**Pathfinder**

Personalized Career Path Recommender System for University Students

Members:

Leshan Chathuranga Kuruppuarachchi – 300363077 (Team Lead)

Hyunhee Kim - 300372531

Course: CSIS-4495 - Applied Research Project

Section: 050

**Table of Contents**

[1. Introduction 4](#_Toc208570389)

[2. Proposed Research Project 6](#_Toc208570390)

[2.1 Methodology 6](#_Toc208570391)

[2.2 Technologies to be used 7](#_Toc208570392)

[2.3 Expected Results 7](#_Toc208570393)

[3. Project Planning and Timeline 9](#_Toc208570394)

[3.1 Responsibilities 10](#_Toc208570395)

[3.2 Project Timeline Gantt Chart 10](#_Toc208570396)

[4. Project Contract 11](#_Toc208570397)

[5. Work Date/Hours logs 12](#_Toc208570398)

[6. References 13](#_Toc208570399)

# **Introduction**

In today’s competitive job market, university students face increasing challenges when choosing their future career path. The traditional methods of career finding, such as personality tests, career questionnaires, or career advisers often fail to understand individual skills and interests, because they are too broad and fail to understand individual differences in skills, academic background and personality.

These limitations have led to interest in the application of intelligent systems in the field of career counseling and education technology.

To overcome these limitations, personalized career recommendation system has emerged as a solution. These systems collect user-provided information such as gender, academic background, skills, interests, and analyze the data to generate personalized recommendations. Based on this context, this research addresses two key questions:

1. How can machine learning models improve the accuracy and related to these recommendations compared to traditional methods?
2. How can student-provided information be used effectively to generate personalized career recommendations?

These questions are important because wrong career choices can lead to poor experience for students and labor market needs. By developing a data-driven and personalized recommendation system, this research aims to help students make confident decisions about their future careers. Recent career-recommendation studies range from classical ML pipelines to modern hybrid ensembles. For example,

* Classical content-based pipelines [1]. Systems preprocess user-provided descriptions, vectorize with TF-IDF, segment users via K-Means, reduce dimensions with PCA, and rank careers using cosine similarity; a full-stack reference implementation integrates these components into a web app (e.g., Django).
* Hybrid stacking ensembles [2]. More recent work for competency-based education creates a structured student dataset (5 features, 5,000 records) and combines Deep Neural Networks + Random Forest via stacking to recommend STEM tracks, reporting ~90.06% accuracy and 92.07% precision with 5-fold CV .
* Mobile applications (CareerX) [3]. CareerX has demonstrated the integration of NLP and machine learning into mobile applications, with user satisfaction rates of around 90%.
* Hybrid filtering approaches [4]. Other studies combine collaborative and content-based methods to generate personalized educational roadmaps.
* Hierarchical multi-tiered systems [5]. These systems address sparsity and cold-start problems, achieving over 99% accuracy across diverse career dataset.

Across this literature, several gaps remain that our project targets.

1. Feature scope and personalization: Classical pipelines often rely on limited or generic features, hybrid models show gain but are frequently tailored to a single context (e.g., STEM) and a fixed, low dimensional schema
2. Feedback lookups: User feedback is rarely closed-looped into model updates in near-real time. Most evaluations are static offline metrics
3. Generalizability: Results validated on a specific region/population or track set may not transfer broadly without careful adaptation.

Based on these gaps, this research assumes that a machine learning based recommendation system using effectively diverse student-provided information will significantly enhance personalization and accuracy. The expected benefit is to provide students with reliable, data-driven guidance for career choices, while contributing to the development of intelligent systems in education technology.

# **Proposed Research Project**

The proposed Pathfinder project is a data-driven experimental design. The main objective is to build a personalized career recommendation system that uses user provided information, such as educational background, gender, skills and interests, to recommend personalized career pathways.

## **Methodology**

Pathfinder follows a five-step methodology, data collection and preprocessing, model training and recommendation engine design, skill gap analysis, job market trend visualization and system implementation.

1. Data Collection and Preprocessing
2. Training Data: Public datasets from Kaggle and web scraping data from LinkedIn, Indeed will be used to train the model.
3. Operational Data: Datasets provided by users through the Pathfinder website, such as major, GPA, gender, interest, and skills will be used for real recommendations.
4. Model Training and Recommendation Engine Design

For the recommendation engine, we will train models using the collected data to come up with possible career paths using similarity. We are going to use Cosine similarity, KNN, or ML classifier.

1. Skill Gap Analysis

For each recommended career, the required and desirable skills will be compared with the user’s current skill set.

1. Job Market Trend Visualization

Job market dataset will be analyzed to highlight the demand for recommended careers in the last few years.

## **Technologies to be used**

This project will integrate several technologies

* Operating System / Platform: The system will run on both Windows and macOS environments.
* Programming Language / Frameworks: Python will be the primary language, using machine learning libraries such as scikit-learn, TensorFlow, and PyTorch.
* Database: MySQL or H2 database will be used to store cleaned data for model and visualization data.
* Frontend: React or Vue.js will be used to build a responsive and interactive user interface that allows users to input their profiles and view career recommendations, skill gap analysis, and job market trends.
* Backend: FastAPI or Flask will serve as the backend to manage model inference, skill gap analysis, and job trend visualization.
* Visualization Tools: Libraries such as Chart.js, D3.js, or Recharts will be used to display skill gaps and job market trends on the dashboard.
* Collaborative Development: GitHub will be used for version control, source code management, and team collaboration. Team members will work together using branches, pull requests and reviews during development.

## **Expected Results**

The expected outcomes of this research:

1. Improved accuracy and personalization: Compared to traditional career counseling methods, the proposed system is expected to provide more accurate and personalized career recommendations.
2. Top N career recommendations: The system will recommend the top 3-5 most suitable careers to each user, providing a clear and actionable career path.
3. Skill gap analysis: Users will be provided with a breakdown of the missing skills required for their desired careers, along with the relative importance and estimated effort needed to close the gap.
4. Job market insights: Visualization of labor market trends over the last five years will enable users to understand the demand for different careers and make more informed decisions.
5. Interactive career platform: A user-friendly web platform will allow students and job seekers to explore potential career paths, and continuously refine their skills based on feedback.
6. Contribution to educational technology: The research will contribute to the advancement of intelligent recommendation systems in the field of career counseling and educational technology.

# **Project Planning and Timeline**

This section presents the planned schedule for completing the personalized Career Path Recommender System. The project is divided into clear phases, each with specific milestones and deliverables, to ensure steady progress and timely completion. The timeline follows the official course deadlines for the proposal, progress reports, midterm and final submissions.

Table 1: Proposed project schedule

|  |  |  |  |
| --- | --- | --- | --- |
| Phase | Duration/  Deadline | Milestones | Deliverables |
| Phase 1:  Literature Review & Project Initialization | 2025/09/04 – 2025/09/11 | Collect references on career recommendation systems | Project Proposal |
| Phase 2:  Data Collection & Preprocessing | 2025/09/08 – 2025/09/25 | * Identify data sources * Implement scraping & preprocessing | Cleaned dataset |
| Phase 3:  System Design | 2025/09/12 – 2025/09/25 | * Define system architecture (frontend, backend) * Design recommendation logic | * System architecture diagram * Progress Report 1 |
| Phase 4:  Recommendation Model Development | 2025/09/17 – 2025/10/16 | * Implement recommendation model * Train/test with sample student data | * Working recommendation engine * Progress Report 2 |
| Phase 5:  Frontend & User Interaction | 2025/10/08 – 2025/11/05 | * Build React UI for student input * Connect frontend to backend API | * Functional prototype * Midterm Report * Progress Report 3 |
| Phase 6:  Testing & Refinement | 2025/11/03 – 2025/11/18 | * Test system with multiple profiles * Debug & bug fixing | * Progress Report 4 * Midterm Video Report |
| Phase 7:  Documentation & Final Report | 2025/11/12 – 2025/11/27 | * Prepare final report and presentation | * Progress Report 5 * Final Report & Implementation * Project Defense |

## **Responsibilities**

Leshan Chathuranga Kuruppuarachchi

* Overall project management and coordination
* Data scraping design and implementation
* Model design and implementation
* Frontend development
* Testing and bug fixing

Hyunhee Kim

* Data collection and preprocessing
* Model design and implementation
* Backend API development
* Data visualization implementation
* Testing and bug fixing
* Documentation

## **Project Timeline Gantt Chart**

Figure 1: Proposed project Gantt chart

# **Project Contract**

**Agreement**

This contract confirms agreement between the team members on the scope of work and timeline as outlined in the proposal.

* The members agree to carry out the project in accordance with the defined phases, milestones, and deliverables.
* All tasks, including data collection, model development, system design, testing, documentation, and presentation, will be completed by the members.
* Project progress will be communicated through official progress reports and consultation sessions.
* The members commit to submitting all deliverables on or before the deadlines specified in the proposal.
* Meetings will be held daily/weekly with prior notice and members should provide progress updates.
* Members should communicate promptly about any challenges or delays.
* Any modifications to scope, schedule, or deliverables will be discussed in advance and inform the instructor.
* Members should maintain professional standards of work and collaboration



Name & Signature: \_Leshan Chathuranga Kuruppuarachchi\_\_\_\_

Date: \_\_2025/09/12\_\_\_\_\_

A black text on a white background

AI-generated content may be incorrect.

Name & Signature: \_\_Hyunhee Kim\_\_\_\_

Date: \_\_2025/09/12\_\_\_\_\_

# **Work Date/Hours logs**

Student Name: Leshan Chathuranga Kuruppuarachchi

Table 2: Work log - Leshan Kuruppuarachchi

|  |  |  |
| --- | --- | --- |
| Date | Number of Hours | Description of work done |
| 2025-09-05 | 3 | Initial research on the topic. Refer previous research papers. |
| 2025-09-07 | 4 | Refer previous research papers.  Research for existing web scraping resources |
| 2025-09-08 | 3 | Research for existing web scraping resources |
| 2025-09-09 | 3 | Work on project proposal |
| 2025-09-10 | 2 | Work on project proposal  Set up git repo |
| 2025-09-11 | 3 | Finalizing project proposal |

Student Name: Hyunhee Kim

Table 3: Work log Hyunhee

|  |  |  |
| --- | --- | --- |
| Date | Number of Hours | Description of work done |
| 2025-09-06 | 4 | Initial research on topic and existing systems |
| 2025-09-07 | 3 | Research on existing systems  Search for datasets in kaggle |
| 2025-09-08 | 3 | Work on project proposal document |
| 2025-09-09 | 4 | Research on existing systems  Work on project proposal document |
| 2025-09-10 | 3 | Work on project proposal document |
| 2025-09-11 | 2 | Finalizing project proposal |

# **References**

|  |  |
| --- | --- |
| [1] | N Deepak, M Gowtham, J Mithilesh and A Naresh Kumar, "Personalized Career Recommendation System Based on Customer Segmentation," *Journal of Emerging Technologies and Innovative Research,* 2024. |
| [2] | F. Kainyu, M. Mwadulo and S. Munialo , "A Hybrid Recommender Model for Career Pathway Selection in Competency-based Education," *International Journal of Computer Applications,* 2025. |
| [3] | Y. Gupt, "CareerX: AI-Powered Career Path Recommender System for College Students," *ResearchGate,* 2025. |
| [4] | H. Kumar, U. P. Pandey and P. Kumar, "AI-Based Career Path Recommendation System," [Online]. Available: https://amity.edu/UserFiles/aijem/293Harsh1.0%20(AJCS).pdf. |
| [5] | M. QAMHIEH, H. SAMMANEH and M. N. DEMAIDI, "PCRS: Personalized Career-Path Recommender," *IEEEAccess,* 2020. |