

Recitation 8

Fairness



Exercise - College Admissions





Goals & Stakeholders

- **Population groups & distribution**
 - Background
 - Country of origin
 - Undergrad degree
 - Work experience
 - Gender
 - Race
 - Financial status
- **Goals**
 - Get the best and qualified applications into the program
 - Promote diversity to enrich the program network / alumni
- **Assumptions**
 - Applicants are filling in information that reflect reality, and are authentic
 - Applications aren't given admits/denied admits outside of what the AI system recommends
 - Past admissions were fair, and the correct/all data is provided



Harms & Biases

- **Harms**
 - Allocation / Representation?
 - Stereotyping?
- **Biases**
 - Historical?
 - Tainted examples?
 - Skewed samples?
 - Limited features?
 - Sample size disparity?
 - Proxies?



Harms & Biases

- **Allocation / Representation**

- Male students are ranked higher / more often offered admission
- Students with only CS background are selected more often
- Students from a only subset of countries are selected more often
- Students from similar background as alumni are selected more often

- **Biases**

- Historical (past biases - admission committee / before AI)
- Tainted examples (dataset biased due to improper human labeling)
- Skewed samples (manual admission biases skews future admissions; ex: selection based on alumni)
- Limited features
- Sample size disparity (insufficient data for applications from certain countries / backgrounds)
- Proxies (ex: type of undergraduate degree can serve as a proxy for gender - not always applicable)
- Population bias (training vs target demographic), Outliers, etc.



Biases in Data Collection & Preprocessing

- **Data**
 - Past applicant's application packets
 - Program's admission records, reasons, etc.
- **Data collection**
 - Acquisition (sufficient applications data was not recorded/kept track of until few years ago)
 - Querying (the whole applicant pool data was not used to train on)
 - Filtering (some application data was considered useless after program structure changed)
- **Data preprocessing**
 - Cleaning (some applicants did not fill in all data required in the packet, so they were removed)
 - Enrichment (we filled in placeholder or average values based on historical data for some fields)
 - Aggregation (multiple areas of study were merged into a single feature)



Fairness

- What do we want to achieve - equality or equity?
- What type of fairness definition/metric is appropriate?
 - Anti-classification
 - ??
 - Independence
 - ??
 - Separation w/ FPR or FNR
 - ??



Fairness

- Do we want equality or equity?
 - Depends on goal
- What type of fairness definition/metric is appropriate?
 - Anti-classification
 - Gender/race, etc. should not be considered at all for admission
 - Recall: proxies
 - Independence
 - Rate of positive predictions (acceptance rate) is the same across groups
 - Vary thresholds for different groups - Equity
 - Separation w/ FPR or FNR
 - False positive and negative rates are the same across groups
 - Vary thresholds for different groups - Equity
 - Ex: FNR - probability of incorrectly being denied admission is equal across groups



Thank You!