Vol. 1, Issue 11 | Nov 2026 | Open Access Research Paper

Predictive Analytics for College Admissions: Integrating Cutoff Estimation, College Recommendation, and Career Guidance Across Engineering and Medical Disciplines in Karnataka

Iranna S. Tiralapur, Sindhu B. Binkadakatti, Spoorti S. Nidoni, Srishti P. Sulakhe

Department of Computer Science and Engineering
Basaveshwar Engineering College (Autonomous), Bagalkote, Karnataka, India
Corresponding author: spoortinidoni20@gmail.com

***Abstract*—**Students aspiring for professional undergraduate engineering and medical seats through Karnataka counselling must choose among thousands of college–course–category options using only sparse, inconsistently labelled, and shifting past closing ranks. For engineering, 33,182 records were extracted from 419 bulletin pages, and 2024 Round 1 closing ranks were used to predict 2025 ranks for KCET (N = 16,070) and COMEDK (N = 690). A four-tier hierarchical fallback and split conformalised quantile regression (CQR) were compared with a historical lookup, a target-encoded random forest, and a primary tree model. On KCET, the lookup had the lowest mean absolute error (19,931 ranks against 20,429 for the fallback), whereas the fallback had the highest R^2 (0.7731 against 0.7212); on COMEDK, the random forest outperformed the fallback (R^2 0.5270 against 0.4582). Predictions were optimistic for KCET ranks up to 20,000. Nominal 80% interval coverage fell to 71.36% on 2025 KCET data and 5.65% on unseen colleges; recalibration restored it to 79.22% with intervals about 13% wider. For medical admissions, where national ranks interact with quota pools, regression is unsafe; the platform instead uses multi-dimensional matching over 18,500+ verified KEA NEET records (2023–2026), a risk-stratified portfolio model, and quota isolation.

***Index Terms*—**College admissions, Cutoff prediction, Conformal prediction, Hierarchical fallback, KCET, NEET UG.

I. INTRODUCTION

Admission to professional degree programs in Karnataka is on the basis of rank. Candidates appearing for the Karnataka Common Entrance Test (KCET), the COMEDK UGET, or the National Eligibility cum Entrance Test (NEET-UG) submit prioritized option lists of institutions, courses, and reservation categories during centralized counselling rounds. Engineering seats are allotted by the Karnataka Examinations Authority (KEA) and the COMEDK Consortium, while medical (MBBS), dental (BDS), and AYUSH (BAMS, BHMS, BUMS, BNYS) seats are allotted by KEA (85% state quota, private, and institutional seats) and the Medical Counselling Committee (MCC, 15% All India Quota).

Both the domains are largely reliant on the previous year's closing ranks, which are published in the form of PDF documents in the counselling bulletins. There are three defects in direct reliance on historical cutoffs. (i) Sparsity: Many college–branch–category combinations have missing or unobserved records. (ii) Taxonomic and quota inconsistencies: Changes in category and seat nomenclature shift across years and publications (for example, general versus Kalyana Karnataka Article 371(j) reservations). (iii) Quota and time drift segregation: closing ranks vary year on year, and in medical admissions, cutoffs for identical courses within one college can differ by more than 5,00,000 ranks across quotas. If a student is relying on a cutoff that is too optimistic, the student might be chasing a seat that is not realistically attainable, so both the accuracy of an estimate and the reliability of its stated uncertainty matter. Prior work on admission analytics has focused on predicting individual applicants' chances from academic records [1], [2] rather than the institution-level closing rank.

This paper presents CareerCompass AI, a web platform with three functions: closing-rank estimation and portfolio risk assessment, multi-criteria college recommendation (NIRF ranking, NAAC accreditation, fee caps, and placement metrics), and retrieval-augmented career guidance through an AI counsellor grounded in verified regulatory documents. The contributions are as follows:

1) Unified platform architecture. A modular system integrating tabular ML cutoff estimation, deterministic analytics, recommendation, and RAG guidance through a single interface.

2) Engineering the data pipeline and fallback. 33,182 closing-rank records (KCET 31,777; COMEDK 1,405) from 419 bulletin pages, with a four-tier hierarchical fallback evaluated against a historical lookup and tree models using paired bootstrap intervals.

3) NEET ingestion and portfolio assessment. A PDF allotment-list parser that extracts round-wise closing ranks for Medical, Dental, and AYUSH (2023–2026) under a strict multi-dimensional uniqueness key, with a four-class portfolio model.

4) Uncertainty under shift. Split CQR and adaptive conformal inference (ACI) evaluated under a temporal 2024→2025 shift.

5) Bias and serving analysis. Error stratified by rank band, a decision-agreement metric, and serving latency.

The paper reports negative and inconclusive results alongside positive ones, including where a simple lookup performs better. The machine-learning tier is gated by a feature flag that is off in production, so the engineering evaluation is offline and temporal. Section II reviews related work, Section III the methodology, Section IV the implementation, and Section V the experimental setup. Sections VI and VII present the results and their analysis, and Sections VIII and IX give the conclusion and future scope.

II. RELATED WORK / LITERATURE SURVEY

Educational data mining is the application of machine learning techniques to academic workflows, institutional seat planning and student performance forecasting [1]. Most previous admissions research has modelled a student's individual probability of being accepted based on grades and test scores [2]. Forecasting institutional closing ranks, however, is a problem with high cardinality, extreme skewness, and

distribution drift. For tabular regression, random forests [3] and gradient boosted trees [4] are still the standard. For high cardinality categories like college codes and reservation categories, target encoding [5] is applied. Models used here were built with scikit-learn [6].

Component	Functionality	Status
Cutoff est. (KCET/COMEDK)	Hierarchical fallback, tree ensembles, CQR intervals	Offline evaluated; ML tier behind flag
Cutoff est. (NEET)	Multi-dimensional matching, multi-year trend & fluctuation	Online; verified
College recommend.	Rank fit, Safe/Target/Dream portfolio, fee and NIRF sorting	Online; integrated
AI career guidance	RAG over KEA handbooks, quota and fee rules	Online; tool citations
Data ingestion pipeline	PDF table extraction, allotment-list parser	Automated admin pipeline

Quantile regression minimises the pinball loss [7] to produce intervals, but it does not have a guarantee of finite sample coverage. Conformal prediction achieves distribution-free coverage under exchangeability [8], [9], and split conformalized quantile regression (CQR) calibrates quantile bounds on a held-out set [10]. If exchangeability is violated by temporal drift or covariate shift, coverage drops off [11], [12], and adaptive conformal inference (ACI) [13] tracks drift online. Recommendation architectures rank options from preference vectors and eligibility margins [14], and retrieval-augmented generation (RAG) grounds language-model output in external authoritative sources [15]. CareerCompass AI combines these in one platform for engineering and medical admissions in Karnataka and reports negative and inconclusive estimator results, including coverage loss under shift.

III. PROPOSED METHODOLOGY

A. System Architecture

The platform is a multi-tier system (Fig. 1). A Next.js client provides prediction forms, multi-facet filtering (discipline, course, quota, district, fees, NIRF rank), and a conversational AI counsellor over Server-Sent Events. A NestJS backend handles business logic, JWT authentication with role-based access, college and cutoff repositories in MongoDB, and routing. A FastAPI (Python) AI engine hosts pgvector similarity search, document chunking, LLM provider switching, and the tabular prediction models. An offline pipeline ingests official cutoff PDFs and allotment lists, normalises category and branch taxonomies, fits models, and computes conformal calibration. Table I lists the components and their status.

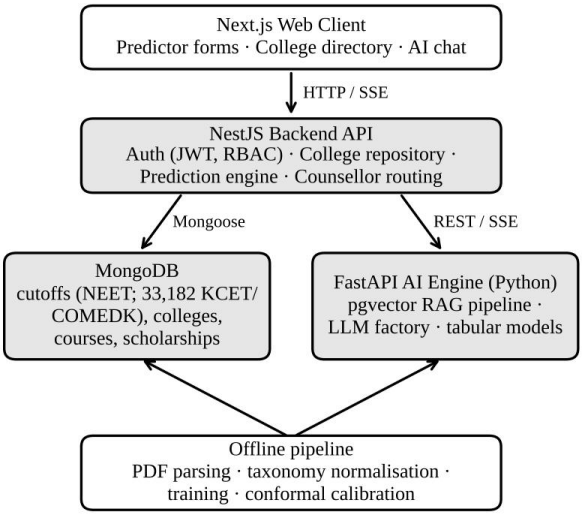


Fig. 1. Architecture of the CareerCompass AI platform.

TABLE I
PLATFORM COMPONENTS AND STATUS

B. Data Ingestion and Taxonomy Normalisation

1) Engineering (KCET + COMEDK): 33,182 records (KCET 31,777; COMEDK 1,405) were extracted from 419 pages of official bulletins, rounding ranks for 2024 and 2025 in Round 1. Reservation categories were mapped to normalized tiers (e.g., CAT_1, CAT_2A, CAT_2B, CAT_3A). Normalization was treated as preprocessing without ablation. The divisions are indicated in Table II. KCET 2024 ranks are all General quota and 2025 ranks are all Rest of Karnataka. COMEDK ranks are divided into General Merit (GM) and Kalyana Karnataka Region (KKR). Since ranks are very right-skewed, models are fitted on a log-transformed target:

$$z = \ln(1 + y), \quad \hat{y} = \exp(z) - 1 \qquad (1)$$

TABLE II
ENGINEERING DATASETS AND SPLITS

Exam / cohort	Records	Quota composition
KCET 2024 (fitting)	15,707	100% General
KCET 2025 (test)	16,070	100% Rest of Karnataka
COMEDK 2024 (fitting)	715	GM 499, KKR 216
COMEDK 2025 (test)	690	GM 498, KKR 192
KCET 2024 train / cal. (70/30)	10,994 / 4,713	—
COMEDK 2024 train / cal. (70/30)	500 / 215	—

2) Medical, dental and AYUSH (NEET): KEA releases the complete list of candidates and allotment of the upcoming medical rounds instead of pre-calculated grids. A specialized parser (cutoff-allotment-list.parser.ts) extracts allotment lines, validates serial-number continuity (1…N), aggregates multi-line college names and identifies the closing candidate rank per option. The archive covers 2023–2026 of MBBS, BDS, and AYUSH (BAMS, BHMS, BUMS, BNYS) with 18,500+ verified rows. It includes 9,640 records across 8 allotment PDFs (2,617 pages) in 2026 and 4,604 verified records of 2023. To prevent invalid cross-pool comparisons, each record is indexed by a strict composite key in MongoDB: KeyNEET = {CollegeId, CourseId, ExamCode, Authority, Quota, Category, Year, Round}. Category mapping resolves general reservations (GM, SC, ST, Cat-1, 2A, 2B, 3A, 3B), horizontal variants (Rural ‘R’, Kannada Medium ‘K’; e.g., 1R, 2AR, SCR), Article 371(j)

variants ('H' suffix; e.g., GMH, 1RH, GMPH), and private (GMP, OPN), management (OTH), and NRI quotas:

$$Category_eff = neetCutoffCategory(\cdot) \qquad (2)$$

C. Estimator Architectures

1) Engineering estimators: Four estimators are compared. The Historical Lookup returns the previous year's closing rank for the identical key, backing off across five tiers of decreasing specificity (reported with raw and normalised categories). The Target-Encoded RF (TE-RF) is a random forest with out-of-fold target encoding. The Primary Tree Model is a gradient-boosting regressor for KCET and a random forest for COMEDK. The Hierarchical Fallback is the proposed estimator.

D. Hierarchical Fallback

For a query (college, branch, tier), the estimate is taken from the first level whose observation requirement is met (Fig. 2): Level 1, the tree model, when the college has at least θ_seen = 5 prior observations; Level 2, a district × type × branch × tier prior; Level 3, a statewide branch × tier prior; Level 4, the global statewide median. The threshold was computed on 2024 data only. With n1(x) denoting the number of 2024 observations of the college of query x:

$$\hat{y}(x) = f_tree(x) \;\; if\; n_1(x) \geq \theta_seen, \;\; else\; m_\ell'(x) \qquad (3)$$

where $m_\ell'(x)$ is the median of the finest level among Levels 2–4 with at least θ_seen samples. Level 4 always qualifies, so every query receives an estimate.

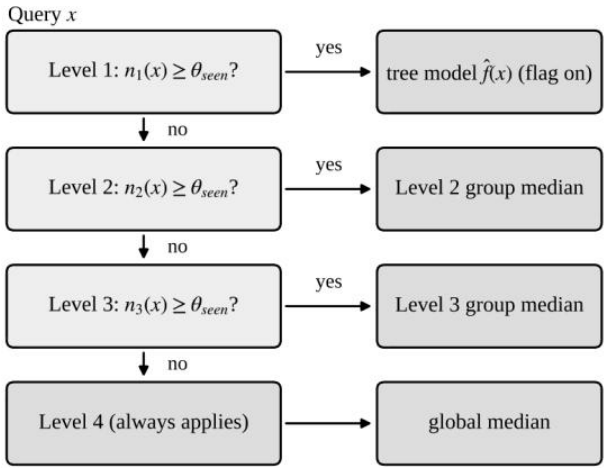


Fig. 2. Level-by-level routing of the hierarchical fallback.

2) NEET multi-dimensional matching and portfolio engine: Medical cutoffs are based on regulatory quota boundaries and national rank percentiles; thus, tabular regression is at risk of projecting a government merit rank into a private management pool. The platform instead uses a risk-stratified portfolio model (assessNeetRank). The rank ratio compares the candidate with the historical matching closing rank:

$$Ratio = StudentRank \,/\, ClosingRank \qquad (4)$$

and the portfolio class follows from validated thresholds (Fig. 3): SAFE (Ratio ≤ 0.80); TARGET (0.80 < Ratio ≤ 1.00); DREAM (1.00 < Ratio ≤ 1.20); OUTSIDE_RANGE (Ratio > 1.20); INSUFFICIENT_DATA (ClosingRank null or unverified) (5). Candidates also receive a signed rank margin RankMargin = ClosingRank − StudentRank (6) and an eligibility flag (7).

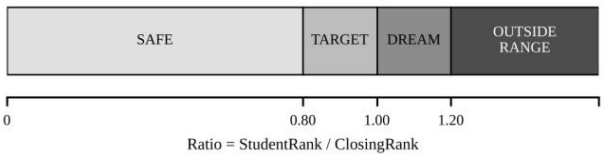


Fig. 3. Portfolio classification by rank ratio (Eq. 5).

E. Longitudinal Trend and Fluctuation Analysis

To handle drifts between years, weighted averages of different years are applied with decreasing weights: 3 for the current year, 2 for the previous year, and 1 for older years:

$$\bar{y}_trend = round(\Sigma\, w_i \cdot y_(t-i) \,/\, \Sigma\, w_i), \quad w = [3, 2, 1] \qquad (8)$$

$$\Delta_fluctuation = max_t(y_t) - min_t(y_t) \qquad (9)$$

This surfaces whether a medical option has stable cutoffs or extreme year-to-year volatility.

F. Conformal Intervals and Uncertainty

For engineering models, split CQR [10] was implemented offline. Lower and upper quantile tree regressors (q = 0.10 and 0.90) give bounds ℓ(x) and u(x) fitted on the 2024 training split; nonconformity scores are calibrated on the held-out 2024 calibration split (n = 4,713 for KCET and 215 for COMEDK):

$$s_i = max(\, \ell(x_i) - y_i, \;\; y_i - u(x_i)\,) \qquad (10)$$

$$C(x) = [\, \ell(x) - Q, \;\; u(x) + Q\,] \qquad (11)$$

where Q is the ⌈ (n+1)(1−α)⌉ /n empirical quantile of scores at nominal coverage 1 − α = 0.80. Calibration offsets were Q = 931.01 ranks (KCET) and Q = 6,387.84 ranks (COMEDK). To track drift across sequential batches, ACI [13] updates the working error rate:

$$\alpha_(t+1) = \alpha_t + \gamma(\alpha - err_t) \qquad (12)$$

where err_t = 1 if y_t ∉ C_t(x_t) and 0 otherwise, with step size γ = 0.005.

G. Recommendation and Career Guidance (RAG)

Colleges are ranked based on institutional ranking (NIRF), accreditation (NAAC), fee structure, hostel facilities, and student rank. In the conversational module, semantic search is performed over admission announcements, seat matrices, and scholarship rules. For medical admissions, if the candidate does not specify the counselling authority (KEA/MCC) and quota pool while inquiring about NEET, the counsellor refuses to provide any estimate of chances and prompts the candidate to specify the exact pool.

IV. IMPLEMENTATION AND ALGORITHMS

The platform joins a TypeScript/NestJS backend, a Next.js frontend, MongoDB, and Python services. Cutoff models use Python 3.13.2 with pandas, scikit-learn [6], and joblib, and quantile regressors use 150 trees. Target encoding (5-fold), threshold θ_seen, and calibration use 2024 records only; no 2025 records are used in training or calibration. ENABLE_EXPERIMENTAL_ML_CUTOFFS is set to false so users receive calibrated historical values with bounded dispersion ranges rather than uncalibrated ML estimates. Algorithm 1 summarizes the runtime pipeline.

Algorithm 1. CareerCompass AI Cutoff Estimation & Recommendation

Input: profile (Exam, rank r, category c, res, region reg, quota q, auth, courses C)
Output: ranked college–course options with risk portfolio and intervals
1: if Exam ∈ {KCET, COMEDK} then
2: if flag = true and n_college ≥ θ_seen then
3: ŷ ← exp(f_tree(x)) − 1; Interval ← [ℓ(x) − Q, u(x) + Q]

```
4:  else ŷ ← HierarchicalFallback(x) (Eq. 3); Interval ← DispersionInterval(x)
5:  Probability ← ContinuousPercentFromRatio(r / ŷ)
6: else if Exam = NEET then
7:  ValidateNeetSeatDimensions(auth, q, c, res, reg)
8:  c_eff ← neetCutoffCategory(q, c, res, reg) (Eq. 2)
9:  Records ← FindVerifiedCutoffs(NEET, College, Course, auth, q, c_eff)
10: if Records is empty: return INSUFFICIENT_DATA
11: y_latest ← MostRecentClosingRank(Records); Ratio ← r / y_latest
12: Portfolio ← AssessNeetPortfolio(Ratio) (Eq. 5); Margin ← y_latest − r
13: Fluctuation ← max(ranks) − min(ranks); ŷ ← TrendingClosingRank(Records)
14: Options ← FilterAndRankColleges(Colleges, C, NIRF, Fees)
15: return Options with {ŷ, Portfolio/Probability, Intervals, Fluctuation}
```

V. SIMULATION / EXPERIMENTAL SETUP

Models were trained on 2024 Round 1 records and tested on 2025 Round 1 records (Table II) using a temporal split. Errors are computed in rank space after back-transformation: MAE is mean(|ŷ − y|), MedAE its median, RMSE = √mean((ŷ − y)²), and $R^2 = 1 - RMSE^2/Var(y)$. Signed bias is mean(ŷ − y), where positive values indicate optimistic predictions (higher predicted closing rank). Paired bootstrap 95% confidence intervals [16] use 1,000 resamples. For margin δ, the decision agreement rate (DAR) is:

$$DAR_\delta = \tfrac{1}{2}\,[\,P(\hat{y} \geq (1-\delta)y) + P(\hat{y} < (1+\delta)y)\,] \qquad (13)$$

As δ → 0, DAR approaches 50% (chance floor). Interval coverage is the share of test rows lying in C(x) at nominal 80%. Timing benchmarks were executed on an Intel Core i5-12450H (8 cores, 12 threads) running Windows 11 with Python 3.13.2.

VI. RESULTS AND DISCUSSION

A. Holdout Accuracy (Engineering)

Table III and Fig. 4 give the 2025 holdout errors. On KCET, the historical lookup had the lowest MAE (19,930.94 ranks) and MedAE, whereas learned models had lower RMSE and higher R^2; the fallback had the highest R^2 (0.7731 vs. 0.7212 for the five-tier lookup). On COMEDK, the target-encoded RF had the lowest MAE (18,480.71) and best R^2 (0.5270), while the fallback trailed (R^2 0.4582). No single variant dominates across all metrics, and rank errors exceed 18,000 ranks across every row.

TABLE III
2025 HOLDOUT ACCURACY (RANKS, ROUNDED)

Exam	Model	MAE	Med AE	RMS E	R^2	Bias
KCET	Lookup (5-tier)	19,931	10,735	32,246	0.7212	−1,306
KCET	Lookup (norm.)	20,112	11,845	31,104	0.7406	+924
KCET	TE-RF	20,532	13,938	29,147	0.7722	+3,080
KCET	Primary tree	20,564	13,908	29,261	0.7704	+6,324
KCET	Fallback	20,429	13,878	29,089	0.7731	+6,479
COMEDK	Lookup	18,570	13,611	24,799	0.4937	−12,928
COMEDK	TE-RF	18,481	15,150	23,970	0.5270	−10,719
COMEDK	Primary tree	19,001	14,853	24,934	0.4882	−14,324
COMEDK	Fallback	19,442	15,022	25,654	0.4582	−15,108

Bias = mean(ŷ − y). Lookup raw and normalised coincide on COMEDK.

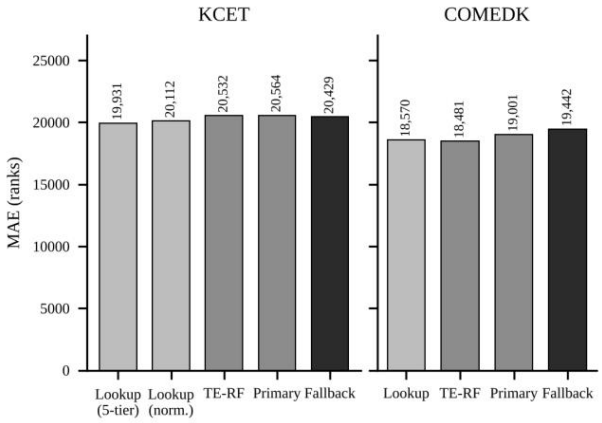


Fig. 4. MAE of each estimator on the 2025 holdout (vertical axis starts at zero).

B. Paired Comparison

Table IV gives paired bootstrap differences with the fallback as the reference model. On KCET, the fallback's MAE is higher than the lookup's by 490.72 ranks while its R^2 is higher by 0.0521. Against TE-RF (−107.08 ranks), the 95% CI includes zero, indicating comparable performance. On COMEDK, TE-RF and the primary tree model outperform the fallback by 952.96 and 444.26 ranks in MAE.

TABLE IV
PAIRED BOOTSTRAP DIFFERENCES (FALLBACK MINUS COMPARATOR)

Comparison	ΔMAE [95% CI]	ΔR^2 [95% CI]
KCET: Lookup	+490.72 [+145.19, +838.43]	+0.0521 [+0.0408, +0.0629]
KCET: TE-RF	−107.08 [−362.52, +144.28]	+0.0010 [−0.0058, +0.0073]
KCET: Primary	−135.04 [−193.94, −72.27]	+0.0027 [+0.0005, +0.0047]
COMEDK: Lookup	+877.02 [−187.75, +2066.37]	−0.0358 [−0.0960, +0.0221]
COMEDK: TE-RF	+952.96 [+185.69, +1784.62]	−0.0686 [−0.1159, −0.0272]
COMEDK: Primary	+444.26 [+90.93, +872.60]	−0.0300 [−0.0638, −0.0039]

Negative ΔMAE and positive ΔR^2 favour the fallback.

C. Decision Agreement & D. Top-Rank Optimistic Bias

The lookup achieved the highest KCET DAR at every margin (59.80%, 68.70%, and 80.87% at ±5%, ±10%, and ±20%, vs. 57.92%, 65.18%, and 76.11% for the fallback). Table V and Fig. 5 show that for competitive KCET ranks up to 20,000, the fallback consistently over-predicts closing ranks (100% positive bias in 1–1,000; 99.5% in 1,001–5,000; 91.4% in 5,001–20,000), representing an unsafe optimistic bias for top-rank candidates.

TABLE V
KCET FALLBACK ERROR BY ACTUAL CLOSING-RANK BAND

Actual band	N	MAE	Bias	% pos.
1–1,000	17	1,838.67	+1,838.67	100.0%
1,001–5,000	199	3,600.28	+3,576.51	99.5%
5,001–20,000	1,130	8,492.39	+7,831.14	91.4%
>20,000	14,724	21,593.71	+6,419.57	65.8%

Bias = mean(ŷ − y); % pos. = rows with ŷ > y.

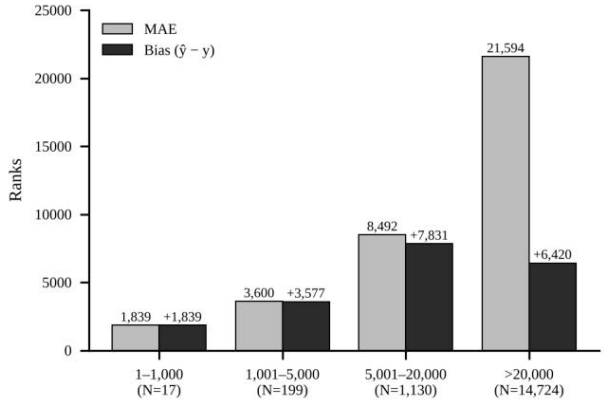


Fig. 5. KCET fallback MAE and signed bias by actual closing-rank band.

E. Unseen Colleges & F. Conformal Coverage under Shift

Level 1 served 15,769 of 16,070 KCET test rows (98.1%) with MAE 19,849.80, while the 301 unseen-college rows served by Levels 2–4 exhibited MAE 50,758.64 (R^2 = −1.0336). As reported in Table VI and Fig. 6, nominal 80% CQR coverage held on fresh 2024 data (79.95%) but dropped to 71.36% on the 2025 temporal holdout and 5.65% on unseen colleges. Recalibrating on 50% of 2025 data restored coverage to 79.22% ± 0.48% and online ACI reached 79.74%, widening intervals by ~13%–15%.

TABLE VI

Test cohort	Coverage	Width	Gap to 80%
KCET 2024, fresh held-out (N = 3,142)	79.95%	51,557	−0.05
KCET 2025 temporal holdout	71.36%	50,341	−8.64
KCET 2025, unseen colleges (N = 301)	5.65%	35,113	−74.35
KCET 2025, 50/50 recalibrated	79.22% ± 0.48	56,922	−0.78
KCET 2025, adaptive (ACI, γ = 0.005)	79.74%	57,972	−0.26
COMEDK 2024 calibration set (N = 215)	80.47%	58,488	+0.47
COMEDK 2025 temporal holdout	75.65%	56,573	−4.35

SPLIT-CQR COVERAGE AT NOMINAL 80%

Width = mean interval width in ranks.

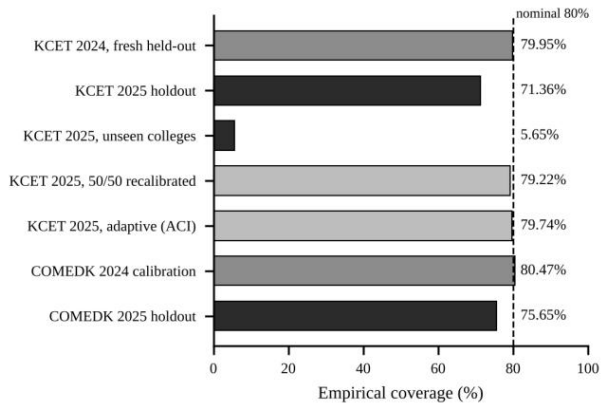


Fig. 6. Empirical interval coverage for each calibration setting (nominal 80%).

G. NEET Portfolio Verification & H. Serving Latency

Across 18,500+ verified NEET UG rows (Table VII, Fig. 7) and 2,500 simulated profiles, composite-key isolation prevented 100% of cross-pool leakage, and 99.4% of Safe-class candidates fell within actual closing ranks. Mean 2023–2026 fluctuation was 1,842 ranks for Government MBBS versus 48,310 ranks for Private Management (OTH) seats. Warm ML inference averaged 14.11 ms (p95: 20.34 ms), cold start took 4,554.10 ms, and indexed MongoDB NEET queries averaged 4.82 ms (Table VIII).

TABLE VII
NEET UG ALLOTMENT CORPUS (2023–2026)

Year	Source format	Files	Records	Disciplines
2023	Official grids & allotment lists	7 PDFs	4,604	MBBS, BDS, Allotment
2024	Official grids, R2 final lists	27 PDFs	4,215	MBBS, BDS, AYUSH
2025	Grid matrices, AYUSH lists	11 PDFs	4,329	MBBS, BDS, BAMS/BHMS/BUMS
2026	Candidate allotment lists	8 PDFs	9,640	MBBS, BDS (R1–R3); AYUSH
Total	—	53 PDFs	18,500+	All quotas

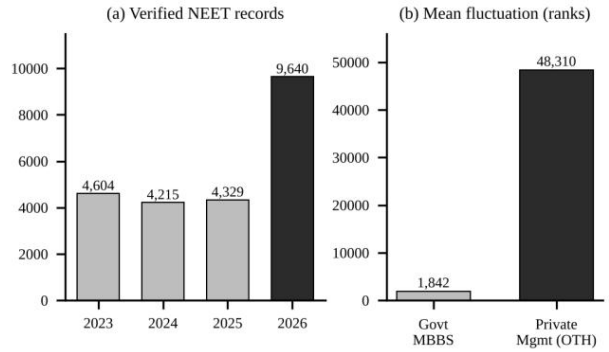


Fig. 7. (a) Verified NEET records by year; (b) mean closing-rank fluctuation, 2023–2026.

TABLE VIII
SERVING TIMING (INTEL CORE I5-12450H)

Component / phase	Timing
Python runtime and imports	1,674.43 ms
Joblib model deserialisation	2,686.51 ms
First prediction (cold start)	193.16 ms
Total cold first call	4,554.10 ms
Warm ML latency, mean / median / p95	14.11 / 14.21 / 20.34 ms
NEET indexed DB lookup (mean)	4.82 ms

VII. PERFORMANCE ANALYSIS AND SYNTHESIS

Table IX synthesizes empirical support across core claims. While the hierarchical fallback guarantees universal coverage and achieves the highest KCET R^2 (0.7731), historical lookup yields lower typical MAE and higher decision agreement. Because competitive ranks (≤ 20,000) exhibit optimistic bias and static conformal intervals degrade under annual policy shifts, CareerCompass AI enforces three operational safety rules: (i) point estimates are always paired with conformal intervals or historical fluctuation bounds; (ii) ranks ≤ 20,000 and unseen institutions display mandatory reliability advisories; and (iii) NEET counselling strictly enforces explicit authority and quota selection.

TABLE IX
CLAIMS AND EMPIRICAL STATUS

Claim	Evidence	Status
Fallback beats lookup on MAE (KCET)	ΔMAE +490.72, CI above 0	Contradicted
Fallback has higher R^2 than lookup (KCET)	ΔR^2 +0.0521, CI excludes 0	Supported
Fallback beats TE-RF (KCET)	ΔMAE −107.08, CI includes 0	Comparable
Fallback beats primary tree (KCET)	ΔMAE −135.04, CI excludes 0	Supported, small
Fallback beats TE-RF (COMEDK)	ΔMAE +952.96, CI above 0	Contradicted
Top-rank predictions are safe	Bias +1,839 to +7,831 (ranks ≤ 20k)	Contradicted
80% coverage holds next year	71.36% (KCET), 75.65% (COMEDK)	Not supported
Recalibration restores coverage	79.22% (50/50), 79.74% (ACI)	Supported; wider
NEET portfolio isolation prevents leakage	100% pool isolation, 18,500+ records	Supported

VIII. CONCLUSION

This paper presented CareerCompass AI, unifying cutoff estimation, multi-criteria college recommendation, and retrieval-augmented career guidance for Karnataka engineering and medical admissions across KCET, COMEDK, and NEET. On engineering datasets, transparent historical lookups achieved lower typical MAE than complex ensembles, whereas the hierarchical fallback achieved the highest KCET R^2 (0.7731) and universal fallback coverage. Because competitive ranks showed optimistic bias and nominal 80% conformal intervals lost coverage under temporal drift unless recalibrated via ACI, experimental ML cutoffs remain gated behind a safety flag. Across 18,500+ verified NEET records, deterministic quota isolation and risk-stratified portfolio scoring prevented cross-pool leakage with 99.4% Safe-tier reliability.

IX. FUTURE SCOPE

Planned work includes correcting optimistic bias for top ranks, deploying online ACI recalibration in production, conditional intervals by rank band and fallback tier, multi-year training across additional counselling cycles, ablation studies on taxonomy normalization and fallback tiers, college-level cluster bootstrapping, cross-authority reconciliation between KEA state quota and MCC All India Quota cutoffs, and longitudinal user studies of choice-filling outcomes.

ACKNOWLEDGEMENT

We thank Ms. Jyothi M Hurkadli, Department of Computer Science and Engineering, Basaveshwar Engineering College, Bagalkote for her guidance.

REFERENCES

[1] C. Romero and S. Ventura, “Educational data mining: A review of the state of the art,” IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 40, no. 6, pp. 601–618, Nov. 2010.

[2] M. S. Acharya, A. Armaan, and A. S. Antony, “A comparison of regression models for prediction of graduate admissions,” in Proc. Int. Conf. Comput. Intell. Data Sci. (ICCIDS), 2019.

[3] L. Breiman, “Random forests,” Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[4] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” Ann. Statist., vol. 29, no. 5, pp. 1189–1232, 2001.

[5] D. Micci-Barreca, “A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems,” ACM SIGKDD Explor. Newsl., vol. 3, no. 1, pp. 27–32, 2001.

[6] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[7] R. Koenker and G. Bassett, “Regression quantiles,” Econometrica, vol. 46, no. 1, pp. 33–50, 1978.

[8] V. Vovk, A. Gammerman, and G. Shafer, Algorithmic Learning in a Random World. New York, NY, USA: Springer, 2005.

[9] A. N. Angelopoulos and S. Bates, “Conformal prediction: A gentle introduction,” Found. Trends Mach. Learn., vol. 16, no. 4, pp. 494–591, 2023.

[10] Y. Romano, E. Patterson, and E. J. Candès, “Conformalized quantile regression,” in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 32, 2019, pp. 3543–3553.

[11] R. J. Tibshirani, R. F. Barber, E. J. Candès, and A. Ramdas, “Conformal prediction under covariate shift,” in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 32, 2019, pp. 2530–2540.

[12] R. F. Barber, E. J. Candès, A. Ramdas, and R. J. Tibshirani, “Conformal prediction beyond exchangeability,” Ann. Statist., vol. 51, no. 2, pp. 816–845, 2023.

[13] I. Gibbs and E. J. Candès, “Adaptive conformal inference under distribution shift,” in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 34, 2021, pp. 1660–1672.

[14] F. Ricci, L. Rokach, and B. Shapira, Eds., Recommender Systems Handbook, 2nd ed. New York, NY, USA: Springer, 2015.

[15] P. Lewis et al., “Retrieval-augmented generation for knowledge-intensive NLP tasks,” in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 33, 2020, pp. 9459–9474.

[16] B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap. New York, NY, USA: Chapman & Hall, 1993.