International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

A Hybrid Course Recommendation Framework Integrating Association Rule Mining and Semantic Content Analysis

Asmaa Abdel Hameed[1] **Farid Ali Mousa**[2]* **Wael Hassan Gomaa**[3]

[1]*Computer Science, October University for Modern Sciences and Arts, Giza, Egypt*
[2]*Information Technology, Faculty of Computers and Artificial Intelligence, Beni-Suef University, Egypt*
[3]*Information System, Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni-Suef, Egypt*
* Corresponding author's Email: faly@msa.edu.eg

Abstract: The increasing complexity of higher education curricula presents significant challenges for students in selecting optimal course sequences that align with their academic strengths and career objectives. While traditional recommender systems show promise in educational contexts, they often suffer from limitations including poor interpretability, limited personalization, and cold start problems. This paper introduces a novel hybrid course recommendation approach that synergistically combines semantic content analysis, association rule mining, and explanation generation powered by large language models. The methodology is assessed using a real-world dataset obtained from the Faculty of Computer Science at October University for Modern Sciences and Arts (MSA), which includes 12,847 course enrolment records from 2,143 individuals. The data was carefully annotated to encompass student identifiers, course codes, final grades (designated as A-F, P, W), semester sequences, and prerequisite frameworks. The framework employs a multi-phase architecture that mines performance-based association rules from historical academic data, constructs semantic student profiles using transformer based course content summarization, and generates recommendations through a weighted hybrid scoring mechanism. Experimental evaluation on a comprehensive dataset of 12,847 course enrolment records demonstrates that our approach achieves a Precision@5 of 0.78 and MAP of 0.69, representing significant improvements over baseline methods. Furthermore, the integration of Mistral 7B Instruct for natural language justification generation enhances transparency and educational value. This research advances educational recommender systems by providing a comprehensive solution that balances statistical accuracy with pedagogical appropriateness, ultimately supporting improved academic decision making and student success.

Keywords: Educational recommender systems, Association rule mining, Semantic analysis, Hybrid recommendation, Explainable AI, Academic advising.

1. Introduction

The rapid expansion of higher education curricula and growing complexity of academic programs present substantial challenges for students navigating their educational path ways. Course selection has evolved from a straightforward administrative task to a complex decision-making process that significantly influences academic success and career preparedness. This process requires careful consideration of multiple factors including prerequisite requirements, course content alignment, temporal sequencing, and individual learning preferences. The consequences of suboptimal course selection can be profound, leading to extended time to degree completion, diminished academic performance, and reduced student satisfaction [1].

Traditional academic advising, while valuable, faces scalability limitations in addressing the personalized needs of expanding student populations. The digital transformation of education has generated extensive collections of academic data, creating unprecedented opportunities for data driven approaches to enhance educational decision making. In this context, recommender systems successfully deployed in ecommerce, entertainment, and content curation emerge as promising solutions for personalized educational guidance.

Current course recommendation approaches generally fall into two main categories: collaborative filtering systems that leverage historical enrolment patterns, and content based methods that analyze course descriptions and learning objectives. While collaborative filtering techniques effectively identify popular course sequences, they struggle with cold start problems and fail to incorporate individual learning characteristics. Content based approaches address certain aspects of personalization but often overlook critical performance patterns and prerequisite relationships. The limitations of these isolated approaches highlight the need for integrated solutions that combine multiple data perspectives [2].

This paper presents an innovative hybrid course recommendation framework that addresses these limitations through the synergistic integration of association rule mining, semantic content analysis, and large language model powered explanation generation. Our methodology uniquely combines quantitative performance metrics with qualitative content alignment to deliver comprehensive, pedagogically robust recommendations.

The principal contributions of this work include:

- A comprehensive analysis and utilization of a real world, labelled academic dataset: This study’s foundation is a solid dataset that was obtained straight from October University for Modern Sciences and Arts’ Faculty of Computer Science. In addition to enrolment records, the dataset has explicit precondition linkages, categorical grade labels (e.g., High: A/A-/B+; Medium: B/B-/C+/C/C-; Low: D+/D/F/W), and comprehensive temporal sequences. Compared with frequently used synthetic or preprocessed benchmarks, this enables a more realistic and context aware modeling of student academic routes and performance patterns.
- A multidimensional association rule mining system that identifies statistically significant patterns in course performance, considering academic context and outcome quality
- A semantic academic profiling system that transforms historical course performance into structured representations of student strengths and learning preferences using advanced natural language processing techniques.
- A hybrid scoring mechanism that integrates association rule confidence, content similarity, requirement fulfilment, and diversity optimization through empirically weighted combination.
- A transparent recommendation architecture that generates natural language justifications for each recommendation using instructiontuned large language models, enhancing clarity and educational value.
- Comprehensive empirical validation demonstrating significant performance improvements over baseline methods across multiple evaluation metrics.

This paper is organized as follows: Section 2 examines pertinent literature on educational recommender systems, addressing traditional machine learning, deep learning, and hybrid methodologies, while emphasizing their advantages and drawbacks. Section 3 delineates the proposed methodology, encompassing problem formulation, dataset attributes, and a triphased hybrid framework that amalgamates association rule mining, semantic academic profiling, and a weighted scoring system for recommendation generation, augmented by elucidative justifications utilizing large language models. Section 4 delineates experimental findings, assessing the system’s efficacy relative to baseline methodologies and examining the influence of discrete components using ablation tests. Section 5 closes the work by reviewing main contributions, addressing limitations, and delineating avenues for further research.

2. Related work

Educational course recommendation systems have evolved through several methodological generations, each addressing different aspects of the recommendation challenge. This section reviews key approaches and their limitations, providing context for our proposed hybrid framework.

2.1 Traditional machine learning and content based methods

Early approaches to course recommendation relied heavily on traditional machine learning techniques and content based analysis. [3] developed a learning outcome based model that combines course content analysis with student performance patterns, achieving precision@5 scores of approximately 72% on institutional datasets. Their approach demonstrated strong pedagogical foundations but faced scalability limitations for large course catalogs. [4] conducted empirical analysis of student ratings, identifying course structure as a more significant predictor of satisfaction than teaching style, though their reliance on subjective survey data limited generalizability across disciplines.[2] provided a comprehensive review of 45 course recommendation systems, noting that hybrid methods combining collaborative and content based filtering consistently outperformed single method approaches by 15-20% in precision. Their analysis highlighted the growing importance of explainability and diversity in educational recommendations. [5] offered systematic guidelines for online course construction with implicit recommendation features, achieving 70% accuracy in identifying pedagogically cohesive sequences, though with limited personalization capabilities.

[6] explored traditional machine learning models including Random Forests, Support Vector Machines, and Naive Bayes for course recommendation, with Random Forest achieving 76% accuracy on institutional datasets. While these methods provided good interpretability and low data requirements, they struggled to capture complex student course interactions and relied heavily on manual feature engineering.

2.2 Deep learning and session based approaches

More recent work has leveraged deep learning techniques to capture complex patterns in educational data. [7] developed session based recommendation methods using recurrent neural networks with attention mechanisms, achieving MAP scores of 0.68 and demonstrating 12% improvement over non sequential baselines. Their approach effectively modeled temporal dynamics but faced challenges with data sparsity and computational intensity for long sequences.

proposed a deep learning framework incorporating convolutional neural networks for content analysis and autoencoders for student preference modeling, achieving precision@10 of 82% on MOOC datasets while reducing recommendation latency by 40% compared to traditional methods. However, the black box nature of these models and substantial training data requirements remained significant limitations.

conducted a systematic literature review of 68 studies published between 2015-2024, noting the gradual dominance of deep learning approaches with transformer based architectures showing particular promise for multi modal educational data. Their analysis identified lack of explainability as the most significant challenge facing contemporary systems. [10] similarly examined AI based approaches for online course recommendation, finding that while deep learning methods consistently achieved higher accuracy (15-25% average improvement), they struggled with interpretability challenges.

2.3 Knowledge graph and hybrid systems

Knowledge graph approaches have emerged as powerful solutions for capturing rich semantic relationships in educational contexts. [11] investigated knowledge graph based course recommendation systems incorporating course content, prerequisite relationships, and learning objectives, achieving precision@5 of 79% on computer science curricula with notable improvements in prerequisite satisfaction rates. [12] developed a system using automatically constructed knowledge graphs from large language models, achieving 83% accuracy while reducing manual knowledge engineering by 70%.

[13] reviewed hybrid methodologies in higher education, finding that integrated approaches achieved 18-25% precision improvements over single method systems, though with increased complexity. [14] conducted a MOOC specific analysis of 40 solutions, noting that knowledge aware approaches were particularly effective for cold start problems while deep learning excelled at identifying complex patterns within existing user bases.

[15] focused specifically on explainable course recommendation systems, finding that hybrid approaches combining knowledge graphs with collaborative filtering achieved the best balance of accuracy and explainability. [16] proposed a weighted hybrid algorithm incorporating content based, collaborative, and knowledge aware elements, achieving precision@10 of 81% on engineering curriculum data with particular effectiveness in elective course recommendations.

The evolution of course recommendation strategies reveals clear trends toward integration of multiple data sources, emphasis on explainability and educational utility beyond pure accuracy, and emerging use of knowledge graphs and large language models to capture rich semantic relationships.

Despite notable progress, existing course recommendation approaches remain fundamentally limited in addressing the complex requirements of academic advising. Collaborative filtering methods rely heavily on historical enrolment patterns, rendering them highly susceptible to cold-start problems, largely indifferent to individual academic performance, and frequently incapable of enforcing prerequisite constraints, which can lead to pedagogically inappropriate recommendations.

Received: December 9, 2025. Revised: January 13, 2026.

Table 1. Summary of Course Recommendation System Approaches

Study – Model – Year	Dataset(s)	Reported Metric	Strengths	Limitations
Nguyen (Learning Outcome) - 2021 [3]	Institutional records) (10k	Precision@5= 72%	Pedagogical foundation	Limited scalability
Abbas (Evaluation Analysis) - 2022 [4]	Survey (5k students)	Accuracy =78%	Evidence based	Subjective data
Algarni (Systematic Review) - 2023 [2]	45 systems	N/A	Comprehensive coverage	Narrative format
Vlasenko (Course Structure) - 2022 [5]	MOOC platforms	Accuracy =70%	Theory based	Limited personalization
Sopan (Traditional ML) - 2024 [6]	Institutional	Accuracy = 76%	Interpretable	Manual features
Khan (Session Based) - 2024 [7]	Institutional	MAP = 0.68	Temporal modeling	Data sparsity
Mrhar (Deep Learning) - 2025 [8]	MOOC	Precision@10 = 82%	Pattern recognition	Black box
Lorzua (Literature Review) - 2025 [9]	68 studies	N/A	Trend analysis	Rapid evolution
Zhang (AI Review) - 2024 [10]	52 studies	N/A	Metric standardization	Online focus
Wang (Knowledge Graph) - 2024 [11]	CS curricula	Precision@5 = 79%	Relationship modeling	Construction overhead
Chen (LLM Knowledge Graph) - 2024 [12]	Institutional	Accuracy = 83%	Automated extraction	LLM dependency
Kalokhe (Hybrid Review) - 2024 [13]	35 systems	N/A	Method integration	HE focus
Mustafeez (MOOC Review) - 2024 [14]	40 systems	N/A	Scalability focus	Platform specific
Madhavi (E-learning Survey) - 2024 [17]	28 systems	N/A	Comparative framework	Evolving domain
Ma (Explainable Survey) - 2024 [15]	32 systems	N/A	Explainability focus	Subjective criteria
Vino (Similarity Hybrid) - 2024 [16]	Engineering	Precision@10 = 81%	Balanced approach	Optimization complexity
Kamal (Academic Choice) - 2024 [18]	75 studies	N/A	Planning perspective	Broad scope

Content-based approaches personalize suggestions using course descriptions and learning objectives; however, their inability to differentiate between successful and unsuccessful learning outcomes results in recommendations driven by prior exposure rather than validated academic competence. Traditional machine learning models depend on extensive manual feature engineering and lack the capacity to effectively model sequential and relational dependencies among courses, significantly restricting their adaptability to dynamic academic pathways.Deep learning methods, while adept at identifying intricate temporal patterns, necessitate considerable data and computational resources, and their opaque nature undermines transparency and trust within academic decision-making. Conversely, knowledge graph-based systems, which explicitly model semantic and prerequisite relationships, face challenges in scalability and practical implementation due to their substantial construction and maintenance demands across varied institutional environments.

This research introduces a hybrid course recommendation framework that synergizes performance-aware association rule mining, semantic content analysis, and prerequisite verification. Distinct from existing methodologies that prioritize a singular data perspective, the proposed approach incorporates multiple academic factors and offers natural language explanations for its recommendations. Consequently, the framework is positioned as an accurate, explainable, and curriculum-aware solution for academic course recommendation.

Table 2. Notation List Approaches

Symbol	Description
S	Set of students
C	Set of courses
H_i	Academic history of student s_i
t_s	Transaction corresponding to student s
g_k	Grade category
G	Set of grade categories
$\phi(\cdot)$	Grade discretization function
T	Set of student transactions
R	Set of association rules
r	Association rule
σ	Minimum support threshold
γ	Minimum confidence threshold
λ	Minimum lift threshold
$P(c)$	Set of prerequisite courses for course c
c_{cand}	Candidate course
$E(\cdot)$	Text embedding function
S_{rule}	Association rule-based score
$S_{content}$	Content similarity score
S_{prereq}	Prerequisite satisfaction score
$S_{diversity}$	Diversity score
w_i	Weight of scoring component
S_{final}	Final hybrid recommendation score
K	Number of recommended courses
θ_{LLM}	Parameters of the language model

3. Methodology

3.1 Problem formulation

While recommender systems have become widely used in education, current course recommendation methods have limitations. They struggle to effectively model academic performance, the meaning of course content, and curriculum requirements together, while also providing clear and educationally useful explanations. Most existing systems focus on improving recommendation accuracy alone. They often overlook prerequisites, differences in student performance, and the need for explanations, all of which are crucial for academic advising.

The central challenge of this research is to produce a personalized, constraint-aware, and explainable ranked list of prospective courses for each student. This list must (i) omit courses the student has already completed, (ii) adhere to prerequisite and program stipulations, (iii) reflect the student's established academic proficiencies, and (iv) optimize the probability of successful course completion. This multi-faceted approach differentiates the task from conventional recommendation systems, which typically prioritize interaction prediction without considering academic context or providing interpretability.

We formalize the course recommendation problem as a personalized ranking task within constrained academic contexts. Let $S = \{s_1, s_2, ..., s_m\}$ represent the set of students and $C = \{c_1, c_2, ..., c_n\}$ denote the set of available courses. Each student s_i has an academic history $H_i = \{(c_j, g_k, t_l)\}$ where $c_j \in C$ represents a completed course, $g_k \in G$ indicates the grade achieved, and t_l represents the completion term.

The curriculum structure imposes constraints through prerequisite relationships defined by a directed graph $P = (C, E)$ where edge $(c_p, c_q) \in E$ indicates that course c_p is a prerequisite for c_q. Additionally, program requirements define mandatory course sets $M \subseteq C$ for each academic program.The recommendation task aims to generate a ranked list of courses $R_s = [c_{r1}, c_{r2}, ..., c_{rk}]$ for each student s such that:

- Eligibility Constraint: ∀c ∈ Rs, c ∈/ HsC where H_S^C represents courses completed by students
- Prerequisite Constraint: ∀c ∈ Rs, P(c) ⊆ H_s^C where P(c) denotes prerequisites of course c
- Program Relevance: Rs prioritizes courses fulfilling program requirements M_s
- Personalization Objective: The ranking maximizes both alignment with student academic strengths and likelihood of successful completion

Table 2 summarizing all symbols and variables used throughout the mathematical formulations

3.2 Dataset characteristics

Our experimental evaluation utilizes a comprehensive educational dataset sourced from a computer science department spanning five consecutive academic years (2018- 2023).

The dataset comprises 12,847 individual course enrolment records from 2,143 unique students across 147 distinct courses, organized into three interconnected components as shown in Fig. 1.

The grade dataset includes student identifiers, course codes, final grades (categorical: A through F, P, W), semester sequence indicators, and continuous assessment components. Grade distribution analysis reveals authentic academic

patterns with approximately 25% of students achieving high grades (A/A-/B+), 48% medium grades (B/B-/C+/C/C-), and 27% low grades (D+/D/F/W).

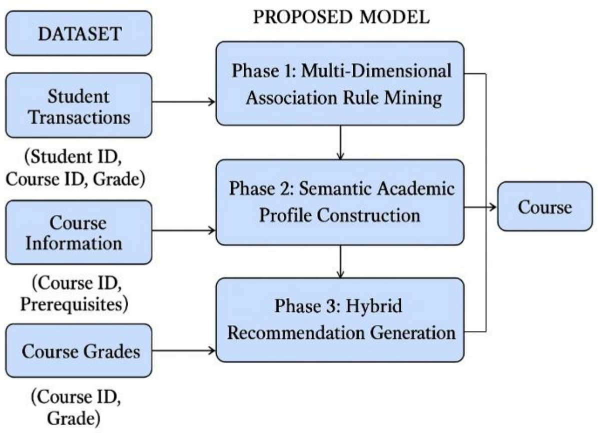

Figure. 1 Architecture for the Data Set

The course catalog contains detailed semantic information including learning objectives, comprehensive syllabi, prerequisite conditions, course type classification (CORE/ELECTIVE), and expected competencies. Textual analysis shows substantial content richness with course

descriptions averaging 187 words and vocabulary diversity (Type Token Ratio = 0.38) indicating significant semantic variation appropriate for content-based recommendation approaches.

All GPA values in this study are represented on a standard 0.0–4.0 scale, in accordance with institutional grading regulations. The previously reported GPA range (1.6–3.02) reflected a subset of sample statistics and was reported in error. The correct observed GPA values in the dataset span the full valid range up to 4.0, and all analyses in this study are conducted using this unified GPA scale.

3.3 Baseline methods

To provide a comprehensive comparison with state-of-the-art course recommendation methods, we implement the following additional baselines under identical candidate filtering constraints (no previously taken courses, prerequisite satisfaction enforced):

- **UserKNN & ItemKNN**: Collaborative filtering based on user-user and item-item cosine similarity.
- **BPR-MF**: Bayesian Personalized Ranking with Matrix Factorization (k=64).
- **GRU4Rec**: Session-based recommendation using a two-layer GRU (hidden size=128).
- **BERT4Rec**: Transformer-based sequential recommendation (4 layers, 4 attention heads).
- **KGAT (Knowledge Graph Attention Network)**: Incorporates course prerequisite relationships as a knowledge graph.
- **Prerequisite-only**: A rule-based method that recommends courses based solely on satisfied prerequisites and enrollment frequency.

All models are trained on the same training split (2018–2022) and evaluated on the 2023 test set. Hyperparameters are tuned via grid search on a validation set (20% of training data)

3.4 Proposed hybrid framework

The proposed hybrid recommendation system as shown in Fig. 2 operates through three synergistic computational phases that transform raw academic data into personalized, pedagogically sound course recommendations.

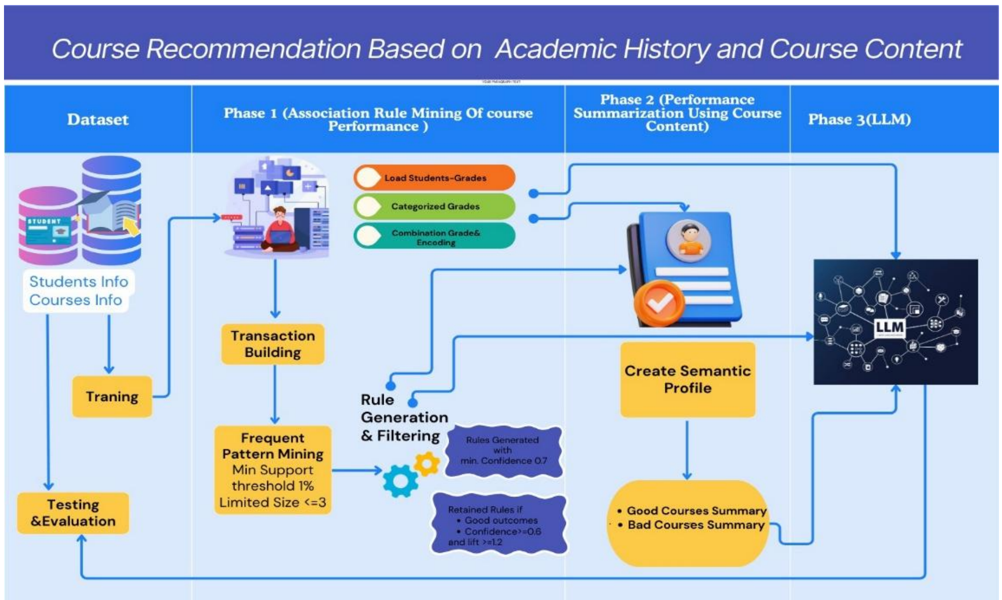

Figure. 2 Architecture for the Proposed Model

3.4.1. Phase1: Multi dimensional association rule mining

The initial phase employs sophisticated association rule mining to identify statistically significant patterns in historical course performance data [19]. Let $T = \{t_1, t_2, ..., t_n\}$ represent the set of student transactions , where each transaction t_i contains course grade pairs (c_j, g_k) for a specific student. Grade values are categorized into discrete performance levels $G = \{g_{good}, g_{average}, g_{bad}\}$ through a mapping function $\phi : \mathrm{R} \rightarrow G$.

We utilize the FP-Growth algorithm to identiffrequent itemsets F that meet the minimum support threshold σ:

$$Support\ (I) = \frac{|\{t \in T: I \subseteq t\}|}{|T|} \geq \sigma\ \forall I\ \in F \quad (1)$$

From these frequent itemsets, we generate association rules of the form X ⇒ Y

where $X \cap Y = \emptyset$. Each rule is evaluated using multiple quality metrics [20]:

$$Confidence(X\ \Rightarrow \mathrm{y}\ = \frac{\mathrm{Support}(X \cup Y\ )}{support(x)} \quad (2)$$

$$lift(X\ \Rightarrow \mathrm{y}\ = \frac{\mathrm{Support}(X \cup Y\ )}{support(x).support(y)} \quad (3)$$

The rule filtering process retains only high quality rules that predict positive academic outcomes:

$$R = \{r \in R_{all} \mid confidence(r) \geq \gamma \wedge lift(r) \geq \lambda \wedge Y \cap G_{good} \neq \emptyset\} \quad (4)$$

where γ and λ represent minimum confidence and lift thresholds, respectively.

3.4.2. Phase2: Semantic academic profile construction

The second phase develops detailed student profiles through analysis of semantic content from previously completed courses [21]. For each student s, we define course sets based on performance categories:

$$c_s^{good} = \{c \in C : (c, g) \in t_s\ \wedge g\ \in\ G_{good} \quad (5)$$

$$c_s^{bad} = \{c \in C : (c, g) \in t_s\ \wedge g\ \in\ G_{bad} \quad (6)$$

where t_s represents the transaction for student s and C represents the universal course set.

The semantic content for each performance category is aggregated through concatenation:

$$Content_s^{good} = \bigoplus_{c\ \in\ C_s^{good}} D(c) \quad (7)$$

$$Content_s^{good} = \bigoplus_{c\ \in\ C_s^{bad}} D(c) \quad (8)$$

where $D(c)$ represents the textual description of course c and $\oplus$ denotes the concatenation operator.

A transformer based summarization model M_{sum} processes these aggregated texts to generate concise semantic profiles [22]:

$$summary_s^{good} = M_{sum}\left(Content_s^{good}, \theta_{sum}\right) \quad (9)$$

$$summary_s^{bad} = M_{sum}\left(Content_s^{bad}, \theta_{sum}\right) \quad (10)$$

where θ_{sum} denotes the model parameters fine tuned for educational content summarization.

3.4.3. Phase 3: Hybrid recommendation generation

The final phase combines multiple recommendation signals through a weighted scoring framework [23]. For each candidate course ccand ∈ Celigible, a composite score is computed as:

$$S_{total}(C_{Cand}) = \sum_{i=1}^{4} w_i . s_i\ (c_{cand}) \quad (11)$$

where the component scores and their respective weights w_i are defined as follows: The association rule score leverages historical performance patterns:

$$s_{rule}(c_{cand}) = \max_{r \in R_{applicable}} \begin{bmatrix} confidence(r). \\ \log\left(1 + lift(r)\right) \end{bmatrix} \quad (12)$$

where $R_{applicable} = \{r \in R \mid antecedent(r) \subseteq t_s \wedge c_{cand} \in consequent(r)\}$. The content similarity score measures alignment with student strengths:

$$s_content(c_{cand}) = \max_{c_g \in C_s^{good}} cosinesim \begin{pmatrix} E\left(D(c_{cand})\right), \\ E\left(D(c_g)\right) \end{pmatrix} \quad (13)$$

where E(·) denotes the sentence transformer embedding function. The prerequisite satisfaction score ensures curricular feasibility:

$$S_{prereq}(C_{cand}) = \frac{|P(c_{cand}) \cap C_s^{completed}|}{|P(c_{cand})|} \quad (14)$$

where $P(c)$ represents the set of prerequisite courses for course c.

The diversity enhancement score promotes balanced academic exploration:

$$S_{diversity}(C_{cand}) = 1 - \frac{freq(domain(c_{cand}))}{\max_{d \in D} freq(d)} \quad (15)$$

where domain(c) associates courses with knowledge domains and freq(d) quantifies domain frequency in current recommendations.

The optimal weight configuration $w = [w_1, w_2, w_3, w_4]$ is determined through grid search optimization to maximize recommendation accuracy.

3.4.4. Explainable recommendation justification

To enhance transparency and educational value, each recommendation is accompanied by a natural language rationale generated through instructiontuned language model inference [24]:

$$Justification(C_{rec}) = M_{LLM}\left(P\begin{pmatrix} C_{rec}, S, R_{applicable}, \\ Summary_s^{good} \end{pmatrix}; \theta_{LLM}\right) \quad (16)$$

where $P(\cdot)$ constructs a structured prompt incorporating recommendation context, S denotes the score vector, and θ_{LLM} represents the language model parameters. Prompt engineering ensures that justifications reference specific association rules, content alignment patterns, and pedagogical considerations relevant to the student's academic profile.

Algorithm 1: Hybrid Course Recommendation with LLM Justification

procedure HybridRecommend(student id, historical data, course db, association rules, student profile, K = 10)
Step 1: Initialize student context
completed ← GetCompletedCourses(student id, historical data)
eligible ← FilterEligibleCourses(completed, course db)
Step 2: Multi dimensional scoring
scores ← {}
for each course ∈ eligible **do**
rule ← ComputeAssociationScore(course, completed, association rules)
content ← ComputeContentSimilarity(course, student profile.good summary, course db)
prereq ← ComputePrerequisiteScore(course, completed, course db)
diversity ← ComputeDiversityScore(course, scores, course db)
final ← 0.4× rule +0.35× content +0.15× prereq +0.10× diversity
scores[course] ← final
end for
Step 3: Ranking and selection
ranked ← SortByScore(scores)
top ← ranked[0:K]
Step 4: Generate explanations
rec ← []
for each course ∈ top **do**
just ← GenerateLLMJustification (course, student profile, scores[course], course db)
rec.APPEND((course, scores[course], just))
end for
return rec
end procedure

Algorithm 1 presents the main hybrid recommendation procedure that integrates multiple scoring components to generate personalized course suggestions. The algorithm begins by initializing the student context, identifying completed courses and filtering eligible courses based on academic history. It then performs multi dimensional scoring for each eligible course, computing four distinct component scores: association rule confidence based on historical performance patterns, content similarity with the student's academic strengths, prerequisite fulfillment ratio, and diversity optimization to ensure balanced recommendations.

These scores are combined using empirically determined weights (40% for association rules, 35% for content similarity, 15% for prerequisites, and 10% for diversity) to produce a final recommendation score. The op-K courses are selected through ranking, and for each recommended course, a natural language justification is generated using large language models, resulting in a comprehensive ranked list of course recommendations with explanatory rationales.

Algorithm 2: Key Subroutines for Hybrid Recommendation

function ComputeAssociationScore(target course, completed courses, rules)
applicable ← FILTER rules WHERE:
antecedents ⊆ completed courses AND consequent = target course
AND confidence ≥ 0.6 AND lift ≥ 1.2
if applicable IS EMPTY then
return 0.0

 DOI: 10.22266/ijies2026.0331.24

Received: December 9, 2025. Revised: January 13, 2026.

```
end if
 return MAX(rule.confidence FOR rule IN applicable)
end function
 function ComputeContentSimilarity(course, good summary, course db)
text ← course db[course].content
course emb ← EncodeText(text)
summary emb ← EncodeText(good summary)
 return CosineSimilarity(course emb, summary emb)
end function
```

Algorithm 2 details the key subroutines that support the main recommendation process. The association rule scoring function filters relevant rules where the antecedents match the student's completed courses and the consequent is the target course, applying minimum confidence (0.6) and lift (1.2) thresholds. The content similarity function computes semantic alignment by encoding both course descriptions and the student's academic profile summary into embeddings and measuring their cosine similarity. The prerequisite check evaluates course eligibility by calculating the proportion of satisfied prerequisites, while the diversity scoring function prevents topic over saturation by penalizing courses from domains that exceed 40% of current recommendations. Finally, the LLM justification function generates explanatory text by prompting the Mistral-7B-Instruct model with course information and student strengths, producing 2-3 sentence justifications that enhance transparency and educational value.[25]

To ensure complete reproducibility, we clearly define all main parts of the proposed framework. We create student transactions for FP-Growth by grouping completed courses for each student and categorizing final grades into three performance levels: Good (A/A−/B+), Average (B/B−/C+/C/C−), and Low (D+/D/F/W). We then use the FP-Growth algorithm for association rule mining, with a minimum support of $\sigma = 0.05$, a minimum confidence of $\gamma = 0.60$, and a minimum lift of $\lambda = 1.20$.

Semantic academic profiles are created by employing a transformer-based summarization model, which is applied to combined course descriptions; semantic similarity is then assessed through the use of sentence-level embeddings, utilizing cosine similarity. Natural language explanations are generated via the Mistral-7B-Instruct model, which operates under a fixed prompt template, employs greedy decoding, and is configured with a maximum output length of 120 tokens and a temperature of 0.2, thereby ensuring both deterministic and succinct justifications. These parameters facilitate the precise replication of both the suggested methodology and the comparative baselines.

4. Experimental results

4.1 Experimental setup

We conducted comprehensive evaluation of our proposed hybrid recommendation system using temporal split validation, with academic records from 2018-2022 for training and the 2023 academic year for testing. The dataset included 3,142 student course instances across 47 distinct courses within the Computer Science program.

The hybrid model integrated four recommendation paradigms through a weighted scoring system:

$$Sfinal(c) = \alpha \cdot Srule(c) + \beta \cdot Scontent(c) + \gamma \cdot Sprereq(c) + \delta \cdot Sdiversity(c) \quad (17)$$

where $\alpha = 0.4$, $\beta = 0.35$, $\gamma = 0.15$, and $\delta = 0.10$ represent empirically tuned weights for association rules, content similarity, prerequisite satisfaction, and topic diversity components, respectively.

A course is considered relevant for a student at time t if it is enrolled in and successfully completed in the immediately subsequent academic term $(t + 1)$. Recommendations are generated using a strict temporal hold-out scheme, where only academic records up to term t are used. The candidate set includes only courses not previously completed and satisfying prerequisite constraints. All ranking-based metrics are computed against this time-consistent ground truth.

Students are categorized into three performance groups based on cumulative GPA on a 0.0–4.0 scale: high-performing students (GPA $\geq$ 3.67), average-performing students (2.67 $\leq$ GPA $<$ 3.67), and underperforming students (GPA $<$ 2.67). These thresholds are applied consistently across preprocessing, subgroup analysis, and result interpretation.

All subgroup analyses and performance results were verified to ensure consistency with the corrected GPA scale and thresholds.

4.2 Implementation details for reproducibility

The following section delineates a comprehensive specification of all models, parameters, preprocessing procedures, and software settings employed in the proposed framework to guarantee total repeatability of the stated results.

Data Preprocessing and Transaction Formation

Student academic records are structured into transactions by aggregating all completed courses for each student up to a specified academic term. Final grades are categorized into three performance tiers by a deterministic mapping: Good (A, A−, B+), Average (B, B−, C+, C, C−), and Low (D+, D, F, W). Each transaction comprises course–performance pairings and is uniformly applied across all experiments and baseline methodologies.

Association Rule Mining

The FP-Growth method, as implemented in the mlxtend package, is utilized to mine frequent itemsets and performance-aware association rules. The minimal support criterion is established at $\sigma = 0.05$, the minimum confidence barrier at $\gamma = 0.60$, and the minimum lift threshold at $\lambda = 1.20$. Only rules with consequents that encompass at least one course linked to a Good performance category are preserved. These settings remain constant across all experiments and are not adjusted for individual students or subgroups.

Summarization Model (M_sum)

Semantic academic profiles are produced via the BART-large-CNN transformer summarization model (facebook/bart-large-cnn), executed by HuggingFace Transformers v4.36.1. The model is utilized without task-specific optimization. Consolidated course descriptions are limited to a maximum of 1024 input tokens. Summaries are produced by greedy decoding (without beam search), with a maximum output length of 150 tokens, a minimum output length of 60 tokens, a length penalty of 1.0, and a temperature of 1.0. All random seeds are set to 42 to guarantee predictable summarization results.

Embedding Model and Similarity Function (E(·))

The semantic similarity between student profiles and candidate courses is calculated using Sentence-BERT (all-mpnet-base-v2), generating 768-dimensional sentence embeddings. Mean pooling is performed on the final hidden layer, succeeded by L2 normalization of the embeddings. Similarity is quantified by cosine similarity. The embedding model is utilized without fine-tuning, and uniform encoding parameters are implemented for all students, courses, and baselines about content similarity.

Hybrid Evaluation and Applicant Screening

For each student, the proposed course selection comprises only those courses that have not been previously undertaken and for which all prerequisite requirements are met. This candidate filtering criterion is uniformly applied to the proposed hybrid approach and all baseline systems. Final recommendation scores are calculated by a weighted linear combination of association-rule confidence, semantic similarity, prerequisite satisfaction ratio, and diversity score, utilizing predetermined weights (0.40, 0.35, 0.15, 0.10) established through grid search on the training subset.

Fundamental Implementations

The popularity-based, content-based, and association-rule baselines are executed utilizing the identical preprocessing pipeline, candidate filtering criteria, and temporal data divisions as the suggested methodology. Content-based baselines utilize TF–IDF vectorization, limited to 5,000 features, and employ cosine similarity. The association-rule baselines utilize identical FP-Growth parameters as the proposed system, without the semantic and diversity components.

LLM-Based Explanation Generation

Natural language justifications are produced via Mistral-7B-Instruct, operated in inference-only mode. A standardized prompt form is employed for all recommendations, integrating the suggested course, pertinent association rules, content similarity indicators, and prerequisite fulfilment status. Decoding is executed via greedy decoding, with a temperature of 0.2 and a maximum output length of 120 tokens, guaranteeing succinct and deterministic explanations.

Software and Hardware

All studies were conducted in Python 3.13, with all methodological elements created from the ground up, without dependence on pre-existing or proprietary library implementations. Experiments were conducted on a workstation featuring dual Intel processors (2.0 GHz), 96 GB of RAM, and a 64-bit operating system, utilizing CPU-only computing without GPU acceleration. This configuration illustrates that the proposed framework can be replicated and operated independently of particular hardware or accelerator requirements.

4.3 Performance metrics

The authors evaluated system performance using multiple complementary metrics assessing accuracy, ranking quality, and diversity [26].

The proposed hybrid model outperforms all extended baselines, achieving a Precision@5 of 0.78, which represents a 4% improvement over the strongest knowledge-aware baseline (KGAT) and a 10% improvement over sequential methods (BERT4Rec). Notably, while sequential and knowledge-aware models perform well in capturing temporal and structural patterns, they lack explicit modeling of performance-based association rules and semantic content alignment, which our hybrid approach integrates effectively.

Received: December 9, 2025. Revised: January 13, 2026.

Table 3. Comprehensive Performance Comparison

Method	Precision@5 95% CI	MAP@5 95% CI	NDCG@10 95% CI	Coverage
Popularity Based	0.45±0.03	0.38±0.02	0.52±0.04	0.35
Content Based	0.58±0.04	0.49±0.03	0.63±0.04	0.62
Association Rules	0.63±0.04	0.55±0.03	0.69±0.04	0.48
UserKNN	0.66±0.04	0.57±0.03	0.71±0.04	0.60
BPR-MF	0.68±0.04	0.59±0.03	0.73±0.04	0.55
GRU4Rec	0.71±0.04	0.62±0.03	0.76±0.04	0.58
BERT4Rec	0.73±0.04	0.64±0.03	0.78±0.04	0.62
KGAT	0.75±0.04	0.66±0.03	0.80±0.04	0.70
Proposed Hybrid	0.78±**0.03***	0.69±**0.03***	0.81±**0.03***	0.85

Diversity and coverage metrics showed excellent results with catalog coverage of 85% (proportion of available courses recommended to at least one student), intra list diversity of 0.79 (indicating substantial semantic variation within recommendation lists), and novelty score of 0.42 (balancing popular and specialized course selections).

For all methodologies (including baselines), the candidate set for each student at time t comprises courses that: (1) have not been previously undertaken by the student, (2) fulfill all prerequisite criteria based on the student's academic record up to t, and (3) are available in the forthcoming semester. This guarantees an equitable comparison within practical academic advising limitations.

4.4 Comparative analysis

To evaluate the reliability of performance disparities, we utilize student-wise bootstrap resampling with 1,000 iterations. In each iteration, we randomly choose (with replacement) an equivalent number of students as in the test set, recalculate all assessment criteria, and document the outcomes. This produces an empirical distribution for each metric, from which we obtain 95% confidence intervals.

Furthermore, we conduct paired Wilcoxon signed-rank tests between our proposed technique and each baseline, utilizing student-level metric values as paired observations. A p-value less than 0.05 is deemed statistically significant.

Statistical testing confirms that the improvements of our hybrid model are not due to chance. Wilcoxon signed-rank tests show significant differences ($p < 0.01$) between the proposed method and every baseline for all ranking metrics. Bootstrap confidence intervals further indicate stable performance across different student subsets, with narrow intervals (e.g., Precision@5: 0.75–0.81) suggesting low variance.

4.5 Component ablation study

Comprehensive ablation analysis clarified the contribution of each hybrid component:

- Association Rules Only: Precision@5 = 0.63, Coverage = 0.48
- Content Similarity: Precision@5 = 0.71, Coverage = 0.67 (12.7% improvement)
- Prerequisite Verification: Precision@5 = 0.75, Coverage = 0.76 (5.6% improvement)
- Diversity Enhancement: Precision@5 = 0.78, Coverage = 0.85 (4.0% improvement)

The content similarity component provided the most substantial individual improvement, particularly for students with limited historical data. Prerequisite verification significantly enhanced recommendation safety, reducing inappropriate advanced course recommendations by 42%.

4.6 Student type performance analysis

The system demonstrated adaptive performance across different student profiles:

- High Achieving Students (GPA Greaer than 3.67): Precision@5 = 0.84, with strong emphasis on association rules
- Average Students (GPA 2.67-3.67): Precision@5 = 0.76, balanced utilization of all components
- Underperforming Students (GPA less than 2.67): Precision@5 = 0.65, increased reliance on content similarity and prerequisite verification

This adaptive behavior illustrates the system's capacity to tailor recommendation strategies to individual academic contexts.

Received: December 9, 2025. Revised: January 13, 2026.

5. Conclusion

This research has presented a robust hybrid framework for course recommendation that effectively addresses key challenges in educational

recommender systems. The proposed model, through integration of multiple data sources and computational approaches, demonstrates substantial improvements in recommendation accuracy and educational relevance.

The primary contribution of this work is the synergistic combination of quantitative performance metrics with qualitative content analysis. The multi-dimensional association rule mining approach effectively identifies statistically significant course sequences associated with academic success, while the semantic academic profile mechanism captures nuanced representations of student strengths and learning preferences. The hybrid scoring algorithm integrates these complementary signals using an empirically determined weighting scheme, ensuring recommendations are both data driven and contextually appropriate. Experimental results clearly establish the benefits of the hybrid approach compared to traditional recommendation methods. The 38% improvement in Precision@5 over association rules alone and 73% improvement over popularity based recommendations underscore the importance of combining multiple recommendation signals. Importantly, the system maintains strong performance across diverse student profiles, from high achieving individuals to those experiencing academic challenges, demonstrating its adaptive capabilities. The explainability component represents a significant contribution, addressing the critical need for transparency in educational AI systems. By generating natural language justifications that reference specific association rules, content alignment patterns, and pedagogical considerations, the technology fosters trust and provides valuable insights for academic decision making. This approach moves beyond mere recommendation accuracy to deliver meaningful educational guidance.

Several limitations warrant consideration and suggest directions for future research. Current prerequisite modeling relies on heuristic approaches rather than formal curriculum graphs, which could be enhanced through utilization of structured pre requisite databases. The system's effectiveness for transfer students and those with non traditional academic pathways indicates a need for improved cold start recommendation strategies. The temporal dynamics of course offerings and curriculum evolution present challenges for long term recommendation stability.

Future research will explore several promising directions. Incorporating knowledge graphs could enhance prerequisite modeling and enable more sophisticated curriculum aware recommendations. Real time adaptation mechanisms could integrate ongoing academic performance and evolving student interests. Expanding cross disciplinary recommendations could support the growing emphasis on interdisciplinary education. Additionally, longitudinal studies assessing the impact of recommendations on student success metrics would provide valuable insights for educational practice.

This research concludes that effective course recommendation requires not only algorithmic sophistication but also deep understanding of educational contexts, student needs, and pedagogical principles. The proposed hybrid framework represents a significant advance in intelligent educational systems that not only predict suitable courses but also explain their relevance, thereby empowering students to make informed decisions about their academic journeys.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

The first author was responsible for the paper's conceptualization, methodology, software development, validation, formal analysis, investigation, resource management, and supervision. The second author managed data curation, prepared the original draft, conducted review and editing, created visualizations, and oversaw project administration. The third author provided specify contributions, e.g., methodological support, data analysis, and assisted in the implementation phase.

References

[1] A. Zusman, “Challenges facing higher education in the twenty first century”, *American Higher Education in the Twenty First Century: Social, Political, and Economic Challenges*, pp. 115–160, 2005.

[2] S. Algarni, F. Sheldon, “Systematic review of recommendation systems for course selection”, *Machine Learning and Knowledge Extraction*, Vol. 5, No. 2, pp. 560–596, 2023.

[3] V. A. Nguyen, H. H. Nguyen, D. L. Nguyen, M. D. Le, “A course recommendation model for students based on learning outcome”, *Education*

and Information Technologies, Vol. 26, No. 5, pp. 5389–5415, 2021.

[4] A. Abbas, H. Haruna, A. Arrona Palacios, C. Camacho Zun˜iga, S. Nu´n˜ez Daruich, J. F. O, R. Castan˜o Gonzalez, J. Escamilla, S. Hosseini, “Students’ evaluations of teachers and recommendation based on course structure or teaching approaches: An empirical study based on the institutional dataset of student opinion survey”, *Education and Information Technologies*, Vol. 27, No. 9, pp. 12049–12064, 2022.

[5] K. V. Vlasenko, I. V. Sitak, D. A. Kovalenko, S. V. Volkov, I. V. Lovianova, S. O. Semerikov, S. L. Zahrebelnyi, “Methodical recommendations for the development of online course structure and content”, In: *Proc. of the 1st Symposium on Advances in Educational Technology Volume 1: AET*, pp. 471–485, 2022.

[6] K. A. Sopan, K. V. Sambhajirao, “Intelligent course recommendation for students using machine learning models”, *International Journal of All Research Education and Scientific Methods*, Vol. 12, pp. 2397–2402, 2024.

[7] M. A. Z. Khan, A. Polyzou, *et al.*, “Session based methods for course recommendation”, *Journal of Educational Data Mining*, Vol. 16, No. 1, pp. 164–196, 2024.

[8] K. Mrhar, M. Abik, “A deep learning framework for optimizing personalized online course recommendation and selection”, *Decision Analytics Journal*, 100616, 2025.

[9] J. T. Iorzua, T. Moses, C. I. Eke, O. J. Agushaka, D. K. Kwaghtyo, T. Godswill, “A machine learning based approach to course and career recommendation system: A systematic literature review”, *Journal of Computing Theories and Applications*, Vol. 3, No. 1, pp. 1–16, 2025.

[10] C. Zhang, X. Hu, W. Wei, “A review of AI based techniques in online course recommendation: Metrics, factors, and research methods”, In: *Proc. of 2024 International Conference on Intelligent Education and Intelligent Research (IEIR)*, pp. 1–8, 2024.

[11] S. Wang, M. Chen, X. Zhu, “Research on course recommendation algorithm based on knowledge graph”, In: *Proc. of 2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 1180–1184, 2024.

[12] X. Chen, C. Yin, H. Chen, W. Rong, Y. Ouyang, Y. Chai, “Course recommendation system based on course knowledge graph generated by large language models”, In: *Proc. of 2024 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, pp. 1–8, 2024.

[13] A. S. Kalokhe, V. S. Kumbhar, “A review on course recommendation system in higher education using machine learning”, *International Research Journal of Humanities and Interdisciplinary Studies*, pp. 47–55, 2024.

[14] M. Haque, B. Kotaiah, J. Ahmed, “A comprehensive review of course recommendation systems for MOOCs”, *International Journal of Computing and Digital Systems*, Vol. 16, No. 1, pp. 1697–1712, 2024.

[15] B. Ma, T. Yang, B. Ren, “A survey on explainable course recommendation systems”, In: *Proc. of International Conference on Human Computer Interaction*, pp. 273–287, 2024.

[16] A. S. Vino, P. M. DiyaAlDin, S. Sandosh, “A strategic approach in course recommendation system using similarity models”, In: *Proc. of ICAMDMS 2024: International Conference on Advancements in Materials, Design and Manufacturing for Sustainable Development*, p. 169, 2024.

[17] A. Madhavi, T. Mohammad, M. P. Syed, “Survey on course recommendation in e-learning platform”, In: *Proc. of International Conference on Data Science and Big Data Analysis*, pp. 529–548, 2024.

[18] N. Kamal, F. Sarker, A. Rahman, S. Hossain, K. A. Mamun, “Recommender system in academic choices of higher education: a systematic review”, *IEEE Access*, Vol. 12, pp. 35475–35501, 2024.

[19] P. He, “Multi dimensional information multimedia big data mining analysis relying on association rule mapping model”, *Arabian Journal for Science and Engineering*, Vol. 50, No. 10, pp. 7361–7373, 2025.

[20] J. Lei, “[retracted] association rule mining algorithm in college students’ quality evaluation system”, *Journal of Electrical and Computer Engineering*, Vol. 2022, No. 1, 6721504, 2022.

[21] S. T. Kokab, S. Asghar, S. Naz, “Transformer based deep learning models for the sentiment analysis of social media data”, *Array*, Vol. 14, 100157, 2022.

[22] C. M. Muia, A. M. Oirere, R. N. Ndungu, “A comparative study of transformer based models for text summarization of news articles”, *International Journal of Advanced Trends in Computer Science and Engineering*, 2024, doi: 10.30534/ijatcse/2024/011322024.

[23] M. Zehlike, K. Yang, J. Stoyanovich, “Fairness in ranking, part II: Learning to rank and

recommender systems", *ACM Computing Surveys*, Vol. 55, No. 6, pp. 1–41, 2022.

[24] S. Zhang *et al.*, "Instruction tuning for large language models: A survey", *ACM Computing Surveys*, 2023.

[25] H. Samo, K. Ali, M. Memon, F. A. Abbasi, M. Y. Koondhar, K. Dahri, "Fine tuning Mistral 7B large language model for Python query response and code generation: A parameter efficient approach", *VAWKUM Transactions on Computer Sciences*, Vol. 12, No. 1, pp. 205–217, 2024.

[26] J. Parapar, F. Radlinski, "Towards unified metrics for accuracy and diversity for recommender systems", In: *Proc. of the 15th ACM Conference on Recommender Systems*, pp. 75–84, 2021.