## Introduction

* 1. Background and Context of the research

In today’s competitive job market, university students face increasing challenges when choosing their future career path. The traditional methods of career finder, such as personality test, career questionnaires, or career advisers often fail to understand individual skills and interests, because they are to 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 to students and labour 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. 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) [1] .
* Hybrid stacking ensembles. 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 [2].
* Mobile applications (CareerX). CareerX has demonstrated the integration of NLP and machine learning into mobile applications, with user satisfaction rates of around 90% [3].
* Hybrid filtering approaches. Other studies combine collaborative and content-based methods to generate personalized educational roadmaps [4].
* Hierarchical multi-tiered systems. These systems address sparsity and cold-start problems, achieving over 99% accuracy across diverse career dataset [5].

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 adaption.

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

This 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.

* 1. Methodology

It 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
   1. Training Data: Public datasets from Kaggle or LinkedIn will be used to train the model. These datasets include features such as major, GPA, gender, interests, and skills.
   2. Operational Data: Datasets provided by users through the Pathfinder website, such as major, GPA, gender, interest, and skills will be used for real recommendations.
   3. Preprocessing:
      * ~~Text data, such as input by users will be vectorized using TF-IDF (Term Frequency – Inverse Document Frequency) or sentence embeddings (Sentence-BERT).~~
      * ~~Numerical data (e.g., GPA) will be standardized and normalized.~~
      * ~~Skills will be mapped to standard skill categories to handle duplicates and variations~~
2. Model Training and Recommendation Engine Design

To enhance both accuracy and personalization, this research will use a two-stage recommendation pipeline

* 1. Retrieval
     + Two-Tower model (user + job) will embed user features and job features into a shared vector space with similarity measured by cosine distance.
     + For the initial version, Approximate nearest neighbour search will be used to efficiently retrieve top 50-100 job candidates
  2. Re-ranking
     + Candidates from the retrieval stage will be re-ranked using a gradient boosting ranking model
     + Cross features such as skills, major-job compatibility and GPA-job gap relations will be included.
  3. Lastly, the system will recommend the top 3-5 most suitable careers to the user.

1. Skill Gap Analysis

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

* 1. Missing skills will be presented on the dashboard, and each skill will be weighted by current weight and required weight
  2. A gap score will be calculated to estimate how much effort is needed to complete them.

1. Job Market Trend Visualization

Job market dataset will be analyzed to highlight the demand for recommended careers in the last 5 years.

1. System Implementation (Frontend, Backend, Database)
   1. Frontend (React/Vue.js): Input forms for user data, results page showing top career recommendations, skill gap analysis, and job trend dashboards.
   2. Backend: Python (FastAPI): Exposes APIs for model inference.
   3. Database (MySQL/MongoDB)
      * Collect and store ‘Users’, ‘Jobs’, ‘Recommendations’, ‘Feedback’
2. Evaluation and Validation
   1. Evaluation: Cross-validation will be used to measure Accuracy, Precision, Recall and F1-score for quantitative evaluation.
   2. Experimental Design: A/B testing will compare the proposed system against baseline approaches, while user studies will assess the effectiveness of the skill gap analysis and market trend visualization.
   3. Technologies to be used

This project will integrate several technologies

1. Operating System / Platform: The system will run on Windows or macOS environments.
2. Programming Language / Frameworks: Python will be the primary language, using machine learning libraries such as scikit-learn, TensorFlow, and PyTorch. FastAPI will be used to implement the backend APIs.
3. Database: MySQL or MongoDB will be used to store user profiles, job dataset, recommendation result.
4. 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.
5. Backend: FastAPI or Django will serve as the backend to manage model inference skill gap analysis, and job trend visualization.
6. 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.
7. Collaborative Development: GitHub will be used for version control, source code management, and team collaboration. Team members will work together using branches, pull request and reviews during development.  
   1. 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 labour 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.

Reference:

[1] JETIR, “Personalized Career Recommendation System,” JETIR2403A44, 2024. [Online]. Available: [https://www.jetir.org/papers/JETIR2403A44.pdf](https://www.jetir.org/papers/JETIR2403A44.pdf?utm_source=chatgpt.com)

[2] P. Kainyu, “Hybrid Stacking Ensemble for Career Recommendation,” International Journal of Computer Applications, vol. 187, no. 2, 2025. [Online]. Available: [https://www.ijcaonline.org/archives/volume187/number2/kainyu-2025-ijca-924808.pdf](https://www.ijcaonline.org/archives/volume187/number2/kainyu-2025-ijca-924808.pdf?utm_source=chatgpt.com)

[3] CareerX, “AI-Powered Career Path Recommender System for College Students,” ResearchGate, 2024. [Online]. Available: [https://www.researchgate.net/publication/390953768\_CareerX\_AI-Powered\_Career\_Path\_Recommender\_System\_for\_College\_Students](https://www.researchgate.net/publication/390953768_CareerX_AI-Powered_Career_Path_Recommender_System_for_College_Students?utm_source=chatgpt.com)

[4] H. Harsh, “AI-Based Career Path Recommendation System,” Amity Journal of Computational Sciences, 2021. [Online]. Available: [https://amity.edu/UserFiles/aijem/293Harsh1.0%20(AJCS).pdf](https://amity.edu/UserFiles/aijem/293Harsh1.0%20(AJCS).pdf?utm_source=chatgpt.com)

[5] IEEE, “Hierarchical and Multi-Tiered Personalized Career Recommender System,” IEEE Xplore, 2020. [Online]. Available: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9268112>