

Using this equation, analysts can predict the number of minutes of weekend cell phone use by summing 12, plus 17.5 times the customer's age, plus 23.7 times the number of months of the account.

As you will learn in your statistics classes, considerable skill is required to interpret the quality of such a model. The regression tool will create an equation, such as the one shown. Whether that equation is a good predictor of future cell phone usage depends on statistical factors, such as *t* values, confidence intervals, and related statistical techniques. **Neural networks** are another popular supervised data mining application used to predict values and make classifications such as "good prospect" or "poor prospect" customers. The term *neural networks* is deceiving because it connotes a biological process similar to that in animal brains. In fact, although the original *idea* of neural nets may have come from the anatomy and physiology of neurons, a neural network is nothing more than a complicated set of possibly nonlinear equations. Explaining the techniques used for neural networks is beyond the scope of this text. If you want to learn more, search <http://kdnuggets.com> for the term *neural network*.

In the next sections, we will describe and illustrate two typical data mining tools—market-basket analysis and decision trees—and show applications of those techniques. From this discussion, you can gain a sense of the nature of data mining. These examples should give you, a future manager, a sense of the possibilities of data mining techniques. You will need additional coursework in statistics, data management, marketing, and finance, however, before you will be able to perform such analyses yourself.

Market-Basket Analysis

Suppose you run a dive shop, and one day you realize that one of your salespeople is much better at up-selling to your customers. Any of your sales associates can fill a customer's order, but this one salesperson is especially good at selling customers items *in addition* to those for which they ask. One day, you ask him how he does it. "It's simple," he says, "I just ask myself what is the next product they would want to buy. If someone buys a dive computer, I don't try to sell her fins. If she's buying a dive computer, she's already a diver and she already has fins. But, these dive computer displays are hard to read. A better mask makes it easier to read the display and get the full benefit from the dive computer."

A **market-basket analysis** is an unsupervised data mining technique for determining sales patterns. A market-basket analysis shows the products that customers tend to buy together. In marketing transactions, the fact that customers who buy product X also buy product Y creates a **cross-selling** opportunity; that is, "If they're buying X, sell them Y" or "If they're buying Y, sell them X."

Figure 9-20 shows hypothetical sales data from 400 sales transactions at a dive shop. The first row of numbers under each column is the total number of times an item was sold. For example, the 270 in the first row of Mask means that 270 of the 400 transactions included masks. The 90 under Dive Computer means that 90 of the 400 transactions included dive computers.

We can use the numbers in the first row to estimate the probability that a customer will purchase an item. Because 270 of the 400 transactions were masks, we can estimate the probability that a customer will buy a mask to be 270/400, or .675. In market-basket terminology, **support** is the probability that two items will be purchased together. To estimate that probability, we examine sales transactions and count the number of times that two items occurred in the same transaction. For the data in Figure 9-20, fins and masks appeared together 250 times, and thus the support for fins and a mask is 250/400, or .625. Similarly, the support for fins and weights is 20/400, or .05. These data are interesting by themselves, but we can refine the analysis by taking another step and considering additional probabilities. For example, what proportion of the customers who bought a mask also bought fins? Masks were purchased 270 times, and of those individuals who bought masks, 250 also bought fins. Thus, given that a customer bought a mask, we can estimate the probability that he or she will buy fins to be 250/270, or .926. In market-basket terminology such a conditional probability estimate is called the **confidence**.