Preface

Data mining is the extraction of readily unavailable information from data by sifting regularities and patterns. These ground breaking technologies are bringing major changes in the way people perceive these inter-related processes: the collection of data, archiving and mining it, the creation of information nuggets, and potential threats posed to individual liberty and privacy. This edited book, *Social Implications of Data Mining and Information Privacy: Interdisciplinary Frameworks and Solutions*, brings a collection of chapters that analyze topics related to competing goals: the need to collect data and information for disparate operational objectives, and the necessity to preserve the integrity of the collected data to protect privacy. Currently, scores of data mining applications technologies exist in marketing products, analyzing election results, identification of potential acts of terrorism, and prevention of such threats; uses of data mining in agricultural, in healthcare, and in education among many others.

The warehousing and data mining of data serve the public interest by improving service, reducing costs, and ultimately satisfying customers. The uneasiness on the use of these technologies arise from the fact that data are collected for different purposes, say, credit card charges for purchases of goods and services, the information provided for the transactional purposes are massaged and analyzed for unrelated uses that deviate from the original purposes. All these seemingly disparate data are collected, warehoused, and mined to find patterns. Aggregated data provides useful information for marketing goods and services to a segment of population. The aggregation of data can provide privacy whereas micro targeting mining does not. Data in the hands of unsavory characters can harm the victims and cause overall negative perceptions in providing personal information to conduct business transactions.

The significance of understanding data mining and the confusion it generates due to the misuse of collected data are explored in this edition. Addressing these concerns both legally and ethically by examining the implications of data mining as it impacts privacy and applications on these issues are explored in the various chapters of this edited book. The objective of the book is to provide the most comprehensive, in-depth, and recent coverage of information science and technology in data mining and information privacy disciplines. Various privacy problems are addressed in the public debate and the technology discourse. The chapters address what problems are critical and suggestions to address the solutions. Furthermore, which forms of privacy policy are adequate and the ramifications involved are covered. Carefully selected chapters in this edition include: information ethics, information privacy on the Web, models for data mining metaphors, information privacy in customer relations management, electronic networking in urban neighborhoods, theory of information security empirical validation, electronic collaboration and privacy preservation, machine learning for Web classification and privacy, U.S. federal government and data mining applications in Home Land Security Agency, legal framework
In order to provide the best balanced coverage of concepts and issues related to the selected topics of this book, researchers from around the world were asked to submit proposals describing their proposed coverage and the contribution of such coverage to the book. All proposals were carefully reviewed by the editor in light of their suitability, researcher’s records of similar work in the area of the proposed topics, and the best proposal for topics with multiple proposals. The goal was to assemble the best minds in the information science and technology field from all over the world to contribute entries to the book. Upon the receipt of full entry submissions, each submission was forwarded to at least three expert external reviewers on a double-blind, peer review basis. Only submissions with strong and favorable reviews were chosen as entries for this encyclopedia. In many cases submissions were sent back for several revisions prior to final acceptance.

The following paragraphs provide a brief synopsis on the chapters covered.

Cultural differences among Western and non-Western societies may have impacts on the attitudes of information privacy. Chapter I covers whether information ethics, and by inference, information privacy is culturally relative. The author argues that there must be concepts and principles of information ethics that are universally valid. It analyzes the normative implications of information ethics in a cross-cultural context. It posits the position between moral absolutism and relativism, which are based on intercultural understanding and mutual criticism that could be helpful in overcoming differences and misunderstandings between cultures in their approach to information and information technologies.

Protecting privacy when surfing the Internet is an ongoing challenge to end-users and system administrators. Chapter II provides taxonomy for regulatory and technological solutions to protect user privacy. Information can be collected either overtly or covertly and the protection and the integrity of the information is an ongoing problem that is baffling technological solution due to the open architecture design of the World Wide Web.

Regulatory bodies have not adapted to the moral norms and models used in data mining technology. Building ethical data mining usage is a framework argued in Chapter III. This is done by sketching three models of game theory: pure conflict, pure collaboration, and a mixed motive cooperation game.

Information privacy and security can be compromised in applications such as customer relations management (CRM). Virtual relationships must be carefully managed between customers and businesses in order to add value and minimize unintended consequences that may harm the relationship. Based upon a customer’s requirements of privacy and an enterprise requirement to establish markets and sell goods and services, there is a value exchange relationship. Chapter IV covers a model of an integration of spheres of customer privacy, security, and their implementations.

From an urban information perspective, emerging trends on information networked individualism via the Internet is the topic of Chapter V. The chapter presents a framework of an urban cohesive social network in the dynamics of their existing communicative ecology, and social behavior of urban dwellers. It challenges the accepted view that providing technology application support only, is not adequate to meet the location and proximity requirements in urban neighborhoods information seeking perspectives.

Using variables like user training, security culture, policy relevance, and policy enforcement, a theoretical model was tested to see the influence of these variables on security effectiveness. Both qualitative and quantitative data was used to analyze the survey response using structural equation modeling. Evidence was found supporting the hypothesized model in Chapter VI. Furthermore, it explores a higher factor version of the model that provided an overall fit and general applicability across various demographics of the collected data.
Information sharing from data mining applications is controversial and protected by privacy regulations and ethical concerns. Chapter VII introduces a method for privacy preserving clustering, called dimensionality reduction-based transformation (DRBT). The method relies on random projection to hide the underlying attribute values subjected to cluster analysis. The authors argue that this method has advantages with little overhead cost in CPU intensive operations, and has a sound mathematical foundation.

The classification of Web pages using machine learning techniques is the focus of Chapter VIII. Different techniques are used to classify Web pages by means the most efficient algorithms with search engines to provide accurate results to the user. Additionally, machine learning classifiers can be trained to preserve privacy from unauthenticated users.

Chapter IX covers issues related to U.S. Federal data mining programs. The authors argue that in the context of 9/11, intelligence gathering within and outside United States is one of the primary strategies of the U.S. Homeland Security Agency. These activities have raised a number of issues related to privacy protection and civil liberties of citizens and non-citizens. The chapter extensively discusses relevant issues on the gathering of intelligence for national security and the war on terror and privacy protection of individuals.

The next chapter explores the legal rights and privacy in the United States as related to data mining in contemporary society. It covers the legal frameworks of data mining and privacy which historically lags to technological developments. The author argues that legal rights as related to privacy in the United States is not keeping pace with the data mining activities of businesses, governments, and other entities which is the topic of Chapter X.

Significant advances in telecommunications and GPS sensors technology are enabling the users of these technologies to track down objects and individuals with remarkable precision. As useful as these technologies are, without strict safeguards on privacy, it becomes problematic in the hands of malevolent users. Chapter XI reviews state-of-the-art work on trajectory data privacy. Furthermore, the authors share views regarding the current state and future trend in trajectory data privacy.

Web pages are classified on content or imbedded contextual information. The pages can contain irrelevant information that may reduce performance of the Web classifiers. The accessed data can be mined to predict user authentication and hide personal details to preserve privacy. Chapter XII comprehensively covers the feature selections techniques used by researchers.

Collection of data related to agricultural produce and subsequent data mining of it, can improve production, handling, and public health safety. Chapter XIII deals with recent advances in technology application in agriculture such as uses of RFID and bar coding to trace backward and forward information of fresh produce on production, handling, storage, repackaging, and final sale to consumer. The authors discuss the promises of these technologies in agriculture and potential benefits to business, government, and world health and food safety monitoring organizations in order to increase safety and service, and reduce costs of such products.

Chapter XIV highlights a case study involving research into the science of building teams. Accomplishment of mission goals requires team members to not only possess the required technical skills, but also be able to collaborate effectively. The authors describe a research project that aims to develop an automated staffing system. Any such system requires a large amount of personal information about the potential team members under consideration. Gathering, storing, and applying this data raises a spectrum of concerns, from social and ethical implications, to technical hurdles. The authors hope to highlight these concerns by focusing on their research efforts which include obtaining and using employee data within a small business.
The final chapter gives a summary of data types, mathematical structures, and associated methods of data mining. Topological, order theoretical, algebraic and probability theoretical mathematical structures are introduced in Chapter XV. The \( n \)-dimensional Euclidean space, the model used most for data, is defined. It notes that the treatment of higher dimensional random variables and related data is problematic.

Since topological concepts are less well known than statistical concepts, many examples of metrics are given. Related classification concepts are defined and explained. Possibilities of their quality identification are discussed. One example each is given for topological cluster and for topological discriminant analyses and their implications on preserving privacy.

The coverage of these chapters provide strength to this book for both information science and technology researchers and also decision makers in obtaining a greater understanding of the concepts, issues, problems, trends, challenges, and opportunities related to this field of study. It is my sincere hope that this publication and its vast amount of information and research will assist researchers, teachers, students, and practitioners in enhancing their understanding of the social implications of data mining usage and information privacy and the frameworks and solutions applied.