

**A PROJECT REPORT ON**

**CAREERSAGE: A CAREER GUIDANCE AI MODEL**

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN THE PARTIAL  
FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE

OF

**BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)**

SUBMITTED BY

STUDENT NAME	Exam No.
Patil Gaurav Suresh	B190074260
Sapkale Buddhabhushan Pramod	B190074272
Kute Vivek Kisan	B190074239

**GUIDE**

Prof. A. D. Karanje



**DEPARTMENT OF COMPUTER ENGINEERING**

**PVG's COLLEGE OF ENGINEERING & TECHNOLOGY and GKPIM**

**44, VIDYANAGARI, PARVATI, PUNE 411009**

**SAVITRIBAI PHULE PUNE UNIVERSITY**

**2023-24**



## CERTIFICATE

This is to certify that the project report entitles

### CAREERSAGE: A CAREER GUIDANCE AI MODEL

Submitted by

STUDENT NAME	Exam No.
Patil Gaurav Suresh	B190074260
Sapkale Buddhabhushan Pramod	B190074272
Kute Vivek Kisan	B190074239

is a bonafide student of this institute and the work has been carried out by him/her under the supervision of **Prof. A. D. Karanje** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering** (Computer Engineering) in the year 2023-24.

**Prof. A. D. Karanje**

Guide

Department of Computer Engineering

**Prof. U. M. Kalshetti**

Head

Department of Computer Engineering

**Dr. Manoj R. Tarambale**

I/C Principal,

PVG's College of Engineering & Technology and GKPIM, Pune – 09

Place:

Date:

## ACKNOWLEDGEMENT

It gives us great pleasure in presenting the Project Work - I report on “**CareerSAGE: A Career Guidance AI Model**” and to express my deep regards towards those who have offered their valuable time and guidance in our hour of need.

I would like to express my sincere and wholehearted thanks to our project guide **Prof. A. D. Karanje** and Head of the department **Prof. U. M. Kalshetti** for contributing valuable time, knowledge, experience and providing valuable guidance in making this project a success.

I am also glad to express my gratitude and thanks to our Principal **Prof. Manoj R. Tarambale** Sir for their constant inspiration and encouragement. Finally, before ending I would like to express my gratitude and thanks to all my colleagues who are involved directly and indirectly in making this project a success.

Patil Gaurav Suresh  
B190074260

Sapkale Buddhabhushan Pramod  
B190074272

Kute Vivek Kisan  
B190074239

## ABSTRACT

CareerSage, an innovative career guidance AI model, addresses the challenge of navigating the complexities of career exploration and decision-making in today's dynamic job market. With the rapid evolution of industries and professions, individuals often face difficulty in identifying suitable career paths aligned with their interests, skills, and aspirations. This problem underscores the need for personalized and effective career guidance solutions to empower individuals in making informed decisions about their professional futures.

CareerSage offers a comprehensive solution by leveraging deep learning and natural language processing techniques to provide personalized recommendations tailored to each user's unique profile. Through interactive features such as genre prediction questionnaires, game recommendations, aptitude quizzes, and career matching algorithms, CareerSage empowers users to explore diverse career options and make informed choices aligned with their individual preferences. By combining advanced technology with user-centric design principles, CareerSage aims to streamline the career exploration process, enhance user engagement, and promote satisfaction and fulfillment in the pursuit of meaningful careers.

In conclusion, CareerSage represents a significant advancement in the field of career guidance, offering a user-friendly and adaptive platform to assist individuals in navigating their career journeys. By providing personalized recommendations, actionable insights, and continuous support, CareerSage enables users to unlock their full potential and achieve success in their chosen fields. As individuals continue to evolve and adapt to the changing demands of the workforce, CareerSage stands as a valuable resource, empowering users to navigate the complexities of the modern job market with confidence and clarity.

## Contents

<b>CHAPTER 1: INTRODUCTION</b>	<b>2</b>
1.1: OVERVIEW . . . . .	2
1.2: MOTIVATION . . . . .	2
1.3: PROBLEM DEFINITION AND OBJECTIVES . . . . .	2
1.3.1: Problem Definition . . . . .	2
1.3.2: Objectives . . . . .	2
1.4: PROJECT SCOPE AND LIMITATIONS . . . . .	3
1.4.1: Project Scope . . . . .	3
1.4.2: Limitations . . . . .	3
1.5: METHODOLOGIES OF PROBLEM SOLVING . . . . .	4
<b>CHAPTER 2: LITERATURE SURVEY</b>	<b>5</b>
2.1: PAPER 1: EXPLAINABLE AI AND MACHINE LEARNING . . . . .	5
2.2: PAPER 2: AI CHATBOT TO SUPPORT CAREER GUIDANCE . . . . .	5
2.3: PAPER 3: APPLICATION EXPERT SYSTEM CAREER GUIDANCE . . . . .	6
<b>CHAPTER 3: SOFTWARE REQUIREMENTS SPECIFICATION</b>	<b>7</b>
3.1: ASSUMPTIONS AND DEPENDENCIES . . . . .	7
3.2: FUNCTIONAL REQUIREMENTS . . . . .	8
3.3: EXTERNAL INTERFACE REQUIREMENTS . . . . .	9
3.3.1: User Interfaces . . . . .	9
3.3.2: Hardware Interfaces . . . . .	9
3.3.3: Software Interfaces . . . . .	9
3.3.4: Communication Interfaces . . . . .	10
3.4: NON-FUNCTIONAL REQUIREMENTS . . . . .	10
3.4.1: Performance Requirements . . . . .	10
3.4.2: Safety Requirements . . . . .	10
3.4.3: Security Requirements . . . . .	10
3.4.4: Software Quality Attributes . . . . .	10
3.5: SYSTEM REQUIREMENTS . . . . .	11
3.5.1: Database Requirements . . . . .	11
3.5.2: Software Requirements (Platform Choice) . . . . .	11
3.5.3: Hardware Requirements . . . . .	11
3.6: ANALYSIS MODEL: SDLC MODEL - ITERATIVE MODEL . . . . .	11
3.6.1: Key Characteristics of Iterative Model . . . . .	12
3.6.2: Phases of Iterative Model . . . . .	12
<b>CHAPTER 4: SYSTEM DESIGN</b>	<b>14</b>
4.1: SYSTEM ARCHITECTURE . . . . .	14
4.1.1: Phase 1: Discovery . . . . .	14

4.1.2: Phase 2: Recommendation . . . . .	14
4.1.3: Phase 3: Career Matching . . . . .	15
4.2: MATHEMATICAL MODEL . . . . .	15
4.3: DATA FLOW DIAGRAMS . . . . .	16
4.3.1: Level 0 Data Flow Diagram . . . . .	17
4.3.2: Level 1 Data Flow Diagram . . . . .	17
4.4: ENTITY RELATIONSHIP DIAGRAMS . . . . .	18
4.5: UML DIAGRAMS . . . . .	18
4.5.1: Use Case Diagram . . . . .	18
4.5.2: Sequence Diagram . . . . .	19
4.5.3: Class Diagrams . . . . .	20
4.5.4: Activity Diagram . . . . .	21
<b>CHAPTER 5: PROJECT PLAN</b>	<b>23</b>
5.1: PROJECT ESTIMATE . . . . .	23
5.1.1: Reconciled Costs . . . . .	23
5.1.2: Project Resources . . . . .	24
5.2: RISK MANAGEMENT . . . . .	25
5.2.1: Risk Identification . . . . .	25
5.2.2: Risk Analysis . . . . .	26
5.2.3: Overview of Risk Mitigation, Monitoring and Management . . . . .	27
5.3: PROJECT SCHEDULE . . . . .	29
5.3.1: Project Task Set . . . . .	29
5.3.2: Task Network . . . . .	30
5.3.3: Timeline Chart . . . . .	30
5.4: TEAM ORGANIZATION . . . . .	32
5.4.1: Team Structure . . . . .	32
5.4.2: Management Reporting and Communication . . . . .	33
<b>CHAPTER 6: PROJECT IMPLEMENTATION</b>	<b>34</b>
6.1: OVERVIEW OF PROJECT MODULES . . . . .	34
6.1.1: Phase 1: Questionnaire and Genre Prediction . . . . .	34
6.1.2: Phase 2: Game Recommendation and Interaction . . . . .	34
6.1.3: Phase 3: Aptitude Quiz and Career Match . . . . .	34
6.2: TOOLS AND TECHNOLOGIES USED . . . . .	34
6.2.1: Frontend Development . . . . .	34
6.2.2: Backend Development . . . . .	35
6.2.3: Libraries and Frameworks . . . . .	35
6.3: ALGORITHM DETAILS . . . . .	36
6.3.1: Data Preprocessing . . . . .	36
6.3.2: NLP Algorithms . . . . .	36

6.3.3: Career Matching Algorithms . . . . .	37
<b>Chapter 7: SOFTWARE TESTING</b>	<b>38</b>
7.1: TYPES OF TESTING . . . . .	38
7.1.1: Unit Testing . . . . .	38
7.1.2: Integration Testing . . . . .	38
7.1.3: System Testing . . . . .	38
7.1.4: Validation Testing . . . . .	38
7.1.5: GUI Testing . . . . .	38
7.2: TEST CASES AND TEST RESULTS . . . . .	39
<b>CHAPTER 8: RESULTS</b>	<b>40</b>
8.1: OUTCOMES . . . . .	40
8.2: SCREENSHOTS . . . . .	41
<b>CHAPTER 9: CONCLUSIONS</b>	<b>46</b>
9.1: CONCLUSIONS . . . . .	46
9.2: FUTURE SCOPE . . . . .	46
9.3: APPLICATIONS . . . . .	47
<b>APPENDIX A: FEASIBILITY ASSESSMENT</b>	<b>48</b>
<b>APPENDIX B: PAPER PUBLICATION</b>	<b>49</b>
<b>APPENDIX C: PLAGIARISM REPORT</b>	<b>50</b>
<b>REFERENCES</b>	<b>52</b>
<b>REPORT DOCUMENTATION</b>	<b>53</b>

## List of Figures

1	Phases of Iterative Model. . . . .	12
2	System Architecture . . . . .	14
3	Level 0 Data Flow Diagram . . . . .	17
4	Level 1 Data Flow Diagram . . . . .	17
5	Entity Relationship Diagram . . . . .	18
6	Use Case Diagram . . . . .	19
7	Sequence Diagram . . . . .	19
8	Class Diagram: User Registration and Authentication . . . . .	20
9	Class Diagram: UI Module . . . . .	20
10	Class Diagram: Career Guidance Model . . . . .	21
11	Activity Diagram . . . . .	22
12	Factors for Cost Estimation . . . . .	23
13	Estimated cost and man-power . . . . .	24
14	Task Network . . . . .	30
15	Gantt Chart: Semester I . . . . .	31
16	Gantt Chart: Semester II . . . . .	32
17	Team Structure . . . . .	32
18	Login Page . . . . .	41
19	Website Interface . . . . .	41
20	Questionnaire Bot . . . . .	42
21	Answer Submission . . . . .	42
22	Game Selection . . . . .	43
23	Game: Dodge A Wrench . . . . .	43
24	Game: Chess . . . . .	44
25	Game Scores . . . . .	44
26	Aptitude Quiz (Trivia API) . . . . .	45
27	Final Report . . . . .	45
28	Plagiarism Report . . . . .	50
29	Plagiarism Percentage . . . . .	51
30	Plagiarism Verification . . . . .	51



## List of Tables

1	Functional Requirements: Questionnaire Chat . . . . .	8
2	Functional Requirements: Game Recommendation . . . . .	8
3	Functional Requirements: Career Aptitude Assessment . . . . .	9
4	Database Requirements . . . . .	11
5	Software Requirements . . . . .	11
6	Hardware Requirements . . . . .	11
7	Risk Table . . . . .	26
8	Risk Probability Definitions . . . . .	27
9	Risk 1: Scope Creep . . . . .	27
10	Risk 2: Technical Challenges . . . . .	27
11	Risk 3: Data Privacy and Security . . . . .	28
12	Risk 4: User Adoption . . . . .	28
13	Risk 5: Integration Challenges . . . . .	28
14	Risk 6: Resource Constraints . . . . .	29
15	Risk 7: Dependency on External Services . . . . .	29
16	Timeline: Semester I . . . . .	30
17	Timeline: Semester II . . . . .	31
18	Reporting and Communication . . . . .	33
19	Test Cases and Test Results . . . . .	39

# CHAPTER 1: INTRODUCTION

## 1.1: OVERVIEW

CareerSage is an innovative career guidance AI model designed to assist individuals in navigating the complexities of career exploration and decision-making. Leveraging deep learning and natural language processing techniques, CareerSage provides personalized recommendations tailored to each user's interests, aptitudes, and preferences. Through interactive features such as genre prediction questionnaires, game recommendations, aptitude quizzes, and career matching algorithms, CareerSage empowers users to explore diverse career options and make informed choices aligned with their unique aspirations. With its user-friendly interface and adaptive functionality, CareerSage aims to enhance user engagement, promote self-awareness, and foster satisfaction and fulfillment in the pursuit of meaningful careers.

## 1.2: MOTIVATION

The motivation behind CareerSage stems from the recognition of the profound impact that career choices have on individuals' lives and well-being. In today's rapidly evolving job market, navigating the multitude of career options can be overwhelming and daunting for many. Traditional career guidance resources often lack personalization and fail to address the diverse interests, talents, and aspirations of individuals. CareerSage seeks to address this gap by harnessing the power of artificial intelligence and data-driven algorithms to provide tailored and relevant career guidance to users. By offering personalized recommendations based on users' unique profiles, CareerSage aims to empower individuals to make informed decisions about their career paths, ultimately leading to greater satisfaction, fulfillment, and success in their professional endeavors.

## 1.3: PROBLEM DEFINITION AND OBJECTIVES

### 1.3.1: Problem Definition

Career exploration and decision-making can be daunting and overwhelming for many individuals, as they often struggle to identify suitable career paths aligned with their interests, skills, and aspirations. Traditional career guidance resources often lack personalization and fail to address the diverse needs and preferences of users, leading to uncertainty and indecision.

### 1.3.2: Objectives

- **Personalized Career Guidance:** Develop an AI-driven platform that provides personalized career guidance tailored to the unique interests, aptitudes, and preferences of each user.
- **Enhanced User Engagement:** Create interactive features such as genre prediction questionnaires, game recommendations, aptitude quizzes, and career matching algorithms to engage users actively in the career exploration process.

- **Empower Informed Decision-Making:** Enable users to make informed decisions about their career paths by offering comprehensive insights, resources, and recommendations based on data-driven analysis and algorithms.
- **Accessibility and Inclusivity:** Ensure that CareerSage is accessible to users from diverse backgrounds and demographics, providing equal opportunities for career exploration and guidance.
- **Continuous Improvement:** Continuously refine and enhance CareerSage based on user feedback, emerging technologies, and advancements in AI and data analytics, ensuring that it remains a valuable and relevant resource for individuals seeking career guidance.

## 1.4: PROJECT SCOPE AND LIMITATIONS

### 1.4.1: Project Scope

- **Career Guidance:** CareerSage aims to provide comprehensive career guidance to users by leveraging artificial intelligence and data-driven algorithms.
- **Personalized Recommendations:** The project will focus on offering personalized recommendations tailored to each user's interests, aptitudes, and preferences.
- **Interactive Features** CareerSage will include interactive features such as genre prediction questionnaires, game recommendations, aptitude quizzes, and career matching algorithms to engage users and facilitate their career exploration process.
- **User-Friendly Interface:** The project will prioritize the development of a user-friendly interface to ensure easy navigation and accessibility for users from diverse backgrounds.
- **Scalability:** CareerSage will be designed with scalability in mind, allowing for future expansion and integration of additional features and functionalities

### 1.4.2: Limitations

- **Data Availability:** The effectiveness of CareerSage may be limited by the availability and quality of data used for training and validation purposes.
- **Algorithmic Accuracy:** The accuracy of career recommendations provided by CareerSage may vary based on the performance of underlying algorithms and models, which are subject to limitations and biases inherent in machine learning approaches.
- **Subjectivity:** CareerSage's recommendations may be influenced by subjective factors such as user input and interpretation, which may not always align perfectly with users' individual preferences and aspirations.
- **Resource Constraints:** The project may face limitations in terms of resources such as computing power, storage, and human expertise, which could impact the scope and scale

of development and deployment efforts.

- **Privacy and Security:** CareerSage will need to adhere to strict privacy and security protocols to protect user data and ensure compliance with regulatory requirements, which may impose limitations on certain features and functionalities.

## 1.5: METHODOLOGIES OF PROBLEM SOLVING

- **Machine Learning and Deep Learning:** Utilize machine learning and deep learning techniques to develop robust models for genre prediction, game recommendation, and career matching. This involves data preprocessing, feature engineering, model training, evaluation, and optimization to ensure accurate and personalized results for users.
- **Natural Language Processing (NLP):** Apply NLP methodologies to analyze user inputs, such as questionnaire responses and feedback, to understand user preferences, interests, and skills. Techniques like tokenization, sentiment analysis, and topic modeling can help extract valuable insights from unstructured text data.
- **Data Analytics and Visualization:** Employ data analytics and visualization techniques to gain insights from user interactions, system usage patterns, and performance metrics. Exploratory data analysis (EDA), statistical analysis, and visualization tools can help identify trends, correlations, and areas for improvement in CareerSage's functionality.

## CHAPTER 2: LITERATURE SURVEY

### 2.1: PAPER 1: EXPLAINABLE AI AND MACHINE LEARNING

Machine Learning concept learns from experiences, inferences and conceives complex queries. Machine learning techniques can be used to develop the educational framework which understands the inputs from students, parents and with intelligence generates the result. The framework integrates the features of Machine Learning (ML), Explainable AI (XAI) to analyze the educational factors which are helpful to students in achieving career placements and help students to opt for the right decision for their career growth. It is supposed to work like an expert system with decision support to figure out the problems, the way humans solve the problems by understanding, analyzing, and remembering. In this paper, the authors have proposed a framework for career counseling of students using ML and AI techniques. ML-based White and Black Box models analyze the educational dataset comprising of academic and employability attributes that are important for the job placements and skilling of the students. In the proposed framework, White Box and Black Box models get trained over an educational dataset taken in the study. The Recall and F-Measure score achieved by the Naive Bayes for performing predictions is 91.2% and 90.7% that is best compared to the score of Logistic Regression, Decision Tree, SVM, KNN, and Ensemble models taken in the study.

**Pros:**

- Integration of Machine Learning and AI Techniques
- Expert System with Decision Support

**Cons:**

- Limited Evaluation Metrics
- Potential Bias and Fairness Issues

### 2.2: PAPER 2: AI CHATBOT TO SUPPORT CAREER GUIDANCE

The thesis has explored, analyzed and understood the difficulties, shortcomings and problems arising in career counseling, enrollment support. Building a chatbot is a perfect solution. A chatbot can be available 24/7 for 365 days. Chatbot also receives and processes requests of students / parents automatically and quickly and uniformly in replying content, especially optimized in repetitive scenarios. Thus we decided to do research to understand this situation. Then creating a dataset supports vocational guidance and advising education enrollment activities. We also design and integrate chatbot into the school system to support the admissions counseling process. Besides, the thesis has successfully built a structured dataset about enrollment orientation and apply of natural language processing, machine learning to build identification models. The thesis will contribute to overcome and improve the performance of university admissions consulting.

**Pros:**

- 24/7 Availability and Uniform Responses
- Structured Dataset and ML Models

**Cons:**

- Dependency on Technology
- Potential Lack of Personalization

**2.3: PAPER 3: APPLICATION EXPERT SYSTEM CAREER GUIDANCE**

The use of information technology and communication (ICT) is widely applied in student career guidance services. In addition to increasing student accessibility in obtaining career information, the use of ICT can also be used as a reference system in decision making known as expert systems. The focus of the research in this paper is to analyze the results of research related to the impact of the application of expert systems in student guidance on aspects of educational guidance, educational evaluation and aspects of academic career guidance and work guidance. Overall, review findings on the topics and sub-topics analyzed show that the application of expert systems in educational guidance greatly helps students achieve: learning success, specialization in education and training, student performance, achievement and self evaluation. Given the lack of development of expert systems in the field of student career guidance, it is very necessary for continuous development, especially in aspects of uncertainty arising from application and user factors.

**Pros:**

- Enhanced Accessibility and Decision-Making Support
- Positive Impact on Educational and Academic Guidance

**Cons:**

- Potential Limitations in Development and Application
- Need for Continuous Development and Improvement

## CHAPTER 3: SOFTWARE REQUIREMENTS SPECIFICATION

### 3.1: ASSUMPTIONS AND DEPENDENCIES

- **Accuracy of External Data Sources:** The project relies on the accuracy and comprehensiveness of data from external sources like Quora and learnEnglishEveryday for question creation and the Student's Interest and Learning Survey of 2019 for genre classification. The success of the project hinges on the quality and relevance of information obtained from these external sources.
- **Validity of Genre Classification:** The project categorizes users into four genres (Artistic, Adventure, Sports, Logical) based on their questionnaire responses. The accuracy of career recommendations depends heavily on the effectiveness of this genre classification system.
- **Effectiveness of Game Recommendations & Difficulty Levels:** The recommended games (e.g., Chess for Logical, Dodge the Wrench for Adventure) effectively engage users and provide a reliable indicator of their skill level. The project depends on a well-defined correlation between game performance and aptitude within each genre.
- **Aptitude Question Source & Relevance:** The free Trivia API provides a comprehensive and relevant set of aptitude questions aligned with the user's predicted genre and difficulty level. The quality of career recommendations relies heavily on the accuracy and relevance of the aptitude questions sourced from the Trivia API.
- **Standard English Dictionary for Career Matching:** A standard English dictionary sufficiently links user interests to potential career paths. The project depends on the chosen dictionary providing accurate mappings between user interests and potential career options.
- **User Engagement & Honesty:** Users will honestly answer questionnaire prompts, actively participate in recommended games, and thoughtfully answer the final MCQ questions. The project's effectiveness relies on user engagement and truthful responses throughout the process.

### 3.2: FUNCTIONAL REQUIREMENTS

Table 1: Functional Requirements: Questionnaire Chat

ID	Requirement
FR01	Every user must create a unique profile by registering with a valid email address and password.
FR02	The system must assign a unique user ID to each registered user for reference and data management.
FR03	The questionnaire chat must be initiated upon user login or registration.
FR04	The system must present 20 questions to the user, selected from authorized websites, covering various aspects of career preferences and interests.
FR05	User responses to the questionnaire must be stored securely and linked to their unique user ID.
FR06	The system must analyze user responses using NLP techniques to predict their genre preference (Artistic, Adventure, Sports, Logical).

Table 2: Functional Requirements: Game Recommendation

ID	Requirement
FR07	Based on the predicted genre, the system must recommend relevant games to the user.
FR08	Each genre (Adventure, Sports, Logical) must have a specific set of recommended games stored in the system database.
FR09	Recommendations should be tailored to the user's interests and hobbies, utilizing their genre preference and previous gaming history.
FR10	Game difficulty levels must be adjusted dynamically based on the user's performance in previous stages and their overall skill level.
FR11	The system must implement a scoring system for each game to track user progress and performance.



Table 3: Functional Requirements: Career Aptitude Assessment

ID	Requirement
FR12	Aptitude questions related to the predicted genre must be presented to the user.
FR13	The difficulty level of questions should be determined based on the user's genre preference and previous gaming performance.
FR14	The system must match user responses with relevant career domains using standard English dictionaries and career databases.
FR15	The top three matched career options must be presented to the user based on their aptitude assessment.
FR16	The system must provide a score or probability for each career option, indicating the user's compatibility and potential success in each field.

### 3.3: EXTERNAL INTERFACE REQUIREMENTS

#### 3.3.1: User Interfaces

- Users will interact with CareerSage primarily through a web-based interface.
- The interface will be designed using HTML, CSS, and Bootstrap for a visually appealing and user-friendly experience.
- The interface will feature interactive chat elements for the questionnaire, game recommendations, and career aptitude assessment.
- Users will also be able to view game recommendations and career options through the interface.

#### 3.3.2: Hardware Interfaces

- CareerSage will be compatible with standard computing hardware such as desktop computers, laptops, tablets, and smartphones.
- Users will need a stable internet connection to access the CareerSage web application.

#### 3.3.3: Software Interfaces

- The backend of CareerSage will be developed using the Flask framework in Python.
- Python libraries such as os, pandas, json, requests, BeautifulSoup, nltk, numpy, sklearn, gensim, and itertools will be utilized for data processing, analysis, and integration.
- The system will interface with external APIs for fetching aptitude questions and possibly game data.

### 3.3.4: Communication Interfaces

- CareerSage will communicate with users through HTTP requests and responses over the internet.
- The system will handle user authentication and session management securely using HTTPS protocols.
- Communication between the frontend and backend components of CareerSage will be facilitated through RESTful APIs.

## 3.4: NON-FUNCTIONAL REQUIREMENTS

### 3.4.1: Performance Requirements

- **Response Time:** The system should respond to user inputs within 1-2 seconds to maintain a smooth user experience.
- **Scalability:** The system should be able to handle concurrent users without significant degradation in performance. It should scale gracefully as the user base grows.
- **Reliability:** The system should have a high level of reliability, with minimal downtime or service disruptions.

### 3.4.2: Safety Requirements

- **Data Privacy:** User data, including questionnaire responses and personal information, must be securely stored and protected against unauthorized access.
- **Error Handling:** The system should gracefully handle errors and exceptions to prevent data loss or corruption.
- **Backup and Recovery:** Regular backups of user data should be performed to ensure data integrity and facilitate recovery in case of system failures.

### 3.4.3: Security Requirements

- **Authentication and Authorization:** User authentication must be implemented to ensure that only authorized users can access the system. Role-based access control should be enforced.

### 3.4.4: Software Quality Attributes

- **Usability:** The system should have a user-friendly interface with intuitive navigation and clear instructions to facilitate user interaction.
- **Maintainability:** The system should be designed with modular components and well-documented code to facilitate maintenance and future enhancements.

- **Portability:** The system should be platform-independent and compatible with different web browsers and operating systems.
- **Performance Efficiency:** The system should optimize resource usage to minimize server load and maximize responsiveness.

### 3.5: SYSTEM REQUIREMENTS

#### 3.5.1: Database Requirements

Table 4: Database Requirements

Requirement	Details
Database Management System	MySQL (version 8.0)

#### 3.5.2: Software Requirements (Platform Choice)

Table 5: Software Requirements

Requirement	Details
Frontend Framework	HTML, CSS, Bootstrap (version 5.1.3)
Backend Framework	Flask (version 2.1.0)
Python Libraries	os (version 3.10.3), pandas (version 1.3.3), json (version 2.0.9), requests (version 2.26.0), BeautifulSoup (version 4.10.0), nltk (version 3.6.5), numpy (version 1.21.4), sklearn (version 0.24.2), gensim (version 4.1.2), itertools (version 3.10)

#### 3.5.3: Hardware Requirements

Table 6: Hardware Requirements

Requirement	Details
Server Specifications	CPU: Intel Xeon or equivalent, RAM: 8GB or higher, Storage: 100GB or higher
Network	Stable internet connection

### 3.6: ANALYSIS MODEL: SDLC MODEL - ITERATIVE MODEL

CareerSage will follow the Iterative Model of the Software Development Life Cycle (SDLC). This model emphasizes a cyclical approach to development, where the project is divided into smaller iterations or increments. Each iteration goes through the phases of planning, design, implementation, and testing. After each iteration, there is a review and feedback process, which allows for adjustments and improvements to be made before proceeding to the next iteration.

### 3.6.1: Key Characteristics of Iterative Model

- **Incremental Development:** CareerSage will be developed incrementally, with each iteration adding new features or enhancements to the system. This allows for a gradual buildup of functionality and allows stakeholders to see progress early on.
- **Feedback-driven:** The iterative model encourages frequent feedback from stakeholders, including users, to ensure that the system meets their needs and expectations. Feedback is used to refine requirements, make design decisions, and prioritize features for future iterations.
- **Flexible and Adaptive:** The iterative model is well-suited for projects with evolving requirements or uncertain constraints. It allows for flexibility in accommodating changes and adapting to new insights or challenges discovered during development.
- **Continuous Improvement:** Each iteration provides an opportunity for learning and improvement. Lessons learned from one iteration can be applied to subsequent iterations, leading to a more refined and polished final product.

### 3.6.2: Phases of Iterative Model

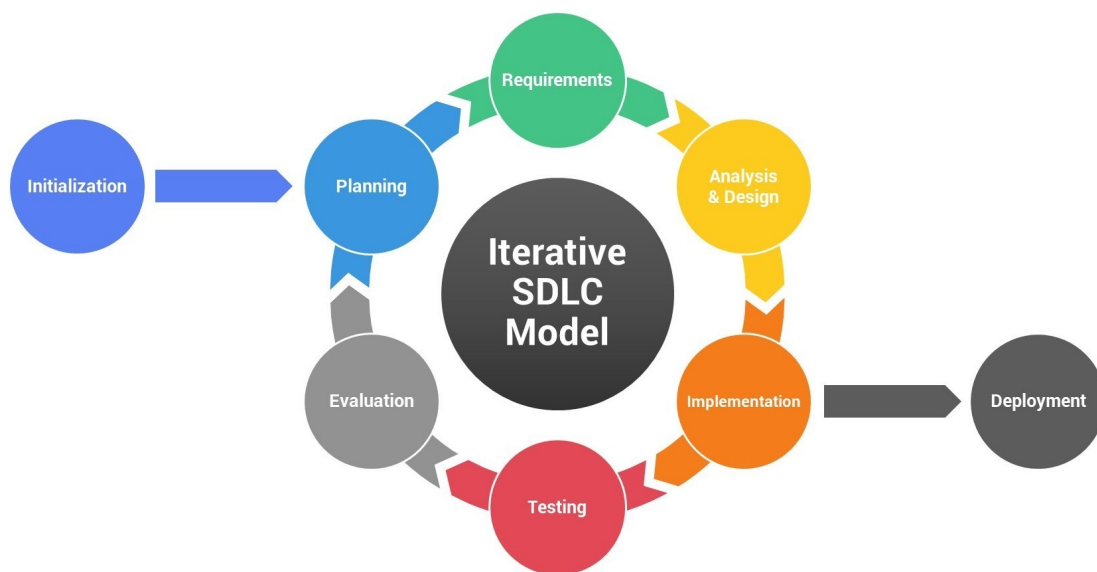


Figure 1: Phases of Iterative Model.

- **Planning:** In this phase, the overall project goals and objectives are defined. Requirements are gathered and prioritized, and a high-level project plan is developed. The scope of the project is defined, and initial architecture and design decisions are made.
- **Design:** The design phase focuses on translating requirements into detailed system specifications. Architectural design, database design, and user interface design are all developed iteratively, with each iteration building upon the previous one.

- **Implementation:** During this phase, the system is built according to the design specifications. Development tasks are divided into smaller units, and coding is done incrementally. Continuous integration and testing ensure that each increment is functional and meets quality standards.
- **Testing:** Testing is an integral part of each iteration, ensuring that the system meets functional and non-functional requirements. Automated testing tools and manual testing techniques are used to identify and fix defects early in the development process.
- **Review and Feedback:** After each iteration, there is a review and feedback session with stakeholders. The system is demonstrated, and feedback is solicited to identify areas for improvement or changes to be made in the next iteration.
- **Incremental Deployment:** As each iteration is completed and validated, it is deployed to a staging environment for user acceptance testing. Once approved, it is deployed to production, allowing users to start benefiting from the new features or enhancements.

## CHAPTER 4: SYSTEM DESIGN

### 4.1: SYSTEM ARCHITECTURE

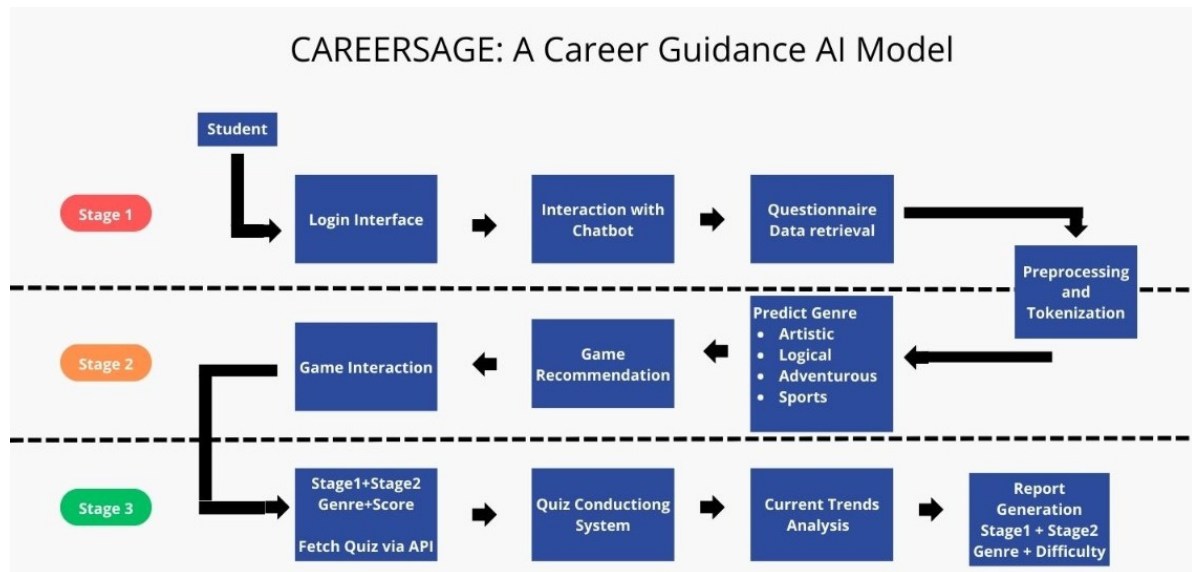


Figure 2: System Architecture

The architecture of CareerSage is designed to facilitate seamless interaction between users (students) and the system, leveraging a combination of frontend and backend components. The system is structured into three main phases, each handling specific functionalities to guide users in exploring their career preferences and options.

#### 4.1.1: Phase 1: Discovery

- **Student Login Interface:** Users access CareerSage through a login interface, providing authentication credentials to access the system.
- **Interaction with Chatbot:** Upon successful login, users engage with a chatbot interface, which serves as the primary mode of interaction. The chatbot guides users through the career exploration process by presenting questions and collecting responses.
- **Questionnaire Data Retrieval:** The chatbot retrieves questions from a predefined questionnaire dataset, possibly sourced from authorized websites. User responses are collected for further processing.

#### 4.1.2: Phase 2: Recommendation

- **Preprocessing and Tokenization:** User responses undergo preprocessing and tokenization, where text is cleaned and organized into tokens for analysis.
- **Genre Prediction:** Using natural language processing (NLP) techniques, the system

predicts the user's genre preference based on their responses. Genres such as Artistic, Logical, Adventure, and Sports are identified.

- **Game Recommendation:** Depending on the predicted genre, the system recommends relevant games to the user. These recommendations are tailored to the user's interests and preferences, encouraging engagement.
- **Game Interaction:** Users interact with the recommended games, exploring different activities and challenges within each genre.

#### 4.1.3: Phase 3: Career Matching

- **Quiz Conduction System:** Users proceed to a quiz conduction system, where aptitude questions are presented based on their genre preference and gaming performance.
- **Trivia API Integration:** The system interfaces with a Trivia API to fetch quiz questions relevant to the user's career interests and genre.
- **Current Trends Analysis:** User responses and quiz performance data are analyzed to identify current trends and patterns in career preferences and aptitude.
- **Report Generation:** Based on the analysis, CareerSage generates personalized reports for users, highlighting their career interests, strengths, and potential opportunities. These reports serve as valuable insights for users to make informed decisions about their future career paths.

## 4.2: MATHEMATICAL MODEL

- **User Profile Representation:**

- Each user  $U_i$  is represented as a vector of features:

$$U_i = (u_{i1}, u_{i2}, \dots, u_{in})$$

- Features may include demographic information, questionnaire responses, gaming behavior, and quiz performance.

- **Genre Prediction:**

- Genre prediction involves mapping user profiles to genre categories using a classifier.
- Let  $G$  be the set of genre categories:  $G = \{g_1, g_2, g_3, g_4\}$  (Artistic, Logical, Adventure, Sports).
- The predicted genre for user  $U_i$ , denoted as  $P_i$ , is determined by:

$$P_i = \operatorname{argmax}_{g \in G} Pr(G = g | U_i)$$

- **Game Recommendation:**

- Game recommendation involves selecting the most suitable games for each genre.
- Let  $R_i$  denote the set of recommended games for user  $U_i$  based on predicted genre  $P_i$ .

- **Aptitude Assessment:**

- Aptitude assessment includes administering quizzes and analyzing user performance.
- Let  $Q$  be the set of quiz questions and  $A_i$  be the set of answers provided by user  $U_i$ .
- Quiz score  $S_i$  for user  $U_i$  is calculated based on the correctness of answers and possibly additional factors like time taken:

$$S_i = \frac{\text{Total correct answers}}{\text{Total questions}} \times 100$$

- **Trend Analysis:**

- Trend analysis involves identifying patterns and trends in user behavior and preferences over time.
- This can be done using statistical methods such as regression analysis, time series analysis, or machine learning algorithms.

- **Report Generation:**

- The final step involves generating a personalized report for each user summarizing their career preferences, game recommendations, and quiz performance.
- The report may include visualizations such as charts or graphs to present the data in an understandable format.

- **System Evaluation:**

- The effectiveness of CareerSage can be evaluated using metrics such as user satisfaction, accuracy of genre prediction, relevance of game recommendations, and improvement in aptitude assessment scores over time.

### 4.3: DATA FLOW DIAGRAMS

he classic visual representation of how information moves through a system is a data flow diagram (DFD). A tidy and understandable DFD can graphically represent the appropriate quantity of the system demand. It can be done manually, automatically, or both. It demonstrates how information enters and exits the system, what modifies the data, and where information is kept. A DFD's goal is to outline the boundaries and scope of the structure as a whole. It can be utilised as a communication tool among a system analyst and any participant in the sequence that serves as the foundation for system redesign. The DFD is also known as a bubble chart or data flow graph



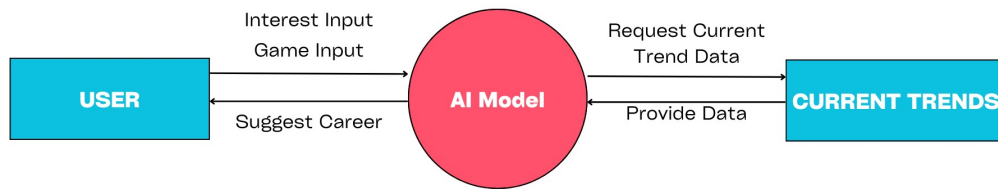
**4.3.1: Level 0 Data Flow Diagram**

Figure 3: Level 0 Data Flow Diagram

In the Level 0 DFD, users input their interests into the AI model, which processes the data to predict their genre preferences and suggest relevant career options. Additionally, users can provide feedback on games they enjoy, which is used to refine future recommendations. The AI model also requests current trends data, which is utilized to keep the system up-to-date and provide users with the latest information relevant to their career exploration journey.

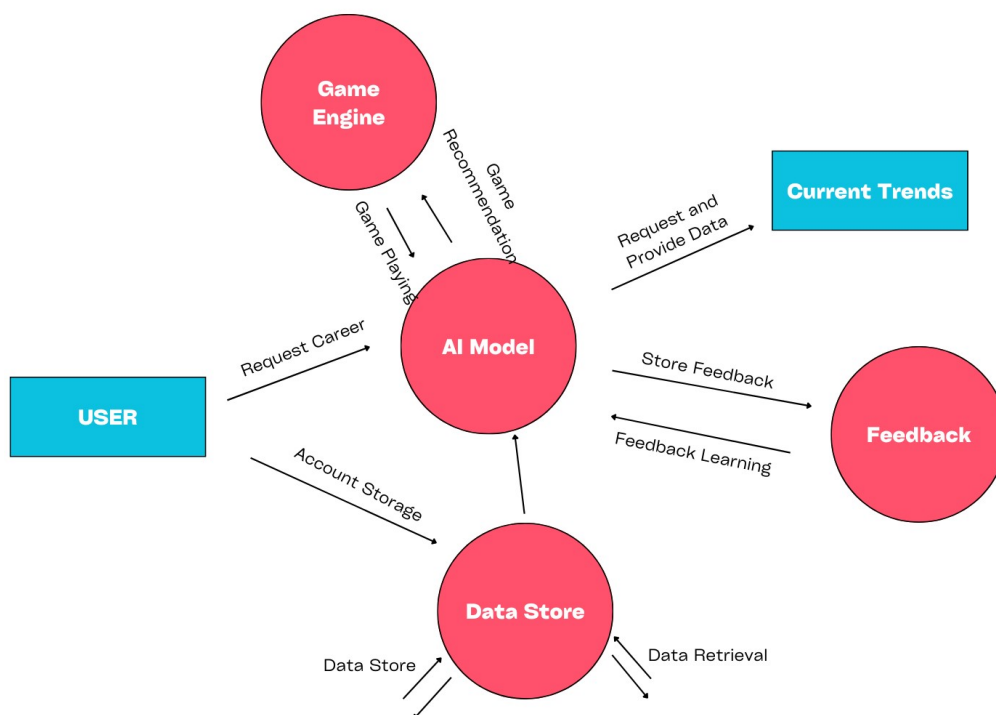
**4.3.2: Level 1 Data Flow Diagram**

Figure 4: Level 1 Data Flow Diagram

In the Level 1 DFD, users interact directly with the AI model, providing input on their interests and receiving recommendations for games and career options. The AI model accesses a data store to retrieve and store user data for analysis and future reference. It communicates with a game engine to provide tailored game recommendations based on user preferences. Additionally, the

AI model fetches current trends data to ensure the system remains updated, while feedback from users is incorporated back into the AI model to improve the accuracy of its recommendations over time.

## 4.4: ENTITY RELATIONSHIP DIAGRAMS

An Entity-Relationship (ER) diagram is a visual representation of the entities, attributes, and relationships within a database or information system. It primarily depicts the logical structure of the data model, showing how different entities relate to each other. Entities represent objects or concepts, attributes describe the properties of those entities, and relationships illustrate the connections between entities. ER diagrams are widely used in database design to help stakeholders understand and communicate the structure and behavior of a system's data.

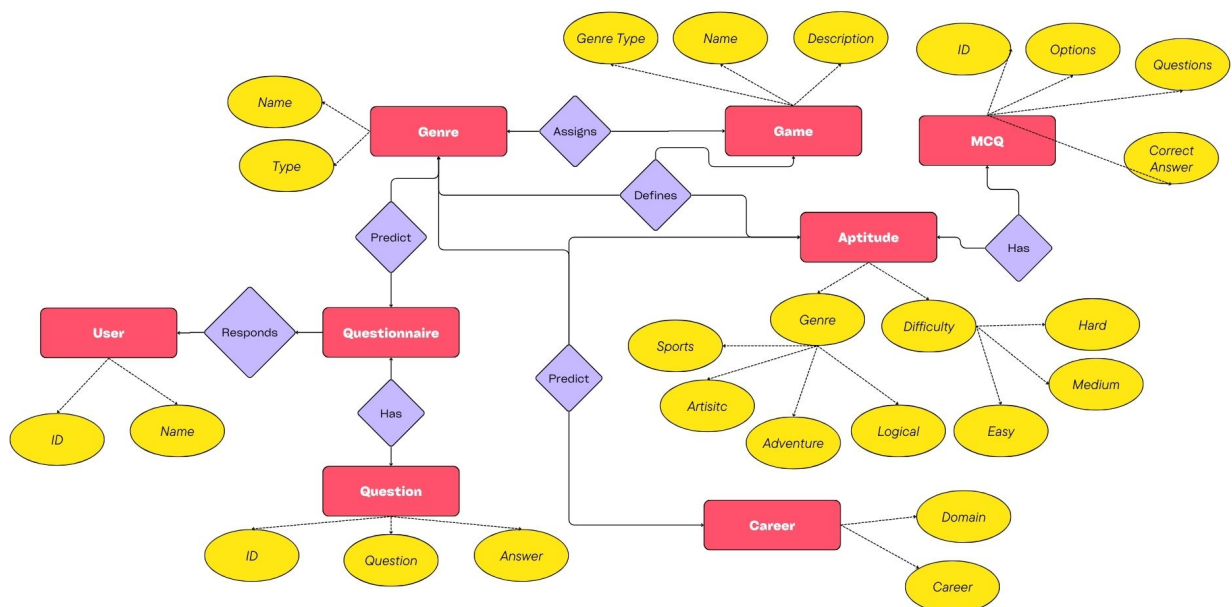


Figure 5: Entity Relationship Diagram

## 4.5: UML DIAGRAMS

### 4.5.1: Use Case Diagram

A use case diagram is a visual representation of interactions between actors (users) and a system to achieve specific goals. It outlines the functionalities or services provided by a system and the interactions between the system and its users. Use case diagrams typically consist of actors, use cases (represented by ovals), and the relationships between them (represented by lines). These diagrams are valuable tools in software development for capturing and illustrating the functional requirements of a system in a clear and concise manner, aiding in communication between stakeholders and guiding the design and development process.

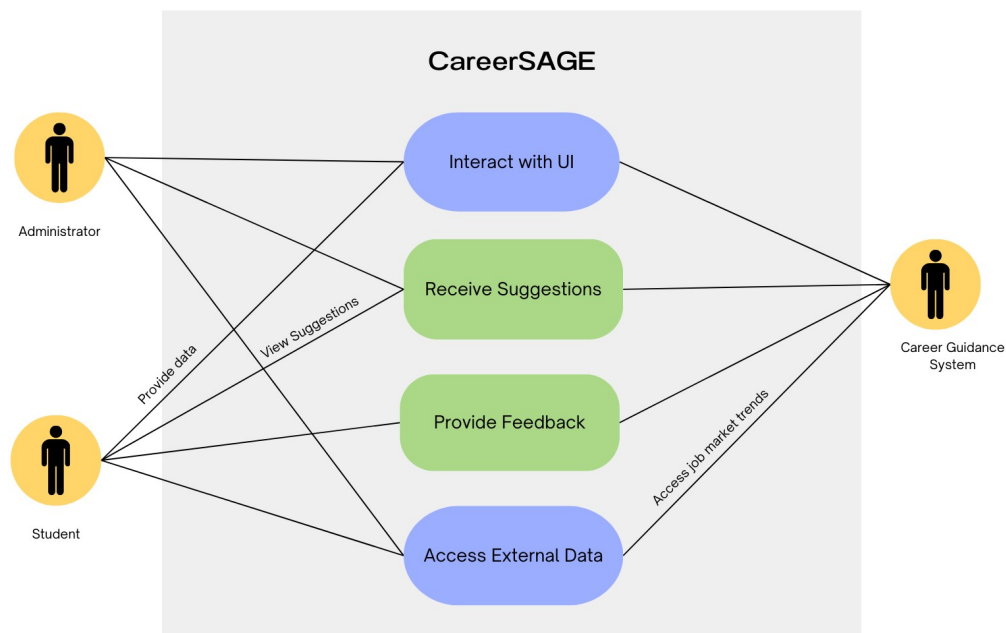


Figure 6: Use Case Diagram

#### 4.5.2: Sequence Diagram

A sequence diagram is a type of interaction diagram that illustrates the flow of messages or interactions between objects or components in a system over time. It shows the sequence of events or actions that occur in response to stimuli or messages, typically depicting the behavior of a single scenario or use case within the system. Sequence diagrams are useful for visualizing and understanding the dynamic behavior of a system, including the order of method calls, the timing of interactions, and the collaboration between different components or objects.

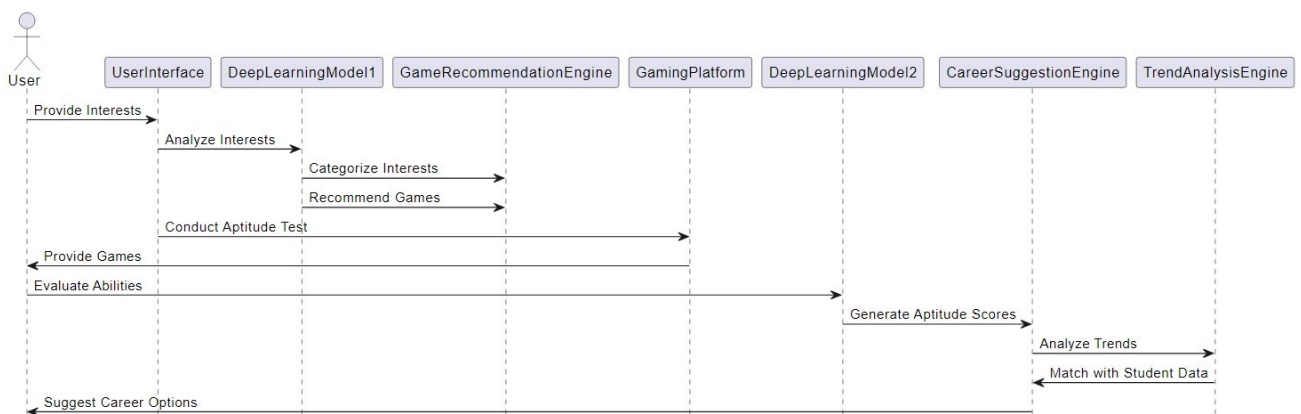


Figure 7: Sequence Diagram

### 4.5.3: Class Diagrams

A class diagram is a type of static structure diagram in the Unified Modeling Language (UML) that represents the structure and behavior of a system or application through classes, interfaces, associations, and their relationships. It provides a visual representation of the system's objects, their attributes, methods, and the associations between them. Class diagrams are widely used in software engineering during the design phase to visualize and communicate the architecture of a system, aiding in understanding, planning, and implementing complex software systems.

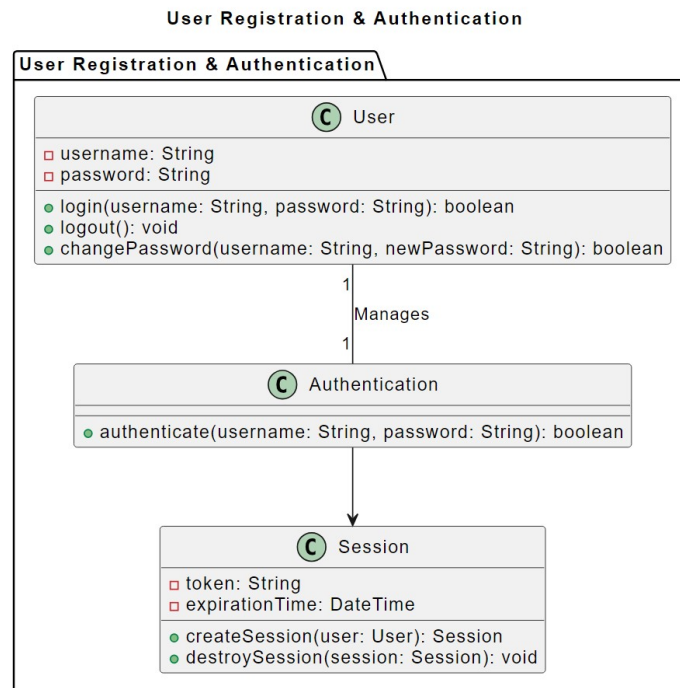


Figure 8: Class Diagram: User Registration and Authentication

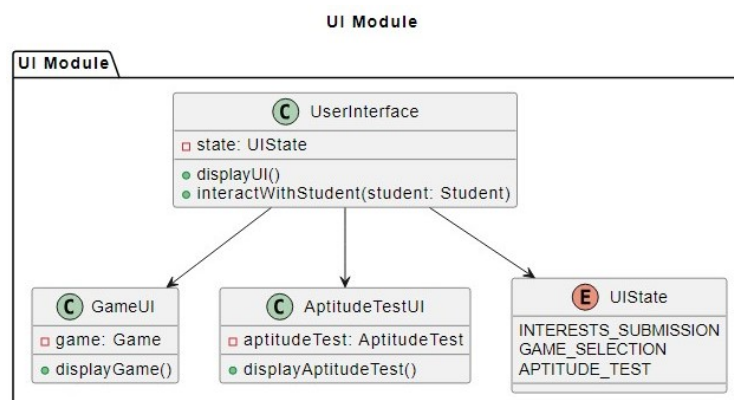


Figure 9: Class Diagram: UI Module

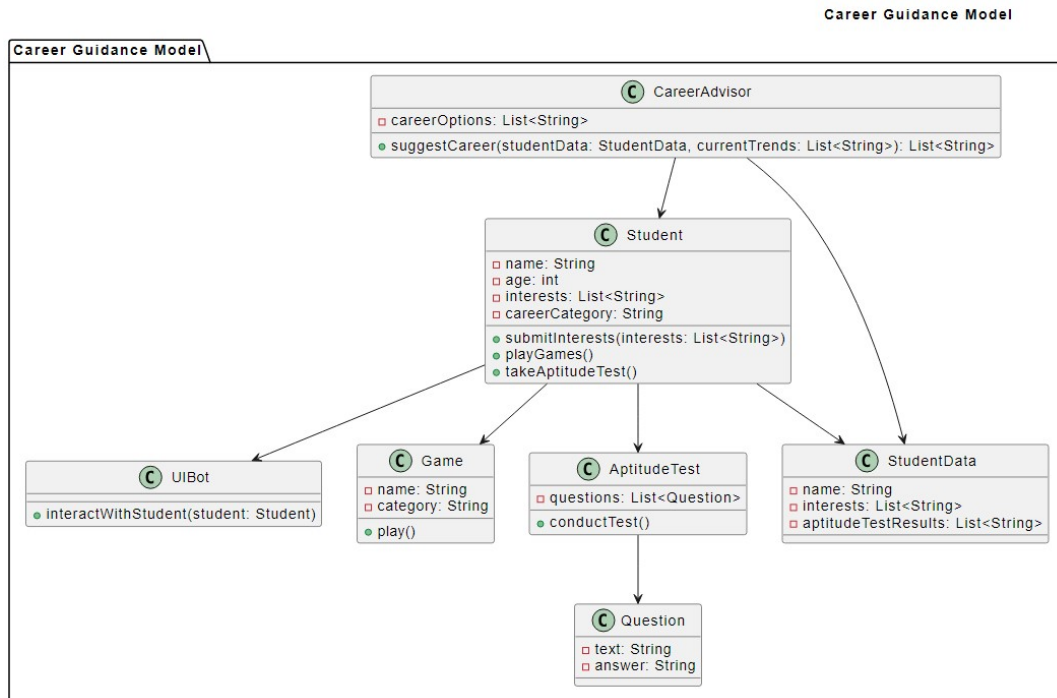


Figure 10: Class Diagram: Career Guidance Model

#### 4.5.4: Activity Diagram

An activity diagram is a graphical representation of workflows, processes, or algorithms, often used in software engineering to model the steps of a system or business process. It depicts the flow of activities, actions, and decisions, showcasing the sequence of actions and the conditions governing their execution. With nodes representing actions and transitions representing the flow between them, activity diagrams provide a visual means to understand, analyze, and communicate complex processes, aiding in system design, documentation, and improvement efforts.

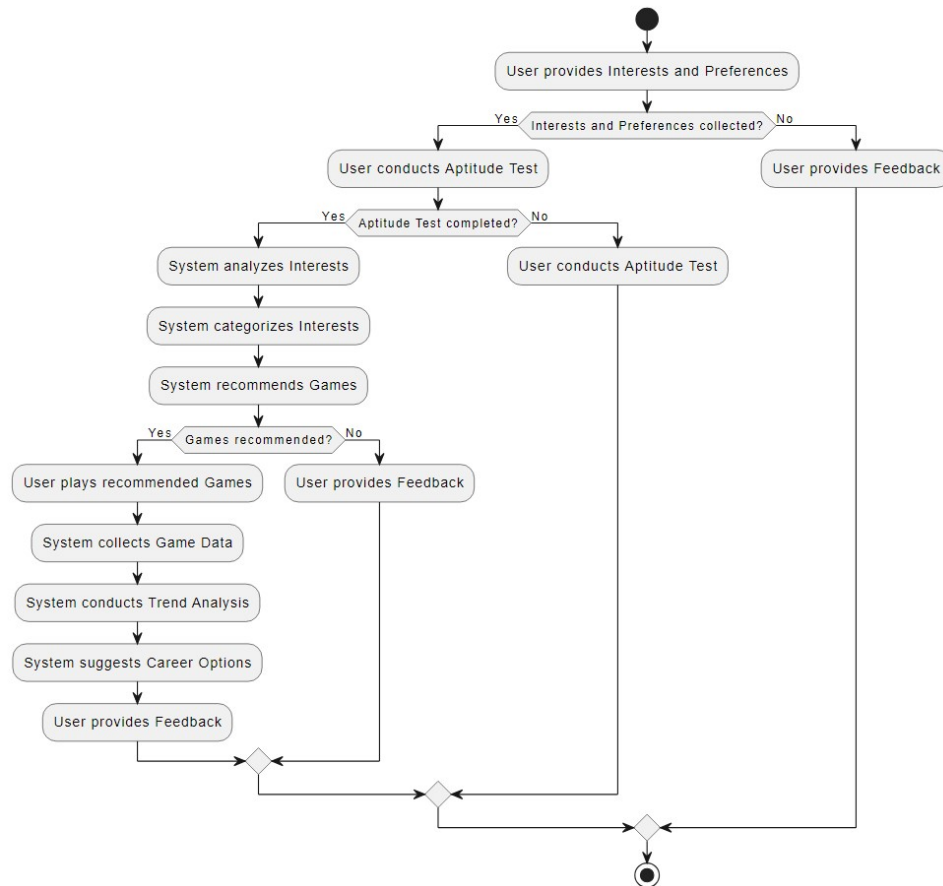


Figure 11: Activity Diagram

## CHAPTER 5: PROJECT PLAN

### 5.1: PROJECT ESTIMATE

#### 5.1.1: Reconciled Costs

The COCOMO (Constructive Cost Model) is a widely-used estimation model in software engineering that provides a structured approach to estimating the effort, time, and cost required for software development projects. Introduced by Barry Boehm in 1981, COCOMO categorizes projects into three modes: Basic, Intermediate, and Advanced. It estimates effort based on lines of code (LOC) or function points and adjusts for various project factors such as complexity, development environment, and team experience. By applying a set of predefined formulas and coefficients, COCOMO assists project managers in making informed decisions regarding resource allocation, scheduling, and budgeting throughout the software development lifecycle.

**Software Size**      Sizing Method Source Lines of Code ▾

[SLOC](#)      % Design Modified      % Code Modified      % Integration Required      Assessment and Assimilation (0% - 8%)      Software Understanding (0% - 50%)      Unfamiliarity (0-1)

New	3486					
Reused	11567	0	0			
Modified	5642	36	49	58	2	31

**Software Scale Drivers**

Precedentedness	Nominal ▾	Architecture / Risk Resolution	Low ▾	Process Maturity	Nominal ▾
Development Flexibility	High ▾	Team Cohesion	Low ▾		

**Software Cost Drivers**

<b>Product</b>		<b>Personnel</b>		<b>Platform</b>	
Required Software Reliability	Low ▾	Analyst Capability	High ▾	Time Constraint	Nominal ▾
Data Base Size	Low ▾	Programmer Capability	High ▾	Storage Constraint	Nominal ▾
Product Complexity	Very High ▾	Personnel Continuity	Nominal ▾	Platform Volatility	Nominal ▾
Developed for Reusability	Low ▾	Application Experience	High ▾	<b>Project</b>	
Documentation Match to Lifecycle Needs	Nominal ▾	Platform Experience	High ▾	Use of Software Tools	Low ▾
		Language and Toolset Experience	Very High ▾	Multisite Development	Nominal ▾
				Required Development Schedule	Low ▾

**Maintenance** On ▾

Annual Change Size (ESLOC)	2364	Maintenance Duration (Years)	1
Software Understanding (0%-50%)	34	Unfamiliarity (0-1)	

**Software Labor Rates**

Cost per Person-Month (Dollars)	85
---------------------------------	----

Figure 12: Factors for Cost Estimation



**Results****Software Development (Elaboration and Construction)**

Effort = 15.3 Person-months  
 Schedule = 7.2 Months  
 Cost = \$1303

**Staffing Profile**

Your project is too small to display a staffing profile due to truncation.

Total Equivalent Size = 6222 SLOC  
 Effort Adjustment Factor (EAF) = 0.68

**Acquisition Phase Distribution**

Phase	Effort (Person-months)	Schedule (Months)	Average Staff	Cost (Dollars)
Inception	0.9	0.9	1.0	\$78
Elaboration	3.7	2.7	1.4	\$313
Construction	11.7	4.5	2.6	\$991
Transition	1.8	0.9	2.0	\$156

**Software Effort Distribution for RUP/MBASE (Person-Months)**

Phase/Activity	Inception	Elaboration	Construction	Transition
Management	0.1	0.4	1.2	0.3
Environment/CM	0.1	0.3	0.6	0.1
Requirements	0.3	0.7	0.9	0.1
Design	0.2	1.3	1.9	0.1
Implementation	0.1	0.5	4.0	0.3
Assessment	0.1	0.4	2.8	0.4
Deployment	0.0	0.1	0.3	0.6

**Maintenance**

Annual Maintenance Effort = 5.2 Person-Months  
 Annual Maintenance Cost = \$443  
 Total Maintenance Cost = \$443

Figure 13: Estimated cost and man-power

**5.1.2: Project Resources**

- **Human Resources:**

- **Software Developers:** Responsible for frontend and backend development, database management, and system integration.
- **Data Scientists/NLP Experts:** Required for implementing the AI model for genre prediction and analyzing user responses.
- **UI/UX Designers:** Responsible for designing user interfaces and ensuring a seamless user experience.
- **Project Manager:** Oversees the project, coordinates tasks, manages timelines, and communicates with stakeholders.

- **Hardware Resources:**

- **Development Machines:** Desktop or laptop computers for software development and testing.
- **Server:** Hosting environment for deploying the web application and database.



- **Software Resources:**
  - **Development Tools:** IDEs (Integrated Development Environments) such as PyCharm, Visual Studio Code, or Sublime Text for coding.
  - **Version Control System:** Git for code versioning and collaboration.
  - **Python Environment:** Required for backend development, utilizing frameworks like Flask and libraries like NLTK, NumPy, and Scikit-learn.
  - **Database Management System:** PostgreSQL, MySQL, or SQLite for storing user data, game recommendations, and questionnaire responses.
  - **Web Server:** Apache or Nginx for hosting the web application.
  - **Virtual Environment:** Tools like virtualenv or conda for managing Python dependencies and environments.
- **Data Resources:**
  - **Questionnaire Dataset:** A dataset containing a variety of questions sourced from authorized websites like Quora and LearnEnglishEveryday.
  - **Game Database:** A collection of games categorized by genre for providing recommendations to users.
  - **Current Trends Data:** Access to APIs or databases for fetching current trends relevant to career options.
- **External Services/APIs:**
  - **Trivia API:** Used for fetching aptitude questions for quiz conduction.
  - **Trend Data API:** Access to APIs providing current trends data for career options.

## 5.2: RISK MANAGEMENT

### 5.2.1: Risk Identification

- Have all stakeholders formally agreed upon the project scope, objectives, and specifications to ensure alignment and avoid scope creep?  
Yes
- Are all required software tools, libraries, and APIs readily available and compatible with project requirements?  
Yes
- Are end-users (students, career seekers) fully committed to using CareerSage for career guidance, ensuring active engagement and feedback throughout the development process?

Yes

- Does the software development team possess the necessary skills in deep learning, NLP, web development, and database management to successfully implement CareerSage?

Yes

- Are project requirements stable, or are there potential changes or additions expected during the development process?

Yes

- Will users find CareerSage intuitive and valuable, or is there a risk of low adoption rates due to usability issues or lack of perceived usefulness?

Yes

- Are there potential challenges in integrating external APIs for game recommendations and current trends data, considering variations in data formats and API availability?

No

- Are there constraints in terms of project timeline, budget, or team resources that could impact the successful completion of CareerSage within the specified timeframe?

Yes

- Are there dependencies on third-party services or data sources for which reliability and availability cannot be guaranteed?

No

### 5.2.2: Risk Analysis

Table 7: Risk Table

Risk	Potential Impact	Likelihood
Scope Creep	Increased project timeline, budget overruns	High
Technical Challenges	Delays in implementation, compromised system quality	Medium
Data Privacy and Security	Reputational damage, legal consequences	High
User Adoption	Low usage, dissatisfaction among users	Medium
Integration Challenges	System functionality issues, delayed project delivery	Medium
Resource Constraints	Project delays, compromised deliverables	Medium
Dependency on External Services	Disrupted functionality, system downtime	Medium

Table 8: Risk Probability Definitions

Likelihood	Probability of occurrence
High	>50%
Medium	>25%
Low	>0%

### 5.2.3: Overview of Risk Mitigation, Monitoring and Management

Table 9: Risk 1: Scope Creep

Risk ID	1
Risk Description	Scope Creep
Category	Requirements
Source	Stakeholder communication
Probability	Medium
Impact	High
Response	Mitigate
Strategy	Regular stakeholder communication and requirement reviews to manage expectations and align scope.
Risk Status	Identified

Table 10: Risk 2: Technical Challenges

Risk ID	2
Risk Description	Technical Challenges
Category	Technical
Source	Project requirements
Probability	High
Impact	High
Response	Mitigate
Strategy	Conduct thorough research, training, and possibly seek external expertise to address technical complexities.
Risk Status	Identified

Table 11: Risk 3: Data Privacy and Security

Risk ID	3
Risk Description	Data Privacy and Security
Category	Security
Source	Data protection regulations
Probability	High
Impact	High
Response	Mitigate
Strategy	Implement robust encryption techniques, access controls, and compliance with data protection regulations.
Risk Status	Identified

Table 12: Risk 4: User Adoption

Risk ID	4
Risk Description	User Adoption
Category	User Experience
Source	User feedback
Probability	Medium
Impact	High
Response	Mitigate
Strategy	Conduct user testing and feedback sessions to ensure usability and address user needs effectively.
Risk Status	Identified

Table 13: Risk 5: Integration Challenges

Risk ID	5
Risk Description	Integration Challenges
Category	Technical
Source	External APIs documentation
Probability	Medium
Impact	High
Response	Mitigate
Strategy	Thoroughly assess API documentation, conduct integration testing, and have backup plans in place.
Risk Status	Identified

Table 14: Risk 6: Resource Constraints

Risk ID	6
Risk Description	Resource Constraints
Category	Resource
Source	Project timeline and budget
Probability	Medium
Impact	High
Response	Mitigate
Strategy	Prioritize tasks, allocate resources efficiently, and communicate transparently with stakeholders.
Risk Status	Identified

Table 15: Risk 7: Dependency on External Services

Risk ID	7
Risk Description	Dependency on External Services
Category	Technical
Source	External service providers
Probability	Low
Impact	High
Response	Mitigate
Strategy	Have contingency plans in place, explore alternative service providers, and maintain open communication.
Risk Status	Identified

## 5.3: PROJECT SCHEDULE

### 5.3.1: Project Task Set

- **Task 1:** Requirement Gathering and Analysis
- **Task 2:** Designing User Interface
- **Task 3:** Developing Backend Infrastructure
- **Task 4:** Implementing AI Model for Genre Prediction
- **Task 5:** Integrating Game Recommendation System
- **Task 6:** Implementing Data Security Measures
- **Task 7:** Developing Frontend Components
- **Task 8:** Testing and Quality Assurance

- **Task 9:** User Acceptance Testing
- **Task 10:** Deployment and Hosting Setup
- **Task 11:** User Training and Documentation
- **Task 12:** Feedback Collection and Iterative Improvements
- **Task 13:** Monitoring and Maintenance Plan
- **Task 14:** Data Backup and Recovery Setup
- **Task 15:** Project Review and Closure

### 5.3.2: Task Network

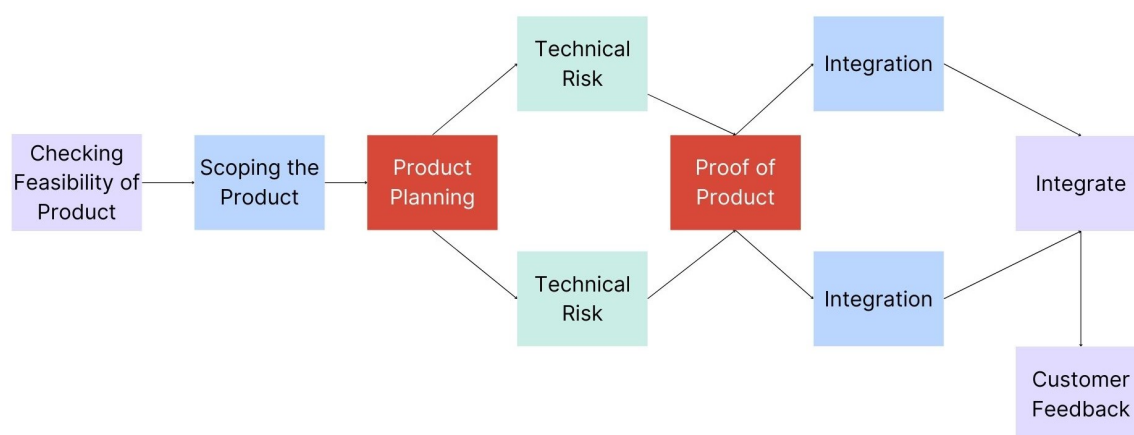


Figure 14: Task Network

### 5.3.3: Timeline Chart

Table 16: Timeline: Semester I

Task	Start Date	End Date	Duration
Domain Selection	25-07-2023	30-07-2023	5
Guide Allocation and Idea Gathering	31-07-2023	02-08-2023	2
Idea Discussion and Abstract Submission	03-08-2023	08-08-2023	5
Project Review I	09-08-2023	14-08-2023	5
Submission of Final Proposal	15-08-2023	18-08-2023	3
Literature Survey	19-08-2023	27-08-2023	8
Domain Scope	28-08-2023	07-09-2023	10
Requirement Analysis	08-09-2023	13-09-2023	5
Features and Objectives	14-09-2023	24-09-2023	10
System Design and Architecture	25-09-2023	10-10-2023	15
Project Review II	11-10-2023	13-10-2023	2
Report Writing and Final Report Submission	14-10-2023	19-10-2023	5

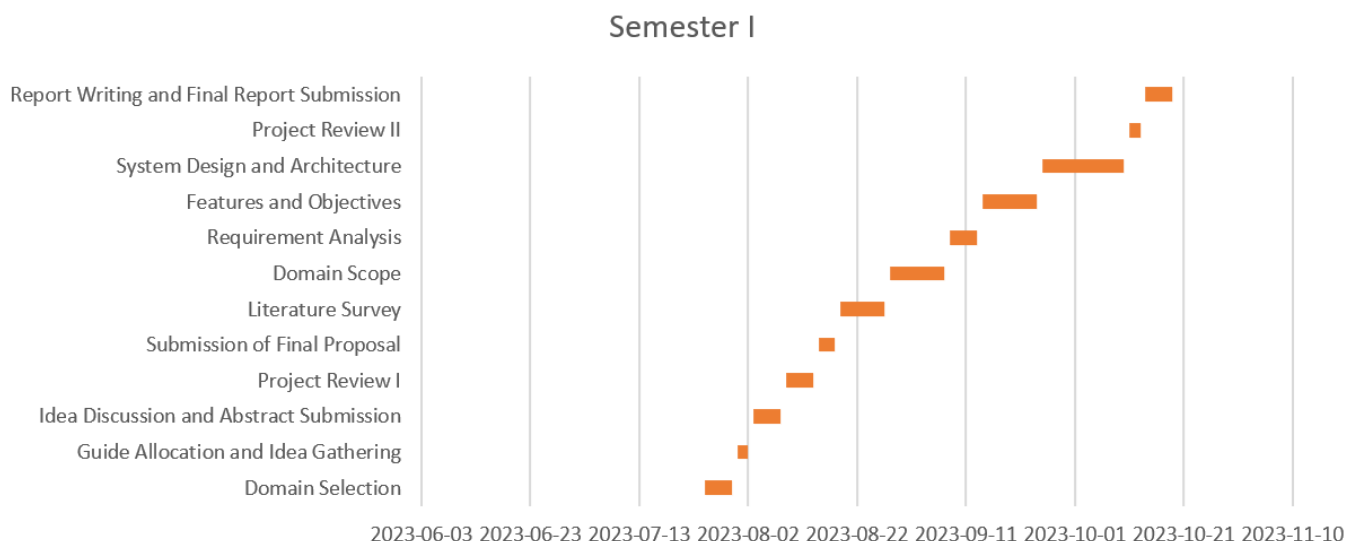


Figure 15: Gantt Chart: Semester I

Table 17: Timeline: Semester II

Task	Start Date	End Date	Duration
Phase 1 Development	05-11-2023	10-11-2023	5
Phase 1 Integration	11-11-2023	23-11-2023	12
Phase 2 Development	24-11-2023	09-12-2023	15
Phase 2 Integration	10-12-2023	16-12-2023	6
Phase 3 Development	17-12-2023	02-01-2024	16
Phase 3 Integration	03-01-2024	15-01-2024	12
Project Review III	16-01-2024	18-01-2024	2
Overall Integration	19-01-2024	31-01-2024	12
Designing Test cases	01-02-2024	05-02-2024	4
Performing Tests	06-02-2024	09-02-2024	3
Applying Changes	10-02-2024	12-02-2024	2
Project Deployment	13-02-2024	16-02-2024	3
Project Demo Presentation	17-02-2024	19-02-2024	2

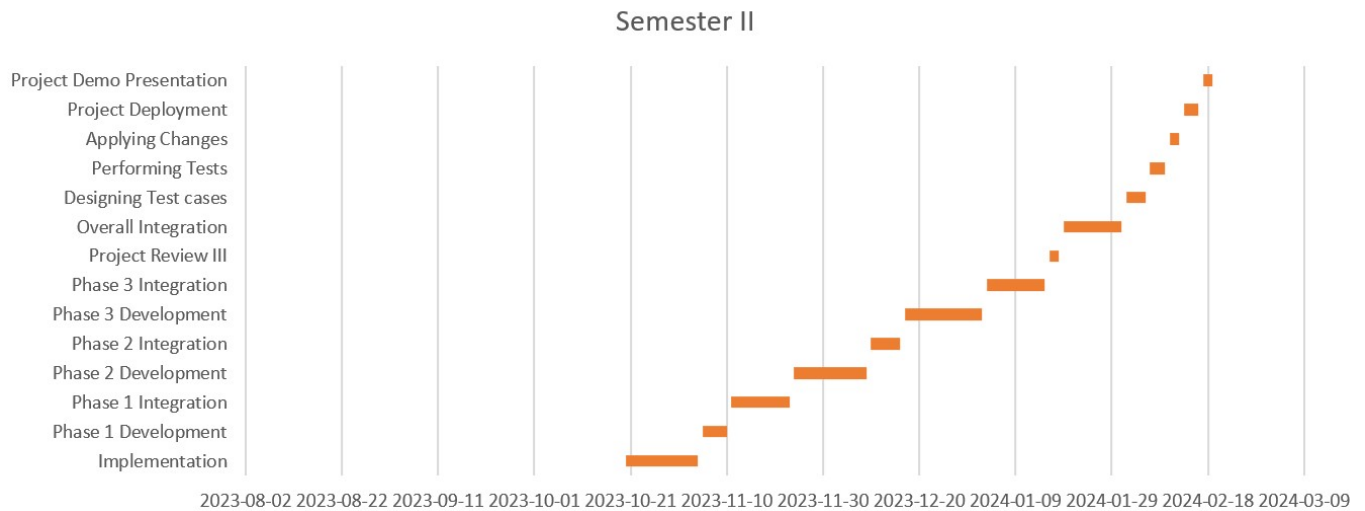


Figure 16: Gantt Chart: Semester II

## 5.4: TEAM ORGANIZATION

Team Members Details:

- **Member 1:** Patil Gaurav Suresh (Team Leader).
- **Member 2:** Sapkale Buddhabhushan Pramod.
- **Member 3:** Kute Vivek Kisan.

### 5.4.1: Team Structure

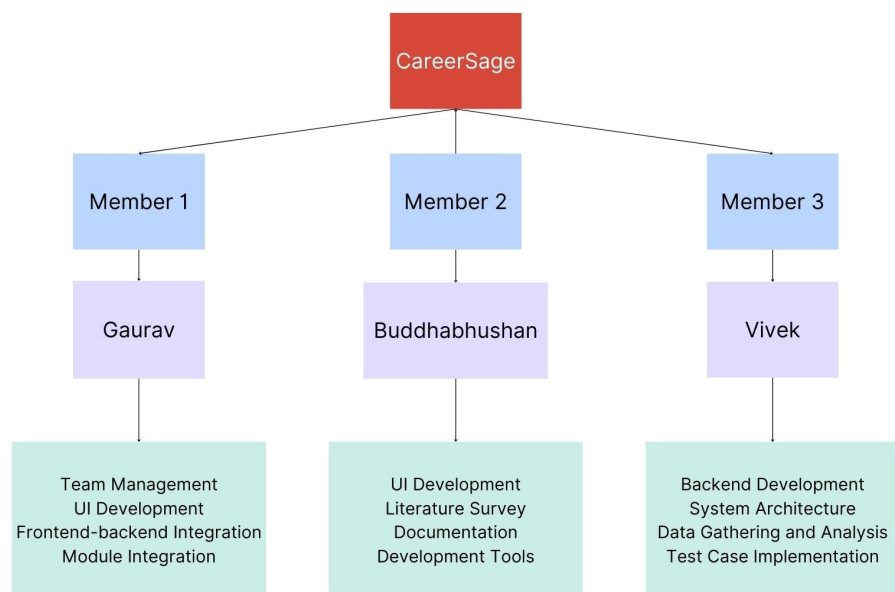


Figure 17: Team Structure



**5.4.2: Management Reporting and Communication**

Table 18: Reporting and Communication

<b>Sr. No.</b>	<b>Reporting Date</b>	<b>Project Activity</b>
1	25-07-2023	Domain Selection
2	03-08-2023	Idea Discussion and Abstract submission
3	09-08-2023	Final Topic selection
4	19-08-2023	Literature Survey
5	09-09-2023	Requirement Analysis
6	25-09-2023	System Design and Architecture
7	14-10-2023	Report Writing and Report Submission
8	15-10-2023	Gather Requirements
9	11-11-2023	Phase 1: Analysis of Questionnaire
10	24-11-2023	Phase 2: Genre Selection
11	10-12-2023	Phase 2: Game Selection
12	03-01-2024	Phase 3: Aptitude
13	19-01-2024	Integration of Modules
14	06-02-2024	Perform Test Cases
15	13-02-2024	Project Deployment
16	17-02-2024	Project Demo

## CHAPTER 6: PROJECT IMPLEMENTATION

### 6.1: OVERVIEW OF PROJECT MODULES

#### 6.1.1: Phase 1: Questionnaire and Genre Prediction

- User Interface for Questionnaire
- Data Collection from User Responses
- Preprocessing and Tokenization of User Responses
- Genre Prediction Model Implementation
- Genre Classification (Artistic, Adventure, Sports, Logical)

#### 6.1.2: Phase 2: Game Recommendation and Interaction

- Game Recommendation Engine
- Integration of Game APIs
- User Interface for Game Recommendations
- Interaction with Recommended Games
- Scoring and Feedback Collection

#### 6.1.3: Phase 3: Aptitude Quiz and Career Match

- Aptitude Quiz Generation
- Quiz Conduction System
- Career Domain Matching
- User Interface for Career Options
- Feedback Collection on Career Choices

### 6.2: TOOLS AND TECHNOLOGIES USED

#### 6.2.1: Frontend Development

- **HTML (Hypertext Markup Language):** HTML is the standard markup language used to create the structure of web pages, defining elements such as headings, paragraphs, and links.
- **CSS (Cascading Style Sheets):** CSS is used to style HTML elements, controlling the presentation and layout of web pages, including aspects like colors, fonts, and spacing.

- **Bootstrap:** Bootstrap is a popular CSS framework that provides pre-designed components and styles, making it easier to create responsive and visually appealing web interfaces.

### 6.2.2: Backend Development

- **Flask:** Flask is a lightweight Python web framework that simplifies backend development by providing tools and libraries for handling HTTP requests, routing URLs, and interacting with databases.

### 6.2.3: Libraries and Frameworks

- **os:** The os module in Python provides functions for interacting with the operating system, allowing tasks like file manipulation, directory navigation, and environment variable access.
- **pandas:** Pandas is a powerful data analysis library in Python, offering data structures and functions for data manipulation, cleaning, and analysis, commonly used for handling structured data.
- **json:** The json module in Python provides functions for encoding and decoding JSON (JavaScript Object Notation) data, facilitating data interchange between different systems and languages.
- **requests:** The requests library in Python simplifies HTTP communication, enabling easy sending and receiving of HTTP requests and responses, commonly used for web scraping and API interaction.
- **BeautifulSoup:** BeautifulSoup is a Python library for parsing HTML and XML documents, providing tools for extracting data from web pages, making it useful for web scraping tasks.
- **nltk:** NLTK (Natural Language Toolkit) is a Python library for natural language processing, offering tools and algorithms for tasks such as tokenization, stemming, and part-of-speech tagging.
- **numpy:** NumPy is a fundamental library for scientific computing in Python, providing support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays.
- **sklearn:** Scikit-learn is a popular machine learning library in Python, offering tools for data mining and data analysis, including algorithms for classification, regression, clustering, and dimensionality reduction.
- **gensim:** Gensim is a Python library for topic modeling and document similarity analysis, providing implementations of algorithms like Latent Dirichlet Allocation (LDA) and word embeddings.

- **itertools:** The itertools module in Python provides functions for creating and manipulating iterators, offering tools for efficient looping and combination generation.

## 6.3: ALGORITHM DETAILS

### 6.3.1: Data Preprocessing

- **Tokenization:** Tokenization is the process of breaking down raw text into smaller units called tokens, which could be words, phrases, or characters. This step divides the text into meaningful chunks for further analysis. Simple tokenization algorithms split the text based on whitespace or punctuation. More advanced tokenizers may consider language-specific rules and handle special cases like contractions and hyphenated words.
- **Text Cleaning:** Text cleaning involves removing unnecessary elements from the text data to improve its quality and reduce noise. This step typically includes removing special characters, punctuation, and stopwords (commonly occurring words that do not carry significant meaning). Simple text cleaning algorithms use regular expressions to remove non-alphanumeric characters and common stopwords. More sophisticated algorithms may also handle stemming (reducing words to their root form) and lemmatization (reducing words to their base or dictionary form).
- **Vectorization:** Vectorization converts text data into numerical vectors that machine learning algorithms can process. This transformation is necessary because most machine learning algorithms require numerical input. Popular vectorization techniques include:
  - **Bag-of-Words (BoW):** Represents text as a sparse matrix where each row corresponds to a document and each column corresponds to a unique word in the corpus. The values in the matrix represent word frequencies or presence/absence indicators.
  - **Term Frequency-Inverse Document Frequency (TF-IDF):** Weights each word in the BoW matrix based on its frequency in the document and its rarity across the entire corpus. This helps prioritize words that are more informative and discriminative.

### 6.3.2: NLP Algorithms

- **Word Embeddings:** Word embeddings are dense vector representations of words in a high-dimensional space, where similar words are closer to each other. Algorithms like Word2Vec or GloVe are commonly used to generate word embeddings. These embeddings capture semantic relationships between words, allowing the algorithm to understand the meaning and context of words in the text data.
- **Text Classification:** Once the text data is preprocessed and represented as numerical vectors using word embeddings, classification algorithms can be applied to classify the text into different categories or genres. For example, Support Vector Machines (SVM),

Naive Bayes, or deep learning models like Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) can be used for text classification tasks.

- **Topic Modeling:** Topic modeling algorithms like Latent Dirichlet Allocation (LDA) can be used to discover latent topics or themes in the text data. This can help in identifying patterns and trends in user responses related to different career interests or preferences.

### 6.3.3: Career Matching Algorithms

- **Pattern Matching:** This algorithm involves matching user input with keywords or phrases associated with specific career domains. For example, if a user expresses interest in "designing," the algorithm might match this input with career domains like "graphic design," "web design," or "interior design." The algorithm may use a predefined dictionary of keywords and phrases associated with different career domains. When a user provides input, the algorithm searches for matches within this dictionary to identify relevant career domains.
- **Semantic Similarity:** Semantic similarity algorithms analyze the meaning and context of user input to identify career domains that closely relate to the user's interests and skills. Techniques such as Word Embeddings or Latent Semantic Analysis (LSA) can be used to represent user input and career domain descriptions in a high-dimensional vector space. By comparing the vectors, the algorithm identifies career domains with similar semantic meanings to the user input.

## **Chapter 7: SOFTWARE TESTING**

### **7.1: TYPES OF TESTING**

#### **7.1.1: Unit Testing**

To assure the quality of each component, each one is tested individually. The goal is to find design and implementation flaws.

#### **7.1.2: Integration Testing**

To assure the quality of the resulting unit, a set of dependent components is evaluated collectively. Utilising a progressive integration strategy, the "big bang" problem is avoided.

#### **7.1.3: System Testing**

The software for the system is tested in its entirety. In order to ensure that it functions properly under severe load and that the system doesn't crash, we tested entire modules one at a time.

#### **7.1.4: Validation Testing**

Various fields, including a valid username, a mobile number, and many more, were subjected to validation testing. Each field has been correctly checked and is undergoing testing.

#### **7.1.5: GUI Testing**

We carried out GUI testing on each module to ensure that it functions as expected and that the GUI is appropriately integrated into all the pages.

**7.2: TEST CASES AND TEST RESULTS**

Table 19: Test Cases and Test Results

<b>Test Case ID</b>	<b>Test Case Description</b>	<b>Expected Result</b>	<b>Actual Result</b>	<b>Pass/Fail</b>
TC001	User accesses the CareerSage website	Homepage should load displaying the login interface	Homepage loaded with login interface	Pass
TC002	User provides valid login credentials	User should be logged in and redirected to the dashboard	User logged in and redirected to dashboard	Pass
TC003	User provides invalid login credentials	Error message should be displayed indicating invalid credentials	Error message displayed: "Invalid username or password"	Pass
TC004	User navigates to game recommendation page	Game recommendation page should load displaying recommended games	Game recommendation page loaded with recommended games	Pass
TC005	User interacts with the aptitude quiz	Aptitude quiz should load displaying questions for the user to answer	Aptitude quiz loaded with questions	Pass
TC006	User submits responses to the aptitude quiz	System should match user responses with career domains and display recommendations	Career recommendations displayed based on quiz responses	Pass
TC007	User provides feedback on career recommendations	System should update recommendations based on user feedback	Career recommendations updated reflecting user feedback	Pass
TC008	User navigates through the frontend interface	Interface should be intuitive and responsive, allowing seamless navigation	Frontend elements display correctly and respond to user actions	Pass
TC009	User logs out of CareerSage	User should be logged out and redirected to the login page	User logged out and redirected to login page	Pass

## CHAPTER 8: RESULTS

### 8.1: OUTCOMES

- **Improved Career Guidance:** CareerSage provides users with personalized career guidance based on their interests, aptitudes, and preferences. By leveraging advanced algorithms and data analysis techniques, CareerSage matches users with suitable career domains, helping them make informed decisions about their future paths.
- **Enhanced User Engagement:** CareerSage offers an interactive and user-friendly platform that engages users through various features such as questionnaire-based genre prediction, game recommendations, aptitude quizzes, and career matching. This increased engagement encourages users to explore different career options and actively participate in the career discovery process.
- **Empowered Decision-Making:** CareerSage empowers users to make informed decisions about their career paths by providing comprehensive insights and recommendations tailored to their individual profiles. By presenting relevant information and options in a structured manner, CareerSage enables users to evaluate their choices effectively and choose careers aligned with their strengths and aspirations.
- **Streamlined Career Exploration Process:** CareerSage streamlines the career exploration process by offering a centralized platform where users can access a wide range of career-related resources and tools. From genre prediction and game recommendations to aptitude quizzes and career matching, CareerSage provides users with everything they need to explore, discover, and pursue their desired career paths efficiently.
- **Increased Career Satisfaction and Fulfillment:** By guiding users towards careers that match their interests, skills, and aspirations, CareerSage helps individuals find roles that they are passionate about and suited for. This alignment between personal preferences and professional choices leads to increased career satisfaction, fulfillment, and long-term success for users in their chosen fields.



## 8.2: SCREENSHOTS

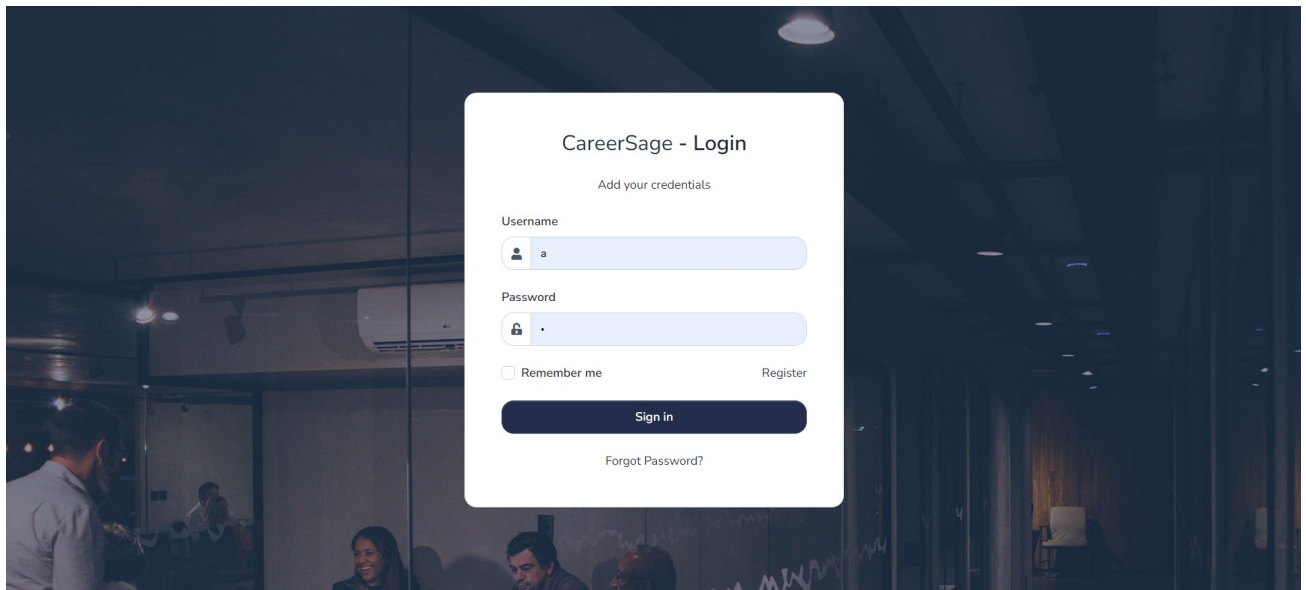


Figure 18: Login Page

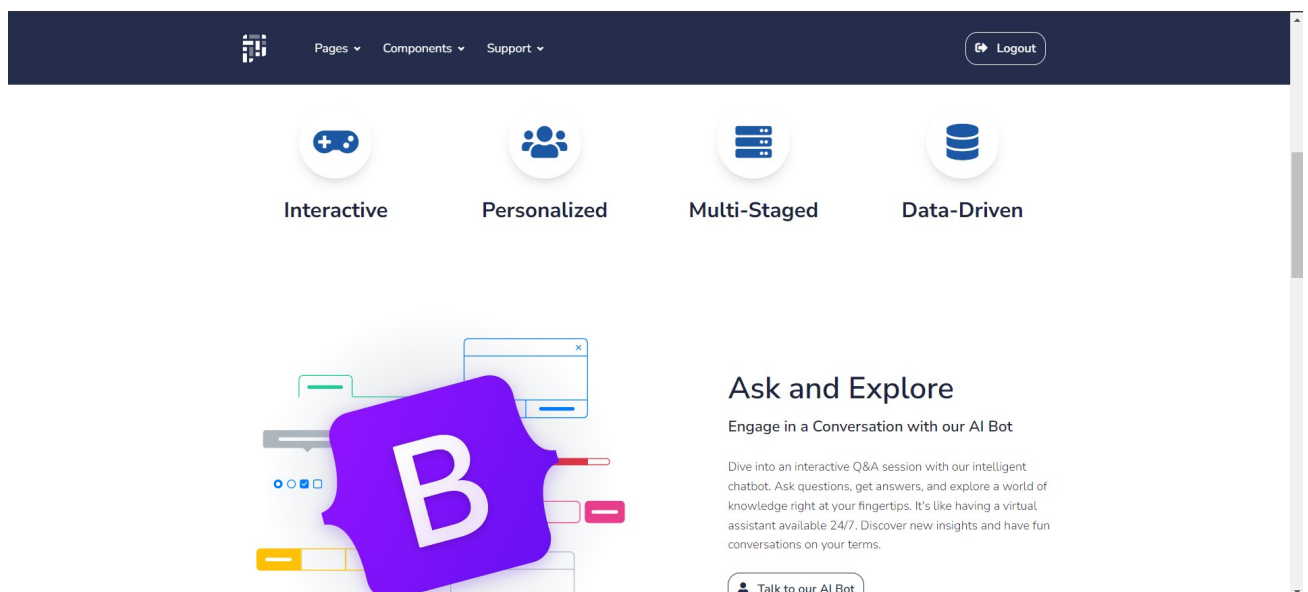


Figure 19: Website Interface



Figure 20: Questionnaire Bot

SUBMITTED ANSWERS		
Question No.	Question	Answer
1	What subjects do you think are most important for daily life	I believe basic mathematics and communication skills are crucial
2	Do you have a favorite subject	Yes I love art and drawing
3	Do you like participating in debates or public speaking	Yes I enjoy expressing my opinions in debates and public speaking
4	Tell me about your favorite school memory	My favorite memory is the school farewell party where we had a lot of fun
5	What activities do you enjoy during your free time	I like playing outdoor games with friends and watching cartoons
6	Did you participate in any school events	Yes I participated in the school talent show and performed a dance
7	What kind of pets do you have or wish to have	I have a pet fish
8	Tell me about any volunteering experiences	I volunteered at a local community center during weekends

Figure 21: Answer Submission

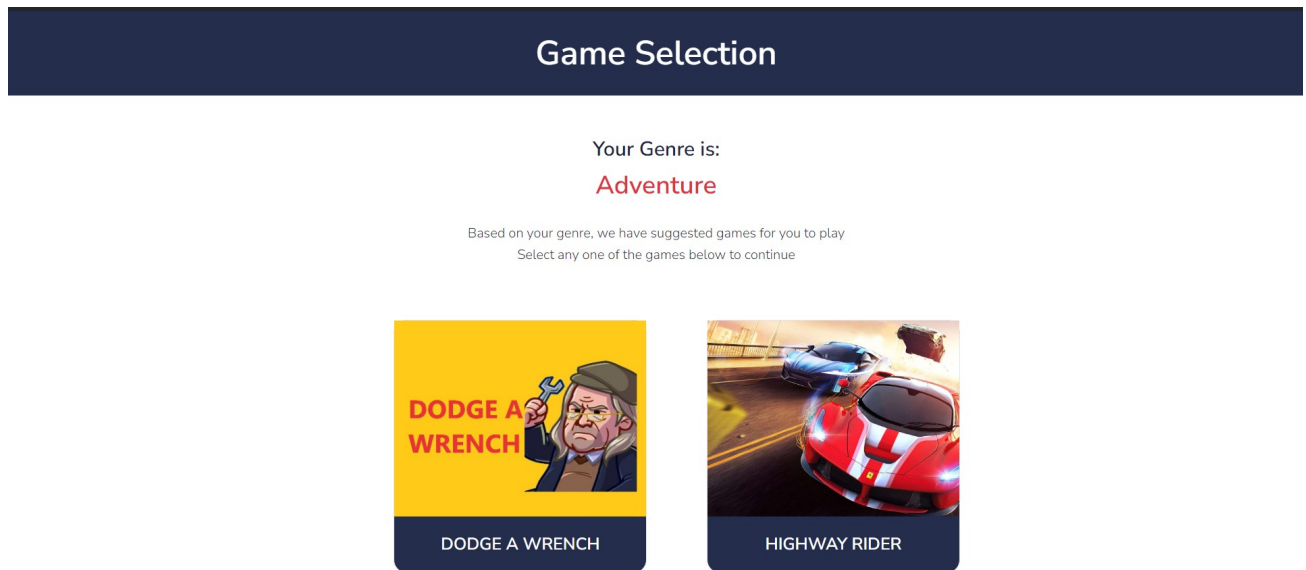


Figure 22: Game Selection



Figure 23: Game: Dodge A Wrench

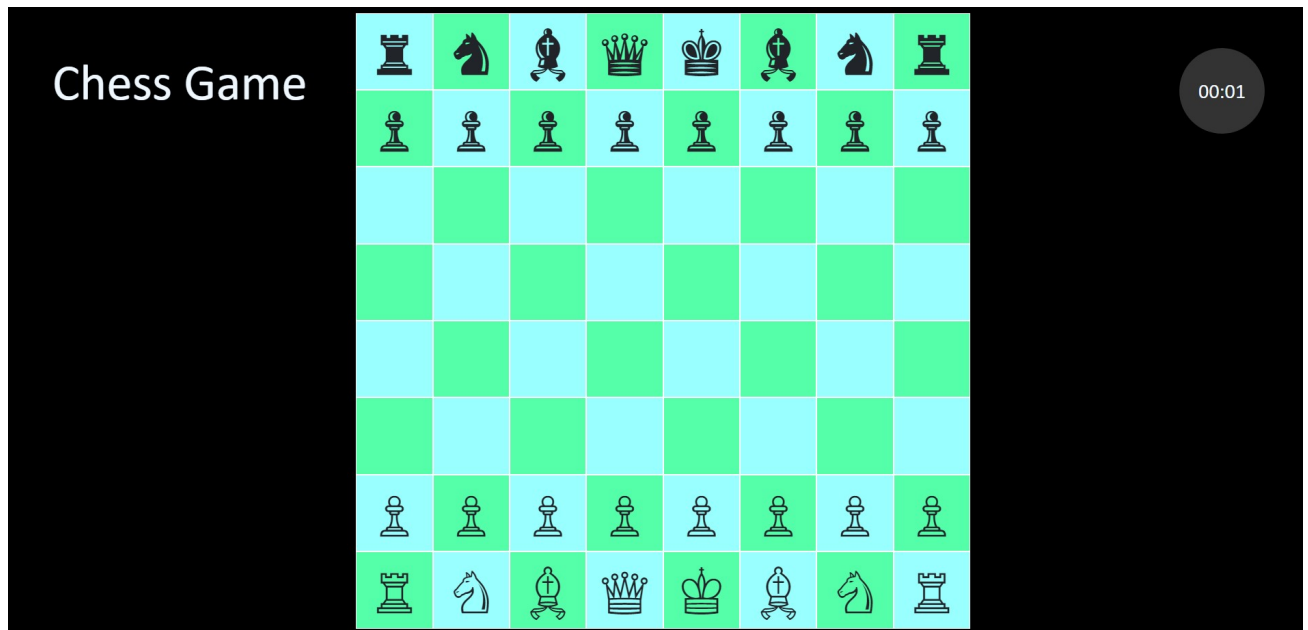


Figure 24: Game: Chess

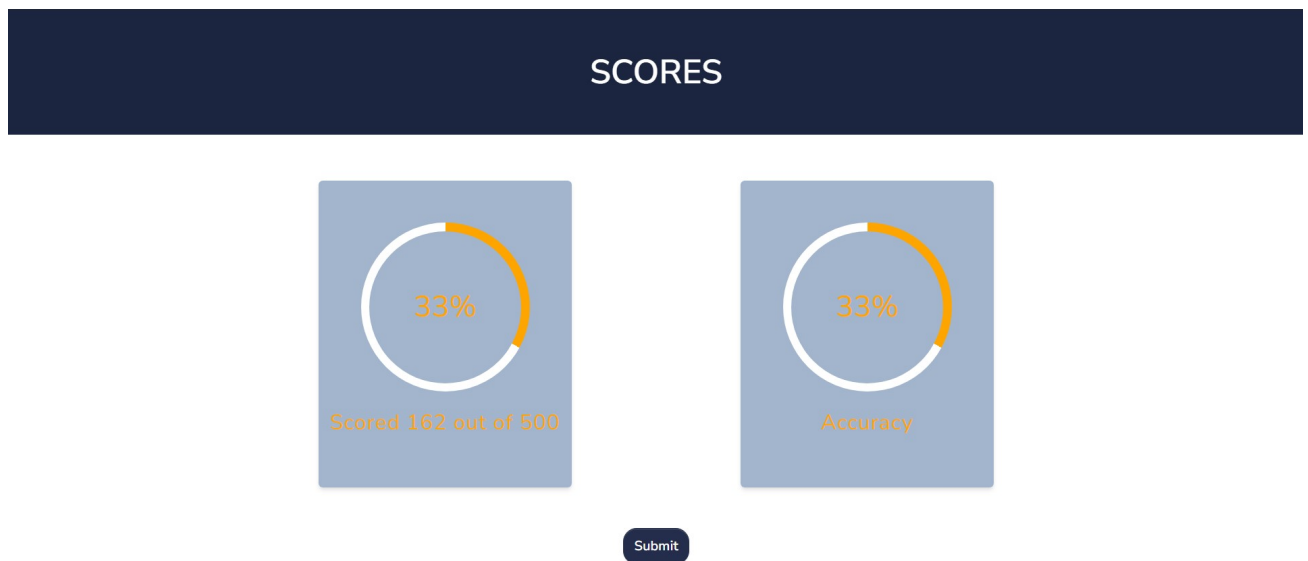


Figure 25: Game Scores

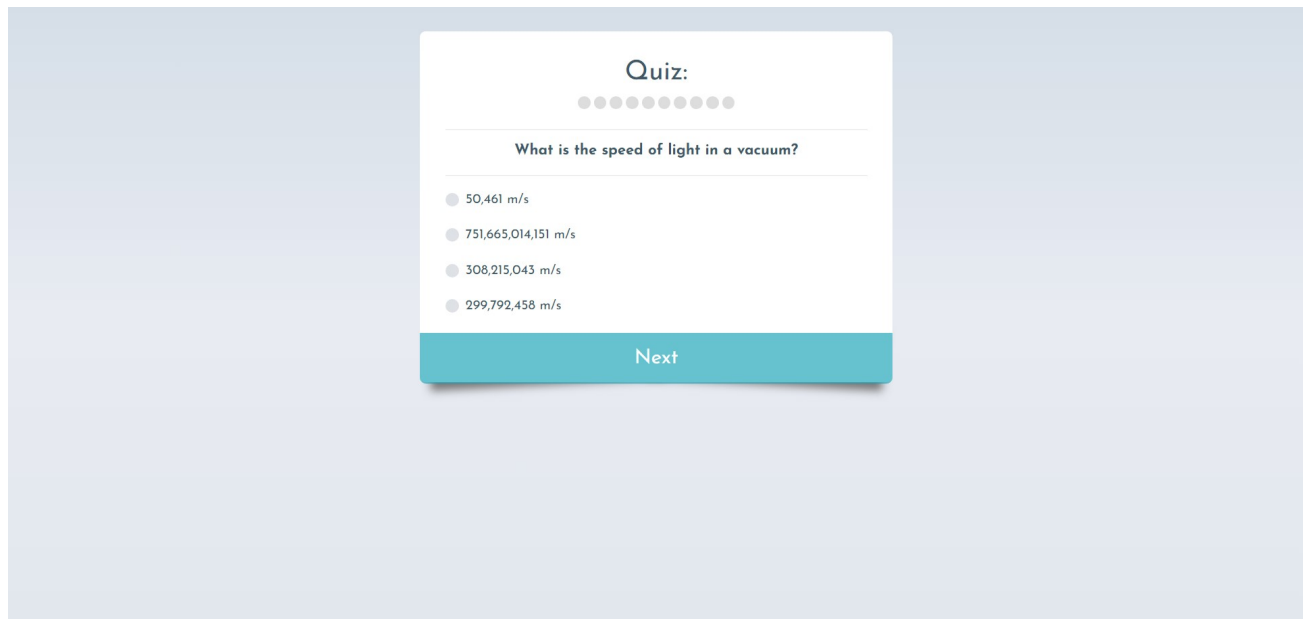


Figure 26: Aptitude Quiz (Trivia API)

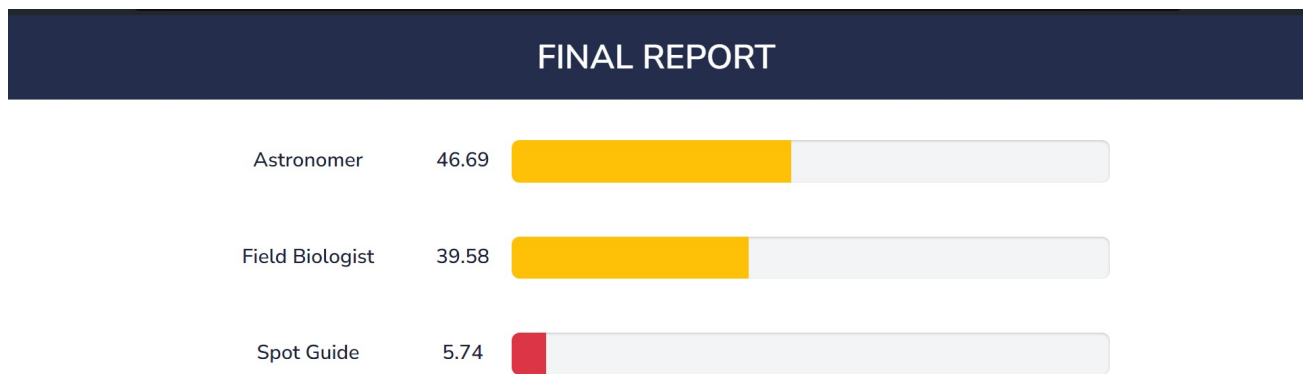


Figure 27: Final Report

## CHAPTER 9: CONCLUSIONS

### 9.1: CONCLUSIONS

In conclusion, CareerSage represents a significant advancement in the realm of career guidance and exploration. By seamlessly integrating deep learning, natural language processing, and recommendation systems, CareerSage offers users a comprehensive platform to navigate the complexities of career decision-making. Through interactive features such as genre prediction questionnaires, personalized game recommendations, and aptitude quizzes, CareerSage empowers individuals to explore diverse career options and make informed choices aligned with their unique interests and talents. Moreover, CareerSage goes beyond traditional career guidance by continuously adapting and refining its recommendations based on user feedback, ensuring that users receive relevant and tailored advice throughout their career exploration journey.

Furthermore, CareerSage not only facilitates individual career discovery but also contributes to societal advancement by promoting a workforce that is more engaged, fulfilled, and aligned with its professional pursuits. By helping individuals find careers that resonate with their passions and strengths, CareerSage fosters a culture of self-awareness, growth, and fulfillment in the workplace. As a result, CareerSage has the potential to drive positive outcomes at both the individual and organizational levels, ultimately contributing to a more vibrant and productive workforce in the ever-evolving landscape of the modern economy.

### 9.2: FUTURE SCOPE

- **Integration of Emerging Technologies:** Explore the integration of emerging technologies such as virtual reality (VR) and augmented reality (AR) to enhance the user experience, particularly in the gaming aspect. By incorporating VR elements into games recommended for the Sports genre, CareerSage can provide users with immersive and interactive experiences that simulate real-world sporting events and activities.
- **Expansion of Career Domains:** Extend the scope of CareerSage to include a broader range of career domains beyond the initial categories of Artistic, Adventure, Sports, and Logical. Conduct research and analysis to identify additional career paths and specialties, incorporating them into the career matching algorithm to provide users with more diverse and relevant recommendations.
- **Enhanced Personalization Features:** Implement advanced personalization features within CareerSage to further tailor recommendations and guidance to individual user preferences and goals. This could involve incorporating machine learning techniques to analyze user behavior, preferences, and feedback over time, allowing CareerSage to adapt and refine its recommendations dynamically.
- **Collaboration with Industry Partners:** Forge partnerships with educational institutions, industry organizations, and employers to enrich CareerSage's database of career-related

resources and opportunities. By leveraging real-world insights and expertise, CareerSage can provide users with up-to-date information on industry trends, job openings, internships, and professional development opportunities.

- **Mobile Application Development:** Develop a mobile application version of CareerSage to expand its accessibility and reach to a broader audience of users. A mobile app would allow users to access CareerSage's features conveniently on their smartphones or tablets, enabling seamless career exploration and guidance anytime, anywhere.

### 9.3: APPLICATIONS

- Career counseling centers and career coaches can utilize CareerSage to supplement their services and provide personalized guidance to their clients.
- Educational institutions can integrate CareerSage into their curriculum to help students explore career options and make informed academic and career-related decisions.
- Human resource departments in organizations can leverage CareerSage to assist employees in career development planning and talent management initiatives.
- Job search platforms and recruitment agencies can incorporate CareerSage to offer additional value to job seekers by providing career exploration tools and resources.
- Non-profit organizations and community centers can deploy CareerSage to support underserved populations in accessing career guidance and mentorship opportunities.

## APPENDIX A: FEASIBILITY ASSESSMENT

- **Problem Statement:** Given a set of users  $U$ , CareerSage aims to match each user  $ui$  with a suitable career domain  $Cj$  based on their interests, aptitudes, and preferences.
- **Logical Formula:** Let  $P$  represent the set of all possible pairings between users and career domains. Each pairing  $pk$  consists of a user  $ui$  and a career domain  $Cj$ . The problem is to find a satisfying assignment of  $P$  such that each user is matched with a suitable career domain, taking into account their individual profiles.

$$\text{Maximize } \sum_{k=1}^{|P|} f(p_k)$$

subject to the constraints:

User Constraint:  $\forall ui \in U, \exists Cj \in \text{CareerDomains: Match}(ui, Cj)$

Career Domain Constraint:  $\forall Cj \in \text{CareerDomains}, \exists ui \in U: \text{Match}(ui, Cj)$

Preference Constraint:  $\forall ui \in U, \forall Cj \in \text{CareerDomains}, \text{Preference}(ui, Cj)$

Here,  $\text{Match}(ui, Cj)$  represents whether user  $ui$  is matched with career domain  $Cj$ , and  $\text{Preference}(ui, Cj)$  denotes the preference of user  $ui$  for career domain  $Cj$ . The function  $f(pk)$  represents the suitability score of pairing  $pk$ , which can be computed based on various factors such as user preferences, aptitudes, and the relevance of the career domain.

- **Feasibility Assessment:** The problem's feasibility heavily depends on the efficiency of algorithms used to compute the maximum satisfying assignment of  $P$ . While the problem's formulation is expressible as a logical formula, finding an optimal solution may be computationally intensive, potentially falling into the class of NP-Hard or NP-Complete problems. Further analysis and algorithm design are necessary to determine the problem's exact complexity and develop efficient solving approaches.



## **APPENDIX B: PAPER PUBLICATION**

PAPER NOT PUBLISHED

MANUSCRIPT READY

## APPENDIX C: PLAGIARISM REPORT

Plagiarism Checked at Plagiarism Checker X software compatible for Windows 11.

**Verify Plagiarism at:** [CLICK HERE](#)

**URL:**

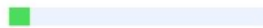
[https://plagiarismcheckerx.com/verify-report?date=2024-05-10&words\\_matched=1196&words\\_original=14190&score=8&wml=4&ver=8.0.7&key=818F4B9461692FDCE38EA8681A66833E](https://plagiarismcheckerx.com/verify-report?date=2024-05-10&words_matched=1196&words_original=14190&score=8&wml=4&ver=8.0.7&key=818F4B9461692FDCE38EA8681A66833E)



### Plagiarism Checker X - Report

Originality Assessment

**8%**



**Overall Similarity**

**Date:** May 10, 2024

**Matches:** 1196 / 14190 words

**Sources:** 52

**Remarks:** Low similarity detected, check with your supervisor if changes are required.

**Verify Report:**

[View Report Online](#)

Figure 28: Plagiarism Report

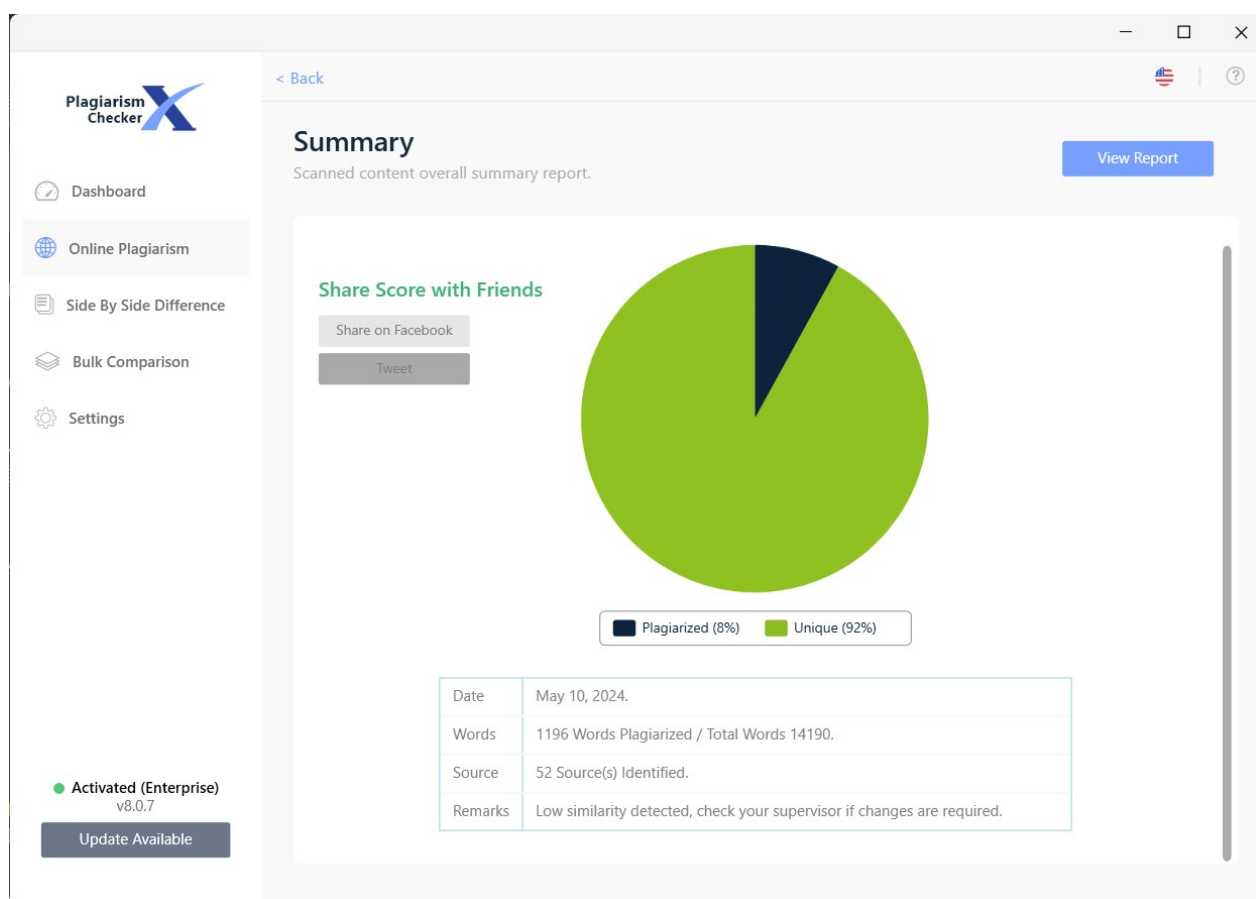


Figure 29: Plagiarism Percentage

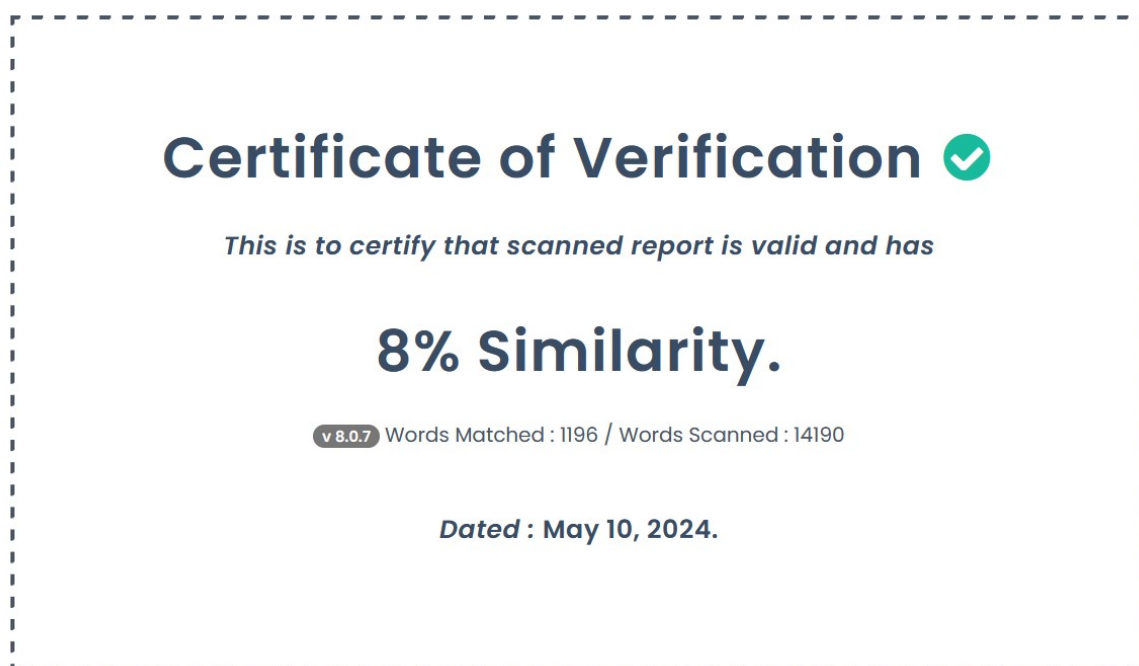


Figure 30: Plagiarism Verification

## REFERENCES

- Artificial Intelligence and Machine Learning by Vinod Chandra S. S and Anand Hareendran S
- Artificial Intelligence – A Modern Approach by StJohn D. Kelleher, Brian Mac Namee, Aoife D’Arcyuart Russell & Peter Norvig
- Make Your Own Neural Network By Tariq Rashid
- The Elements of Statistical Learning: Data Mining, Inference and Prediction by Trevor Hastie, Robert Tibshirani, Jerome Friedman
- Pattern Recognition and Machine Learning by Christopher Bishop
- Artificial Intelligence for Career Guidance – Current Requirements and Prospects for the Future Westman, Stina; Kauttonen, Janne; Klemetti, Aarne; Korhonen, Niilo; Manninen, Milja; Mononen, Asko; Niittymäki, Salla; Paananen, Henry IAFOR Journal of Education, v9 n4 p43-62 2021
- Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling
- Multi-expert system design for educational and career guidance: an approach based on a multi-agent system and ontology Essaid EL HAJI , Abdellah AZMANI , Mohamed EL HARZLI
- PCRS: Personalized Career-Path Recommender System for Engineering Students
- Career Guidance System Using Ensemble Learning Proceedings of the Advancement in Electronics & Communication Engineering 2022
- A Systematic Review on the Application of Ontologies to Improve Career Guidance

## REPORT DOCUMENTATION

<b>Report Documentation</b>				
<b>Report Code:</b> CS-BE-Project 2023-2024		<b>Report Number:</b>		
<b>Report Title:</b> CareerSAGE: A Career Guidance AI Model				
<b>Address (Details):</b> Pune Vidyarthi Griha's College Of Engineering and Technology, Pune				
<b>Author 1 [with Address, phone, E-mail]:</b> Gaurav Suresh Patil  <b>Address:</b> Pune Vidyarthi Griha's College Of Engineering and Technology, Pune  <b>E-mail:</b> <a href="mailto:21110019@pvgcoet.ac.in">21110019@pvgcoet.ac.in</a>  <b>Roll No.:</b> 1019  <b>Exam Seat No.:</b> B190074260  <b>Cell No.:</b> 8007124774	<b>Author 2 [with Address, phone, E-mail]:</b> Buddhabhushan Pramod Sapkal  <b>Address:</b> Pune Vidyarthi Griha's College Of Engineering and Technology, Pune  <b>E-mail:</b> <a href="mailto:20110081@pvgcoet.ac.in">20110081@pvgcoet.ac.in</a>  <b>Roll No.:</b> 0081  <b>Exam Seat No.:</b> B190074272  <b>Cell No.:</b> 9028043654	<b>Author 3 [with Address, phone, E-mail]:</b> Vivek Kisan Kute  <b>Address:</b> Pune Vidyarthi Griha's College Of Engineering and Technology, Pune  <b>E-mail:</b> <a href="mailto:20110056@pvgcoet.ac.in">20110056@pvgcoet.ac.in</a>  <b>Roll No.:</b> 0056  <b>Exam Seat No.:</b> B190074239  <b>Cell No.:</b> 8080825764		
<b>Year:</b> 2023-2024				
<b>Branch:</b> Computer Engineering				
<b>Keywords:</b>				
<b>Type of Report:</b> FINAL	<b>Report Checked By:</b>	<b>Report Checked Date:</b>	<b>Guide Complete Name:</b> Prof. A. D. Karanje	<b>Total Copies:</b> 4