Comprehensive College Ranking System: A DEA-ANP Hybrid Model

Vedant Joshi (23280045)

Introduction

This project develops a **two-stage decision-support system** to help students prioritize engineering colleges using a combination of Data Envelopment Analysis (DEA) and Analytic Network Process (ANP). The system first screens colleges based on objective efficiency and then ranks the shortlisted options using subjective, student-centric preferences.

Problem Statement and Objectives

Many students select colleges based on fragmented information such as brand reputation, anecdotes, or single metrics like placement rate, which can lead to suboptimal decisions. There is a clear need for a structured, transparent, and customizable decision-support tool that integrates multiple academic, financial, logistical, and campus-life factors.

The **primary objective** of this project is to design and validate a hybrid DEA–ANP framework that:

- Screens a large set of colleges to identify **efficient institutions** using resource–output relationships.
- Ranks the shortlisted colleges based on **student-specific preference structures** captured through ANP.
- Provides **sensitivity analysis** to test the robustness of rankings under different weighting scenarios.

Data Description

The current implementation uses **synthetically generated data** representing 30 engineering colleges with realistic ranges for inputs and outputs. This allows testing of the full analytical pipeline without relying on confidential or hard-to-access institutional data.

DEA Input Variables (Resources)

The DEA stage models each college as a Decision-Making Unit (DMU) consuming multiple inputs to generate desirable outputs. The following five input variables are used:

- Faculty FTE: Full-time equivalent faculty strength.
- PhD Faculty Count: Number of doctoral-qualified faculty members.
- Hostel Beds: Available residential capacity for students.
- Infrastructure Score: Composite score for academic and non-academic facilities.
- Operating Expenditure: Annual operational cost of the institution.

DEA Output Variables (Outcomes)

To capture institutional performance and student outcomes, six output metrics are modelled.

- Placement Rate: Percentage of students successfully placed.

- Average Package: Mean salary of placed students.
- Research Publications: Annual research output.
- Student Satisfaction: Survey-based satisfaction index.
- Graduation Rate: Percentage of students graduating on time.
- Inverted Cutoff Rank: Transformed selectivity metric, restructured in the final design as a **constraint** rather than a DEA output.

The student's entrance examination rank is assumed to be 4,312 for the scope of this report and is used for **feasibility screening and rank-fit scoring** rather than raw DEA efficiency computation.

Methodology

The proposed framework follows a **two-stage hybrid decision-making process**, where DEA handles objective efficiency screening, and ANP handles subjective, preference-driven ranking.

Stage 1: Data Envelopment Analysis (DEA)

Model Specification

- Model Type: Input-oriented CCR (Charnes–Cooper–Rhodes) model.
- Objective: Maximize the ratio of weighted outputs to weighted inputs for each college subject to efficiency scores not exceeding 1.
- Efficiency Scores: Bounded between 0 and 1, where 1 indicates a technically efficient college on the DEA frontier.

For each college j, the efficiency score θ_j is conceptually defined as:

$$\theta_j = \frac{\text{weighted sum of outputs}}{\text{weighted sum of inputs}}$$

with the weights chosen optimally under standard DEA constraints.

Cutoff Rank Handling

- Cutoff rank is used as a **pre-filter constraint** to include only colleges that are realistically attainable for the given student rank.
- Rank fit is modelled separately as a **preference score** rather than an efficiency output.

DEA Selection

- Total colleges evaluated: 30.
- Efficient colleges with θ = 1.0: 10.
- The top 7 efficient colleges are retained for downstream ANP analysis to reduce cognitive and computational burden.

The shortlisted colleges (all with θ = 1.0) include IDs C24, C14, C29, C28, C10, C9, and C2, spanning both medium and large institutions.

Stage 2: Analytic Network Process (ANP)

ANP generalizes AHP by allowing **interdependencies** among criteria and feedback loops, which better reflect real-world college selection decisions.

Criteria Clusters and Weights

Five main clusters and their default global weights are defined as follows:

Cluster	Default Weight	Description (Sub-Criteria)
Logistics	7.9%	Distance from home, travel time, hostel availability.
Academic	24.4%	Branch availability, faculty–student ratio, curriculum relevance, rank-fit score.
Financial	13.7%	Annual fees, scholarships, payment flexibility.
Campus Life	13.7%	Safety, extracurricular activities, health facilities.
Reputation	40.3%	Alumni base, industry linkages, accreditations.

- Criteria clusters and weights

These weights represent a baseline balanced view where **reputation** is the most important cluster, followed by **academic** quality.

ANP Steps

The ANP process implemented in this project follows standard practice:

- Construction of pairwise comparison matrices using Saaty's 1–9 scale for both clusters and sub-criteria.
- Derivation of local priority vectors via the principal eigenvector method of each comparison matrix.
- Supermatrix construction, followed by weighted and limit supermatrix computation to derive long-run priority weights.
- Calculation of final priority scores for the 7 shortlisted colleges.

Consistency Validation

To ensure logically coherent judgments, the **Consistency Ratio (CR)** is computed for each pairwise comparison matrix.

$$CR = \frac{CI}{RI}, CI = \frac{\lambda_{\max} - n}{n - 1}$$

where RI is the Random Index and n is the matrix order. Matrices with CR ≥ 0.10 trigger a warning and require revision of judgments.

Rank Fit Scoring

The rank fit component links the student's rank (4,312) to college cutoff ranks via a **continuous scoring function**.

- If the student's rank is better than the cutoff, a percentile-based formula yields a high rank-fit score, with 1 indicating comfortably within the cutoff.
- If the student's rank is worse than the cutoff, an exponential decay penalty is applied, ensuring rank fit is never negative and decays smoothly as the gap widens.

Implementation Details

Technology Stack

The implementation is done in **Python**, using the following libraries:

- NumPy: Linear algebra and matrix operations.
- SciPy: Optimization routines (where needed).
- Pandas: Data ingestion, transformation, and tabular output management.
- PuLP: Linear programming solver for DEA efficiency computation.

The codebase is modular, with separate components for DEA modelling, ANP weight processing, consistency checking, rank-fit computation, and scenario-based sensitivity analysis.

Key Algorithmic Components

- DEA: For each DMU, a linear program is solved to maximize efficiency subject to DEA constraints and $\theta \leq 1$ for all units.
- ANP: Supermatrix operations iterate until convergence, producing stable priority vectors for clusters, sub-criteria, and alternatives.
- Consistency Ratio: A reusable function computes CI and CR, referencing standard RI values for matrix sizes.
- Rank Stability: A utility counts how often each college appears in the top 3 across multiple weighting scenarios.

Experimental Results

DEA Outcomes

The DEA stage reveals that 10 of the 30 colleges lie on the efficiency frontier, with scores $\theta = 1.0$. These colleges achieve high outputs relative to their resource consumption, across both medium and large size categories.

Seven of these efficient colleges are retained for detailed preference-based ranking in the ANP stage, ensuring that only **resource-efficient** institutions are compared for final selection.

ANP Final Ranking

The ANP framework yields a prioritized list of the 7 shortlisted colleges with the following final priority scores:

Rank	College	Priority Score	Notable Strengths
1	C29	0.1542	Strong financial position and excellent campus facilities.
2	C24	0.1531	Best logistics (closest to home) with strong reputation.
3	C14	0.1481	Very good campus life and solid academic environment.
4	C28	0.1448	Balanced performance across all major criteria.
5	C10	0.1421	Best rank fit, with a generous cutoff rank of 35,453.
6	C9	0.1394	Strong brand reputation and industry connections.
7	C2	0.1183	Good academic programs with moderate strength in other clusters.

For the given student rank, college C10 offers the **largest rank safety margin**, with the cutoff about 31,141 ranks worse than the student's rank, indicating a highly safe admission option.

Sensitivity and Scenario Analysis

To examine robustness, the project evaluates five different weighting scenarios:

- Balanced (default weights).
- Academic-focused (≈50% weight on academic criteria).
- Financial-focused (≈40% weight on financial criteria).
- Reputation-focused (≈60% emphasis on reputation).
- Campus life-focused (≈40% emphasis on campus life).

A **rank stability metric** tracks how often each college appears in the top three across these scenarios. Colleges C29, C24, and C14 consistently remain in the top three, suggesting that they are robust choices under diverse preference profiles.

Conclusion

This project demonstrates a **production-ready** hybrid DEA–ANP framework for engineering college selection that combines objective efficiency analysis with subjective preference modelling. All major methodological issues—such as misuse of cutoff rank, binary rank-fit scoring, and lack of consistency checks—have been addressed, and a comprehensive sensitivity analysis confirms the robustness of top recommendations.

The remaining step before real-world deployment is integration of high-quality institutional data from authoritative sources. Once this is in place, the system can provide transparent, customizable, and analytically sound decision support for 12th-grade students navigating complex college choices.