**Handouts of Lecture 20 Professional Practices (IT)**

**Lecture Title: Data Mining**

**Data Mining Examples**

Data mining is put to a wide variety of uses in modern society. Here are a few examples.

**Credit reports**

Credit reports are a great example of how information about customers has itself become a commodity. A credit bureau is a company that keeps track of an individual’s assets, debts, and history of paying bills and repaying loans, using this information to determine the creditworthiness of that person. Credit bureaus sell credit reports to banks, credit card companies, and other potential lenders.

Thanks to the national credit bureau system, you can get a credit card from a bank or store that you’ve never done business with. When you want to borrow money to purchase a home, you don’t have to go to your local bank. You can get the money from a bank across the country that trusts you to repay the loan because of your high credit score. Competition among banks leads to lower interest rates, a definite advantage for consumers.

Of course, if you have a poor record of paying bills on time and repaying loans, your credit score will be low. People with low credit scores have a harder time getting loans and pay higher interest rates on the loans they do get.

A poor credit report can come back to haunt people in ways they might not have anticipated. Many employers conduct a credit check late in the job interview process as a way of double-checking a candidate before making a job offer. A credit report received by an employer doesn’t give the applicant’s credit score, but it does list the applicant’s debts.

**Targeted direct mail**

Years ago direct mail marketing meant mass mailings of advertisements. Today the trend is toward targeted direct mail, in which businesses mail their advertisements only to the leads most likely to be interested in purchasing their products. Customized mailing lists are available from a variety of sources, including Experian (one of the three largest credit report companies in the United States). From its database of more than 200 million consumers, Experian can provide mailing lists of people in a particular geographical area who meet certain criteria, including new parents, people who have just moved, new homeowners, renters with estimated incomes greater than $100,000, health enthusiasts, sports enthusiasts, and “green” consumers. A good example of direct mail marketing is Target’s efforts to reach pregnant women. Retailers know that the habits of shoppers—where they buy certain goods and the brands they select—are difficult to change. However, when people graduate from college or move to a new town or get married, their shopping habits are more malleable. New parents are particularly open to changes in their shopping habits. For that reason Target asked its statisticians to find ways to predict which of Target’s women customers were in their second trimester of pregnancy. The company’s goal was to use direct mail offers to get these women into the habit of buying a wide variety of items at Target stores.

**Microtargeting**

Since 2004 dire

ct marketing based on data mining has become part of US presidential campaigns. In a technique called microtargeting, a campaign combines data about voter registration, voting frequency, and contributions with consumer data and information available from a geographic information system to gain insights into which candidate the voter is likely to favor. The campaign then uses direct mailings, email, text messages, or home visits to encourage likely supporters to vote.

**Connecting the dots**

Data mining can be surprisingly powerful. Suppose a government agency managing tollbooths were to sell information records of the following form:

The agency does not reveal the names of the owners of the cars, so it believes it is protecting their anonymity. However, many people have an account set up so that their tollbooth payments are automatically charged to their credit cards. If a credit card company buys these records from the tollbooth agency, it can match the date, time, and amount of the tollbooth payments with the date, time, and charge on its credit cards to determine the identity of the person driving a vehicle with a particular transponder number. Once this has been done, the credit card company can figure out which customers are driving the most miles and are likely to purchase new cars more frequently. It can then sell this information to banks interested in soliciting automobile loan applications.

**Price customization**

Retailers are using information gathered about their customers to offer different prices to different people. You read about this practice earlier in the chapter in the description of grocery store loyalty cards. Another example: data aggregation firms sell profiles of shoppers to some online merchants, who use this information to determine who should be offered a discount and who should be charged full price and offered more expensive options.

**Social Network Analysis**

A promising new area in data mining is the incorporation of information collected from social networks. Here are a few examples of how organizations have used social network analysis to meet their objectives:

In ultracompetitive cell phone markets, it’s crucial that companies keep their subscribers from defecting to rival firms. Bharti Airtel, India’s largest cell phone company, uses software to analyze phone records and identify “influencers,” those subscribers most likely to be able to persuade their friends and family members to follow them when they switch carriers. It then offers the influencers special promotions to keep them loyal. How can Bharti Airtel identify influencers from their phone records? They are the ones whose calls are quickly returned, who call other people late at night more frequently, and who get more calls on Friday afternoons when parties are often organized.

Police in Richmond, Virginia, monitor Facebook and Twitter messages to determine where parties are happening. Data-mining software identifies the party locations mentioned most frequently. By deploying officers more strategically on big party nights, the department saves about $15,000 on overtime pay, and the community has seen a big drop in criminal activity.

Banks are combining data collected from social networks with credit card statements and other information to evaluate the riskiness of loans. For example, someone applying for a loan to start a new business may be a bad risk if the proposed business has no connection with their social network, educational background, travel history, or previous business dealings.

**Examples of Consumer Backlash**

Advances in information technology have led to a drop in the cost of acquiring information. Meanwhile, the value of information continues to rise, as organizations refine their data-mining techniques. The result of these trends is that corporations have an incentive to acquire more information, making it more difficult for individuals to protect their privacy. Still, people can and do fight back when they feel a corporation has gone too far.

**Marketplace: Households**

Lotus Development Corporation teamed up with credit reporting company Equifix to develop a database on 120 million people and a program that would allow the user to produce direct mailing lists based on various criteria, such as household income. Lotus hoped to sell the package, which it called “Marketplace: Households,” to small businesses. Soon after the product was announced in the spring of 1990, there was a considerable backlash. Consumers complained loudly and vigorously, with more than 30,000 letters, phone calls, and emails. Lotus dropped plans to sell the CD. How times have changed! A little more than two decades later, credit reporting company Experian has a thriving business selling direct mailing lists based on much more detailed information about consumers.

**Facebook Beacon**

In November 2007, Facebook announced Beacon, “a core element of the Facebook Ads system for connecting businesses with users and targeting advertising to the audiences they want” [58]. Beacon promised to be an important way for Facebook to earn advertising revenue. Fandango, eBay, and 42 other online businesses paid Facebook to do “word-of-mouth” advertising of their products and services through Beacon. For example, after a Facebook user bought movie tickets on Fandango, Fandango would send this information to Facebook so that Facebook could broadcast it to that user’s friends.

Beacon was based on an opt-out policy, meaning that it was in effect unless a user explicitly asked to be excluded. That decision was good for Facebook because advertising revenue depends on the size of the audience. However, the decision to make the system opt-out upset many Facebook users, who were unaware of Beacon until it revealed information they thought was private. For example, after Sean Lane purchased what was supposed to be a surprise Christmas gift, the following news headline was broadcast to his wife and more than 700 other people in his Facebook network: “Sean Lane bought 14K White Gold 1/5 ct Diamond Eternity Flower Ring from overstock.com”.

Beacon soon attracted strong criticism from a variety of sources. A spokesman for MoveOn.org said, “Sites like Facebook are revolutionizing how we communicate with one another and organize around issues together in a 21st century democracy. The question is: Will corporate advertisers get to write the rules of the Internet or will these new social networks protect our basic rights, like privacy?” MoveOn.org created an online group calling for Beacon to require an explicit opt-in from users, and it attracted the support of more than 50,000 Facebook users. A few weeks later, Facebook decided to switch to an opt-in policy for Beacon. “I’m not proud of the way we’ve handled this situation, and I know we can do better,” said Mark Zuckerberg, CEO of Facebook.

**Netflix Prize**

Netflix is a popular subscription service for movies and television shows. An important feature of Netflix is its movie recommendation service. After a subscriber has rated several movies, Netflix uses a collaborative filtering algorithm to predict how well the subscriber will like other movies in its collection. It then recommends to the subscriber movies the subscriber is likely to enjoy. In 2006 Netflix offered a $1 million prize to any group that could come up with a collaborative filtering algorithm that was at least 10 percent better than Netflix’s own algorithm at predicting user ratings for movies. Netflix released more than 100 million movie ratings from nearly half a million customers, stripped of private information in an attempt to make the records anonymous. Each movie rating consisted of four pieces of information—subscriber, movie, date of grade, and grade—where each subscriber was represented by an integer.

However, a group of researchers at the University of Texas-Austin demonstrated how it was possible, with just a little information about movies seen by an individual, to identify the movie rating records associated with that person, revealing potentially sensitive information, such as their political leanings or sexual orientation. The revelation that the release of “anonymous” movie ratings information could compromise the privacy of Netflix subscribers led to a complaint by the US Federal Trade Commission and a lawsuit. On March 12, 2010, Netflix announced that it was canceling a Netflix Prize sequel.

**Malls Track Shoppers’ Cell Phones**

On Black Friday, 2011 two malls in the United States—the Promenade Temecula in California and the Short Pump Town Center in Virginia—began recording the movement of shoppers through the stores by tracking the locations of their cell phones. The malls hoped to be able to answer such questions as these:

How much time do people spend in store X?

How many people who shop at store Y also shop at store Z?

Are there unpopular areas of the mall that do not attract enough shoppers?

Small signs posted throughout the malls informed shoppers of the study. It let them know that the data were being collected anonymously; the mall assigned an ID code to each phone and tracked the movements of the phone without knowing anything about the person carrying the phone. To prevent data about their movements from being collected, shoppers had to turn off their cell phones. The malls had planned to continue their study through the Christmas buying season, but an intervention by Senator Charles Schumer of New York prompted them to stop collecting data after only three days. The senator issued a statement that said, in part, “Personal cell phones are just that—personal. If retailers want to tap into your phone to see what your shopping patterns are, they can ask you for your permission to do so”.

**iPhone Apps Uploading Address Books**

In February 2012, programmer Arun Thampi in Singapore discovered that the app for the social networking site Path was uploading his iPhone’s address book without first getting his permission. Thampi wrote on his blog, “I’m not insinuating that Path is doing something nefarious with my address book but I feel quite violated that my address book is being held remotely on a third-party service”. David Morin, the CEO of Path, initially replied to Thampi’s post by saying that the purpose of collecting the information was to make it easier for people to connect with family and friends on Path and to let them know when people in their address book join Path. Morin’s statement also labeled Path’s actions as an “industry best practice.” However, it didn’t take long for the Internet community to weigh in and point out that Apple’s guidelines require apps to ask for permission before uploading information from address books. Facing a storm of negative publicity, Morin issued another statement in which he apologized for what Path had done, promised that the company would destroy the data it had collected, and announced that the app would be changed so that it would no longer upload address books without permission.

As the Path controversy erupted, the media pointed out that other popular iPhone apps, including Twitter, Foursquare, and Instagram, were also gathering information from address books without asking for permission. All the aforementioned companies responded by announcing that they, too, would release new versions of their apps that explicitly asked users for permission before uploading contact information from address books.

**Instagram’s Proposed Change to Terms of Service**

In December 2012, the popular photo-sharing service Instagram announced an upcoming change in its privacy policy and terms-of-service agreement. The terms-of-service agreement appeared to change how Instagram and its parent company, Facebook, could use photographs uploaded by Instagram users. The proposed agreement included the following statement: You agree that a business or other entity may pay us to display your username, likeness, photos (along with any associated metadata), and/or actions you take, in connection with paid or sponsored content or promotions, without any compensation to you.

After some legal experts said that the new terms-of-service agreement would allow Instagram or Facebook to use photos on the Instagram site in advertisements without compensating or even getting the permission of the person who uploaded them, the reaction was swift. The hashtag #Boycottinstagram began trending on Twitter, and many Instagram users downloaded alternative photo-sharing apps. The number of people using the Pheed and Flickr increased significantly, though the total number of users of these services remained far below the 100-plus million using Instagram [68]. Responding to the uproar, Instagram cofounder Kevin Systrom issued a statement saying that the new privacy policy and terms-of-service had been misunderstood. He also announced that the advertising section in the terms of service agreement was being changed back to its original version.

***Reference***

***Lecture 20 slides: Information Privacy (cont.)***

***Gao, Y. (2012). Ethics for the Information Age by Michael J. Quinn. World Libraries, 20(1).***