Ethics

CSCI 347, Data Mining

Professionalism

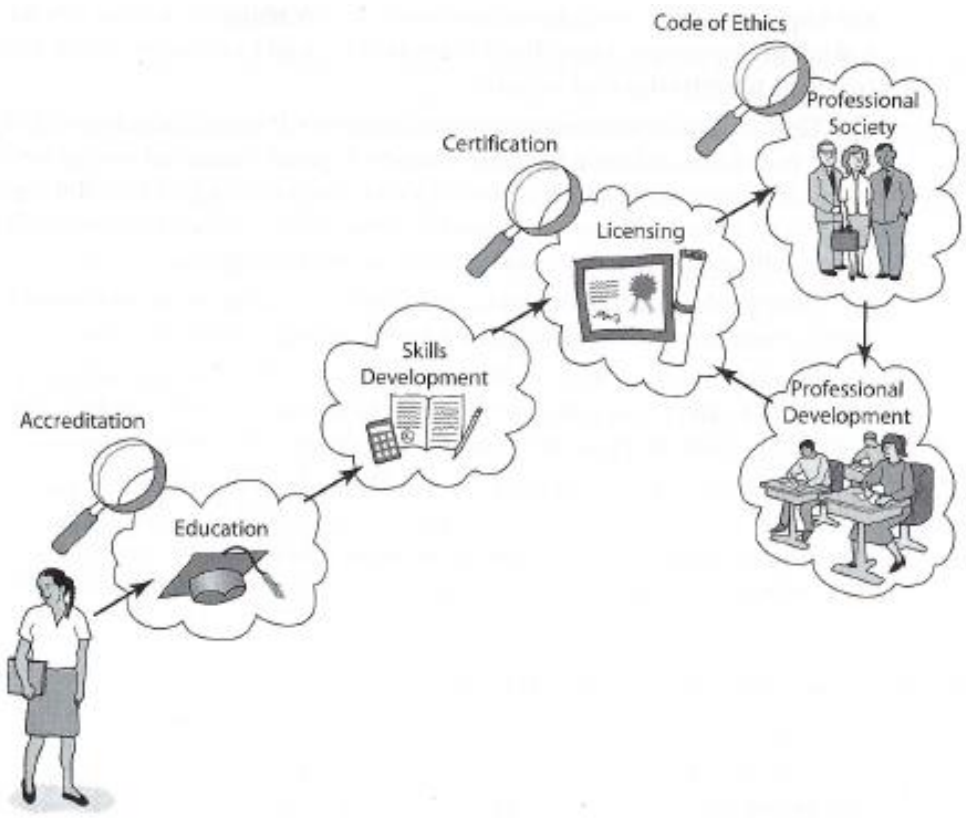
Professionals (doctor, lawyers, engineers)

- Have a duty to safeguard life, health, property, the public welfare and the environment.
- There is a need for self-regulation.
- Have a stake in ensuring that fellow professionals are capable and act appropriately.

Characteristics of a Profession

Components of a mature professional infrastructure:

- Education
- Accreditation
- Skills development
- Licensing
- Professional development
- Code of ethics
- Professional society



Code of Conduct

Data Science Association, 9 rules:

1. Terminology
2. Competence
3. Scope of professional services between client and data scientist
4. Communication with clients
5. Confidential information

Code of Conduct (continued)

6. Conflicts of interest
7. Duties to prospective client
8. Evidence, quality of data and evidence
9. Misconduct

https://www.datascienceassn.org/code-of-conduct.html

Ethics for Data Mining

Ethics particular to data science:

- Rule 5 - Confidential Information
- Rule 8 - Data Science Evidence, Quality of Data and Quality of Evidence

Rule 5 - Confidential Information

The use of data, particularly about people, for data mining has serious ethical implications.

When people give personal information, they have a right to know:

- How their information will be used,
- Who will use their information
- What is the purpose of using their information
- How will their information be protected
- How can they change the information kept if they discover it is wrong.

Reidentification

Reidentification – discovering the identification of someone from the data

- 85% of Americans can be identified just from zip code, birth date and gender
- 50% of Americans can be identified from city, birth date and gender
- Search terms (woman searched for landscapers in her home town and for people with her same last name)

Anonymizing Data

Anonymizing data – Making personal data so that reidentification is impossible

Rule 8: Data Science Evidence

Rule 8 - Data Science Evidence, Quality of Data and Quality of Evidence

(d) If a data scientist reasonably believes a client is misusing data science to communicate a false reality …. take reasonable remedial measures,…. if necessary, disclosure to the proper authorities.

(g) … use reasonable diligence when designing, creating and implementing algorithms to avoid harm.

(n) … use reasonable diligence to detect, recognize, disclose and factor real, perceived and potentially hidden risks in using data science. …

Potentially Harmful

Mathematical models used for the following can be harmful:

- Rank colleges
- Employment application screeners, resume sorters
- Determine credit worthiness used to grant/deny loans
- Determine likelihood of recidivism used to set parole, policing and sentencing algorithms
- Evaluate workers, teacher quality
- Target voters

https://blogs.scientificamerican.com/roots-of-unity/review-weapons-of-math-destruction/

Harmful Algorithms

Poor algorithms have 3 things in common:

- Opacity – black box, often proprietary
- Scale – the affect large numbers of people, increasing the chances that they get it wrong for some of them
- Damaging- encoding racism or other biases and/or enabling predatory advertisements