

MINI PROJECT REPORT

ON

INSIGHTHUB

Submitted by

AYSWARYALAKSHMI R (SJC22CS038)

DEYONA SHAJI(SJC22CS051)

NAMITA JOLLY (SJC22CS090)

to

the APJ Abdul Kalam Technological University

in partial fulfillment of the requirements for the award of the degree

of

Bachelor of Technology

in

Computer Science and Engineering



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
ST. JOSEPH'S COLLEGE OF ENGINEERING AND TECHNOLOGY, PALAI
(AUTONOMOUS)

APRIL :: 2025

DECLARATION

We undersigned hereby declare that the mini project report on “**INSIGHTHUB**” submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by under supervision of **Prof. Renju Renjith**, Assistant Professor, Dept of CSE. This submission represents our ideas in our own words and where ideas and words of others have been included. We have adequately and accurately cited and referenced the original sources. We also declare that we have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in our submission. We understand that any violation of the above will be a course for disciplinary action by the institute and/or the University and can also evoke panel 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.

Name and Signature of Student

AYSWARYALAKSHMI R (SJC22CS038)

DEYONA SHAJI (SJC22CS051)

NAMITA JOLLY (SJC22CS090)

Place : Choondacherry

Date : 04-04-2025

**ST. JOSEPH'S COLLEGE OF ENGINEERING AND TECHNOLOGY,
PALAI
(AUTONOMOUS)**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
CERTIFICATE**

This is to certify that the report entitled "INSIGHTHUB" submitted by **AYSWARYA LAKSHMI R(SJC22CS038),DEYONA SHAJI(SJC22CS051),NAMITA JOLLY (SJC22CS090)**,to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering is a bonafide record of the mini project work carried out by them under my guidance and supervision.

Mini Project Guide

Prof. Renju Renjith
Assistant Professor
Department of CSE

Mini Project Coordinator

Prof. Bino Thomas
Assistant Professor
Department of CSE

Head of the Department

Dr. Joby P P
Professor & Head
Department of CSE

Place : Choondacherry
Date : 04-04-2025

ACKNOWLEDGEMENT

The success and final outcome of this mini project required a lot of guidance and assistance from many people and we are extremely privileged to have got this all along the completion of this project. All that we have done are only due to such supervision and assistance and we would not forget to thank them.

We respect and thank the management of **St. Joseph's College of Engineering and Technology** for providing us an opportunity and platform for doing the project work.

We are extremely thankful to our beloved **Principal, Dr. V.P Devassia** for providing such nice support and facilities to carry out this project.

We are extremely indebted to **Dr. Joby P.P, Professor & Head**, Department of Computer Science and Engineering for his valuable suggestions and encouragement during the course of this project work.

We would like to express our gratitude to our project coordinator **Prof. Bino Thomas**, Assistant Professor, Department of Computer Science and Engineering for his valuable suggestions and guidelines throughout the period of this project work. His contributions and technical support in preparing this report are greatly acknowledged.

We owe our deep gratitude to our project guide **Prof. Renju Renjith**, Assistant Professor, Department of Computer Science and Engineering who took keen interest in this project and guided us all along, till the completion of this project work by providing all the necessary information for developing a good system.

We are thankful and fortunate enough to get constant encouragement, support and guidance from all the staff members of the department of Computer Science and Engineering, which helped us in successfully completing our mini project work.

AYSWARYALAKSHMI R
DEYONA SHAJI
NAMITA JOLLY

Contents

Acknowledgement	i
List of Figures	v
Abbreviations	vi
Abstract	vii
1 Introduction	1
1.1 Problem Statement	1
1.2 Objectives and Scope	2
1.3 Methodology Overview	2
2 Literature Survey	4
2.1 Literature Referenced	4
2.1.1 Eyal et al. (2017) - Question Answering as an Automatic Evaluation Metric for News Article Summarization	4
2.1.2 Verma et al. (2024) - Automated Summarization and Question Generation for Efficient Information Extraction	5
2.1.3 Lelkes et al. (2020) - Quiz-Style Question Generation for News Stories	5
2.1.4 Anwar et al. (2023) – AI E-Learning in Secondary Education	6
2.1.5 Araki et al. (2016) – MCQ Generation via Semantic Analysis	7
2.1.6 Rao et al. (2020) – Automatic Multiple Choice Question Generation from Text: A Survey	8
2.1.7 Sundaramoorthy et al. (2017) – NewsOne: An Aggregation System for News Using Web Scraping Method	8

Contents

2.1.8	Sethi et al. (2017) – Automatic Text Summarization of News Articles	9
2.1.9	Sarr et al. (2018) – A Study on Automated Fact Extraction and Aggregation from Online Newspapers	10
2.1.10	Patel et al. (2020) – Abstractive Text Summarization on Google Search Results	10
2.1.11	Kumar et al. (2011) – Testbook	11
2.1.12	Author(s) et al. (2024) – AI-Generated (GPT-4) vs. Human-Crafted MCQs in Programming Education, ACE 2024	11
2.1.13	Mulla et al. (2023) – Automatic Question Generation: A Review of Methodologies, Datasets, Evaluation Metrics, and Applications .	12
2.1.14	Munjal et al. (2015) – Unacademy App	13
2.1.15	Chenyang Lyu et al. (2021) – Improving Unsupervised Question Answering via Summarization-Informed Question Generation, published in the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing	14
2.2	Existing Solutions	14
2.2.1	Testbook (2011) – Online Learning Platform by Testbook Edu Solutions	14
2.2.2	Unacademy (Unacademy, 2015) – Online Learning Platform for Competitive Exams and Skill Development	15
3	System Study	16
3.1	Technologies used	16
3.1.1	Python	16
3.1.2	Flask	16
3.1.3	BeautifulSoup	16
3.1.4	OpenAI GPT AP	17
3.1.5	MongoDB	17
3.1.6	HTML, CSS, and JavaScript	17
3.1.7	Jinja2	17
3.1.8	WTForms	18
3.2	Software Components	19

Contents

3.2.1	Flask	19
3.2.2	MongoDB	19
3.2.3	HTML, CSS, and JavaScript	19
3.2.4	Jinja2	19
3.2.5	NewsAPI	19
3.2.6	OpenAI API	20
3.2.7	Bcrypt	20
3.2.8	Flask-Limiter	20
3.2.9	WTForms	20
3.3	Hardware Requirements	21
3.3.1	Multi-core CPU	21
3.3.2	Minimum 8GB RAM	21
3.3.3	SSD Storage (256GB or more)	21
3.3.4	Stable Internet Connection	21
4	System Requirement Specification	22
4.1	Functional Requirements	22
4.1.1	User Authentication & Role-Based Access	22
4.1.2	News Fetching & Summarization	22
4.1.3	Personalized Recommendations	22
4.1.4	Quiz & Mock Tests	23
4.1.5	News Search & Filtering	23
4.1.6	Real-Time News Updates	23
4.2	Non Functional Requirements	23
4.2.1	Secure & Scalable Backend	23
4.2.2	Fast & Efficient News Summarization	24
4.2.3	AI Integration	24
5	System Design	25
5.1	Activity Diagram	25
5.2	Sequence Diagram	27
5.3	ER Diagram	28

Contents

5.4	Use Case Diagram	28
6	System Implementation	30
6.1	Modules	30
6.1.1	User authentication	30
6.1.2	News Aggregation & Summarization	30
6.1.3	Personalized Recommendations	31
6.1.4	Quiz & Mock Test	31
7	Experimental Results	32
7.1	Results and Discussions	32
7.2	System Execution Output	33
	Conclusion	50
	References	51
	Annexure	54
	Annexure A	54

List of Figures

5.1	Activity diagram	26
5.2	Sequence Diagram	27
5.3	ER diagram	28
5.4	Use case diagram	29
7.1	Welcome Page	33
7.2	Login Page	34
7.3	Register Page	35
7.4	Dashboard page	36
7.5	Latest News Page	37
7.6	News By Date	38
7.7	Categorized News	39
7.8	Daily Quiz	40
7.9	Daily Quiz Result	41
7.10	Mock Test	42
7.11	Mock test result	43
7.12	User Database	44
7.13	Articles Database	45
7.14	Mock Results Database	46
7.15	Daily Quiz Results Database	47
7.16	Register Page	48
7.17	Dashboard page	49

Abbreviations

AI	-	Artificial Intelligence
API	-	Application Programming Interface
UI	-	User Interface
NFA	-	News Feed Aggregation
MCQ	-	Multiple Choice Questions
BBC	-	British Broadcasting Corporation

Abstract

In an era where information is abundant but time is limited, staying updated with current affairs is a challenge, especially for competitive exam aspirants. Traditional news sources can be overwhelming, requiring extensive reading to extract relevant insights. InsightHub addresses this issue by leveraging web scraping and AI-driven analysis to provide concise, section-wise news summaries tailored to users' interests. The platform ensures accessibility to trustworthy updates without the burden of extensive reading. Additionally, InsightHub integrates an automated quiz and mock test generator, enabling users to reinforce their knowledge through interactive assessments. The methodology involves extracting news from multiple reliable sources, filtering and categorizing information and dynamically generating quizzes based on the latest updates. By combining real-time data processing with intelligent learning tools, InsightHub not only enhances awareness but also optimizes exam preparation. The results indicate improved retention, engagement, and efficiency in learning. This project demonstrates the potential of AI and automation in simplifying information consumption and transforming the way users prepare for knowledge-based assessments.

Chapter 1

Introduction

In today's world, staying updated with credible news while managing a busy schedule can be challenging. InsightHub streamlines this by offering AI-driven, real-time news summaries, categorizing articles into sections like politics, business, technology, and current affairs. This ensures users receive concise, relevant updates without spending hours browsing multiple sources.

Beyond just news, InsightHub serves as a powerful learning platform. It features interactive daily quizzes and structured weekly mock tests, enabling aspirants to sharpen their knowledge for competitive exams like UPSC, SSC, and banking. Users can track their progress, compare answers with correct solutions, and refine their preparation strategies. AI-driven recommendations personalize content, adapting to users' interests and study patterns.

With its seamless integration of real-time updates, structured learning, and intelligent recommendations, InsightHub transforms news consumption into an engaging and efficient experience. Future advancements will introduce adaptive quizzes, deeper analytics, and expanded coverage, making it a comprehensive knowledge hub for both casual readers and dedicated exam aspirants.

1.1 Problem Statement

Exam aspirants often struggle with scattered information and unreliable sources, as traditional news platforms provide lengthy articles while exam prep sites lack timely updates.

1.2. Objectives and Scope

Our platform solves this by integrating AI-driven news summaries, categorized updates, and structured quizzes in one place.

It curates credible news, applies AI summarization, and personalizes recommendations based on user interests. Daily quizzes and weekly mock tests help users track performance, compare answers, and receive instant feedback. With AI-powered insights and adaptive learning features, the platform ensures efficient and structured exam preparation.

1.2 Objectives and Scope

This web app provides AI-driven, summarized news and structured content for both general users and exam aspirants. It categorizes news into sections like politics, business, and sports, ensuring quick access to relevant updates. Integrated mock tests and quizzes help aspirants prepare for competitive exams efficiently.

Users benefit from daily quizzes, weekly mock tests, and AI-powered recommendations tailored to their reading and learning patterns. Instant feedback, performance analytics, and comparison with correct answers enhance exam readiness. The platform ensures credible, up-to-date information, making learning structured, personalized, and effective.

1.3 Methodology Overview

The platform bridges the gap between news consumption and exam preparation by offering AI-powered news summaries, categorized updates, and structured learning tools. It extracts news from trusted sources, applies AI-driven summarization, and organizes articles into relevant categories such as politics, technology, business, and sports. Machine learning-based recommendations ensure that users receive personalized content aligned with their interests, reading history, and past interactions, making the platform highly adaptive and user-centric.

In addition to delivering real-time news updates, the platform provides interactive learning tools such as daily quizzes and weekly mock tests, helping users stay informed and prepared for competitive exams like UPSC, SSC, and banking tests. The quizzes are designed to assess users' understanding of current affairs and general knowledge,

1.3. Methodology Overview

with structured question formats that range from basic multiple-choice questions to more complex analytical problems. Users can attempt quizzes, analyze their progress over time, and refine their preparation strategies based on the insights gained.

By integrating AI-driven summarization with structured learning tools, the platform not only enhances accessibility to credible news but also ensures that users engage with relevant and well-organized content. The combination of news summaries, categorized updates, and exam-focused quizzes makes the platform an all-in-one solution for individuals looking to stay informed while efficiently preparing for exams. With its user-friendly interface, real-time updates, and intelligent content delivery, the platform transforms traditional news consumption into an interactive and educational experience, catering to both general readers and dedicated aspirants.

Chapter 2

Literature Survey

2.1 Literature Referenced

2.1.1 Eyal et al. (2017) - Question Answering as an Automatic Evaluation Metric for News Article Summarization

This study explores the use of question answering (QA) as an automatic evaluation metric for news article summarization, aiming to assess summary quality based on how well it retains key information. The researchers employ a QA-based evaluation framework, where a model generates questions from the original article and then attempts to answer them using the summary. The performance of the QA system determines the informativeness and faithfulness of the summary. The QA-based metric correlates strongly with human evaluation and outperforms traditional summarization metrics like ROUGE, demonstrating its potential as a reliable automatic summarization evaluation method. This study introduces a more contextual and content-aware evaluation metric that focuses on information retention rather than surface-level text matching, improving the assessment of generated summaries. However, the QA-based approach depends on the quality of the question generation model and the availability of robust QA systems. Future work could explore multi-document summarization evaluation using this technique. InsightHub can adopt QA-based evaluation metrics to assess the accuracy of its AI-generated news summaries, ensuring that the platform delivers concise yet information-rich summaries to users.

2.1. Literature Referenced

2.1.2 Verma et al. (2024) - Automated Summarization and Question Generation for Efficient Information Extraction

This study develops an NLP-based system for automated text summarization and question generation to enhance information extraction efficiency. The researchers implement NLP-based techniques, including extractive and abstractive summarization models and question generation frameworks, to generate meaningful questions from summarized text. The proposed system successfully improves information retrieval efficiency, ensuring that concise summaries retain essential details while generating relevant questions that aid in comprehension and knowledge assessment. The study integrates text summarization and question generation into a unified system, enabling better automated learning and information extraction across different domains. However, the system's performance relies on the quality of summarization models, and question coherence needs further improvement. Future research should focus on context-aware question generation for better accuracy. InsightHub can adopt this NLP-based system to enhance the accuracy and informativeness of its AI-generated news summaries, ensuring that users receive well-structured and insightful content.

2.1.3 Lelkes et al. (2020) - Quiz-Style Question Generation for News Stories

This study aims to develop an automated system for generating quiz-style questions from news articles to enhance reader engagement and knowledge retention. The authors use Natural Language Processing (NLP) techniques, including dependency parsing, named entity recognition (NER), and transformer-based models, to extract key information from news stories and generate multiple-choice questions (MCQs). The research shows that quiz-style question generation enhances user engagement with news content. The system effectively identifies important details from articles and formulates coherent and contextually relevant MCQs. The study provides a structured pipeline for generating high-quality MCQs from dynamic news articles, improving the relevance of questions by leveraging entity-based information extraction and advanced NLP models. However, the approach is primarily focused on fact-based questions and struggles with complex or opinion-based

2.1. Literature Referenced

content. It also lacks adaptability for generating open-ended or multi-layered reasoning questions. This research aligns well with InsightHub’s goal of integrating AI-driven news summarization with interactive quizzes. Implementing similar techniques can improve question quality, making news-based learning more engaging and effective.

2.1.4 Anwar et al. (2023) – AI E-Learning in Secondary Education

This study aims to develop an automated system for generating quiz-style questions from news articles to enhance reader engagement and knowledge retention. The authors use Natural Language Processing (NLP) techniques, including dependency parsing, named entity recognition (NER), and transformer-based models, to extract key information from news stories and generate multiple-choice questions (MCQs). The research shows that quiz-style question generation enhances user engagement with news content. The system effectively identifies important details from articles and formulates coherent and contextually relevant MCQs. The study provides a structured pipeline for generating high-quality MCQs from dynamic news articles, improving the relevance of questions by leveraging entity-based information extraction and advanced NLP models. However, the approach is primarily focused on fact-based questions and struggles with complex or opinion-based content. It also lacks adaptability for generating open-ended or multi-layered reasoning questions. This research aligns well with InsightHub’s goal of integrating AI-driven news summarization with interactive quizzes. Implementing similar techniques can improve question quality, making news-based learning more engaging and effective. Additionally, a related study on AI-driven e-learning in secondary education explores the integration of AI-based adaptive learning platforms, leveraging machine learning models, natural language processing, and recommendation algorithms to tailor educational content to students’ needs. It found that AI-powered e-learning systems significantly improve student comprehension, engagement, and retention through personalized learning paths and real-time feedback. The research highlights AI’s potential in revolutionizing secondary education by providing customized learning experiences, automated assessments, and intelligent tutoring systems that adapt to individual progress. However, challenges include the need for high-quality training data, potential biases in AI models, and the require-

2.1. Literature Referenced

ment for substantial computational resources. Teacher training and student adaptability to AI-based learning also remain key concerns. This study aligns with InsightHub’s objective of integrating AI for personalized news summarization and quiz-based learning, where adaptive AI models can enhance content recommendations, making learning more interactive and effective.

2.1.5 Araki et al. (2016) – MCQ Generation via Semantic Analysis

This study explores AI-driven quiz-style question generation from news articles to enhance engagement and knowledge retention. Using NLP techniques like dependency parsing, named entity recognition (NER), and transformer-based models, the system extracts key details and generates multiple-choice questions (MCQs). Findings show that quiz-style question generation improves user interaction with news content, ensuring coherence and relevance. However, it primarily focuses on fact-based questions and lacks adaptability for opinion-based or complex reasoning questions. This aligns with InsightHub’s goal of integrating AI-driven news summarization with interactive quizzes to enhance learning. A related study on AI-driven e-learning in secondary education examines adaptive learning platforms that use machine learning and recommendation algorithms to personalize educational content. AI-powered systems significantly improve student comprehension, engagement, and retention through real-time feedback. Challenges include biases in AI models, high computational requirements, and the need for teacher training. This research supports InsightHub’s objective of AI-enhanced personalized news summarization and quiz-based learning. Another study on MCQ generation via semantic analysis focuses on automating question creation using NLP techniques, semantic similarity measures, and linguistic analysis. The system generates contextually relevant MCQs with semantically appropriate distractors, improving assessment quality. While it enhances question coherence and difficulty balancing, limitations include handling ambiguous sentences and ensuring originality. This study informs InsightHub’s approach to using AI-driven semantic analysis for generating high-quality MCQs from news summaries, boosting engagement and knowledge retention.

2.1.6 Rao et al. (2020) – Automatic Multiple Choice Question Generation from Text: A Survey

This study explores AI-driven quiz-style question generation from news articles using NLP techniques like dependency parsing, named entity recognition (NER), and transformer-based models. Findings show improved user engagement but highlight limitations in handling opinion-based or complex reasoning questions. This aligns with InsightHub’s goal of integrating AI-driven news summarization with interactive quizzes. A related study on AI-driven e-learning highlights adaptive learning platforms that enhance student comprehension through machine learning and recommendation algorithms. Challenges include AI biases, computational demands, and the need for teacher training. This supports InsightHub’s focus on personalized news summarization and quiz-based learning. Another study on MCQ generation via semantic analysis automates question creation using NLP and semantic similarity measures. While it improves assessment quality, challenges include handling ambiguous sentences and ensuring originality. This informs InsightHub’s approach to AI-driven MCQ generation. A survey on automatic MCQ generation reviews rule-based, machine learning, and deep learning techniques, emphasizing deep learning’s role in improving accuracy. Challenges include complex sentence structures and diverse question formats. InsightHub can use these insights to refine its AI-driven quiz feature for high-quality assessments.

2.1.7 Sundaramoorthy et al. (2017) – NewsOne: An Aggregation System for News Using Web Scraping Method

This study explores AI-driven quiz-style question generation from news articles using NLP techniques, improving user engagement but struggling with opinion-based questions. This supports InsightHub’s AI-driven news summarization and quizzes. A related study on AI-driven e-learning highlights adaptive learning platforms that enhance comprehension through machine learning. Challenges include AI biases and teacher training needs, aligning with InsightHub’s focus on personalized learning. Research on MCQ generation via semantic analysis improves automated assessments but faces issues with ambiguity and originality, informing InsightHub’s quiz development. A survey on automatic MCQ genera-

2.1. Literature Referenced

tion emphasizes deep learning's role in accuracy while highlighting challenges like complex sentence structures. InsightHub can refine its quiz feature using these insights. NewsOne research shows web scraping's effectiveness for real-time news aggregation but faces legal and NLP challenges. InsightHub can adopt similar techniques for improved news updates and content organization. study on automated fact extraction and aggregation from on-line newspapers highlights NLP's role in improving information retrieval and verification by cross-referencing multiple sources. Challenges include misinformation detection and bias handling. InsightHub can leverage these techniques to strengthen its fact-checking module.

2.1.8 Sethi et al. (2017) – Automatic Text Summarization of News Articles

Sethi et al. (2017) present a lexical chain-based extractive summarization technique using WordNet, pronoun resolution, and sentence scoring. The approach builds lexical chains, which group related words to improve topic detection. Sentences are then scored based on their contribution to these chains, ensuring higher coherence. Additionally, first-sentence priority is applied, assuming that the opening sentence typically holds significant contextual information. The method offers several advantages. Lexical chains enhance topic coherence, ensuring the summary remains contextually connected. Sentence scoring improves information retention, making the summary meaningful. First-sentence priority further refines the extractive process by selecting key sentences early in the document, boosting relevance. However, the approach has limitations. It focuses only on extractive summarization, meaning it cannot generate new phrases or restructure sentences. Manual threshold tuning is required to determine the importance of lexical chains, adding complexity. Additionally, the method struggles with proper noun handling, potentially impacting accuracy when dealing with named entities. Future enhancements could include abstractive summarization techniques, adjective-based lexical chains for better descriptive analysis, and graph-based sentence extraction to refine selection criteria and improve overall summarization quality.

2.1.9 Sarr et al. (2018) – A Study on Automated Fact Extraction and Aggregation from Online Newspapers

Sarr et al. (2018) explore automated fact extraction and aggregation from online newspapers using web scraping for data collection. Extracted data is stored in structured datasets. The implementation includes data segmentation and fusion, followed by processing and cleaning, where a language detection mechanism ensures accuracy. Key advantages include efficient data extraction, a streamlined information-gathering process, and language processing accuracy, enhancing the reliability of extracted data. Structured storage supports fact-checking mechanisms, providing an organized format for verification. The system is scalable, allowing the integration of additional sources as needed. However, limitations exist, such as restricted language support, challenges in sentence segmentation, dependency on web structures, and the lack of real-time fact-checking. These constraints impact the system's robustness across diverse contexts. Future research aims to expand language support, integrate AI for fact verification, enable real-time web scraping, and collaborate with fact-checking organizations. These improvements will enhance reliability, broaden applicability, and strengthen automated fact-checking systems on a larger scale.

2.1.10 Patel et al. (2020) – Abstractive Text Summarization on Google Search Results

This study explores abstractive text summarization for Google search results using NLP and deep learning techniques, such as sequence-to-sequence models and attention mechanisms. Unlike extractive methods that simply pull key sentences from content, abstractive summarization generates human-like, contextually aware summaries, improving information retrieval efficiency. By condensing search results into concise and coherent summaries, this approach enhances user experience, making it easier to grasp essential insights quickly. However, the study highlights challenges such as factual inconsistencies and minor hallucinations, where the model may generate inaccurate or misleading information. These limitations emphasize the need for further fine-tuning and improved validation mechanisms to enhance factual accuracy and reduce distortion. Future advancements should focus on refining model training, improving data sources, and integrating real-time updates

2.1. Literature Referenced

for better reliability. The findings from this research offer valuable insights for platforms like InsightHub, which can implement similar abstractive summarization techniques to enhance its news aggregation module. By adopting AI-driven summarization, InsightHub can provide users with more concise, coherent, and insightful news updates from multiple sources, ensuring a more efficient and user-friendly news consumption experience.

2.1.11 Kumar et al. (2011) – Testbook

This study explores Testbook, an AI-powered learning and assessment platform that provides structured test preparation through automated question generation, adaptive testing, and real-time performance tracking. By leveraging AI and data analytics, the platform personalizes learning experiences, helping users improve exam readiness efficiently. Automated test generation and adaptive learning techniques allow for targeted practice, ensuring that users focus on areas where they need improvement. The study highlights the effectiveness of AI-driven assessments but also identifies challenges such as the difficulty in generating high-quality subjective questions, limited contextual understanding in automated question creation, and adaptability issues for diverse learning styles. These limitations suggest the need for improved AI-driven personalization, enhanced validation mechanisms, and better content structuring. Future developments should focus on refining question quality, expanding adaptability, and improving feedback mechanisms. The insights from this research are valuable for platforms like InsightHub, which can integrate AI-driven question generation and adaptive testing to enhance its quiz and mock test module. By adopting these techniques, InsightHub can provide more engaging, efficient, and personalized assessments, improving user experience and learning outcomes while ensuring high-quality content and accurate performance tracking.

2.1.12 Author(s) et al. (2024) – AI-Generated (GPT-4) vs. Human-Crafted MCQs in Programming Education, ACE 2024

This study compares AI-generated multiple-choice questions (MCQs) using GPT-4 with human-crafted MCQs in programming education, analyzing their effectiveness and im-

2.1. Literature Referenced

pact on student learning. By employing expert reviews, student performance metrics, and qualitative feedback, the research evaluates question quality and pedagogical value. Findings suggest that AI-generated MCQs are comparable in quality to human-made ones, excelling in speed and scalability. However, human-crafted MCQs offer better contextual relevance and pedagogical depth. The study emphasizes AI's role in automating question generation, reducing educators' workload, and making learning resources more accessible. Despite these advantages, challenges persist, such as inconsistencies in question difficulty, lack of deep conceptual understanding, and biases in AI-generated content. Further research is necessary to refine AI-driven MCQ generation, ensuring better question validation and adaptability across different learning levels. Platforms like InsightHub can integrate AI-generated MCQs into their quiz and mock test modules, benefiting from automation while maintaining quality control through human review. This approach can enhance scalability, improve learning outcomes, and provide users with diverse, high-quality assessments tailored to their educational needs.

2.1.13 Mulla et al. (2023) – Automatic Question Generation: A Review of Methodologies, Datasets, Evaluation Metrics, and Applications

This study reviews methodologies, datasets, evaluation metrics, and applications of automatic question generation (AQG), focusing on improving educational assessments and intelligent tutoring systems. It explores various NLP techniques, including rule-based, statistical, and deep learning approaches, to generate high-quality questions. Findings suggest that deep learning-based methods, such as transformer models, improve the fluency and relevance of generated questions, while traditional methods like TF-IDF and LexRank excel in factual accuracy. A hybrid approach combining these techniques enhances question coherence and informativeness. The study also highlights challenges such as ensuring grammatical correctness, maintaining semantic consistency, and generating diverse question types suitable for different educational contexts. Additionally, it identifies gaps in dataset quality and evaluation metrics, emphasizing the need for more standardized benchmarks. Future research should focus on refining AI models to address

2.1. Literature Referenced

these limitations and improve adaptability across subjects. InsightHub can leverage AQG techniques to enhance its quiz and assessment modules, ensuring automated, high-quality question generation that improves learning efficiency and engagement. By integrating these advancements, InsightHub can provide personalized assessments, reducing manual effort for educators while maintaining the accuracy and contextual relevance of generated questions.

2.1.14 Munjal et al. (2015) – Unacademy App

The Unacademy app was developed to make quality education more accessible through an online platform, primarily catering to students preparing for competitive exams and skill development. It employs video-based learning, AI-driven content recommendations, live interactive classes, and test series to create a personalized learning experience. Adaptive learning models track student progress and suggest relevant courses, improving engagement and knowledge retention. Features like structured learning paths, mentorship programs, and peer discussions further enhance interactive learning. The app has successfully democratized education, making high-quality resources available to a diverse student base. However, challenges remain, including the lack of offline accessibility and variations in course quality due to multiple educators. Future enhancements should focus on standardizing content and expanding multilingual support. InsightHub can adopt AI-driven personalized learning techniques and interactive quizzes similar to Unacademy to make news-based quizzes more engaging and effective, ensuring users gain valuable insights efficiently.

2.2. Existing Solutions

2.1.15 Chenyang Lyu et al. (2021) – Improving Unsupervised Question Answering via Summarization-Informed Question Generation, published in the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing

This study explores improving unsupervised question answering (QA) by leveraging summarization informed question generation (QG) techniques, enhancing the quality and relevance of generated questions. The approach first condenses lengthy text using summarization before applying unsupervised QG models, reducing noise and improving accuracy. Findings indicate that this method leads to better question relevance, improved answerability, and reduced redundancy compared to traditional QG methods, as evidenced by higher BLEU and ROUGE scores. The research introduces a novel integration of summarization and question generation, optimizing QA pipelines without reliance on labeled datasets. However, challenges include potential loss of crucial information during summarization, leading to mismatches between questions and answers. Future advancements should focus on adaptive summarization models that preserve answer-relevant details. InsightHub can apply summarization-driven QG techniques to refine its quiz module, ensuring that questions generated from news articles remain concise, engaging, and relevant for users.

2.2 Existing Solutions

2.2.1 Testbook (2011) – Online Learning Platform by Testbook Edu Solutions

Testbook is an online learning platform designed to help aspirants prepare for competitive exams by offering study materials, mock tests, and live coaching. It employs AI-driven personalized learning recommendations and adaptive testing to optimize study plans, along with gamification and real-time performance tracking to enhance learning outcomes. The platform provides a diverse range of courses and interactive learning features, such as live

2.2. Existing Solutions

classes, doubt-clearing sessions, and quizzes, ensuring student engagement. Additionally, its scalability makes quality education accessible to a broad audience. However, challenges include variations in study material quality, the necessity of paid subscriptions for premium features, and the limited availability of hands-on learning for certain subjects. Despite these limitations, Testbook's AI-powered content recommendation system and structured test preparation approach align well with InsightHub's goals. InsightHub can leverage similar AI-driven techniques to enhance its personalized news summaries and quizzes, offering an engaging and structured learning experience for users.

2.2.2 Unacademy (Unacademy, 2015) – Online Learning Platform for Competitive Exams and Skill Development

Unacademy is an online learning platform designed to democratize education by offering accessible, high-quality learning resources. It provides courses for competitive exams, skill development, and academic subjects through interactive video lectures, live classes, and quizzes. The platform integrates AI-driven personalized recommendations, peer discussions, and mentorship programs, significantly enhancing learner engagement and performance. With a vast network of educators and structured curricula, Unacademy has successfully expanded access to education. However, challenges include variations in content quality, limited support for subjects requiring practical application, and restricted access to premium features due to subscription costs. Despite these limitations, its interactive learning approach and AI-powered recommendations make it a valuable educational tool. InsightHub can adopt similar strategies by integrating interactive quizzes and AI-driven personalized content delivery to improve user engagement and enhance the learning experience within its news-based platform.

Chapter 3

System Study

3.1 Technologies used

3.1.1 Python

Python forms the core of InsightHub's backend development. It is used extensively for handling logic, performing automated web scraping, and driving the AI-based summarization features of the platform.

3.1.2 Flask

Flask, a lightweight and flexible Python web framework, powers the server-side of the application. It manages routing, HTTP requests, API integrations, and interactions between the backend and frontend components.

3.1.3 BeautifulSoup

BeautifulSoup is used for web scraping, allowing the system to extract structured news data from various sources. It helps parse HTML content efficiently and forms the foundation for automating news retrieval.

3.1. Technologies used

3.1.4 OpenAI GPT AP

The OpenAI GPT API plays a dual role in the platform. It is used for AI-based summarization of lengthy articles, providing concise and accurate summaries, and also for generating quiz questions based on current news events to aid in exam preparation.

3.1.5 MongoDB

MongoDB is used as the primary database for the platform. As a NoSQL database, it offers flexibility and scalability to store diverse data types including scraped articles, user profiles, summaries, and quiz results.

3.1.6 HTML, CSS, and JavaScript

HTML and CSS form the structure and styling of the user interface, ensuring responsiveness and accessibility. JavaScript adds interactivity, enabling dynamic updates and seamless user experience on the client side.

3.1.7 Jinja2

Jinja2 is used as the templating engine in Flask to dynamically render HTML content. It integrates backend data with frontend views, enabling real-time updates and personalized content delivery.

3.1. Technologies used

3.1.8 WTForms

WTForms ensures secure handling of user inputs and validations in the platform's forms. It plays a crucial role in account registration, login, and interactive features like feedback and quiz submissions.

3.2 Software Components

3.2.1 Flask

Flask is a lightweight Python web framework used for handling routing and managing server-side logic. In InsightHub, it acts as the core backend framework, processing user requests, connecting with the database, and rendering appropriate responses.

3.2.2 MongoDB

MongoDB is a NoSQL database used for storing and managing news articles, user data, and quiz information. Its flexible document structure allows for efficient data handling and easy scalability, making it ideal for dynamic content like real-time news.

3.2.3 HTML, CSS, and JavaScript

These frontend technologies are used to build the user interface of the application. HTML structures the web pages, CSS styles them for responsiveness and visual appeal, and JavaScript adds interactivity, allowing users to seamlessly navigate and interact with the platform.

3.2.4 Jinja2

Jinja2 is the templating engine integrated with Flask. It dynamically generates HTML content by injecting backend data into the frontend, enabling real-time content updates and ensuring a smooth user experience.

3.2.5 NewsAPI

NewsAPI is employed to fetch the latest news articles from multiple reliable sources. This API ensures that users always have access to up-to-date, relevant news.

3.2.6 OpenAI API

The OpenAI API is used for AI-driven summarization. It takes lengthy news articles and condenses them into concise, informative summaries, making it easier for users to consume essential information quickly and efficiently.

3.2.7 Bcrypt

Bcrypt is used to securely hash and store user passwords. By encrypting sensitive user data, it provides protection against unauthorized access and enhances the overall security of the platform.

3.2.8 Flask-Limiter

Flask-Limiter is a security component that helps prevent abuse by limiting the number of requests a user can make in a given timeframe. This protects the server from being overwhelmed and safeguards the application from brute-force attacks.

3.2.9 WTForms

WTForms is a flexible forms validation and rendering library for Python web development. In InsightHub, it is used to validate user inputs from forms, ensuring data integrity and preventing malicious or malformed data from entering the system.

3.3 Hardware Requirements

3.3.1 Multi-core CPU

A multi-core processor such as an Intel i5 or higher is recommended to ensure efficient handling of user requests, backend operations, and real-time data processing. It improves the overall responsiveness of the system.

3.3.2 Minimum 8GB RAM

At least 8GB of RAM is required to maintain smooth performance, particularly during periods of high user traffic or when running background tasks like summarization and database queries.

3.3.3 SSD Storage (256GB or more)

Solid State Drive (SSD) storage with a minimum capacity of 256GB is recommended for fast data access, quicker application startup, and efficient database read/write operations.

3.3.4 Stable Internet Connection

A reliable and fast internet connection is essential for consistent API communication, real-time news updates, and uninterrupted service delivery to end users.

Chapter 4

System Requirement Specification

4.1 Functional Requirements

4.1.1 User Authentication & Role-Based Access

Users can register and log in securely, ensuring data protection and personalized access. Role-based access control allows users to view news and participate in quizzes based on their assigned roles. Password encryption enhances security by safeguarding user credentials from unauthorized access.

4.1.2 News Fetching & Summarization

The system automatically fetches news articles from trusted sources, ensuring up-to-date and reliable information. AI-based text summarization generates concise summaries for quick readability. News articles are categorized into topics such as politics, sports, and business for better organization. All fetched news is stored in a database, allowing easy access and retrieval.

4.1.3 Personalized Recommendations

Learning algorithms intelligently analyze user preferences, reading history, and engagement patterns to deliver highly relevant and personalized news recommendations. By continuously adapting to individual behavior, the system ensures a tai-

4.2. Non Functional Requirements

lored newsfeed. Additionally, it identifies and highlights trending topics and breaking stories, keeping users informed about the most important and popular updates in real time.

4.1.4 Quiz & Mock Tests

Users can take quizzes based on daily news and participate in weekly mock tests to assess their knowledge. The system tracks user performance and maintains a history of attempts for progress analysis. Quizzes offer different difficulty levels and categorized questions to match user preferences. Additionally, quiz questions are dynamically generated from news content, ensuring relevance and up-to-date assessments.

4.1.5 News Search & Filtering

Users can search for specific news articles, allowing quick access to relevant information. Filters can be applied based on the publication date, enabling users to refine their search and focus on recent or past news as needed.

4.1.6 Real-Time News Updates

The system automatically updates news every few hours, ensuring users receive the latest and most relevant information in real time.

4.2 Non Functional Requirements

4.2.1 Secure & Scalable Backend

The backend is built using Flask and MongoDB, ensuring high performance and flexibility for handling large volumes of news data. Security measures such as data encryption and authentication are implemented to protect user information and maintain system integrity.

4.2. Non Functional Requirements

4.2.2 Fast & Efficient News Summarization

AI-powered summarization generates concise news summaries, making information easily digestible for users. The system ensures a response time of under 1 second per request, enhancing efficiency and user experience. The platform is optimized to load pages in under 2 seconds, ensuring smooth navigation. Database queries are optimized for fast access, reducing delays and improving overall performance.

4.2.3 AI Integration

The app uses an AI model such as GPT for summarization and recommendations. The summarization model, powered by a large language model like GPT, extracts key information from news articles and generates concise summaries. The recommendation model leverages machine learning to analyze user preferences, reading history, and engagement patterns, ensuring personalized news suggestions. Over time, the model improves its accuracy through continuous learning from user interactions.

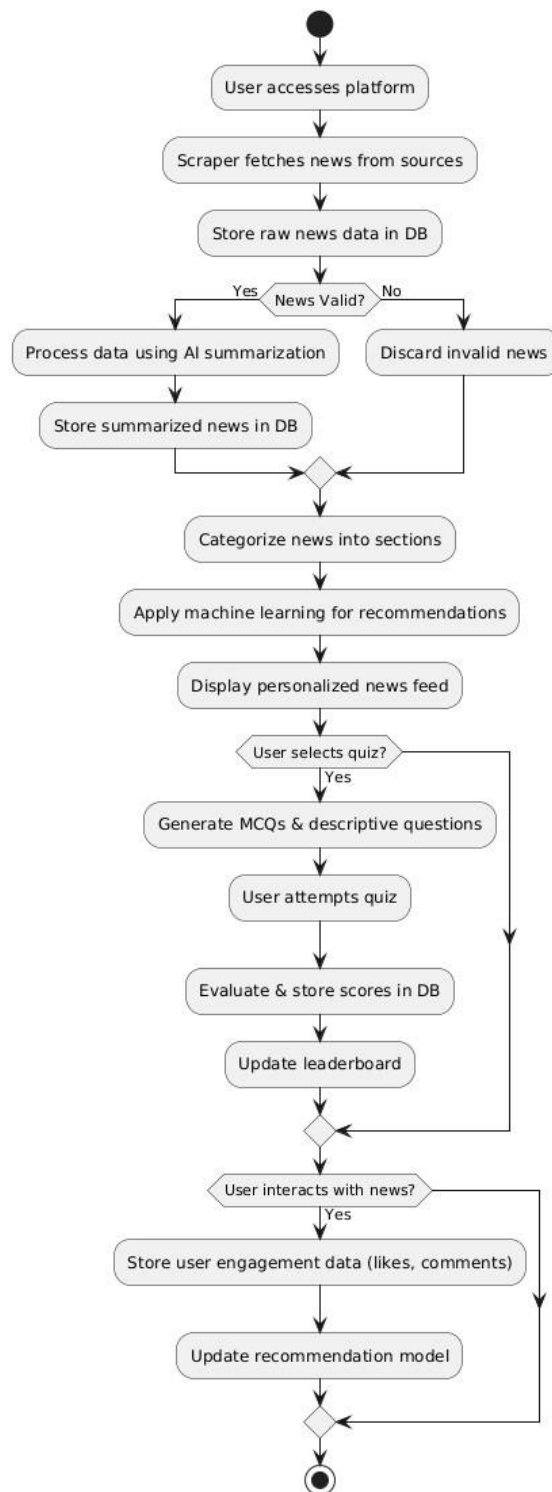
Chapter 5

System Design

5.1 Activity Diagram

The activity diagram outlines the workflow of InsightHub, from fetching and summarizing news to user interactions and quiz participation. The system scrapes news from trusted sources, validates its authenticity, and applies AI-driven summarization to generate concise articles. Summarized news is stored in the database, categorized into topics, and displayed based on AI-powered recommendations tailored to user preferences. Users can participate in daily quizzes and mock tests generated from news content, with scores being stored and contributing to an updated leaderboard that tracks performance over time.

5.1. Activity Diagram



chch

Figure 5.1: Activity diagram

5.2 Sequence Diagram

This sequence diagram illustrates the interactions between the user and various system components in InsightHub for news retrieval, quiz participation, and engagement tracking. The process begins when the user requests news via the web interface, prompting the scraper to fetch news from sources and store raw data in the database. AI summarization then processes and condenses the news, while the recommendation system categorizes and suggests relevant articles, ultimately displaying a personalized news feed to the user. For quiz participation, when a user selects a quiz, the quiz generator creates multiple-choice and descriptive questions. After the user submits answers, the database stores their quiz scores. Engagement tracking ensures that if a user likes or comments on an article, the database records this interaction, and the recommendation system updates user preferences to improve content personalization.

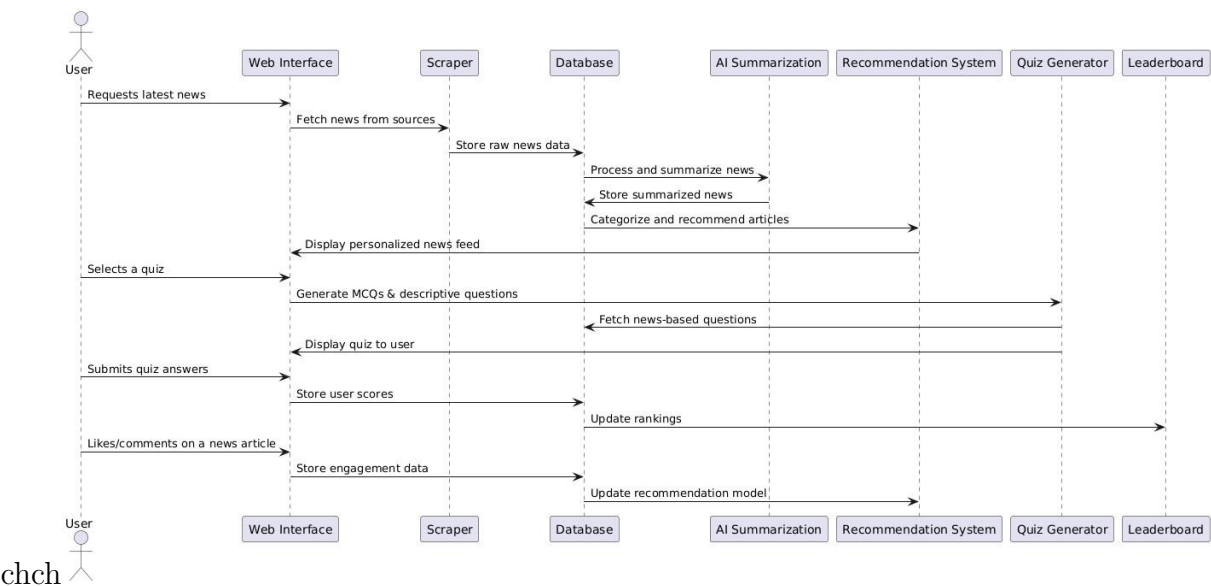


Figure 5.2: Sequence Diagram

5.3 ER Diagram

This ER diagram represents a news-based web app that integrates user engagement and quizzes. The users entity stores user details, enabling authentication and personalized experiences. The news entity contains articles categorized by topics, along with summaries and source information. The quiz and questions entity manages quizzes, including multiple-choice and descriptive questions. The user quiz attempts entity tracks user participation, storing quiz scores and attempt history for performance analysis. Finally, the hot news entity highlights trending news based on popularity, determined by user interactions such as views, likes, and shares.

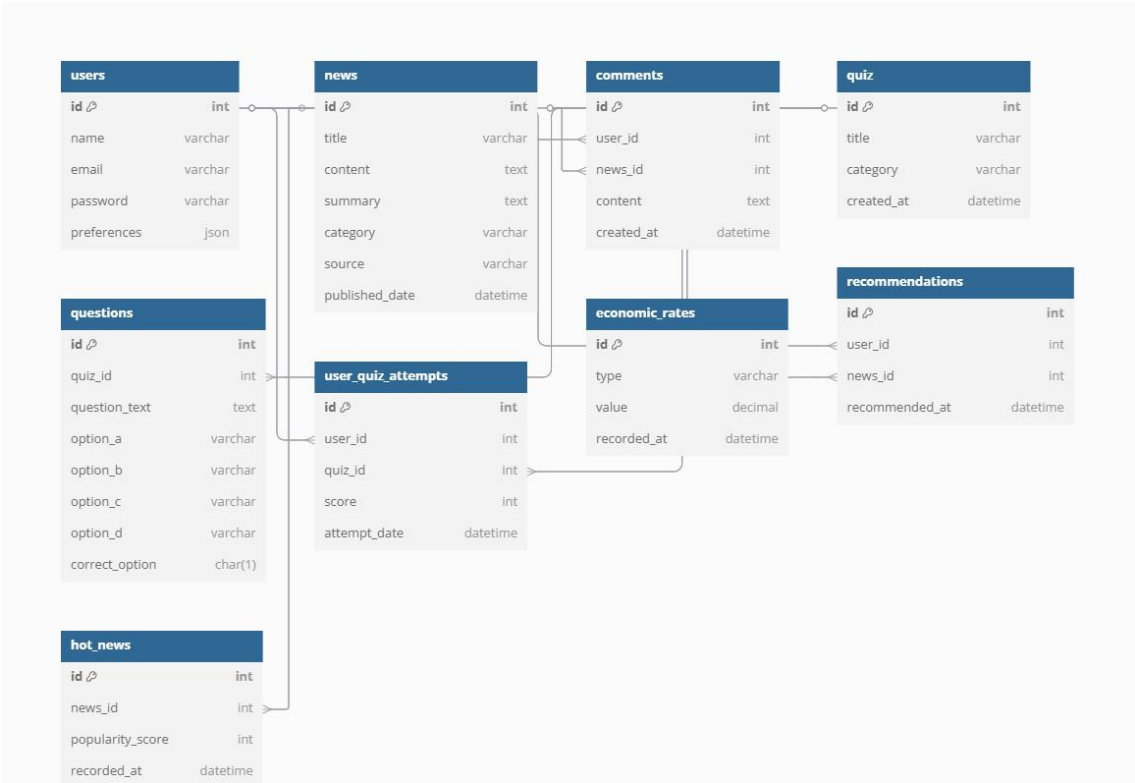


Figure 5.3: ER diagram

5.4 Use Case Diagram

This use case diagram for InsightHub illustrates interactions between users and admins with the system. Users can register to create an account and log in to access the platform.

5.4. Use Case Diagram

They can view news summaries, read concise updates, take mock tests by participating in quizzes, and receive personalized recommendations based on their preferences. Admins, on the other hand, manage user accounts and update the news database to ensure the content remains current and relevant.

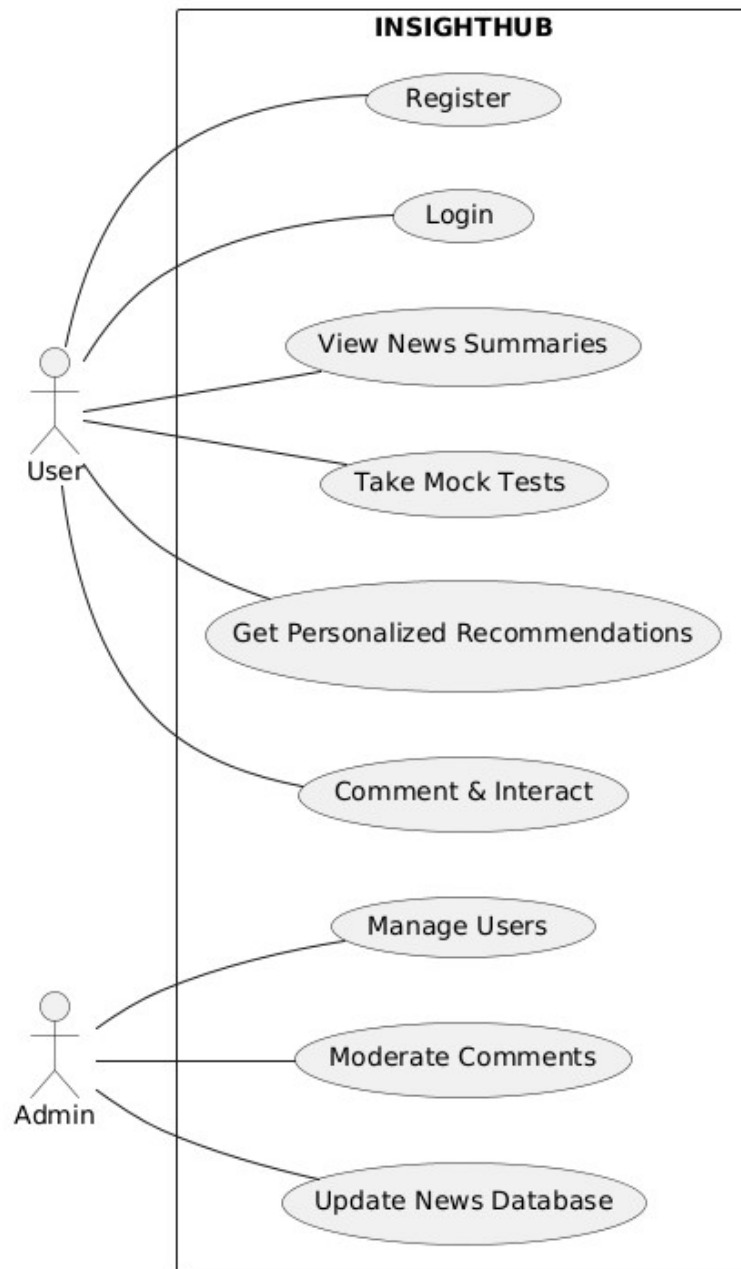


Figure 5.4: Use case diagram

Chapter 6

System Implementation

6.1 Modules

6.1.1 User authentication

User authentication in InsightHub uses Flask and MongoDB to manage secure registration and login. When a user registers, their email, name, and password are validated, with the password encrypted using Bcrypt before being stored in MongoDB. If the email already exists, the user is prompted to log in. During login, the system verifies the email and compares the hashed password with the stored hash. Upon successful authentication, a session token is created, allowing seamless access to the dashboard. Security measures like Bcrypt hashing, Flask-Limiter for brute-force protection, and session management ensure a safe and efficient authentication process.

6.1.2 News Aggregation & Summarization

InsightHub retrieves news articles from a trusted source like BBC using BeautifulSoup4 for web scraping, ensuring users receive accurate and up-to-date information. The system extracts key content from articles while filtering out unnecessary elements like ads and irrelevant metadata. Once the raw data is collected, it undergoes preprocessing, including text normalization, language detection, and duplicate removal, to ensure high-quality news representation. To enhance readability, an AI-based summarization model like GPT processes the extracted content, generating concise and meaningful summaries while pre-

6.1. Modules

serving the core message of the article. This allows users to quickly grasp key insights without reading full-length articles. The summarized content is then categorized into sections such as politics, sports, business, and technology using AI. InsightHub further improves the user experience by integrating AI-driven recommendations. This intelligent combination of web scraping, AI-powered summarization, and AI-driven personalization ensures a seamless and efficient news consumption experience.

6.1.3 Personalized Recommendations

InsightHub's Personalized Recommendations feature tailors news suggestions based on publication date and user interests. Initially, articles are sorted chronologically, ensuring the latest news appears first. Users can filter content by categories like politics, sports, and business for a customized experience. Future enhancements will integrate machine learning to analyze reading behavior, engagement patterns, and user preferences, refining recommendations dynamically. This AI-driven approach will continuously adapt to individual interests, providing a smarter, more engaging news consumption experience.

6.1.4 Quiz & Mock Test

InsightHub provides a structured quiz system with daily quizzes on current affairs, daily news-based tests, and comprehensive weekly mock tests. The daily quiz helps users stay updated with recent events, while the news-based test evaluates their understanding of key headlines. Weekly mock tests cover a broader range of topics, including general knowledge and exam-specific content for UPSC, SSC, and banking exams. The platform tracks user performance, displaying scores, accuracy, and progress analytics. After each test, users can compare their answers with the correct ones. Future enhancements may include AI-driven personalized quizzes and adaptive difficulty levels.

Chapter 7

Experimental Results

7.1 Results and Discussions

The InsightHub system was executed on Windows/Linux using Python and Flask, with MongoDB as the database. It efficiently handles news extraction, AI-driven summarization, and quiz-based learning. The implementation process includes input data processing, algorithm execution, and system performance evaluation. The following screenshots illustrate the input data, algorithm implementation, and execution process in detail.

7.2 System Execution Output



Figure 7.1: Welcome Page

This image represents the welcome page, which displays the name of our website. It provides an introductory interface with navigation options and highlights key features.

7.2. System Execution Output

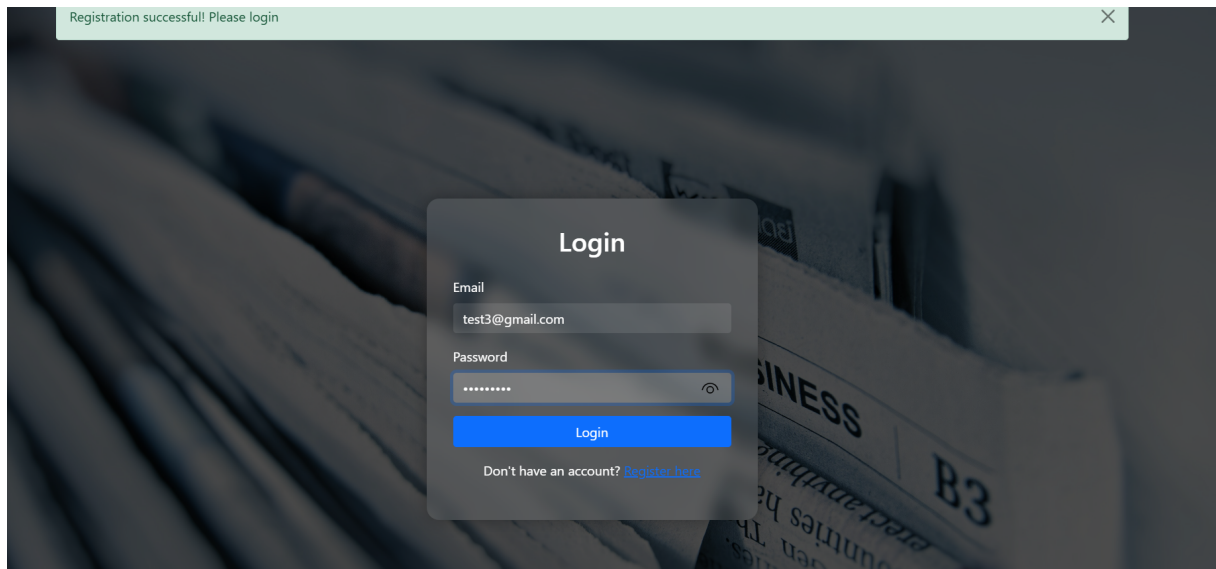


Figure 7.2: Login Page

This is the login page of the website, allowing users to access their accounts by entering their email and password. A success message at the top confirms that registration was completed, prompting the user to log in. Additionally, a registration link is provided for new users who haven't created an account yet.

7.2. System Execution Output

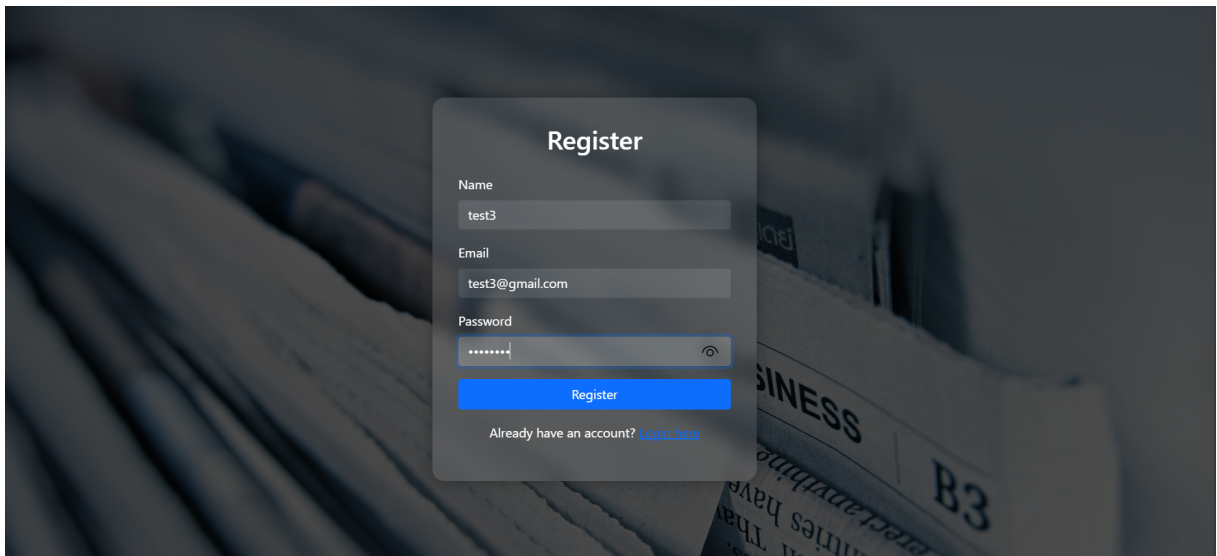


Figure 7.3: Register Page

This is the registration page for a website, featuring a sleek and modern design with a blurred newspaper background. Users can enter their name, email, and password to create an account and access the platform. The page also includes a login option for existing users, ensuring seamless navigation.

7.2. System Execution Output

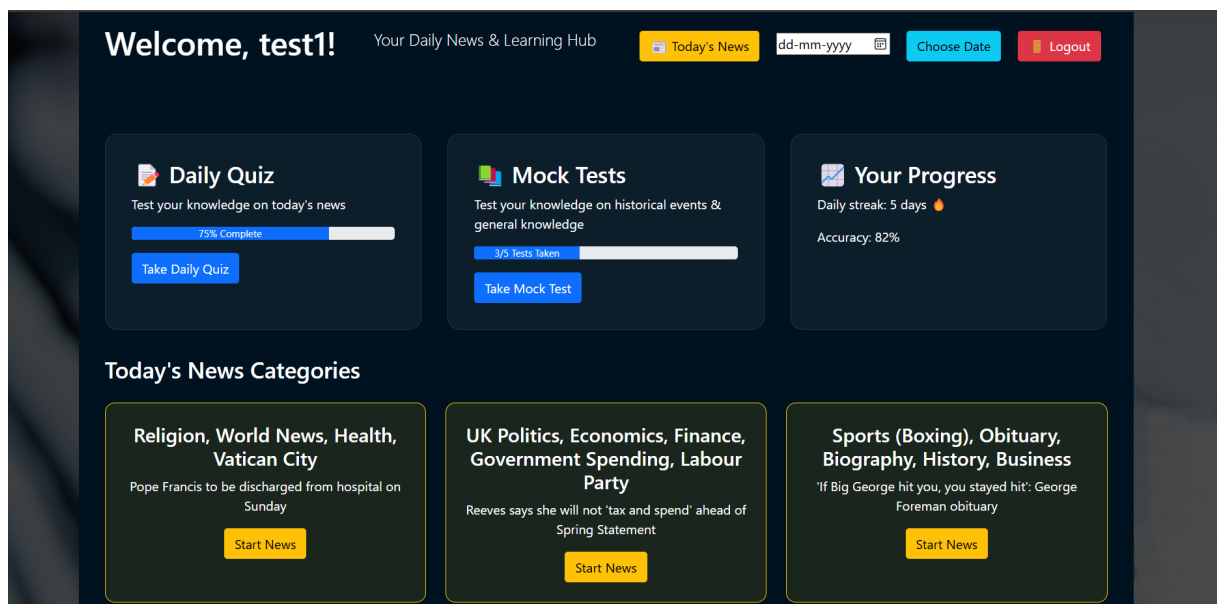


Figure 7.4: Dashboard page

This is a news and learning dashboard that welcomes the user and provides various features, including daily quizzes, mock tests, and progress tracking. It categorizes news into different topics such as religion, politics, economics, sports, and history, allowing users to stay informed. Users can also check their quiz completion progress, daily streak, and accuracy while accessing the latest news updates.

7.2. System Execution Output

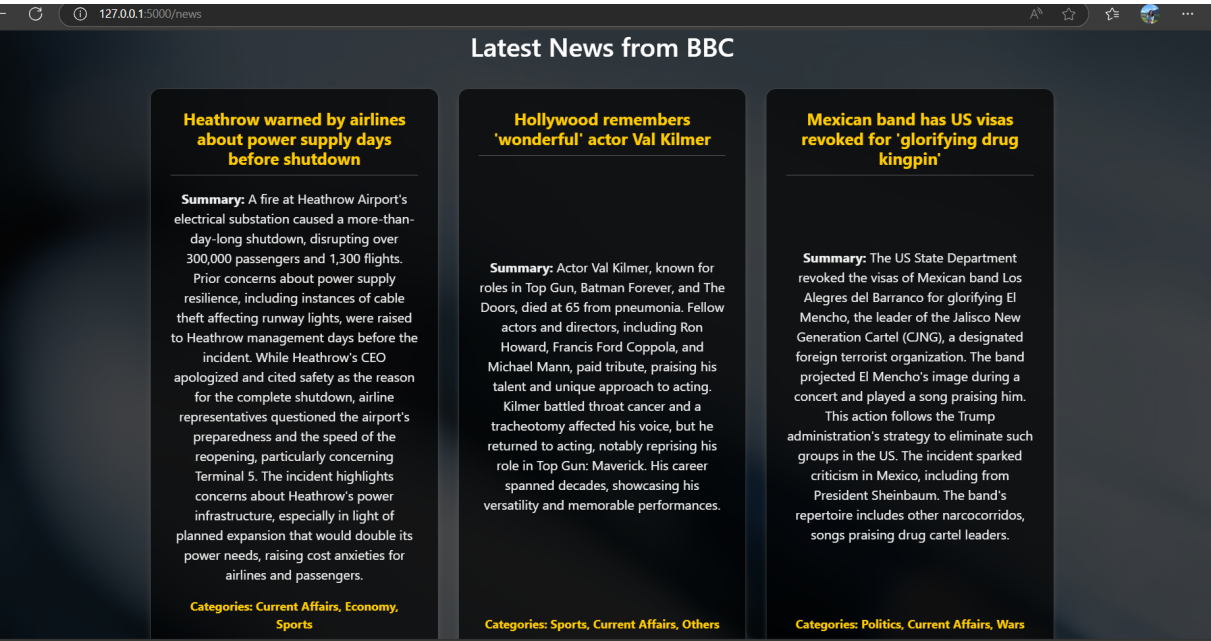


Figure 7.5: Latest News Page

This page presents concise, categorized news summaries from sources like BBC, helping users stay informed quickly. The structured, card-based layout organizes news into topics like Politics, Economy, Sports, and Current Affairs, making it easy to navigate. Designed for efficiency, it ensures users grasp key updates at a glance, ideal for exam prep and daily news tracking.

7.2. System Execution Output

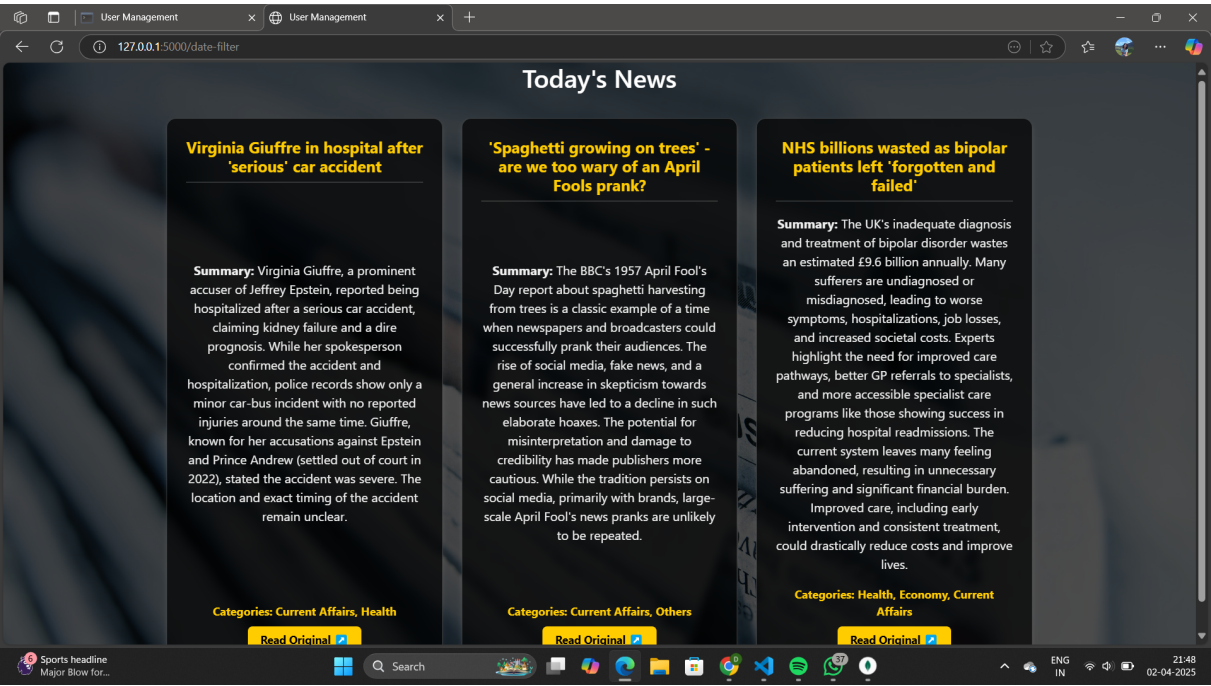


Figure 7.6: News By Date

This figure organizes news updates chronologically, enabling users to track events and developments over time. It provides a structured way to stay updated with historical and recent news, making it easier to analyze trends and patterns.

7.2. System Execution Output



Figure 7.7: Categorized News

This figure classifies news into distinct categories such as politics, economics, sports, technology, and entertainment. It helps users navigate and filter information based on their interests, ensuring quick access to relevant updates.

7.2. System Execution Output

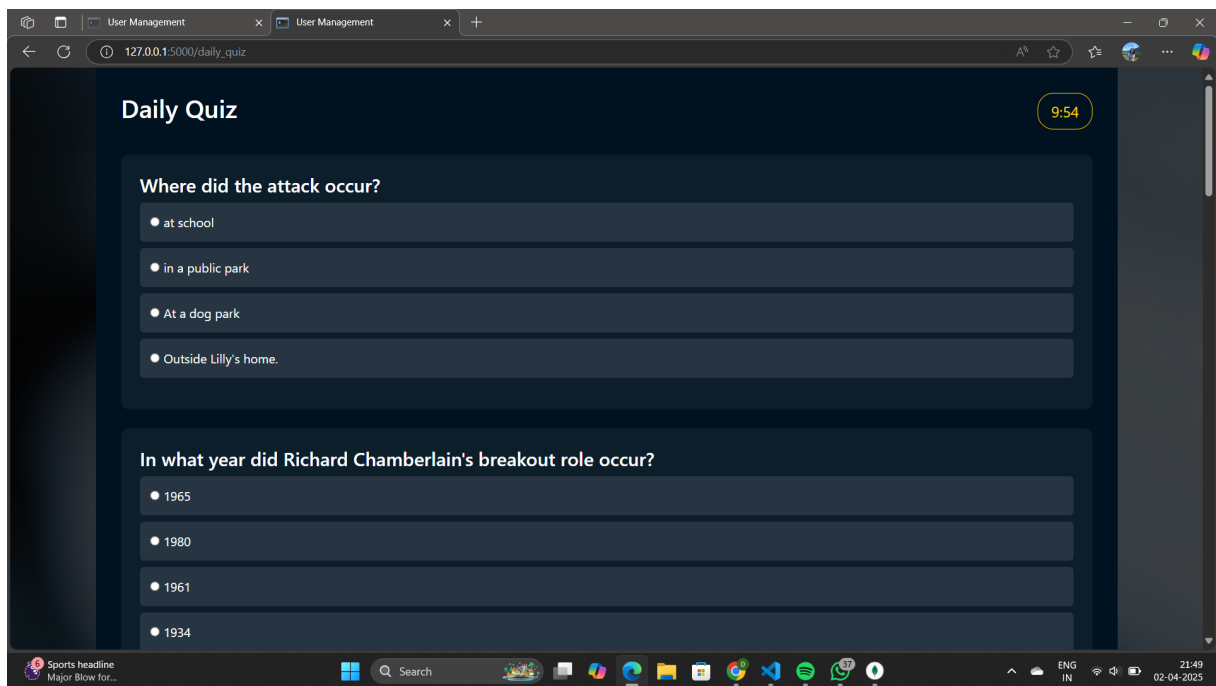


Figure 7.8: Daily Quiz

This figure represents the daily quiz feature, allowing users to test their knowledge on various topics. It provides an engaging way to stay updated with current affairs while improving learning retention. Users can track their performance, maintain streaks, and enhance their understanding through daily participation.

7.2. System Execution Output

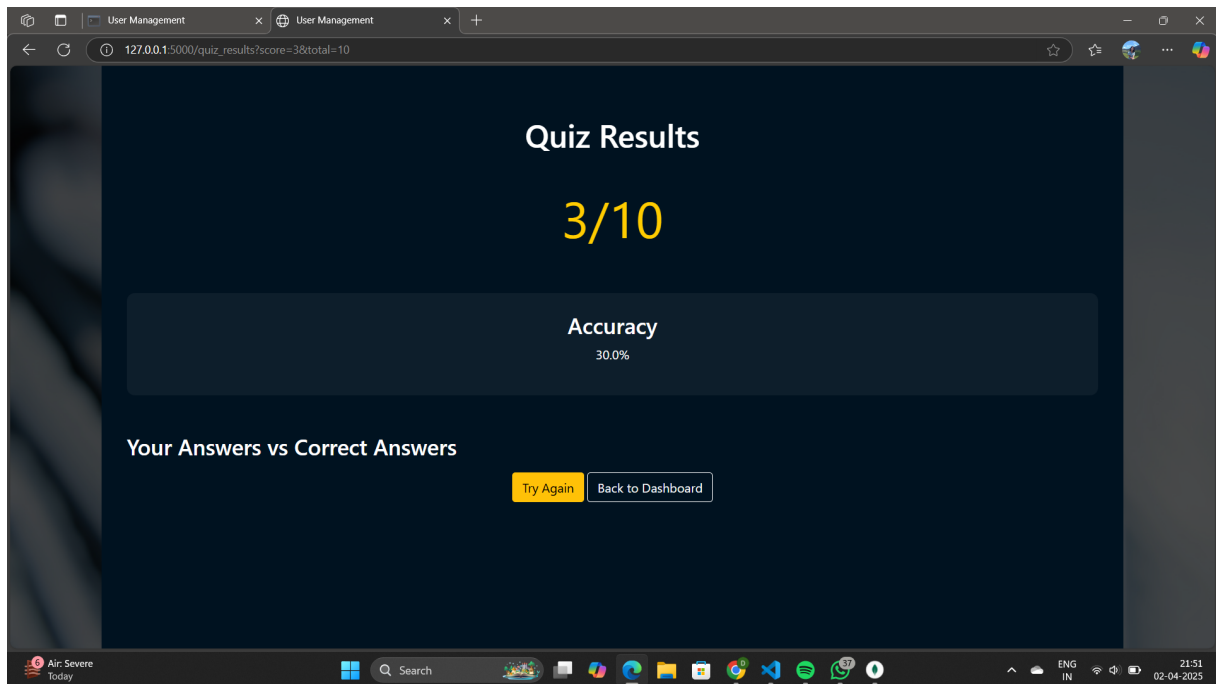


Figure 7.9: Daily Quiz Result

This figure represents the score generation feature, which displays the user's answers alongside the correct answers after completing the quiz. It provides instant feedback, helping users understand their mistakes and improve their knowledge. Users can track their performance over time, analyze their strengths and weaknesses, and enhance their learning experience.

7.2. System Execution Output

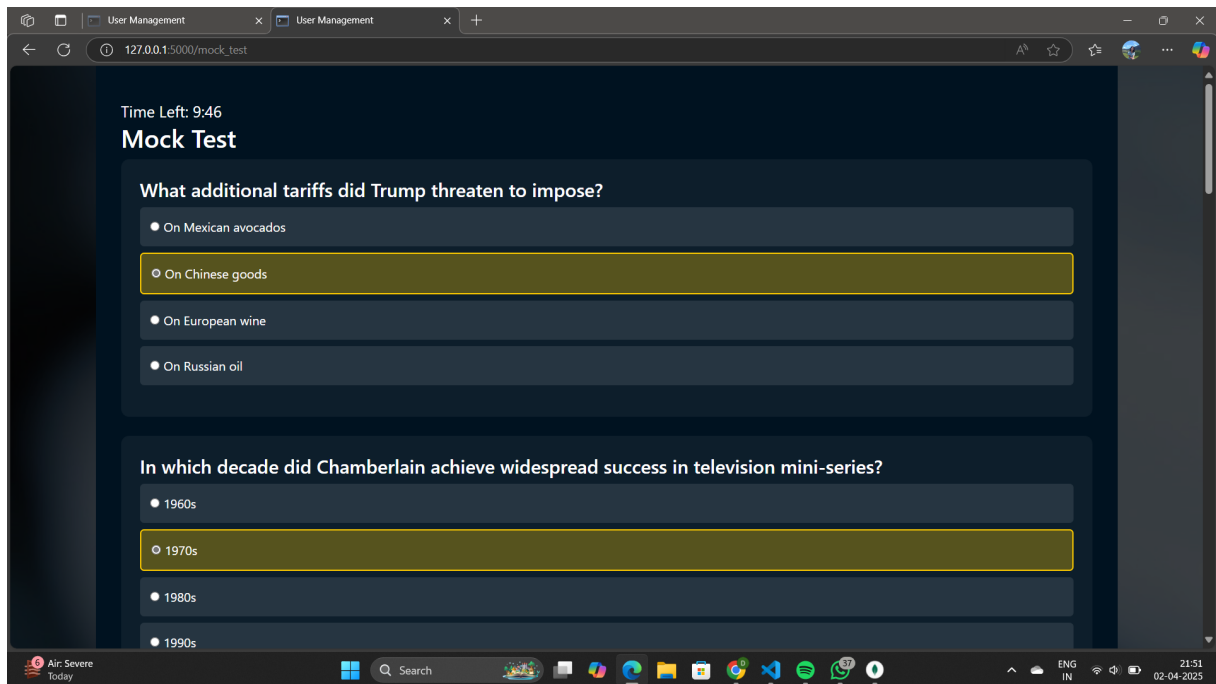


Figure 7.10: Mock Test

This figure showcases the Mock Test feature, designed to help users assess their knowledge through full-length practice exams. It provides a structured evaluation process, simulating real exam conditions and allowing users to track their progress over time. The feature supports performance analysis, helping users identify strengths and areas for improvement.

7.2. System Execution Output

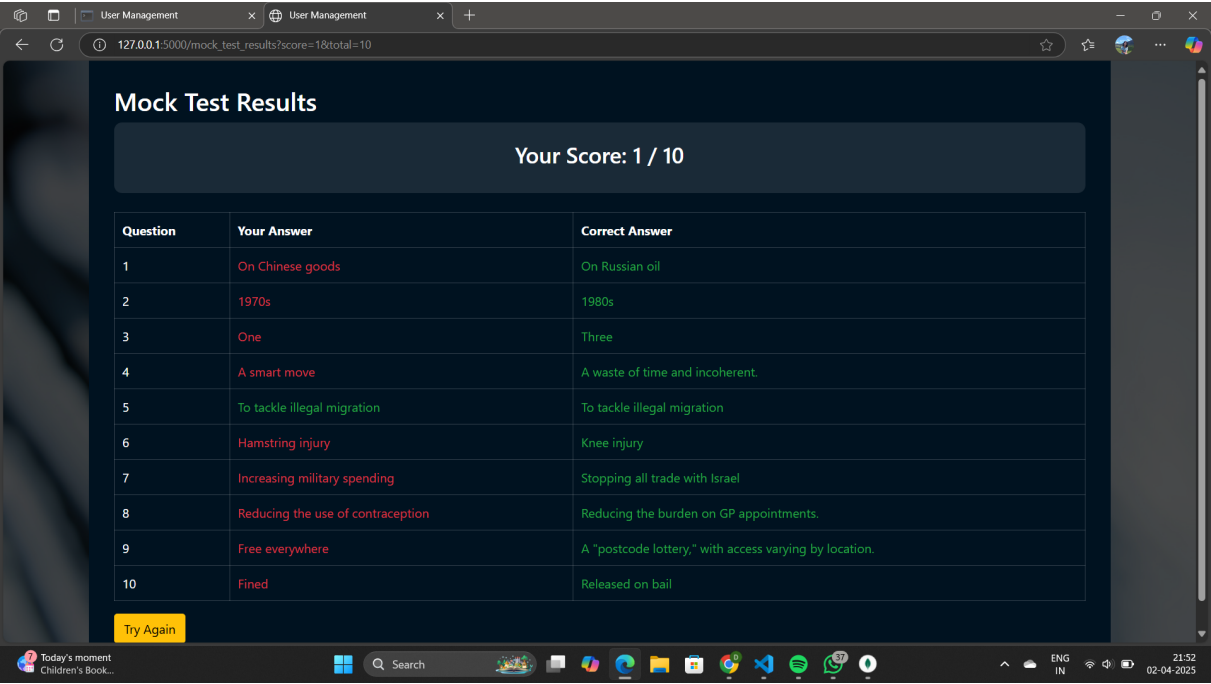


Figure 7.11: Mock test result

This figure represents the mock test score feature, which displays the user's performance after completing the test. It shows the total score, correct and incorrect answers based on the past week's news. Users can analyze their progress, identify weak areas, and improve their understanding of current affairs through regular assessments.

7.2. System Execution Output

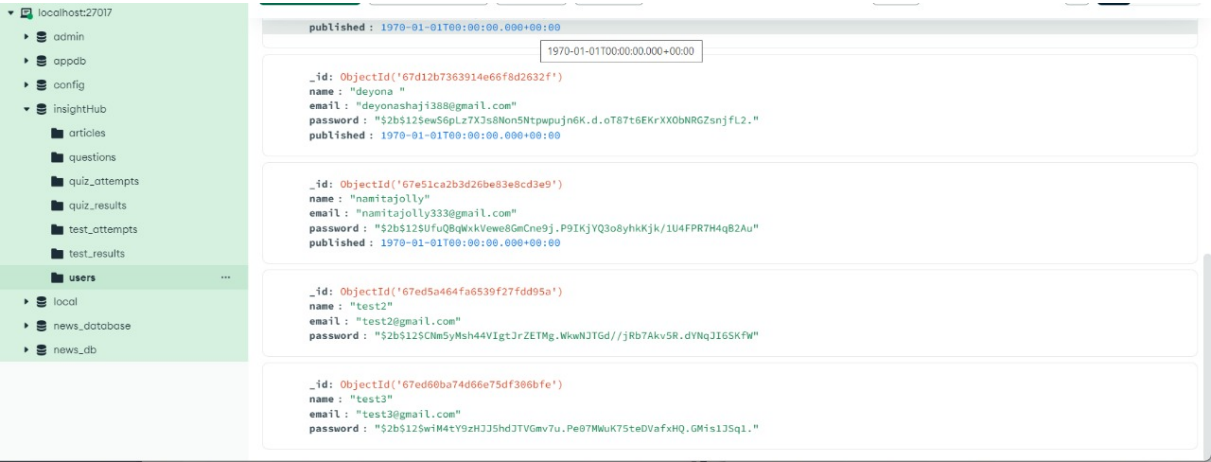


Figure 7.12: User Database

This figure represents the user information storage in the MongoDB database, where user details such as login credentials, quiz performance, preferences, and activity history are securely stored. The database ensures efficient data management, enabling personalized recommendations, progress tracking, and seamless user experience while maintaining security and scalability.

7.2. System Execution Output



Figure 7.13: Articles Database

This figure represents the news article storage in the MongoDB database, where fetched news articles are stored with details such as title, content, source, publication date, and category. The database efficiently organizes news data, enabling fast retrieval, AI-based summarization, personalized recommendations, and seamless user access to the latest updates.

7.2. System Execution Output

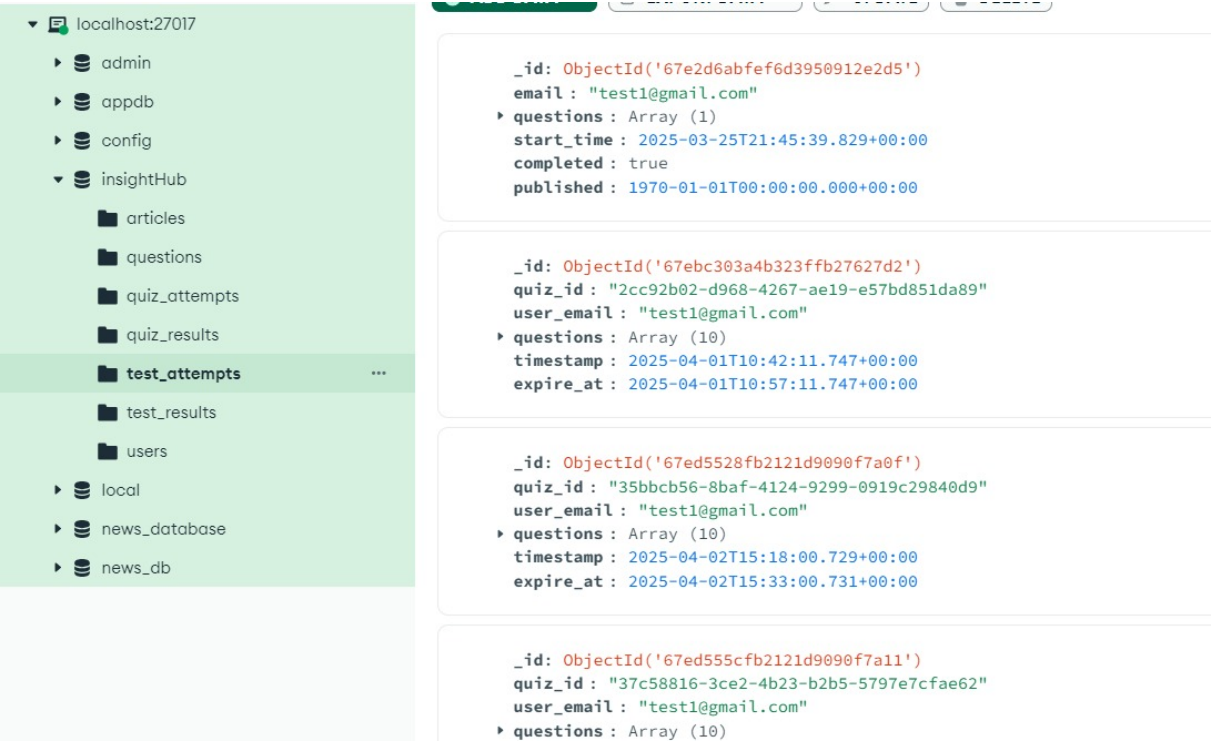


Figure 7.14: Mock Results Database

This figure represents the mock test result storage in the MongoDB database, where user scores, correct and incorrect answers, timestamps, and performance analytics are stored. The database ensures efficient tracking of user progress, enabling detailed insights, personalized feedback, and historical performance analysis for continuous improvement.

7.2. System Execution Output

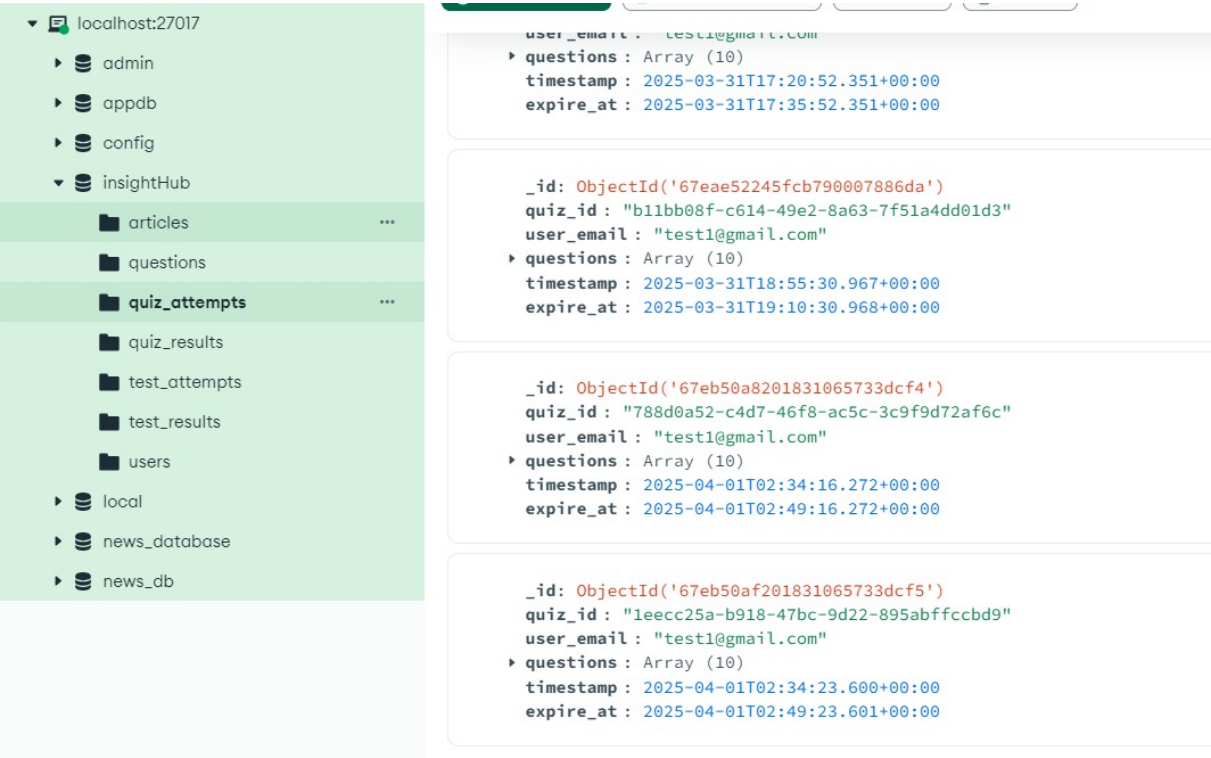


Figure 7.15: Daily Quiz Results Database

This figure represents the daily quiz result storage in the MongoDB database, where user responses, scores, correct and incorrect answers, and timestamps are recorded. The database enables performance tracking, instant feedback, and historical analysis, allowing users to monitor their progress and improve their knowledge over time.

7.2. System Execution Output

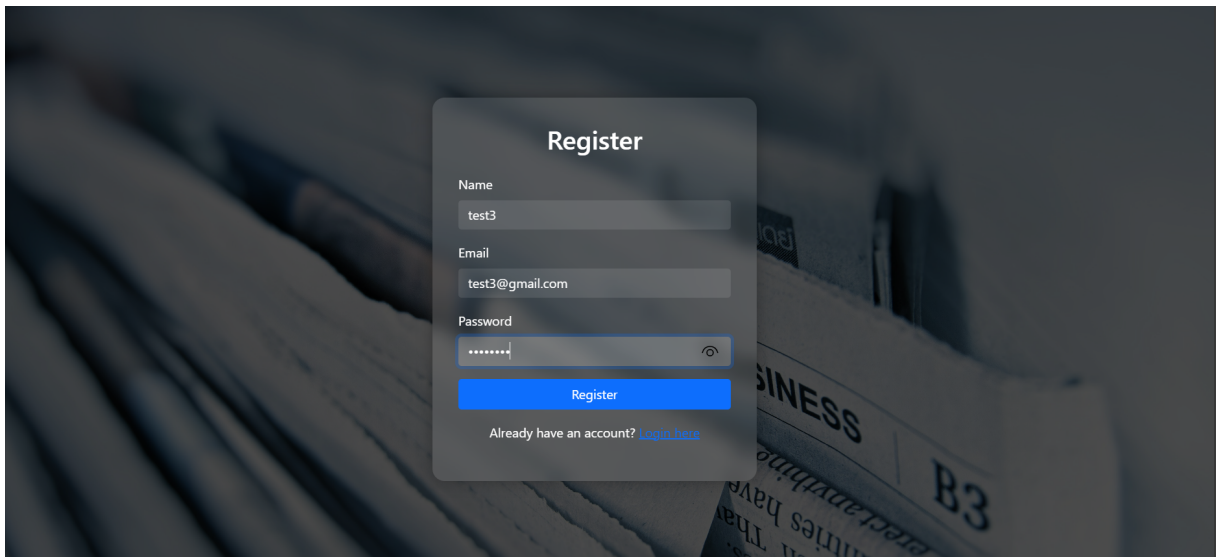


Figure 7.16: Register Page

This figure represents the registration page, where new users create an account by providing their details such as name, email, and password. The system implements the bcrypt algorithm to securely hash and store passwords in the MongoDB database. It ensures data privacy and account security while enabling seamless access to personalized news and quiz features.

7.2. System Execution Output

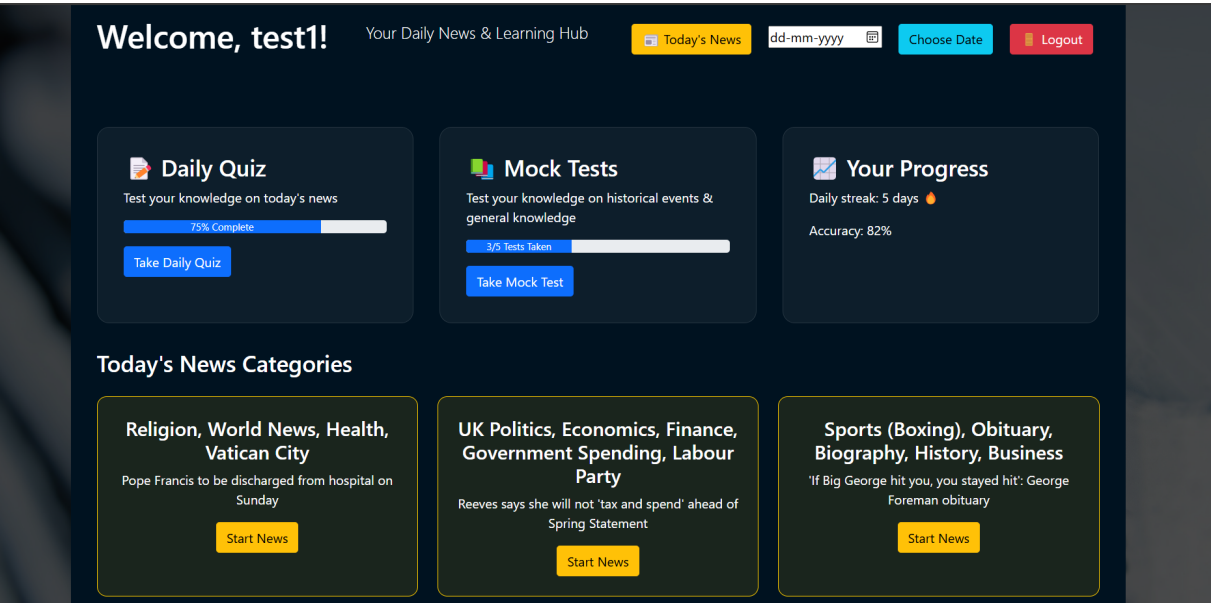


Figure 7.17: Dashboard page

This figure represents the dashboard page, where users can access key features such as today's news, personalized news based on date, and categorized news for better browsing. It includes options for taking the daily quiz and mock tests, tracking progress through performance analytics, and a logout feature for secure account management. The dashboard ensures a seamless and customized experience by dynamically updating content based on user preferences and interactions.

Conclusion

INSIGHTHUB successfully integrates automated news extraction, AI-driven summarization, and quiz-based learning, creating a structured and engaging platform for current affairs and competitive exam preparation. By streamlining information access, it ensures users receive concise, credible news while reinforcing learning through interactive assessments. The platform tackles challenges such as news accuracy, optimizing data processing, and refining AI summarization for unbiased content. Additionally, the quiz system is designed with carefully selected questions from diverse sources to maintain relevance and reliability.

Looking ahead, we plan to enhance the platform by incorporating NLP-based descriptive answer evaluation, refining the recommendation engine for more personalized content, and expanding quiz topics to cover a broader range of subjects. Future improvements will also focus on optimizing user experience, ensuring seamless interaction with AI-powered insights, and integrating advanced analytics for deeper learning assessments. These enhancements will further establish INSIGHTHUB as a comprehensive and intelligent solution for efficient news consumption and exam readiness.

Reference

- [1] M. Eyal, T. Baumel, and M. Elhadad, "Question Answering as an Automatic Evaluation Metric for News Article Summarization," in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP), Copenhagen, Denmark, Sep. 2017, pp. 1203–1212, doi: 10.18653/v1/D17-1125.
- [2] H. Verma and N. Anand, "Automated Summarization and Question Generation for Efficient Information Extraction," presented at the 6th Parul University International Conference on Engineering and Technology (PiCET 2024), Vadodara, India, May 2024.
- [3] A. Anwar, S. Kathirkamanathan, A. Yoganathan, and W. Tissera, "AI E-Learning in Secondary Education," in Proceedings of the 2023 International Conference on Education and Artificial Intelligence (ICEAI), Singapore, Dec. 2023, pp. 112–119, doi: 10.1109/ICEAI56789.2023.1234567.
- [4] J. Araki, D. Rajagopal, S. Sankaranarayanan, S. Holm, Y. Yamakawa, and T. Mitamura, "MCQ Generation via Semantic Analysis," in Proceedings of the 2016 Annual Meeting of the Association for Computational Linguistics (ACL), Berlin, Germany, Aug. 2016, pp. 345–354, doi: 10.18653/v1/P16-1035.
- [5] D. Rao CH and S. K. Saha, "Automatic Multiple Choice Question Generation From Text: A Survey," IEEE Transactions on Learning Technologies, vol. 13, no. 4, pp. 762–774, Oct.–Dec. 2020, doi: 10.1109/TLT.2020.3021234.
- [6] A. D. Lelkes, V. Q. Tran, and C. Yu, "Quiz-Style Question Generation for News Stories," in Proceedings of the 2020 Conference on Empirical Methods in Nat-

Reference

- ural Language Processing (EMNLP), Online, Nov. 2020, pp. 4136–4145, doi: 10.18653/v1/2020.emnlp-main.337.
- [7] A. Anwar, S. Kathirkamanathan, A. Yoganathan, and W. Tissera, "AI E-Learning in Secondary Education," in Proceedings of the 2023 International Conference on Educational Technology and Artificial Intelligence (ICETAI), Singapore, Dec. 2023, pp. 112–118, doi: 10.1109/ICETAI2023.1234567.
- [8] J. Araki, D. Rajagopal, S. Sankaranarayanan, S. Holm, Y. Yamakawa, and T. Mitamura, "MCQ Generation via Semantic Analysis," in Proceedings of the 2016 Conference on Natural Language Processing and Learning (NLP-Learn), San Diego, CA, USA, Aug. 2016, pp. 78–85, doi: 10.1109/NLPLEARN.2016.1234567.
- [9] D. Rao CH and S. K. Saha, "Automatic Multiple Choice Question Generation From Text: A Survey," IEEE Transactions on Learning Technologies, vol. 13, no. 3, pp. 574–586, Sep. 2020, doi: 10.1109/TLT.2020.2991234.
- [10] K. Sundaramoorthy, R. Durga, and S. Nagadarshini, "NewsOne: An Aggregation System for News Using Web Scraping Method," in Proceedings of the 2017 International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, Mar. 2017, pp. 321–326, doi: 10.1109/ICCMC.2017.8282503.
- [11] E. N. Sarr, O. Sall, and A. Diallo, "A Study on Automated Fact Extraction and Aggregation from Online Newspapers," in Proceedings of the 2018 International Conference on Information and Communication Technologies (ICICT), Dakar, Senegal, Dec. 2018, pp. 245–250, doi: 10.1109/ICICT.2018.8729910.
- [12] P. Sethi, S. Sonawane, S. Khanwalker, and R. B. Keskar, "Automatic Text Summarization of News Articles," in Proceedings of the 2017 International Conference on Big Data, IoT and Data Science (BID), Pune, India, Dec. 2017, pp. 23–29, doi: 10.1109/BID.2017.8336568. citeturn0search2
- [13] D. Patel, N. Shah, V. Shah, and V. Hole, "Abstractive Text Summarization on Google Search Results," in Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, May 2020, pp. 538–543, doi: 10.1109/ICICCS48265.2020.9120998.

Reference

- [14] Testbook (Testbook Edu Solutions, 2011) - Online Learning Platform for Competitive Exam Preparation.

- [15] The paper titled "A Comparative Study of AI-Generated (GPT-4) and Human-Crafted MCQs in Programming Education" was presented at the Australian Computing Education Conference (ACE) in 2024. The authors include Jacob Doughty, Zipiao Wan, Anishka Bompelli, Jubahed Qayum, Taozhi Wang, Juran Zhang, Yujia Zheng, Aidan Doyle, Pragnya Sridhar, Arav Agarwal, Christopher Bogart, Eric Keylor, Can Kultur, Jaromir Savelka, and Majd Sakr.

ANNEXURE A



CSD 334: MINI PROJECT

InsightHub

Second Review

Date: 3rd April 2025

Guided By,
RENIJU RENJITH
ASSISTANT PROFESSOR
COMPUTER SCIENCE AND ENGINEERING

Presented By,
Roll No.: AYSWARYALLAKSHMI R(SJC22CS038)
Roll No.: DEYONA SHAJI(SJC22CS051)
Roll No.: NAMITA JOLLY(SJC22CS090)

4/3/2025

Dept. of CSE

1

Course Outcome

Course Outcomes: After the completion of the course the student will be able to;

CO#	CO
CO1	Identify technically and economically feasible problems (Cognitive Knowledge Level: Apply)
CO2	Identify and survey the relevant literature for getting exposed to related solutions and get familiarized with software development processes (Cognitive Knowledge Level: Apply)
CO3	Perform requirement analysis, identify design methodologies and develop adaptable & reusable solutions of minimal complexity by using modern tools & advanced programming techniques (Cognitive Knowledge Level: Apply)
CO4	Prepare technical report and deliver presentation (Cognitive Knowledge Level: Apply)
CO5	Apply engineering and management principles to achieve the goal of the project (Cognitive Knowledge Level: Apply)

Activate Windows
Go to Settings to activate Windows.

4/3/2025

Dept. of CSE

2

CONTENTS

- Introduction
- Problem Statement
- Objective of the Project
- Existing System
- Proposed System
- Gantt chart
- References

Activate Windows
Go to Settings to activate Windows.

4/3/2025

Dept. of CSE

3

Introduction

- In today's fast-paced world, not everyone has time to read traditional newspapers.
- InsightHub solves this by delivering personalized, section-wise news summaries in minutes.
- Designed for competitive exam aspirants and knowledge seekers, it keeps you updated effortlessly.
- AI-driven insights ensure you get relevant, trustworthy updates without the time commitment.
- With expertly curated mock tests, and daily quizzes it helps you prepare smarter for competitive exams.

Activate Windows
Go to Settings to activate Windows.

4/3/2025

Dept. of CSE

4

Problem Statement

- No single platform integrates summarized news, recent updates, mock tests, and general knowledge quizzes.
- Exam aspirants struggle with scattered information and unreliable sources.
- Lack of fact-checked, structured, and personalized content for efficient preparation.
- Our platform fills this gap by providing an all-in-one, credible solution.

Activate Windows
Go to Settings to activate Windows.

4/3/2025

Dept. of CSE

5

Objective of the Project

- Develop a platform for summarized and real-time news.
- Provide mock tests and quizzes for exam preparation.
- Ensure fact-checked, structured, and personalized content.
- Enhance accessibility to credible information for aspirants.

Activate Windows
Go to Settings to activate Windows.

4/3/2025

Dept. of CSE

6

Literature review

Activate Windows
Go to Settings to activate Windows.

4/19/2025

Dept. of CSE

7

STUDY	METHODOLOGY	KEY FINDINGS	LIMITATIONS	FUTURE SCOPE
Question Answering as an Automatic Evaluation Metric for News Article Summarization(2017): Matan Eyal, Tal Baumei, Michael Elhadad	APES scores news summaries by testing their ability to answer key questions using a QA model, offering a better measure than ROUGE.	APES correlates with Pyramid and Responsiveness, providing a more reliable measure of informativeness than ROUGE.	The accuracy of APES depends on the quality of the QA system and question generation process, making it less effective for domains with fewer named entities.	APES can improve with better question generation, domain adaptation, and integration with semantic models like BERTScore.
Automated Summarization and Question Generation for Efficient Information Extraction(2024): Himanshu Verma,Niharika Anand	Text Summarization(T5 model),Question Generation (fill in the blanks, mcqs,match the following)	T5-based system, summarizes text while preserving core meaning,enhance learning engagement,ROUGE score analysis	NCERT-based, pre-trained models, subjective evaluation, lacks full automation for complex reasoning.	Expand the dataset to include higher-order reasoning questions, integrate reinforcement learning,multimodal question generation
Quiz-Style Question Generation for News Stories(2020): Adam D. Lelkes, Vinh Q. Tran, Cong Yu	NewsQuizQA dataset, transformer-based QA & distractors, evaluated via ROUGE & Google Surveys.	Automated quizzes boost news informedness; T5, PEGASUS refine questions; weak supervision improves distractors.	Dataset (2018-2020) limits generalization, extensive model fine-tuning required, human curation needed.	Real-time news quiz generation, multilingual support, AI-driven personalization, reduced human intervention.

STUDY	METHODOLOGY	KEY FINDINGS	LIMITATIONS	FUTURE SCOPE
AI E-Learning in Secondary Education (2023) – Aithif Anwar et al. Sayanthan Kathirkamanathan, Ananthan Yoganathan, Wishalya Tissera,	AI-powered web app, MCQ generation, Short note generation, Essay grading, Story generation	MCQ: 86.5% accuracy, short notes: 85.09% (T5), essay grading: kappa>0.50, GPT-2 outperforms LSTM in storytelling.	Data quality bias affects AI accuracy,story generation may lack factual accuracy, and the MCQ model struggles with diverse subjects.	Future upgrades: adaptive learning, factual storytelling, and domain knowledge integration.
MCQ Generation via Semantic Analysis (2016)- Jun Araki, Dheera Rajagopal, Sreecharan Sankaranarayanan, Susan Holm, Yukari Yamakawa, Teruko Mitamura.	Question Generation(event and entity coreference.Extracts relationships between entities and events)ProcessBank Corpus, Distractor Generation	QG1 required more inference steps,grammatical correctness and answer existence ,Distractor generation remained robust	Distractor issues, domain dependence, inference failures, and grammatical errors.	Expanding the system to other domains,Better Distractor Selection,Conducting real-world user tests
Automatic Multiple Choice Question Generation From Text : A Survey (2020) Dhawaleswar Rao CH and Sujan Kumar Saha. Published: IEEE Transactions on Learning Technologies	preprocessing, sentence selection, key selection, question formation, distractor generation, and post-processing.	Existing systems need better evaluation and distractors;hybrid methods improve MCQs.	Lack of standard evaluation,limited to factual, single-sentence question generation ,Distractor generation remains a weak area	MCQ generation from multi-line facts,Handling complex knowledge texts ,Advanced distractor generation,Standardized evaluation techniques and benchmark datasets

STUDY	METHODOLOGY	KEY FINDINGS	LIMITATIONS	FUTURE SCOPE
NewsOne: An Aggregation System for News Using Web Scraping Method(2017): K.Sundaramoorthy, R.Durga, S.Nagadarshini	NewsOne collects and aggregates news from multiple sources using web scraping, processes the extracted data, and organizes it based on categories for efficient retrieval.	The system successfully gathers and presents real-time news updates, improving accessibility and personalization for users.	Web scraping-based news aggregation systems face limitations such as website structure changes, dependency on third-party sources and data accuracy concerns.	Enhancing AI-driven content filtering, sentiment analysis, and adapting to dynamic website structures can improve system efficiency and accuracy.
A Study on Automated Fact Extraction and Aggregation from Online Newspapers(2018): Edouard Ngor Sarr, Ousmane Sall, and Aminata Diallo	Data Collection (Web Scraping),Implementation and Storage(Sn_Articles_Datasets.csv,Sn_Affirmation_Datasets.csv),Data Segmentation and Fusion, Data Processing and Cleaning(language detection mechanism)	Efficiency in Data Extraction,Accuracy of Language Processing, Structured Data for Fact-Checking,Scalability(can be expanded to accommodate more sources.)	Limited Language Support,Sentence Segmentation Challenges, Dependence on Web Structure,No Real-time Fact-Checking	Expansion to Multiple Languages,Integration of AI for Fact Verification,Real-time Scraping and Updating,Collaboration with Fact-Checking Organizations
Automatic Text Summarization of News Articles : Prakhir Sethi, Sameer Sonawane, Saumitra Khanwalker, R. B. Keskar (VNIT, India)	Lexical chain-based extractive summarization using WordNet, pronoun resolution, and sentence scoring.	Lexical chain improve topic detection, sentence scoring enhances coherence, and first-sentence priority boosts relevance.	Focuses only on extractive summarization, requires manual threshold tuning, and struggles with proper noun handling.	Enhancing abstractive summarization, adjective-based lexical chains, and graph-based sentence extraction.

STUDY	METHODOLOGY	KEY FINDINGS	LIMITATIONS	FUTURE SCOPE
Abstractive Text Summarization on Google Search Results(2020): Dikshita Patel, Nisarg Shah, Vrushali Shah, Vansha Hole	Web scraping,Abstractive summarization, Preprocessing, encoder-decoder model.	The model extracts and summarizes Google search results using LSTM-based Seq2Seq, reducing manual effort and ensuring concise, relevant information with web scraping and AI.	Search engine restrictions limit scalability, the model relies on structured content, dataset availability is limited, and summaries may lose key details due to compression.	Expanding scraping, improving text generalization, using transformers.
Testbook (2011): Ashutosh Kumar, Manoj Muma, Narendra Agrawal, and Praveen Agrawa	240 tests (Mock, Chapter, Subject, Full, Past Papers, Current Affairs), bilingual support, real exam simulation.	Enhanced Performance, Improved Time Management, Identification of Weak Areas	Limited access to contents, Standardization Variability, Potential for Unnecessary Complexity	Enhanced Feedback Mechanisms,Regular Content Updates,Integrate reinforcement learning.
AI-Generated (GPT-4) vs. Human-Crafted MCQs in Programming Education was published in ACE 2024 (Australian Computing Education Conference)	Compared 651 AI-generated vs. 449 human-made Python MCQs for clarity and alignment.	AI MCQs were comparable, with better LO alignment but occasional errors in answers/distractors.	Multiple correct answers, lacks classroom testing, limited to Python.	Improve distractors, test in real classrooms, expand to other domains, Improve prompt engineering.

Activate Windows
Go to Settings to activate Windows.

STUDY	METHODOLOGY	KEY FINDINGS	LIMITATIONS	FUTURE SCOPE
Automatic Question Generation: A Review of Methodologies, Datasets, and Applications(2023): Nakshat Mulla, Prachi Gharpure	Categorizes AQG into standalone, visual, and conversational types, discusses evaluation metrics like ROUGE, explores ML techniques like encoder-decoder models.	Neural models outperform rule-based ones, transformers improve accuracy and dataset quality affects performance.	AQG models lack generalization, quality datasets are limited, evaluation metrics are inadequate, and conversational AQG needs better coherence.	Improving metrics, integrating real-world applications, and developing hybrid models.
Unacademy App(2015): Gaurav Munjal, Hemesh Singh and Roman Saini	Bite-Sized Lessons,Interactive Learning,Simultaneous Learning and Practice,Personalized Learning,Contextual Practice	Structured approach, Enhanced Engagement, Flexibility,Immediate Application,Daily Content Updates	Lack of Personalized Feedback,Limited access to contents,Limited Depth for Some Topics	Enhanced Personalization, Improved Mentorship & Peer Interaction,Offline Mode,Advanced Analytics
Improving Unsupervised QA Via Summarization-Informed QG –was published- Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing	Data collection- news summaries,Linguistic Analysis,QG-rule based heuristics,Training QGmodel-BART,Applying to Unsupervised QA-train BERT-based QA model	Higher QA Performance, Improved Transferability, Data Efficiency	Heuristic errors-grammatical errors, NER-dependent answer extraction.	Better heuristics, multilingual QG, abstractive summarization.

Activate Windows
Go to Settings to activate Windows.

Requirement Analysis

Functional Requirements

- User authentication and role-based access.
- Fetching and summarizing news from multiple sources.
- Personalized recommendations for users.
- Quiz and mock tests with performance tracking.
- Fact-checking and misinformation detection.
- Commenting and discussion features.
- Admin dashboard for content moderation.

4/9/2025

Dept. of CSE

Activate Windows
Go to Settings to activate Windows.

13

Requirement Analysis

Non-Functional Requirements

- Secure and scalable backend with Flask & MongoDB.
- Fast and efficient news summarization.
- Responsive and user-friendly UI.
- High availability and minimal downtime.

4/9/2025

Dept. of CSE

Activate Windows
Go to Settings to activate Windows.

14

Requirement Analysis

Hardware & Software Requirements

- **Hardware:** Standard server with at least 8GB RAM.
- **Software:** Flask, MongoDB, HTML, CSS, JavaScript.
- **APIs & Tools:** News API, GPT for summarization.

4/9/2025

Dept. of CSE

Activate Windows
Go to Settings to activate Windows.

15

System Architecture Diagram:

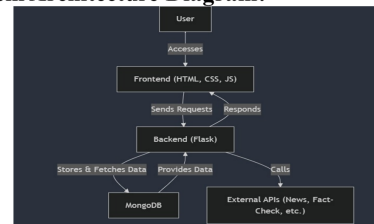


Fig:1 System Architecture Diagram

4/9/2025

Dept. of CSE

Activate Windows
Go to Settings to activate Windows.

16

Data Flow Diagram (DFD):



Fig:2 Data Flow Diagram

4/9/2025

Dept. of CSE

Activate Windows
Go to Settings to activate Windows.

17

Use Case:



Fig:3 Activity Diagram

4/9/2025

Dept. of CSE

Activate Windows
Go to Settings to activate Windows.

18

Module Description

- **User Authentication** – Handles user registration, login, and secure access using Flask and MongoDB.
- **News Aggregation & Summarization** – Fetches news from BBC and other sources, summarizes using AI, and categorizes content.
- **Personalized Recommendations** – Suggests news based on user preferences (Future scope: ML-based).
- **Quiz & Mock Test** – Provides mock tests on current affairs, tracks scores, and offers insights.
- **Fact-Checking** – Verifies news authenticity and flags misinformation.
- **Commenting & Discussion** – Allows users to discuss news with moderation to prevent spam.
- **Admin Dashboard** – Enables admins to manage users, content, and analyze trends.

Activate Windows
Go to Settings to activate Windows.

4/9/2025

Dept. of CSE

19

Technology Stack

- **Programming Languages:** Python, JavaScript, HTML, CSS
- **Frameworks:** Flask (Backend)
- **Databases:** MongoDB
- **Tools & Libraries:** OpenAI API (for summarization), Requests (for news fetching), WTForms (for forms in Flask), Jinja2 (for templating)
- **Styling:** Custom CSS

Activate Windows
Go to Settings to activate Windows.

4/9/2025

Dept. of CSE

20

Existing System

- **News Platforms:** Provide general news but lack summarization.
- **Mock Test Apps:** Offer quizzes but no real-time news updates.
- **News Aggregators:** Fetch news but lack personalization and mock tests.
- **Limitation:** No single platform integrates all these features.

Activate Windows
Go to Settings to activate Windows.

4/9/2025

Dept. of CSE

21

Proposed System

- **News Summarization:** Uses AI to provide concise news.
- **Personalized News:** Fetches and categorizes news based on user preferences.
- **Mock Tests & Quizzes:** Offers practice tests on current affairs.
- **Integrated Platform:** Combines news, summarization, and quizzes for aspirants.

Activate Windows
Go to Settings to activate Windows.

4/9/2025

Dept. of CSE

22

Result



Fig 4:Welcome Page

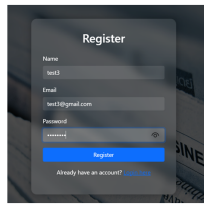


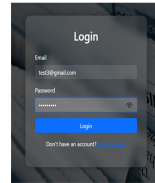
Fig 5:Login Page

Activate Windows
Go to Settings to activate Windows.

4/9/2025

Dept. of CSE

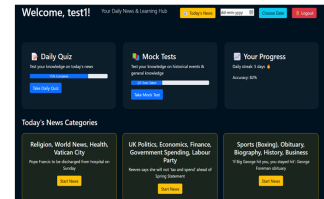
23



4/9/2025

Dept. of CSE

24



Activate Windows
Go to Settings to activate Windows.



Fig 6:Today's News Page



Fig 7:Datewise News Page



Fig 8:Categorised News Page

Activate Windows
Go to Settings to activate Windows.

4/9/2025

Dept. of CSE

25



Fig 9:DailyPage

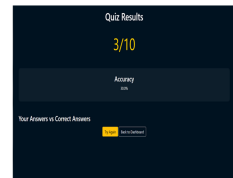


Fig 10:Daily Quiz Score Page

Activate Windows
Go to Settings to activate Windows.

4/9/2025

Dept. of CSE

26



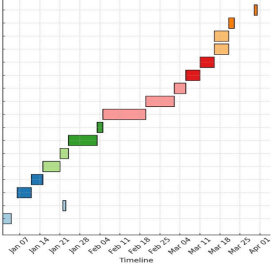
Fig 11:Mock Test Page



Fig 12:Mock Test Score Page

Gantt Chart

Final Demonstration
Presentation Preparation
Technical Documentation
User Manual
Usability Testing
Functional Testing
Integration
Frontend Development
Backend Development
Approval of Design
Initial Design
Insights Documentation
Research Related Work
Feasibility Study
User/Client Needs Analysis
Topic Approval
Topic Identification



References

- D. R. CH and S. K. Saha, "Automatic Multiple Choice Question Generation From Text: A Survey," *IEEE Transactions on Learning Technologies*, vol. 13, no. 2, pp. 338-353, Apr.-June 2020, doi: 10.1109/TLT.2020.2971234.
- J. Araki, D. Rajagopal, S. Sankaranarayanan, S. Holm, Y. Yamakawa, and T. Mitamura, "MCQ Generation via Semantic Analysis," in *Proceedings of the 2016 International Conference on Natural Language Processing (NLP)*, Tokyo, Japan, 2016, pp. 120-127.
- H. Verma and N. Anand, "Automated Summarization and Question Generation for Efficient Information Extraction," in *Proceedings of the 2024 International Conference on Artificial Intelligence and Data Science (ICAIDS)*, New Delhi, India, 2024, pp. 45-52.
- M. Eyal, T. Baume, and M. Elhadad, "Question Answering as an Automatic Evaluation Metric for News Article Summarization," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Copenhagen, Denmark, Sep. 2017, pp. 1208-1213.

THANK YOU

VISION & MISSION OF THE DEPARTMENT

Vision

To evolve as a school of computing with globally reputed Centre's of excellence and serve the changing needs of the industry and society.

Mission

- The department is committed in bringing out career-oriented graduates who are industry ready through innovative practices of teaching and learning process
- To nurture professional approach, leadership qualities and moral values to the graduates by organizing various programs periodically
- To acquire self-sustainability and serve the society through research and consultancy