Ranking postsecondary institutions in the US

Team 16

Problem Description

Applications are a long process

Important step in education

Rankings can be subjective

Needs a holistic way of ranking universities

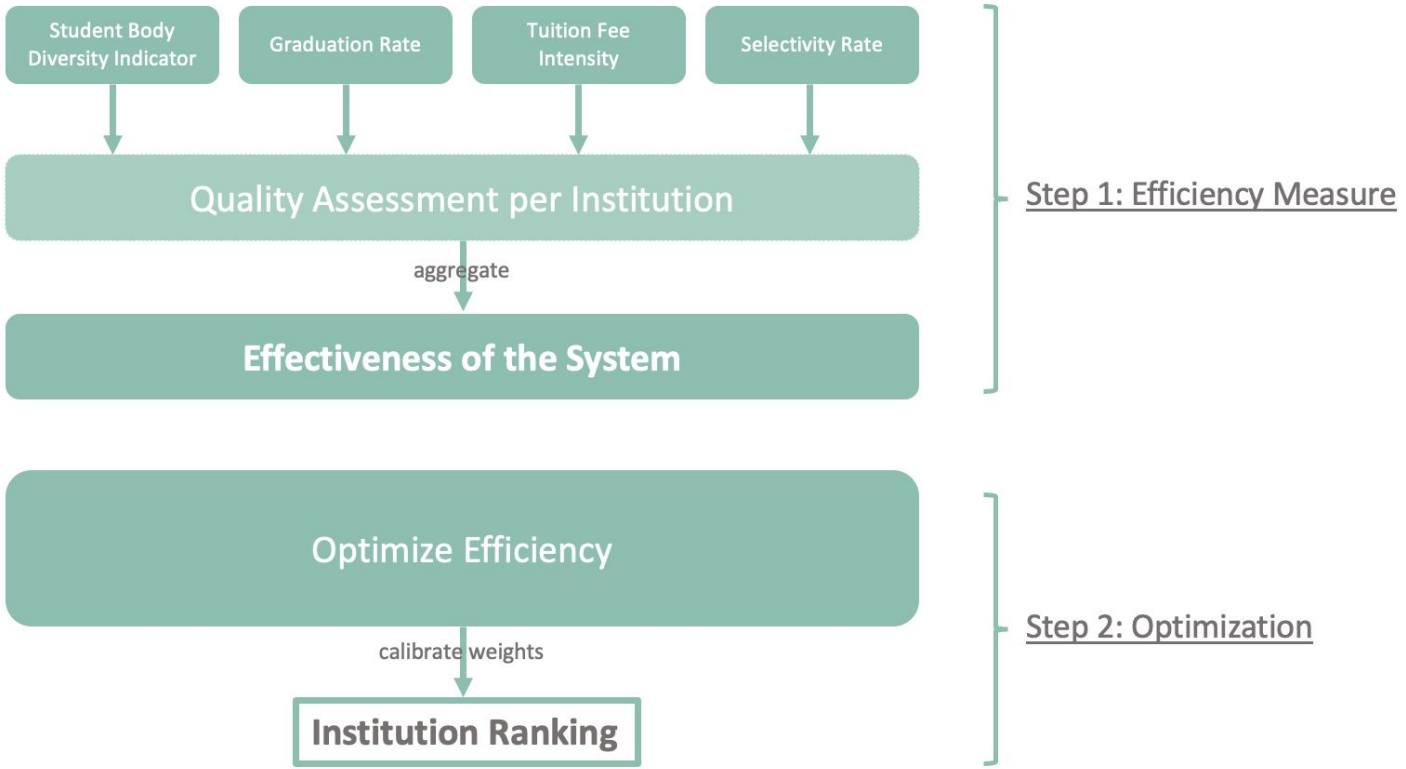
How can we rate universities in a more inclusive way?

Data Description

IPEDS

01	Fall Enrollment	• Ethnicities
02	Outcome Measures	• Graduation Rate after 2-4 years
03	Admissions and Test Scores	• Admission Rate (Proportion of admissions out of all applications)
04	Institutional Finances	• Total profit made on tuition fees

Method and Workflow



Efficiency Measure

What characterizes a good institution?

- diverse student body
- high graduate rate
- low tuition fees
- low selectivity rate

→ Assumptions based on **historical and current facts** about the US postsecondary system: high tuition fees, racial segregation etc.

Define quality metric per institution*:

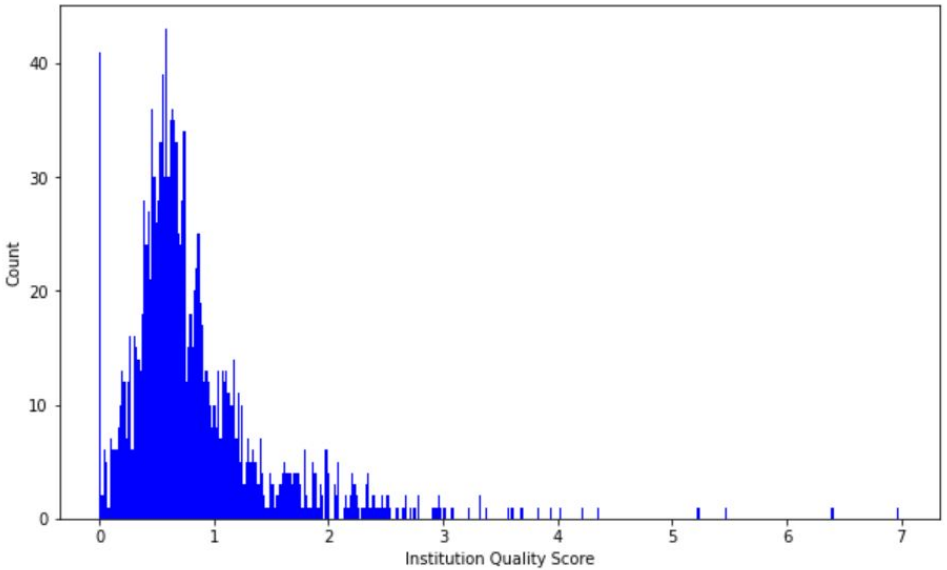
$$\frac{\text{student body diversity indicator} \cdot \text{graduation rate}}{\text{tuition fees intensity} \cdot \text{selectivity rate}}$$

- **Student body diversity:** based on JS distance measure
- **Tuition fees intensity:** based on CDF of university total profit from tuition fees

→ All these metrics belong to [0, 1] for comparability purposes

*see Appendix for more details

Key Insights - Efficiency Measure



- Initial “unoptimized” results ranking universities
- Distribution is **skewed towards low quality**
- High peak around 0: system is indeed inefficient

→ Optimize efficiency features

Optimization of the system

- Calibrate **proportion of ethnicity** per institution
- **Feature parametrization:** feed graduation and selection rates into a ML model
  - Causal inference for diversity onto those features

Objective: maximize overall effectiveness

Efficiency Measure

$$\left(p_1^{(q)*}, \ldots, p_e^{(q)*}\right)_{q\in\mathcal{Q}} = \underset{\left(p_1^{(q)}, \ldots, p_e^{(q)}\right)_{q\in\mathcal{Q}}}{\operatorname{Argmin}} -\frac{1}{|\mathcal{Q}|}\sum_{q\in\mathcal{Q}} \frac{\left(1-\Delta\left(p_1^{(q)}, \ldots, p_e^{(q)}\right)\right) g(q)}{\mathbf{F}_{\text{fee}}(q)s(q)}$$

  - Numerator: **maximize** the quality of ethnic diversity of the school, graduation rate
  - Denominator: **minimise** selection rates, tuition fees
- Parametrization of features is very computationally time-costly

Limitation & Future Work (1)

	Limitation	Future Work
Model Assumptions	**Limited number of features** chosen from dataset, using undergraduate data only	**Add more features from data:** faculty-student ratio, student aid, library budget and **classical features** (e.g. number of publications, awards)
Model Validation	Get **niche results** of unknown small institutions that do well on one particular metric	**Weight features** to calibrate model
Optimization Tractability	**Does not converge** in short period	**New objective function** (linearized version)

Future Work (2)

- **Customising Rankings:** Develop personalised rankings for students
- **Predict-Optimize Framework on Quality:** Find the most important features from a predictive model on system quality and re-parameterize model
- **Optimization Idea:** Weight each feature in the objective function and find the optimal weights
- **Holistic Rankings:** Weight our model with current ranking models (FT etc.) and formulate as a multi objective optimization
- **Price of Diversity:** With constrained optimization, we could estimate the impact of controlling diversity on quality of cohorts

Appendix - Efficiency Measure

$$\Delta(q) := \sqrt{\frac{\mathcal{D}(\left(\overline{p_1}, ..., \overline{p_e}\right)||(p_1^{(q)}, ..., p_e^{(q)})) + \mathcal{D}\left((p_1^{(q)}, ..., p_e^{(q)})||(\overline{p_1}, ..., \overline{p_e})\right)}{2}} \in [0, 1],$$

with $\mathbf{D}(p||q)$ the Kullback-Leibler divergence defined by:

$$\mathbf{D}(p||q) := \sum_x p(x) \log\left(\frac{p(x)}{q(x)}\right)$$

$$\forall q \in \mathcal{Q}, \qquad \mathbf{F}_{\text{fees}}(a_q) = \frac{1}{|\mathcal{Q}|} \sum_{q' \in \mathcal{Q}}^{n} \mathbf{1}_{(a_q \leq a_{q'})}.$$

- **Quality of an institution q:**

$$\frac{\text{student body diversity indicator} \cdot \text{graduation rate}}{\text{tuition fees intensity} \cdot \text{selectivity rate}}$$

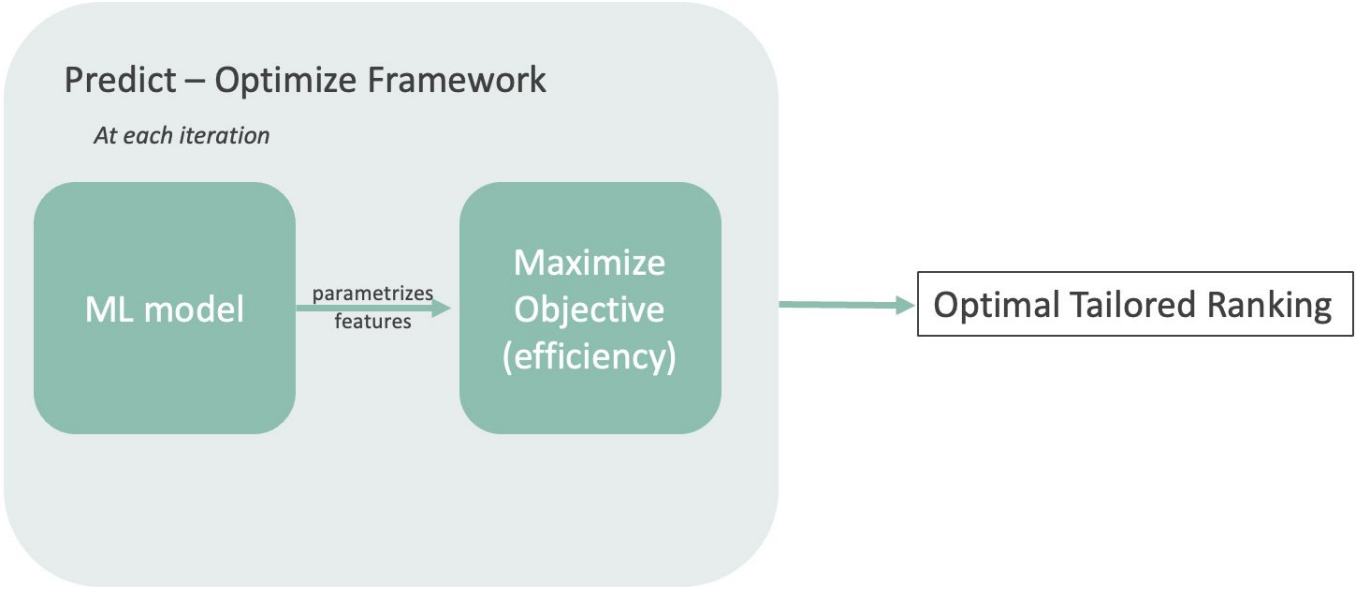
equivalent to

$$\pi(Q) := \frac{\left(1 - \Delta\left(p_1^{(q)}, ..., p_e^{(q)}\right)\right) g(q)}{\mathbf{F}_{\text{fee}}(q) s(q)}$$

- **Effectiveness of the education system:**

$$\pi(\mathcal{Q}) = \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} \pi(q).$$

Appendix - Optimization



Q&A

- price of diversity?
- mix of this ranking and current ones
- why did we try to work on this question?
    - take a global question rather than focusing on one specific part of the data
        - introduction of the paper: find strength and weakeness
            - build metrics to assess those strengths and weakeness

- if we add more time, what other parameters would have we added (from the dataset)?

faculty/student ratio (max that), student aid (max that), library budget / number of available literature (max)

from not the dataset: could add metrics for tailored choices (publications, famous alumni etc)