STAT 413/613: Ethics in Data Science

Richard Ressler

2020-03-29

College of Arts and Sciences, Department of Mathematics and Statistics

Learning Outcomes

- Differentiate Legal, Professional, and Ethical Considerations in Decision Making
- Identify Sources of Ethical Issues in Data Science
- Apply Frameworks for Ethical Reasoning in Data Science
    - Philosophical Frameworks
    - Professional Codes of Conduct
    - Technical Guidelines
- References: See the end notes >CAVEAT: This presentation is intended to provide general background and facilitate educational discussion. It does not offer any form of legal advice or recommendations.

We make choices across the data science life cycle.

- What kids of projects do I work on? What questions do I analyze?
- How do I get my data?
- What data can I share and with whom?
- What do I do about “bad” or missing data? What to do with outliers that mess up my results?
- How do I leverage/credit other people’s work to improve my own?
- Towards what goals do I optimize my models?
- How much effort do I put into checking my results?
- How do I report my results - what is personal intellectual property and what should be public domain?
- and many more...
- Choices can involve Legal, Professional, or Ethical Considerations

Choices have consequences, with benefits and risks

Legal considerations address the risks of legal (criminal/civil) consequences for violating laws

- Governments (organizations) establish laws (statutes) to specify permissible and impermissible activities as well as potential punishments - (judicial, e.g., imprisonment)
- Policies (regulations) provided guidance (within the constraint of laws) for regulating activities along with adjudication procedures and potential punishments - (non-judicial, e.g., debarment)
- Legal issues in big data include how you gather, protect, and share data, and increasingly, how you use it.
- However, laws are “local” not universal - consider the distinction between Pirates and Privateers
- Laws are often slow to keep pace with technology and difficult to interpret

When in doubt, ask a lawyer competent in the issue

US, International and State Laws Limit Potential Data Science Activities

- *Ignorantia_juris_non_excusat*
- Health Insurance Portability and Accountability Act (HIPPA)), regulates the protection and useage of health-related information and data.
- Family Educational Rights and Privacy Act (FERPA)) regulates the privacy records of students in schools. This act also limits data how third parties obtain and use data, e.g., for analysis.
- Fair Credit Reporting Act
- Fair Housing Act
- Equal Employment Opportunity Act
- European General Data Protection Regulation (GDPR))
- California Consumer Protection Act("Everything You Need to Know About the CCPA WIRED" n.d.)
- **And many more...**

Professional Considerations address activities related to organizations with which you affiliate

- Organizations establish bylaws or policies to identify and manage risk to the institution
- These may include "Professional Codes of Conduct"
  - American Statistical Association ("Ethical Guidelines for Statistical Practice," n.d.)
  - Association of Computing Machinery("The Code Affirms an Obligation of Computing Professionals to Use Their Skills for the Benefit of Society." n.d.)
- Organizational behavior and ethics may conflict with individual ethics
- Choices can include trying to change the organization or individuals, or leaving or being forced to leave the organization, with potential legal issues as well.

When in doubt, ask a mentor or manager you trust

Ethical Considerations arise when asking “What should I do?” or “What is”Right“?”

- Ethical Choices are often hard choices
  - Usually not covered by law or may include violating a law, regulation, or professional guideline
- Individual principles and cultural norms shape options and guide choices in complex situations
- Often no universally-accepted or even a good “right answer”
- May have to choosing between two bad outcomes or
- May have to choose between competing individual and group outcomes
- Choices can lead to “guilt”, reprobation, or even civil action (torts)
- Many, Many, frameworks attempt to guide Ethical Choices

- Broad Types of Philosophical Frameworks for Ethical Choices ("A Framework for Making Ethical Decisions Science and Technology Studies" n.d.)

- **Consequentialist**: Utilitarianism: greatest balance of good over harm (groups/individual)
  - Focus on the future outcomes to produce the most good
  - Compromise is expected as *the end justifies the means*.
- **Duty**: Do your Duty, Respect Rights, Be Fair and Just, Follow Divine Guidance:
  - Do what is "right" regardless of the consequences or emotions involved
  - Everyone has the same duties and ethical obligations regardless of circumstances
- **Virtues**: Live a Virtuous Life by developing the proper character traits
  - Ethical behavior is whatever a virtuous person would do
  - Tends to reinforce current cultural norms as the standard of ethical behavior

No single right answer

An Approach to Ethical Decision Making

- Recognize There May Be an Ethical Issue
    - Assess the underlying definitions, facts, and assumptions and constraints
- Consider the Parties Involved
    - Who are the individuals or groups who might be harmed or who might benefit and by when.
- Gather all of the Relevant Information
    - Are you missing key facts? Are they knowable?

An Approach to Ethical Decision Making (Continued)

- Formulate Actions and Consider Alternatives
  - Which action will produce the most good and do the least harm? (The Utilitarian Approach)
  - Which action respects the rights of all who have a stake in the decision? (The Rights Approach)
  - Which action treats people equally or proportionately? (The Justice Approach)
  - Which action serves the community as a whole, not just some members? (The Common Good Approach)
  - Which action leads me to act as the sort of person I want to be? (The Virtue Approach)
- Examine Alternatives and Make a Decision
- Act and Observe
- Assess and Reflect on the Outcome

Biased Brains Helped Us Survive (Cognitive Biases)

- Our brains have evolved mechanisms to make quick decisions
  - Us versus Them, Friend or Threat, Familiar or Strange
- Under stress we tend to bypass the higher-level cognitive centers.
- Humans get comfortable with patterns which can lead to systematic deviations from making rational judgments.
- We have many different ways to make decisions
- ... that may be bad in the long run...

Many Sources of Ethical Issues in Data Science

- Just a few of the possible sources ...
- Contradictions and competition among legal, professional and ethical guidelines
- Using biased data (even unknowingly)
- Using Proxies for “Protected” Attributes (even unknowingly)
- Protection of Intellectual Property versus Transparency and Accountability
- Law of Unintended Consequences - people will use your products and solutions in “creative” ways

Breakout Group Examples - 10 minutes

- Read the article
- Discussion Questions
  - Is there an ethical issue? What is it?
  - Who is affected and who is responsible?
  - What would you do differently or recommend?
- Report back to the main group when we return
- Groups

1. Software created by brand-name tech firms such as Amazon uncovered much higher error rates in classifying the gender of darker-skinned women than for lighter-skinned men.
2. Big Data used to generate unregulated “e-scores” in lieu of FICO scores for Credit in Lending
3. When Learning Analytics Violate Student Privacy Raths and 05/02/18 (n.d.)

UK Government Principles for Ethical Data Science

1. Start with clear user need and public benefit
2. Be aware of relevant legislation and codes of practice
3. Use data that is proportionate to the user need
4. Understand the limitations of the data
5. Ensure robust practices and work within your skillset
6. Make your work transparent and be accountable
7. Embed data use responsibly

Royal Statistical Society's A Guide for Ethical Data Science

What can you do about it?

- Recognize legal, professional and technical considerations surround your work
- Ask Questions of Peers, Managers, Mentors and Experts -
    - Seek Permission and Guidance, don't bank on forgiveness
- Document your sources - data, code, references
- Protect your data: especially Personally Identifiable Information or HIPPA information
- Produce Reproducible Analysis and solicit appropriate review
- Produce Results that Reveal, Don't Conceal
- Assess for bias in your data, analysis, and reporting
- To paraphrase American Frontiersman Davy Crockett -

"Try to be sure you are right, then Go Ahead!"

Additional Resources for Ethical Considerations in Data Science

- Awesome AI Guidelines
- Princeton University Center for Human Values and the Center for Information Technology Policy
    - Case Studies (“Case Studies” 2018)
- Markkula Center for Applied Ethics @ Santa Clara University
- Fast.AI

References

"A Framework for Making Ethical Decisions Science and Technology Studies." n.d. Accessed February 24, 2020. https://www.brown.edu/academics/science-and-technology-studies/framework-making-ethical-decisions.

"Case Studies." 2018. *Princeton Dialogues on AI and Ethics*. https://aiethics.princeton.edu/case-studies/.

"Ethical Guidelines for Statistical Practice." n.d., 7.

"Everything You Need to Know About the CCPA WIRED." n.d. Accessed February 17, 2020. https://www.wired.com/story/ccpa-guide-california-privacy-law-takes-effect/.

Raths, By David, and 05/02/18. n.d. "When Learning Analytics Violate Student Privacy -." *Campus Technology*. Accessed February 17, 2020. https://campustechnology.com/articles/2018/05/02/when-learning-analytics-violate-student-privacy.aspx.

"The Code Affirms an Obligation of Computing Professionals to Use Their Skills for the Benefit of Society." n.d. Accessed February 17, 2020. https://www.acm.org/code-of-ethics.