Case Studies

Bias and Fairness in ML: KDD 2020 Tutorial

# CASE STUDY A – Student Support Programs

**Description**: You’re working for a large, diverse school district with about 25,000 students in the graduating class each year. As with any district, students failing to graduate on time (or at all) is a considerable concern -- currently, about 10% of students who enter 9th grade fail to graduate within 4 years. Recently, a large foundation provided a generous grant to provide additional tutoring to 10% of the students throughout their entire time in high school. In order to help identify which students should receive this additional help as they enter their freshman year, your team has built a model to predict the students most at risk of not graduating on time.

**Goal**: Improve graduation rates for students

**Data**: student records from different school districts and states, national student clearinghouse data (which gives us information about college outcomes)

**Analysis**: predict risk of not graduating on time

**Actions**: assign after-school programs to most at-risk students

## Breakout Session 1 –– Sources of Bias:

### What are some potential sources of bias in the underlying data?

[The bullet points below are filled out answers from the participants in the KDD 2020 Tutorial]

* Sampling for national student clearinghouse data
* Unknown/different sampling and completion rates based on the district/state that student records come from
* Reasons for failing to complete or nature of failing to graduate (e.g., pursue GED)
  + Reasons could differentially relate to SES
* **Teacher might be biased to certain students**
* Special students need special attention
* Assumption that failure to complete is because of academic problems
* Demographic mismatch between national and local data
* Student records may not contain the important features for making predictions (related to above “assumption that failure to complete…”)
* Gender, race, demographic and income-related features in the data can create biases
* The criteria for fail/pass maybe biased towards certain group that has social/economic advantages (measurement bias)
* Are the school districts selected representative of the entire sample? (representation bias)
* Different economic conditions of the family leading to improper nutrition and different brain capabilities (also foetus got improper nutrition data can’t be captured)
* Why four years?

### How might biases be introduced in the data science pipeline? (Think about ETL, record linkage, feature engineering, labels, modeling, and model selection)

[The bullet points below are filled out answers from the participants in the KDD 2020 Tutorial]

* Features implicitly dependent on continued performance; eg, if you used results of standardized tests like the PSAT that are typically administered in 11th grade, there will be no results from students who left school prior to taking the test. (*cf. survival analysis*)
* Record linkage can often have problems for students with non-Western names
* Features have built in assumptions about cause of failing to graduate and might only capture a certain subset of them
* Eg Mental health can’t be measured, labels would be biased
* Features to capture socio-economic or cultural factors that may affect school completion rate (e.g. child marriage, child labour, family responsibilities); however, we’ll need these features at a very granular level (i.e. sub district)

### 

### 

### What are the risks to fairness in downstream applications and deployment of the model described?

[The bullet points below are filled out answers from the participants in the KDD 2020 Tutorial]

## Breakout Session 2 –– Bias Metrics:

### How would you describe a **false positive in this problem** to a policymaker or business owner? What’s the potential harm of cost of having one?

### How would you describe a **false negative** to a policymaker or business owner? What’s the potential harm or cost of having one?

### What confusion matrix metric (e.g., FPR, FNR, TPR, FDR, etc.) would you choose to focus on in terms of equity for this case? Think of the fairness tree here.

Add table here - metric , rank, reason