The Team

Griffin McCauley

M.S. Data Science

Sc.B. Applied Mathematics, A.B. Economics

Theo Thormann

M.S. Data Science

B.S. Environmental Science and Policy

Eric Tria

M.S. Data Science

B.S. Computer Science

Jake Weinberg

M.S. Data Science

B.S. Commerce

Client and Project Background

hum

- Customer data platform (CDP)

- Use that data to help publishers understand their customers and content

- Focus on educational publishers

Project Background

- Publishers need insight on clients
  - Want to maximize user retention
  - Want to know possible reviewers

- Utilize user "events" to create models
  - Pageviews
  - Citations
  - Scrolling

Goal: Create two models using user events to maximize user retention and suggest article reviewers

Project Timeline

	Fall	Winter	Spring
Goals	• Complete kickoff and onboarding • Gain subject matter expertise • Complete EDA ✓	• Model design • Begin model buildout	• Model refinement and testing • Prepare for deployment of models to clients
Critical Activities	• Gain access to client data • Understand necessary analysis tools ✓	• Implement RNN for retention model • Perform regression and clustering analysis for reviewer model	• Validate model performance • Ensure models are malleable for specific clients

Data and Tools

The Modeling Data

- First-Party Customer Data
  - Significant user events and actions stored on the platform
  - User profile information

Data Access

- Cloud-based platforms
  - Data hosted on Snowflake
  - Python integration through Snowpark API
  - Machine Learning using AWS SageMaker

Defining Success

Client Side

- User Retention Model
  - Identify users with high churn risk
  - Recommend ways to re-engage

- Reviewer Recommendations Model
  - Assess the qualifications of candidate reviewers
  - Provide suggestions to editors

Student Side

- Apply concepts from our coursework to a real-world problem domain

- Use teamwork skills to design and manage a project from start to finish

- Learn about a unique industry

Overarching Mission
Make human knowledge more interconnected and accessible