From the Expert: Data Science and Social Justice

Week 1: From the Expert - Data Science

Data Science

Data Science is the study of data, and the use of tools and techniques to turn data into information. The data can come from many different sources and can be used in many different ways. Data is being collected about each and every one of us every day. Who collects the data? What do they know about us? What are they going to do with the data? Who has access to the data? How can I limit what is shared about me? What is my data footprint? These are a few of the questions we will be answering in this class. Review the following white paper on data science and ethics in higher education: Ethical Breaches – Tomorrow's Security Breach?

Data and Information

It is important to first have a good understanding of data versus information. Data and information are very different, although they have similarities. We'll first take a look at data, information, and how they fit into the data science picture.

Data

Data in its simplest form consists of 1's and 0's. Bits and bytes, consisting of 1's and 0's represent letters and numbers. Letters form words, words form sentences, and sentences begin to tell a story about the data. Data in itself doesn't tell us much without analyzing it. For example, a listing of 5000 customers that a company has sold goods to in the past that includes their names, addresses, and zip codes. At first glance a listing of 5000 names doesn't seem to be very useful, it is hard to draw much information from it. This is a good example of how data itself may not be very useful, it is the analysis of data, and representation of the results that enables us to describe it, identify data points, present results, and use it to make decisions. Data analysis techniques such as sorting, grouping, counting, graphing, and applying statistics can be used by the data scientist, and a picture of the data can start to emerge. Data represents a fact or statement of an event without relation to other things. Ex: It is raining. (Bellinger, 2004).

Information

Information describes data so it can be understood, further analyzed, and used to make decisions. Sometimes these are called "actionable insights". For example, given the data containing a list of 5000 customers, you can analyze it and discover that 20% of the customers are from Western U.S., 30% from the Central U.S., and 50% from the Eastern U.S. You could also do a count and see how many are from each area, and how many come from each state. These counts or percentages could be put into a bar chart and included in a report that provides customer source information, including the geographic locations with the highest customer density, as well as specific state and city statistics. This information can be used by company decision makers to influence future business decisions. This is just one example of how data that is turned into information can be used in data science.

Information embodies the understanding of a relationship of some sort, possibly cause and effect, or at least correlation. Ex: The temperature dropped 15 degrees and then it

started raining. (Gene Bellinger, 2004)

Privacy

With the advent of the internet, a plethora of private messages and information is available to organizations and individuals. You may end up working at an organization that has access to private information. It's important to be mindful of privacy laws and to respect private information. Without the owner's consent, their information may not be shared. Hermmann (2007) elaborates:

"The right to privacy is protected by privacy laws. Although exact provisions of privacy laws in each country differ, the common ground is restricting access to private residences and property, personal information, and personal communications...A person or organization commits a breach of privacy whenever they knowingly and without legal authority or consent from the individuals involved obtain, release, or disseminate the contents of private communications, regardless of the means of communications. Whenever a breach of privacy or invasion of privacy occurs, the victim has the right to pursue a legal remedy based on the contents of the applicable privacy laws and regulations. Both the individuals and the organization(s) responsible for the privacy violation can be prosecuted."

Ethics

Ethics are rules or standards governing the conduct of a person or the conduct of the members of a profession. Data science is a field that is growing and changing very quickly. The ability to leverage how data is analyzed and interpreted is vital to many fields such as medical, financial, business, social sciences, weather, government, and engineering. The way data is being gathered, used, stored, and reported is changing rapidly, as a result of ever-changing advances in technology the ethical standards for data science professionals are in the early stage of development. The Association for Computing Machinery (ACM) has established some ethics and privacy codes of conduct. ACM claims the title of "world's largest educational and scientific computing society." ACM was founded in 1947, it now includes over 100,000 members focused on computing and advancing the field both as a science and a profession. The ACM Code of Ethics and Profesional Conduct is relevant for understanding ethics and privacy in data science.

What is Social Justice?

We will explore data science and social justice in detail in upcoming weeks. For now, it is important to understand what social justice is, so we can later explore how data science can have a positive impact on the world. Read more about social justice from Professor Gorzycki (Gorzycki, 2017) to understand why incorporating social justice into education is important.

Social Media and the Internet

Data scientists of the past had to rely on structured tests and the solicitation of test subjects for their experiments. While these methods are still being used, the internet and social media have provided researchers with a wealth of information at their fingertips.

One way this has been used is to study human psychology. In the controversial Facebook study (Goel, 2017b), users' news feeds were manipulated in order to study how it affected emotions. As the New York Times put it, "To Facebook, we are all lab rats." (Goel, 2017b).

The study left many people feeling "violated", which was a surprise to Professor Jeffrey Hancock, the supervising professor of the study (Goel, 2017a).

History of Ethics Codes

This article by Metcalf (2014) provides background on the history and development of ethics codes. This article considers whether a data ethics plan should be required in data science projects like data management plans already are, and what they should address.

References

- Herrmann, D. S. (2007). Complete Guide to Security and Privacy Metrics: Measuring Regulatory Compliance, Operational Resilience, and ROI. CRC Press.
- ACM Ethics. (n.d.). Retrieved December 7, 2018, from https://ethics.acm.org/

- ACM Ethics. (n.d.). Retrieved December 7, 2018, from https://ethics.acm.org/
- Goel, V. (2017a, December 20). As Data Overflows Online, Researchers Grapple With Ethics. The New York Times. Retrieved from https://www.nytimes.com/2014/08/13/technology/the-boon-of-online-data-puts-social-science-in-a-quandary.html
- Goel, V. (2017b, December 20). Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry. The New York Times. Retrieved from https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html
- Gene Bellinger, D. C. (2004). *Data, Information, Knowledge, and Wisdom.* Gene Bellinger.
- Gorzycki, Meg. Social Justice in the Curriculum - The Center for Teaching and Faculty Development - SF State. (2017, November 21). Retrieved December 7, 2018, from https://web.archive.org/web/20171121062908/http://ctfd.sfsu.edu:80/content/social-justice-curriculum
- Metcalf, J. (2014). Ethics Codes: History, Context, and Challenges. Council for Big Data, Ethics, and Society. Retrieved from http://bdes.datasociety.net/council-output/ethics-codes-history-context-and-challenges/.

Additional Optional Resources

Videos

- O'Reilly. (n.d.). O'Reilly Webcast: An Introduction to Ethics of Big Data. Retrieved from https://www.youtube.com/watch?v=PsC9CMgyTxY&feature=youtu.be

Readings

- O'Reilly Media. (n.d.). The Ethics Of Big Data. Retrieved December 9, 2018, from https://www.forbes.com/sites/oreillymedia/2012/06/21/the-ethics-of-big-data/
- Regulation of big data in the United States - Taylor Wessing's Global Data Hub. (n.d.). Retrieved December 9, 2018, from https://globaldatahub.taylorwessing.com/article/regulation-of-big-data-in-the-united-states

Activity Details

Task: View this topic