Case Study

A Machine-Learning Approach to University Rankings

University of Rochester
Spring 2019

The Opportunity

University of Rochester's Office for Global Engagement connected with the Rochester Data Science Consortium (RDSC) to develop a data-driven university ranking methodology, one that more accurately assesses the strengths and weaknesses of research institutions in the U.S. and worldwide.

The Challenge

The current university ranking methodologies are incredibly influential, however, still contain significant blind spots. Current methodologies often rely upon reputational data extracted from individual surveys performed across populations of interest, including educators. However, the outcomes of such studies often favor a few select institutions due to intrinsic human bias. Additionally, these survey results only provide a static snapshot of the academic landscape – a landscape that is better understood as a dynamic entity.

A more accurate (fairer) university ranking methodology would replace subjective survey questionnaires sent to select individuals, with tangible datasets (publications, grants, patents, and clinical trials) and incorporate temporal developments in the global academic research landscape.

"There is a need within higher education to validate and qualify rankings through the lens of comprehensive scholarly productivity. The application of machine learning methods to curated research databases holds promise as a potential solution." - Dale P. Hess, Ph.D., University of Rochester
Why RDSC?

RDSC scientists have the expertise needed to develop a sophisticated, data-driven approach to the question of university rankings: a method that incorporates more data, ongoing research trends, and better avoids the pitfalls of human bias reflected in popular university rankings. These novel data-driven methods include:

- Supervised and unsupervised machine learning
- Deep Learning
- Long-short term memory prediction models

Data Science in Action

In order to devise a more robust university ranking system, RDSC Scientists utilized the Dimensions dataset, a collaborative global data platform that catalogs publications, grants, patents, and clinical trials, as well as the connections among them.

- First, RDSC Scientists queried the Dimensions dataset across 22 fields of research to build global- and institution-level temporal research trends.
- Next, RDSC Scientists compared university research data, across schools and across time, to determine whether an institution is ahead, on par, or behind a given reference trend.
- Finally, RDSC Scientists leveraged long-short term memory networks (LSTM) to predict the future path of research trends. This technology allows the ranking model to anticipate the evolution of a given institution’s standing in a particular field, and factor this trajectory into a university’s overall ranking (in a given field).

Dimensions Dataset Overview

<table>
<thead>
<tr>
<th>Quick facts on Dimensions - the total record count and more</th>
<th>PUBLICATIONS</th>
<th>GRANTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of publication records</td>
<td>Publication references</td>
<td>Resulting publications</td>
</tr>
<tr>
<td>Number of funded grants</td>
<td>Publication citations</td>
<td>Resulting patents</td>
</tr>
<tr>
<td>Number of clinical trials</td>
<td>Supporting grants</td>
<td>Resulting clinical trials</td>
</tr>
<tr>
<td>Number of patents</td>
<td>Patent citations</td>
<td>PATENT</td>
</tr>
<tr>
<td>Number of records with Atometric data in Dimensions</td>
<td>Linked clinical trials</td>
<td>Patient references</td>
</tr>
<tr>
<td>Total number of documents in Dimensions</td>
<td>Linked publications</td>
<td>Publications references</td>
</tr>
<tr>
<td>…but what matters even more: number of links between these records</td>
<td>Supporting grants</td>
<td>Supporting grants</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patent citations</td>
</tr>
</tbody>
</table>

What’s Next?

University Rankings are, for better or worse, akin to fate for top research institutions, influencing everything from student populations, tuition fees, research dollars, faculty hires, strategic planning decisions, and more. Given the out-sized impact rankings have on higher education, devising a more accurate and data-driven ranking methodology is both vital and necessary; to provide a more transparent picture of the field, and to ensure the integrity of the field of higher education.

RDSC Scientists will continue to improve this university ranking methodology, comparing the results with publicly available rankings, and devising prescriptive analytics to better map actions and outcomes.

Want to learn more?
Visit ROCDataScience.com

©2019 Rochester Data Science Consortium