Arnoldas Kurbanovas INF 202

**End of Module Assignment 4: Big Data and Privacy**

Big data is a term for the gathering of big and complex data sets and the analysis of these data sets for relationships. Because big data is so large, thousands of gigabytes, that of the data collected in these sets, it prevents traditional methods of analysis from being effective. Instead of concentrating on specific relationships between individual pieces of data, big data uses several different algorithms and techniques to assume general trends over the entire set. What people need to understand is that what counts is not the quantity or how much data you have gathered, but it is the quality, and how useful will this data be. While there are definitely benefits of analyzing big data and using it, the current methods for doing so often tend to invade individual privacy and fail to protect individual rights. The challenges to privacy arise because technologies collect so much data so fast. By an individual using a smart phone, using credit cards, singing in at the apple store, all these things make up big data.

This FBI video of Hasan Elahi shows how big data was used in a way to put him on the FBI watch list by mistake. During this video Hasan went ahead and to protect himself documented all of his adventures. By taking pictures of everywhere he went you can cross reference that to see what the location was and kind of follow along and back track his footsteps to see where he was at what time. This is pretty much what the FBI did to watch him and to see if the information he provided was accurate. Because Hasan does thing differently then most people in the United States, or at least in a sense of how data is stored about each individual, Hasan stood out and was put on the FBI watch list because he didn’t fit the correct patterns. The patterns of the rest of Americans that created these sets that a computer essentially analyzes and has a base of what to compare it against. Which is the large majority.

This is a perfect example of how the data that was collected to put him on the FBI watch list was not accurate enough and made a mistake. The algorithms and the systems that are used to turn this big data into information that people can look at and understand is not bullet proof and not even close to perfect. These systems that analyze big data rely on imperfect inputs of data, flawed logic and inaccurate probabilities. Because these systems are created by humans it needs to be understood that they wont be perfect and can’t be taken as correct. This article/video is also a great example of how the government knows everything about us. Where we shop, were we travel, what we eat and how we eat it. While big data can be great for some things like businesses seeing what the majority likes to eat and do, it also faces the invasion of individual privacy. Makes a good point that because all our information is out there and stored, the best way to protect it and to protect ourselves is to take control over the information. By taking control over your individual information and making everything available, then in a sense you are devaluing that information. Meaning that if everybody was sharing everything, including all the nonsense and useless information, which is starting to happen in the world with social media. Then the value of the information gets devalued were as if everything was private and needed to use inaccurate algorithms to obtain it.

While there are many benefits to the growth of big data analytics, traditional methods of privacy protections often fail. Big data focuses more on the what, rather than the why in a sense when it comes to information. Many purposes of privacy rely on informed consent for the disclosure and use of an individual's private data. However, because big data can be used and reused again this means that data is often used in ways that were un-thought of at the time the data was collected. Security of personal information is also gone in a big data theory that we have today. Even if everyone’s data that is collected is anonymous and the individual pieces of information are stripped of anything that could be lined to a specific person, an individual's identity can still be revealed just by the relationships alone between the pieces that are still left that were gathered.

Article used: <http://www.ted.com/talks/hasan_elahi#t-500036>