Fundamental Concepts in Data Insight:

Data Ethics

Fundamentals for a General Audience

Data Ethics

- Why Data Ethics?
    - What is Data Ethics?
- Policing Case Studies: Motivations
    - Data Storage
    - Illegal Actions
    - Bias
- Mini Case Studies: Motivations & Problems
    - Case Study: Tesla Crash (Responsibility)
    - Case Study: `care.data` (Importance of Ethics)
    - Case Study: Airport Threat Detection (Risks)
    - Case Study: Unintended behaviour
    - Case Study: Street Bump (Inequality)
- Topics in Data Ethics
    - Ethical Systems
    - The Ethics of (Opaque) Algorithms
    - Problem in Data Ethics
        - Problems: Privacy
        - Problems: Discrimination
        - Problems: Auditing & Responsibility
- Topic Focus: Accountability & Explanation
    - Case Study: Compass Recidivisism System
    - Why don't we get explanations?
    - What is an Explanation of a Model?
    - What is an Explanation?
    - Counterfactual explanations of models
- Group Exercise

Why Data Ethics?

- ethical implications of data analysis
  - we need algorithmic systems to process data
  - novel practices requried, driven by data
- data ethics concerned with data itself
  - producing large volumes of data
- data combined with other sources
  - suprising cross-domain inferences
- concerns
  - fairness, responsibility, respect for HR
  - counter-productive ot ignore
    - back-reactions & downsides

What is Data Ethics?

- ethics of data
    - privacy
    - trust
    - transparency
- ethics of algorithms
    - accountability
    - design
    - auditing
- ethics of practices
    - deontological code
    - consent
    - privacy
- laws
    - seperate legal structures
    - smart robotics, AI, data protection

Policing Case Studies: Motivations

Data Storage

Investment in databases and their functionalities has met opposition from oversight commission, human rights, and public interest groups. In the U.K "The Biometrics Commissioner has warned [the police] that ***many of the 20 million custody photographs currently stored on their systems are being held unlawfully*** *and might need to be destroyed" (Loeb, 2018).*

Merging the three databases into the NLEDS, not only accumulates these existing concerns it also creates new ones related *to proportionality, legitimacy, and ownership of data. While "proportionality will be a design feature of the system with permission-based access, with a full audit trail and a description of purpose of access. There is much work to do in terms of exact detail" (Surveillance Camera Commissioner, 2016: 23).*

Illegal Actions

A clear example of this was during the ***G20 summit in Hamburg****, Germany when the* ***accreditation of 32 journalists was revoked as a result of them being labeled as ‘left motivated violent offenders’*** *in a BKA database (Monroy et al, 2018).* ***Legal challenges revealed that the decision to revoke their accreditation was based on ‘old’ data which was never deleted****. According to the BKA, it is the responsibility of the authorities, who enter data on suspects into the system, to delete old records. In most cases, data entry is done by the German states, which fail to remove ‘old’ data (Fiedler, 2017).*

In Cardiff, the South Wales Police received £2.6m from the Police Transformation Fund to lead on the testing and deployment of AFR, and in some instances LFR. ***“South Wales Police has admitted it has used AFR technology to target*** *petty criminals, such as ticket touts and pickpockets outside football matches, but they have also used it on* ***peaceful protesters****” (Liberty, 2018).*

Bias

> ***The AFR systems used in by U.K police were found to perform abdominally poorly, with “on average, a staggering 95% of ‘matches’ wrongly identified innocent people”*** *(Ferris, 2018: 13), who were subsequently* ***stopped and asked to prove their identity****. These findings of flawed facial recognition technology are in line with research showing* ***inherent bias in facial recognition*** *systems towards women and people of colour (Buolamwini et all, 2018; Ferris, 2018).*

***The London MET developed the Gang Matrix to identify potential gang members** and score them according to the risk they pose to society. Research by Amnesty International (2018) and Scott (2018) revealed the discriminatory nature of this predictive identification program, in which **the majority of individuals were young black men**. The MET was ordered to radically reform the matrix within a year by the Mayor of London, and are currently working on a new program called the 'Concern Hub' (Mayor or London, 2018a: Dodd, 2018; Crisp, 2019).*

*The use of predictive policing technologies has raised many concerns. Scholars in the USA have demonstrated how **historically biased police data** in predictive policing programs is **erpetuating the over-policing of African American neighbourhoods** (Lum et al, 2016). In the U.K Liberty found a similar negative feedback loop, people from "back, Asian and minority ethnic (BAME) communities are disproportionately more likely to be arrested, leading the program to assume, wrongly that the area in which they live or spend time are the areas where there is more crime” (Couchman, 2019: 4).*

Mini Case Studies: Motivations & Problems

Case Study: Tesla Crash (Responsibility)

- tesla in autopiolt mode
    - who is responsible for crash?
- oversight is very difficult
    - many parts of systems
- "distributed responsibility"
    - how does liability apply to sofwtare?

Case Study: `care.data` (Importance of Ethics)

- https://en.wikipedia.org/wiki/Care.data
- sharing health data in UK
- limited consent
- didnt go ahead

Case Study: Airport Threat Detection (Risks)

- robot patrol or CCTV
- people suffer harm from these systems, ie., detention
- who can you hold to account for this?
- where is redress?

Case Study: Unintended behaviour

- (hidden) competition between wikiepdeia bots
    - unexpected interactions
- high-frequency trading crashes
- tay chatbot
    - effect of *who* was training system!

Case Study: Street Bump (Inequality)

- areas with smartphones could register potholes
    - so those areas got fixed first -- exagerate inequality
- app could have used smartphones as *a* source of data
    - but it was *the* source, and so exclusionary

Topics in Data Ethics

Ethical Systems

- denotology = principles = motivations & intentions
    - autonomy = ends in themselves
    - procedural justice
        - do no harm
    - not complete, eg., "unintended concequences"
- concequentialism = outcomes, effects
    - better for
        - unitended concequences
        - broader impact of data sci
    - also limited
        - can over-look effects on invididuals
- "environmental approach"
    - respect & care are due to " an environment "
    - individual duties to protect and foster
    - gardners?
    - trying to combine deont + concq.
- virute ethics
- causitry

The Ethics of (Opaque) Algorithms

- data-processing decision-making algorithms
  - specification
  - implementation
  - configuration
- concerns
  - evidence is
    - inconcluive (unjustified)
    - inscruitable (unexplainable, opacity)
    - misguided (stat. biased)
  - unfair outcomes (morally biased, discrimination)
  - transformative effects (autonomy, privacy)
  - traceability (audit/responsibility)
    - de-respsonsibilisation
      - "computer says so"
- when is correlation sufficinety reliable?
  - limit use of opaque methods in some contexts?
    - credit, policing,...

Problem in Data Ethics

Problems: Privacy

- EU data protection
  - identifiable individuals
    - groups are not protected
    - "consent"
  - remove/hide in datasets
    - within law, individuals lose control over anonymised data
  - however,
    - in ml/pa you are grouped into sets of similar people
    - their actions affect you
    - these are ad-hoc groupings based on organizations' interest
      - not collective orgs or ascriptive
        - eg., union, genetic group
  - a group privacy right may be required

Problems: Discrimination

- conclusive, transparent, well-founded (automated) decisions can be discimrinative
- bias *in human labelling!* of training data
- "racial" correaltions found *after*wards

Problems: Auditing & Responsibility

- who makes the ethical choice?
  - auto-car: driver, machine?
    - moral effects obvious
  - many systems have non-obvious moral effects
  - organizational policy...
    - programmer?

Topic Focus: Accountability & Explanation

Case Study: Compass Recidivisism System

- More false-positive for black defendents than for white defendents.
- Related court case
  - Wisconsin vs. .... -- do you have a right to an explanation?
  - judge said: system is trade secret!?

Why don’t we get explanations?

- infringe trade secrets
- technically infeasible
- not meaningful for individual cases
- manipulation of system by third parties
    - eg., Tay chatbot

What is an Explanation of a Model?

- holistic or local explanantion
    - how model works in general?
    - or your partiuclar decision?
- form of explanation
    - visual
    - audience: child, adult, expert
- what innformation?
    - source code
    - input data
    - weights
    - model
    - ...?

What is an Explanation?

- number of ways of providing explanations...
  - case-based explanations
    - find a clear historical case similar to current one
    - provide that justification
  - local explanations
    - train a simpler model within reigion of complex model (eg., linear)
    - eg., "How do I drive north?"
      - forwards *only*? or to a destination?
  - counter-factual explanations
    - understand, challenge or alter a decision
    - if q were false, S would not believe p
      - so q *explains* "belief that p"
    - multiple counterfactual explanations are possible

Counterfactual explanations of models

- (credit score) pv was returned because
  - variables V
  - had values v1, ..vn
- if they had instead had U (u1..un)
  - score pu would be returned
- eg., you were denied a loan because your income was £35k, if your income was £45k then you would have been offered a loan
- can be used with blacbox algs
  - you can "find nearest" input variable values to arrive at desired output, ie., alternative decision

Group Exercise

- consider one of the case studies where you were a victim of an automated decision
- what information would you want? why?
- what would constitute a meaningful explanation?
- does this change depending on who you are?

The Law

Concerns

Changing Laws

> *In 2019 the German ANPR systems became contested, the German Constitutional Court ruled that mass registration of license plates infringes on the right to informational self-determination and limits individual freedoms*

International Regulartory Concerns

Police forces across Europe are mainly turning to Palantir , a big data analytics company from the US with a track record of contracting for police and secret services across the world, for the integration and searching of heterogeneous databases, and social graphing of suspects

European human rights groups, oversight committees, and scholars have been less vocal on the use of Palantir technologies by European police forces then they have been on the creation of databases, real time identification, and predictive policing systems. Concerns have been raised on the use of an American company for crunching French police data (Samama, 2018), how the use of Palantir software might not be in-line with national and European data protection regulation (EDRi, 2017; Technology World, 2018) or more fundamentally how the use of Hessendata by police to track suspects blurs the separation between the German police and secret service (Kurz, 2018).

GDPR

does not provide a right to decision making!

you have a right to receive *some* information, and what *type* of data is used, and what *of your data* is being used

but not right to a decision on your individual decision.

a "right to be informed" vs. a right to an explanation.

Data Protection Directive (Not Legally Binding, presently) – Art 12a

- right to obtain informaton from controller
  - "knowledge of the logic involved in any automatic processing"

- judges have concluded: very lmimted
  - broad overview, no detail on tradescretes

Art 22 GDPR – Automated Individual Decision-Making

teh data subject shall have the right..

right to contest -- meaningless

- **not explanation** of the *rational*

only for *soley* automated -- any human involvement drops right; only where *significant effects* (undefined(