Data Mining

Craig Chomsky, Marek Dvorak

Introduction to Data Mining

The use of data mining has become an issue of ethics over the past decade. Technology is allowing new forms of information extraction with the use of data mining. Data mining is used for different purposes and the possibilities are almost endless. Data mining is being used to exploit certain groups of people with disadvantages. Politicians use it during campaigns; and through its use, personal privacy is invaded. Data mining is a dangerous tool that can be used to take advantage of anyone at any time.

"Data mining is the process of extracting hidden knowledge from a large volume of data."[1] Companies can obtain types of data through large servers that store the data. Individual businesses have their own storage facilities, which thousands of other companies are granted access to. It provides a competitive advantage for companies when selecting groups of people to advertise to. In the past, it may have been more difficult for companies to advertise to specific groups of people. But with data mining, all that has changed. In today's world, enormous amounts of data are readily available to companies due to data mining. The process of analyzing data becomes automatic with data mining because it finds relationships and patterns in raw data. With the newly provided information, decisions can be made either automatically by a support system or evaluated by a human.[1]

The Benefits of Data Mining and How Information is Obtained

Data mining can provide many types of advantages for companies to use in their daily activities. Data mining can be used to help identify a company's best prospect and to maintain them as a customer for a long period of time. It will also help them to save money and time by concentrating their marketing activities only on those who are interested in their products. Therefore, their success in marketing increases. Data mining can help companies in their selling

activities by increasing their cross-selling opportunities. Information is provided from many different sources, which allows predictions to be made on what products customers will like. Data mining can also help companies to segment their markets and to personalize their communications between them. Wide varieties of customers require companies to take different approaches while advertising to these different groups. Data mining makes the task of targeting different groups of customers easier by providing so much information.[1]

Data mining consists of five elements which provide the information needed to create the database. The first stage is to take data and transform it so that is can be put into a data warehouse system. The next step is to store and control the data in a multidimensional database system. Then, professional analysts are allowed access to the data so they can take the next step, which is analyzing information with software. Finally, the data can be displayed in legible formats (tables or graphs) in which the correct decisions can be made.[2]

Overall, data mining is used to provide an easier way for businesses to access information on a wide scale. This information is used to analyze external and internal factors that allow companies to maximize their profits. Some internal factors are product, price, and promotion. External factors are customer demographics, economic conditions, or the level of competition.[2] All of these elements are analyzed through the process of data mining and this new technology is being used worldwide.

Thesis

Data mining is known to be beneficial to businesses, politicians, and consumers, though it also plays a major role in influencing the lives and decisions of U.S. citizens on a daily basis.

Gathering Information with Data Mining Can be Harmful

Data mining is an invasion of an individual's privacy, specifically through the privacy from corporations. Through the use of gathering information on an individual, much can be compiled to create an image one would not want known to the public or to the business world. A substantial number of companies across the nation use data mining in order to collect information on consumers. All the information gathered between these companies gets stored on large databases which many different people can gain access to. In some cases, the data that is being used by a company becomes public information in which anyone can gain access to. The

problem is personal information about consumers is being released to the public without the consent or knowledge of the person being exploited. Many people do not see how this could be an ethical dilemma because they do not realize the potential harm it can cause to an individual. However, if one were to compare the situation to a more visible case, they might see how devastating data mining can really be.

Let us imagine a similar situation where John allows Rick to borrow his car for a month. Rick decides he wants to make some money with the car so he began lending it for money to Sandra, Mike, and Joe. When John gets his car back, he would not know that other people used the car if everything looked fine. Though, if the car were wrecked, he would go to Rick and ask what happened. Rick would explain how he lent the car out to friends of his and they messed up the car and it is not his fault. Rick's actions were obviously unethical and he should be responsible for the damage done to the car. Data mining is a new technology that is lending information out to other businesses for money or sometimes for free. Consumers are unaware of how the information collected about them is being used. The information can be used to help benefit customers, but more importantly, it can cause a lot of harm and damage to their personal lives.

Evidence of the Misuse of Information Obtained Through Data Mining

In October of 2004, information was stolen from a nationwide database that contained personal information on almost every U.S citizen. The information was gathered with data mining and was put into a supposedly secure database. However, hackers were able to penetrate the security systems of the Georgia based ChoicePoint firm. Nearly 145,000 U.S. citizens had information such as bank account, credit cards, and social security numbers stolen without their knowledge. The people who had their personal lives invaded were not even made aware of the situation until four months after the fact.[3]

A few months after the ChoicePoint situation arose, LexisNexis also had one of their main databases broken into. Hackers stole personal information on over 300,000 customers.[4] California is the only state that requires in writing that the company tell customers their personal information was stolen. The other forty-nine states do not have this law so many of the customers did not find out for a long time that they had information stolen.[3]

Ever since ChoicePoint and LexisNexis databases were hacked into, "public and congressional concerns have increased"[4] concerning the use of data mining and how information is stored. With the increased technology used for data mining, it has become much easier to obtain personal information about U.S citizens. The companies who use databases to analyze customer behaviors and patterns are not protecting information properly.[4]

Taking Advantage of Specific Groups Through Data Mining

Data mining forms internal groups separating people based off specific characteristics from each individual. These groupings are created to make assumptions about how people will act in response to advertisements and other forms of incentive based promotions. Companies target those groups who can be easily influenced. Companies can use databases, which were created because of data mining, and select specific information about groups they want to exploit.[5] One group that can be easily influenced and has high buying power is the elderly. The elderly are retired and in many cases have a lot of money. They sometimes live alone, are often times bored, and are looking for any contact they can find. Most importantly, their decision-making processes have weakened over the years. In the Techtalk video, data mining was one of the main topics discussed. Professor John Riedl of the Computer Science and Engineering, University of Minnesota, talked with Susan McKinnell about an example of how data mining is used to target groups with specific traits in common. The elderly became a target of one company because they compiled a list of all the old people with the suspected symptoms of bladder control issues.[6]

Data mining is once again shown to be unethical because it is used to take advantage of a specific group whose decisions are easily influenced. Data mining gives companies the ability to separate people into very specific groups that all act in different ways. Companies take that specific information about individual groups and exploit any angle they can get. Data mining allows the companies to play an unethical game against their customers. Companies gain the ability to predict how customers will react to certain types of advertisements and other promotions.

Why Certain Groups Need Protection

Taking advantage of the elderly with the use of data mining is not the first time they have been exploited by advertising. In 1987, Janet Shikles wrote the Chairman of the Select Committee on Aging House of Representatives, Edward R. Roybal on the subject of direct advertising to the elderly. Janet Shikles was outraged about how the "aging organizations attempted to solicit funds, sell insurance, and offer direct mail advertising of products in the manner that frightened, threatened, or otherwise coerced the elderly into contributing money or buying products from these organizations."[7] Shinkles was saying to the chairman that the "highly vulnerable group such as the elderly and minorities"[7] need to be protected from being taken advantage of. There are types of enforcement that have helped to prevent direct advertising through the mail to the elderly. However, today with the Internet, there is no protection; so companies keep using data mining to influence these groups.[7]

Politicians Use of Data Mining to Influence Voter Decisions

Data mining has entered the political field and makes it easier for politicians to win elections. Politicians use endless amounts of information on voters to see what the voters are thinking. They have recently started using data mining to help sway the votes of those undecided about whom they are going to elect into office. Voters' ability to make a decision on who they will vote for is strongly influenced by advertising based off the information from data mining. The information is used in many ways, but is specifically targeted at the small percentage of voters who are undecided about their vote. Politicians can gather enough information about those people so they can sway their vote one way or the other to win an election. Advertising might push those who are easily influenced or less educated into making decisions they would not normally make.

The use of data mining is unethical and will most likely not change. The politician who uses data mining has a much greater chance of swaying votes in their favor and winning elections. However, laws against data mining will most likely not change because politicians will never admit they did anything unethical to win their way into office. Also, the politicians do not want to lose their ability to sway votes with the use of data mining for their next political campaign.

Examples of Politicians Using Data Mining

In the 2004 elections for President of the United States, President Bush's campaign data mined over 10 million voters to find out personal information. The types of information they looked for ranged from "the car you drive, the magazines you read, the catalogs you buy from, the house and neighborhood where you live, and your voting pattern which helps them know which button to push to win your vote."[8]

In 2006, Governor Rick Perry's campaign modeled themselves after President Bush's with the use of data mining to win his election. Governor Perry's campaign used data mining to influence voters. He used information based off previous elections and "the campaign targeted those that, without the right push, might not have voted."[8] Governor Perry's campaign ran extensive tests. They looked at every bit of information they could to make the right types of advertisements to influence voters. Data mining looks for very specific information which helps the campaign to direct funds to the right places. In both President Bush's and Governor Perry's campaign, they used data mining to influence voters.[8]

Concluding thoughts

Data mining gathers all types of private information about consumers and uses it to make business decisions for the future. The act of taking people's information through the Internet without their knowledge or permission is called the invasion of Internet privacy.[9] Many people are not aware of the information being circulated about them or how it is being used, which makes many feel that data mining is unethical. Data mining is also used to exploit certain groups of people who are easily influenced by advertising. These groups usually have disposable income and data mining provides a way to take advantage of them. Companies gather certain pieces of information about the elderly, which can lead to a larger profit. The elderly are being taken advantage of and the name for it is privacy from corporations. Politicians also use data mining in an unethical manner for their own personal gains.[9] They use the information from data mining to sway votes in their favor during elections. Data mining has been proven to be a useful piece of technology for businesses to maximize profit, but it has also been proven to be a dangerous tool that influences the minds of people everyday. Unfortunately, many businesses today would choose to use unethical practices to make financial gains even at the cost of invading the privacy of their own consumers. These companies risk the safety of the consumers' information and put forth little effort to keep the data from being stolen. Data mining is an

unethical piece of technology that takes advantage of consumers all over the world by invading their privacy, putting their personal information at risk, and warping their ability to make decisions for themselves.

Works Cited

[1] "Data Mining," [Online Document] No Date, [February 28, 2007], Available at: http://www.megacomputer.com/dm101.php3?gclid=CLvgx6CjyooCFQUgggodxAFOgw

[2] "Data Mining: What is Data Mining?," [Online Document] No date, [February 28, 2007], Available at: http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm

[3] "Private Information Stolen from Nationwide Consumer Database," [Online Document] February 16, 2005, [April 8, 2007], Available at: http://www.consumeraffairs.com/news04/2005/choicepoint.html

[4] C. Strohm, "Federal data-mining efforts fail to fully safeguard privacy, GAO says," [Online Document], August 30, 2005, [April 8, 2007], Available at: http://www.govexec.com/story_page.cfm?articleid=32114&ref=rellink

[5] "Data Mining," [Online Document] No Date, [April 8, 2007], Available at: http://www.sentient.nl/dataminframe.html

[6] "Data Mining Video," [Online Document] No Date, [March 3, 2007], Available at: http://techtalk.umn.edu/episodes/season5/508.mp4

[7] J. Shikles, "Letter to the Chairman," General Accounting Offices, ed 3, no. 4, pp76-79.

[8] "Politicians, the Car You Drive," [Online Document], No Date, [March 3, 2007], Available at: http://www.dallasnews.com/sharedcontent/dws/news/elections/2006/stories/072606dntexvotetarget.36ee531.html

[9] "Privacy," [Online Document], January 24, 2006, [March 6, 2007], Available at: http://en.wikipedia.org/wiki/Right_to_privacy