

Honors Case Study Challenge Entry Form

Are you submitting this entry as an individual chapter member or as a chapter?

Individual Member Entry

Chapter Entry

Please fill out the corresponding entry fields below.

Individual Member Entry Information

Title of Case Study	
Individual Member's Full Name	
Address	
City, State, Zip	
Phone #	
Email Address	
Phi Theta Kappa Chapter Name	
Advisor Name	
Advisor Phone #	
Advisor Email	
School Name	
School City and State	

Chapter Entry Information

Title of Case Study	Big Data: Mother Lode Or Minefield
Phi Theta Kappa Chapter Name	Beta Theta Omicron
Advisor Name	Christine Case
Advisor Phone	650-738-4376
Advisor Email	case@smccd.edu
School Name	Skyline College
School City and State	San Bruno CA

Sponsored by Phi Theta Kappa.

For questions or additional information email honors.programs@ptk.org or call 800.946.9995.

Case Study Title:

Big Data: Mother Lode Or Minefield

Article Information:

USA Today

Consumers have been giving their information away

Matt Krantz

January 22, 2014

MONEY

page B4

USA Today

They're studying you

Sharon Jayson

March 11, 2014

MONEY

page B6

USA Today

NatureBox delivers; buyers eat up

Laura Baverman

April 15, 2014

MONEY

page B6

USA Today

White House raises concerns about 'big data'

Aamer Madhani

May 2, 2014

NEWS

page A5

USA Today

Patients' social media don't lie

John Shinal

May 15, 2014

MONEY

page B2

Summary Statement:

Recording information dates back to 3000 B.C.E. when the Sumerians began using pictures to represent ideas. In 1300 B.C.E., the Chinese began to form letters that are used today. Scribes recorded information on large clay tablets until the development of paper in 500 B.C.E. In the fifteenth century, the Gutenberg Press allowed widespread communication of discoveries and concepts through mass produced books and journals.

In the 1980s, magnetic and electronic storage devices enabled storage of more data than could be collected on paper. And digital communication including web sites, blogs, and tweets resulted in an unprecedented publication of information. Electronic media and cloud servers enable businesses, economists, scientists, and health care providers to store large amounts of data. However, these data are uncataloged and therefore inaccessible. These unmanageable digital data are now being called Big Data and offer a frontier of exploration with the new science of data mining. Data sets so large, complex, or rapidly-generated that they can't be processed by traditional technologies have been dubbed big data. Data mining or Knowledge Discovery in Databases (KDD) uncovers patterns and correlations in big data using statistical analysis and computers. Web search engines provide a familiar example of the benefits and current limitations of accessing data. The user types a few keywords and gets back links that need to be read and evaluated. KDD methods in development would provide more relevant results to the user's query.

One of the most visible examples of big-data is the efficiency of targeted advertisement — that is, pop-up ads targeted to a consumer's buying or even Internet browsing. Search engines can gather detailed information about potential customers and their specific product interests from social media. Brick-and-mortar retailers can access the Wi-Fi connection on customers' smartphones to pinpoint their location and amount of time spent in a particular department. "Keyword tagging" allows advertisers to tweet based on a Twitter user's immediate behavior.

In health care, big data is changing the way that patients, medical professionals, and researchers communicate. Traditionally, physicians have used their judgment when making treatment decisions. Currently, stored health care data are being made searchable and usable. The integration of data from clinical trials and millions of patients can reduce subjectivity in decision-making and provide new useful information to health care workers. Additionally, with access to analyses of patient responses to treatments, patients can have informed discussions with their physicians.

Since 1791, the fourth amendment has protected citizens from unreasonable search and seizure of the property by the government. Today, a person's "virtual property" is a combination of telephone, Internet, and cloud storage. The fourth amendment specifically protects people and their physical places, such as a home or car. The 1986 Electronic Communications Privacy Act protects electronic storage. Nevertheless, in 2013, a computer specialist working for the CIA reported that the United States government was conducting mass-surveillance of electronic records of private citizens who were not under suspicion for any criminal activity.

Discussion Questions:

1. Grocery stores analyze individual's purchasing from their store loyalty cards. The Starfish grocery store sends discount coupons to customers who spend more money but does not provide discounted coupons to customers who spend less money. Are poorer income people and smaller households being discriminated against? Discuss whether loyalty cards are an advantage for the consumer.
2. Predictive data mining analyzes historical data to make predictions about the future.

BigBox Mart assigns every customer a Guest ID number tied to their email address, Facebook page, or IP address that becomes a basket that stores a history of searches, purchases, friends, and information purchased from other sources. An analysis of their data revealed that women on the baby registry buy larger quantities of unscented lotion around the beginning of their second trimester. And sometime in the first 20 weeks, pregnant women buy vitamins and calcium. Someone who suddenly starts buying scent-free soap and large bags of cotton balls, in addition to hand sanitizers and washcloths, is close to their delivery date. Analysts developed a BEST (baby estimator) score to estimate when the baby would be born. Then BigBox started sending out advertisements and coupons for baby supplies.

Very upset, Mr. Perry told the store manager "My daughter got this [advertisement] in the mail! She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?" The manager didn't have any idea what the man was talking about. And the man didn't know that his daughter was pregnant.

Was the girl's privacy violated? Should ads be targeted to your online behavior? ? Should ads be targeted to your physical location (from the GPS on your smartphone or car)?
3. An Internet service company offers free email accounts to everyone. The company scans its clients' emails to block pornography and other illegal activity. It also mines client emails looking for words and context for sending target advertisements. Should these ads go to a child's email? A teen's? An adult's? A compulsive shopper's? Is scanning emails a justifiable procedure?
4. The English author, George Orwell, was influenced by events of his time including Hitler's Germany, Stalin's Soviet Union, and Franco's Spain. To prevent future repressive dictatorships, in 1949 Orwell wrote *1984* to warn people to strive for a society that fosters individual freedoms. In *1984*, Orwell created a technologically advanced world where people are watched at all times by "telescreens"— at home, at work, in restaurants, in bathroom stalls. The "thought police" monitor every move. Individuals who did not conform to the government's values disappeared and were presumably killed.

In 2014, data from license plate readers, web searches, and cell phone calls can be combined with facial recognition. These technologies are currently being used to identify terrorists and locate stolen cars.

Discuss whether current data mining technology will lead to Orwell's *1984*. Include a comparison to the political climate in the world during the time when Orwell wrote.
5. In 1950, the United Nations Educational, Scientific, and Cultural Organization issued a statement that "race is not so much a biological phenomenon as a social myth" and

Sponsored by Phi Theta Kappa.

For questions or additional information email honors.programs@ptk.org or call 800.946.9995.

suggested to “drop the term ‘race’ altogether.” Genetic evidence shows that 90% of the DNA in all humans is the same but about 10% of the human genome is more common in certain ancestral groups. For example Jews are offered prenatal tests for the inherited Tay-Sachs disease gene. The heart-disease drug BiDil is marketed exclusively to African-Americans, in which a modification of the ABCA1 gene is common. Some asthma medications can shut down the lungs of people with a certain mutation in the beta2AR gene that is more common in African Americans. Han Chinese are more likely to have adverse reactions to methotrexate, a drug commonly used to treat rheumatoid arthritis.

Should genetic data mining be used to define races in order for drug therapy to be useful? Will this lead to discrimination?

Future Implications:

We are at the edge of a new frontier. The Federal Big Data Research and Development Initiative plans to make these enormous quantities of digital data more useful to researchers, businesses, and policy makers. Disease outbreaks, carcinogens, and unexpected drug side effects could be detected early. For example, Google Flu aims to predict influenza outbreaks in real time based on queries about flu symptoms.

However, privacy issues are the primary concern about Big Data. Data brokers collect personal information about consumers and sell that information. A private company uses data mining to help state agencies find fraudulent health care claims. Who will have access to your health and financial information? Will your employer know about your job searches? Federal legislators have begun questioning the reach of private individuals and companies and government agencies in data gathering. Over 80 countries have data protection laws. In the U.S., a patchwork of federal and state laws and industry guidelines address individual privacy. The 2014 White House privacy report recommends that consumers have access to their personal data collected and be allowed to correct inaccurate data.

Will we need to give up some individual privacy to get the advantages of knowledge from Big Data?

Additional Resources:

Furnas, A. Everything you wanted to know about data mining but were afraid to ask. *The Atlantic Magazine* 3 April 2012. Web. 6 June 2014. <<http://www.theatlantic.com/technology/archive/2012/04/everything-you-wanted-to-know-about-data-mining-but-were-afraid-to-ask/255388/>>

Mayer-Schönberger, V. and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt, 2013.

Office of the President. *Big Data: Seizing Opportunities, Preserving Values*. May 2014. Washington, D.C.: The White House. Web. 9 September 2014. <[big_data-privacy_report_may_1_2014.pdf](#)>.

Pappalardo, J. “NSA Data Mining: How It Works.” *Popular Mechanics* October 2013: 59-62.

Sponsored by Phi Theta Kappa.

For questions or additional information email honors.programs@ptk.org or call 800.946.9995.

Stein, J. "Your Data, Yourself." *Time* 21 March 2011: 40-46.

Tsisis, A. "The Right to Erasure: Privacy, Data Brokers, and the Indefinite Retention of Data." *Wake Forest Law Review* 49 (2014): 433-484 Web. 3 October 2014.
<<http://lawcommons.luc.edu/facpubs/494/>>.

Sponsored by Phi Theta Kappa.

For questions or additional information email honors.programs@ptk.org or call 800.946.9995.