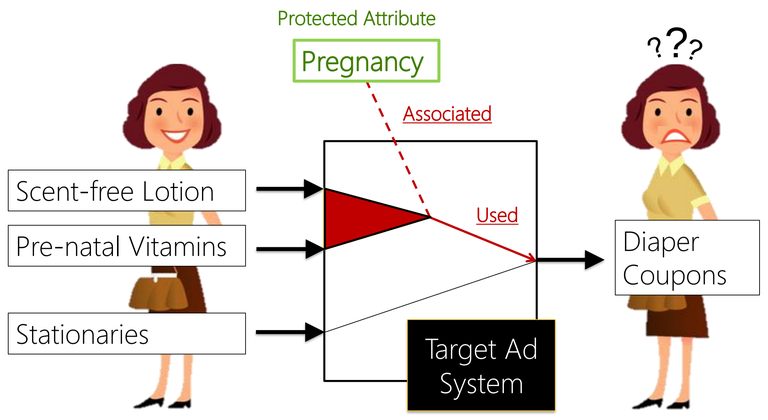
Use Privacy in Data-Driven Systems

Theory and Experiments with Machine Learnt Systems

Anupam Datta, Matt Fredrikson, Gihyuk Ko, Piotr Mardziel, Shayak Sen
Carnegie Mellon University

Learning Systems Threaten Privacy

Credit	Web systems	Healthcare
Education	Law Enforcement	Personalized privacy assistance
IoT Applications	…	…

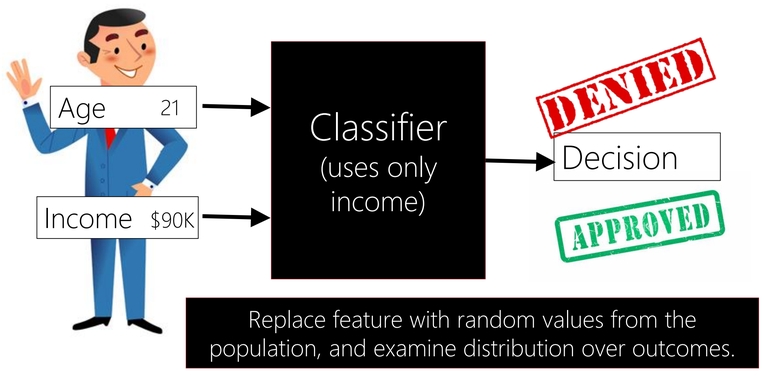


Using pregnancy status (inferred via past purchases) for marketing [Target 2012]

Need theories to define **'use restrictions'** for the protected information type in algorithmic decision making systems.

Explicit Use

Quantitative Input Influence*

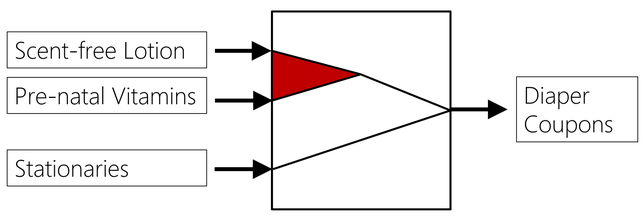


Replace feature with random values from the population, and examine distribution over outcomes.

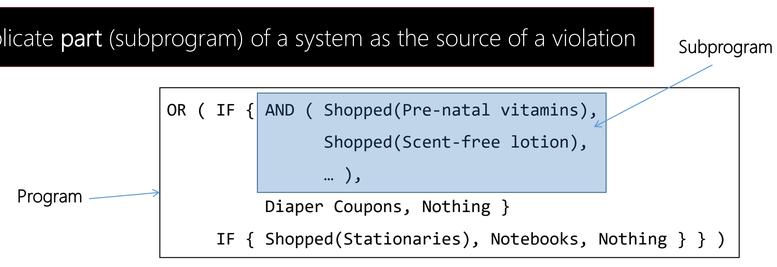
*Anupam Datta, Shayak Sen, Yair Zick. *Algorithmic Transparency via Quantitative Input Influence*. Oakland'16

Proxy (or implicit) Use

Learning Systems as Programs



Implicate **part** (subprogram) of a system as the source of a violation



Two-Phase Definition

Association — *Is a subprogram associated to the protected attribute?*

➔ Well-studied association measures (e.g., Mutual Information)

Use / Influence — *Is a subprogram influential to the output of the program?*

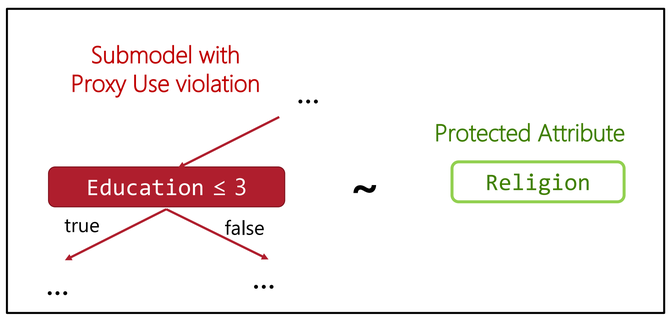
➔ QII for subprograms

(ϵ, δ)-Proxy Use: A subprogram with association level above ϵ, and influence measure above δ exists

Experiments

contra

Advertisement targeting using the Indonesian Contraception Dataset
- Features: `Education, Children, Husband's Job`, etc
- Classification: `Contraception Methods`
- Protected attribute (removed in training phase): `Religion`
- ~1,500 individuals

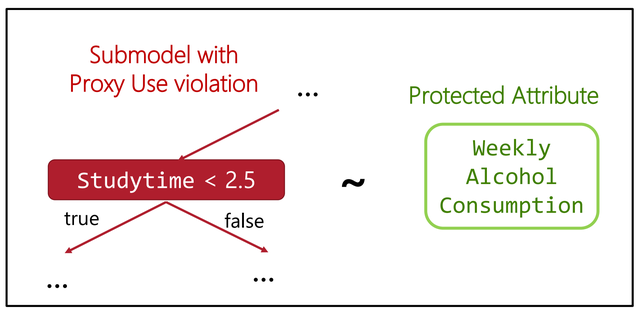


Education level used to target people in specific religion

➔ Concerning Use

student

Academic performance prediction using Portuguese Student Alcohol dataset
- Features: `Failures, Studytime, Father's education level, Health status`, etc
- Classification: `Grade`
- Protected attribute : `Weekly alcohol consumption`
- ~7,000 individuals



Study time used as a predictor for the academic performance

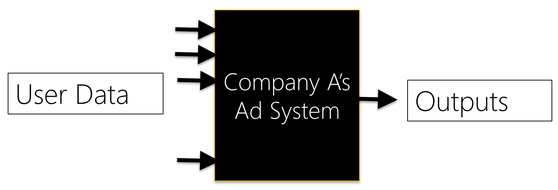
➔ Acceptable Use

Summary

- Use Privacy restricts use (*explicit* or *proxy)* of protected information type for *certain purposes* with *some exceptions*
- We implemented a software tool for detection and repair of use violations
  - Experimental validations with a few real dataset

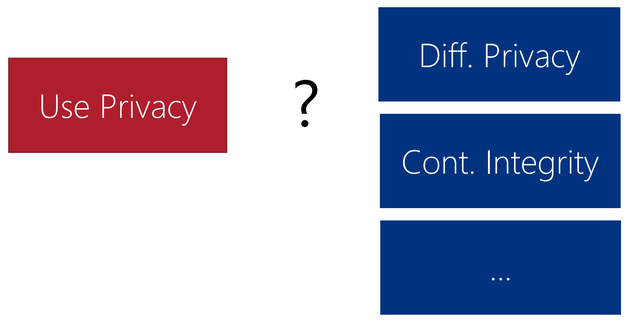
Directions

Black-box Models



Explore Utility Tradeoffs

Explore Relationship with Other Privacy Notion



Carnegie Mellon University

Carnegie Mellon University CyLab Security and Privacy Institute