

---

---

# Student Dropout Prediction

— Rui Chen, James Fantin, Kun Yi —

---

---

# Introduction

We want to use machine learning to predict if a student will drop out or not

Why is this important?

- At UW, 22% of students dropout after their first year <sup>1</sup>
- Nationwide, the rate of student dropout is 28% <sup>1</sup>
- US taxpayers spending nearly \$2 billion annually on educating non-returning first-year students alone <sup>2</sup>

1. "University of Wyoming Graduation & Retention." College Factual, 13 Sept. 2019, <https://www.collegefactual.com/colleges/university-of-wyoming/academic-life/graduation-and-retention/>.

2. M. Schneider. Finishing the rst lap: The cost of first year student attrition in America's four year colleges and universities. American Institutes for Research, 2010.

# Transfer Learning

Transfer learning is a problem in machine learning that attempts to apply knowledge from one problem to solve a related problem

We want to build a model using information from one college and transfer that prediction to other colleges

# Fairness In Educational Data

We want to work on increasing fairness between groups in an educational data setting

**Fairness** between groups means that each group is treated equally

i.e. The dropout prediction rate for a female student should be similar to the dropout prediction rate for a male student

**Intersectional bias** means that the model does discriminate based on several attributes of a student

i.e. The algorithm may be fair between gender and race alone, but may discriminate against black men

# What We Will Do

## Major Tasks

- How do we transfer prediction from one-school model to multi-school model?
- How do we detect bias from a single attribute to bias from the intersection of several attributes and remove the bias?
- How do we create a fair algorithm if we do not have access to private attributes about a student?

# Plans

- Meet with Dr. Lan bi-weekly to give presentation on what we have done in that period of time
- Survey current literature on student dropout prediction
- Replicate results from research papers and find bias in our educational datasets by the end of semester
- Acquire more educational data online and possibly from the University of Wyoming

# Goals

Build a machine learning algorithm that combines all three of our tasks together

We will evaluate our success based on

- Accuracy of our model compared to other models in research
- Reduction of intersectional bias in our model compared to others
- Effectiveness of our transfer learning model

# Tools

Write our code in python and use th Scikit Learn for some machine learning models

Use LaTeX to write a publishable research paper



# Roles

**James** will start by investigating removing bias in data without access to the private attribute information about a student

**Kun** will start researching transfer prediction from one-school to multi-school using domain adaptation

**Rui** will start by investigating removing bias from a single attribute and removing bias from intersectional attributes

# Progress

Mostly finished our literature review of student dropout prediction

Have a dataset from Massively Open Online Courses (MOOC) with 500,000 students

Have a dataset of 700 students from a Portugal school

Building basic models to replicate results from research

# Questions