Privacy, Security and Professional Ethics

UBCO Master of Data Science – DATA 553

SOME SLIDES FROM DSCI 541: ED KNORR, FEB-MAR 2020, MDS PROGRAM UBC VANCOUVER

SOME SLIDES FROM DATA 553: FATEMEH FARD, FEB-MAR 2020, MDS PROGRAM UBC OKANAGAN

Introduction

I respectfully acknowledge this university is located on theunceded territory of the Okanagan Nation.

Instructor: Dr. Mostafa Mohamed, Mostafa.Mohamed@ubc.ca

Lecturer, Computer Science

Research area: Computer Vision, Data Science

Current Office: SCI 200E

TA: Yining Zhou, zhou258@student.ubc.ca

Introducing yourself – Private message on Slack / Canvas

- **Name, one thing about you (where are you now? What is your background? What you plan to do after the MDS…)**

Why this course?

The overall goal of DATA 553 is for you to:

Practice and Understand security, privacy, and ethics concepts, techniques, and tools as a data scientist

This course will cover essential knowledge and topics on data security, privacy, and professional ethics required for data scientist.

How to excel in this course

Read/Listen to the advanced material

Attend lectures and labs

Interact

Practice

Solve assignments yourself

Course Format

At home :

- Videos
- Readings

During class time:

- Content and Case studies
- Guest lectures
- Final Quiz

Expect up to 30 minutes of work prior to a class.

Evaluation

Weekly Quizzes (3)	15%	due on Fridays
Lab 1	15%	due Oct 22nd
Lab 2	15%	due Oct 29th
Lab 3	15%	due Nov 5th
Lab 4	15%	due Nov 12 ***
Final Quiz	25%	

Tools / Other details

Lab 3: Python

Case studies (class/lab):

- Don’t expect RIGHT / WRONG answers all the time
- Can have arguments for and against
- Reflection-based

All course materials and deadlines will be on Canvas

Guest Lectures

Oct 11/2023 (Security Threats)

Larry Carson, Associate Director – Information Security Management

Oct 23/2023 (Privacy and Responsible AI)

Liza Wood, Vice President, Data Science, Products and Services – Two Hat Security [acquired by Microsoft]

Nov 1/2023 (Security measures)

Don Thompson, Chief Security Officer – UBC; and Larry Carson

Nov 6/2023 (DB Security)

Dr. Ramon Lawrence, Computer Science UBC

Course objectives

Upon successfully completing this course, you will be able to:

- Explain security, privacy and professional ethics
- Identify situations in which data is sensitive, assess the risks, and articulate a reasoned response.
- Apply ethical considerations to case studies. Consider privacy, human dignity, harm, the public good, legal issues, the role of ethics boards, and various forms of consent.
- Apply some privacy algorithms on your data
- Identify algorithmic bias, and recommend practical solutions to promote fairness.
- Explain why good security is not a product, but rather a process and a mindset.
- Argue for why security is complex and difficult, and why perfect security may be unachievable.

Motivation

There is a lot of valuable research data that the public could benefit from, if it were made available to more researchers.

Much of this data is either private or "sensitive", and contains Personally Identifiable Information (PII), such as:

- Tax records
- Medical records
- School records
- Employment data
- Government records
- Census data

How can we gain the benefit of publishing and using this valuable data *without compromising privacy*?

⇒ "Privacy-Preserving Data Publishing"

Motivation

By law, some organizations MUST share their data or make it public

- Can you cite some?

Why?

- Freedom of information
- Public interest
- Research interest:
  - Sharing of medical data can lead to
    - better health for all, now and in the future
    - More efficient system for the benefit of all

Motivation

New technologies and analysis techniques allows for better analysis of larger and larger datasets

Interconnection of datasets create potential for greater results but also greater risk

- Medical field: connection of IoT devices to medical records? DNA records?
  - What is the potential and risk?
  - Eric Topol, “The patient Will See You Now: The FUTURE of MEDECINE is in YOUR HANDS” (2015)

Privacy and Risks

Berman, Jules. *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*, 2013, p. 211:

"Unfortunately, zero risk is not obtainable. However, it is quite possible to reduce the risk of privacy violations to something far, far below the known risks of identity theft that occur with every charge card transaction, every cashed check, and every Facebook posting."

"It's easy to focus on how data is collected by corporations and governments, but that gives a distorted picture. The real story is how the different streams of data are processed, correlated, and analyzed."

Goals

Preserve privacy

Provide useful data to researchers, governments, consumers, etc. for a purpose.

Health care:

- Heath benefits, reduce costs, new drugs
- Faster to market
- Avoid treatment that are unlikely to work

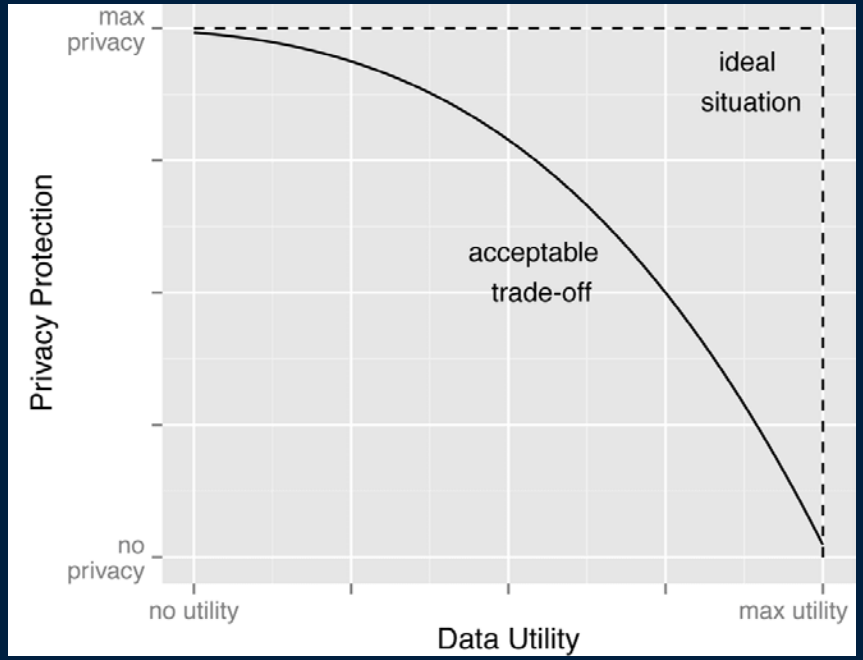


Figure from: Khaled El Emam and Luk Arbuckle. *Anonymizing Health Data: Case Studies and Methods to Get You Started. O'Reilly, 2013.*

Goals

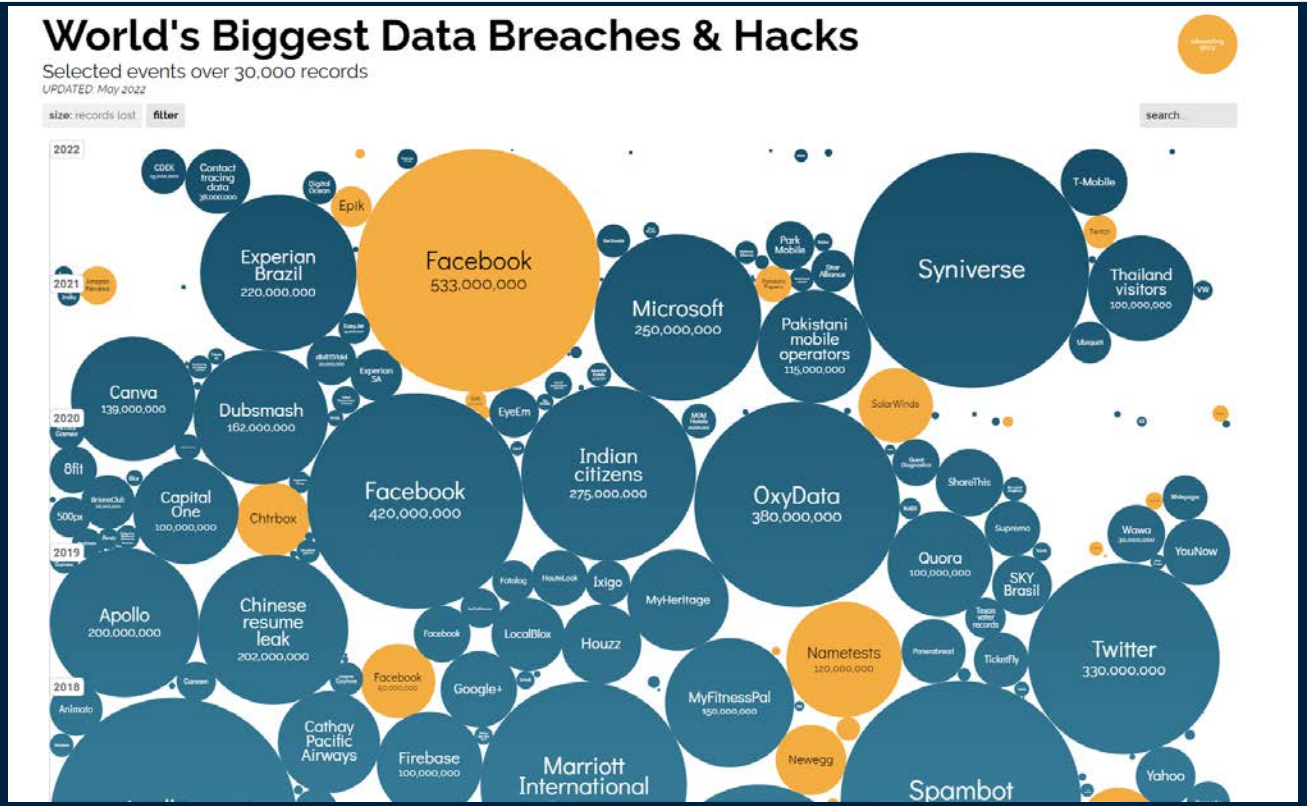
Help organizations

- Develop new services (e.g., social media sites, search engines)
- Respond to trends
- Support reviews and evaluations: Reddit, Yelp
- Amazon recommendations, Netflix recommendations
- Avoid “re-inventing the wheel” by sharing prior work

But it has to be done ethically

The social dilemna

Motivation



https://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/

Guest Lecture – Larry Carson

Associate Director – Information Security Management
UBC
Has stewardship responsibilities over a substantial volume of information assets used for the teaching, research, and administrative functions of the institution.
Leads the information security program for UBC.

Next

Lab on Tuesday:

- Lots of reading and videos.
- Some should be done at home prior to coming to the lab

Lecture:

- Video: The end of Privacy, Dr. Michal Kosinski
- Short reading: case study for class


THE UNIVERSITY OF BRITISH COLUMBIA