In response to increasingly high customer demands and formidable competition, today’s organizations have implemented new technologies that generate and collect massive volumes of data. However, a large majority of the data collected goes to waste or becomes obsolete. Inconsistency, repetition, time-sensitivity, disparate data sources and data overkill are just a few of the problems overwhelming many decision makers.

Fortunately, the need for digestible and useful data has created a new generation of business intelligence technologies. This includes software that sifts through mountain ranges of data until the most useful nuggets of information are extracted and identified, much the same way that California prospectors of the 19th century refined their techniques to strike gold more quickly.

This new technology, commonly known as Data Mining, has taken on a life of its own and continues to evolve.

Data Mining is the science of revealing useful patterns and relationships in vast data stores. Data Mining — considered by many as the new name for statistics — draws on many disciplines. With statistics as its foundation, it incorporates computer science, machine learning, artificial intelligence and other specialties. Data Mining ultimately provides a
framework for dealing with uncertainty. As organizations and the global economy become more complex, sources of uncertainty become more plentiful. To make decisions confidently, new and sophisticated approaches are required.

The vast quantities of data that businesses produce, gather and store fuel the demand for more effective means of deriving value from them. Technologies that distill value from data have been available for some time. We have long been able to ask simple questions of the data: What were sales last quarter? How did the western region do compared to the eastern region? This book is not about such query and reporting or OLAP. These tools, while necessary, do not generate the big returns, nor do they provide answers to the really challenging questions. Business managers have often shied away from high-end data analysis, perhaps fearing it to be overly complex. With ever-increasing quantities of data, and lots of questions needing better answers, a growing number of decision makers want to know what Data Mining can do for them.

Today’s powerful multivariate modeling techniques provide answers to the hard questions: Which patterns and relationships are truly significant and which are merely chance? What are the really important factors affecting quality and sales? Which customers are likely to leave? Why and when? Given the probabilities of default, what is the exposure for a given portfolio of loans? Analytical rigor is required to effectively address the wide range of growing business needs, such as assessing risk more accurately, improving price optimization, and anticipating customer demand and preferences. Such applications of Data Mining are increasingly possible—and valuable—in business, as the examples in this book clearly demonstrate.

Data Mining delivers on the promise of helping executives make, save and spend money more effectively. Enlightened decision makers now enjoy new peace of mind. They can chart their courses more strategically and with greater precision over longer periods, rather than simply reacting,
in fire-drill fashion, to either tactical or strategic issues that flare up. Savvy executives harness the power of Data Mining in new areas of business for superior performance and sustained competitive advantage. No longer is high-end data analysis considered a rare, specialized activity with limited use in business. Instead, corporate survival depends on how well Data Mining is applied to vast and growing data stores.

Data Mining has led to highly refined processes, especially where the stakes are high. Minimizing the downside often piques management’s interest first — reducing write-offs, losses and fraud — since these improvements are typically easy to measure. Significant returns can also come from the upside where the potential may be less obvious — increasing customer response, sales and market share. Success breeds success, and the interest in high-end data analysis continues to grow. Only high-end data analysis keeps delivering big returns across increasing areas of business; ROI in excess of 1,000% is not uncommon. Data Mining has become integral to many aspects of business with returns that are significant, yet often incalculable.

In today’s data-driven economy, Data Mining is an essential tool in the pursuit of enhanced productivity, reduced uncertainty, delighted customers, mitigated risk, maximized returns, refined processes and optimally allocated resources. To thrive, performance must be measured along the way so that the factors that contributed to success — or failure — can be understood. By applying a more scientific approach to business decision making, the stage is set for continuous learning and improvement.

This book highlights the value of making sound decisions in the face of uncertainty. Real-world examples span industry sectors and illustrate how Data Mining drives optimal decision making for a variety of business processes. The business leaders and experts contributing to this book provide unique insights on a variety of business processes, challenges, solutions and best practices. As Data Mining becomes more ubiquitous in
business, let this book serve as an invitation to reap the benefits for those who have not yet begun. And for those who have already embarked, may this book allow greater success.

Dr. James H. Goodnight is CEO, chairman, co-founder and president of SAS Institute, the world’s largest privately held software company. Chief executive since the company’s incorporation in 1976, Goodnight continues to focus on strategic planning for the global business, which provides software and services that enable customers to transform data from all areas of their business into intelligence. An accomplished programmer, Goodnight has authored many of the procedures that comprise SAS® software.

SAS passed the $1 billion revenue mark in 1999 and revenues continue to grow. According to Goodnight, the key to the company’s success has been its ability to listen to more than 3.5 million software users and respond to their needs. SAS customers represent numerous industries and can be found in more than 100 countries. More than 200 SAS offices around the globe support this large customer base. SAS responds to customer needs by staying near the top of the software industry in the percentage of revenue reinvested in research and development, devoting over a quarter of total revenue to R&D.

In addition to this significant investment in technology, Goodnight also invests in people — SAS employees and their families. The company’s work environment is designed to nurture and encourage creativity, innovation and quality. Since the early 1980s, Goodnight has supported on-site child care, health care and recreation and fitness centers. His commitment to these progressive work-life
programs has earned SAS national recognition in publications such as The Wall Street Journal as well as Fortune, Fast Company, BusinessWeek and Working Mother magazines.

Goodnight’s passion, and the focus of SAS’ philanthropic efforts, is education. In August 1997, the doors opened at Cary Academy, an independent college preparatory day school for students in grades 6 through 12. Goodnight co-founded the school in 1996 as a model school — one that integrates technology into all facets of education. Shortly after Cary Academy opened, SAS launched SAS inSchool, which develops content-based educational software that is helping move schools into the next millennium. The software contains the framework for a new generation of teaching courseware that will further the use of technology as a learning tool.

A native of Wilmington, N.C., Goodnight holds bachelor’s and master’s degrees as well as a doctorate in statistics from North Carolina State University. He served on the faculty of NCSU from 1972 to 1976, and continues to serve as an adjunct professor. Goodnight is a Fellow of the American Statistical Association, and has authored numerous papers on statistical computing.