Using Machine Learning to Combat Inequity & Predict Student Success

Nathan Greenstein
Grant Crider-Phillips

April 13, 2023
Welcome

Nathan Greenstein
Assistant Director of Machine Learning Academic Data Analytics, University of Oregon
ngreenst@uoregon.edu

Grant Crider-Phillips
Machine Learning Analyst Academic Data Analytics, University of Oregon
criderg@uoregon.edu
Session
Roadmap

1. Organizational background
2. Winter retention case study
3. Approach to machine learning
4. Discussion
Organizational Background
Academic Data Analytics

• Office of the Provost
• Culture of data-driven decision-making
  • Shape policy
  • Prioritize equity
  • Increase transparency
• Focus areas:
  • Predicting student success
  • Understanding student feedback
  • Visualizing complex data
  • Understanding student and faculty progression

https://provost.uoregon.eduanalytics
Winter Retention Case Study
Prediction Task

Which incoming students will not persist to their second term?

- Predict before students matriculate
- Include all incoming first-time first-year students
- Each year, use predictions to target early advising intervention
Central Challenge

Non-Retention

- **Damaging** to students and university
- Disproportionately impacts **most vulnerable** students

Timely Intervention

- **Difficult to recover** from early negative experiences
- **Proactive interventions are more effective** than reactive ones

Finite Resources

- **Fewer advisors** than students
- **Must choose** who receives a given intervention first
Central Challenge

Can we predict which incoming students will not persist to their second term?

How early can we make our predictions?

Non-Retention  Timely Intervention  Finite Resources
Model Performance, 2021 Cohort

ADA Model
2021 Cohort
Model Performance, 2021 Cohort

- ADA Model
  2021 Cohort

- GPA Alternative
  2021 Cohort

- Random Lottery
  2021 Cohort
Approach to Machine Learning
Machine Learning

Promises
• Greater **predictive power**
• Better equipped for **challenging outcomes**
• Harnesses **bigger, messier data**

Concerns
• Will **human stakeholders** lose their voice?
• Might the algorithm be **biased or inequitable**?
• How much **transparency** will be offered?
Process Commitments

- **PARTICIPATORY**
  Engage meaningfully with a range of stakeholders

- **TRANSPARENT**
  Report honestly and accessibly on process and outcomes

- **EQUITY-ORIENTED**
  Apply lens throughout; demonstrably advance equity
Process Highlights

**Participatory**
- Partner closely with Undergraduate Education and Student Success
- Converse with other offices
- Reflect student body through diverse data sources

**Transparent**
- Report actively to UESS throughout
- Publicly disseminate methods and results
- Acknowledge strengths and limitations

**Equity-Oriented**
- With stakeholders, define equity standards
- Ground our work in existing scholarship
- Thoroughly vet model for equity and revise as necessary
Discussion
Thank you!

Nathan Greenstein
Grant Crider-Phillips

Academic Data Analytics
Office of the Provost
University of Oregon

https://provost.uoregon.edu/analytics
Appendix
• Many varieties; today’s focus is **predictive analytics**
• Harnesses large amounts of **data** and **computing power**
• Searches for **relationships** between inputs and outputs
• Finds patterns **more complex** than human eyes and traditional methods can handle
• Not magical, but **powerful in the right situation**
• Early advising already in place
• Mathematical model already in use
  o Predicts first-term GPA
  o Traditional linear regression
  o Unable to predict second-term retention
  o A useful tool, but a compromise
• Not evaluated for equity
Initial Guardrails

Do No Harm

• Focus on model performance
• Compare specific groups to everyone
• Ensure that:
  • Vulnerable groups: ≥ 85%
  • Privileged groups: ≥ 75%*
URM Performance, 2021 Cohort

ADA Model
All Students

ADA Model
URM Students

GPA Alternative
URM Students
## Equity Performance Comparison

<table>
<thead>
<tr>
<th>Model Performance (% of Non-Returners Targeted)</th>
<th>All Students</th>
<th>First-Generation</th>
<th>URM Race/Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>All</td>
</tr>
<tr>
<td>Training Cohorts (2010 thru 2020)</td>
<td>100%</td>
<td>149%</td>
<td>146%</td>
</tr>
<tr>
<td>Validation Cohort (2021)</td>
<td>100%</td>
<td>129%</td>
<td>149%</td>
</tr>
</tbody>
</table>

Cells representing fewer than 10 students are not reported.
• Equity depends on **real-world use**
  • Sometimes need parity
  • Sometimes need **disparity**
• Withholding race **does not ensure equity**
  • Proxy variables
  • Unbalanced data
• Model bias can be **measured** and **corrected**
  • **Empirical** metrics quantify bias
  • Corrective technologies **adjust** model
  • Use together for real **accountability**