

EXPLORATORY ANALYSIS OF UNIVERSITY RANKING

By Yanxi Lu



Background:

- Different versions of rank.
- The Ranks change every year.

Objectives:

- Finding the correlation of following features to the **overall score** of universities.

Teaching Level

Research Level

International Outlook

Number Of Students

Citation

Average Salary Of Graduates

Endowment

Student To Faculty Ratio

Local GDP (State)

Ratio Of International Student

Student Diversity

Acceptance Rate

Rank	Name	Overall
1	Harvard University United States	100.0
2	Massachusetts Institute of Technology United States	80.2
3	Stanford University United States	76.2

• Data Sources

- Times Higher Education (THE)
- Wikipedia
- Google Scholar, etc.

• Datasets

- 156 Universities
- 12 Variables

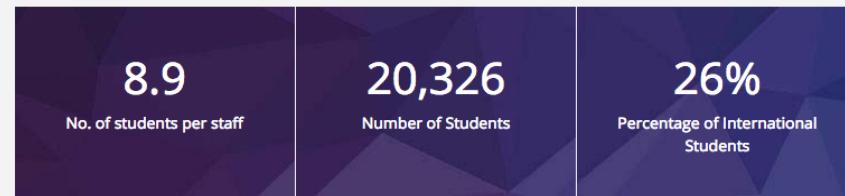
• Machine Learning Models

- Basic Regression
- Regression with Regularization
- Single Decision Tree
- Random Forest
- Gradient Boosting

• Indicator

- R Score
- Feature Importance, etc.

RANK	UNIVERSITY	COUNTRY	CITATIONS
1	Harvard University	USA	1734533
2	Stanford University	USA	1197114
3	Johns Hopkins University	USA	1152185
4	University of California Berkeley	USA	1106930



#23 Most Diverse Colleges in America

Carnegie Mellon University

4 Year • Pittsburgh, PA • ★★★★☆ 1,307

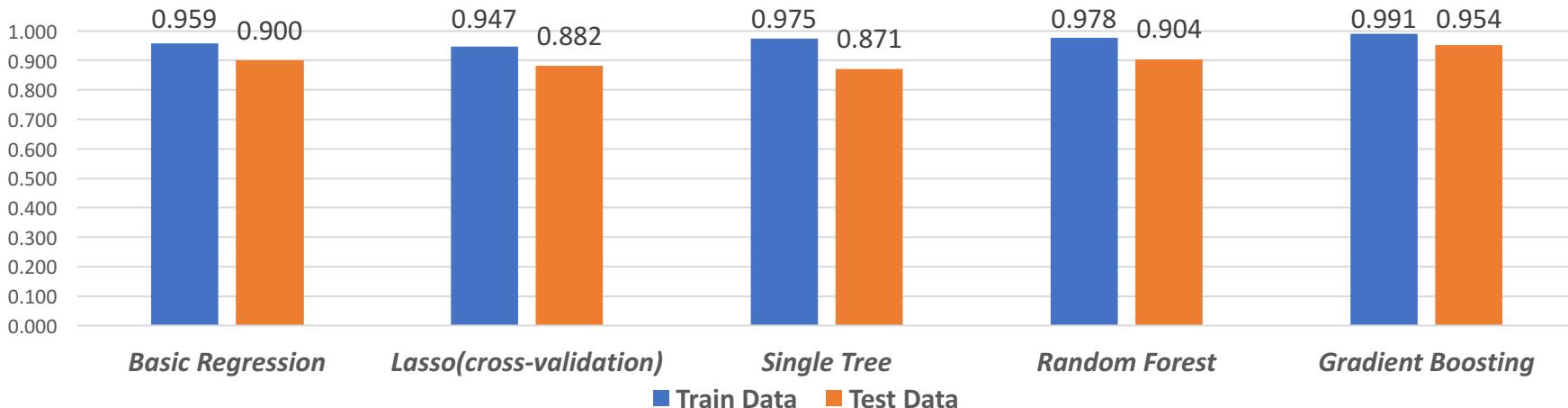
Alum: Carnegie Mellon is tough, no doubt about it. But it's also a place where you'll be pushed and surrounded [Read 1,307 Reviews](#)

A+ Overall Niche Grade • **A** Diversity • Acceptance Rate 24% • Net Price \$33,657 • SAT Range 1360-1540

Methodology

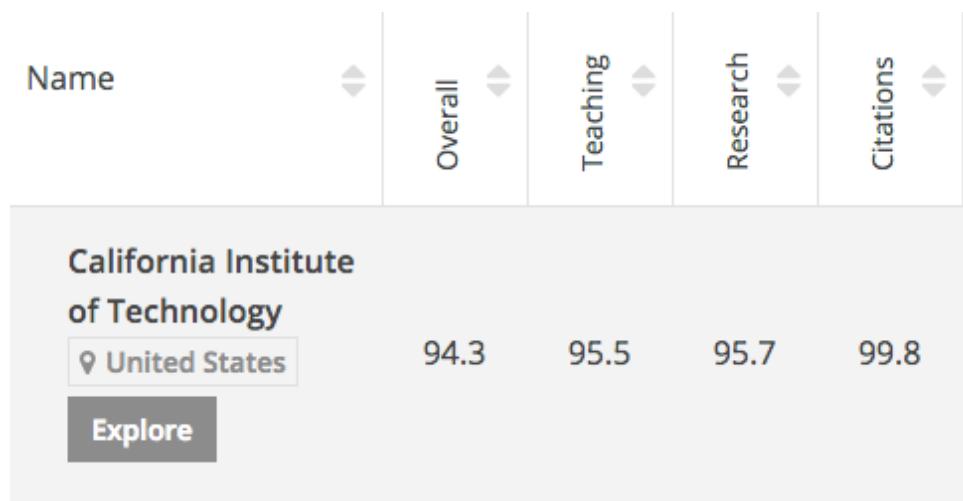
Results

- R Score:



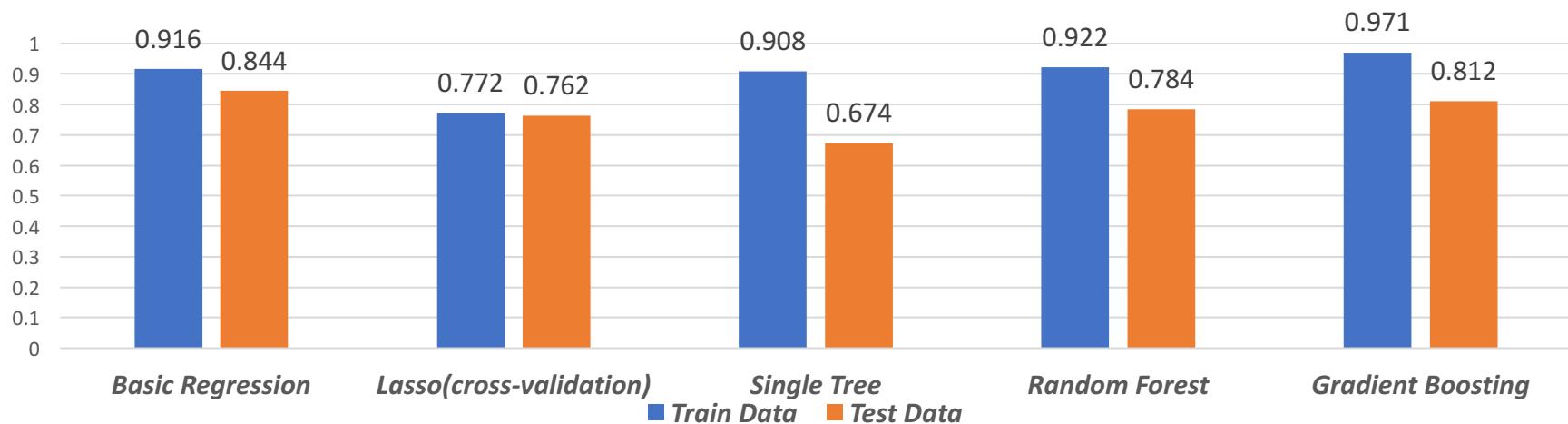
Feature Importance:

Research Level	14.60%
Teaching Level	13.20%
Citation Number	11.40%
International Outlook	10.71%
Number of Student	9.50%
Graduates Salary	8.85%
Local GDP	6.20%
International Student Ratio	6.63%
Student to Faculty Ratio	6.08%
Student Diversity	5.34%
Acceptance Rate	5.47%
Endowment	1.59%



Results

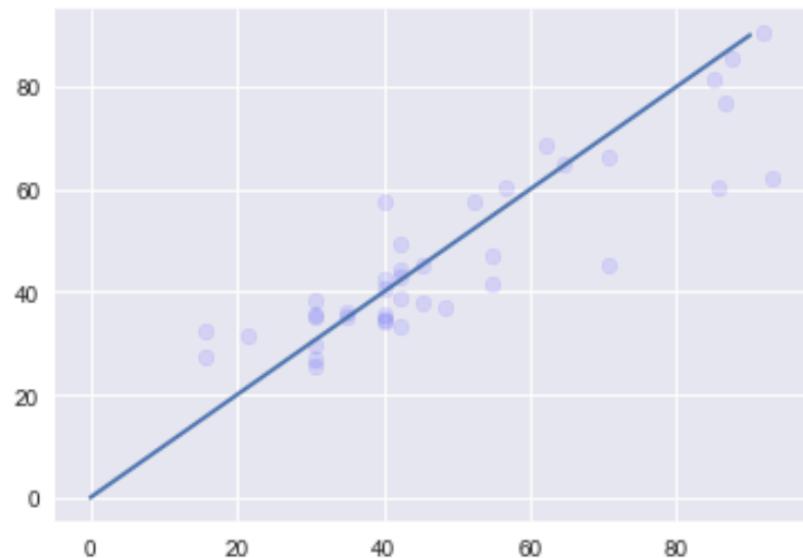
- R Score:



Features Importance:

Citation Number	22.84%
Number of Student	16.90%
Graduates Salary	12.26%
Local GDP	12.23%
International Student Ratio	8.01%
Acceptance Rate	7.91%
Student to Faculty Ratio	7.76%
Endowment	6.39%
Student Diversity	5.83%

Prediction vs. Test



Interpretation and Validation

- Gradient Boosting works best.
- Research Level is the key factor.
- Graduates Salary, Local GDP, Number of Student are highly correlated factors.

Interpretation above are partially supported by other sources:

- Wiki:

...Various rankings consider combinations of measuresVarious rankings mostly evaluating on institutional output by research...

- Wall Street Journal:

... methodology explores four key areas...

... The Resource area represents 30 percent of the overall ranking...

... The Engagement area represents 20 percent of the overall ranking...

...The Outcomes area represents 40 percent of the overall ranking...

...The Environment area represents 10 percent of the overall ranking...

- Getting More Data Points
- Screening Variables
 - International Student Ratio < Diversity Index
- Finding Causation

Future Work



Thank you!

