Responsible Data Science

Lab 12: Final Exam Review and Q&A
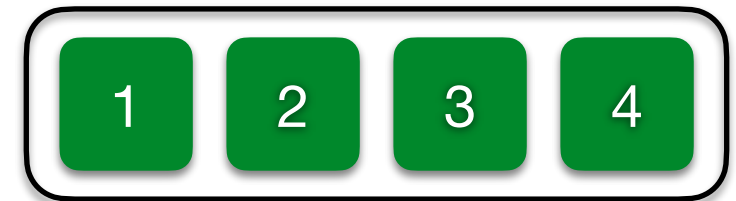
Center for Data Science
New York University

data *RESPONSIBLY*

NYU

Fairness survey

(1) Types of bias:

- Pre-existing
- Technical
- Emergent

Fairness

| 1 | 2 | 3 | 4 |

(2) Fairness in classification:

- Group fairness vs individual fairness
- Disparate impact vs disparate treatment
- FNR, FPR, Calibration
- Trade-offs
- Fairness from the perspective of different stakeholders

(3) Fairness as equality opportunity

- Formal EO (and Formal+)
- Substantive/Rawlsian EO
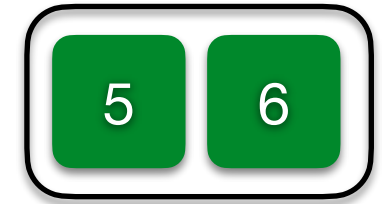- Luck-egalitarian EO

r/ai

Data science lifecycle survey

(1) Data lifecycle of an ADS:

- Acquisition
- Technical
- Emergent

(3) Dimensions of technical bias:

5	6

DS Lifecycle

- Interpolating missing values
- Data filtering and joins
- Data distribution debugging

(2) Data profiling:

- Statistics (e.g. mean, median) and distributions (e.g. histograms)
- Structure (business rules, schema of a dataset)
- Data types
- NULL and missing values
- Different types of missingness

r/ai

Data protection survey

(1) Privacy:

- Randomized response
- Do we need randomization?
- Linkage attacks and differencing attacks
- Trusted and untrusted parties
- Differential privacy
- Privacy budget
- Privacy guarantees vs noisy estimates
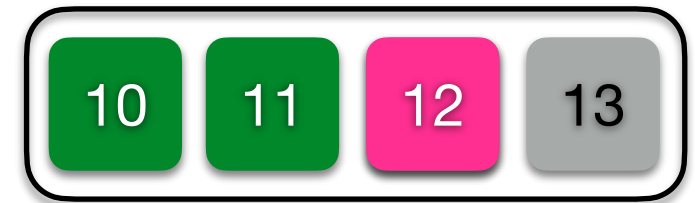
Data Protection

7 8 9

(2) Ethical frameworks:

- Consequentialism
- Deontology
- Belmont report, principles:
  - Respect for persons
  - Beneficence
  - Justice
- Menlo report, principles:
  - Respect for law and public interest
- Informed consent

r/ai

Transparency and intepretability survey

(1) Explaining black-box models:

- What are we explaining, to whom, and why?
- Explanations based on features
- LIME, SHAP, QII
- Limitations/insufficiency of accuracy for understanding models
- Importance of transparency & interpretability for trust

| 10 | 11 | 12 | 13 |

Transparency & Interpretability

r/ai