Preface

In today's data-driven environment, the need for the organization and analysis of a constant flow of information presents a challenge for professionals and researchers. The development, advancement, and implementation of data mining and warehousing techniques have profoundly impacted the ability of researchers and decision makers to identify much more relevant data related to a particular subject, issue, or phenomenon, and, as a result, heighten their understanding and comprehension. With applications in medical research, banking, security, and beyond, data mining and warehousing is both an invisible and powerful force in modern society.

As research projects on data mining and warehousing have grown in both number and popularity, researchers and educators have devised a variety of techniques and methodologies to develop, deliver, and, at the same time, evaluate the effectiveness of their use. The explosion of methodologies in the field has created an abundance of new, state-of-the-art literature related to all aspects of this expanding discipline. This body of work allows researchers to learn about the fundamental theories, latest discoveries, and forthcoming trends in the field of data mining and warehousing.

Constant technological and theoretical innovation challenges researchers in data mining and warehousing to stay abreast of and continue to develop and deliver methodologies and techniques utilizing the latest advancements. In order to provide the most comprehensive, in-depth, and current coverage of all related topics and their applications, as well as to offer a single reference source on all conceptual, methodological, technical, and managerial issues, Information Science Reference is pleased to offer a six-volume reference collection on this rapidly growing discipline. This collection aims to empower researchers, students, and practitioners by facilitating their comprehensive understanding of the most critical areas within this field of study.

This collection, entitled *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* is organized into eight distinct sections, which are as follows: (1) Fundamental Concepts and Theories; (2) Development and Design Methodologies; (3) Tools and Technologies; (4) Utilization and Application; (5) Organizational and Social Implications; (6) Managerial Impact; (7) Critical Issues; and (8) Emerging Trends. The following paragraphs provide a summary of what is covered in each section of this multi-volume reference collection.

Section I, **Fundamental Concepts and Theories**, serves as a foundation for this exhaustive reference tool by addressing crucial theories essential to understanding data mining and warehousing. Opening this elemental section is "Administering and Managing a Data Warehouse" by James E. Yao, Chang Liu, Qiyang Chen, and June Li, which provides a framework for data warehouse design and management while also explaining the pivotal role of data warehouses in shaping business decision making. Similarly, "Conceptual Modeling Solutions for the Data Warehouse" by Stefano Rizzi presents a practical methodology for designing a successful data warehouse. Additional contributions, such as "Knowledge Structure and Data Mining Techniques" by Rick L. Wilson, Peter A. Rosen, and Mohammad Saad Al-Ahmadi,

offer a broad overview of procedures and methods essential to data mining. The expertly researched contributions within this section also present an essential introduction to the innumerable applications of data mining and warehousing. "Introduction to Data Mining in Bioinformatics" by Hui-Huang Hsu illustrates how researchers can discover new biological knowledge using data mining techniques, and "Introduction to Data Mining and its Applications to Manufacturing" by Jose D. Montero provides insight into the promising integration of business applications in the field of manufacturing. This foundational section enables readers to learn from expert research on the elemental theories underscoring data mining and warehousing.

Section II, **Development and Design Methodologies**, contains in-depth coverage of conceptual architectures and frameworks, providing the reader with a comprehensive understanding of emerging theoretical and conceptual developments within the field of data mining and warehousing. "A Framework for Organizational Data Analysis and Organizational Data Mining" by Bernd Knobloch offers a methodology for more efficient and beneficial data mining within business organizations. Security, specifically as it relates to the design of data warehouses, is a major concern of researchers and professionals. Within their contribution "Designing Secure Data Warehouses," Rodolfo Villarroel, Eduardo Fernández-Medina, Juan Trujillo, and Mario Piattini compare six different designs for secure data warehouses. From basic designs to abstract development, chapters such as "Metric Methods in Data Mining" by Dan A. Simovici and "An Ontology-Based Data Mediation Framework for Semantic Environments" by Adrian Mocan and Emilia Cimpian serve to expand the reaches of development and design methodologies within the field of data mining and warehousing.

Section III, **Tools and Technologies**, presents extensive coverage of the interaction between various technologies and the field of data mining and warehousing. This symbiotic relationship encourages the advancement and invention of new technologies as well as continual innovation witnessed within the development of new data mining and data warehousing techniques. The rapid expansion of data mining to encompass the extraction of images, audio, and video is evaluated in "Video Data Mining" by JungHwan Oh, JeongKyu Lee, and Sae Hwang. "VRMiner: A Tool for Multimedia Database Mining With Virtual Reality" by Hanene Azzag, Fabien Picarougne, Christiane Guinot, and Gilles Venturini provides a more in-depth analysis of multimedia data mining, explaining how a 3-D method for visualizing multimedia data can be applied to real-world situations. Data mining's role in making technology more efficient is explored in selections such as "Mining E-Mail Data" by Steffen Bickel and Tobias Scheffer and "Mobile User Data Mining and Its Applications" by John Goh and David Taniar. Through these rigorously researched contributions, the reader is provided with countless examples of the up-and-coming tools and technologies that emerge from or can be applied to the multi-dimensional field of data mining and warehousing.

Section IV, **Utilization and Application**, explores the ways in which data mining and warehousing has been adopted and implemented in all facets of society. This collection of innovative research begins with "Strategic Utilization of Data Mining" by Chandra S. Amaravadi, which documents the emergence of data mining as an essential tool in decision support. Specific applications of data warehousing and mining, which are described in selections such as "Biological Data Mining" by George Tzanis, Christos Berberidis, and Ioannis Vlahavas and "The Utilization of Business Intelligence and Data Warehousing and Analytics in Banking: Concepts" by L. Venkat Narayanan relate the importance of data mining and warehousing in specific fields. The essential use and continual evolution of Web data mining is discussed in a set of informative selections, which include "Data Mining in Web Services Discovery and Monitoring" by Richi Nayak and "Traversal Pattern Mining in Web Usage Data" by Yongqiao Xiao and Jenq-Foung Yao. From established applications to forthcoming innovations, contributions in this

section provide excellent coverage of today's global community and demonstrate how data mining and warehousing impacts the social, economic, and political fabric of our present-day global village.

Section V, **Organizational and Social Implications**, includes a wide range of research pertaining to the organizational and cultural implications of data mining and warehousing. Introducing this section is "Data Mining in Practice" by Sherry Y. Chen and Xiaohui Liu, a selection that describes how different data mining approaches, such as classification and visualization, make data analysis within organizations more intelligent and automatic. Governmental data mining is explored in "Data Mining in the Federal Government" by Les Pang, which analyzes different federal data mining projects and the knowledge they generated for future organizational implementation. Additional chapters included in this section, such as "Data Mining and Knowledge Discovery in Healthcare Organizations: A Decision-Tree Approach" by Murat Caner Testik, George C. Runger, Bradford Kirkman-Liff, and Edward A. Smith, discuss the use of data mining techniques to both improve and predict an individual's healthcare system use. Organizational data mining and warehousing, however, is not always employed without difficulty, as is illustrated in the concluding chapter of this section, "Impediments to Exploratory Data Mining Success" by Jeff Zeanah. Within this chapter, the author assesses barriers to data mining implementation and provides organizations with guidelines for anticipating and preventing problems. Overall, the discussions presented in this section offer insight into the integration of data mining and warehousing techniques in society and how these techniques can be better structured and implemented in modern-day organizations.

Section VI, **Managerial Impact**, presents contemporary coverage of the more formal implications of data mining and warehousing, which are, more specifically, related to the corporate and managerial utilization of data mining and warehousing within organizations. Core ideas such as integration, evaluation, and potential strategies for increasing the effectiveness of modern organizations are discussed in this collection. "Data Mining and Business Intelligence: Tools, Technologies, and Applications" by Jeffrey Hsu emphasizes the importance of data mining for transforming raw data into useful knowledge that promotes business intelligence. Equally essential to this examination of managerial impact is evaluating the effectiveness of already-implemented data mining and warehousing programs, which is examined at length in chapters such as "The Business Data Warehouse: The Case of Wal-Mart" by Indranil Bose, Lam Albert Kar Chun, Leung Vivien Wai Yue, Li Hoi Wan Ines, and Wong Oi Ling Helen and "Data Warehousing: The 3M Experience" by Hugh J. Watson and Barbara H. Wixom. Considered together, these selections provide a framework for data warehouse design and construction within corporations and also analyze both the successes and pitfalls that such large-scale operations have encountered. This section concludes with a brief overview of data mining for both customer prediction and marketing within a business. Selections such as "Marketing Data Mining" by Victor S.Y. Lo explain how data mining can be used to both explain and predict customer behavior and, therefore, establish long-term customer relationships and increase profit within an organization.

Section VII, **Critical Issues**, presents readers with an in-depth analysis of the more theoretical and conceptual issues within this growing field of study by addressing topics such as the ethics of data mining, mining with incomplete data, and the various theories that have been applied to and derived from the study of data warehousing and mining. Specifically, certain myths and questions regarding how consumer data is extracted and what organizations do with this data are presented in selections such as "Ethical Dilemmas in Data Mining and Warehousing" by Joseph A. Cazier and Ryan C. LaBrie, "Privacy and Confidentiality Issues in Data Mining" by Yücel Saygin, and "Privacy in Data Mining Textbooks" by James Lawler and John C. Molluzzo. Later chapters in this section explore a different aspect of data mining and warehousing—the specific theories, approaches, and algorithms researchers use to both create and analyze data mining and data warehousing tools. "Data Mining with Incomplete Data" by Hai

Wang and Shouhong Wang conceptualizes the common problem of trying to extract information from an incomplete survey data set, while "A Single Pass Algorithm for Discovering Significant Intervals in Time-Series Data" by Sagar Savla and Sharma Chakravarthy introduces a particular algorithm for extracting significant information from time-series data. In all, the theoretical and abstract issues presented and analyzed within this collection form the backbone of revolutionary data mining and warehousing research and inquiry.

The concluding section of this authoritative reference tool, **Emerging Trends**, highlights research potential within the field of data mining and warehousing while exploring uncharted areas of study for the advancement of the discipline. New trends in data mining and warehousing research discussed in this section include pattern mining, which is explored within "Semantic Data Mining" by Protima Banerjee, Xiaohua Hu, and Illhoi Yoo, integrating data mining with data warehousing, which is studied in "Toward Integrating Data Warehousing with Data Mining Techniques" by Rokia Missaoui, Ganaël Jatteau, Ameur Boujenoui, and Sami Naouali, and mining in medical applications, which is discussed by researchers at Pfizer in the selection "Mining Clinical Trial Data." The connection between data analysis and homeland security is explored within "Homeland Security Data Mining and Link Analysis" by Bhavani Thuraisingham and "Data Mining and Homeland Security" by Jeffrey W. Seifert. The latter of these selections documents how "dataveillance," or the monitoring of data related to an individual's activities, can be both a beneficial and controversial addition to national security. The future of data mining and warehousing, with its infinite potential for growth and change, promises to be as influential as its storied history.

Although the contents of this multi-volume book are organized within the preceding eight sections which offer a progression of coverage of the important concepts, methodologies, technologies, applications, social issues, and emerging trends, the reader can also identify specific contents by utilizing the extensive indexing system listed at the end of each volume. Furthermore, to ensure that the scholar, researcher, and educator have access to the entire contents of this multi-volume set, as well as additional coverage that could not be included in the print version of this publication, the publisher will provide unlimited, multi-user electronic access to the online aggregated database of this collection for the life of the edition, free of charge when a library purchases a print copy. In addition to providing content not included within the print version, this aggregated database is also continually updated to ensure that the most current research is available to those interested in data mining and warehousing.

Data mining and warehousing as a discipline has witnessed fundamental changes during the past two decades, allowing researchers and decision makers around the globe to have access to data and information which, two decades ago, was inaccessible. In addition to this transformation, many traditional organizations and business enterprises have taken advantage of the technologies offered by the development of data mining and warehousing in order to expand and augment their existing ability to make the most use of their databases. This has allowed practitioners and researchers to serve their customers, employees, and stakeholders more effectively and efficiently in the modern information age. With continued technological innovations in information and communication technology and with on-going discovery and research into newer and more innovative techniques and applications, the data mining and warehousing discipline will continue to witness an explosion of information within this rapidly evolving field.

The diverse and comprehensive coverage of data mining and warehousing in this six-volume, authoritative publication will contribute to a better understanding of all topics, research, and discoveries in this developing, significant field of study. Furthermore, the contributions included in this multi-volume collection series will be instrumental in the expansion of the body of knowledge in this enormous field, resulting in a greater understanding of the fundamentals while also fueling the research initiatives in

emerging fields. We at Information Science Reference, along with the editor of this collection, hope that this multi-volume collection will become instrumental in the expansion of the discipline and will promote the continued growth of data mining and warehousing.