Adam Forestier

Professor Chen

IST 3420

August 27, 2017

Data Science Concerns Essay

Data science is a valuable method for companies to gain useful customer insights to give them a competitive edge in both retaining and acquiring new customers. To gain these insights, consumer data must be gathered, stored, and analyzed. Many legal, policy, and ethical concerns may arise during all steps of the data science process. Companies must be careful to be compliant with privacy laws, clearly state what information they will be gathering and storing in their user agreement statements, and have strong company values about what data is ethical to utilize for business purposes.

Privacy laws regarding personally identifiable information (PII) can be a real concern for data scientists. “The U.S. Senate proposed the Privacy Act of 2005, which attempted to strictly limit the display, purchase, or sale of PII without the person’s consent” (wikipedia.org). Two special areas of concern regarding privacy laws are protected health information (PHI) and social security numbers. Protected Health Information includes a person’s health status, the payment for their health care, and also what care was provided to the individual. This is to keep personal health information completely confidential between the patient and provider. Information may only be released with the authorization of the patient. The government also tries to limit the collection and sharing of individual’s social security numbers. Social security numbers can be used to steal identities; thus data leaks, hacking, or sharing this information could compromise customer’s identities. Data scientists must be careful to remain completely compliant from a legal standpoint when collecting information, storing it, and performing analyses on personal data.

Companies must tread carefully from a policy standpoint during the data science process. Usage agreements from companies and websites should warn users what type of data will and will not be stored. It is law in European Union countries for websites to inform users that they are collecting user data through cookies. However, I believe it should not only be businesses in these nations that inform users that their preferences and data are being stored to the web browser. In my opinion, I think that all companies should inform users on their site that cookies are being used to store preferences and some data. Businesses should gain consent from consumers to store and utilize data, to avoid lawsuit and anger from their customers.

From an ethical standpoint, companies should outline what data can and cannot be gathered and used by its data scientists. This keeps the data science process from being grey and ambiguous. I believe that a good rule of thumb is that if you would not feel comfortable with a business having a certain form of your own personal data, then you should not collect similar data from others. One of the most important parts of customer loyalty is trust of the brand. If consumers feel like they are constantly being spied on by a corporation, their level of trust in the brand may diminish to the point where they take their business to competitors. It is vital for data scientists to gather, store, and analyze data; however, they must remain privy to individual’s feelings of privacy and avoid data that could damage the reputation of the company.

In the modern world, data science is a necessity for large corporations to gain competitive advantages over competitors through improved marketing techniques, product promotions, and making decisions about joining new industries. As with every great stride in technology, there are concerns. There are legal, policy, and ethical questions that must be answered in each stage of the data science process. While gathering, storing, and analyzing individual’s information, data scientists must remain legally compliant, clearly state what data will gathered and the purpose for it, and contemplate what is ethical data to work with.