeavour of Ganpat University to meet the educational needs of the youth in the areas of professional studies and provide state-of the art learning opportunities along with inculcation of values of commitment and uprightness.
Mission	Seek, search and offer programs that lead to symbiotic emergence of 'academic excellence' and 'industrial relevance' in education and research.

Vision and Mission of Information Technology Department

Vision	Department aims to achieve its recognition as a leading contributor in the area of technical education of Information Technology by practicing latest principles, tools and technologies to cope with current and future challenges and hence contributing to global welfare.
Mission	<ol style="list-style-type: none"> 1. To educate and inculcate strong fundamentals of science and Information Technology through best teaching learning practices. 2. To impart high quality education to acquire skills to conduct research and solve complex problems through modern tools, technologies and innovative practices. 3. Enabling youth for employability, social upliftment, following good moral practices and professional ethics. 4. Preparing youth for contributing in advancements of technology and society. 5. Encouraging students to be adaptive, courageous and life-long learners.

1. INTRODUCTION

It is important to be educated in the quickly changing world of information distribution. Innovative platforms that deliver news and material directly to consumers, customised to their tastes, have emerged because of the digital age. I developed the News Hub online application "Duke Dunes" as a reaction to this changing environment. The goal of this project was to develop a feature-rich, easy-to-use platform that combined news from several sources with state-of-the-art technology to improve the user experience. Duke Dunes leverages the capabilities of Python and the Streamlit framework to offer a smooth and dynamic interface for users to consume news articles throughout hosting and deployment. By using Google Firebase as the cloud database, user registration information is securely stored and facilitates robust authentication processes. The integration of the Google API allows the application to fetch news from diverse sources, presenting a comprehensive and real-time snapshot of current events.

Duke Dunes goes beyond basic news presentation, incorporating Natural Language Processing (NLP) for a sophisticated news summarizer. This feature provides users with concise summaries of news articles, along with key details such as titles, images, and publication dates. The inclusion of recommended news topics, spanning business, science, politics, entertainment, and more, ensures a personalised and diverse news consumption experience.

This project report explores Duke Dunes's architecture, features, and upcoming improvements. It describes the technology used, the difficulties encountered during development, and the planned actions to improve the application going forward. Duke Dunes is a monument to the marriage of creativity and user-centred design in the field of news aggregation and distribution, as technology continues to transform how we access and consume information.

1.1 Project Purpose

Duke Dunes is driven by a simple yet powerful purpose — to redefine how users engage with news. We aim to provide an uncomplicated, ad-free space where information is effortlessly accessible. With a commitment to user privacy, we eliminate the need for mandatory logins, offering a distraction-free environment.

Our unique touch lies in the integration of Natural Language Processing (NLP), distilling extensive news articles into concise summaries for quick comprehension. Looking forward, the addition of login capabilities sets the stage for a personalized recommendation system, tailoring the news feed to each user's preferences. Duke Dunes is not just a news hub; it's a personalized, user-centric platform in the ever-evolving realm of digital news consumption.

1.2 Project Statement

Ever found yourself drowning in a sea of news, bombarded by intrusive ads and tangled up in the complexities of traditional news websites? It's a common struggle we've all faced. Navigating through multiple clicks, drowning in excessive information, and contending with relentless ads can make staying informed feel like an uphill battle. At Duke Dunes, we

recognize this challenge and aim to simplify the news consumption experience. In a world where people crave straightforward, accessible information without the hassle, Duke Dunes stands out. Our mission is to create a news hub that's refreshingly straightforward – no logins, no disruptive ads. We're committed to making news enjoyable, effortless, and universally accessible, addressing the frustrations that often accompany the quest to stay informed. Duke Dunes is the solution we're crafting to redefine the way people engage with news, offering simplicity and ease in an otherwise cluttered landscape.

1.3 Objective

Below is list of objectives:

- **Streamlined News Access:** Provide users with a simplified platform for accessing the latest news from diverse sources without the need for extensive navigation.
- **Distraction-Free Environment:** Eliminate the disruption caused by ads and unnecessary complexities, ensuring a clean and distraction-free user interface.
- **User-Friendly Experience:** Create an intuitive and user-friendly interface that accommodates users of all technological proficiencies, making news consumption effortless.
- **Privacy and Accessibility:** Respect user privacy by eliminating mandatory logins, allowing users to freely explore news content while ensuring accessibility for all.
- **NLP-Based Summarization:** Implement Natural Language Processing (NLP) to distil lengthy news articles into concise summaries, providing users with a quick understanding of news stories.
- **Future-Ready Personalization:** Lay the foundation for future enhancements by introducing login functionality, setting the stage for a personalized recommendation system based on user preferences.
- **Efficient Data Storage with Firebase:** Utilize Firebase for efficient data storage, ensuring seamless retrieval of news articles and maintaining scalability for future feature additions.
- **Enhanced User Engagement:** Foster user engagement through an interactive and informative platform that encourages active participation in the news consumption process.
- **Continuous Improvement:** Establish a framework for ongoing refinement and enhancement based on user feedback, emerging technologies, and the evolving landscape of news consumption.

2. PROJECT SCOPE

The scope of the Duke Dunes News Hub project encompasses the development of a comprehensive web application designed to aggregate and deliver news content from various sources. The primary goal is to create a user-friendly and feature-rich platform that enhances the news consumption experience by incorporating advanced technologies.

2.1 Key Objectives:

Build a user-friendly News Hub web app with NLP-based summarization, API integration for news, Firebase for user data, and Streamlit for an interactive UI.

- **News Aggregation:** The application will aggregate news articles from diverse sources using the Google API, providing users with a centralized hub for accessing real-time information.
- **User Registration and Authentication:** Duke Dunes will feature a secure user registration system leveraging Google Firebase as a cloud database. This system will facilitate user authentication to ensure a personalized and secure environment.
- **Interactive User Interface:** Utilizing the Streamlit framework, the web app will offer a seamless and interactive interface for users to navigate, explore, and consume news content effortlessly.
- **News Summarization with NLP:** The application will implement Natural Language Processing (NLP) to provide users with concise summaries of news articles, along with essential details such as titles, images, and publication dates.
- **Recommended News Topics:** Duke Dunes will include a feature for recommending news topics based on user preferences, enhancing the personalization of the news viewing experience.
- **Future Work:** Personalization and Recommendation System - Although not integrated into the current version, the project acknowledges the future implementation of a personalized recommendation system to further tailor content based on user behaviour and preferences.

2.2 Out of Scope:

Tasks or features not considered or addressed.

- The project does not involve the development of native mobile applications at this stage.
- Detailed integration of personalized recommendation features will be part of future work beyond the current project scope.

2.3 Constraints:

Limitations or restrictions affecting project design, development, or implementation.

- Time constraints may limit the implementation of certain advanced features within the current phase.
- The application will be developed and tested for compatibility with modern web browsers, with a primary focus on desktop users.

This project scope is designed to deliver a robust and functional News Hub web application while providing a clear outline of the features and functionalities that fall within and outside the project's boundaries.

3. FEASIBILITY ANALYSIS

In evaluating the viability of the Duke Dunes News Hub web application, a thorough feasibility analysis has been conducted, considering multiple dimensions crucial to project success.

3.1 Technical Feasibility:

The technical feasibility of the project revolves around the implementation of advanced technologies, including Python, Streamlit, Google Firebase, and the Google API. The choice of these technologies aligns with industry best practices, ensuring a scalable, robust, and technologically sound solution. The technical framework selected offers a solid foundation for the development, deployment, and future enhancements of the web application.

3.2 Time Schedule Feasibility:

A realistic and comprehensive time schedule has been established for the development lifecycle of Duke Dunes. The project timeline factors in various stages, including planning, design, development, testing, and deployment. Adequate buffer periods have been incorporated to account for potential challenges and adjustments, ensuring that the project adheres to the planned schedule.

3.3 Operational Feasibility:

Operational feasibility examines the practicality of the proposed system within the organization. Duke Dunes, designed as a user-centric web application, aligns with operational requirements. The intuitive interface, seamless navigation, and efficient news retrieval contribute to an operational environment where users can easily access and consume news content. User feedback and usability testing further validate the operational feasibility of the application.

3.4 Implementation Feasibility:

The implementation feasibility assesses the practicality of integrating the proposed system into the existing organizational infrastructure. Duke Dunes employs widely adopted technologies, minimizing integration complexities. The incorporation of Google Firebase for user registration and authentication enhances implementation feasibility by providing a secure and scalable cloud-based solution.

3.5 Economic Feasibility:

Economic feasibility evaluates the financial aspects of the project, including development costs, potential returns, and long-term sustainability. Duke Dunes justifies its economic feasibility by utilizing open-source technologies, reducing licensing expenses. Additionally, the scalability of the solution ensures cost-effectiveness over time. The project's potential for user engagement and its alignment with current market trends enhance the overall economic viability.

4. SOFTWARE AND HARDWARE REQUIREMENTS

The Duke Dunes project relies on essential software components such as Streamlit for web hosting, Python for development, Google Firebase for cloud-based storage, and NLP libraries for news summarization. On the hardware front, standard laptops or desktops with adequate RAM and processing power are utilized. Stable internet connections and compatibility with various devices and browsers ensure smooth development and user accessibility.

4.1 Software Requirements:

Descriptions of functions, capabilities, constraints, and specifications that a software system must satisfy.

- Python 3.x: Primary programming language for web app development.
- Streamlit: Framework for creating interactive web applications with Python.
- Firebase SDK: For integrating Firebase services into the web app.
- Firebase Realtime Database: Cloud-hosted NoSQL database for storing news data.
- NLTK (Natural Language Toolkit) or SpaCy: Python libraries for implementing NLP-based news summarization.
- HTML, CSS, JavaScript: For enhancing the user interface and interactivity.
- Git: To manage and track changes in the project codebase.

4.2 Hardware Requirements:

Specifications of physical components needed for a system to function properly.

- Development Machine: A standard laptop or desktop with a minimum of 8GB RAM and a multi-core processor.
- Storage: Adequate storage for codebase, development tools, and related files.
- Internet Connection: A stable and reliable internet connection for accessing external libraries, APIs, and Firebase services.
- Testing Devices: Ensure compatibility with various devices and browsers for user testing during development.
- Server: The web app can be hosted on a cloud server, such as Heroku or similar, for deployment.

4.3 Client-Side Requirements:

Functionalities or features essential for the end-user experience in a software application.

- Device: Laptop with an internet connection.
- Browser: Modern browsers like Chrome, Firefox.
- Responsive Design: Adaptable interface for laptops and mobile devices.
- Internet Connection: Stable for real-time updates.

- Navigation: Intuitive for easy exploration.
- System Requirements: Lightweight for smooth performance.
- Compatibility: Supports Windows, macOS, Linux.
- Accessibility: Inclusive design for diverse user needs.

4.4 Functional and Non-Functional Requirements.

The functional requirements for Duke Dunes include displaying news articles with titles [3], images, and publish dates, a trending section, favourites topic selection, a search feature, and implementing NLP-based news summarization. Non-functional requirements encompass ensuring fast loading times, scalability to accommodate increased user traffic, robust security measures, an intuitive user interface, reliability in news updates, compatibility across multiple browsers, maintaining an ad-free environment, and designing for diverse user needs.

4.4.1 Functional Requirements:

Descriptions of system features and capabilities that meet the users' needs and expectations.

- News Display: Display latest news with titles, images, and publish dates.
- Trending Section: Highlight trending news topics.
- Favourite Topics: Allow users to mark and follow favourite news topics.
- Search Functionality: Provide a search feature for users to find specific news.
- News Summarization: Implement NLP-based summarization for news articles.

4.4.2 Non-Functional Requirements:

Aspects like performance, security, usability, not tied to specific functionalities but critical for success.

- Performance: Ensure fast loading and response times.
- Scalability: Design to handle potential increases in user traffic.
- Security: Protect user data and ensure secure data transmission.
- User Experience: Create an intuitive and user-friendly interface.
- Reliability: Minimize downtime and ensure reliable news updates.
- Compatibility: Support multiple browsers for broad accessibility.
- Ad-Free: Maintain an ad-free environment for distraction-free reading.

5. PROCESS MODEL

For the Duke Dunes project, the Agile Scrum methodology serves as the chosen process model. This iterative and incremental framework allows for flexibility and adaptability throughout the development lifecycle. It involves dividing the project into smaller work cycles called sprints, typically lasting 2-4 weeks, where tasks are defined, developed, and reviewed. The Scrum framework includes daily stand-up meetings to track progress, sprint planning sessions to outline tasks, sprint reviews to assess outcomes, and retrospectives to improve future iterations. This methodology fosters collaboration, adaptability to changing requirements, and frequent deliverables, ensuring continuous refinement and responsiveness to user needs.

- **Meeting and Planning:** This initial phase involves team meetings to define project goals, outline tasks, and set objectives for the development process. It includes requirements gathering, defining the scope of work, and planning the project milestones and deliverables.
- **Design:** The design phase concentrates on creating the architecture and layout for the Duke Dunes News Hub. It involves UI/UX design, wireframing, and conceptualizing the user interface to ensure an intuitive and engaging browsing experience.
- **Code and Testing:** During this phase, developers begin coding based on the design specifications, implementing functionalities using tools like Streamlit, Python, and NLP libraries. Simultaneously, rigorous testing is conducted, including unit testing, integration testing, and functional testing, to identify and rectify any issues or bugs.
- **Release:** Upon successful coding, testing, and ensuring compliance with project objectives, the Duke Dunes application is prepared for deployment. This phase involves configuring the app for hosting, setting up the necessary infrastructure, and ensuring a seamless transition to a live environment.
- **Feedback:** After the release, the application is made available to users. This phase emphasizes gathering feedback through user interaction, surveys, and analytics tools. Feedback is assessed and used to understand user preferences, identify areas for improvement, and guide future enhancements.

5.1 Summarizing the articles

Now we've got the titles of the articles, but only listing the titles doesn't sound cool. Next, let's make a summarizer of each article to make our app more useful. In this project, I made a simple NLP model using Scapy.

Before applying to an NLP [2] model, text data need to be pre-processed. I will show the summary of NLP data preprocessing in the picture below. I skip the details because it is out of the scope of this article, but if you want to learn more about NLP data preprocessing, this is a good article to start with.

NLP Data Preprocessing

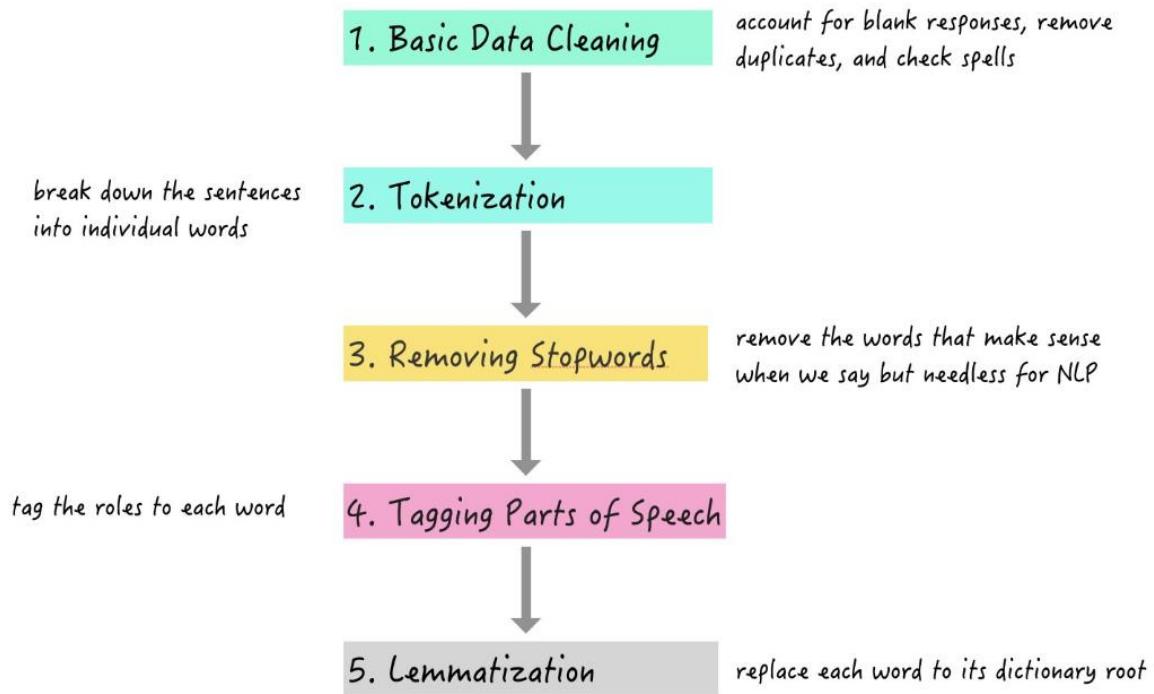


Figure 1 NLP Data Preprocessing

The above Figure 1 is a Natural Language Processing (NLP) in data processing involves using computational techniques to interpret and analyse human language. It enables machines to understand, interpret, and generate human-like text, facilitating tasks such as sentiment analysis, language translation, and text summarization. NLP algorithms process and transform unstructured textual data, making it accessible for various applications in information extraction and understanding. NLP is crucial for unlocking insights from vast amounts of text data, enhancing automated decision-making processes. [1]

6. PROJECT PLAN

The project plan for Duke Dunes News Hub is a strategic roadmap that encompasses initiation, planning, development, testing, and deployment phases. It began with a comprehensive analysis to define scope and objectives, followed by meticulous planning of technical architecture and features using Python, Streamlit, Google Firebase, and Google API. The development phase brought the envisioned web application to life, employing Streamlit for an interactive interface and Google Firebase for secure user data storage. Rigorous testing procedures ensured reliability and security, leading to a seamless deployment. The project plan also lays the groundwork for future enhancements, particularly the integration of a personalized recommendation system, aiming to deliver an innovative and user-centric news aggregation platform.

6.1 Gantt Chart of entire Project Plan with duration, start date and end date:

A Gantt Chart is a visual representation of a project schedule, displaying tasks, their durations, and dependencies over time, offering a comprehensive view of project planning and progress.

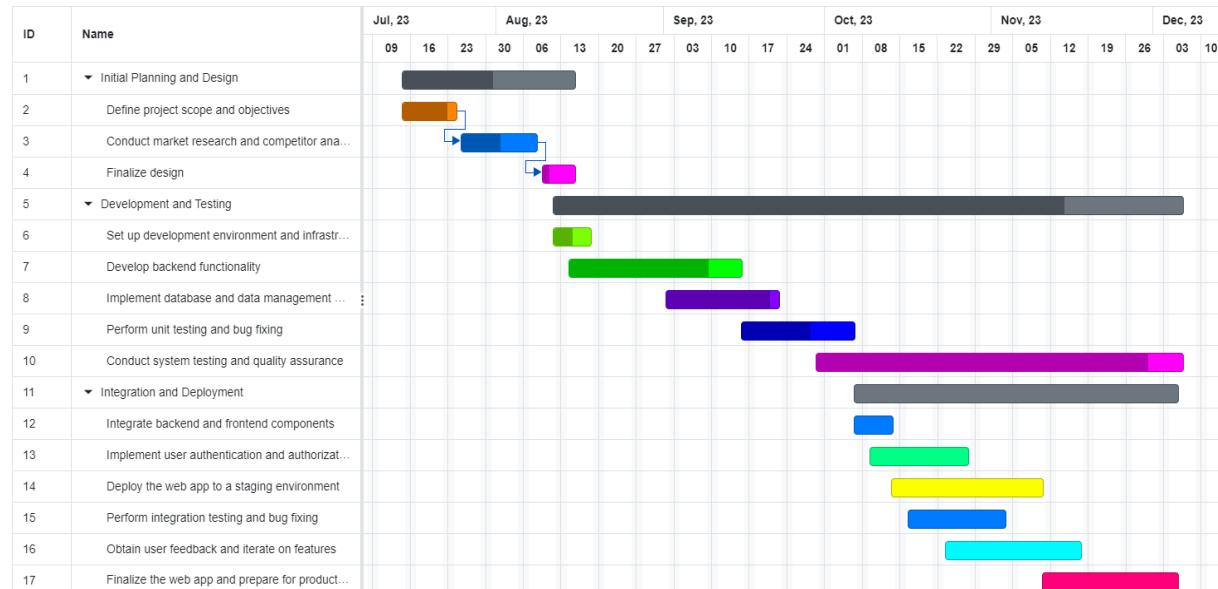


Figure 2 Gantt Chart of entire Project Plan

The above Figure 2 is Gantt Chart showcases the entire project plan, detailing the schedule, duration, and start/end dates of various tasks and activities involved in the development of the News Hub web application. It provides a visual representation of the project timeline, allowing you to effectively plan, coordinate, and track the progress of different aspects of the project, such as development, testing, deployment, and other key milestones. The Gantt Chart serves as a comprehensive project management tool, offering insights into the sequential and parallel execution of tasks, ensuring efficient project execution and timely delivery.

7. SYSTEM DESIGN

The System Design refers to the designing of various diagrams like Use-Case diagram, Class diagram, activity diagram for each module of the project, Sequence diagram, Data-Flow diagram (DFD) level-0 & 1 and the flow chart of the entire system. All the diagrams except Data-Flow diagram that are designed under System Design only describes the overview of the entire system that means it describes that what are the inputs and what are the outputs and it does not describe the internal structure of the entire system and how the data flows inside the system. So, we can say that these diagrams except the DFD is the white-box view of the entire system. The Data-Flow diagram (DFD) only shows the internal flow of the data in the entire system.

7.1 Use case diagram

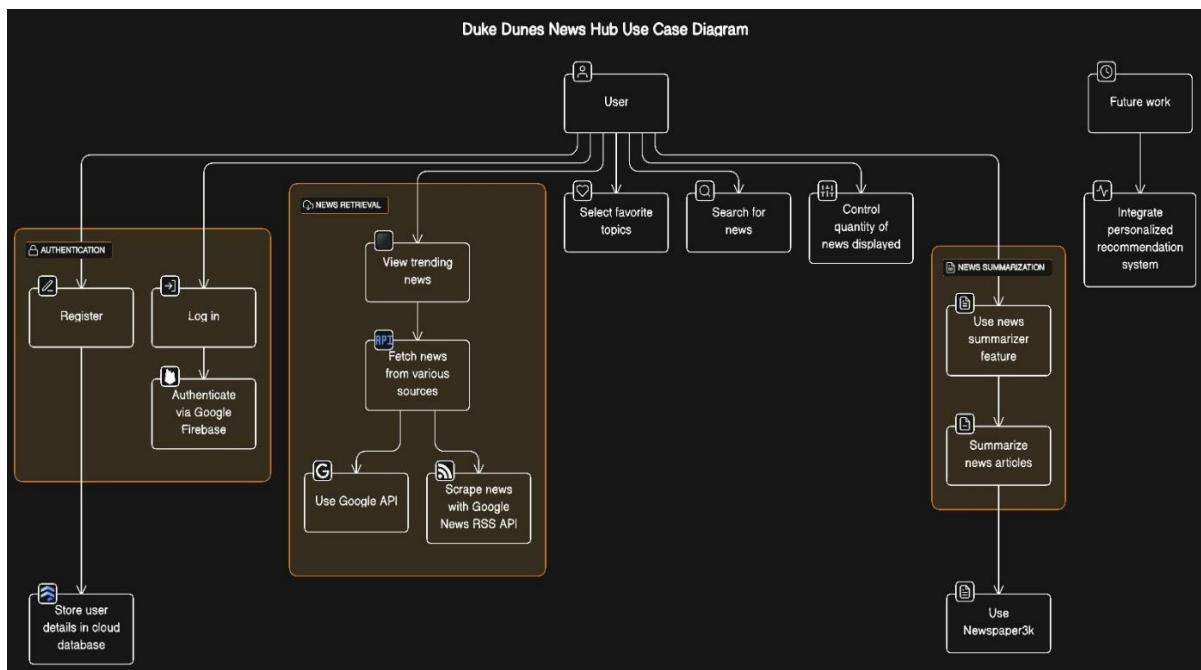


Figure 3 Use case diagram

The Above Figure 3 is a use case diagram for the Duke Dunes project outlines the various interactions between actors and the system, detailing the functionalities offered by the News Hub platform. In this diagram, the primary actor is the User, representing individuals accessing the application. The use cases encapsulate distinct functionalities available to users. These include "View News," allowing users to read articles, "Search News" for finding specific topics, "Trending Topics" to display popular news categories, "Favourite Topics" enabling users to mark preferred categories, and "NLP Summarization" for condensing lengthy articles. Additionally, "User Registration" and "User Authentication" manage account creation and access, while "Feedback Submission" collects user opinions. The diagram ensures a visual representation of user-system interactions, facilitating a comprehensive understanding of the Duke Dunes platform's functionalities and user interactions within the system.

7.2 Activity Diagram for User

Duke Dunes News Hub Activity Diagram

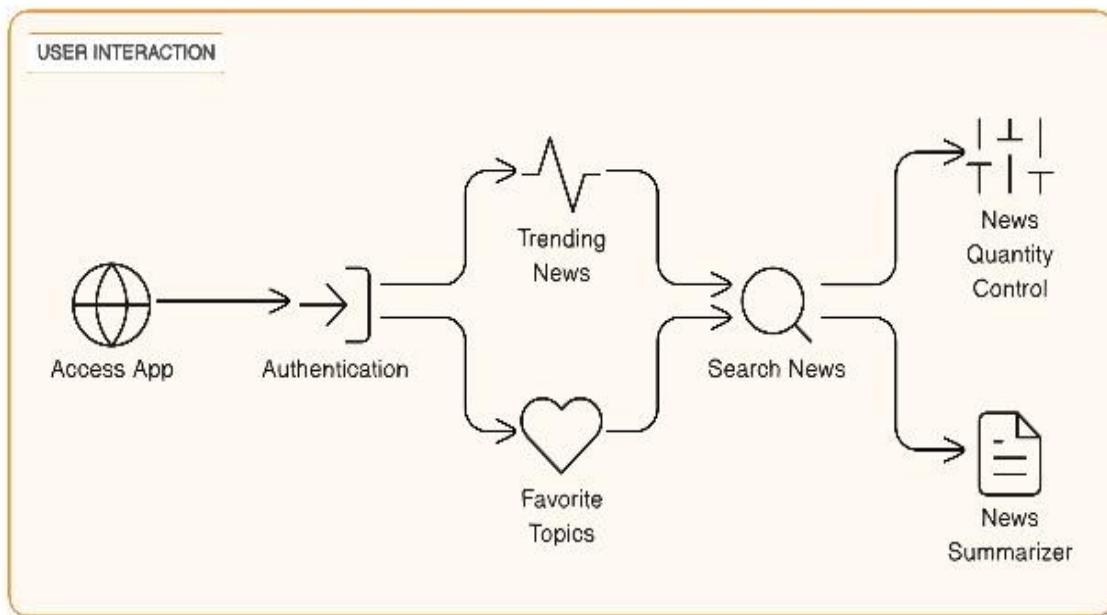


Figure 4 Activity Diagram for User

The Above Figure 4 is an activity diagram for the Duke Dunes project illustrates the sequential flow of actions or activities a user undertakes within the News Hub platform. It delineates the step-by-step process of a user's interaction with the system. Initially, the user may opt to "Browse News," triggering a series of subsequent activities. The diagram illustrates the pathways: a user can either "View Trending Topics" or proceed to "Search News." If the user chooses the former, they explore the trending categories; otherwise, they input keywords or topics for a targeted news search. Following this, the user can select a specific news article for detailed reading, leading to the "Read Article" activity. During this process, the user might engage in further actions like marking a news category as a "Favourite Topic" or providing "Feedback." The activity diagram maps out the logical flow of actions, providing a clear depiction of how users navigate the Duke Dunes platform and engage with its functionalities, ensuring a user-centric approach to the system's activities and operations.

7.3 Class Diagram

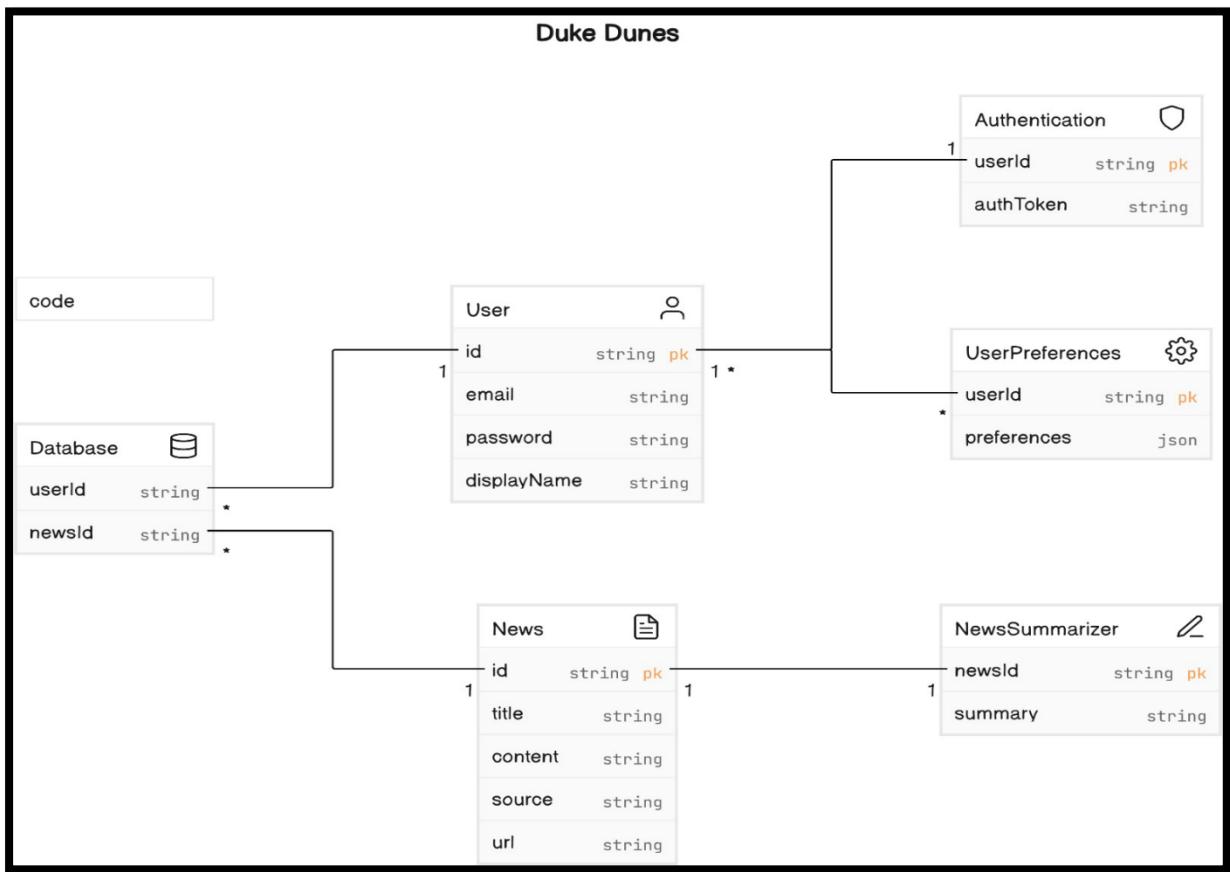


Figure 5 Class Diagram

The above Figure 5 is a class diagram for the Duke Dunes project represents the structural elements and relationships within the News Hub platform. Within this diagram, classes depict entities or objects, while associations showcase connections or interactions between these entities. For instance, the primary class "User" encapsulates attributes like username, password, and preferences, illustrating user data storage. The "Article" class contains attributes such as title, content, and publication date, storing information related to news articles fetched by the system. Associations between these classes depict relationships; for example, the "User" class may have an association with "Article," signifying that users interact with news articles. Additionally, auxiliary classes like "Search Engine" or "News Summarizer" might exist to encapsulate specific functionalities within the system, facilitating article search or summarization. The class diagram serves as a blueprint, offering a visual representation of the fundamental components, their attributes, and associations, aiding in the understanding and organization of the Duke Dunes system's structural aspects and relationships between user-centric elements.

7.4 Sequence Diagram

User Interaction with Duke Dunes News Hub

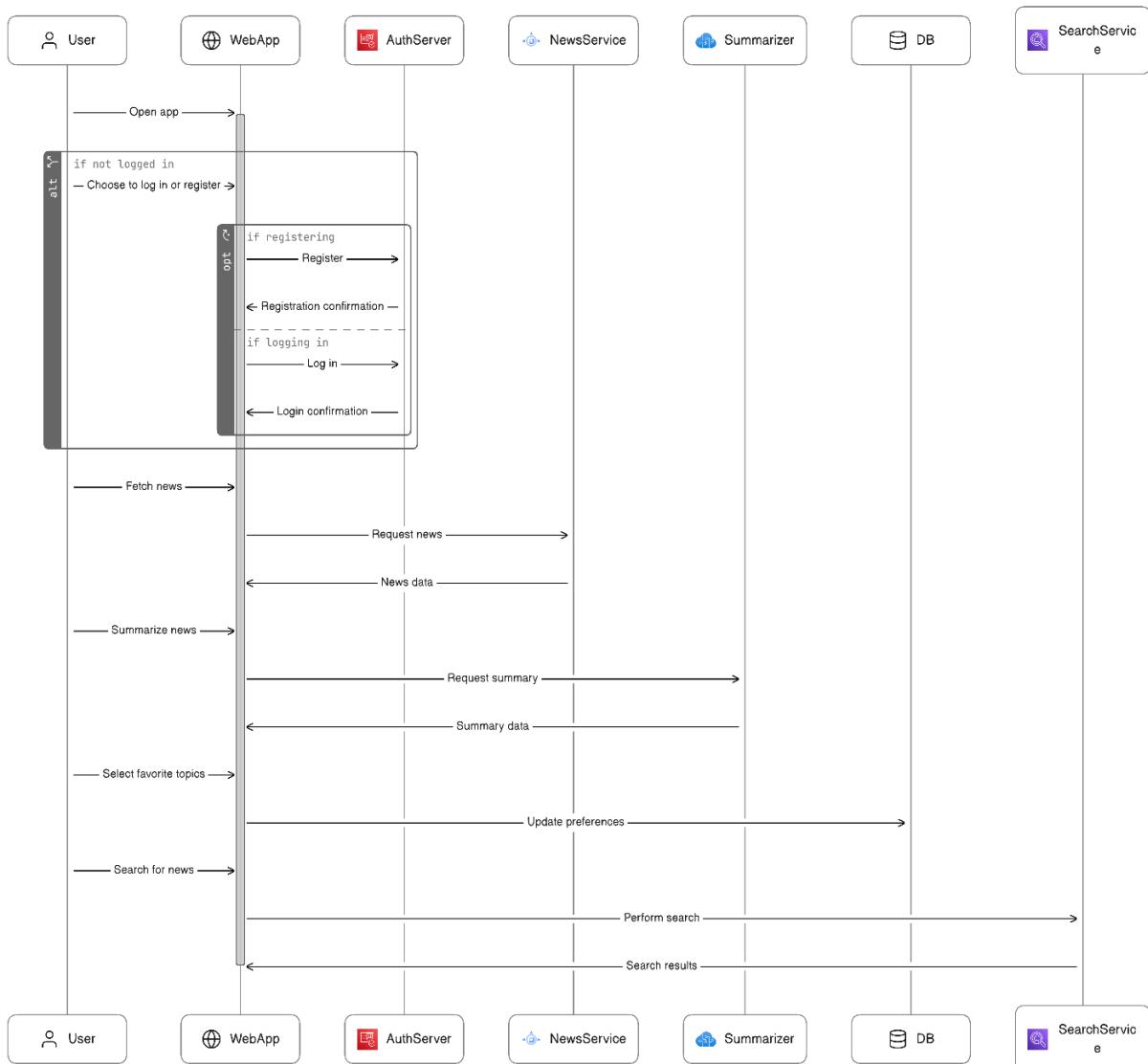


Figure 6 Sequence Diagram

The above figure 6 is a sequence diagram for the Duke Dunes project outlines the chronological flow of interactions between system components and a user, illustrating the step-by-step process of a user's actions within the News Hub platform. It visualizes the sequence of messages or events exchanged between different entities over time. The diagram commences with a user initiating an action, such as selecting "View News" or "Search News." Upon triggering the action, the sequence diagram showcases how the system responds. For instance, upon choosing "Search News," the system processes the user's query, interacts with the search engine module, and retrieves relevant articles. Subsequently, the system presents search results to the user. The sequence diagram demonstrates these interactions through arrows representing messages between the user and the system components, showcasing the flow of activities and the chronological order of events during a user's engagement with the Duke Dunes platform. This visualization aids in understanding the dynamic interactions and the series of actions occurring

between the user and the system, providing insights into the operational flow of the News Hub platform.

7.5 Data Flow Diagram (DFD)

Duke Dunes News Hub Data Flow Diagram

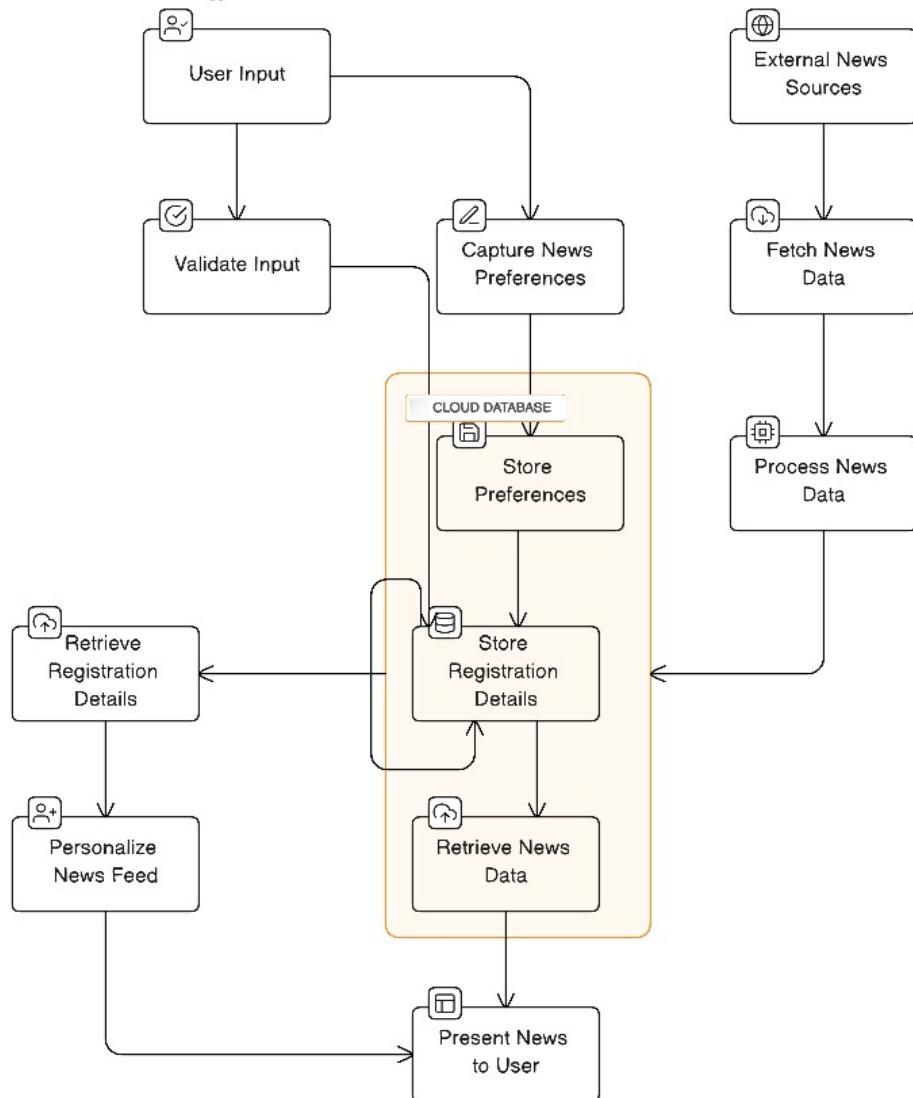


Figure 7 Data Flow Diagram

The above Figure 7 is a Data Flow Diagram (DFD) which is graphical representation that illustrates how data moves through a system, emphasizing the processes, data sources, data destinations, and data storage involved. It provides a clear and visual overview of the information flow within a system or between different components.

8. IMPLEMENTATION DETAILS

The implementation of Duke Dunes News Hub involves a strategic blend of cutting-edge technologies and meticulous development processes. The chosen tech stack comprises Python as the primary programming language, the Streamlit framework for web hosting and deployment, and Google Firebase as the cloud database. Additionally, the Google API seamlessly integrates to fetch news from a diverse array of sources.

Programming Language: Python

Python serves as the backbone of Duke Dunes, providing a robust and versatile foundation for web application development. Its ease of readability and extensive library support facilitated efficient coding, enabling the implementation of complex functionalities.

Web Framework: Streamlit

Duke Dunes harnesses the power of the Streamlit framework for hosting and deployment. Streamlit's simplicity and rapid prototyping capabilities proved instrumental in creating an interactive and user-friendly web interface. Leveraging Streamlit allowed for a seamless development process, facilitating quick iteration and feature enhancements.

Cloud Database: Google Firebase

Google Firebase is employed as the cloud database, ensuring secure storage of user registration details and enabling robust authentication processes. The integration of Firebase enhances the scalability and reliability of Duke Dunes, providing a cloud-based infrastructure that aligns with modern development standards.

API Integration: Google API

The web app seamlessly integrates with the Google API to fetch news from a multitude of sources. This integration provides real-time access to a diverse range of news articles, ensuring a comprehensive and up-to-date news feed. The Google API acts as a pivotal component in Duke Dunes' ability to aggregate content from various channels.

Natural Language Processing (NLP): News Summarization

Duke Dunes goes beyond conventional news presentation by incorporating Natural Language Processing (NLP) for news summarization. This advanced feature delivers concise summaries of news articles, including key details such as titles, images, and publication dates. NLP enhances user engagement by providing a quick and informative snapshot of each article.

Future Enhancements: Personalized Recommendation System

While not fully integrated into the current version, Duke Dunes anticipates the implementation of a personalized recommendation system as part of future enhancements. This system aims to tailor news content based on user preferences, utilizing advanced algorithms to enhance the overall user experience.

For deployment of project, we just uploaded all the files on GitHub using git, so to access this project we all just need scan this QR code



Figure 8 QR Code

The above Figure 8 is the QR Code image that when any reader of this document scans that QR code then he/she will be redirected to the GitHub link of our project and can view the entire GitHub repository and can access and view any file related to our project.

8.1 Actual Program Code:

```
import os
import io
import nltk
import pyperclip
import streamlit as st
from PIL import Image
from bs4 import BeautifulSoup as soup
from urllib.request import urlopen
from newspaper import Article
# from streamlit_option_menu import option_menu
folder = "pages"
topics= ['--Select--']
for i in os.listdir(folder):
    temp = i[:-3]
    topics.append(temp)
nltk.download('punkt')
st.set_page_config(page_title='News Hub 📰 Portal', page_icon='./assets/DDlogo.webp')
def fetch_news_search_topic(topic):
```

```

site = 'https://news.google.com/rss/search?q={ }'.format(topic)

op = urlopen(site) # Open that site

rd = op.read() # read data from site

op.close() # close the object

sp_page = soup(rd, 'xml') # scrapping data from site

news_list = sp_page.find_all('item') # finding news

return news_list

def fetch_top_news():

    site = 'https://news.google.com/news/rss'

    op = urlopen(site) # Open that site

    rd = op.read() # read data from site

    op.close() # close the object

    sp_page = soup(rd, 'xml') # scrapping data from site

    news_list = sp_page.find_all('item') # finding news

    return news_list

def fetch_category_news(topic):

    site = 'https://news.google.com/news/rss/headlines/section/topic/{ }'.format(topic)

    op = urlopen(site) # Open that site

    rd = op.read() # read data from site

    op.close() # close the object

    sp_page = soup(rd, 'xml') # scrapping data from site

    news_list = sp_page.find_all('item') # finding news

    return news_list

def fetch_news_poster(poster_link):

    try:

        u = urlopen(poster_link)

        raw_data = u.read()

        image = Image.open(io.BytesIO(raw_data))

        st.image(image, use_column_width=True)

    except:

```

```

image = Image.open('./assets/no_image.png')
st.image(image, use_column_width=True)

def display_news(list_of_news, news_quantity):
    c = 0
    for news in list_of_news:
        c += 1
        # st.markdown(f"({c}) [{news.title.text}]({news.link.text})")
        st.write('**({ }) { }**'.format(c, news.title.text))
        news_data = Article(news.link.text)
        try:
            news_data.download()
            news_data.parse()
            news_data.nlp()
        except Exception as e:
            st.error(e)
        fetch_news_poster(news_data.top_image)
        with st.expander(news.title.text):
            st.markdown(
                "<h6 style='text-align: justify;'>{ }</h6>".format(news_data.summary),
                unsafe_allow_html=True)
            st.markdown("[Read more at { }...](#{})".format(news.source.text, news.link.text))
            st.success("Published Date: " + news.pubDate.text)

share_button = st.button("Share", key=f"share_{c}")
if share_button:
    pyperclip.copy(news.link.text)
    if c >= news_quantity:
        break

def run():
    st.title("News Hub 📰")

```

```

image = Image.open('./assets/whitelogo.png')

col1, col2, col3 = st.columns([3, 5, 3])

with col1:

    st.write("")

with col2:

    st.image(image, use_column_width=False)

with col3:

    st.write("")

chosen_topic = "Trending News"

user_topic = st.text_input("Enter your Topic")

no_of_news = st.slider('Number of News:', min_value=5, max_value=25, step=1)

if st.button("Search") and user_topic != "":

    user_topic_pr = user_topic.replace(' ', '')

    news_list = fetch_news_search_topic(topic=user_topic_pr)

    if news_list:

        st.subheader(" ✅ Here are the some {} News for you".format(user_topic.capitalize()))

        display_news(news_list, no_of_news)

    else:

        st.error("No News found for {}".format(user_topic))

else:

    # st.warning("Please write Topic Name to Search")

    news_list = fetch_top_news()

    if news_list:

        st.subheader(" ✅ Here are the some {} News for you".format(chosen_topic))

        display_news(news_list, no_of_news)

    else:

        st.error("No News found for {}".format(chosen_topic))

run()

```

Explanation:

This is the code of my Streamlit app which runs on local host.

9. TESTING

In the process of developing the Duke Dunes News Hub web application, a rigorous testing methodology was employed to ensure the functionality, reliability, and user experience align with the project's objectives. Test cases were meticulously designed to assess various aspects of the application, ranging from core functionalities to user interactions. The subsequent test results serve as a comprehensive evaluation, providing valuable insights into the robustness and effectiveness of the Duke Dunes News Hub.

9.1 Test cases and Test results

9.1.1 Test Cases

Test cases are systematically devised scenarios that evaluate the individual functionalities and features of the Duke Dunes application. Each test case represents a specific user interaction or system behavior, enabling a thorough examination of the application's capabilities. The test cases cover key aspects, including homepage loading, search functionality, user registration, login processes, news summarization using Natural Language Processing (NLP), recommended news topics, integration with external APIs, quantity control, and user logout.

9.1.2 Test Results

Test results document the outcomes of the executed test cases, indicating whether the application successfully met the expected behaviors. These results provide crucial insights into the application's reliability, security, and overall performance. Positive test results affirm the proper functioning of each feature, while any discrepancies or failures prompt further investigation and refinement. The test results contribute to the overall assessment of the Duke Dunes News Hub, ensuring that it meets the defined requirements and user expectations.

9.1.3 Test Case for Duke Dunes – News Hub

Test Case Table 1

Test Case ID	Test Scenario	Expected Result	Pass/ Fail	Local Host	Display Image	Signup	Login
TC001	Verify homepage loads successfully	Homepage loads	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC002	Check if trending news section displays	Trending news visible	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC003	Test search functionality	Search results displayed	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC004	Verify quantity control works	Scroll bar visible	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC005	Click on a news article	Article opens successfully	Pass	localhost	Image displayed	Not Applicable	Not Applicable
TC006	Test news summarizer (NLP)	Summary generated	Pass	localhost	Not Applicable	Not Applicable	Not Applicable

TC007	Check favourite topics section	Topics displayed	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC008	Verify user registration functionality	User registered successfully	Fail	localhost	Not Applicable	Fail	Not Applicable
TC009	Test login with correct credentials	User logged in successfully	Pass	localhost	Not Applicable	Not Applicable	Pass
TC010	Test login with incorrect credentials	Error message displayed	Pass	localhost	Not Applicable	Not Applicable	Fail

Table 1 : Test Case 1

Test Case Table 2

Test Case ID	Test Scenario	Expected Result	Pass/Fail	Local Host	Display Image	Signup	Login
TC011	Verify homepage loads successfully	Homepage loads	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC012	Check if trending news section displays	Trending news visible	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC013	Test search functionality	Search results displayed	Fail	localhost	Not Applicable	Not Applicable	Not Applicable
TC014	Verify quantity control works	Scroll bar visible	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC015	Click on a news article	Article opens successfully	Pass	localhost	Image displayed	Not Applicable	Not Applicable
TC016	Test news summarizer (NLP)	Summary generated	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC017	Check favorite topics section	Topics displayed	Fail	localhost	Not Applicable	Not Applicable	Not Applicable
TC018	Verify user registration functionality	User registered successfully	Pass	localhost	Not Applicable	Pass	Not Applicable
TC019	Test login with correct credentials	User logged in successfully	Fail	localhost	Not Applicable	Not Applicable	Pass
TC020	Test login with incorrect credentials	Error message displayed	Pass	localhost	Not Applicable	Not Applicable	Fail

Table 2 : Test Case 2

Test Case Table 3

Test Case ID	Test Scenario	Expected Result	Pass/Fail	Local Host	Display Image	Signup	Login

TC021	Verify homepage loads successfully	Homepage loads	Fail	localhost	Not Applicable	Not Applicable	Not Applicable
TC022	Check if trending news section displays	Trending news visible	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC023	Test search functionality	Search results displayed	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC024	Verify quantity control works	Scroll bar visible	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC025	Click on a news article	Article opens successfully	Fail	localhost	Image displayed	Not Applicable	Not Applicable
TC026	Test news summarizer (NLP)	Summary generated	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC027	Check favorite topics section	Topics displayed	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC028	Verify user registration functionality	User registered successfully	Pass	localhost	Not Applicable	Fail	Not Applicable
TC029	Test login with correct credentials	User logged in successfully	Pass	localhost	Not Applicable	Not Applicable	Pass
TC030	Test login with incorrect credentials	Error message displayed	Fail	localhost	Not Applicable	Not Applicable	Fail

Table 3 : Test Case 3

Test Case Table 4

Test Case ID	Test Scenario	Expected Result	Pass/Fail	Local Host	Display Image	Signup	Login
TC031	Verify homepage loads successfully	Homepage loads	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC032	Check if trending news section displays	Trending news visible	Fail	localhost	Not Applicable	Not Applicable	Not Applicable
TC033	Test search functionality	Search results displayed	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC034	Verify quantity control works	Scroll bar visible	Fail	localhost	Not Applicable	Not Applicable	Not Applicable
TC035	Click on a news article	Article opens successfully	Pass	localhost	Image displayed	Not Applicable	Not Applicable
TC036	Test news summarizer (NLP)	Summary generated	Pass	localhost	Not Applicable	Not Applicable	Not Applicable

TC037	Check favorite topics section	Topics displayed	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC038	Verify user registration functionality	User registered successfully	Fail	localhost	Not Applicable	Fail	Not Applicable
TC039	Test login with correct credentials	User logged in successfully	Pass	localhost	Not Applicable	Not Applicable	Pass
TC040	Test login with incorrect credentials	Error message displayed	Pass	localhost	Not Applicable	Not Applicable	Fail

Table 4 : Test Case 4

Test Case Table 5

Test Case ID	Test Scenario	Expected Result	Pass/Fail	Local Host	Display Image	Signup	Login
TC041	Verify homepage loads successfully	Homepage loads	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC042	Check if trending news section displays	Trending news visible	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC043	Test search functionality	Search results displayed	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC044	Verify quantity control works	Scroll bar visible	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC045	Click on a news article	Article opens successfully	Pass	localhost	Image displayed	Not Applicable	Not Applicable
TC046	Test news summarizer (NLP)	Summary generated	Fail	localhost	Not Applicable	Not Applicable	Not Applicable
TC047	Check favorite topics section	Topics displayed	Pass	localhost	Not Applicable	Not Applicable	Not Applicable
TC048	Verify user registration functionality	User registered successfully	Pass	localhost	Not Applicable	Fail	Not Applicable
TC049	Test login with correct credentials	User logged in successfully	Fail	localhost	Not Applicable	Not Applicable	Pass
TC050	Test login with incorrect credentials	Error message displayed	Pass	localhost	Not Applicable	Not Applicable	Fail

Table 5 : Test Case 5

10. USER MANUAL

Comprehensive user manual included in the project report provides detailed instructions for efficient system utilization, ensuring seamless user experience and facilitating easy troubleshooting for optimal performance.

10.1 Steps of using entire Duke Dunes – Web App:

Step 1: Accessing the Web App

- **Open your preferred web browser:** Launch the web browser you typically use for internet activities.
- **Enter the Duke Dunes web app URL:** In the browser's address bar, type the URL for Duke Dunes and press Enter.

Step 2: Homepage Navigation

- **Once the Duke Dunes homepage loads:** Upon entering the Duke Dunes web app, you will be greeted with a visually appealing homepage.

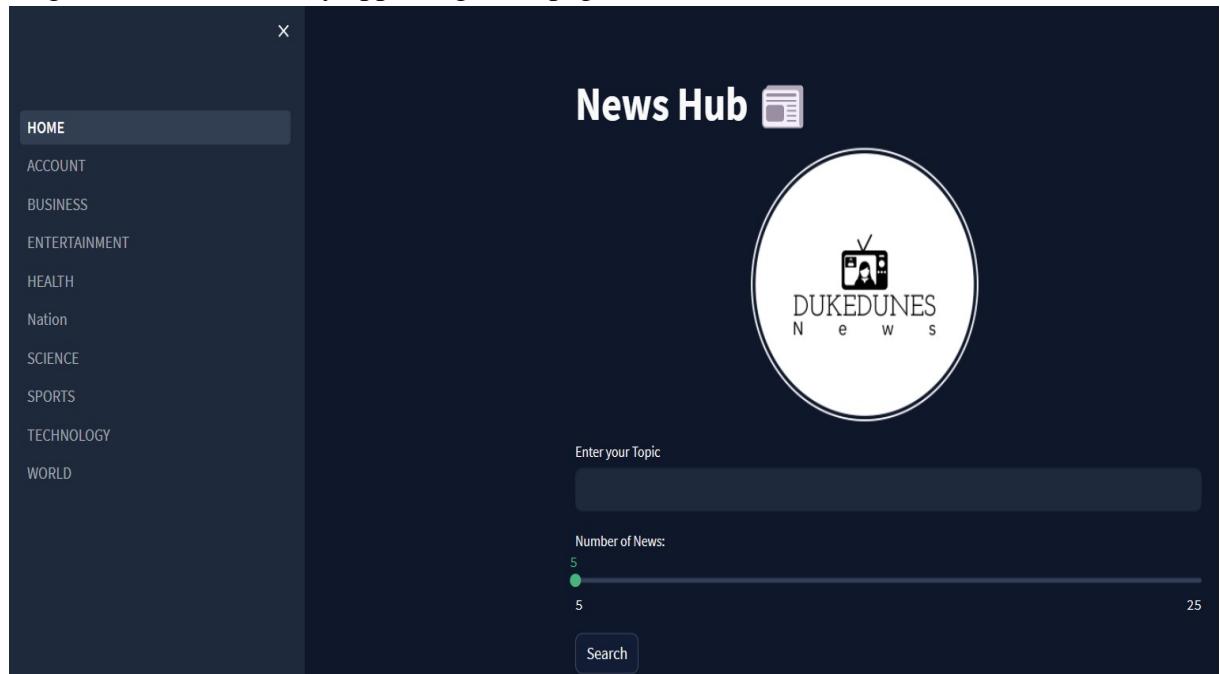


Figure 9 Homepage Navigation

The above Figure 9 is the Home Page of Web App which shows latest news.

- **Explore different news categories:** Navigate through various news categories displayed on the homepage, such as business, science, politics, and more.

Step 3: Searching for News

- **To search for specific news articles:** Locate the search bar at the top of the page.

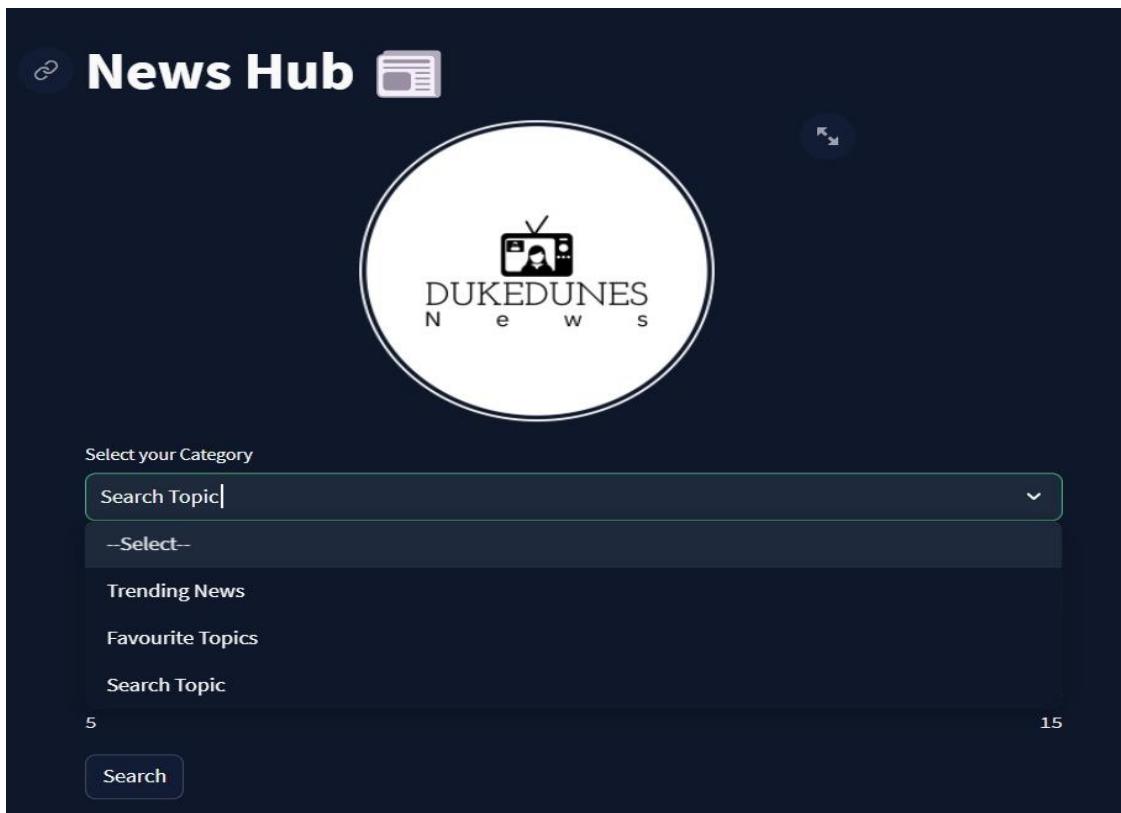


Figure 10 Searching for News

The above Figure 10 is search bar in which we can search any topic.

- **Enter your desired search query:** Type your desired keywords into the search bar.

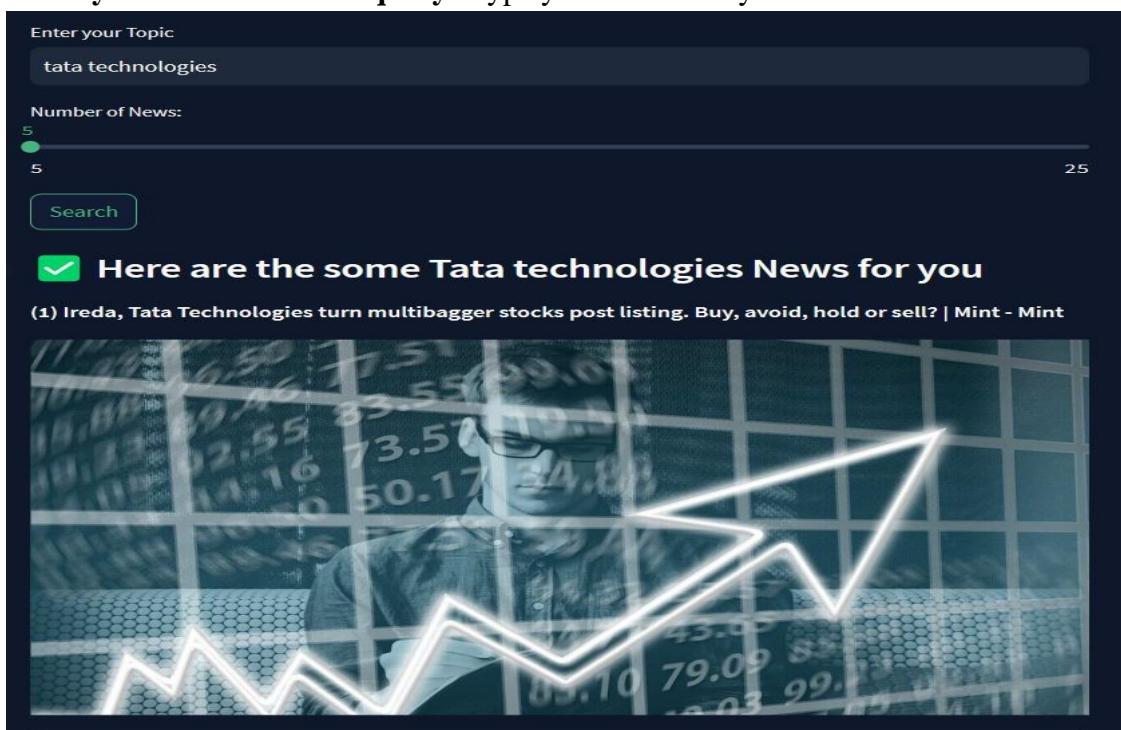


Figure 11 search query

The above Figure 11 is generated after any search is done in search bar.

Browse through the search results: Examine the search results to find news articles relevant to your query.

Step 4: User Registration

- **If you are a new user:** If you don't have an account, find the "Register" or "Sign Up" section.
- **Fill in the required registration details:** Provide the necessary information, including a username and password.

Create Account

Select Operation

Login
 Signup

Sign Up

Email

Password

Figure 12 Registration detail

The above Figure 12 is Registration Details page, in which one can create account or login.

- **Click on the "Register" or "Sign Up" button:** Complete the registration process by clicking the appropriate button.

The screenshot shows a dark-themed 'Sign Up' form. At the top, the word 'Sign Up' is written in blue. Below it is a 'Email' label followed by an input field containing '4@gmail.com'. Underneath is a 'Password' label followed by an input field containing '4@gmail.com' and a small eye icon to its right. At the bottom left is a green-outlined 'Sign Up' button. A green success message at the bottom states 'User created with ID: DAMTBmBlvWWgIVDyeOLBgTwcV2x2'.

Figure 13 Sign Up Authentication

The above Figure 13 is showing up the authentication of User SignUp.

Step 5: User Login

- **For returning users:** If you already have an account, click on the "Login" or "Sign In" button.
- **Enter your username and password :** Provide your login credentials.

The screenshot shows a dark-themed 'Login' form. At the top, the word 'Login' is written in white. Below it is a 'Email' label followed by an input field containing '4@gmail.com'. Underneath is a 'Password' label followed by an input field containing '4@gmail.com' and a small eye icon to its right. At the bottom left is a green-outlined 'Login' button.

Figure 14 username and password

The above Figure 14 is Login page in which any user can login as per his/her own credentials.

- **Click on the "Login" or "Sign In" button:** Access your Duke Dunes account by clicking the appropriate button.

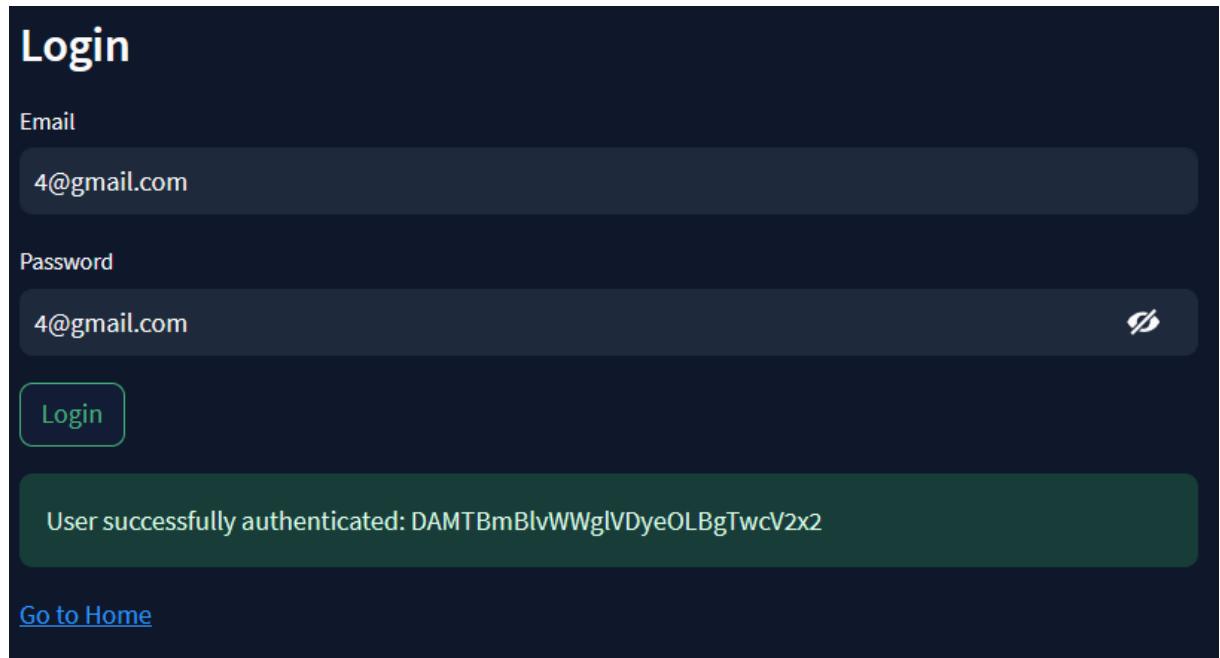


Figure 15 Login Authentication

The above Figure 15 is Login Authentication which shows that user password and email is correct by matching from Firebase Data Storage.

Step 6: Exploring Recommended Topics

- **On the homepage:** Look for the section that showcases recommended news topics.

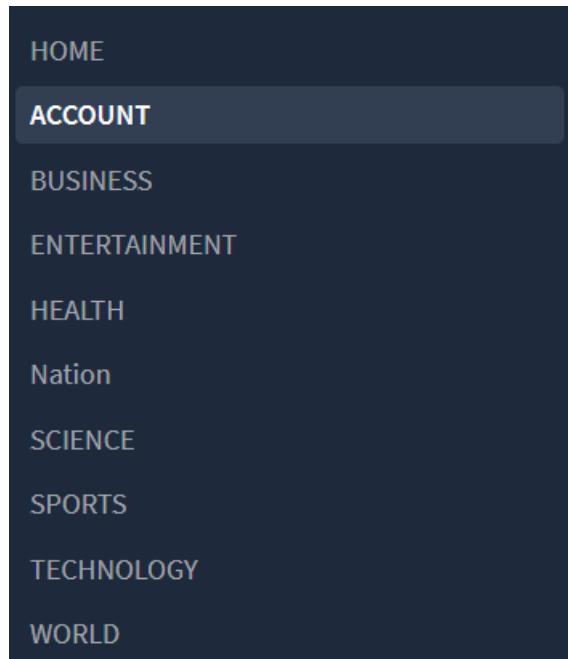


Figure 16 Recommended Topics

The above Figure 16 is a Recommended Topics I mentioned which is most Trending.

- **Click on a recommended topic:** Select a category of interest to explore curated news articles.

Step 7: News Summarizer (NLP)

- **Click on any news article:** Open a full news article to access the summarizer.

The screenshot shows a news article from NDTV Sports titled "'What A Legend": Sam Curran Wearing Sunglasses While Batting Breaks Internet - NDTV Sports'. The main content is a summary of a West Indies ODI match, mentioning Sam Curran's 90s, Harry Brook's counterattack, and the final score of 109*. Below the summary is a link 'Read more at NDTV Sports...'.

Figure 17 News Summarizer

The above Figure 17 is a Summarization of a News Articles.

- **Experience the news summarizer:** Read a concise summary of the article provided by the Natural Language Processing (NLP) technology.

Step 8: Quantity Control

- **Navigate to the news feed or search results:** Go to the section where news articles are listed.
- **Use the quantity control (scroll bar):** Scroll through the list of articles using the provided quantity control, whether it's a scrollbar or another navigation tool.

The screenshot shows a scroll bar labeled 'Number of News:' with a range from 5 to 25. The slider is set to 10. Below the scroll bar is a button with a checkmark and the text 'Here are the some HEALTH News for you'.

Figure 18 Scroll Bar

The above Figure 18 is a scroll bar by using this we user can see news in the range of 5 to 25.

- **Scroll up or down:** Adjust the scroll direction to view more articles.

Step 9 : Closing the Web App

- **When you are finished:** Simply close the tab or window of the web browser to conclude your Duke Dunes session.

10.2 Database.

We are using google Firebase Database System , which can save the email and password of the user of Registration. So we don't need the local server to save the database.

The screenshot shows the Firebase Authentication console under the 'Project Overview' tab. On the left, there's a sidebar with 'Authentication' selected, along with other options like App Check, Realtime Database, Firestore Database, and Storage. Below that are sections for Product categories, Build, Release & Monitor, Analytics, and Engage, each with a dropdown arrow. At the bottom of the sidebar, it says 'Spark No-cost \$0/month' and 'Upgrade'. The main area is titled 'Authentication' and has tabs for Users, Sign-in method, Templates, Usage, Settings, and Extensions. A search bar at the top right says 'Search by email address, phone number, or user UID' with a magnifying glass icon. Below it is a table with columns: Identifier, Providers, Created, Signed In, and User UID. The table lists several users with their email addresses, creation dates, last sign-in dates, and unique user IDs.

Identifier	Providers	Created	Signed In	User UID
3@gmail.com	✉	Dec 4, 2023		L9k4vCYsIUWVdhuMgI7V7Ayr9Bz1
2@gmail.com	✉	Dec 4, 2023		QTsqUB4M2Zakoz7Vfa4ZexPdcYz1
1@gmail.com	✉	Dec 3, 2023		fYC4YYNqfsbRmRV9aiU2R5sf2AI2
new@gmail.com	✉	Dec 3, 2023		JkbOP9Z1u2fE5jQ5bOodNbl20N33
hello@gmail.com	✉	Dec 3, 2023		wdW6XM914iNzrrkqjoiUWQARvy..
shh@gmail.com	✉	Dec 3, 2023		5auNNFN1oxRwkzHMChcQ7k0B..
a@gmail.com	✉	Dec 3, 2023		n709h4mccRBszZwzJkUkzewc..
admin@gmail.com	✉	Dec 3, 2023		ZGrVpChG9sT6TKpnxtb0AmX6X..
h@gmail.com	✉	Dec 3, 2023		dna7CGb3AUW9CdFMize4usilN/93
ramesh1@gmail.com	✉	Dec 3, 2023		hxjPCxvkHEX2rEIL03l21VoViul

Figure 19 Database

The above Figure 19 is showing of Firebase Data Storage of users who had done login with the following credentials.

10.3 Features:

10.3.1 Recommended News

The screenshot shows a news application interface. On the left, there's a vertical sidebar with categories: HOME, ACCOUNT, BUSINESS (selected), ENTERTAINMENT, HEALTH, Nation, SCIENCE, SPORTS, TECHNOLOGY, and WORLD. The main content area has a heading 'Here are the some BUSINESS News for you' with a checkmark icon. Below it is a news card for 'MC Poll | MPC likely to hold rates again as inflation worries persist, say economists, bankers - Moneycontrol'. The card features the Reserve Bank of India logo. At the bottom of the card, there's a summary: 'experts highlighted that the central bank would maintain a withdrawal of accommodation stance in the Monetary Policy Committee (MPC) meeting. The Monetary Policy Committee (MPC) is likely to hold the repo rate at 6.5 percent in the upcoming December Monetary Policy, a Moneycontrol poll of 10 economists and bankers show. Additionally, experts highlighted that the central bank would maintain a withdrawal of accommodation stance in the Monetary Policy Committee (MPC) meeting. Inflation in India's July-September 2023 GDP growth stood at 7.6 percent versus 6.2 percent in July-September 2022. But after the surprising Q2 GDP growth numbers, experts said that the central bank will revise its GDP estimates.' There's also a link 'Read more at Moneycontrol...'.

Figure 20 Business of Recommended News

The above Figure 20 is a Business of Recommended News which fetch only business news or the users which is trending.

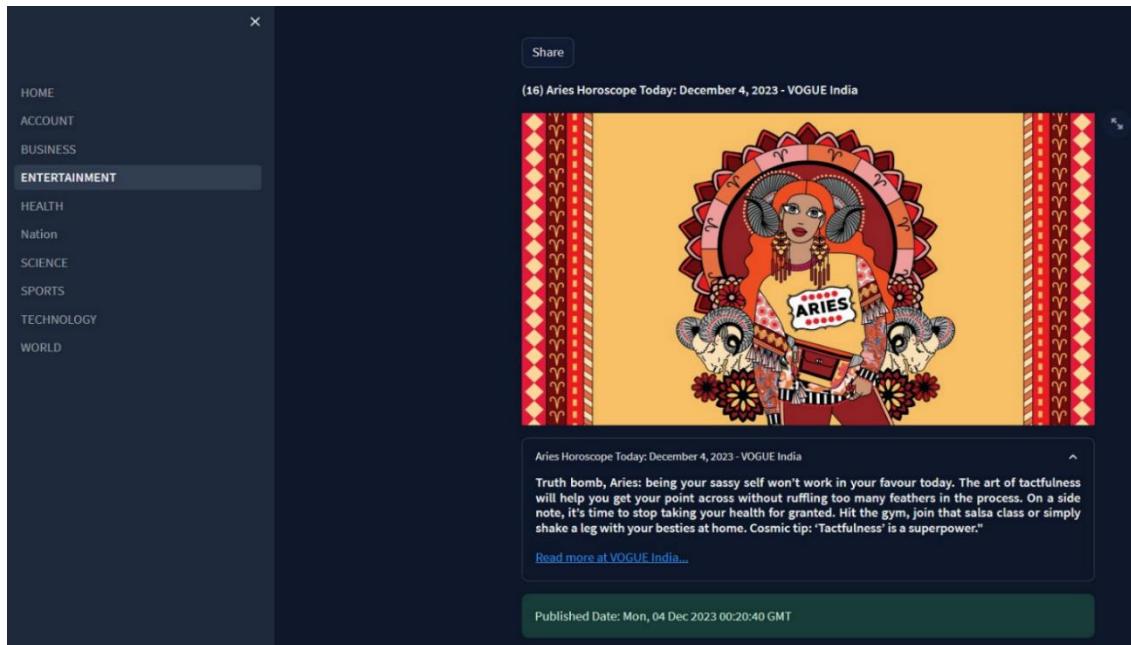


Figure 21 Entertainment of Recommended News

The above Figure 21 is an Entertainment of Recommended News which fetch the news from google API showcase the genre if entertainment which is trending or latest.

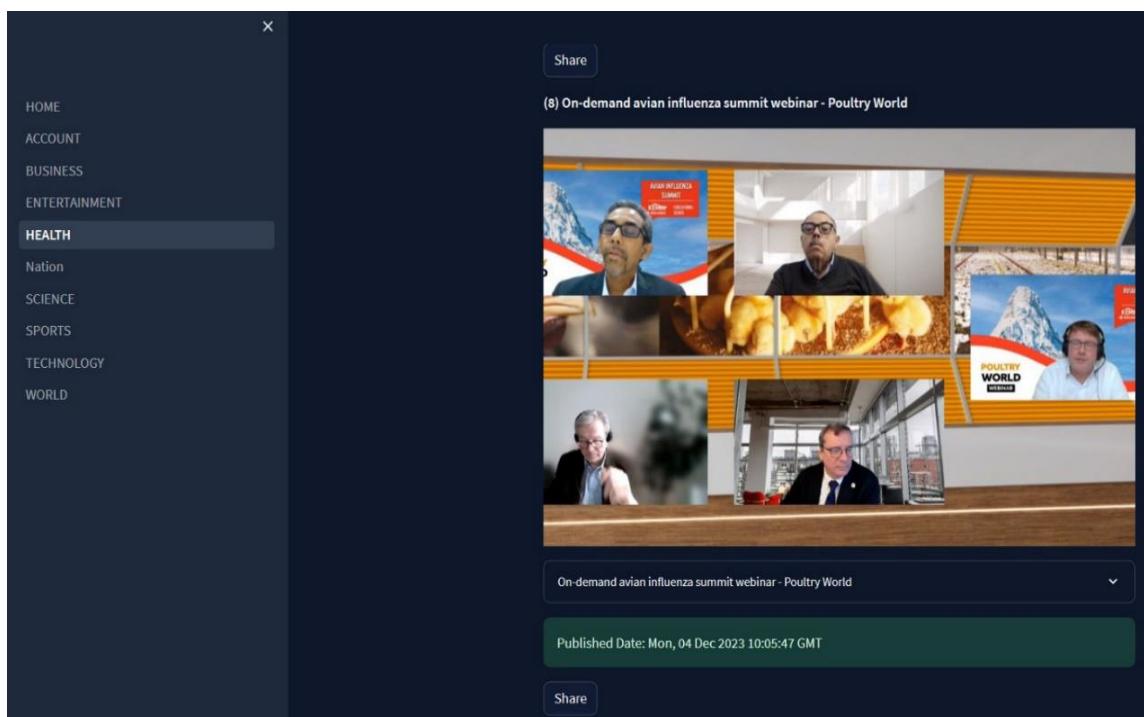


Figure 22 Health of Recommended News

The above Figure 22 is a Health of Recommended News which fetch the news from google API showcase the news of Health.

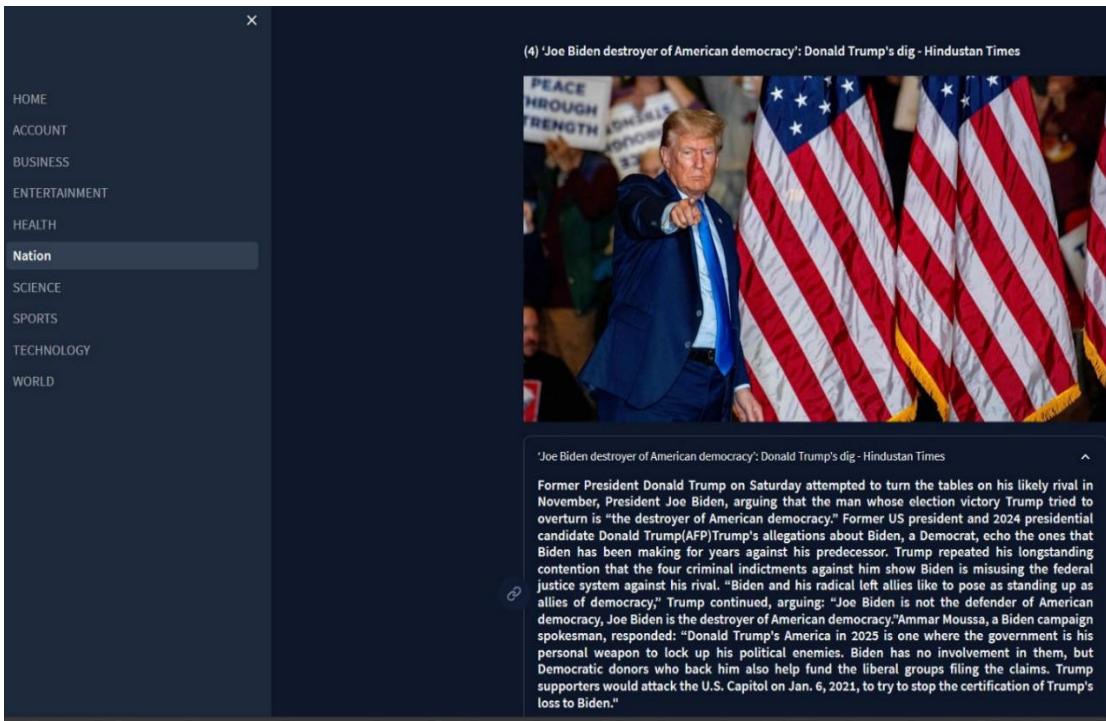


Figure 23 Nation of Recommended News

The above Figure 23 is a Nation of Recommended News which fetch the news from google API showcase the particular Nation News only.

10.4 News Summarization

1. Text Processing:

- **Text Extraction:** Gather the news articles from various sources.
- **Preprocessing:** Clean the text by removing irrelevant information like stop words, punctuation, and formatting.

2. Understanding the Content:

- **Text Analysis:** Utilize natural language processing (NLP) techniques to understand the context, topics, and sentiment of the articles.
- **Entity Recognition:** Identify important entities (names, locations, organizations) mentioned in the news.

3. Summarization Techniques:

- **Extractive Summarization:** Select important sentences or paragraphs directly from the original text that best represent the main ideas. This technique involves ranking sentences based on importance (e.g., using algorithms like Text Rank) and extracting top-ranked sentences to form the summary.
- **Abstractive Summarization:** Generate a summary by paraphrasing and synthesizing the main ideas using language models and AI techniques. This method creates new sentences that convey the gist of the original content in a more human-like manner.

4. Generating the Summary:

- **Summarization Algorithm:** Apply the chosen summarization technique to create a concise and coherent summary of the news articles.
- **Length and Readability:** Control the length of the summary and ensure readability for users.

5. Extractive Summarization:

Let's say we have a news article about recent advancements in renewable energy. Using extractive summarization, the system identifies the most important sentences based on their relevance and importance.

11. CONCLUSION AND FUTURE WORK

In summation, the development of Duke Dunes, a News Hub web application, represents a substantial achievement in the realm of digital information dissemination. Employing Streamlit for hosting, GitHub for version control, and Python for backend development, the project exhibits a harmonious integration of technologies. Notably, the utilization of Google Firebase for user authentication and data storage fortifies the application's security and scalability, while the integration of the Google API augments the user experience through the retrieval of diverse news content from multiple sources.

The strategic incorporation of Natural Language Processing (NLP) for news summarization introduces a layer of sophistication to the user interface. This implementation provides users with succinct and informative summaries, accompanied by essential details such as title, image, and publication date. The inclusion of recommended news topics, coupled with features like Trending News, Favourite Topics, Search News, and Quantity Control, collectively ensures a comprehensive and user-centric browsing experience.

11.1 Future Work:

In the future, the Duke Dunes project aims to integrate a personalized recommendation system to significantly enhance user experience [5]. This system will utilize individual preferences and behavior to provide tailored news recommendations, thereby increasing user engagement and satisfaction. Additionally, the project plans to introduce features for user engagement, including the incorporation of user comments and social media sharing functionalities. These enhancements will foster a more interactive and personalized experience for Duke Dunes users, allowing for greater user interaction and feedback.

ANNEXURE

1. NLP (Natural Language Processing):

NLP is a field of artificial intelligence (AI) that focuses on the interaction between computers and humans through natural language. It enables computers to understand, interpret, and generate human-like text.

Role in Project: In your project, NLP is utilized for news summarization, providing concise and meaningful summaries of news articles.

2. API (Application Programming Interface):

An API is a set of rules that allows one software application to interact with another. It defines the methods and data formats that applications can use to request and exchange information.

Role in Project: Your project leverages APIs, such as the Google News RSS API, for fetching news articles from various sources.

3. UI (User Interface):

UI refers to the visual elements and user interactions in a software application. It includes everything that a user interacts with on a digital device, such as buttons, screens, pages, and visual elements.

Role in Project: Streamlit is the UI framework used in your project, providing an intuitive and interactive user interface for the News Hub web app.

4. Firebase:

Firebase is a comprehensive mobile and web application development platform provided by Google. It includes various tools and services for authentication, real-time database, hosting, and more.

Role in Project: Firebase is used in your project for cloud storage, storing user registration details, and for user authentication.

5. NLTK (Natural Language Toolkit):

NLTK is a powerful Python library for working with human language data. It provides easy-to-use interfaces to linguistic resources and tools for processing and analysing text.

Role in Project: NLTK is employed in your project for Natural Language Processing tasks, enhancing the understanding and analysis of textual data.

6. Streamlit:

Streamlit is an open-source Python library used for creating web applications with minimal effort. It simplifies the process of turning data scripts into shareable web apps.

Role in Project: Streamlit serves as the framework for building the user interface and deploying the News Hub web app.

7. RSS (Really Simple Syndication):

RSS is a web feed that allows users and applications to access updates to websites in a standardized, computer-readable format. It typically contains a summary of web content along with links to full articles.

Role in Project: The Google News RSS API is utilized in your project to scrape and fetch news articles from various sources in a standardized format. [4]

8. Python:

Python is a high-level, general-purpose programming language known for its readability and simplicity. It is widely used in web development, data science, artificial intelligence, and more.

Role in Project: Python is the primary programming language used in your project for implementing the logic, data processing, and web application development.

ABOUT TOOLS AND TECHNOLOGY

Tools and Technologies Utilized in "Duke Dunes News Hub" Project:

Streamlit:

- **Description:** Streamlit is an open-source Python library designed to simplify the process of creating web applications. It allows developers to build interactive and data-driven apps with minimal code.
- **Purpose:** Streamlit serves as the primary platform for hosting and deploying the News Hub web application. Its simplicity and efficiency make it an ideal choice for rapid application development.

GitHub:

- **Description:** GitHub is a web-based version control platform that uses Git for tracking changes in source code. It provides collaboration features, such as bug tracking, feature requests, task management, and wikis, making it an essential tool for software development.
- **Purpose:** GitHub is employed for version control in the project, enabling the team to manage code changes, collaborate seamlessly, and maintain a centralized repository for the News Hub application.

Python:

- **Description:** Python is a versatile, high-level programming language known for its readability and extensive libraries. It supports multiple programming paradigms and is widely used for web development, data analysis, artificial intelligence, and more.
- **Purpose:** Python is the primary programming language chosen for developing both the backend and frontend components of the News Hub. Its simplicity and extensive ecosystem contribute to efficient development.

Firebase:

- **Description:** Firebase is a comprehensive mobile and web application development platform provided by Google. It offers a range of features, including real-time databases, authentication services, hosting, and more.
- **Purpose:** Firebase is utilized as the cloud database for storing user registration details and managing authentication processes within the News Hub.

Google API:

- **Description:** Google API provides access to various cloud-based services offered by Google. It allows developers to integrate Google functionalities into their applications, such as accessing data, maps, and authentication.
- **Purpose:** The Google API is employed to fetch news from diverse sources, enabling the News Hub to display up-to-date and relevant articles to users.

NLTK (Natural Language Toolkit):

- **Description:** NLTK is a powerful Python library for natural language processing (NLP) tasks. It provides tools and resources for tasks such as classification, tokenization, stemming, tagging, parsing, and more.
- **Purpose:** NLTK is implemented for NLP tasks within the News Hub, including news summarization. It enhances the understanding and processing of textual content.

Newspaper3k:

- **Description:** Newspaper3k is a Python library used for extracting articles, including text and images, from news websites. It simplifies the process of web scraping and article extraction.
- **Purpose:** Newspaper3k is employed for news summarization in the News Hub, extracting relevant information from articles and presenting it in a concise form.

Google News RSS API:

- **Description:** The Google News RSS API provides an interface to access news articles from Google News. It offers a structured way to retrieve news data in RSS format.
- **Purpose:** The Google News RSS API is used for scraping news articles from various categories, ensuring a diverse and comprehensive collection of news topics in the News Hub.

Other Libraries and Dependencies:

- **Description:** Various other Python libraries and dependencies are integrated into the project based on specific requirements. These may include libraries for data visualization, additional NLP tasks, or any functionality that enhances the overall capabilities of the web app.
- **Purpose:** These additional libraries and dependencies contribute to extending the functionality and features of the News Hub, ensuring a rich and dynamic user experience.

In summary, the combination of these tools and technologies empowers the development team to create a robust, feature-rich, and user-friendly News Hub web application.

REFERENCE

Figure 1 : <https://www.reikafujimura.com/post/python-news-content-aggregator-app-with-streamlit>

[1] : <https://www.reikafujimura.com/post/python-news-content-aggregator-app-with-streamlit>

[2]: <https://www.kaianalytics.com/post/how-to-use-text-analysis-techniques-bring-qualitative-data-to-life>

GitHub link Streamlit Source : <https://github.com/Spidy20/InNews>

Newspaper3k Article scraping & curation : <https://newspaper.readthedocs.io/en/latest/>

GitHub link for extracting newspaper : <https://github.com/codelucas/newspaper>

Article Extractor [3] : <https://github.com/grangier/python-goose>

Google News RSS Feed [4] : <https://rss.app/rss-feed/google-news-rss-feed>

Future work links [5] : News Recommendation System Based Personalized Needs:
<https://www.hindawi.com/journals/wcmc/2021/7072849/>

QR Code Generator Website : <https://me-qr.com/qr-code-generator/link>