Data Mining as a Business Tool

Spring 2009
MIS304 (21004)
Dr. Fang

Prepared By:
Jason Baltazar
Philippe Cademas
Jillian Latham
Rachel Peeler
Kamila Singh

May 11th, 2009
INTRODUCTION

Today’s business is globalized, and technology has been the catalyst. Boundaries are being pushed farther than ever and businesses everywhere are finding new and innovative ways to use technology. The increase in the use, and popularity, of the Internet has created huge benefits as well as potential threats to users. Data Mining is a business intelligence tool that, depending on how it’s used, has this dual potential. Sorting through massive amounts of data produced by consumers today is impossible without the use of data mining. So, the logical question is: how can businesses make use of, and decipher, all this information? This paper sets out to explain what data mining is, the various ways it can be used as a business tool, the advantages and disadvantages of using such tools, and the ethical implications that follow.

DATA MINING BASICS

There are two basic types of data mining: unsupervised and supervised. Unsupervised Data Mining is sometimes called descriptive modeling because it is used to find patterns and relationships among data. This is accomplished when analysts apply data mining techniques to the data and examine the results (Kroenke 341). This is done to reveal any relationships that may exist between variables. One common technique is cluster analysis.

Unsupervised Data Mining Techniques

Cluster analysis, also known as automated data mining, can be used to discover the segments or groups within a customer data set. A common use for cluster analysis is to find
groups of similar customers from customer orders and demographic data. This type of analysis is appropriate when those groups of similar customers (or classes) are not already known. These groups are identified by the program, not the analysts. Within the context of business, this can be used for marketing and advertising. Developing the “clusters” or groups of customers that naturally fit together is the first step in data mining. Once customers are segmented based on cluster analysis, companies will often develop different marketing and advertising strategies for each cluster. This is obviously based on further analysis and predicting buyer behavior.

**Supervised Data Mining Techniques**

Supervised Data Mining is sometimes called predictive modeling. Data miners set goals and research parameters before starting the data mining process (Keating, 2008) and then apply statistical techniques to the data. This is done in order to concentrate only on the patterns that are relevant and important. Supervised data mining is different from unsupervised data mining in that it already has a pattern identified and is used to predict outcomes based on those predetermined parameters. There are several subcategories of supervised data mining which we will discuss further. We will discuss anomaly detection, classification and prediction, regression, and attribute importance and we will also illustrate ways in which these methods can be used within the business sector to make sound business decisions.

**Anomaly Detection**

Anomaly Detection is a process by which models are built to specify a “normal” range of results and anything falling outside of the normal range is atypical and is identified as either an outlier or an example of a previously unseen category. This type of data mining is particularly useful for detecting fraud within the credit card, tax, and insurance industries (Oracle, 2005). For instance, a credit card company already knows, based on previous analysis, which transactions
are likely to be fraudulent. They can build models to predict these fraudulent transactions and when one comes through their system with all the likely attributes, it will be flagged for further investigation. Anomaly detection can be also be used to detect and prevent identity theft and breaches in computer security.

PayPal is used by numerous websites in 190 different markets to facilitate financial transactions between two parties. During the fourth quarter of 2008, PayPal facilitated nearly $16 billion in transactions worldwide (Wikipedia). So, the need for fraud detection should be obvious. PayPal, in an effort to be proactive, is using a sophisticated data-mining program that creates a “risk profile” for each customer based on previous associations made from lessons learned with regard to fraudulent transactions (Jennings). Once the customer is profiled and given a score, PayPal has the option to initiate one of several actions: allowing full access and service if deemed to be authentic, limiting activity, or requesting further verification of identity (Jennings). PayPal also creates profiles for their merchants. This is done to prevent cases where customers purchase merchandise and never receive it because the vendor is deceptive or non-existent. Another way PayPal fights fraudulent transactions is by verifying all PayPal accounts. Accounts can be either verified by adding and confirming a bank account or credit card. PayPal also recommends to all sellers/merchants to ship to a verified address only.

Classification and Prediction

Classification, or “Prediction”, is the most commonly used data mining tool because it seeks to answer the all-important questions: Who will buy, what will they buy, and how much will they buy (Mena, 10)? This is accomplished by using historical data (patterns) to predict outcomes within the framework of new data. Classification can be used in customer segmentation, business modeling, credit analysis, and many other applications (Oracle, 2005).
Credit analysis differs from our credit example above because credit analysis seeks to identify customers who are likely to default based on parameters within a model such as customer’s spending habits, income, and/or demographic characteristics. It is important to note, however, that there are ethical implications to consider when making determinations based on demographics as this may be perceived as profiling or discrimination if that data is used to preclude customers based on their demographic attributes. Credit scoring (or credit analysis) is widely used within the business sector because businesses want customers who are fiscally responsible and are “good credit risks.” The past financial behavior of these customers can be used as a good predictor for future financial behavior.

Another interesting business application for data mining through the use of classification is in the area of human resources. Data mining can be applied to affect turnover analysis, employee development, recruiting, training, and employee retention. Recently Cataphora, a company normally tasked with detecting statistical signs of fraud, embarked on a study to apply data mining to employee performance in order to determine which employees come up with winning ideas (Baker). In these tough economic times, companies are often faced with decisions with regard to downsizing and retention. These companies are applying data mining techniques to assist in the decision making process and ensure that they are retaining the leaders of tomorrow. One way to conclude who these “valuable” employees are is to classify groups of “thought leaders”, “networked curators” and “bottlenecks.” Thought leaders are employees whose thoughts (via text or software) are copied most often. Networked curators are those employees who spot valuable content and pass it on, and bottlenecks hold up ideas (Baker). So, if you’re a corporation, like so many today that need fresh ideas and innovative thinkers, this new way of determining the value of employees might be worth your consideration.
Another way in which data mining is benefitting business is in regard to recruitment. When looking to fill management positions, most companies look outside their organization or branch to find leadership candidates. Data mining is now being used to identify candidates from within the organization who might not have otherwise been considered (Ng). Correlations are established between top performers and their particular skills, attitudes, and competencies. It is concluded that people with similar attributes are likely to perform at the same caliber. This has opened the door to a wealth of candidates who possess the same core competencies as top performers but may not have the experience or pedigree as those outside the organization. This is an excellent method for determining which employees to retain, which ones to consider rewarding with incentives and pay raises, and when to start grooming them for upper management. This is the most cost effective way to solve the talent shortage that many organizations face today (Ng).

Regression Analysis

Regression analysis measures the impact of one set of variables on another variable (Kroenke, 342). This is another technique of data mining in which statistics are applied to data to make predictions about the future. This is especially useful in the field of business because it can be applied to various different aspects of business in order to assist in decision making. Some of the common areas in which regression analyses are performed are marketing, pricing, product positioning, sales forecasting, advertising, human relations, customer service and customer relations, customer choice modeling, and even credit scoring and fraud detection (Miller, 25) to name a few. The following graph is an illustration of how a regression analysis might look when searching for a correlation between new home sales and baby food sales.
Market response modeling and sales forecasting are several objectives when using regression models. Companies should want to analyze how the market and subsequent sales change as a result to changes in product price, advertising, product placement, promotions, and distribution (Miller, 46).

TEXT MINING

Text mining is the process of automatically processing text and extracting information from it. Algorithms are developed that determine resemblances between strings of words from many different sources. This type of data mining can be useful to help identify the types of stories that are most popular and the websites that report on them first. President Obama made history when he used this type of technique during his campaign. Data mining systems helped him better understand what people were talking about and what people wanted to know.
Text mining can be seen as consisting of two phases: text refining and knowledge distillation. Text refining transforms free text documents into a chosen intermediate form. Knowledge distillation then deduces patterns or knowledge from the intermediate form.

Intermediate form (IF) can be semi-structured or structured. It can be document-based where each entity represents a document, or concept based where each entity represents an object or concept of interests in a specific domain. Mining a document-based IF deduces patterns and relationship across documents. Document clustering, visualization and categorization are examples of mining from a document-based IF. Mining a concept-based IF derives patterns and relationships across objects or concepts. (Tan, 1999)

**Text Mining Applications**

There are many applications for text mining. One is security applications. Text mining can be used as a surveillance system to see what people are talking about including social and economic actions. Today’s criminals and terrorists are using new technologies. Text mining is a knowledge management technology that allows intelligence analysts to automatically analyze the content of online data banks, suspected websites, blogs, emails, chats and other media. This type of information can be used to help detect links between individuals and organizations or other people. One of the largest text mining systems that exist today is the classified ECHELON surveillance system. Also, many text mining software packages such as AeroText, Attensity, SPSS and Expert System are beginning to be marketed towards security applications, particularly analysis of plain text sources such as Internet news. In 2007, the European Union’s criminal intelligence agency, Europol, began using text mining. The Serious Crime division developed an analysis system for tracking transnational organized crime. This was called the Overall Analysis System for Intelligence Support, or OASIS. It operates and integrates information among the
most advanced text analytics and text mining technologies available on today's market. This system led Europol to make the most significant progress to support law enforcement objectives at the international level. More nations and organizations are likely to follow this model in the future.

Another application for text mining is for biomedical use. There has recently been an increasing interest in text mining and information extraction strategies for application to the biomedical and molecular biology literature due to the increasing number of publications that are available electronically through databases such as PubMed. It can be used for the identification of biological entities, such as protein and gene names in free text. One example is GoPubMed which retrieves PubMed abstracts for the users search query, then detects ontology terms from the Gene Ontology and Medical Subject Headings in the abstracts and allows the user to browse the search results by exploring the ontologies and displaying only papers mentioning selected terms, their synonyms or related descendents of the terms. (Krallinger, 2005) Text mining is a tool that can be used to obtain manually annotated relationships for proteins, diseases, drugs, and biological processes as they get published. The medical field advances very quickly and involves a massive amount of information. Programs are being developed to help search through this information more efficiently and solve medical issues faster.

Text mining can also be used for online media applications. It is being used by large media companies to clarify information and provide readers with better search experiences. Editors are also benefitting from this because they are able to share, associate and package news across different media. Lastly, text mining is being used for academic applications. It can be useful to publishers who hold large databases of information requiring indexing for retrieval.
Libraries can benefit from the text mining process to improve the process of searching and comparing relevant literature.

Text mining is being used by many different businesses to improve their intelligence and efficiency. Text mining is believed to have a commercial potential higher than that of data mining. In fact, a recent study indicated that 80% of a company’s information is contained in text documents (Tan, 1999). Businesses may already have unstructured data sources such as documents on their intranet, e-mails that contain internal research, and groupware applications that house customer comments. These business operations may also need to tap external, non-relational data and this is where text mining technology can come into play. Businesses can benefit from text mining in many ways. They see an increase in the value of existing data assets. Staff is able to quickly draw upon a much wider range of information, which helps increase productivity. They can also potentially step ahead of competition by having more complete information to make better-informed decisions (Biggs, 2000).

**Text Mining Considerations**

Businesses should take some factors into consideration before choosing a text mining tool to implement into their company, mainly because the text mining market is still developing and evolving. They should consider choosing a tool that does not require a huge first-time categorization, tagging, or integration effort. The tool that companies select should be capable of automatically identifying and indexing unstructured data concepts. Companies should also expect some form of graphical interface that supports a high-quality, large view of the data as well as the ability to down to a very detailed level. Text mining tools are not supposed to be static in nature, given the changing data involved. The tool selected should let users explore data paths that form new relationships. Some of the text mining tools available now include Semio's
SemioMap, IBM's Intelligent Mining for Text, Autonomy's Agentware, and Megaputer's TextAnalyst. Text mining is not necessarily for every company. You would not use the technology to locate a single piece of information. Text mining is ideal for finding related information in huge amounts of unstructured data. It is also a good way for businesses to learn about a subject, inspect changes in the market, or to identify ideas to pursue (Biggs, 2000).

ADVANTAGES OF DATA MINING

The advantages to data mining are infinite, but you first and foremost must have access to large amounts of information. The driving force behind data mining is the ability to increase profit. Data mining can also help provide better customer service, reduce cost, and also help marketing effectiveness through analyzing data patterns. Data mining is virtually used by every service industry and the two most popular industries are retail and marketing. There are several other industries that use data mining procedures which include: Banking/Crediting, Health/Car Insurance, Transportation, Mobile communication, and Law Enforcement.

Marketing

Arguably, the marketing industry receives the most benefits from data mining. “Data mining helps marketing users to target marketing campaigns more accurately; and also to align campaigns more closely to the needs, wants, and attitudes of customers and prospects” (Benson, Smith, Thearling). Through data mining applications, marketers can find patterns which usually are good predictors of consumer spending. With these results, marketers can target either a wide range of customers or a specific segment by feeding the results into campaign management software. Some of the more popular targeted segments are age, gender, income and lifestyle. Let’s take, for example, a startup online computer retailer vs. Dell. Dell predominately has
control of the market share for online computer sales. Dell keeps an extensive record of all sales transactions, returns, and repairs. With this data, Dell can employ data mining to target customers based on previous purchases, age, income and gender. Imagine if the startup retailer was granted access to this database. With the proper resources, sales would likely increase tenfold or greater for the startup company. They could use data mining and apply it to a certain demographic, and cater a strategic marketing plan. Sir Francis Bacon once said, “Knowledge is power” and that is what the startup company would be obtaining if they were granted access to Dell’s database. Another benefit is helping to predict which products previous customers will purchase. Satisfying the needs of previous clientele will certainly help in retaining these customers.

Retail

Retail stores are also a huge benefactor of data mining. Ever think of why retailers have rewards cards? Personally, I thought it was for receiving “member discounts”. That is only part of the equation, as the other half is keeping track of all purchases through these member programs. “The usual rationale for loyalty cards is that it makes stores more efficient at stocking their shelves with products you want, and allows them to pass the savings on to you” (Pierce). Let’s take for example Ralphs which has a free reward’s program for customers. Members can receive in store discounts with their purchases as long as they use their card. A database is compiled with all of these member purchases. Ralph’s can use data mining to discover trends in items purchased, frequency of purchasing items, and time in between visits. This information can be valuable since Ralphs’ can arrange their inventory and shelves to cater towards these purchasing patterns. Data mining can find items that are frequently purchased together, which
gives the store insight into how the store should be organized to make shopping easier. One downfall to this program is, your personal information can be sold or traded to third parties.

**Banking and Crediting**

The banking and crediting industries have recently incorporated data mining technology to their advantage. The biggest advantage is by examining patterns of income and spending data figures and comparing it other customers with these same attributes. With these results a banker or creditor can analyze risk, thus giving information on whether or not to approve a line of credit or loan. Recent home loans have been given to customers with high levels of risk, who under normal standards wouldn’t qualify; as a result our economy has taken a turn for the worst. Do you think these banks based their decision solely off of data mining results? The answer is more than likely no, but they definitely should have! With the recent increase in fraudulent credit card activity, data mining can help decrease these fraudulent purchases. This can be accomplished by comparing recent spending activity with spending in previous months or by comparing this specific customer to other similar attributed customers and their spending patterns.

**Customer Satisfaction and Retention**

Identifying loyal previous customers is essential, and through data mining they are easily discovered. Being able to determine spending by customer groups is also important. Most banks and creditors will often reward these customers with additional incentives to help retain their services. Another unlikely benefit is targeting customers who simply use their bank to cash their checks. Let’s take for example customers who wait for their checks to clear, then quickly withdraw the funds to be reinvested elsewhere. This is obviously a loss of business to the bank. To counter this, data mining and campaign management software can help immediately identify large deposits and trigger some sort of response. The system might automatically schedule some
sort of promotion as soon as a customer’s balance exceeds a certain predefined amount. “Based on the size of the deposit, the triggered promotion can provide an appropriate incentive that encourages customers to invest their money in the bank’s other products” (Benson, Smith, Thearling). By keeping their own customers happy, through whatever various services, banks can make a substantial profit as long as they keep funds internally.

**Insurance Industry**

There are some questions that have arisen about the usage of data mining in the insurance and health care industries. Many argue that, since these industries are purely profit driven they will only use data mining to their advantage. Many of these companies will try to entice you into purchasing the most expensive plan, when in reality basic plans are more than adequate. Data mining can help these companies target customers that will enroll in their policies. More specifically, it would help them target specific customers for specific policies.

Detecting fraudulent claims is by far the biggest benefit of data mining to the insurance industry. Dennis Parker, marketing manager at SPSS, remembers coming across claims that he labeled “no duh” fraud. The indicators were so obvious, as even an inexperienced claim professional would be able to point this out. The only reason claims were being perpetrated, Parker explains because, “Insurance companies were paying them” (Conz). The insurance industry has grown since then, and many of them have special investigation units that specifically target fraud detection. Insurance companies employ both claims-scoring algorithms and Special Investigation Units (SIUs) to help fight fraudulent claims. According to Accenture’s 2007 global claims study, only 17% of insurers currently utilize advanced IT tools to detect fraudulent claims. Fraud detection is left mostly to the claims handler, at a surprising 37%.

“Farmers Insurance is moving rapidly towards adding text-mining capabilities and predictive
modeling to the front of its claims and fraud-fighting platforms, says Doug Ashbridge, director of special investigations at Farmers. Text analytics solutions extract factual information from unstructured text found in documents such as police reports, medical records, and adjuster notes to establish patterns and identity trends” (Conz).

John Viray, a senior claims analyst at Infinity, explains that his company filters its claims initially by data mining techniques. After the filtering is done, claims with a specific score are handed to claims analyst for further investigation. “We catch a lot of suspicious claims that are discovered by simply cross referencing information with an external database” states John Viray. Not everybody is granted access to these databases since highly sensitive information can be obtained. Infinity hasn’t fully integrated data or text mining into their claims process, but hopes to in the near future.

Cell Phone Industry

Another industry that has benefited from data mining is the cell phone industry. In the late 90’s cell phones were considered more of a luxury item than a necessity. Cell phone plans were very simplistic catering to low, medium or high usage. Since the boom of the cell phone industry, providers have collected large amounts of data. Data pertaining to monthly minutes, time of day, text messaging and multimedia messaging have been stored, maintained, and analyzed by cell phone providers. Through data mining techniques providers can change the plans they offer frequently, catering towards all types of users.

DISADVANTAGES OF DATA MINING

Everyone would agree that personal privacy is of extreme importance. As useful and revolutionary as the internet is considered to be, it is one of the tools that poses a major threat to
our personal privacy. When used in conjunction with data mining, these two can prove to be extremely powerful when put in the wrong hands. From the consumer’s perspective, data mining benefits everyone else except them. Consumers’ value of privacy far exceeds that of customized products and services. Customized products and services are a luxury that consumers can live without, and a luxury that is not worth the possible outcome of identity theft.

**Privacy Issues**

Data mining is a method that allows organizations to breach consumers’ privacy and access personal information. Considering the technological advances that we have achieved over the years, it should not be surprising how easy it is to access and acquire someone’s personal information. Personal information ranging from your name, address, and birthday, all the way to your personal intuitive preferences (Harper) is all available online. Whether it is done legally or illegally, companies make huge amounts of money from the use of consumers’ personal information.

The government also has data mining programs created to track terrorists. The use of predictive data mining in an attempt to find terrorists or terrorism planning among Americans can only be premised on using massive amounts of data about Americans’ lifestyles, purchases, communications, travels, and many other facets of their lives (Harper). Yet, it is found that data mining is not, and cannot be, a useful tool in the anti-terror arsenal. The incidence of terrorism and terrorism planning is too low for there to be statistically sound modeling of terrorist activity (Harper).

There are many benefits in using data mining as a tool. Yet, these benefits can just as simply be turned into problems by cyber criminals and predators. What exactly can they do with this information? The possibilities are almost endless.
Security Issues

Companies have plenty of personal information about us available online, yet some of these companies do not allow sufficient funding for protecting this valuable information. Cyber criminals are increasingly busy. Attempts to access sensitive government information increased 40 percent last year, USA Today reports. Someone makes a serious attempt to steal secret government data 15 times a day (Soat 56). We don't know how many times they succeed. The best company at turning the ocean of Internet data into useful facts has been Google, which began as a search engine and grew into a company that has its electronic eyes on just about everything, including your house as with the Google Earth application (Soat 56).

For example, the Ford Motor credit company had to inform 13,000 of the consumers that their personal information including Social Security number, address, account number and payment history were accessed by hackers who broke into a database belonging to the Experian credit reporting agency (Advantages...). This is a perfect example of how companies are not taking the necessary precautions and security measures to prevent incidents such as this from happening. When an incident such as this is at an even larger scale, consequences can be disastrous.

Misuse of Information/Inaccurate Information

The trends and information gathered can be misused in so many ways. For example, if an insurance company manages to get their hands on a person’s medical records and finds out the person is diagnosed with a terminal disease, such as cancer or AIDS/HIV, they can deny that person eligibility for insurance. As a matter of fact, they can discriminate against a whole population of diagnosed individuals and deny every person eligibility (Advantages). Taking this
example even further, what if, through the use of data mining and analyses of data correlations, a correlation was found with occurrences of cancer and a particular race? Can you imagine how much of an impact this would have? These insurance companies could possibly increase prices and premiums for a particular race based on this information that they have discovered. This can and will have terrible consequences. Another possibility is if a scam artist finds this same information and decides to send this person emails about a “cure” for the disease. Considering how vulnerable this person might be, they’ll probably take the bait and purchase this “cure”. And as we all probably know, there are no cures for many diseases out there today.

Demographics used in data-mining could lead to real problems depending on how it is applied.

Nothing in this world is perfect, and data mining is not an exception (Advantages…). No matter how many precautions one takes to prevent mistakes, there is always that possibility. Misfiling errors and other data type errors can and will occur. Data mining is not error-proof and will probably never be error-proof. Even if the probability of it occurring is absolutely minimal, the possibility is still present.

**ETHICAL CONSIDERATIONS**

Many consumers are not aware of the vast information that companies are able to track on them over time, and in addition many businesses are not aware of how important keeping the collected information confidential can be to the customer. “Users of data mining should start thinking about how their use of this technology will be impacted by legal issues related to privacy” [Thearling]. It comes as no shock that over the years there have been a growing number of disputes and lawsuits filed against businesses who have not respected customer’s privacy due to data mining. A perfect example of this can be found between CVS Pharmacy and Elensys.
CVS Pharmacy worked with Elensys, “a Massachusetts direct marketing company that sends reminders to customers who have not renewed their prescriptions” [Thearling]. CVS decided to work side by side with Elensys as an attempt to increase the quality of their customer service; however, it did not work out that way. Some customers appreciated the reminders to refill their prescriptions that were sent in the mail, but there were also customers that were outraged. The consumers that were angered disapproved because they felt that it was a breach of privacy. In order to send out the mail reminders, Elensys, must have had access to the customers personal information and their medical records as well. In the end, CVS Pharmacy terminated their contract with Elensys because of this issue.

Since privacy is such a touchy issue, there are ethics that both consumers and businesses need to be aware of when data mining is being used. Taking precautionary steps and being aware of the information that is collected can be a solution to keeping data mining beneficial and protected.

**Consumer Ethics in Data Mining**

As technology increases, consumers need to be more aware and careful about what types of information they are releasing to businesses and companies. Not all businesses are trustworthy when handling personal information. Consumers should be well aware of the dangers of giving out information. This could potentially mean reading all fine print before signing papers releasing personal facts.

There are a few steps that consumers can take to ensure the privacy of their personal information. One important thing to do is read through all contracts and documents thoroughly, including the fine print. Sometimes the fine print has disclosure notifications in them. In addition, consumers should also keep a look out for “opt-out” boxes. Consumers have the right
to know what personal information is being used and what it is being used for. Another aspect that will give consumers control will be the access to an online database that they can log into in order to view and change what information a company has on them.

Business Ethics in Data Mining

Over the years, more and more businesses have been turning to data mining in order to increase the quality of customer service as well as profits. However, there are some businesses that overlook the ethical implications of using programs of that nature. As consumers become more aware of data mining, they are also becoming more concerned about the privacy of their personal information. Businesses have to be ethical and trustworthy when handling data mining statistics, which means they should not be selling or releasing personal information of customers and stereotyping groups of people based on data mining results.

When businesses partner with marketing companies or other corporations, they need to have agreements when it comes to sharing personal consumer information from data mining. Businesses need to be aware that consumers are not the only people susceptible scams, so are businesses as a whole. Some marketing companies have worked with businesses to not only benefit the company they are working for, but themselves as well. In the past some businesses released customer’s personal information to marketing companies in order to have them help promote their business or customer service, however sometimes after the marketing company is done the job for the business, they sell or use the customers information for their own benefit. This can cause problems for not only the consumer, but both the businesses involved as well. If something like this happens and the customers’ information gets into the wrong hands because of one of these businesses, there is a good chance the consumer can sue the companies.
In order for businesses to avoid legal problems, there are a few steps that they can take when it comes to safer data mining. First, they should let customers know they will be using their information for internal use only, and they should also be notified what information they will be using. Following this proposal, the company should provide an opt-out box, so the consumer has a personal choice of releasing the information. This will cut down the number of legal problems that some companies face when using data mining.

After acquiring information from customers, businesses need to use the information for the intended purposes. An important issue that is raised with this is misguided stereotyping. When companies get the results from their data mining, they need to make sure they are not categorizing or making assumptions to certain groups or demographics. By doing this, it can seem like businesses are practicing discrimination. All information should be evaluated and looked at equally.

**The Direct Marketing Association**

The Direct Marketing Association [DMA] has more than 3,600 members in 50 countries. One of their goals is to help end any discrepancies that they may make while marketing from data mining that may be unsafe for businesses and consumers. Ultimately, following these guidelines will help benefit both parties. This association has created numerous guidelines to help protect personal data and specify correct usage.

The guidelines are as follows: (1) personal data should be collected by fair and lawful means for a direct marketing purpose. (2) Direct marketers should limit the collection of data to only that deemed necessary for direct marketing. (3) The data should be accurate and complete and should be kept no longer than necessary. (4) Individuals can request personal data about themselves, as well as challenge the accuracy of the personal data. (5)
Consumers who provide data that may be rented or sold should be told of that potential and given an opportunity to delete their data. (6) The collection, rental sale and use of consumer data should be constrained to direct marketing purposes. (7) Each direct marketer is responsible for the security of their data. (8) Visitors to personal data processing and storage sites should only be allowed to visit if they have express permission of the direct marketer and are accompanied by an employee at all times. (9) When transferring data between direct marketers, the receiver is responsible for the security of the data during the transfer. (10) An ethics committee of the DMA has jurisdiction to review individual complaints in violation of these guidelines. [Morris, Manning & Martin]

The guidelines listed above can help protect both businesses and consumers from lawsuits or the misuse of personal information. These simple precautions can increase privacy and boost ethics, making data mining a safe and profitable marketing tool.

CONCLUSION

In conclusion, data mining is a great business tool. Businesses have discovered countless data mining techniques and methods that have proved to increase profit and success with the use of valuable information. These data mining techniques have allowed us to detect fraud, terrorist activity, and so many other benefits in business including Human Resources, loss prevention, and many other sectors and branches in the business world. The information gathered has been proven to be beneficial in many ways, but has also been demonstrated to have the potential to be detrimental to our society. This tool that has allowed businesses to develop and succeed tremendously has done the same for cyber predators and other criminals. DMA regulations and guidelines are a step in the right direction toward protecting businesses and consumers from
unethical use of sensitive information. Following these guidelines will lead data mining in the ethical direction, which is what it was initially intended for.
References


http://xiangyun86.wordpress.com/2006/12/05/advantages-disadvantages-of-data-mining/


Conz, Nathan. "Data Mining Improves Fraud Mitigation Efforts.” 06 September 2007.


