CIVITAS LEARNING & JISC
Helping Every Student Achieve
February 22, 2018
How Many Data Points Are In The Models for the UK Innovation Partners?

5 Innovation Partners
- University of Edinburgh
- University of West London
- University of East London
- Northumbria University
- Bucks New University

76,930 Active Students

168,476,700,000 total data points are available to the model
PUTTING IT ALL TOGETHER

Raw data → Derived Features → Models & personalized insights

**Institutional data**
- Canonicalize historical & daily data
- Transform data & extract features
- AWS Redshift & EMR processing infrastructure for cost-effective scalability
- Preprocessing & model objects AWS
- Train or QA alert?
  - Yes: Automated & manual data-availability segmentation
  - No: Production scoring
- Feature optimization @ segment-level
- Feature optimization @ segment-cluster level
- Clustering to find homogeneous groups
- Feature optimization @ segment-cluster level
- Models @ segment-cluster level
- IFR @ student

**Canonicalized data in secure AWS**
- DB dump, csv, tsv files + nightly
- SIS, LMS, CRM, application, national data, intervention/outreach, Inspire

**App + Intervention data**

**ILLUMETM ADMINISTRATORS**
DATA AVAILABILITY SEGMENTATION

Data adaptive model building to account for student diversity and variability in data footprint

Data availability segmentation
- Maximizes potential of available data - Improves model accuracy and robustness as well as providing greater insights
- Typically 5-20 segments per institution - Efficiently evaluates patterns of data availability to identify segments
- Handles missing data without imputation

Who are these students?
- We can typically look at DA segmentation and see which segments are new students (no grade and no prior terms completed) vs. experienced students (most data available)
DATA SCIENCE: FEATURE COMPETITION

- Cluster level competition to identify best modeling approach and features
- Training/test datasets for validation
- Rank-order curve to identify the point of diminishing returns → Prevents over-fitting
Predictive power is measured by the multi-modal overlap measure (MOM). MOM is the measure of the shared area in a feature graph between those who persisted and did not persist. Same idea when transitioning from categorical features to looking at continuous features. As in a t-test, powerful predictors have the most separation.
A MEASURE OF PERFORMANCE: ROC CURVE

The horizontal axis measures the amount of incorrect predictions—how many students the model predicts are at risk of failure, but in fact succeed. This is the “false alarm rate.”

Vertical axis measures the amount of correct predictions—how many students the model predicts are at risk of failure, and do fail. This is the “detection rate.”

Test Population:
- 10,334 students
- 9,812 continued
- 522 did not continue

<table>
<thead>
<tr>
<th>Civitas Model (Test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000 student contacts</td>
</tr>
<tr>
<td>280 detections</td>
</tr>
<tr>
<td>720 false alarms</td>
</tr>
<tr>
<td>= 28% correct ID rate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Civitas Model (Train)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000 student contacts</td>
</tr>
<tr>
<td>51 detections</td>
</tr>
<tr>
<td>949 false alarms</td>
</tr>
<tr>
<td>= 5% correct ID rate</td>
</tr>
</tbody>
</table>
INSIGHTS FROM THE UK INNOVATION COHORT
Different Universities Create Different Choice Architectures

- “Modules Attempted” is a good variable for understanding student choice relative to student performance.
- Looking at modules attempted (versus modules earned) can help us understand student burn-out and resiliency at different universities.
- All these universities have students who are leaving early in their program (first year) but the dynamics of those stop-outs is different.
Institutions don’t always lose students at the start of their journey.

What is the context at this university and what structures can they enable to help these students?

Some institutions lose students at a consistent rate for a longer time through their programmes.

How would this university provide consistent support? Or should they look at a different variable?
How might we use mindset principles to help students persist and thrive?

- Belonging
- Normalizing
- Goal setting
- Empathy
- Mattering
- Growth mindset
"Nudges are **small pushes in the right direction** that do not require prescribed actions, but **encourage certain behaviors**. When students are presented with a nudge sent **from a trusted person** at your institution, they have the freedom to **make their own choices** with information **about behaviors** we know are more strongly **associated with positive persistence and graduation outcomes**.”

“A fundamental principle of our student success initiative is this: we believe all students can and will succeed if given the opportunity to do so.

Our students have demonstrated that. We are here to help students. We are here to help them earn a degree. Instead of data making students nameless, I actually know more students and know their stories.”

-- Dr. Paul Dosal, Vice Provost for Student Affairs & Student Success
JISC & CIVITAS INSPIRE PILOT
CREATING AN INSTITUTION SPECIFIC PLATFORM

CIVITAS DATA PLATFORM
DATA PROCESSING, PREDICTIVE MODELING, UPDATED DAILY

Jisc Learning Data Hub
WORK WITH US!

Mara Richard
Senior Director of International
Mara.Richard@civitaslearning.com
+447704308857