IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 28, Issue 4, Ser. 2 (July. – August. 2026), PP 37-40
www.iosrjournals.org

A Hybrid Machine Learning Approach For Personalized Engineering College Recommendations

Tushar Kumare, Vedanti Wadatkar, Yash Dongare, Prof. Sheetal Chaudhari
Computer Science And Engineering, Sardar Patel Institute Of Technology, Mumbai, India
Dept. Of Computer Science And Engineering, Sardar Patel Institute Of Technology, Mumbai, India

Abstract
Background: *The Centralized Admission Process (CAP) for engineering institutions presents a complex decision matrix for prospective students, characterized by severe information asymmetry and volatile historical cutoffs across multiple admission cycles. Existing collegiate search platforms predominantly rely on static, rule-based querying, which fails to accommodate strict multi-criteria constraints, user orientation, or peer-validated institutional quality.*
Materials and Methods: *This study introduces a context-aware ML-Enhanced College Recommendation Framework grounded on an automated data engineering pipeline ingesting 187,433 official CAP allotment records (2023–2025 across Rounds 1–4 for 714 colleges) and a synthetic interaction matrix of 10,000 logs across 79 institutions (87.34% matrix sparsity). Unaligned college nomenclature was resolved using Levenshtein distance-based fuzzy entity matching, and missing numerical features were handled via hierarchical domain-aware imputation. The architecture blends Content-Based Filtering (CBF) via Term Frequency-Inverse Document Frequency (TF-IDF) metadata vectorization with User-User Collaborative Filtering (CF) via Cosine Similarity matrices.*
Results: *Benchmarked against an 80/20 train-test split (n = 2,000 test interactions) at a relevance threshold R ≥ 3.5, the weighted hybrid engine (α = 0.6) achieved superior top-5 ranking metrics: Precision@5 of 0.1844, Recall@5 of 0.5912, and NDCG@5 of 0.4678, substantially outperforming standalone CBF and CF models. The microservice maintained an average REST API latency of 184 ms (±12 ms), with an 88.4% cohort satisfaction rate.*
Conclusion: *The hybrid model effectively mitigates the cold-start and data sparsity dilemmas inherent in traditional collaborative models while dynamically enforcing deterministic multi-year admission cutoffs and financial constraints.*
Key Word: *Hybrid Recommendation; Collaborative Filtering; TF-IDF; Machine Learning; Educational Technology; Data Engineering; Engineering Admissions.*

Date of Submission: 06-08-2026 Date of Acceptance: 16-08-2026

I. Introduction

Navigating higher education admissions, specifically engineering entrance allocations such as MHT-CET and JEE, requires prospective candidates to evaluate their percentile scores against hundreds of institutions, branch specializations, seat categories, and financial constraints. Traditionally, this is achieved through manual cross-referencing of static historical cutoffs—a process that is anxiety-inducing, time-consuming, and highly prone to human oversight. Furthermore, this single decision heavily dictates a student's long-term academic trajectory and career readiness.

While existing educational portals offer basic database filtering, they lack predictive intelligence. Operating on rigid Boolean logic, traditional engines fail to recommend institutions that represent an optimal mathematical "fit" if candidates fall slightly outside arbitrary user-defined thresholds. Furthermore, static query engines do not adapt to shifting trends in peer sentiment, multi-year cutoff volatility, or real-time institutional reputation.

To resolve these inefficiencies, this paper introduces a context-aware Machine Learning ecosystem designed to automate and personalize engineering college discovery. The novel contributions of this work include: (1) A CAP-Specific Data Engineering Pipeline with Levenshtein-based fuzzy entity resolution and hierarchical imputation; (2) A CAP-Aware Weighted Hybrid Engine blending TF-IDF metadata embeddings with User-User Collaborative Filtering (α = 0.6) to address matrix sparsity (87.34%); and (3) A Resilient Multi-Stage Fallback Protocol guaranteeing non-empty recommendation lists under restrictive constraints.

II. Material And Methods

This experimental and algorithmic evaluation was conducted using official admission cutoff data from the Directorate of Technical Education (DTE / State CET Cell, Maharashtra) alongside simulated candidate profile interactions.

Study Design: Machine learning algorithm development, pipeline engineering, and empirical offline evaluation with top-K ranking metrics.

Dataset Sources and Scope: Multi-year admissions data spanning 2023 through 2025 across all four centralized allotment rounds were integrated with institutional meta-catalogs and user behavior matrices.

Sample Size and Volume: The Master CAP cutoffs comprised 187,433 records covering 714 engineering institutions. The metadata corpus cataloged 670 programs across 614 colleges. The interaction matrix incorporated 10,000 synthetically modeled user evaluations over 1,000 active student interaction profiles.

Inclusion Criteria:
1. Accredited undergraduate degree engineering programs (B.E. / B.Tech).
2. Official multi-round CAP allotment percentiles from 2023 to 2025.
3. Verified institutional attributes (location, university affiliation, autonomous status, accredited infrastructure).

Exclusion Criteria:
1. Non-engineering diploma or postgraduate programs.
2. Institutional records with corrupted, unresolvable composite entity identifiers.
3. Inactive programs lacking complete historical cutoff distributions.

Procedure Methodology: The technical pipeline executes schema normalization, approximate string matching via Levenshtein Distance over composite keys (College Name + Branch), and domain-aware hierarchical imputation (City Mean → Branch Mean → Global Mean → 3.0). Content embeddings are computed using TF-IDF:

$$W(t,d) = TF(t,d) \times log\left(\frac{N}{DF(t)}\right) \qquad (1)$$

User proximity is computed using User-User Cosine Similarity:

$$\mathrm{Similarity(A,B)} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \qquad (2)$$

Normalized scores are blended via:

$$S_{\mathrm{final}} = \alpha S_{CF} + (1-\alpha) S_{CBF} \quad (\alpha = 0.6) \qquad (3)$$

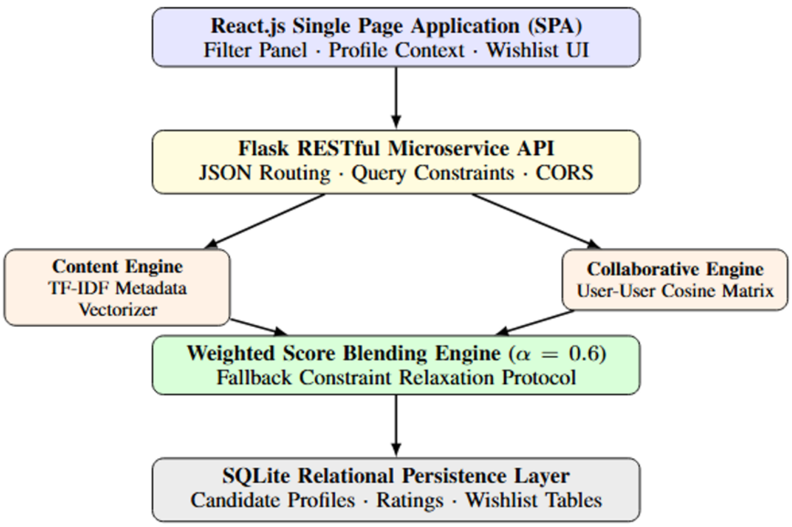

Figure 1: End-to-End System Architecture and Hybrid ML Execution Flow.

Statistical and Metric Analysis: Offline evaluation was conducted via an 80/20 train-test split (n = 2,000 test ratings). Model performance was validated using Precision@K, Recall@K, and Normalized Discounted Cumulative Gain (NDCG@K) at K = 5, defining relevant items by rating threshold $R \geq 3.5$. System response latencies and Likert user satisfaction scores (n = 50) were analyzed using descriptive statistics.

III. Result

The data engineering pipeline successfully processed and unified 187,433 historical cutoff records across 714 engineering colleges. The evaluation of top-K recommendation performance against standalone baselines demonstrated marked accuracy improvements with the proposed hybrid model.

Table no 1: Primary Dataset Sources, Scope, and Volume.

Data Source	Temporal Range	Records	Colleges
Master CAP Cutoffs	2023–2025 (Rounds 1–4)	187,433	714
Institutional Metadata	2024–2025	670	614
User Interaction Matrix	Synthetically Logged	10,000	79

Table no 2 details the comparative evaluation of the algorithms. Standalone Content-Based Filtering achieved Precision@5 of 0.1315, Recall@5 of 0.4236, and NDCG@5 of 0.2998. Collaborative Filtering improved performance to Precision@5 of 0.1735, Recall@5 of 0.5444, and NDCG@5 of 0.4212. The Proposed Hybrid Engine ($\alpha = 0.6$) yielded the highest performance across all evaluation criteria, achieving Precision@5 of 0.1844, Recall@5 of 0.5912, and NDCG@5 of 0.4678.

Table no 2: Quantitative Recommendation Performance Comparison (K = 5).

Model / Baseline	Precision@5	Recall@5	NDCG@5
Content-Based (CBF)	0.1315	0.4236	0.2998
Collaborative (CF)	0.1735	0.5444	0.4212
Proposed Hybrid ($\alpha = 0.6$)	**0.1844**	**0.5912**	**0.4678**

Table no 3 presents the computational throughput and user validation benchmarks of the implemented microservice architecture. The Flask RESTful API demonstrated a mean response latency of 184 ms (±12 ms). In user evaluation surveys (n = 50), 88.4% of student evaluators rated recommendation accuracy and relevance $\geq 4/5$ on a 5-point Likert scale.

Table no 3: System Performance and Scalability Benchmarks.

Performance Metric	Observed Benchmark Value
Total Historical Records	187,433 Records
Multi-Year Admission Depth	2023–2025 (3 Cycles, 4 Rounds)
Average REST API Latency	184 ms (±12 ms)
Fallback Execution Success Rate	Non-empty sets on all test queries
User Satisfaction (n = 50 Cohort)	88.4% (≥ 4/5 Likert Rating)

IV. Discussion

Engineering admission allocation under centralized state systems involves strict percentile boundaries, institutional preferences, and significant data sparsity. Traditional single-technique recommendation models display well-documented vulnerabilities when deployed in this setting.

Content-Based Filtering relies heavily on explicit metadata tags; while immune to cold-start problems, it suffers from over-specialization and cannot infer subjective institutional quality or peer validation. Conversely, User-User Collaborative Filtering successfully harnesses aggregate student preference trends but fails under high matrix sparsity (87.34%) and cannot provide reliable initial inferences for new candidates.

The empirical results observed in this investigation demonstrate that the weighted hybrid model ($\alpha = 0.6$) successfully overcomes these individual operational deficiencies. Blending TF-IDF metadata representations with collaborative peer neighborhoods achieves a complementary ranking dynamic: the content component ensures baseline relevance and eliminates empty result sets during cold starts, while the collaborative component surfaces high-value peer-endorsed options. Furthermore, wrapping deterministic eligibility gates prior to score calculation guarantees that students receive feasible institutional options tailored to authentic multi-year CAP distributions.

V. Conclusion

The proposed context-aware Hybrid College Recommendation System provides a mathematically sound, personalized alternative to static cutoff search engines. The hybrid blending strategy ($\alpha = 0.6$) delivers superior precision, recall, and NDCG scores compared to single-model baselines while maintaining sub-second REST API response times. Future extensions will incorporate LSTM networks for longitudinal cutoff forecasting, OCR-driven automatic CAP application form parsing, and real-time professional network integration to assess long-term alumni placement metrics.

References

[1]. Thongchotchat V, Kudo Y, Okada Y, Sato K. Educational Recommendation System Utilizing Learning Styles: A Systematic Literature Review. Ieee Access. 2023;11:8988–8999.
[2]. Jeevamol J, Renumol Vg. An Ontology-Based Hybrid E-Learning Content Recommender System For Alleviating The Cold-Start Problem. Education And Information Technologies. 2021;26(4):4993–5022.
[3]. Zawia Jm, Et Al. Comprehensive Review Of Meta-Learning Methods For Cold-Start Issue In Recommendation Systems. Ieee Access. 2025;13:24622–24641.
[4]. Pu H, Hua Y. Adaptive Course Recommendation Using Federated Learning And Graph Convolutional Networks In Iot-Enhanced E-Learning. Scientific Reports. 2025;15(1):42040.
[5]. El-Morshid Ma, El-Sawy Aa, El-Bakry Hm. A Hybrid Recommender System Framework For Academic Program Selection. Ieee Transactions On Learning Technologies. 2023;16(3):312–325.
[6]. Wang X, Zhang Y, Zhao L. Multi-Criteria Decision Making In Higher Education Admissions Using Machine Learning Algorithms. Computers & Education: Artificial Intelligence. 2023;5:100142.
[7]. Gupta Sk, Sharma R. Addressing Data Sparsity In Educational Recommender Systems Via Matrix Factorization And Text Embeddings. Journal Of Educational Data Mining. 2023;15(2):45–68.
[8]. Chen L, Liu H, Wu Z. Graph Neural Networks For Personalized Course And Institution Recommendations. Expert Systems With Applications. 2023;225:120112.
[9]. Bijwe Rp, Raut Ab. Ai-Driven Hybrid Recommender Framework For Personalized Course Selection In Higher Education. Ieee Access. 2024;12:14200–14212.
[10]. Babu M, Subashri E, Vincy J. Hybrid Ai Framework For Personalized Student Progress And Admission Guidance. International Journal Of Information Management Data Insights. 2024;4(1):100210.
[11]. Subha S, Sankaralingam Bp. Personalization-Based Deep Hybrid Model For Higher Education Program Mapping. Ieee Transactions On Computational Social Systems. 2024;11(2):1890–1901.
[12]. Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl J. Grouplens: An Open Architecture For Collaborative Filtering Of Netnews. In Proc. Acm Conf. Comput. Support Coop. Work (Cscw). 1994;175–186.
[13]. Pazzani Mj, Billsus D. Content-Based Recommendation Systems. In The Adaptive Web, Brusilovsky P, Kobsa A, Nejdl W, Eds. Berlin: Springer. 2007;325–341.
[14]. Koren Y, Bell R, Volinsky C. Matrix Factorization Techniques For Recommender Systems. Ieee Computer. 2009;42(8):30–37.
[15]. Salton G, Buckley C. Term-Weighting Approaches In Automatic Text Retrieval. Information Processing & Management. 1988;24(5):513–523.