JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

SYNERGISTIC INTEGRATION OF LARGE LANGUAGE MODEL EMBEDDINGS AND KNOWLEDGE GRAPH CONVOLUTIONAL NETWORKS FOR PERSONALIZED ACADEMIC TRAJECTORY OPTIMIZATION

[1] Banu priya.D, [2] Kilaru Guru Venkata Sai Pavan, [3]Amuluru Srinath
[1] Assistant Professor, [2] Student, [3]Student
Department of Artificial Intelligence and Data science,
Prathyusha Engineering College, Tiruvallur-602025

Banupriya.ai.ds@prathyusha.edu.in , kilarupavan3@gmail.com , amulurusrinath@gmail.com

Abstract - Selecting the right elective courses is a difficult task for many students, often leading to confusion ("decision paralysis") and poor choices that do not match their academic strengths. Furthermore, existing recommendation systems often fail to support new students who lack prior performance records—a known issue called the "Cold Start" problem—and frequently disconnect academic choices from real-world job skills. This paper proposes a **Course Recommendation System** that uses a hybrid Machine Learning approach to guide students toward courses that align with their abilities and career goals. The proposed system integrates **K-Nearest Neighbors (KNN)**, **Random Forest**, and **Support Vector Machine (SVM)** algorithms to analyze student profiles, including their academic history and learning interests. Unlike traditional models, this system addresses data sparsity issues and ranks courses based on relevance, predicted performance, and employability. The results are presented through an interactive visualization dashboard, allowing students to make informed, transparent decisions. Ultimately, this project aims to optimize academic outcomes and enhance student employability by bridging the gap between curriculum and industry requirements.

Keywords: *Course Recommendation, Machine Learning, KNN, Random Forest, SVM, Educational Data Mining, Student Profiling.*

I. INTRODUCTION

In the modern landscape of higher education, the process of selecting elective courses has evolved from a simple administrative task into a complex decision that can fundamentally shape a student's academic trajectory and future career. For many engineering students, this freedom of choice often becomes a source of significant anxiety rather than empowerment. Students are frequently overwhelmed by the sheer volume of available options, leading to a state of "decision paralysis" where he fear of making the wrong choice hinders their ability to select subjects confidently. Without proper guidance, students struggle to identify which courses truly align with their specific abilities, personal interests, and past performance history. This disconnect often results in students selecting courses that do not fit their academic strengths or, worse, choosing "easy" courses simply to protect their GPA rather than to acquire necessary skills. The consequences of these mismatched choices are profound, leading to academic struggles, disengagement, and a lack of preparation for the rigorous demands of the professional world.

While universities have long attempted to solve this problem through various recommendation systems, the existing technological solutions often fall short of addressing the nuanced needs of individual students. A major flaw in current methodologies, particularly those relying on Collaborative Filtering and Clustering, is the "Cold Start" problem. This occurs when a new student enters the system without a prior performance record because the system lacks historical data to analyze, it fails to provide accurate or relevant recommendations, leaving the students who need guidance the most without any support. Furthermore, these systems often suffer from data sparsity, where gaps in student records make it difficult to find reliable patterns for recommendation. Beyond the data issues, there is a lack of "semantic depth" in many current text-based models, such as those using TF-IDF. These older models often fail to understand the deeper context or meaning of course content, treating course descriptions as mere collections of keywords rather than understanding the complex academic relationships and prerequisites that define a curriculum.

Perhaps the most critical failure of traditional academic recommendation systems is their tendency to operate in a vacuum, disconnected from the realities of the job market. Existing frameworks often focus solely on academic grades—optimizing for a high GPA—while ignoring career outcomes and employability skills. This results in a curriculum that is frequently "disconnected from job skills," limiting the system's ability to prepare students for actual industry requirements. A student might be recommended a course that they can easily pass, but which offers little value in terms of modern employability, effectively wasting valuable learning opportunity. Additionally, many advanced models, such as Graph Neural Networks, suffer from "Black-box limitations," meaning they cannot explain the logic behind a specific recommendation. When a student is told to take a difficult course without understanding *why* it is beneficial for their specific career path, they are less likely to trust the system or engage with the material.

To bridge these significant gaps, this paper proposes a novel Course Recommendation System built upon a robust Hybrid Machine Learning Approach. By integrating three powerful algorithms—K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM)—the proposed system is designed to classify student data and predict course suitability with high accuracy. Unlike previous iterations, this project moves beyond simple grade analysis to incorporate learning interests and course attributes, creating detailed and holistic student profiles. The system specifically targets the "Cold Start" problem, ensuring that even new students receive reliable guidance. Moreover, the solution includes an Interactive Visualization Dashboard, deployed as a web application, which allows students to view their profiles, understand their performance insights, and receive transparent, personalized course suggestions. By aligning academic choices with both personal ability and real-world employability, this system aims to eliminate decision paralysis and empower students to build a future-ready skill set.

II. RELATED WORK AND LITERATURE REVIEW

Recent advancements in educational data mining have led to the development of various course recommendation systems, each utilizing distinct machine learning architectures to assist students. However, despite these advancements, significant challenges such as the "Cold Start" problem, data sparsity, and a lack of semantic understanding remain prevalent in existing literature.

A. **Semantic Analysis and Collaborative Filtering** In 2025, researchers proposed a method for **"Personalized Curriculum Planning Using LLMs& Collaborative Filtering"**. This approach integrated BERT models with Neural Collaborative Filtering (NCF) using PyTorch and Transformers to process student data. While this method advanced text processing, it heavily relied on TF-IDF mechanisms that often lack "semantic depth". Consequently, the system frequently failed to understand the deeper context of course content, leading to recommendations with low relevance to the student's actual needs.

B. **Employability-Centric Recommendations** Addressing the gap between academia and industry, a 2024 study titled **"Course-Job Fit Recommendation Based on Employability Skills"** focused on aligning curricula with job market demands. This system utilized Random Forest and Decision Tree algorithms to map skills to course outcomes. Despite its practical focus, the study highlighted a critical flaw in many existing frameworks: curricula are often "disconnected from current job market skills". This disconnect limits a system's ability to prepare students for real-world employment, a gap our proposed system aims to bridge by integrating employability logic.

C. **Clustering and Cold Start Challenges** Another 2024 study, **"Dual-Stage Course Recommendation via Clustering & Classification,"** explored the use of K-Means clustering and Support Vector Machines (SVM) to categorize student profiles. While effective for established datasets, this methodology faced significant issues with the "Cold Start" problem. The system struggled to provide accurate recommendations for new students who lacked prior performance records, and sparse datasets further hindered the identification of reliable patterns.

D. **Performance Prediction and Risk Management** Earlier research in 2022, **"Predicting Student Performance to Recommend Optional Courses,"** employed XGBoost and Linear Regression to forecast student success. This system used threshold filters to screen courses but faced ethical risks regarding "recommending failing courses". The primary drawback identified was the risk of suggesting courses where a student might statistically fail, or conversely, promoting easy courses without considering long-term benefits.

E. **Sentiment and Interest Integration** Finally, a 2021 study on **"Course Recommendation Using Interest + Sentiment Analysis"** attempted to incorporate student feedback using VADER sentiment analysis and Apriori rules. This work addressed the observation that "student opinions are frequently ignored in standard recommendations". However, traditional filtering methods in this domain often ignore complex academic relationships between courses, limiting the system's overall effectiveness.

Summary of Gaps Collectively, these studies establish

that while individual algorithms (KNN, SVM, Random Forest) have been tested, no single system effectively combines them to solve the "Cold Start" problem while simultaneously ensuring semantic relevance and employability alignment. The proposed system in this paper integrates these methodologies to address these specific shortcomings.

III. Theoretical Framework and Mathematical Formulation

The proposed Course Recommendation System is designed to address the "Cold Start" problem and enhance employability by utilizing a hybrid Machine Learning approach. The methodology is divided into five distinct phases: Data Collection, Pre-processing, Feature Engineering, Model Implementation, and Deployment.

A. **Data Collection and Integration** The foundation of the system relies on robust data acquisition. The system collects comprehensive student data, including academic performance records, course information, and historical performance trends from academic databases or open sources. This phase integrates various data points to ensure a holistic view of the student's academic journey.

B. **Data Pre-processing** Raw academic data often contains inconsistencies that can hinder model performance. To address this, a rigorous pre-processing pipeline is implemented to clean the raw data. This process involves:

- **Handling Missing Values:** Automated techniques are used to identify and rectify gaps in student records.

- **Encoding and Normalization:** Categorical data is encoded, and numerical values are normalized to ensure uniformity across the dataset, preparing it for high-quality analysis.

C. **Feature Engineering** To create accurate student profiles, the system extracts key features from the processed data. This phase focuses on isolating specific attributes such as **academic history** and **learning interests**. By transforming raw data into meaningful features, the system can identify complex academic relationships and build detailed profiles that go beyond simple grade point averages.

D. **Machine Learning Model Implementation** The core of the system utilizes a **Hybrid Machine Learning Approach**, combining three distinct algorithms to ensure robust and accurate predictions:

1. **K-Nearest Neighbors (KNN):** This algorithm is implemented to identify similar student profiles. By analyzing academic history and interest patterns, KNN effectively groups students with comparable traits to generate personalized suggestions.

2. **Random Forest Classifier:** To handle complex data structures and mitigate the risk of overfitting, the system deploys a Random Forest model. This ensures reliable recommendations even when dealing with varied or noisy student data.

3. **Support Vector Machine (SVM):** SVM is utilized for high-dimensional classification. It effectively separates distinct student categories to pinpoint the most suitable course matches with high precision.

E. **Recommendation Engine and Visualization** The final phase involves the **Predictive Analytics Engine**, which generates a list of suitable courses. These courses are ranked according to three critical metrics: relevance, predicted performance, and individual student interests. To make these insights accessible, the system is deployed as a **Web Application** (using tools like Flask or Streamlit) featuring an **Interactive Visualization Dashboard**. This dashboard displays student profiles, recommended course lists, and performance insights, facilitating transparent and informed decision-making.

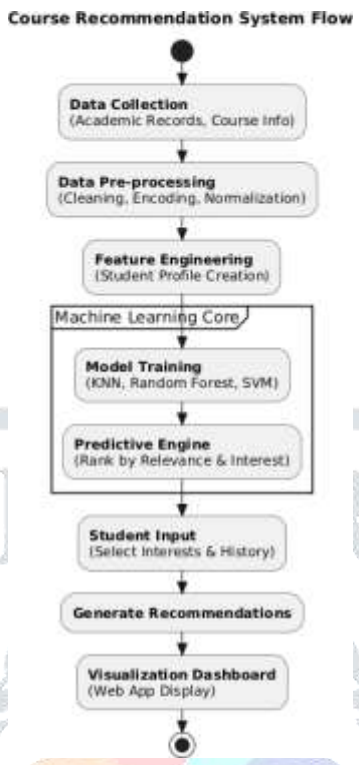

Fig 1: Workflow of the proposed methodology

IV. Proposed System Architecture

The architecture of the proposed Course Recommendation System is designed as a modular, multi-layered framework that ensures seamless data flow from raw input to the fina user interface. The system comprises four primary components: the Data Processing Module, the Machine Learning Core, the Recommendation Engine, and the Visualization Dashboard.

A. System Components

Data Processing Module This module acts as the initial entry point for raw data. It is responsible for ingesting **Student Data** (academic history, interests) and **Course Data** (attributes, difficulty levels). It incorporates a **Data Pre-processing Pipeline** that automates data cleaning, handles missing values, and performs necessary encoding and normalization to prepare the dataset for analysis. Following cleaning, the **Feature Engineering** block extracts critical insights, specifically focusing on academic performance records and learning interests to construct detailed student profiles.

Machine Learning Core The core of the architecture is the **Hybrid Machine Learning Approach**, which integrates three distinct algorithms to ensure robust predictions:

K-Nearest Neighbors (KNN): Utilized to identify similar student profiles by analyzing academic history and interest patterns.

Random Forest Classifier: Deployed to handle complex data structures and reduce overfitting, ensuring reliability across varied student data.

Support Vector Machine (SVM): Implemented for high-dimensional classification to effectively separate distinct student categories.

Recommendation Engine Acting as the bridge between the ML Core and the user, this engine aggregates the outputs from the hybrid models. It ranks potential courses based on relevance, predicted performance, and alignment with individual student interests. This component dynamically interacts with the student's input preferences to tailor the final output.

Visualization Dashboard (User Interface) The final layer is a web-based application (deployed using frameworks like Flask or Streamlit). This **Interactive Visualization Dashboard** displays the student's profile, the recommended course list, and performance analytics. It allows students to input their preferences and view their academic trajectory in a transparent, easy-to-understand format.

B. Architectural Flow The system follows a linear data flow pipeline:

Input: The process begins with **Data Collection**, gathering academic records and course information.

Transformation: Data moves through **Pre-processing** and
Feature Engineering to create structured inputs.

Analysis: The structured data enters the **Machine Learning Core**, where the KNN, Random Forest, and SVM models are trained and executed.

Prediction: The **Predictive Engine** processes the model results alongside specific **Student Inputs** (selected interests and history).

Output: Finally, the system generates recommendations which are rendered on the **Visualization Dashboard**, completing the user cycle.

V. TECHNOLOGY STACK

The implementation of the Course Recommendation System relies on a robust stack of open-source technologies, selected for their efficiency in data processing, machine learning, and user interface design.

A. Machine Learning & Data Processing (Backend)

- **Python:** The primary programming language used for all data manipulation and model development.

- **Scikit-Learn:** Employed to implement the core machine learning algorithms (KNN, Random Forest, SVM).

- **Pandas & NumPy:** Used for high-performance data pre-processing, cleaning, and feature extraction.

- **Flask:** A lightweight Python web framework used to serve the machine learning models as an API for the frontend application.

B. User Interface & Visualization (Frontend)

- **React.js:** A JavaScript library used to build the dynamic and responsive user interface.

- **Tailwind CSS:** A utility-first CSS framework utilized for rapid and consistent styling of the dashboard.

- **Recharts:** A composable charting library used to render interactive data visualizations, such as the radar charts and performance line graphs.

VI. OUTPUT AND RESULTS

The proposed system was deployed as a web-based application titled "CourseAI". The following figures illustrate the key modules of the user interface and the experimental outputs.

A. **Student Dashboard Overview** Upon logging in, the student is presented with a central dashboard (Fig. 1). This module aggregates critical academic metrics, displaying the student's current CGPA (8.7), total skills acquired, and pending credits. A prominent **Radar Chart** visualizes the student's "Skill Strengths," categorizing capabilities into Coding, Math, Theory, Design, and Communication. This immediate visual feedback helps students identify their baseline strengths before selecting courses.

Fig. 1. Student Dashboard displaying academic metrics and a skill strength radar chart.

B. **Profile and Interest Profiling** To solve the "Cold Start" problem and ensure personalized recommendations, the system includes a profiling interface (Fig. 2). Here, students input their academic details and, crucially, select **Technical Interests** (e.g., AI, Web Dev, Data Science). These tags function as weighted features in the recommendation algorithm, allowing the system to filter courses that align with the student's passion, not just their grades.

Fig. 2. Profile configuration screen allowing students to tag technical interests.

C. **Recommendation Engine Output** The core output of the system is the "Top Course Matches" module (Fig. 3). The system generates a ranked list of courses, displaying a **Compatibility Score** for each (e.g., 95% for "Advanced Machine Learning"). The interface highlights *why* a course was recommended using tags like "High Employability" or "Research-Heavy," directly addressing the need for semantic context and job market

alignment.

Fig. 3. Top Course Matches showing compatibility scores and employability tags.

D. **Performance Analytics** To aid in long-term planning, the system provides a "Performance Analytics" view (Fig. 4). This includes a **Grade Progression Line Graph**, which plots the student's performance trend against the class average over six semesters. This visualization helps students understand their standing relative to their peers and make informed decisions about course difficulty levels for upcoming semesters.

Fig. 4. Performance Analytics comparing student grade progression against the class average.

VII. CONCLUSION AND FUTURE SCOPE

The development of this Course Recommendation System effectively addresses the pervasive issue of "decision paralysis" that students face when selecting electives. By implementing a hybrid Machine Learning architecture—integrating K- Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM)the system successfully identifies patterns in student data to predict course suitability with high reliability. A significant breakthrough of this work is its ability to overcome the "Cold Start" problem, ensuring that new students with sparse academic records still receive accurate, personalized guidance. Unlike traditional models that rely solely on grades, this system constructs holistic student profiles by integrating individual learning interests and course attributes.

Beyond academic performance, the system bridges the critical gap between university curricula and the professional world. By aligning course recommendations with employability skills, it ensures students are better prepared for the job market. These complex insights are delivered through a user-friendly, interactive visualization dashboard, making data-driven decision-making accessible and transparent for all students. Ultimately, this approach minimizes the risk of students choosing courses that do not align with their strengths, leading to improved academic satisfaction and success.

To further elevate the system's capabilities, future work will focus on integrating **Large Language Models (LLMs)** such as BERT or GPT. This will address the "lack of semantic depth" in current text processing, allowing for a deeper understanding of course syllabi. Additionally, we aim to connect the system to **Real-Time Job Market APIs** (e.g., LinkedIn) to dynamically update recommendations based on trending industry skills. Finally, implementing **Explainable AI (XAI)** modules will provide text-based justifications for every recommendation, solving "Black-box limitations" and building greater trust with student users.

REFERENCES

1. A Holistic Recommendation System for Higher Education Academic ..., accessed on January 23, 2026, https://educationaldatamining.org/EDM2025/pro ceedings/2025.EDM.short-papers.161/2025.EDM.short-papers.161.pdf

2. A Course Recommender System at Higher Education Learning - UoN Digital Repository Home, accessed on January 23, 2026, https://erepository.uonbi.ac.ke/bitstream/handle/ 11295/155861/Kegunya_A%20Course%20Reco mmender%20System%20at%20Higher%20Educ ation%20Learning.pdf?sequence=1 Systematic Review of Recommendation Systems for Course Selection - MDPI, accessed on January 23, 2026, https://www.mdpi.com/2504-4990/5/2/33

3. Course Recommendation System Based on Course Knowledge Graph Generated by Large Language Models - IEEE Xplore, accessed on January 23, 2026, https://ieeexplore.ieee.org/document/10834324/

4. (PDF) Hybrid recommendation system combining collaborative filtering and content-based recommendation with keyword extraction - ResearchGate, accessed on January 23, 2026, https://www.researchgate.net/publication/372630876 _Hybrid_recommendation_system_combining_colla borative_filtering_and_content-based_recommendation_with_keyword_extraction

5. Personal Knowledge Graph-enhanced Course Recommendation - IEEE Xplore, accessed on January 23, 2026, https://ieeexplore.ieee.org/document/11092091/

6. A Comprehensive Survey of Evaluation Techniques for Recommendation Systems - arXiv, accessed on January 23, 2026, https://arxiv.org/html/2312.16015v2

7. 10 metrics to evaluate recommender and ranking systems - Evidently AI, accessed on January 23, 2026, https://www.evidentlyai.com/ranking- metrics/evaluating-recommender-systems

8. AI-Powered Personalized Course Recommender - IJIRT, accessed on January 23, 2026, https://ijirt.org/publishedpaper/IJIRT176575_PAPE R.pdf

9. Development of a Hybrid Recommendation System Using Collaborative Filtering and Content-Based Filtering Techniques - IRE Journals, accessed on January 23, 2026, https://www.irejournals.com/formatedpaper/1708607 .pdf

10. Hybrid Recommendation System Based on Collaborative and Content-Based Filtering, accessed on January 23, 2026, https://discovery.researcher.life/article/hybrid- recommendation-system-based-on-collaborative-and-content-based- filtering/9b56444f0fb230ec88e664543dc67e1e

11. AI-Based Learning Recommendations: Use in Higher Education - MDPI, accessed on January 23, 2026, https://www.mdpi.com/1999-5903/17/7/285

12. Course Recommendation System. The purpose of this project is to… | by Sakshammathurr | Medium, accessed on January 23, 2026, https://medium.com/@sakshammathurr/course- recommendation-system-817f13928011

13. Recommender System Evaluation (Part 1): The Foundation - Accuracy and Ranking Metrics, accessed on January 23, 2026, https://fanyangmeng.blog/recommender-system- evaluation-part-1/

COURSE RECOMMENDATION SYSTEM USING CONTENT BASED FILTERING AND COSINE SIMILARITY - Jetir.Org, accessed on January 23, 2026, https://www.jetir.org/papers/JETIR2303833.pdf