

other personal data (e.g., anniversary, pregnancy, illness, loss in the family, bankruptcy filings, etc.). Most of these data can be accessed through some third-party data providers. The main question here is the privacy of the person to whom the data belongs. In order to maintain the privacy and protection of individuals' rights, data mining professionals have ethical (and often legal) obligations. One way to accomplish this is the process of de-identification of the customer records prior to applying data mining applications, so that the records cannot be traced to an individual. Many publicly available data sources (e.g., CDC data, SEER data, UNOS data, etc.) are already de-identified. Prior to accessing these data sources, users are often asked to consent that, under no circumstances, will they try to identify the individuals behind those figures.

There have been a number of instances in the recent past where the companies shared their customer data with others without seeking the explicit consent of their customers. For instance, as most of you might recall, in 2003, JetBlue Airlines provided more than a million passenger records of their customers to Torch Concepts, a U.S. government contractor. Torch then subsequently augmented the passenger data with additional information such as family sizes and social security numbers—information purchased from a data broker called Acxiom. The consolidated personal database was intended to be used for a data mining project in order to develop potential terrorist profiles. All of this was done without notification or consent of passengers. When news of the activities got out, however, dozens of privacy lawsuits were filed against JetBlue, Torch, and Acxiom, and several U.S. senators called for an investigation into the incident (Wald, 2004). Similar, but not as dramatic, privacy-related news came out in the recent past about the popular social network companies, which allegedly were selling customer-specific data to other companies for personalized target marketing.

There was another peculiar story about privacy concerns that made it to the headlines in 2012. In this instance, the company did not even use any private and/or personal data. Legally speaking, there was no violation of any laws. It was about Target and is summarized in Application Case 4.7.

Application Case 4.7

Predicting Customer Buying Patterns—The Target Story

In early 2012, an infamous story appeared concerning Target's practice of predictive analytics. The story was about a teenage girl who was being sent advertising flyers and coupons by Target for the kinds of things that a new mother-to-be would buy from a store like Target. The story goes like this: An angry man went into a Target outside of Minneapolis, demanding to talk to a manager: "My daughter got this in the mail!" he said. "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. He looked at the mailer. Sure enough, it was addressed to the man's daughter and contained advertisements for maternity clothing, nursery furniture, and pictures of

smiling infants. The manager apologized and then called a few days later to apologize again. On the phone, though, the father was somewhat abashed. "I had a talk with my daughter," he said. "It turns out there's been some activities in my house I haven't been completely aware of. She's due in August. I owe you an apology."

As it turns out, Target figured out a teen girl was pregnant before her father did! Here is how they did it. Target assigns every customer a Guest ID number (tied to their credit card, name, or e-mail address) that becomes a placeholder that keeps a history of everything they have bought. Target augments this data with any demographic information that they had collected from them or bought from other information sources. Using this information,

(Continued)

Application Case 4.7 (Continued)

Target looked at historical buying data for all the females who had signed up for Target baby registries in the past. They analyzed the data from all directions, and soon enough some useful patterns emerged. For example, lotions and special vitamins were among the products with interesting purchase patterns. Lots of people buy lotion, but what they noticed was that women on the baby registry were buying larger quantities of unscented lotion around the beginning of their second trimester. Another analyst noted that sometime in the first 20 weeks, pregnant women loaded up on supplements like calcium, magnesium, and zinc. Many shoppers purchase soap and cotton balls, but when someone suddenly starts buying lots of scent-free soap and extra-large bags of cotton balls, in addition to hand sanitizers and washcloths, it signals that they could be getting close to their delivery date. In the end, they were able to identify about 25 products that, when analyzed together, allowed them to assign each shopper a "pregnancy prediction" score. More important, they could also estimate a woman's due date to within a small window, so Target could send coupons timed to very specific stages of her pregnancy.

If you look at this practice from a legal perspective, you would conclude that Target did not use any information that violates customer privacy; rather, they used transactional data that almost every other retail chain is collecting and storing (and perhaps analyzing) about their customers. What was disturbing in this scenario was perhaps the targeted concept: pregnancy. There are certain events or concepts that should be off limits or treated extremely cautiously, such as terminal disease, divorce, and bankruptcy.

QUESTIONS FOR DISCUSSION

1. What do you think about data mining and its implication for privacy? What is the threshold between discovery of knowledge and infringement of privacy?
2. Did Target go too far? Did it do anything illegal? What do you think Target should have done? What do you think Target should do next (quit these types of practices)?

Sources: K. Hill, "How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did," *Forbes*, February 16, 2012; and R. Nolan, "Behind the Cover Story: How Much Does Target Know?" *NYTimes.com*, February 21, 2012.

Data Mining Myths and Blunders

Data mining is a powerful analytical tool that enables business executives to advance from describing the nature of the past to predicting the future. It helps marketers find patterns that unlock the mysteries of customer behavior. The results of data mining can be used to increase revenue, reduce expenses, identify fraud, and locate business opportunities, offering a whole new realm of competitive advantage. As an evolving and maturing field, data mining is often associated with a number of myths, including the following (Zaima, 2003):

Myth	Reality
Data mining provides instant, crystal-ball-like predictions.	Data mining is a multistep process that requires deliberate, proactive design and use.
Data mining is not yet viable for business applications.	The current state of the art is ready to go for almost any business.
Data mining requires a separate, dedicated database.	Because of advances in database technology, a dedicated database is not required, even though it may be desirable.
Only those with advanced degrees can do data mining.	Newer Web-based tools enable managers of all educational levels to do data mining.
Data mining is only for large firms that have lots of customer data.	If the data accurately reflect the business or its customers, any company can use data mining.