

CDS6214

Data Science Fundamentals

Lecture 12
Ethics in Data Science

Outline

- ❖ Definitions of Ethics, Security and Privacy
- ❖ Code of Conduct for Data Scientists

Definitions

- ❖ **Ethics:** moral principles that govern a person's behavior or the conducting of an activity.
- ❖ **Data Security:** means protecting data, such as those in a database, from destructive forces and from the unwanted actions of unauthorized users.
- ❖ **Data Privacy:** is the relationship between the collection and dissemination of data, technology, the public expectation of privacy, and the legal and political issues surrounding them.

Code of Professional Conduct

The Code of Data Science Professional Conduct defines 9 sections of rules:

Rule 1: Terminology

Rule 2: Competence

Rule 3: Scope of Data Science Professional Services between Client and Data Scientist

Rule 4: Communication with Clients

Rule 5: Confidential Information

Rule 6: Conflicts of Interest

Rule 7: Duties of Prospective Clients

Rule 8: Data Science Evidence, Quality of data, and Quality of Evidence

Rule 9: Misconduct

* We shall discuss those highlighted in *blue*

<http://www.datascienceassn.org/sites/default/files/datasciencecodeofprofessionalconduct.pdf>



Communication with Clients

- ❖ A data scientist shall:
 - ❖ Consult with client to accomplish the objectives
 - ❖ Act with diligence and promptness in providing service
 - ❖ Keeping client informed of the status of the services
 - ❖ Comply with reasonable requests for information
 - ❖ Consult with client on the potential hidden risks in relying on the data science results
 - ❖ Consult with client about any limitation on the data scientist's conduct when client expects assistance not permitted by the code of conduct or other law.

Confidential Information

- ❖ **Confidential Information:** **Information** that the data scientist creates, develops, receives, uses or learns in the course of employment as a data scientist for a client (directly or in-house) as an employee of an organization or as an independent professional.
Includes:
 - ❖ Information not known to public about the client
 - ❖ Information about affiliates, employees, customers or the parties the client has a relationship with
- ❖ A data scientist has a **professional duty** to protect all confidential information, regardless of form, from time of creation until its authorized disposal

Confidential Information

- ❖ A data scientist shall make reasonable efforts to prevent the inadvertent or unauthorized disclosure or access to information relating to the client:
 - ❖ Not displaying, reviewing, or discussing confidential info in public places or in presence of third parties
 - ❖ Not emailing confidential info outside of an organization to a personal email account, or removing confidential info from the client by acts of disposal or copy
 - ❖ Communicating confidential info only to authorized agents (attorneys, auditors) or those with legitimate reason to know

Quality of Data / Evidence

- ❖ A data scientist should inform the client of all results and material facts known to the data scientist that will enable the client to make informed decisions, whether or not the evidence are adverse
- ❖ A data scientist should persuade the client to use data science appropriately. If client is misusing or misrepresenting data science to communicate a false reality or promote an illusion of understanding • take reasonable measures:
 - ❖ Disclose to the client
 - ❖ Disclose to proper authorities

Quality of Data / Evidence

- ❖ A data scientist shall use reasonable diligence when designing, creating, implementing machine learning systems to avoid harm.
- ❖ Other Don't's
 - ❖ “Cherry-picking” results
 - ❖ Non-reproducibility of results
 - ❖ Claiming weak/uncertain evidence as strong
 - ❖ Using bad data or disclosing incorrect evidences
 - ❖ Etc.
 - ❖ (for an entire list of Do's and Dont's!...refer to:)

<http://www.datascienceassn.org/sites/default/files/datasciencecodeofprofessionalconduct.pdf>

End of Lecture 12

