UNIVERSITY INSTITUTE OF ENGINEERING AND TECHNOLOGY

MDU ROHTAK

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

B.tech.(CSE-AIML) 2025-26

AI Based Career Guidance System

A Project Report ***Submitted by***

Sanjay (2211038)

Parveen Kumar(2211056)

Nishchay (2211037)

Under the Guidance of

Dr. Dheerdhwaj

DECLARATION

Us, Sanjay(2211038), Parveen Kumar(2211056), Nishchay(2211037), Exam Roll Nos. 7113705, 7113687, 7113684, B.Tech (CSE-AIML), VII semester understand that plagiarism is defined as anyone or combination of the following:

1. Un-credited verbatim copying of individual sentences, paragraphs or illustration (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.

2. Un-credited improper paraphrasing of pages paragraphs (changing a few words phrases, or rearranging the original sentence order)

3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did write what. (Source:IEEE, The institute, Dec. 2004)

4. Ihavemadesurethatalltheideas,expressions,graphs,diagrams,etc.,thatarenotaresultof my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

5. I affirm that no portion of my work can be considered as plagiarism and I take full responsibility if such a complaint occurs. I understand fully well that the guide of the seminar/ project report may not be in a position to check for the possibility of such incidents of plagiarism in this body of work.

Signature of the Students:

Sanjay	Parveen Kumar	Nishchay

Name:

Sanjay	Parveen Kumar	Nishchay

Roll No.

7113705	7113687	7113684

Place: UIET,MDU Rohtak, India Date: 23rd Dec, 2025

CERTIFICATE

This is to certify that the project entitled "AI Based Career Guidance System" is the bonafide work carried out by Sanjay(2211038), Parveen Kumar(2211056), Nishchay(2211037), of B.Tech (CSE-AIML), UIET,MDU (Rohtak), during the VIIth semester of the academic year 2025-2026, in partial fulfillment of the requirements for the award of the Degree of Bachelors of Engineering as per the norms prescribed by UIET,MDU Rohtak. The project work has been assessed and found to be satisfactory.

Dr. Dheerdhwaj

Internal Mentor

Examiner

TABLE OF CONTENTS

List of Figures

List of Abbreviations

Abbreviation	Description
AI	Artificial Intelligence
ML	Machine Learning
SVM	Support Vector Machines
MLP	Multi Layer Perceptron
RBF	Radial Basis Function
SMOTE	Synthetic Minority Over-sampling Technique
CV	Curriculum Vitae
AMCAT	Aspiring Minds Computer Adaptive Test
CART	Classification and Regression Trees
API	Application Programming Interface
VAK	Visual, Auditory, and Kinesthetic
PHP	Hypertext Preprocessor
SQL	Structured Query Language
UML	Unified Modeling Language
SAT	Scholastic Assessment Test
MNSS	Monotonic Nonlinear State Space
MBTI	Myers–Briggs Type Indicator
RMSE	Root Mean Square Error

Abstract

With rapid growth of the internet, there has been a huge amount of information being made available to the user which often leads to an information overload and thus needs to be organized efficiently. Students often face a dilemma in deciding to choose a career in their life. There are several factors that influence the students when choosing their career path such as their personal aptitudes, educational achievements and their skill sets. The following report proposes an AI based career guidance and recommendation system where the authors leverage the use of AI to help candidates choose the right career early in their professional lives. The proposed system makes use of Machine Learning techniques on collected and preprocessed career information data gathered from various sources and builds a predictive model to solve the purpose. The dataset is based on the proficiency in computer science and related subjects along with psychological traits of the individual which are then mapped to the appropriate career roles that the individual possessed.

Chapter 1
Introduction

1.1 Introduction

Career selection is one of the most essential decisions a student must make in his or her lifetime. However, the decision is frequently highly impacted by different variables surrounding the student, such as job uncertainties, friends, parents, etc. In such a situation, students choose a profession that may or may not be to their taste. In such a situation, obtaining a degree is a question of survival for them. They have nowhere to hide and hence struggle semester after semester. Therefore, this approach is the ideal aid for such individuals, as it enables them to select courses and areas of interest that would make their pursuit of survival somewhat more joyful. There are other students that choose to major in engineering but do not know what to do after they begin their education. This occurs because people have several interests or, in some situations, are unsure of their interests and hence struggle to make the best decisions. Again, the suggested approach would assist these students in making informed decisions based on their individual interests.

1.2 Problem Definition

Our approach is to create an AI based web application which enables users to avail career counseling facilities and career guidance at the early stage of their career and professional life. We propose creation of a novel dataset for the same and apply traditional and state of the art techniques to suggest career options.

1.3 Project Overview

Students often face a dilemma in deciding to choose a career in their life. There are several factors that influence the students when choosing their career path such as their personal aptitudes, educational achievements and their environment. Hence, by means of this project we will aim to provide students with career counseling and career exploration platforms. ML based models are used to create our prediction engine for career prediction using various personality traits.

1.4 Impact

With rapid growth of the internet, there has been a huge amount of information being made available to the user which often leads to an information overload and thus needs to be organized efficiently. Students often face a dilemma in deciding to choose a career in their life. There are several factors that influence the students when choosing their career path such as their personal aptitudes, educational achievements and their environment. Students have always faced challenges when it comes to getting effective but free guidance with respect to career choices and advice, especially in developing or underdeveloped countries. This large disparity leads to an increasing number of students, who lack the resources to network with other professionals in their dream career roles.

Chapter 2
Literature Review

Previous research [1] on Career Recommendation Systems (CRS) has revealed hybrid approach to be the most frequently implemented approach, followed by approach of collaborative-filtering, approach of content-based filtering and knowledge base approach. The most common AI techniques used in CRS were found to be text mining, clustering and fuzzy based techniques. Course profile, influenced factors, personality test and academic performance were among the factors that were found to be the most relevant to the career recommendation process. Studies of the application of AI in career guidance [2] have identified four major ways that may be utilized as a coach, a collaborator, an assistant, and a tool, each of which involves a specific purpose and level of involvement in the career guidance process. Most concerns regarding the use of AI for career guidance were linked to the lack of sufficient pertinent data and the potential for bias in the data used.

The focus of our study is to analyze and understand the various approaches to career recommendation and guidance systems. The collective approaches were categorized into five broad classes: (i) data mining and machine learning, (ii) recommendation systems, (iii) expert systems and fuzzy logic, (iv) reinforcement learning, and (v) NLP and Chatbot Approach.

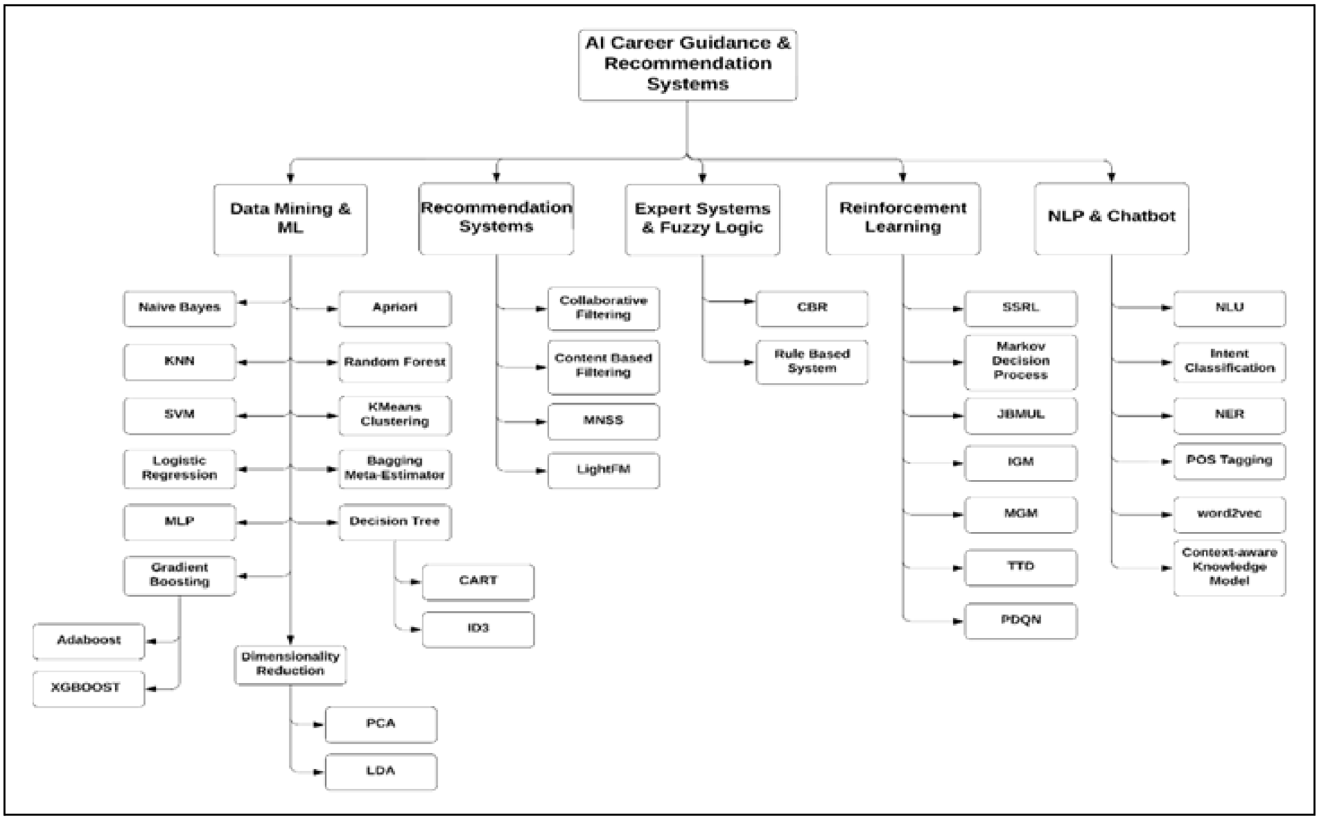

Fig 1. Concept Map Depicting Various Approaches To Career Guidance And Recommendation Systems

2.1 Data Mining and ML Based Approaches

Adapa, S. et. al [3] employed supervised machine learning techniques to recommend learning styles using VAK (Visual, Auditory, Kinaesthetic) technique. Participants were given 30 Self-Assessment questions, each with multiple answer choices, to determine their sensory acuity in terms of VAK. There were a total of 900 responses and they were converted to numeric values. Various machine learning algorithms were applied which include linear and logistic regression, Naive Bayes, LDA, KNN, SVM, CART, stochastic gradient boosting and Random Forest Classifier (RFC). The highest accuracy is given by Naive Bayes(63.37%) followed by SVM(61.34%).

The disadvantages of collaborative filtering in recommendation systems have been discussed by Shi, Z. et al. [4] and an approach has been adopted to overcome the same. Apriori and ACO (Ant Colony Optimization) algorithms are studied and a framework named Position-Apriori-ACO is brought forward to plan the learning path. Recruitment and job related information is gathered using web crawlers, university curriculum in talent training is extracted. Apriori is used to mine the most frequently occurring recruitment requirements and results are encoded into two-dimensional coordinates. The ACO algorithm gives an optimal plan of action. Basic courses are labelled as low level. Manual review is conducted on a sample test case and it is determined that the result produced by the system is accurate and efficient.

Personality traits by A. Ghimire et al. [5] have been used to determine educational preferences. The data set was acquired via a questionnaire containing ten personality-based questions. TIPI (Ten-Item Personality Inventory) were mapped to the big five personality traits. Academic majors were classified into 14 broad categories and the survey was filled by two groups of students. Decision trees were implemented on different feature vectors: ten question survey, big five personality traits and five-dimensional index using PCA. The accuracy of the model is 96% for the TIPI dataset and 95% for the OCEAN dataset.

P. D. Schalk et al. [6] use well established instruments like SAT along with other standard tests designed by them to predict mathematical aptitude using a machine learning model. From an earlier analysis, they found a strong correlation between performance in physics and mathematics, hence data from both the departments is used. Multiple techniques, including neural networks, decision trees, random forests, and support vector machines, were extensively evaluated. Neural networks, decision trees couldn't produce accurate results on new datasets, hence random forest was used for classification.

The research by C. D. Casuat and E.D.Festijo [7] aims at identifying inputs that are most important to predict employability signals of students. The dataset comprised 9 attributes ranging from General appearance, body language to interview evaluation results, students' performance rating, general percentage average (GPA). For the prediction model; SVM, DT, RF are used out of which SVM gives highest (91.22%) accuracy. For feature selection, univariate selection (US), principal component analysis (PCA) and recursive feature elimination (RFE) are employed and compared. PCA with SVM with SMOTE had the greatest accuracy of any combination, 93%. According to the results of the PCA, mental acuity, speaking style, the ability to express ideas, and self-confidence were the most influential elements affecting employment.

L. Xiaomin et al. [8] focus on designing an AI based system performing the function of providing personalized solutions and long-term companionship in finding a career direction they are interested in along with improving their vocational skills. The proposed system makes use of a content-based recommendation algorithm and incremental learning. Since the proposed framework is separated into two broad models namely user and resource, the cosine similarity algorithm is used to compare between the two and the Naive Bayes algorithm calculates the probability of suitability of the personalized recommendation.

R. H. Rangnekar et.al [9] propose a career counseling system that integrates career guidance with artificial intelligence. It comprises a chatbot along with a career guidance system that utilizes decision trees and Support Vector Machines (SVM) for the classification of users based on their responses to a questionnaire. The questionnaire was composed of questions regarding the users' skills as well as their interests and were curated separately for different domains.

Md. Y. Arafath et.al [10] uses personality traits alongside aptitude and background information as input for career prediction. Top 3 suitable career choices are recommended based on the responses to a quiz to determine the aptitude and background information entered by the user. The personality of users is gauged by accessing their Facebook profile using the Facebook Graph API, which is then used to determine a measure of how easy it would be for the user to pursue the recommended careers. The Myers-Briggs Type Indicator is used as the personality metric.

B. Harsha et.al [11] implemented various prediction models to the sample data containing information about academic, technical, and interpersonal characteristics of students in order to obtain the result and utilized five classification algorithms: ID3, CART (Classification and Regression Tree), Multi-Layer Perceptron (MLP), Random Forest and Neural Networks. Highest accuracy of 95.24% was given by CART.

K. Joshi et.al [12] propose a system which is evaluated using a dataset which contains 12 attributes and 700 plus records that contains details like marks in SSC, percentage in physics, math and chemistry along with their intermediate marks. Three classifiers, namely Support Vector Machines (SVM), Random Forest and Decision Trees are used and their accuracies are compared to show that SVM gave the highest accuracy of 90.3%, followed by Random Forest at 88.3% and Decision tree with 86.53% .

H. Al-Dossari et.al [13] proposes a career recommendation system that is trained on a dataset created from a survey of 1707 IT employees in Saudi Arabia.. Respondents were asked to rate a set of 20 skills consisting of both soft and technical skills, as none, low, moderate, or high. The survey also included questions regarding their respondents' job title, gender, specialization and programming languages they are familiar with. The job titles were grouped into 3 categories. The recommendation model was trained to recommend either one of the three categories: analyst, developer and engineer. The model was trained using five machine learning algorithms: K-Nearest Neighbors (KNN), Decision-Tree (DT), Gradient Boosting, Bagging meta-estimator and XGBoost. XGBoost outperformed the other algorithms with an accuracy of 70.47%.

The system proposed by M. S. Khan and M. Darbari [14] leverages Myers-Briggs Type Indicator (MBTI) and social media mining. MBTI classifies the personality of the user in 4 categories: Extraversion or Introversion, Sensing or Intuition, Feeling or Thinking and Perceiving or Judging. The combination of the traits indicated the inclination towards goals. The user's comments and posts on Facebook and Twitter accounts are extracted as input. 3 inputs are used as input to a decision matrix: Social Media mining based individual interest, individual personality traits predicted via Social Media and the MBTI questionnaires-based personality traits. Random forest classifier is used to recommend career options.

S. Vignesh et al. [15] proposed a career recommendation system comprising three modules. These modules consist of questionnaires, machine learning algorithms, and statistical analysis to build a career recommendation system. The authors predicted careers using K-nearest neighbors and K-means clustering and compared the confusion matrices of two other methods, SVM and Naive Bayes. Later, F-measure values were compared for various departments.

Thomas et al. [16] predict students' careers in technology, commerce, and the humanities department using 8 different methods of approach. While creating the dataset, intellect (IQ) and emotional intelligence (EQ) factors are also considered. The paper compares the following algorithms: Naive Bayes, Decision Tree, Logistic Regression, AdaBoost, Multilayer Perceptron, and SVM with the RBF kernel. The algorithm was selected based on its accuracy and the amount of time it took to get the result. The author further compares the accuracy of all the algorithms applied to different datasets in science, commerce, and the humanities. The author suggests feeding more real-time data in order to increase accuracy in the future.

J. Britto et al. [17] base their research on graduate-level study. The algorithm for recommending courses considers the GPA of students and the subjects they scored well in. Feedback of the students was taken in order to assess the system. The paper uses a neural network for recommending courses, which uses the concept of forward and backward propagation with gradient descent.

A. Kamal et al. [18] presented 3 algorithms which are XGboost, Random Forest and Random Forest in OneVsRest. The algorithm suggests top 5 recommendations most suitable for the student. The paper uses a marketing scheme for multiple intelligence which helps to find relevant data and the validity of the dataset is then evaluated using Holland's score. SMOTE is used to handle an imbalanced dataset and k-fold cross validation is later used. The results of the algorithms were compared with field-wise comparison.

K. Joshi et al. [19] aim to provide an overview of the artificial intelligence techniques that were used to predict the student's performance. The model that was implemented is put to the test, after which a report is generated, and additional recommendations are made. The paper shows the implementation of algorithms like decision trees and SVM on the dataset collected from different sources.

T. Tayade et al. [20] study various job recommendation systems in practice which include prediction of personalities through CVs and stream analysis using an online aptitude test. The method of stream analysis helps fresh undergraduates choose a stream they are most capable of. Online occupation recommender systems are examined from 4 angles namely- client profiling, inputs, suggestion systems and yield. The proposed approach involves a job applicant and the recruiter. The applicant gives an aptitude and personality test along with submission of CV. The recruiter module generates results based on user skill sets and company requirements. The authors only introduce a work proposal framework.

R. Ajoodha et al. [21] attempt to provide a data-driven solution to the data-saturated environment of attributes (information overload) relevant to student performance and contribute to the prevention of the rising dropout rates in South African higher education institutions. The study identifies the discrepancy between the necessary skills required for success in a science programme (determined using data-driven methodologies) and the current learner's skill profile (derived from the learners' assessment results). On the basis of the results of the forecast, the improvement in skills required for success in that programme is determined. The dataset collected was trained on six algorithms namely:

Decision Trees, K*, Naive Bayes, Support Vector Machines, Multilayer Perceptron, and Linear Logistic Regression. The Multi-layer Perceptron classification model outperformed the other five models with 62% accuracy.

2.2 Recommendation System Based Approach

A system was proposed by A. H. A. Rashid et al. [22] that recommends careers for students in the field of Computer Science. Data is extracted by scraping a career information website. The recommendation system is based on a content-based filtering method which takes a career ID (based on the user's preferred career choice) as input and produces results that are available career options based on the user's interests. The system's usability was tested using a system usability scale (SUS) which gave an average score of 81.25. A score greater than 68 indicates high usability.

T. V. Yadalam et al. [23] aim to create a job-recommendation system for graduate students based on their interests and skill sets. For prediction, the algorithm considers a student's grade point average in various subjects, ratings on communication skills, number of hours worked per day, rating on the logical quotient, hackathons won, coding skills, public speaking ability, and self-learning capability, among other variables. In addition to including an area for feedback and comments, the author employed NLP techniques to determine the nature of each comment. Cosine similarity function is used to determine the similarity between previous user selections and available jobs and then recommend the best positions based on the score.

A. Gugnani et al. [24] suggested an unique framework that uses text extraction techniques to generate personalized skill graph representations of candidate profiles.The proposed system utilizes the concept of skills to generate skill graphs that can serve as the basis for career path suggestions. The authors concluded that such skill graphs, which record both spatial and temporal correlations, facilitate the generation of accurate career path suggestions. The model takes the candidate's profile to create a skill graph using the skill data retrieved from the candidate's profile. The model also takes feedback and continuously improves using new data, using NLP concepts to understand the feedback.
I. Dutta et al. [25] aim to create a robust recommendation system that can connect students seeking guidance with the most qualified professionals for their questions. The authors make use of the CareerVillage Competition Dataset on Kaggle and implement a data pipeline consisting of processes which include gathering of data, preprocessing, implementing algorithms and its evaluation followed

by testing on sample data. The algorithms implemented include neural network and LightFM Hybrid Recommendation model. The proposed LightFM algorithm achieved an AUC score of 91%.

H. Zhuang and Z. Zheng [26] described a personalized recommendation system for entrepreneurship. The basic information and personal interests of a college student are represented by feature vectors, which provide beneficial theoretical support for career planning, employment, and entrepreneurship among college students. A Deep Learning based information recommendation model is created with an accuracy of 98.99% in contrast to 76.84% of a traditional model. Employment Recommendation and Career Planning have been discussed and experimental analysis included conduction of simulations, function testing and performance testing. The applicability is deduced by scoring per capita satisfaction of college students who have used the information recommendation system.

A. Ghosh et al. [27] propose an interpretable and novel monotonic, non-linear state space model to analyze user profiles and generate feedback and recommendations. The authors perform a series of experiments on LinkedIn and Indeed job resumes data. The proposed MNSS I,e Monotonic Nonlinear State Space model outperforms traditional models in job title, company and skills recommendation. Since the model is interpretable, extended use cases like skill gap identification and career path planning can also be performed.

2.3 Expert System & Fuzzy Logic Based Approach

Fathian Brojeny M. [28] presents a model for designing a career consultation system using case-based reasoning. A CBR (Case Based Reasoning) system deals with retrieving new or similar cases and applying that knowledge to propose solutions and solve problems. This paper deals with finding the most suitable job for an applicant based on his characteristics by determining whether the applicant is ideal or a negative ideal. The similarity metric is calculated as closeness to ideal applicant and remoteness to negative ideal candidate. The model was applied to an Iranian dataset composed of 200 jobs and the validity for the system was found to be 80%.

R. Singh et al. [29] aim at building a testing system for undergraduates which will evaluate the best suited career for users based on their inputs by conducting an aptitude test following which the student will receive his result along with a detailed explanation and advice. The explanation will state the reason as to why we feel a certain career is better for him/her in comparison with other options. A rule-based approach is employed for the data gathered to build the system.

M. Qamhieh et al. [30] design a personalized career recommendation system which will act as an equivalent to professional help. The best engineering discipline for students is predicted based on their academic performance, personality traits and their participation in extracurricular activities. Fuzzy logic is employed where for each student, corresponding processed data of a specific branch of engineering is entered into the fuzzy system and a personalized rate is the output. Cohen's kappa value was used, and it reveals that there is slight agreement between recommender output and student's specializations.

W. M. Aly et al. [31] propose a system wherein an expert system will aid students in deciding registration in a specific course. This is designed as a fuzzy expert system and implemented as a mobile application that runs under the Android operating system. The system accepts six inputs (Perceived Teaching Efficiency of lecturer, Past Performance, Perceived Difficulty of Course, Appeal of Course Topic, Friends in Course, cost of Course, Recommendation of Registering the course) and produces a single output. The RMSE of the proposed Fuzzy recommender system was calculated to be 6.64% and this was based on a sample of 40 students.

R. E. Wulansari et al. [32] address the problem of designing a tool for students in higher education to recognize their potential and abilities. An expert system is developed by the authors by following the Software Development Lifecycle that consists of: problem identification, feasibility study, project planning, knowledge acquisition, knowledge representation, knowledge implementation, verification and validation, installation, transition and training, operation, evaluation and maintenance. The authors use Multiple Intelligence concept which mentions that people not only possess intellectual capacity but also other kinds of intelligences like- musical, social, linguistic, etc.

2.4 Reinforcement Learning Based Approach

Guo et al. [33] propose an intelligent sequential career planning system with a career path rating mechanism and the stochastic subsampling reinforcement learning (SSRL) framework for reinforcement learning (RL). They implemented five baseline models, including JBMUL, IGM, MGM, TTD, and PDQN, with a summary of their benefits and drawbacks. The author believes that this technique can determine the optimal path for various career statuses. A gaussian distribution-based "good career pathway" has a path score higher than 66.62, which is the top 2.2% of real-world (human-decision) data.

M. Kokkodis and P. G. Ipeirotis [34] adopt the ideology of value of a skill being dependent on the market conditions that are dynamic in nature. A constant need for upskilling and choosing those skills is a difficult task to remain relevant. The authors propose a system that integrates reinforcement learning, Bayesian inference, and gradient boosting for recommending new skills to acquire. The proposed system does not learn from past behavior to create future recommendations; rather, it uses a Markov decision process that acts on a graph of viable actions to dynamically recommend profitable career paths.

2.5 NLP & Chatbot Based Approach

A. Nair et al. [35] implement the approach of using chatbots for AI career counselling. The authors mention that the chatbots are able to receive and remember user inputs which helps them grow as more and more responses are received. The Rasa framework has been used to implement the chatbot which is built on data collected by scraping TheStudentSuccessapp.com website created by nSmiles. A psychometric engine powers the chatbot which generates an intermittent and final report containing the list of courses and college recommendations based on the user's interests.

A. Khan et al. [36] propose and educational agent system for recommending resources, professional personalized suggestions and guidance for interviews by building a virtual assistant which has the provision of being multilingual with the addition of languages like English, Hindi, Marathi and Gujarati. The authors make use of APIs by Google, fine tuning data scraped from sites like Reddit and Quora. The end user is benefitted through FAQ solving and recommendation of technology stacks based on skill sets.

D. Nguyen et al. [37] build a personalized career counselling chatbot named "ITCareerBot" in order to address the challenge of quick changing environments in the IT industry. A chatbot framework is built based on CAK (context aware knowledge model) and recommendation methodology by collecting data from professional social networking platforms and online education platforms through which the chatbot matches current employee skills and their interests.

Chapter 3
Analysis and Design

3.1 Proposed System

The system will have a user-interface to allow for smooth user interaction with the recommendation model. The different components of the UI will include a questionnaire, the result page with the result of the recommendation model and detailed career profiles/description pages that will be connected with the backend.

Career Recommendation

This module will consist of 2 elements. First is the carefully curated questionnaire that has to be filled in by the users. The questions in the questionnaire will range from rating for proficiency in various computer science subjects and questions with respect to personality traits. These questions will be created to maintain a specific ratio of the aptitude/skill-related questions and background information according to the requirements of the machine learning model used for career recommendation.

3.2 Hardware Specifications

The hardware specifications for the project have been considered from the perspective of the Developers and the end-users. The Hardware Specifications have been divided into two aspects as follows:

For Developers: The Hardware specifications for developers include as follows:

- Processor: 2.4 gigahertz (GHz) or faster processor
- RAM: 2GB and above
- OS: Windows 7/8/10 and later versions
- Display:800 x 600
- Python 3.7 and later versions
- Graphics: DirectX 9.0 and later versions

For End-Users: The Hardware Specifications for end-users: Since the final product is aimed to be deployed on the web, the end user needs no additional hardware apart from the medium of accessing the application with a stable internet connection.

The aforementioned hardware includes Mobile Phones, Laptops, Desktops and Tablets and the long term supported browsers include:

- Microsoft Edge
- Safari Browser
- Google Chrome

3.3 Software Specifications

The software specifications include the proposed technology stack used to build the end to end application of "AI Based Career Guidance System".

The software specifications are divided into the following categories:

1. Frontend: The frontend of the application consists of the technologies used to build the UI of the application. The proposed stack for Frontend includes: HTML, CSS and Streamlit Library
2. Backend: The backend of the application consists of the technologies used for building ML models and the interaction with the database and authentication. The proposed stack for Backend includes: Python for Machine Learning Applications and PHP for interaction with database and authentication.
3. Database: The proposed stack for Database includes PHPMyAdmin and MySQL
4. Project Delivery and Management: Project Delivery and Management includes the platforms and tools used for complete end to end software development and includes GitHub for Version Control.

The following figure summarizes the complete Technology Stack of our project:

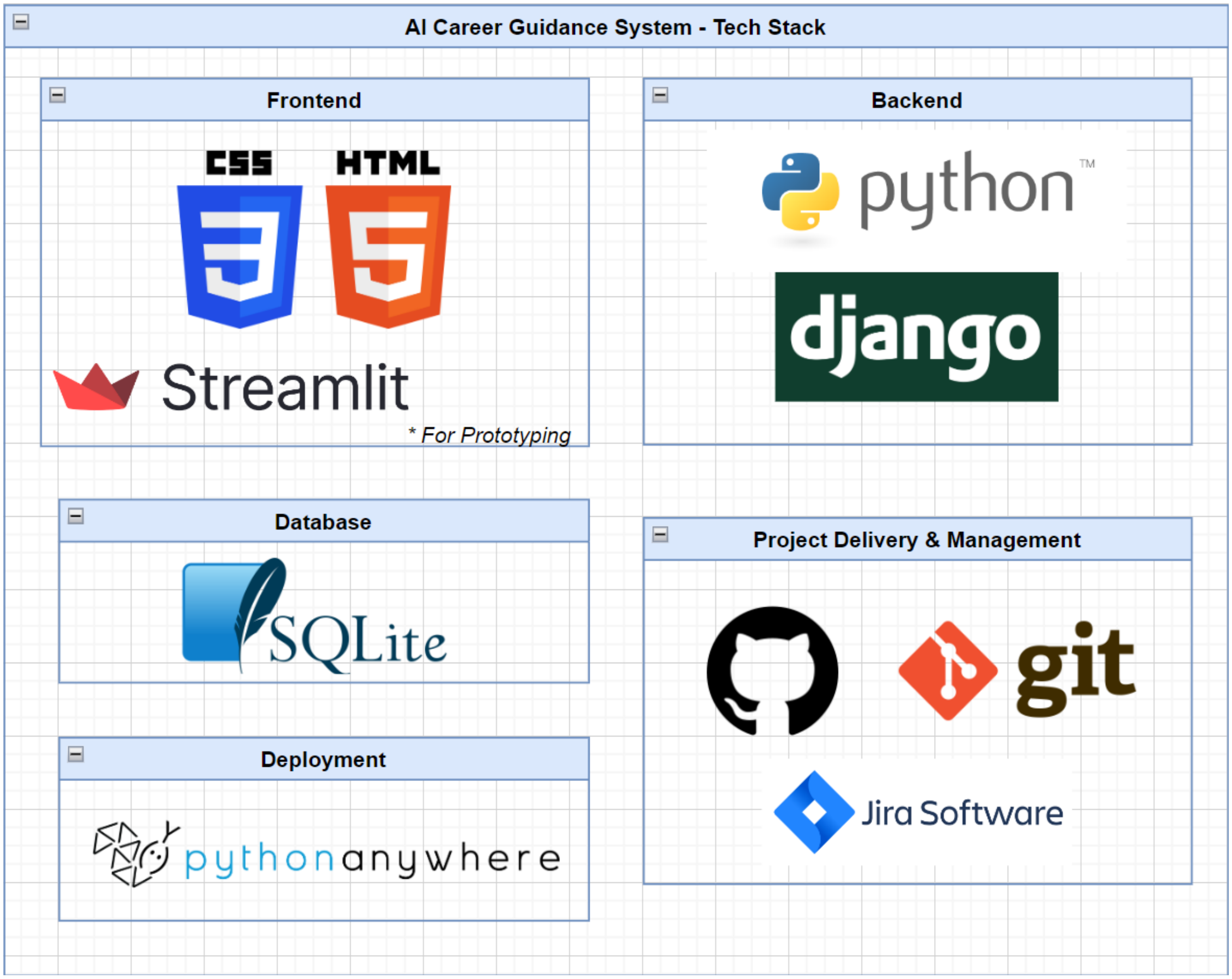

Fig 2. Technology Stack for the Project

3.4 Algorithms

3.4.1 K-Nearest Neighbor(KNN)

K-Nearest Neighbor is one of the most straightforward Machine Learning algorithms based on Supervised Learning.The K-NN method considers the similarity between the new case/data and the existing cases and places the new case in the category that is most similar to the existing categories. The K-NN method may be used for both Regression and Classification, however it is often applied to Classification issues. K-NN is a non-parametric method, hence it makes no assumptions about the underlying data.It is also known as a lazy learner algorithm since it does not instantly learn from the training set. Instead, it saves the dataset and takes an action on it at the time of classification.

The K-NN method may be used for both Regression and Classification, however it is often applied to Classification issues.K-NN is a non-parametric method, hence it makes no assumptions about the underlying data.It is also known as a lazy learner algorithm since it does not instantly learn from the training set. Instead, it saves the dataset and takes an action on it at the time of classification.

3.4.2 Decision Trees

Decision Trees are a non-parametric supervised learning method used for both classification and regression. They are primarily used for classification and have a flowchart-like tree structure. Within a tree each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

It aims at creating a model that learns simple decision rules inferred from the data features and predicts the value of a target variable. A tree can be seen as a piecewise constant approximation. In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree. This algorithm compares the values of the root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node. For the next node, the algorithm again compares the attribute value with the other sub-nodes and moves further. It continues the process until it reaches the leaf node of the tree.

3.4.3 Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be applied to both classification and regression problems in Machine Learning.

Random Forest is a classifier that comprises decision trees on different subsets of the input dataset and averages the results to increase the dataset's predicted accuracy. Instead of depending on a single decision tree, the random forest uses forecasts from each tree and predicts the result based on the votes of the majority of predictions.

Instead of depending on a single decision tree, the random forest uses forecasts from each tree and predicts the result based on the votes of the majority of predictions.

The greater number of trees in the forest prevents the problem of overfitting and leads to higher accuracy. Random Forest works in two-phase first is to create the random forest by combining N decision trees, and second is to make predictions for each tree created in the first phase.

3.4.4 Support Vector Machines

Support Vector Machines (SVMs) are a popular supervised machine learning algorithm used for classification and regression analysis. SVMs classify data by finding the optimal hyperplane that separates the data into different classes. The hyperplane is chosen such that the margin between the hyperplane and the nearest data points of each class is maximized. SVMs are effective in high-dimensional spaces and can handle non-linearly separable data through the use of kernel functions.

SVMs have several advantages over other classification algorithms, such as high accuracy, robustness to noise and outliers, and the ability to handle large datasets. However, they can be computationally expensive and require careful selection of hyperparameters. SVMs have many real-world applications, including image classification, text classification, and bioinformatics.

SVMs are a powerful machine learning algorithm that can be used for both classification and regression tasks. They are particularly useful for datasets with high dimensionality and non-linear relationships between variables. Despite some computational challenges, SVMs are widely used in many fields for their accuracy and robustness.

3.4.5 KNN classifier

KNN (K-Nearest Neighbor) classifier is a simple and effective classification algorithm in machine learning. It belongs to the supervised learning technique, which is used for both classification and regression problems. The KNN classifier works on the principle of finding the K nearest data points in the training set to a given input data point and classifying it based on the majority class of those K neighbors.

K is a hyperparameter that is chosen by the user and determines the number of neighbors to consider. The classifier can use various distance metrics to measure the similarity between the data points, such as Euclidean distance, Manhattan distance, or Cosine distance.

The KNN classifier has several advantages, such as its simplicity, ability to handle multi-class problems, and its non-parametric nature, which means it does not make any assumptions about the underlying data distribution. However, it can be sensitive to noisy data and may require a larger dataset for better accuracy.

KNN classifier is a useful algorithm in machine learning and has been applied in various fields such as image recognition, natural language processing, and recommendation systems.

3.4.6 Multi Layer Perceptron

A Multi-Layer Perceptron (MLP) is a type of artificial neural network that is widely used in machine learning for supervised learning tasks such as classification, regression, and prediction. It consists of multiple layers of interconnected nodes (also known as neurons) that process input data and generate output. The input is fed into the first layer, and each neuron in that layer performs a mathematical computation on the input and sends the result to the next layer. This process continues until the output layer is reached, which produces the final result.

The MLP is trained using backpropagation, a supervised learning algorithm that adjusts the weights of the connections between neurons to minimize the difference between the predicted output and the actual output. The MLP has several advantages over other machine learning algorithms, such as the ability to learn complex nonlinear relationships between input and output and the ability to generalize unseen data. However, it also has some limitations, such as the need for a large amount of training data and the risk of overfitting. MLPs are a powerful tool in the field of machine learning and have applications in various industries such as finance, healthcare, and image processing.

3.4.7 Logistic Regression

Logistic regression is a statistical algorithm used for binary classification problems, where the outcome is either yes or no, true or false, or 0 or 1. It is a supervised learning technique that models the relationship between a dependent variable and one or more independent variables. The dependent variable is categorical, and the independent variables can be categorical or continuous.

The Logistic Regression model calculates the probability of the dependent variable being a particular class. It uses the logistic function, also known as the sigmoid function, to transform a continuous input into a value between 0 and 1, which represents the probability of the positive class.The model estimates the coefficients of the independent variables to maximize the likelihood of the observed data. The coefficients are interpreted as the impact of each independent variable on the outcome.

Logistic regression is widely used in many fields, including healthcare, finance, and marketing, to predict the probability of an event occurring. It is a simple and interpretable algorithm that can provide insights into the relationship between the independent and dependent variables.

3.4.8 Gaussian Naive Bayes

Gaussian Naive Bayes is a classification algorithm used in machine learning. It is based on Bayes' theorem, which states that the probability of a hypothesis (class) given evidence (features) is proportional to the probability of the evidence given the hypothesis and the prior probability of the hypothesis.

Gaussian Naive Bayes assumes that the probability distribution of the features is Gaussian (normal). This assumption simplifies the calculation of the probabilities and makes the algorithm computationally efficient. The "naive" in the name refers to the assumption that the features are independent of each other given the class.

The algorithm works by first calculating the prior probability of each class and the conditional probability of each feature given the class. Then, given a new instance with unknown class, it calculates the probability of the instance belonging to each class using Bayes' theorem and selects the class with the highest probability as the predicted class.

Gaussian Naive Bayes is commonly used in text classification and spam filtering tasks. However, it may not perform well when the features are not normally distributed or when there is strong correlation between the features.

3.4 Data Flow for System Implementation

DFD(data flow diagram) is drawn to represent the system of different levels of abstraction. For the purpose of understanding the structure of the system and the flow of data through various components for different functionalities,we have designed DFD levels 0 and 1.

DFD level-0

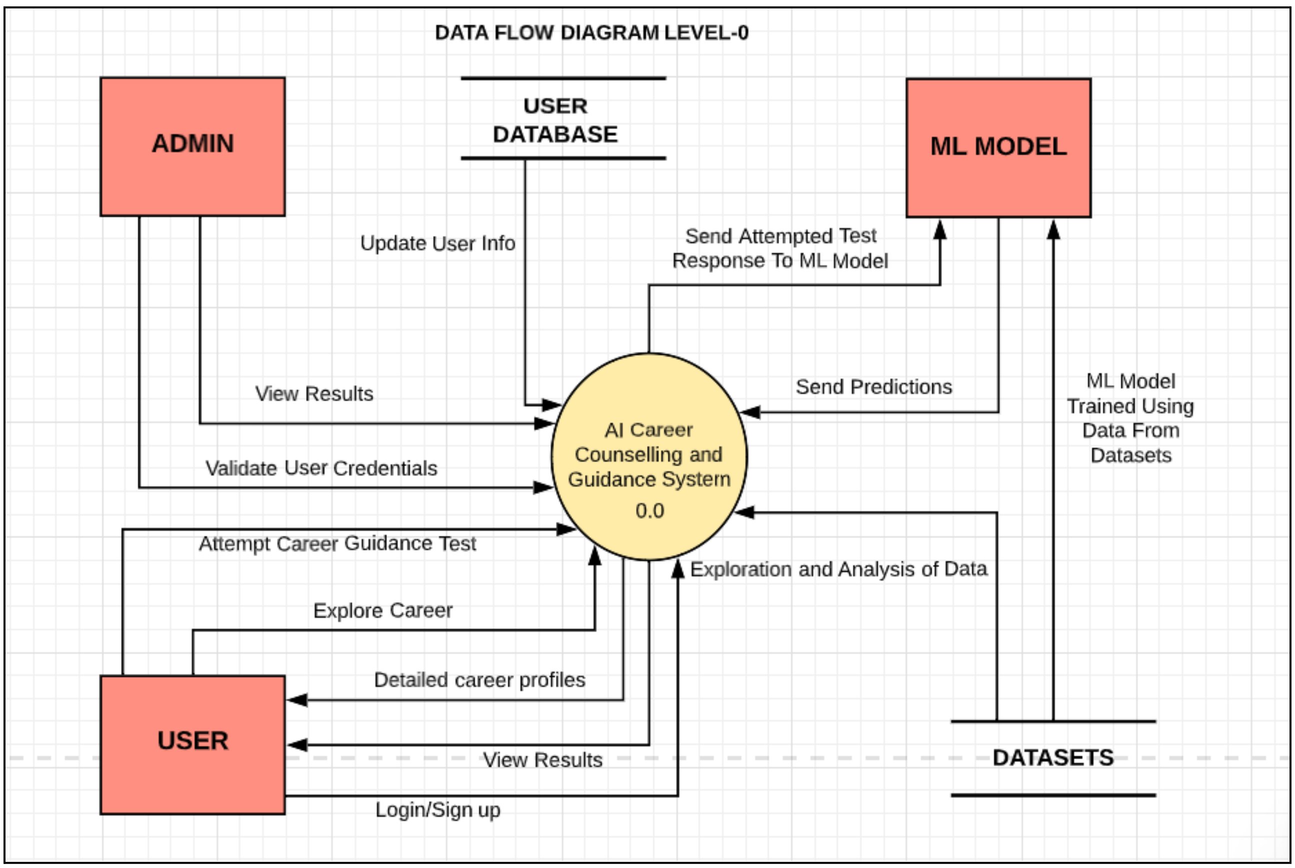

Fig 3. Data Flow Diagram (Level 0)

DFD level-1

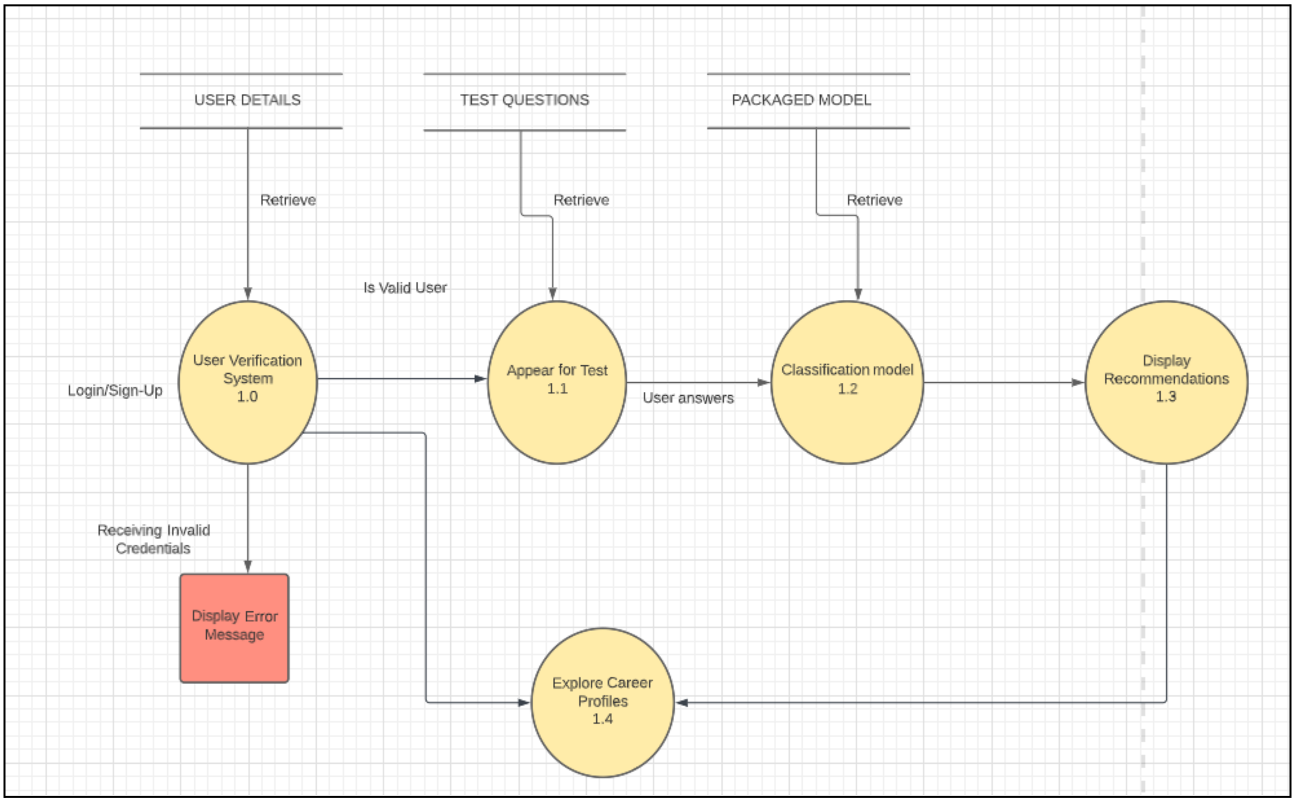

Fig 4. Data Flow Diagram (Level 1)

3.5 UML Diagrams

3.5.1 Use Case Diagram

Use case diagram is a graphical depiction of a user's possible interactions with a system. The diagram shows the various use cases and different possible users.

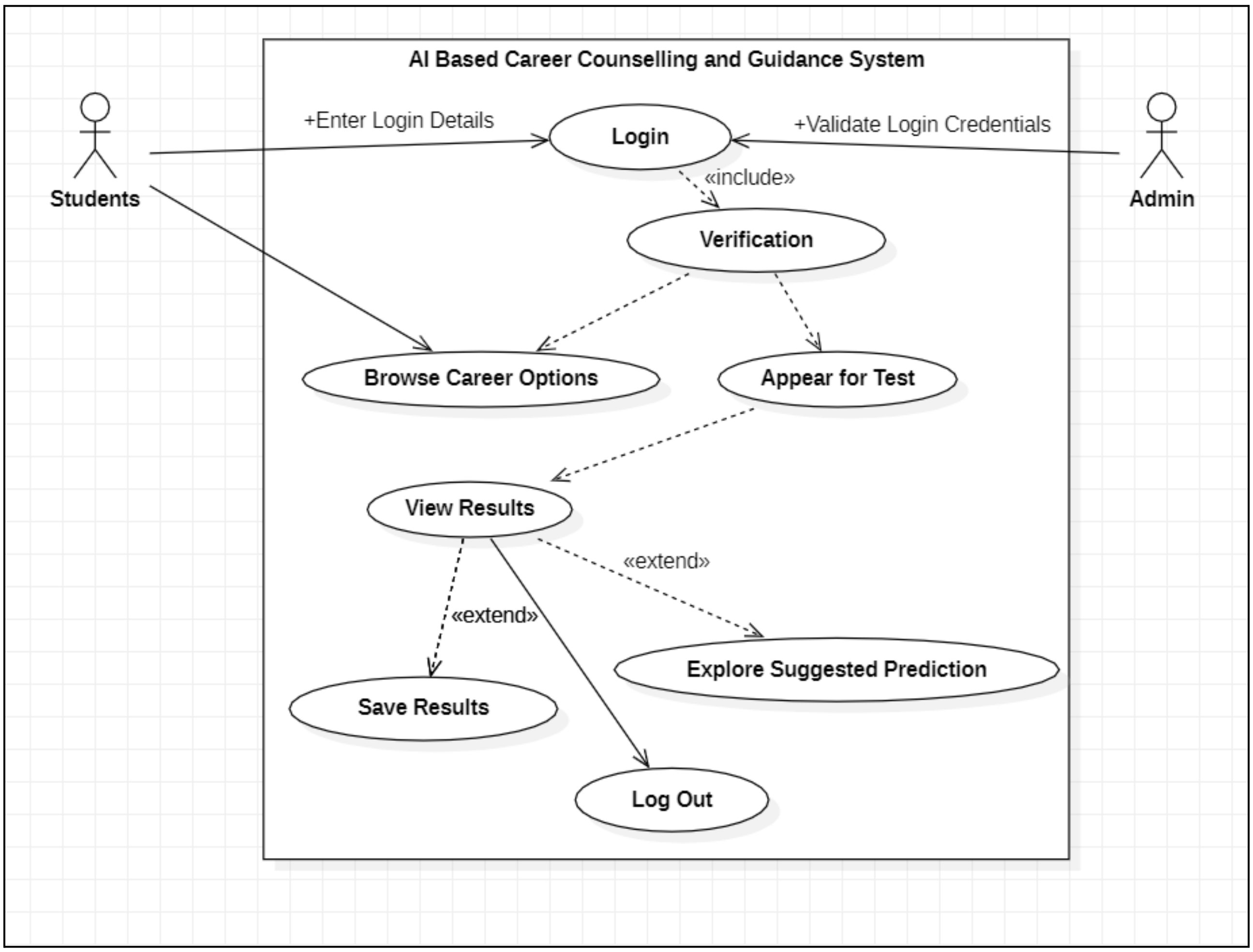

Fig 5. Use Case Diagram

3.5.2 Class Diagram

Class diagrams are the blueprints of your system or subsystem. You can use class diagrams to model the objects that make up the system, to display the relationships between the objects, and to describe what those objects do and the services that they provide. Class diagrams are useful in many stages of system design.

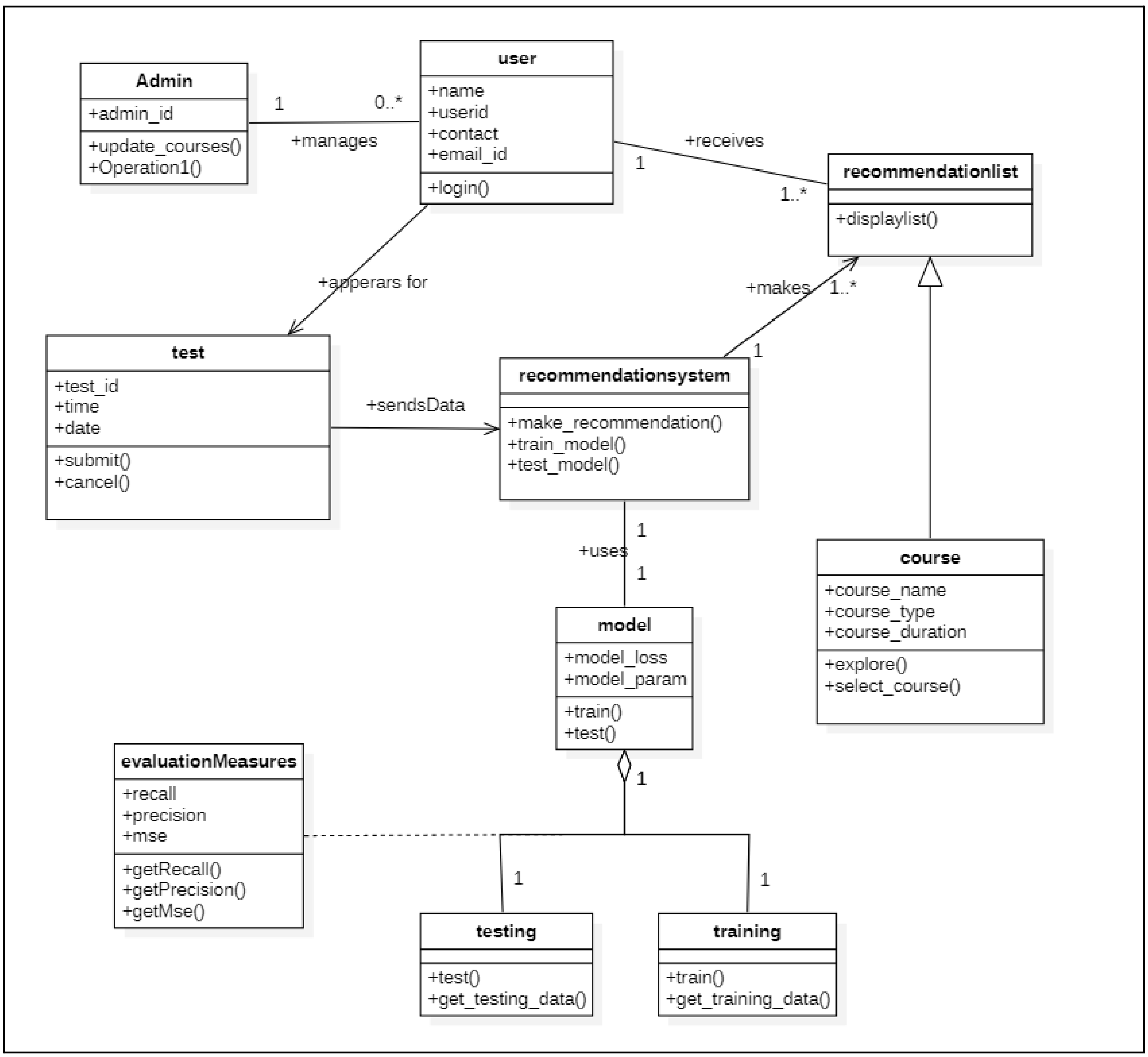

Fig 6. Class Diagram

3.5.3 Sequence Diagram

A sequence diagram is a Unified Modeling Language (UML) diagram that illustrates the sequence of messages between objects in an interaction. A sequence diagram consists of a group of objects that are represented by lifelines, and the messages that they exchange over time during the interaction.

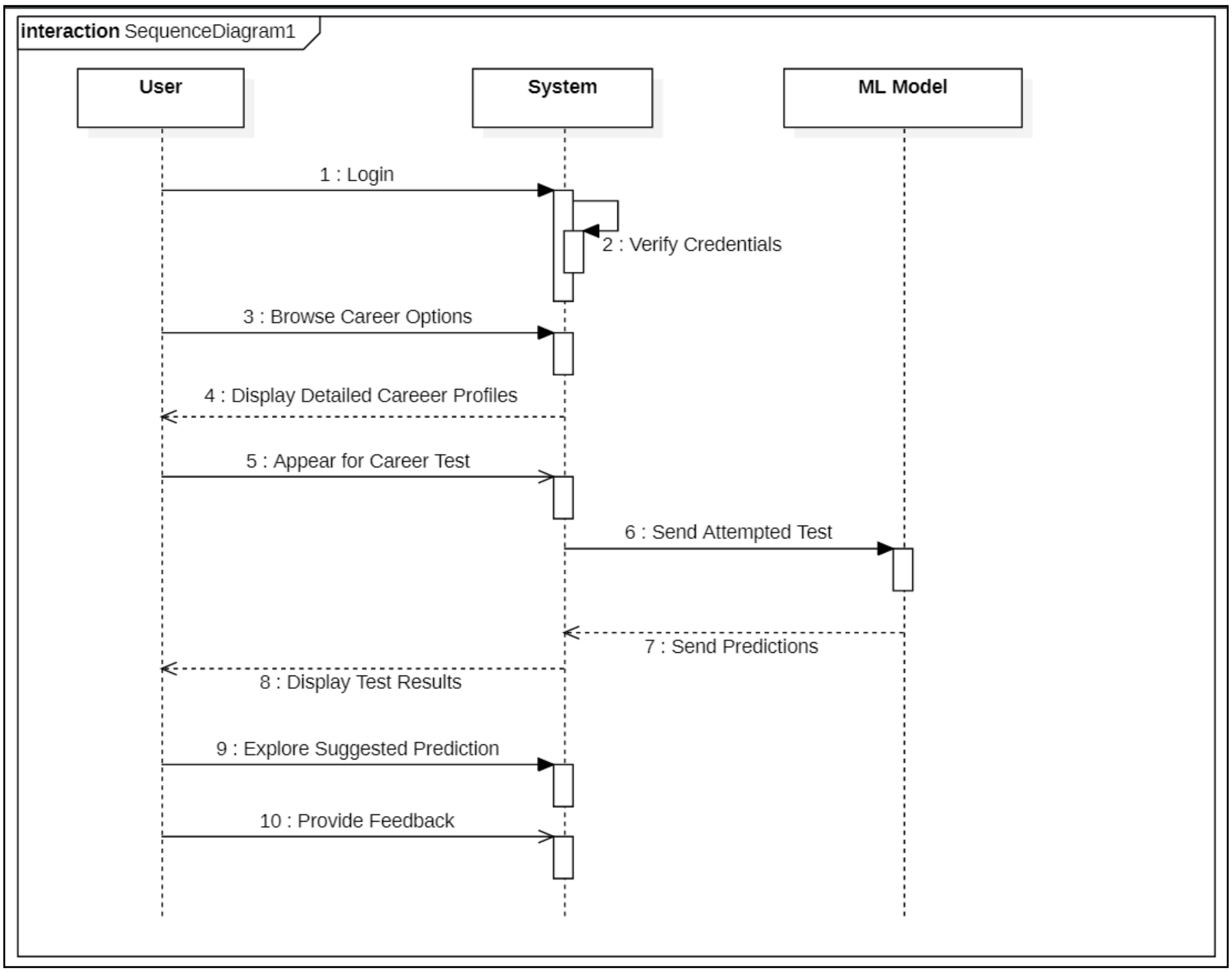

Fig 7. Sequence Diagram

3.5.4 Activity Diagram

An activity diagram visually presents a series of actions or flow of control in a system similar to a flowchart or a data flow diagram. Activity diagrams are often used in business process modeling. They can also describe the steps in a use case diagram.

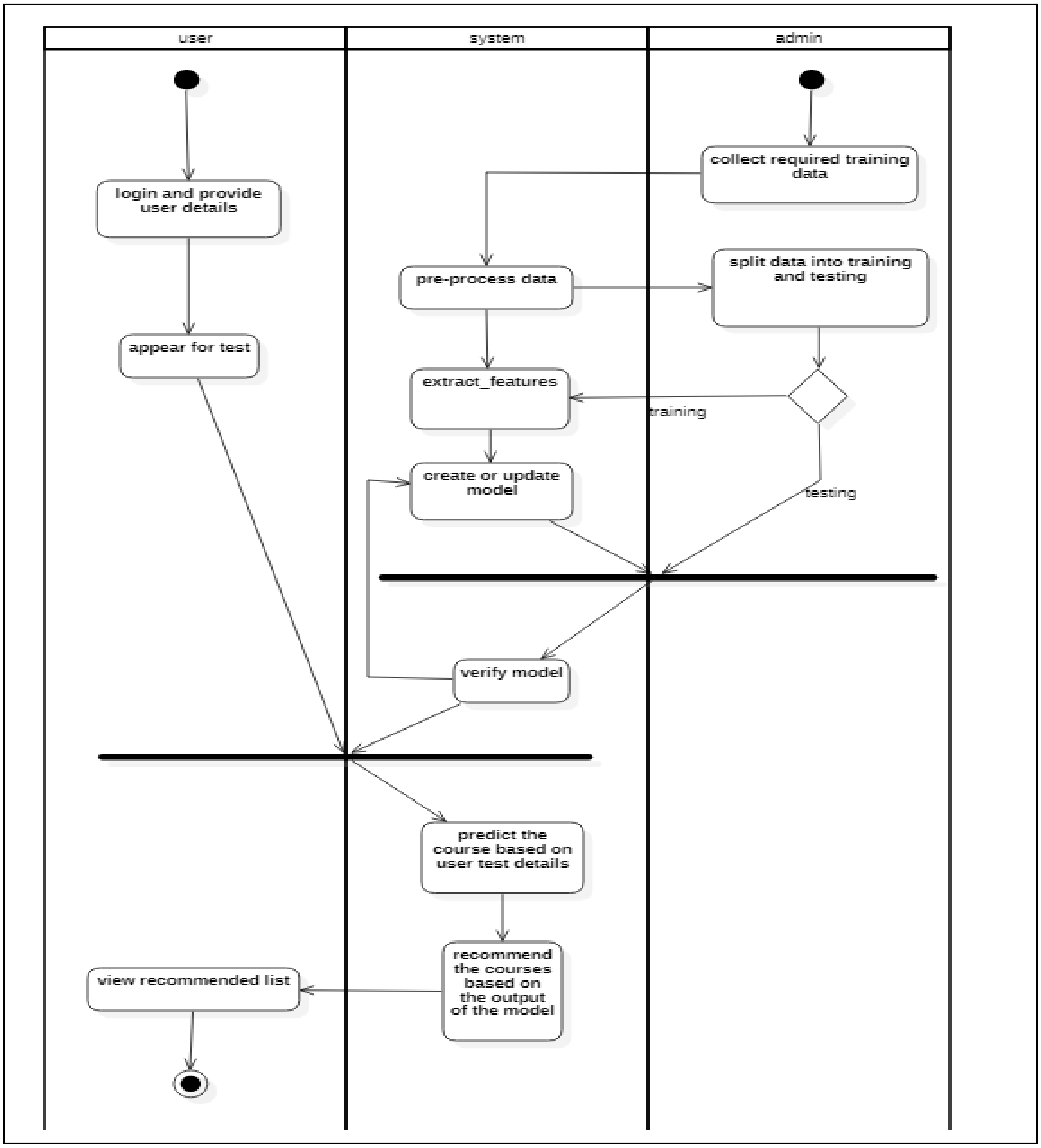

Fig 8. Activity Diagram

Chapter 4

Implementation

4.1 Methodology

The methodology adopted to carry out the implementation follows the machine learning pipeline approach starting from problem definition, dataset creation, dataset selection, data preprocessing, feature selection, implementation of various algorithms, choosing the best performing model, packaging the model and deploying the same in a user interface.

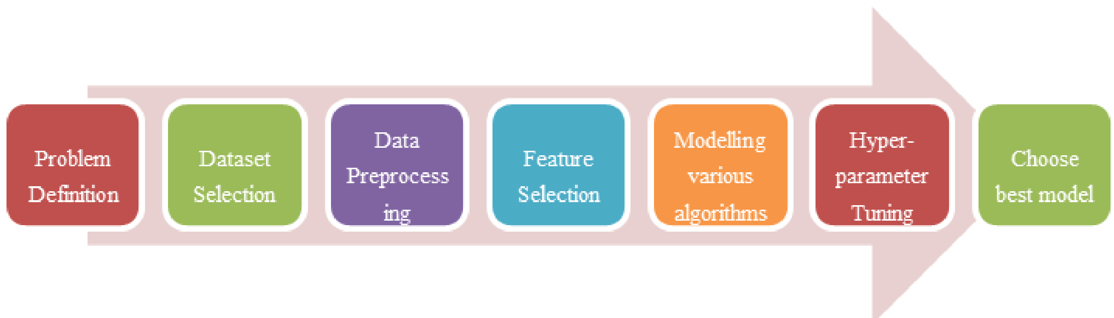

Fig 9. Workflow of the Machine Learning Pipeline

4.1.1 Problem Definition

The current research done in the field of Career Guidance system has been very nascent and limited in its approaches. In terms of ML Based approaches required to build an AI Career Guidance System, we require a good amount of labeled data. The requirements for which were not being able to be fulfilled in a single dataset. Hence, we took the ensemble approach of combining datasets to solve the problem of building a Career Guidance system.

Based on Literature Review conducted, it has been observed that apart form proficiency skills in various domains and concepts of Computer Science, a candidate's psychological traits also played an important role in predicting their career.

Thus, with sufficient literature reviewed and individual datasets obtained, we form our problem definition as:

"To combine datasets of psychological traits and proficiency in technological skills and apply Machine Learning models to predict the career of an individual in the field of Computer Science delivered through an interactive frontend."

4.1.2 Dataset Creation

The dataset used for the training of the classification model was produced by combining two distinct datasets that were discovered through a research process.

4.1.2.1 Dataset for Proficiency in Technical Skills

In the first dataset, Technical Skills in the domain of Computer Science were included as features with each skill being the independent variable and the role assigned based on the scores as the dependent variable. This dataset consisted of 18 skills each having 7 classes, hence there are 7 labels (1-7). These 7 labels were categorical and were encoded to numerical as follows:

1 - Not Interested

2 - Poor

3 - Beginner

4 - Average

5 - Intermediate

6 - Excellent

7 - Professional

The given dataset consists of several features related to various computer science domains and skills. Here's a brief description of each feature:

1. Database Fundamentals: This feature represents the knowledge and understanding of the basics of database management systems, including data modeling, querying, and database design.
2. Computer Architecture: This feature pertains to the understanding of computer hardware, including the design and organization of computer systems and their components.
3. Distributed Computing Systems: This feature represents knowledge and skills related to the design, implementation, and maintenance of distributed computing systems, which involves multiple computers working together to achieve a common goal.
4. Cyber Security: This feature pertains to knowledge and skills related to protecting computer systems, networks, and data from unauthorized access, theft, or damage.

5. Networking: This feature represents knowledge and skills related to the design, implementation, and maintenance of computer networks, including protocols, routing, and network security.
6. Software Development: This feature pertains to knowledge and skills related to the development of computer software, including programming languages, software design patterns, and software testing.
7. Programming Skills: This feature represents the proficiency in writing code in one or more programming languages.
8. Project Management: This feature represents the knowledge and skills related to managing software development projects, including project planning, scheduling, and resource allocation.
9. Computer Forensics Fundamentals: This feature pertains to the knowledge and skills related to collecting, analyzing, and preserving digital evidence in the context of computer crime investigations.
10. Technical Communication: This feature represents the ability to effectively communicate technical information, including documentation, user manuals, and technical reports.
11. AI ML: This feature pertains to the knowledge and skills related to artificial intelligence and machine learning, including algorithms, data analysis, and statistical modeling.
12. Software Engineering: This feature represents the knowledge and skills related to the application of engineering principles to the design, development, and testing of software systems.
13. Business Analysis: This feature represents the knowledge and skills related to analyzing business requirements and translating them into software solutions.
14. Communication skills: This feature pertains to the ability to effectively communicate with others, including verbal and written communication, listening, and interpersonal skills.
15. Data Science: This feature represents knowledge and skills related to the analysis, processing, and visualization of large datasets, including statistics, data mining, and machine learning.
16. Troubleshooting skills: This feature pertains to the ability to diagnose and solve problems related to computer hardware, software, or networks.
17. Graphics Designing: This feature represents the knowledge and skills related to graphic design, including design principles, color theory, and image editing software.
18. Role: This feature represents the role or job title of the individual, such as software developer, network administrator, or data analyst.

4.1.2.2 Dataset for Skills mapped to Psychological Traits

The second dataset taken into consideration contains various psychological traits that have been scraped from twitter profiles and bios and scored on certain parameters of psychological traits which are then mapped to the appropriate career roles that the individual possessed.

A total of 11 features are present in the dataset with 10 features being the independent variables and the 11th feature being the dependent variable- “Profession”.

The given dataset consists of several features related to personality traits and behavior. Here's a brief description of each feature:

1. Openness: This feature represents the degree to which a person is open to new experiences, ideas, and perspectives. Individuals with high openness tend to be imaginative, creative, and curious, while those with low openness tend to be more conventional and conservative.
2. Conscientiousness: This feature pertains to the degree to which a person is organized, responsible, and reliable. Individuals with high conscientiousness tend to be efficient, hardworking, and detail-oriented, while those with low conscientiousness tend to be more careless and disorganized.
3. Extraversion: This feature represents the degree to which a person is outgoing, sociable, and assertive. Individuals with high extraversion tend to be talkative, energetic, and confident, while those with low extraversion tend to be more reserved and introspective.
4. Agreeableness: This feature pertains to the degree to which a person is cooperative, empathetic, and compassionate. Individuals with high agreeableness tend to be kind, considerate, and trusting, while those with low agreeableness tend to be more competitive and skeptical.
5. Emotional_Range: This feature represents the degree to which a person experiences a wide range of emotions, from positive to negative. Individuals with high emotional range tend to be more emotionally volatile, while those with low emotional range tend to be more stable and consistent in their emotional experiences.
6. Conversation: This feature pertains to the degree to which a person is talkative and outgoing in social situations. Individuals with high conversation tend to enjoy talking and interacting with others, while those with low conversation tend to be more reserved and introverted.
7. Openness to Change: This feature represents the degree to which a person is open to change and innovation. Individuals with high openness to change tend to be flexible and adaptable, while those with low openness to change tend to be more resistant to new ideas and ways of doing things.

8. Hedonism: This feature pertains to the degree to which a person seeks pleasure and enjoyment in life. Individuals with high hedonism tend to be more focused on experiencing pleasure, while those with low hedonism tend to be more focused on other values, such as achievement or altruism.
9. Self-enhancement: This feature represents the degree to which a person seeks to enhance their own status, reputation, and self-esteem. Individuals with high self-enhancement tend to be more focused on achieving personal success and recognition, while those with low self-enhancement tend to be more focused on the needs of others.
10. Self-transcendence: This feature pertains to the degree to which a person is focused on the well-being of others and the world around them. Individuals with high self-transcendence tend to be more empathetic, altruistic, and spiritual, while those with low self-transcendence tend to be more focused on their own needs and desires.
11. Profession: This feature represents the role or job title of the individual, such as software developer, network administrator, or data analyst.

4.1.2.3 Preparing datasets to merge

Before combining datasets 1 and 2, there were numerous processes taken to prepare the datasets including the following:

Elimination of non-technical roles:

Due to the fact that the dataset also included non-technical roles, the non-technical roles were manually removed in order to conform to the functionality for predicting just computer science roles.

Role mapping in dataset 2 to dataset 1

In order to successfully combine the datasets, it was necessary for the predicted roles (dependent variable) to be consistent across both the datasets. It was decided that the roles established in dataset 1 would make up the final dataset since the roles defined in dataset 2 had a greater number of occurrences and the roles defined in dataset 1 were more general. As a result, the roles in dataset 2 were mapped to the roles in dataset 1 in such a way that each role in dataset 2 was mapped to the role in dataset 1 that was the most similar to it.

For example the following set of roles in dataset 2 were mapped to their corresponding roles in dataset 1:

.NET Developer ⟶ Software Developer

Role as in the Personality Dataset		Role according to the Proficiency Dataset
Project Executive	→	Application Support Engineer
Network Security Engineer	→	Cyber Security Specialist
Business Intelligence Analyst	→	Business Analyst
Internet Designer	→	Graphics Designer
Compliance Specialist	→	Information Security Specialist
Network Director	→	Networking Engineer
Computer Architect	→	Hardware Engineer
Computer Consultant	→	Helpdesk Engineer
Business Developer	→	Project Manager
Business Programmer	→	Data Scientist
Oracle Developer	→	Database Administrator
Software Writer	→	Technical Writer

Moreover, in order to normalize the dataset, the values of the skill ratings from dataset 1 were divided by the maximum value in order to bring the values to the range of 0-1, as in dataset 2.

4.1.2.4 Merging (includes techniques considered)

Several methods of merging were studied and tried:

1) Merging each data row in dataset 1 with the average value of corresponding role in dataset2
2) Merging each data row in dataset 1 with all data rows with that role in dataset 2.
3) Merging each row in dataset 1 with a random data row with corresponding role in dataset 2.

After implementation of all three methods, it was found that option 3 was the most viable given that the other 2 options led to overfitting.

4.1.3 Data Preprocessing

This step involveD preparing the data before performing an analysis on it. This involves removal of null values, dropping missing rows and normalizing the dataset.

4.1.4 Label Encoding

Label Encoding involves conversion of labels/words into numeric form so as to make them machine readable. For our dataset, we have 7 non-numeric classes, hence there are 7 labels (1-7).

4.1.5 Feature Selection

In this step, the important features within the dataset that are most important for analysis are chosen. Since the dataset contains survey questions, every feature is important in terms of generating a questionnaire for the user to attempt and predict their career choices. Hence, no feature selection is required for the given use-case. The dataset is split in the ratio of 80:20, where 80 is for training and 20 is for testing. The list of features selected are as follows:

- Database Fundamentals
- Computer Architecture
- Distributed Computing Systems
- Cyber Security
- Networking
- Development
- Programming Skills
- Project Management
- Computer Forensics Fundamental
- Technical Communication
- AI ML
- Software Engineering
- Business Analysis
- Communication skills
- Data Science
- Troubleshooting skills
- Graphics Designing
- Openness

- Conscientiousness
- Extraversion
- Agreeableness
- Emotional Range
- Conversation
- Openness to Change
- Hedonism
- Self-enhancement
- Self-transcendence

4.2 Results

For creating the baseline model, we used the following algorithms:

i) **KNN Algorithm:** Within the KNN algorithm, the hyperparameter of the number of neighbors was tested for three different values, namely- 5, 7 and 9.

We noticed that the optimal value of neighbors is 7 since the accuracy of the model with n_neighbors as 5 as well as 9 is less than the case where n_neighbors is taken as 7. The training accuracy is **98.52%** and the testing accuracy is **91.78%**.

ii) **Decision Trees:** A tree is composed of nodes, and those nodes are chosen looking for the optimum split of the features. In the decision tree Python implementation of the scikit-learn library, this is made by the parameter 'criterion'. This parameter is the function used to measure the quality of a split and it allows users to choose between 'gini' or 'entropy'. For our decision tree classifier, we used the **gini criterion**. At a **maximum depth of 14** we get the training accuracy as **94.36%** and testing accuracy as **93.14%.**

iii) **Random Forest:** RFC gives 100% training and testing accuracy, since there is no way of explaining the high accuracy, we understand that the model is overfitting.

iv) **Logistic Regression:** This class implements regularized logistic regression using the 'liblinear' library and 'newton-cg' solvers. Regularization is applied by default and newton-cg supports only L2 regularization. We get a training accuracy of 99.93% and a testing accuracy of 99.89%.

v) **Multilayer Perceptron:** This model optimizes the log-loss function using LBFGS or stochastic gradient descent. With hidden layer size as (12,4) the model gives a training accuracy of **91.04%** and testing accuracy of **90.14%**.

vi) **Support Vector Machines:** SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable. A separator between the categories is found, then the data are transformed in such a way that the separator could be drawn as a hyperplane. With kernel type **rbf** we get an training accuracy of **98.98%** and testing accuracy of **95.70%**

vii) **Gaussian Naive Bayes:** GaussianNB implements the Gaussian Naive Bayes algorithm for classification. The likelihood of the features is assumed to be Gaussian. The training accuracy comes out to be **99.96%** and testing accuracy comes out as **100%**.

4.3 Screenshots

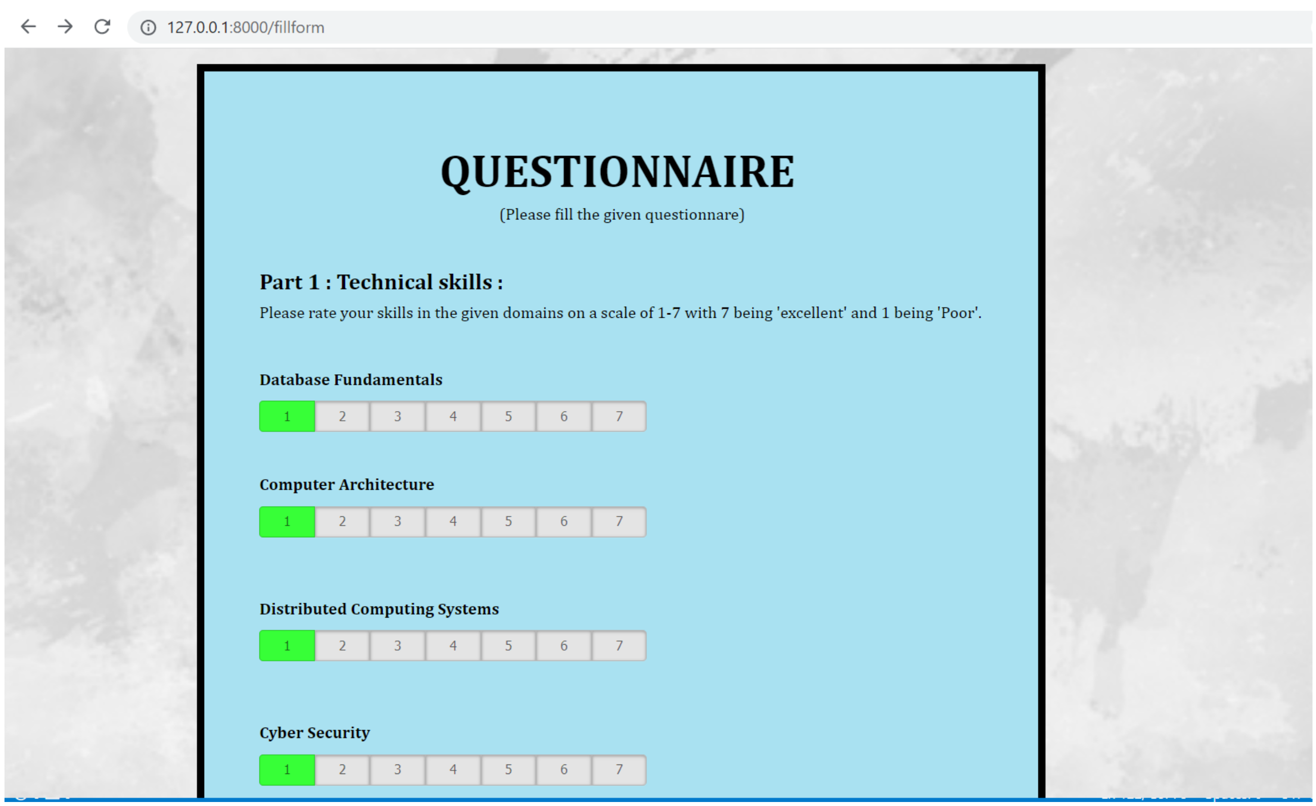

Fig 10: User Interface and Working (1)

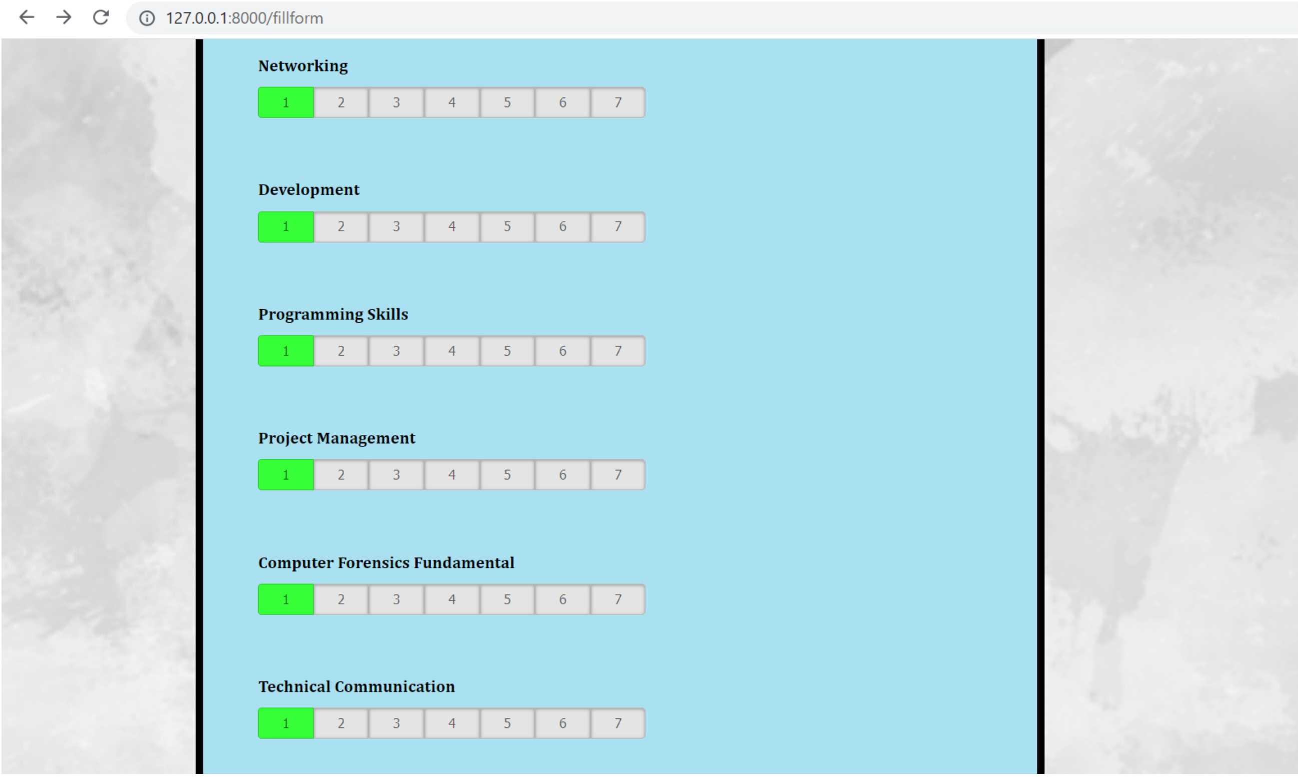

Fig 11: User Interface and Working (2)

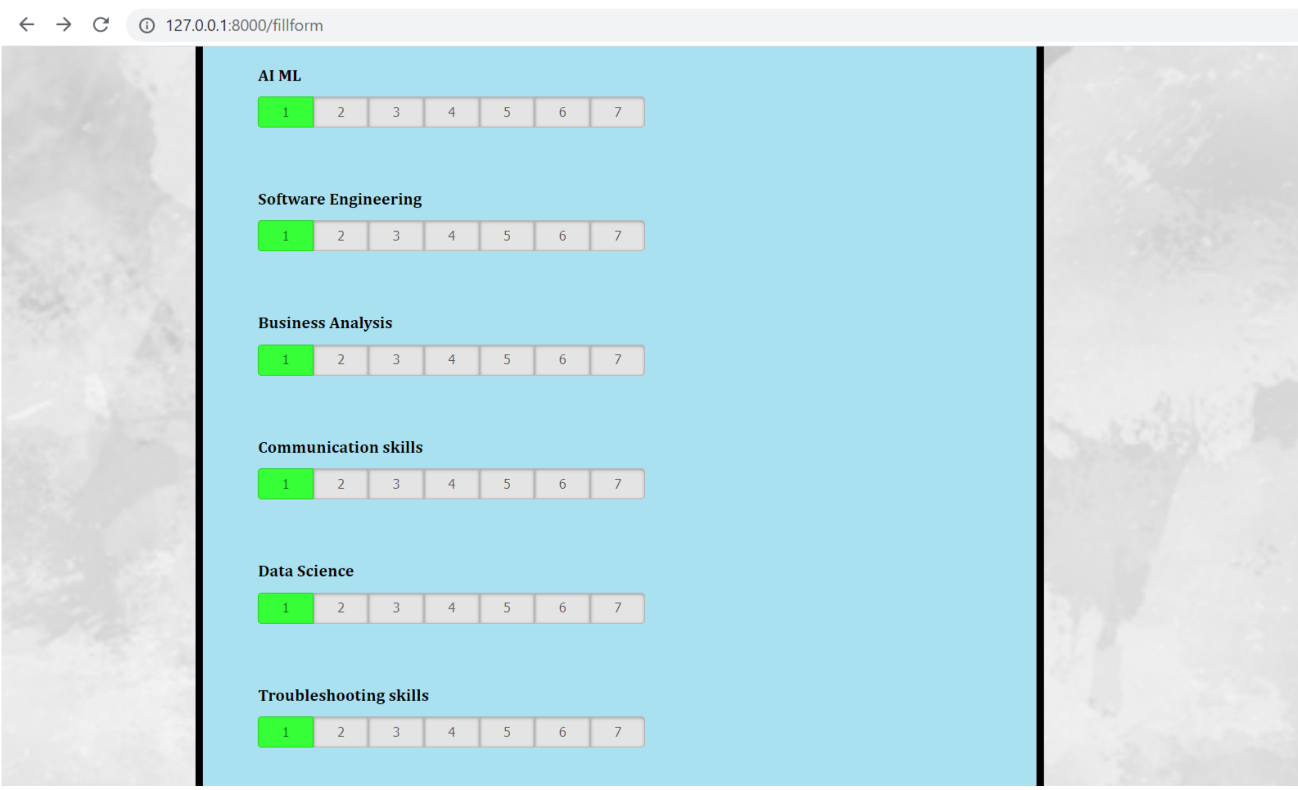

Fig 12: User Interface and Working (3)

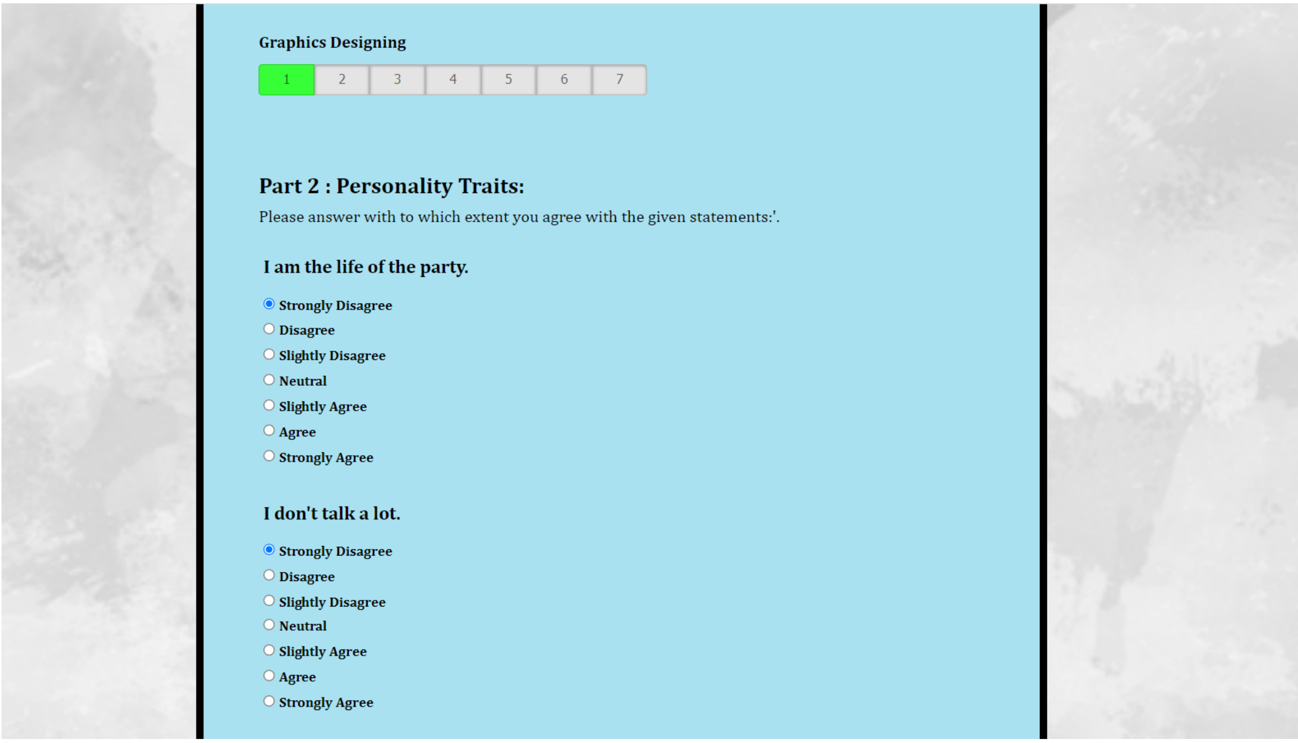

Fig 13: User Interface and Working (4)

Fig 14: User Interface and Working (5)

Fig 15: User Interface and Working (6)

Fig 16: User Interface and Working (7)

Fig 17: User Interface and Working (8)

Fig 18: User Interface and Working (9)

Fig 19: User Interface and Working (10)

Fig 20: User Interface and Working (11)

Fig 21: User Interface and Working (12)

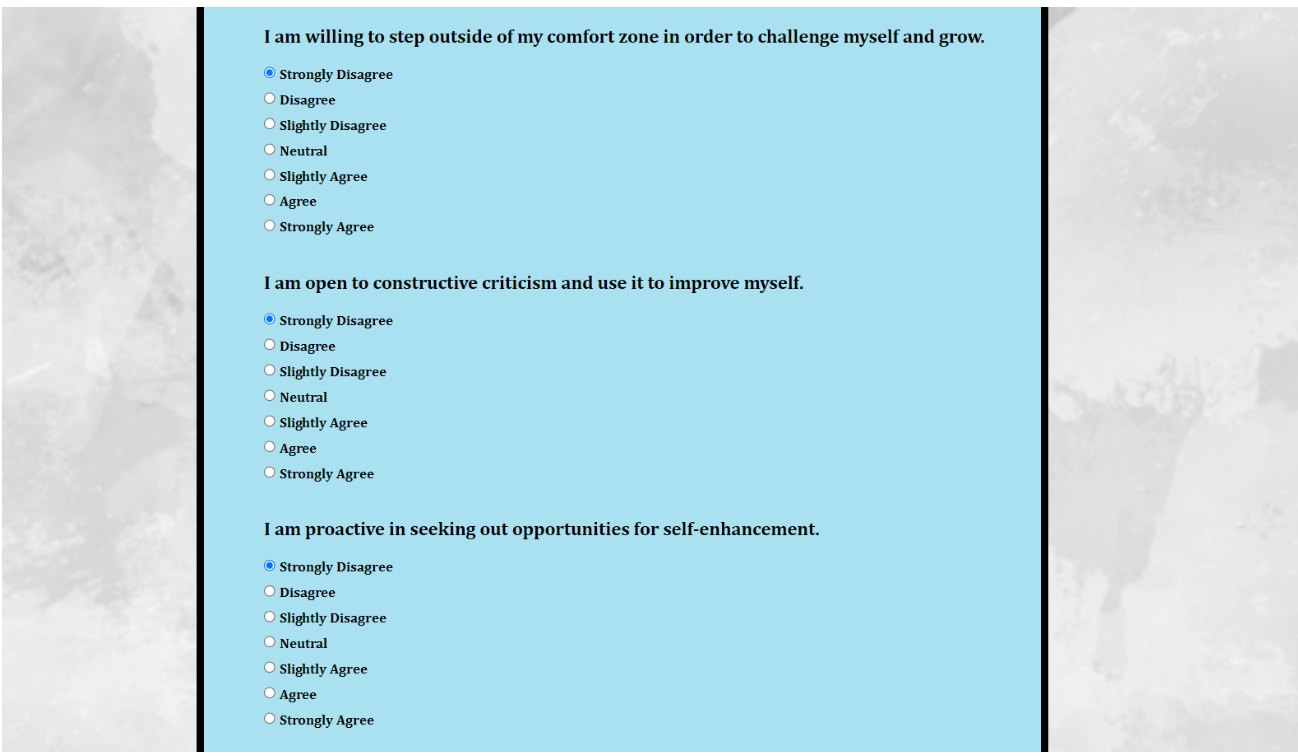

Fig 22: User Interface and Working (13)

References

[1] Siswipraptini, Puji & Spits Warnars, Harco Leslie Hendric & Ramadhan, Arief & Budiharto, Widodo. (2022). Trends and Characteristics of Career Recommendation Systems for Fresh Graduated Students. 355-361. 10.1109/ICIET55102.2022.9779037.

[2] Westman, Stina & Kauttonen, Janne & Klemetti, Aarne & Korhonen, Niilo & Manninen, Milja & Mononen, Asko & Niittymäki, Salla & Paananen, Henry. (2021). Artificial Intelligence for Career Guidance – Current Requirements and Prospects for the Future. IAFOR Journal of Education. 9. 43-62. 10.22492/ije.9.4.03.

[3] Adapa, S., Gandi, N., Alekya, V., Lakshmi Durga, G. (2021). Learning Style Recommender System Using VAK Technique and Machine Learning. In: Satapathy, S.C., Bhateja, V., Ramakrishna Murty, M., Gia Nhu, N., Jayasri Kotti (eds) Communication Software and Networks. Lecture Notes in Networks and Systems, vol 134. Springer, Singapore. https://doi.org/10.1007/978-981-15-5397-4_20

[4] Shi, Z., Wu, Z., Zhang, Z., Chen, Y., & Liu, X. (2022). Learning Path Planning Algorithm Based on Career Goals and Artificial Intelligence. International Journal of Emerging Technologies in Learning (iJET), 17(10), pp. 256–272. https://doi.org/10.3991/ijet.v17i10.28455

[5] A. Ghimire, T. Dorsch and J. Edwards, "Introspection with Data: Recommendation of Academic Majors Based on Personality Traits," 2022 Intermountain Engineering, Technology and Computing (IETC), Orem, UT, USA, 2022, pp. 1-6, doi: 10.1109/IETC54973.2022.9796766.

[6] P. D. Schalk, D. P. Wick, P. R. Turner and M. W. Ramsdell, "Predictive assessment of student performance for early strategic guidance," 2011 Frontiers in Education Conference (FIE), Rapid City, SD, USA, 2011, pp. S2H-1-S2H-5, doi: 10.1109/FIE.2011.6143086.

[7] C. D. Casuat and E. D. Festijo, "Identifying the Most Predictive Attributes Among Employability Signals of Undergraduate Students," 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), Langkawi, Malaysia, 2020, pp. 203-206, doi: 10.1109/CSPA48992.2020.9068681.

[8] Li, Xiaomin & Zhang, Jianrong & Wan, Jiabing & Zhang, JinKai & Zhu, Chenchao & Guo, Nianfeng & Xu, Hao. (2019). Implementation of system for college students' career planning based on user interest model. Journal of Physics: Conference Series. 1237. 022067. 10.1088/1742-6596/1237/2/022067.

[9] R. H. Rangnekar, K. P. Suratwala, S. Krishna and S. Dhage, "Career Prediction Model Using Data Mining and Linear Classification," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697689.

[10] M. Y. Arafath, M. Saifuzzaman, S. Ahmed and S. A. Hossain, "Predicting Career Using Data Mining," 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 2018, pp. 889-894, doi: 10.1109/GUCON.2018.8674995.

[11] B. Harsha, N. Sravanthi, N. Sankeerthana and M. Suneetha, "Career Choice Using Machine Learning Algorithms," 2022 International Conference on Inventive Computation Technologies (ICICT), Nepal, 2022, pp. 171-176, doi: 10.1109/ICICT54344.2022.9850697.

[12] K. Joshi, A. K. Goel and T. Kumar, "Online Career Counsellor System based on Artificial Intelligence: An approach," 2020 7th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2020, pp. 1-4, doi: 10.1109/ICSSS49621.2020.9202024.

[13] Al-Dossari, Hmood & Nughaymish, F. & Al-Qahtani, Z. & Alkahlifah, M. & Alqahtani, A.. (2020). A Machine Learning Approach to Career Path Choice for Information Technology Graduates. Engineering, Technology & Applied Science Research. 10. 6589-6596. 10.48084/etasr.3821.

[14] Khan, Sahid & Darbari, Manuj. (2022). Myers-Briggs Type Indicator and Social Media Mining Based Career Counselling System. Volume-01. 10.5281/zenodo.6726646.

[15] S. Vignesh, C. Shivani Priyanka, H. Shree Manju and K. Mythili, "An Intelligent Career Guidance System using Machine Learning," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 987-990, doi: 10.1109/ICACCS51430.2021.9441978.

[16] Thomas, A., Varghese, A.K., Alex, P.L., Mathews, B.J., Dhanya, L.K. (2022). Analysis of Machine Learning Algorithms for Predicting the Suitable Career After High School. In: Bindhu, V., Tavares, J.M.R.S., Du, KL. (eds) Proceedings of Third International Conference on Communication, Computing and Electronics Systems . Lecture Notes in Electrical Engineering, vol 844. Springer, Singapore. https://doi.org/10.1007/978-981-16-8862-1_7

[17] J. Britto, S. Prabhu, A. Gawali and Y. Jadhav, "A Machine Learning Based Approach for Recommending Courses at Graduate Level," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 117-121, doi: 10.1109/ICSSIT46314.2019.8987815.

[18] A. Kamal, B. Naushad, H. Rafiq and S. Tahzeeb, "Smart Career Guidance System," 2021 4th International Conference on Computing & Information Sciences (ICCIS), Karachi, Pakistan, 2021, pp. 1-7, doi: 10.1109/ICCIS54243.2021.9676408.

[19] K. Joshi, A. K. Goel and T. Kumar, "Online Career Counsellor System based on Artificial Intelligence: An approach," 2020 7th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2020, pp. 1-4, doi: 10.1109/ICSSS49621.2020.9202024.

[20] Tayade, T., Akarte, R., Sorte, G., Tayade, R., Khodke, P. (2021). Data Mining Approach to Job Recommendation Systems. In: Raj, J.S. (eds) International Conference on Mobile Computing and Sustainable Informatics . ICMCSI 2020. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-49795-8_48

[21] R. Ajoodha, S. Dukhan and A. Jadhav, "Data-Driven Student Support for Academic Success by Developing Student Skill Profiles," 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC), 2020, pp. 1-8, doi: 10.1109/IMITEC50163.2020.9334109.

[22] A. H. A. Rashid, M. Mohamad, S. Masrom and A. Selamat, "Student Career Recommendation System Using Content-Based Filtering Method," 2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS), IPOH, Malaysia, 2022, pp. 60-65, doi: 10.1109/AiDAS56890.2022.9918766.

[23] T. V. Yadalam, V. M. Gowda, V. S. Kumar, D. Girish and N. M., "Career Recommendation Systems using Content based Filtering," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 660-665, doi: 10.1109/ICCES48766.2020.9137992.

[24] A. Gugnani, V. K. Reddy Kasireddy and K. Ponnalagu, "Generating Unified Candidate Skill Graph for Career Path Recommendation," 2018 IEEE International Conference on Data Mining Workshops (ICDMW), Singapore, 2018, pp. 328-333, doi: 10.1109/ICDMW.2018.00054.

[25] I. Dutta, A. Shah, M. M. Safa and K. Jayavel, "Building a Recommendation System for Career Advice for Students from Professionals," 2022 International Conference on Computer Communication and Informatics (ICCCI), 2022, pp. 1-10, doi: 10.1109/ICCCI54379.2022.9740837.

[26] Hui Zhang, Zhuonan Zheng, "Application and Analysis of Artificial Intelligence in College Students' Career Planning and Employment and Entrepreneurship Information Recommendation", Security and Communication Networks, vol. 2022, Article ID 8073232, 8 pages, 2022. https://doi.org/10.1155/2022/8073232

[27] Ghosh, Aritra & Woolf, Beverly & Zilberstein, Shlomo & Lan, Andrew. (2020). Skill-based Career Path Modeling and Recommendation. 1156-1165. 10.1109/BigData50022.2020.9377992.

[28] Fathian Brojeny M., "Applying Artificial Intelligence to Career Consultation: A Case Based Reasoning Paradigm for Knowledge Management", EIJH,11 (3), pp. 81-90

[29] R. Singh, R. Pandita, K. Kalyanaraman, G.S. Chhabra, "Career Guidance System", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 7, Issue 2, February 2018, pp 171-175, doi: 10.17148/IJARCCE.2018.7230

[30] M. Qamhieh, H. Sammaneh and M. N. Demaidi, "PCRS: Personalized Career-Path Recommender System for Engineering Students," in IEEE Access, vol. 8, pp. 214039-214049, 2020, doi: 10.1109/ACCESS.2020.3040338.

[31] W. M. Aly, K. A. Eskaf and A. S. Selim, "Fuzzy mobile expert system for academic advising," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, Canada, 2017, pp. 1-5, doi: 10.1109/CCECE.2017.7946846.

[32] Wulansari, R. E., Sakti, R., Ambiyar, A., Giatman, M., Syah, N., & Wakhinuddin, W. . (2022). Expert System For Career Early Determination Based On Howard Gardner's Multiple Intelligence . Journal of Applied Engineering and Technological Science (JAETS), 3(2), 67–76. https://doi.org/10.37385/jaets.v3i2.568

[33] Guo, P., Xiao, K., Ye, Z. et al. Intelligent career planning via stochastic subsampling reinforcement learning. Sci Rep 12, 8332 (2022). https://doi.org/10.1038/s41598-022-11872-8

[34] Marios Kokkodis, Panagiotis G. Ipeirotis (2020) Demand-Aware Career Path Recommendations: A Reinforcement Learning Approach. Management Science 67(7):4362-4383. https://doi.org/10.1287/mnsc.2020.3727

[35] Nair, A., Padhy, I., Nikhil, J.K., Sindhura, S., Vandana, M.L., Vijay Krishna, B.S. (2022). AI-Based Career Counselling with Chatbots. In: Raj, J.S., Shi, Y., Pelusi, D., Balas, V.E. (eds) Intelligent Sustainable Systems. Lecture Notes in Networks and Systems, vol 458. Springer, Singapore. https://doi.org/10.1007/978-981-19-2894-9_46

[36] Khan, A., Masrani, V., Ojha, A., Hamdare, S. (2021). Personal Assistant for Career Coaching. In: Sharma, H., Saraswat, M., Yadav, A., Kim, J.H., Bansal, J.C. (eds) Congress on Intelligent Systems. CIS 2020. Advances in Intelligent Systems and Computing, vol 1334. Springer, Singapore. https://doi.org/10.1007/978-981-33-6981-8_38

[37] Nguyen , D., Dinh , N., Pham-Nguyen, C., Le Dinh, T., Nguyen Hoai Nam, L. (2022). ITCareerBot: A Personalized Career Counselling Chatbot. In: Szczerbicki, E., Wojtkiewicz, K., Nguyen, S.V., Pietranik, M., Krótkiewicz, M. (eds) Recent Challenges in Intelligent Information and Database Systems. ACIIDS 2022. Communications in Computer and Information Science, vol 1716. Springer, Singapore. https://doi.org/10.1007/978-981-19-8234-7_33