

Comparative Analysis of Multi-Criteria Recommender Algorithms for College Choice

Haleemah Amisu

Department of Computer and Information Sciences

Towson University

Towson, USA

hamisu1@students.towson.edu

Abstract—College choice is a multi-criteria decision problem in which institutional outcomes, affordability, resources, and equity do not align uniformly across all options. This study evaluates three recommendation algorithms designed to retrieve public U.S. colleges using percentile-normalized performance indicators from IPEDS (Integrated Post-secondary Education Data System). The models tested include: a weighted additive scorer that ranks institutions based on user-defined importance weights, a cosine similarity model that selects institutions proportional to preference alignment, and a cluster-aware approach that distributes recommendations across institutional profile groups. Twenty-five student priority profiles were constructed to observe model behavior under controlled variation. Results show that the weighted method consistently prioritizes high-performing institutions across selected attributes, cosine similarity favors alignment in relative preference structure even with moderate absolute performance, and cluster-aware retrieval promotes institutional diversity without discarding fit quality. The findings demonstrate how algorithm design influences recommendation composition when applied to the same dataset, establishing a comparative reference for college-choice recommender construction and evaluation.

I. INTRODUCTION

Recommender systems are designed to reduce complexity in decision environments where available options exceed an individual's capacity to evaluate them manually. Most recommender literature has developed in domains such as retail, media, and digital streaming, where user preferences are modeled from behavior or stated interests. The governing principle remains consistent across contexts: a system selects, ranks, or filters information on behalf of a user based on measurable attributes.

College selection represents a similar multi-attribute decision problem. Public institutions differ in measurable ways, including graduation outcomes, cost of attendance, funding and staffing, and equity in student support. Students typically do not evaluate these variables simultaneously, and standard ranking lists aggregate them into a single score that may obscure underlying trade-offs. A recommender-based framing allows institutions to be compared across independent dimensions rather than collapsed into one composite measure.

In this study, we evaluate three recommendation approaches: weighted additive scoring, cosine similarity to an ideal profile, and cluster-aware sampling, using a dataset of U.S. public colleges represented by percentile-scaled indicators of student

success, affordability, academic resources, and access and equity. Each algorithm produces a ranked subset of institutions from the same data, and differences in their outputs reflect the behavior of the underlying decision logic rather than changes in the dataset.

The contribution of this work is an examination of recommender behavior under controlled variation. Results highlight where algorithms converge, where they diverge, and how each prioritization method distributes institutional characteristics. The purpose is to document how these models select, sort, and emphasize options when given identical inputs, generating a basis for understanding recommender suitability in college-choice systems.

II. RELATED WORK

Recommender systems have evolved through several methodological eras, each reflecting different assumptions about user preference modeling, item representation, and objective functions. The foundational survey by Adomavicius and Tuzhilin [1] describes the classical division between content-based recommenders, collaborative filtering approaches, and hybridized models that combine multiple signals. This taxonomy remains the dominant conceptual scaffold in recommender-systems research, and it provides the basis for interpreting the three recommendation strategies evaluated in this work. A subsequent meta-review by Park et al. [2] further categorizes recommendation research by domain, input type, and data structure, emphasizing the distinction between rating-prediction systems and ranking-based systems. Our study aligns with the latter category: instead of predicting a satisfaction score, we directly rank institutions according to student-aligned criteria.

Much of the modern recommender-systems literature is grounded in collaborative filtering, with particular emphasis on modeling interactions in high-dimensional spaces. Koren, Bell, and Volinsky [3] illustrate how matrix-factorization techniques surpassed neighborhood-based similarity for platforms like Netflix by capturing latent preference structure. Although this paradigm demonstrates state-of-the-art performance in implicit-feedback environments—often utilizing deep learning architectures as reviewed by Zhang et al. [9]—the approach is less suitable for contexts requiring interpretability and transparent attribute trade-offs. Bobadilla et al. [4] similarly review

collaborative filtering algorithms and evaluate them according to accuracy, coverage, diversity, novelty, and robustness. Their evaluation framework is relevant to the present work: in addition to selecting appropriate algorithms, our study measures recommendation diversity across models and examines how different weighting functions alter institutional representation.

The design of Ranking With Purpose places the user in control of multiple institutional dimensions (student success, affordability, academic resources, and equity) which situates the system within the class of multi-criteria recommenders. Adomavicius and Kwon [5] formalize multi-criteria recommendation as an extension of traditional utility modeling, where recommender outputs must reflect user-dependent importance weights rather than a single aggregated rating. Our weighted-sum model follows this design directly: a student profile is mapped to a vector of criterion weights, and institutions are ranked according to a normalized multi-attribute score. Earlier work by Manouselis and Costopoulou [6] similarly categorizes multi-criteria recommenders by their aggregation functions, highlighting that explicit-weight, additive models remain effective when transparency is essential to user trust. This principle aligns with the ethos of RWP, where stakeholders must be able to trace recommendations back to measurable variables instead of opaque latent embeddings.

Hybrid recommendation strategies have been proposed as a solution to the limitations of any single method. Burke [7] surveys hybrid architectures (weighted combinations, mixed recommenders, switching models, and cascaded pipelines) and demonstrates that blending signals can increase robustness and mitigate overspecialization. The cluster-aware model evaluated in this paper reflects this lineage: it layers institutional grouping on top of numerical scoring to prevent homogeneity in top-N outputs. Finally, diversity-aware recommendation research (Vargas & Castells [8]) argue that relevance alone is not an adequate optimization target, since narrow lists may exclude valuable alternatives. Our cluster-aware approach is an intentional trade-off mechanism, encouraging recommendation sets that reveal multiple institutional types rather than repeatedly selecting the same cluster.

These works contextualize the three algorithms studied in this paper. The weighted-sum approach draws from multi-criteria utility modeling, cosine similarity represents classical similarity-based recommendation, and the cluster-aware approach extends hybrid and diversity-oriented thinking into a domain where fairness, transparency, and exploration matter as much as accuracy.

III. DATA SOURCE AND DESCRIPTION

The study uses institutional data sourced from the U.S. Department of Education's Integrated Postsecondary Education Data System (IPEDS, 2023 release). The raw dataset contains indicators of student outcomes, affordability, resources, and demographic composition across accredited public four-year institutions in the United States.

The raw IPEDS file included fields representing:

- **Institutional identity and classification:** UnitID, State, Carnegie Classifications, HBCU/Tribal flags.
- **Academic output:** Degree counts across bachelor's, master's, and doctoral levels.
- **Student aid and affordability:** Pell Grant percentages, average net price, loan/grant award amounts.
- **Graduation and retention metrics:** 4-year, 5-year, and 6-year rates; transfer-out rate.
- **Revenues and expenditures:** Tuition share of revenue, state/local appropriations, research spending per FTE.
- **Faculty and learning environment:** Student-to-faculty ratio, library holdings, expenditures per FTE.
- **Enrollment characteristics:** Gender distribution, race categories, distance education enrollment.

1) *Feature Consolidation and Metric Construction:* From the raw dataset, selected variables were cleaned, converted to numeric format where necessary, and normalized into comparable scales. Outcome-relevant fields were then used to derive four conceptual performance dimensions, detailed in Table I.

TABLE I
CONSTRUCTED PERFORMANCE DIMENSIONS

Dimension	Example Inputs
Student Success	4-, 5-, 6-year graduation rates, retention rate
Affordability	Average net price, Pell share, aid coverage
Resources	Instructional spending per FTE, research spending
Equity	Pell completion gap, demographic representation

Each dimension was aggregated into a continuous score, then scaled into percentiles (0–100) to enable independent weighting inside the recommender.

2) *Clustering Procedure:* To explore structural similarity between institutions, the cleaned performance dimensions were used for unsupervised grouping. K-Means clustering was selected after experimenting with alternatives (hierarchical, silhouette-based, agglomerative). The clustering step produced a categorical label (`cluster_id`, `cluster_name`) describing institutional profile types, later used in the recommender as one of the selection models.

IV. METHODOLOGY

The recommender system operates on the final dataset described in the data description section. Three retrieval models were implemented to generate institutional recommendations based on student preferences. Each method receives a user-defined preference vector $W = \{w_1, w_2, w_3, w_4\}$, where w_i corresponds to one of the four percentile fields. The models then produce a ranked list of colleges, and the study evaluates how these lists align with user-defined priorities. The intent is to observe how algorithmic behavior shifts when different assumptions are embedded into the scoring process.

A. Weighted Score Recommender

The first model computes a composite ranking score based on user-assigned importance weights for each percentile metric. Input weights are normalized to sum to one. Each institution receives a composite score computed as:

$$Score_{weighted} = \sum_{i=1}^4 w_i \cdot p_i \quad (1)$$

and instantiated as:

$$\text{Score} = w_s \cdot \text{success} + w_a \cdot \text{affordability} + w_r \cdot \text{resources} + w_e \cdot \text{equity} \quad (2)$$

Where w_s, w_a, w_r, w_e are user-specified priorities. The output is the top- k institutions ranked by score. This model assumes additive trade-offs. High strength in one area may compensate for weakness in another.

B. Cosine Similarity Recommender

The cosine model represents each institution as a four-dimensional vector and computes the similarity between the institution vector P and a normalized preference vector W (i.e. it computes the cosine of the angle between the institution vector P and the student preference vector W):

$$\text{similarity} = \frac{P \cdot W}{\|P\| \|W\|} \quad (3)$$

This method emphasizes proportional alignment. A school that mirrors a student's relative priorities (e.g., success > affordability > equity > resources) may rank highly even if its absolute percentiles are moderate.

C. Cluster-Aware Recommender

The third model first filters institutions through the pre-assigned institutional clusters derived from efficiency and outcome patterns. This ensures that retrieval occurs within conceptually similar peer groups. The recommender then applies the weighted additive score as in method (1) to produce results within each cluster before ranking all clusters at the end.

This approach reduces cross-profile comparison. Highly resourced institutions are not directly competing with access-focused institutions unless a student's preferences explicitly rank those traits highly. Recommendations emerge as within-profile best fits.

D. Evaluation Strategy

Model behavior was examined by constructing distinct student profiles representing varied real-world priorities. Each profile was run through the three recommenders to observe ranking differences. For each profile, the three recommender models produced a top- k list of institutions. Outputs were compared using:

- Overlap proportion between model lists
- Average percentile composition of recommended institutions
- Cluster diversity count (number of clusters represented in the top- k)

All tests were executed on the same dataset with identical k values of $k = 10$ to maintain comparability.

V. RESULTS & ANALYSIS

The three recommendation models were evaluated on a set of 25 synthetic student profiles. Each profile encoded a distinct pattern of priorities over the four percentile-based metrics (student success, affordability, resources, and equity), along with optional constraints such as preferred states. The profiles were designed to approximate realistic decision scenarios, including low-income and first-generation students, geographically constrained students, resource-seeking students, and students who explicitly prioritized equity or institutional diversity.

For each profile, the weighted, cosine, and cluster-aware recommenders produced a top- k list with $k = 10$. The analysis focused on three aspects of model behavior: (1) overlap in recommended institutions across models, (2) the average metric composition of the recommended sets, and (3) the institutional diversity of each list measured by cluster coverage.

A. Illustrative Profile: Low-Income First-Generation Student

One profile is used here as a detailed example. The profile represents a low-income, first-generation student who places the highest importance on affordability and completion outcomes, with moderate concern for resources and equity. The corresponding weight vector favors affordability and student success, while thresholds eliminate institutions below a minimum affordability and success percentile.

Table II shows the top ten institutions returned by each model for this profile. The weighted and cluster-aware recommenders share several institutions, while the cosine-based recommender yields a markedly different list.

TABLE II
TOP-10 INSTITUTIONS FOR A LOW-INCOME FIRST-GENERATION PROFILE
UNDER THREE MODELS.

Weighted score	Cosine similarity	Cluster-aware
University of Minnesota–Morris	Univ. of Wisconsin–Stevens Point	CUNY Bernard M Baruch College
CUNY Bernard M Baruch College	Louisiana Tech University	University of Florida
New College of Florida	Murray State University	Univ. of North Carolina at Greensboro
Univ. of North Carolina at Greensboro	Western Carolina University	University of Minnesota–Morris
North Carolina A&T State Univ.	CUNY Bernard M Baruch College	University of Minnesota–Rochester
University of Minnesota–Rochester	Univ. of North Alabama	UNC School of the Arts
Univ. of Illinois Springfield	Western Kentucky Univ.	North Carolina A&T State Univ.
Cal State Univ.–Stanislaus	Cal State Poly Pomona	New College of Florida
CUNY Hunter College	Cal State Univ.–Fullerton	Western Carolina University
University of Florida	West Liberty University	SUNY Oneonta

Overlap between model outputs was quantified as the fraction of shared institutions in the respective top-10 lists. For this

profile, the weighted and cluster-aware lists exhibited moderate overlap, while cosine similarity produced a largely disjoint set:

- Weighted vs. cosine: very low overlap.
- Weighted vs. cluster-aware: moderate overlap.
- Cosine vs. cluster-aware: low overlap.

Cluster diversity was computed as the number of distinct institutional clusters represented in each top-10 list. The weighted and cluster-aware recommenders each covered four clusters, whereas the cosine recommender concentrated its recommendations in a single cluster. This indicates that the cosine model tended to favor a narrow band of institutions whose relative profile closely matched the input weights, even when those institutions did not score highly on the underlying metrics.

Table III reports the mean percentile values for the institutions recommended by each model under this profile. The weighted and cluster-aware methods produced recommendation sets with higher average success and affordability than the cosine-based method. The cluster-aware model also preserved relatively strong resource percentiles while maintaining moderate equity scores.

TABLE III
MEAN PERCENTILES OF RECOMMENDED INSTITUTIONS FOR THE
LOW-INCOME FIRST-GENERATION PROFILE.

Model	Success	Afford.	Resources	Equity
Weighted	74.9	82.3	69.6	63.0
Cosine	67.5	71.4	26.7	37.0
Cluster-aware	80.7	76.0	73.3	49.1

In this example, the weighted model emphasizes affordability while retaining reasonably high success and resource scores. The cosine model aligns with the relative preference pattern but selects institutions with lower absolute resources and equity. The cluster-aware model yields institutions with very strong success and resource percentiles and moderate affordability, reflecting the influence of cluster structure on the final ranking.

B. Behavior Across 25 Profiles

The same evaluation procedure was repeated for 24 additional profiles spanning different combinations of priorities. Each profile specified a distinct weight vector over the four metrics and, in some cases, state-level constraints that restricted consideration to a subset of regions.

- For each profile, three summary quantities were computed:
- 1) The overlap between model outputs, measured as the proportion of shared institutions across top-10 lists.
 - 2) The mean percentile scores of recommended institutions in each model, computed separately for success, affordability, resources, and equity.
 - 3) The number of distinct clusters represented in each model's recommendations.

These quantities characterize three dimensions of recommender behavior: agreement, attribute emphasis, and institutional diversity.

Overall, the weighted and cluster-aware models showed the highest agreement in recommended institutions, particularly in profiles where student success and affordability carried similar weight. Cosine similarity often diverged from both, especially in profiles with highly skewed preferences. In those cases, cosine similarity tended to prioritize proportional matches to the preference vector even when absolute scores were modest.

Across profiles, lists produced by the weighted and cluster-aware models tended to maintain higher average success and affordability percentiles than cosine-based lists. Cluster-aware recommendations frequently selected institutions with strong resources as well, especially in profiles where resources were non-trivial components of the preference vector. Equity scores varied more widely across profiles, reflecting the fact that some profiles did not explicitly prioritize equity, while others assigned it primary importance.

Cluster diversity patterns were consistent with the illustrative example. The cosine recommender often concentrated recommendations within one or two clusters, while the weighted and cluster-aware models more frequently drew from three or more clusters in their top-10 lists. The cluster-aware model was explicitly designed to prevent over-concentration in a single institutional profile type, and the results reflected this design: institution sets were more varied in terms of cluster membership while still aligned with the student's stated priorities.

C. Summary of Comparative Behavior

The experiments indicate that the weighted and cluster-aware models are better aligned with scenarios where students care about both absolute performance and transparent trade-offs between metrics. The weighted model is straightforward and easy to explain, and it tends to reward institutions that are strongest on the most heavily weighted dimensions. The cluster-aware model further distributes recommendations across multiple institutional types, which can be useful when the goal includes exploration of different but still compatible options.

The cosine similarity model is most responsive to the relative pattern of preferences rather than absolute magnitudes. This can be useful when the goal is to find institutions whose profiles mirror the shape of a student's priorities, but it may underemphasize overall performance when percentiles are modest across all dimensions. The evaluation across 25 profiles shows that this behavior frequently leads to lower resource and equity percentiles in recommended sets compared to the other two methods.

Taken together, the results show that the choice of recommendation model has a direct effect on which institutions appear as top options for a given profile. When the same dataset and the same preference inputs are used, the three algorithms surface different trade-offs between concentration, diversity, and absolute performance.

VI. DISCUSSION AND CONCLUSION

The comparative evaluation illustrates that recommendation outcomes vary substantially depending on how preference

signals are interpreted. The weighted model showed consistent alignment with profiles that prioritized affordability and completion outcomes, and it maintained higher average institutional percentiles across most dimensions. Its behavior reflects a straightforward optimization: the institutions scoring highest on the most heavily weighted metrics rise to the top. This makes the model suitable for decision scenarios where the goal is to rank institutions by measurable performance rather than pattern similarity.

The cosine-based model's selections were based on proportional alignment between institutional profiles and preference vectors rather than absolute percentile strength. In several test profiles, this resulted in recommendations that matched preference patterns but scored lower in resources or affordability compared to the other models. This model is therefore most appropriate when directional alignment is more important than magnitude, such as identifying institutions structurally similar to a student's prioritization pattern even when absolute performance varies.

The cluster-aware model extends the weighted approach by preserving institutional diversity. Instead of drawing repeatedly from one high-scoring institutional type, it distributes recommendations across clusters representing different institutional strategies. This reduces homogeneity in suggested options and increases the likelihood of surfacing institutions that balance strength in outcomes, affordability, resources, and equity in different proportions. Across the 25 tested profiles, this method frequently returned colleges with strong performance while still maintaining cluster variety.

The combined findings show that Weighted scoring offers clarity and measurable trade-offs; cosine similarity emphasizes directional alignment; cluster-aware selection maintains diversity without discarding performance. The choice among these models depends on the context of use. Systems prioritizing transparency and comparative strength benefit from additive scoring. Systems designed to help students explore unfamiliar options may favor cluster-aware sampling. Systems seeking structural similarity rather than absolute strength align with cosine-based retrieval.

In college recommendation contexts, where students operate with incomplete information and constraints vary across individuals, transparency and diversity both carry practical value. The models evaluated in this work demonstrate three viable ways to express institutional fit, each suited to different decision priorities. The results suggest that college-choice recommenders benefit from multiple retrieval logics rather than a single ranking mechanism, and that model selection should be treated as an intentional design decision rather than a technical detail.

REFERENCES

- [1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005.
- [2] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim, "A literature review and classification of recommender systems research," *Expert Syst. Appl.*, vol. 39, no. 11, pp. 10059–10072, 2012.
- [3] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
- [4] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowl.-Based Syst.*, vol. 46, pp. 109–132, 2013.
- [5] G. Adomavicius and Y. Kwon, "New recommendation techniques for multicriteria rating systems," *IEEE Intell. Syst.*, vol. 22, no. 3, pp. 48–55, 2007.
- [6] N. Manouselis and C. Costopoulou, "Analysis and classification of multicriteria recommender systems," *World Wide Web*, vol. 10, no. 4, pp. 415–441, 2007.
- [7] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Model. User-Adapt. Interact.*, vol. 12, no. 4, pp. 331–370, 2002.
- [8] S. Vargas and P. Castells, "Rank and relevance in novelty and diversity metrics for recommender systems," in *Proc. 5th ACM Conf. Recommender Syst.*, 2011, pp. 109–116.
- [9] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Comput. Surv.*, vol. 52, no. 1, pp. 1–38, 2019.