Cloud technologies have revolutionized the way we store information and perform various computing tasks. With the rise of this new technology, the ability to secure information stored on the cloud becomes a concern. The Handbook of Research on Securing Cloud-Based Databases with Biometric Applications explores the latest innovations in promoting cloud security through human authentication techniques. Exploring methods of access by identification, including the analysis of facial features, fingerprints, DNA, demographics, and voice patterns, this publication is designed especially for IT professionals, academicians, and upper-level students seeking current research surrounding cloud security.

This volume constitutes the proceedings of the 18th Industrial Conference on Advances in Data Mining, ICDM 2018, held in New York, NY, USA, in July 2018. The 146 regular papers presented in this book were carefully reviewed and selected from 146 submissions. The topics range from theoretical aspects of data mining to applications of data mining, such as in multimedia data, in marketing, in medicine and agriculture, and in process control, industry, and society.

Knowledge discovery and data mining have become areas of growing significance because of the recent increasing demand for KDD techniques, including those used in machine learning, databases, statistics, knowledge acquisition, data visualization, and high performance computing. In view of this, and following the success of the five previous PAKDD conferences, the sixth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2002) aimed to provide a forum for the sharing of original research results, innovative ideas, state-of-the-art developments, and implementation experiences in knowledge discovery and data mining among researchers in academic and industrial organizations. Much work went into preparing a program of high quality. We received 128 submissions. Every paper was reviewed by 3 program committee members, and 32 were selected as regular papers and 20 were selected as short papers, representing a 25% acceptance rate for regular papers. The PAKDD 2002 program was further enhanced by two keynote speeches, delivered by Vinip Kumar from the Univ. of Minnesota and Rajeev Rastogi from AT&T. In addition, PAKDD 2002 was complemented by three tutorials, XML and data mining (by Kyuseok Shim and Surajit Chaudhuri), mining customer data across various customer touchpoints at- commerce sites (by Jaideep Srivastava), and data clustering analysis, from simple groupings to scalable clustering with constraints (by Osmar Zaiane and Andrew Foss).

A grand challenge for science is to understand the human implications of global environmental change and to help society cope with those changes. Virtually all the scientific questions associated with this challenge depend on geospatial information (geoinformation) and on the ability of scientists, working individually and in groups, to interact with that information in flexible and increasingly complex ways. Another grand challenge is how to respond to calamities-terrorist activities, other human-induced crises, and natural disasters. Much of the information that underpins emergency preparedness, response, recovery, and mitigation is geospatial in nature. In terrorist situations, for example, origins and destinations of phone calls and e-mail messages, travel patterns of individuals, dispersal patterns of airborne chemicals, assessment of places at risk, and the allocation of resources all involve geospatial information. Much of the work addressing environment- and emergency-related concerns is how people integrate, distill, and create a wide range of seemingly unrelated information. In addition to critical advances in location-aware computing, databases, and data mining methods, advances in the human-computer interface will couple new computational capabilities with human cognitive capabilities. This report outlines an interdisciplinary research roadmap at the intersection of computer science and geospatial information science. The report was developed by a committee convened by the Computer Science and Telecommunications Board of the National Research Council.

The Office of Industrial Technologies (OIT) of the U. S. Department of Energy commissioned the National Research Council (NRC) to undertake a study on required technologies for the Mining Industries of the Future Program to complement information provided to the program by the National Mining Association. Subsequently, the National Institute for Occupational Safety and Health also became a sponsor of this study, and the Statement of Task was expanded to include health and safety. The overall objectives of this study are: (a) to review available information on the U.S. mining industry; (b) to identify critical research and development needs related to the exploration, mining, and processing of coal, minerals, and metals; and (c) to examine the federal contribution to research and development in mining processes.

This book provides comprehensive coverage of fundamentals of database management system. It contains a detailed description on Relational Database Management System Concepts. There are a variety of solved examples and review questions with solutions. This book is for those who require a better understanding of relational data modeling, its purpose, its nature, and the standards used in creating relational data model.

With the rapid advancement of information discovery techniques, machine learning and data mining continue to play a significant role in cybersecurity. Although several conferences, workshops, and journals focus on the fragmented research topics in this area, there has been no single interdisciplinary resource on past and current works and possible paths for future research in this area. This book fills this need. From basic concepts in machine learning and data mining to advanced problems in the machine learning domain, Data Mining and Machine Learning in Cybersecurity provides a unified reference for specific machine learning solutions to cybersecurity problems. It supplies a foundation in cybersecurity fundamentals and surveys contemporary challenges—delineating cutting-edge machine learning and data mining techniques. It also: Unveils cutting-edge techniques for detecting new attacks Contains in-depth discussions of machine learning solutions to detection problems Categorizes methods for detecting, scanning, and profiling intrusions and anomalies Surveys contemporary cybersecurity
problems and unveils state-of-the-art machine learning and data mining solutions. Details preservation data mining methods. This interdisciplinary resource includes: techniques review tables that allow for speedy access to common cybersecurity practices and data mining techniques. Numerous illustrative figures help readers visualize the workflow of complex techniques and more than forty case studies provide a clear understanding of the design and application of data mining and machine learning techniques in cybersecurity.

As cameras become more pervasive in our daily life, vast amounts of video data are generated. The popularity of YouTube and similar websites such as Tudou and Youku provides strong evidence for the increasing role of video in society. One of the main challenges confronting us in the era of information technology is to - fectively rely on the huge and rapidly growing video data accumulating in large multimedia archives. Innovative video processing and analysis techniques will play an increasingly important role in resolving the difficult task of video search and retrieval. A wide range of video-based applications have benefited from - vant, in video surveillance, mining in video data, multimedia interactions, interaction, security and surveillance, copyright protection, and personal entertainment, to name a few. This book provides an overview of emerging new approaches to video search and mining based on promising methods being developed in the computer vision and image analysis community. Video search and mining is a rapidly evolving discipline whose aim is to capture interesting patterns in video data. It has become one of the core areas in the data mining research community. In comparison to other types of data mining (e.g., text), video mining is still in its infancy. Many challenging research problems are facing video mining researchers.

Pattern recognition in data is a well known classical problem that falls under the ambit of data analysis. As we need to handle different data, the nature of patterns, their recognition and the types of data analyses are bound to change. Since the number of data collection channels increases in the recent time and becomes more diversified, many real-world data mining tasks can easily acquire multiple databases from various sources. In these cases, data mining becomes more challenging for several essential reasons. We may encounter sensitive data originating from different sources; those cannot be amalgamated. Even if we are allowed to place different data together, we are certainly not able to analyze them when local identities of patterns are required to be retained. Thus, pattern recognition in multiple databases gives rise to a suite of new, challenging problems different from those encountered before. Association rule mining, global pattern discovery and mining patterns of select items provide different patterns discovery techniques in multiple data sources. Some interesting item-based data analyses are also covered in this book. Interesting patterns such as exceptional patterns, icebergs and periodic patterns have been recently reported. The book presents a thorough influence analysis between items in time-stamped databases. The recent research on mining multiple related databases is covered while some previous contributions to the area are highlighted and contrasted with the most recent developments.

Multi-database mining has been recognized recently as an important and strategically essential area of research in data mining. In this book, we discuss various issues regarding the systematic and efficient development of multi-database mining applications. It explains how systematically one could prepare data warehouses at different branches. As appropriate multi-database mining technique is essential to develop better applications. Also, the efficiency of a multi-database mining application could be improved by processing more patterns in the application. A faster algorithm could also play an important role in developing a better application. Thus the efficiency of a multi-database mining application could be enhanced by choosing an appropriate multi-database mining model, an appropriate pattern synthesizing technique, a better pattern representation technique, and an efficient algorithm for solving the problem. This book illustrates each of these issues either in the context of a specific problem, or in general.

Data Mining: Opportunities and Challenges presents an overview of the state of the art approaches in this new and multidisciplinary field of data mining. The primary objective of this book is to explore the myriad issues regarding data mining, specifically focusing on those areas that explore new methodologies or examine case studies. This book contains numerous chapters written by an international team of forty-four experts representing leading scientists and talented young scholars from seven different countries.

Activities in data warehousing and mining are constantly emerging. Data mining methods, algorithms, online analytical processes, data mart and practical issues consistently evolve, providing a challenge for professionals in the field. Research and Trends in Data Mining Technologies and Applications focuses on the integration between the fields of data warehousing and data mining, with emphasis on the applicability to real-world problems. This book provides an international perspective, highlighting solutions to some of researchers' toughest challenges. Developments in the knowledge discovery process, data models, structures, and design serve as answers and solutions to these emerging challenges.

This book offers a clear and comprehensive introduction to broad learning, one of the novel learning problems studied in data mining and machine learning. Broad learning aims at fusing multiple large-scale information sources of diverse varieties together, and carrying out synergistic data mining tasks across these fused sources in one unified analytic. This book takes online social networks as an application example to introduce the latest alignment and knowledge discovery algorithms. Besides the overview of broad learning, machine learning and social network basics, specific topics covered in this book include network alignment, link prediction, community detection, information diffusion, viral marketing, and network embedding.

This book offers state-of-the-art research and development outcomes on methodologies, techniques, approaches and successful applications in domain driven, actionable knowledge discovery. It bridges the gap between business expectations and research output.

Data Analytics Applied to the Mining Industry describes the key challenges facing the mining sector as it transforms into a digital industry and the potential of data analytics to provide data driven insights across the value chain of a mining company. This book aims to provide a comprehensive guide to the application of data analytics in mining.

Internet usage has become a normal and essential aspect of everyday life. Due to the immense amount of information available on the web, it has become obligatory for users to sift through and categorize the overload of data while removing redundant material. Collaborative Filtering Using Data Mining and Analysis evaluates the latest patterns and trending topics in the utilization of data mining tools and filtering practices. Featuring emergent research and optimization techniques in the areas of opinion mining, text mining, and sentiment analysis, as well as their various applications, this book is an essential reference source for researchers and engineers interested in collaborative filtering.

The U.S. mining sector has the highest fatality rate of any industry in the country. Fortunately, advances made over the past three decades in mining technology, equipment, processes, procedures, and workforce education and training have
Data Mining: Concepts and Techniques provides the concepts and techniques in processing gathered data or information, which will be used in various applications. Specifically, it explains data mining and the tools used in discovering knowledge from the collected data. This book is referred as the knowledge discovery from data (KDD). It focuses on the feasibility, usefulness, effectiveness, and scalability of techniques of large data sets. After describing data mining, this edition explains the meaning of data, knowledge, pre-processing, processing, and warehousing data. It then presents information about data warehouses, online analytical processing (OLAP), and data cube technology. Then, the methods involved in mining frequent patterns, associations, and correlations for large data sets are described. The book details the methods for data classification and introduces the concepts and methods for data clustering. The remaining chapters discuss the outlier detection and the trends, applications, and research frontiers in data mining. This book is intended for Computer Science students, application developers, and professionals who seek information on data mining. Presents dozens of algorithms and implementation examples, all in pseudo-code and suitable for use in real-world, large-scale data mining projects Addresses advanced topics such as mining object-relational databases, spatial databases, multimedia databases, time-series databases, text databases, the World Wide Web, and applications in several fields Provides a comprehensive, practical look at the concepts and techniques you need to get the most out of your data

Sentiment analysis is the computational study of people's opinions, sentiments, emotions, and attitudes. This fascinating problem is increasingly important in business and society. It offers numerous research challenges but promises insight useful to anyone interested in opinion analysis and social media analysis. This book gives a comprehensive introduction to the topic from a primarily natural-language-processing point of view to help readers understand the underlying structure of the problem and the language constructs that are commonly used to express opinions and sentiments. It covers all core areas of sentiment analysis, includes many emerging themes, such as debate analysis, intention mining, and fake-opinion detection, and presents computational methods to analyze and summarize opinions. It will be a valuable resource for researchers and practitioners in natural language processing, computer science, management sciences, and the social sciences.

Drawn from the US National Science Foundation's Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation (NGDM 07), Next Generation of Data Mining explores emerging technologies and applications in data mining as well as new knowledge challenges faced by both communities. Gathered by leading data mining researchers to discuss key challenges and debates upcoming challenges and outlines computational methods. The contributors look at how ecology, astronomy, social science, medicine, finance, and more can benefit from the next generation of data mining techniques. They examine the algorithms, middleware, infrastructure, and privacy policies associated with ubiquitous, distributed, and high performance data mining. They also discuss the impact of new technologies, such as the semantic web, on data mining and provide recommendations for privacy-preserving mechanisms. The dramatic increase in the availability of massive, complex data from various sources is creating computing, storage, communication, and human-computer interaction challenges for data mining. Providing a framework to better understand these fundamental issues, this volume surveys promising approaches to data mining problems that span an array of disciplines.

Mine valuable insights from your data using popular tools and techniques in R About This Book Understand the basics of data mining and why R is a perfect tool for it. Manipulate your data using popular R packages such as ggplot2, dplyr, and so on to gather valuable business insights from it. Apply effective data mining models to perform regression and classification tasks. Who This Book Is For If you are a budding data scientist or a data analyst with a basic knowledge of R, and want to get into the intricacies of data mining in a practical manner, this is the book for you. No previous experience of data mining is required. What You Will Learn Master relevant packages such as dplyr, ggplot2 and so on for data mining Learn how to effectively organize a data mining project through the CRISP-DM methodology Implement data cleaning and validation tasks to get your data ready for data mining activities Execute Exploratory Data Analysis both the numerical and the graphical way Develop simple and multiple regression models along with logistic regression Apply basic ensemble learning techniques to join together...
results from different data mining models. Perform text mining analysis from unstructured pdf files and textual data. Produce reports to effectively communicate objectives, methods, and insights of your analyses. In Detail R is widely used to leverage data mining techniques across many different industries, including finance, healthcare, pharmaceuticals, and retail. This book will empower you to produce and present impressive analyses from data, by selecting and implementing the appropriate data mining techniques in R. It will let you gain these powerful skills while immersing in a one of a kind data mining crime case, where you will be requested to help resolving a real fraud case affecting a commercial company, by the mean of both basic and advanced data mining techniques. If you are new to R, worry not: you will already learn a thing or two on real data the various R packages commonly employed for this kind of tasks. You will also get the chance of apply some of the most popular and effective data mining models and algo, from the basic multiple linear regression to the most advanced Support Vector Machines. Unlike other data mining learning instruments, this book will effectively expose you the theory behind these models, their relevant assumptions and when you can apply to the data you are facing. By the end of the book you will hold a complete toolset of knowledge and how to engineer them, examine your data mining problems and get the most out of your data. Finally, to let you maximize the exposure to the concepts described and the learning process, the book comes packed with a reproducible bundle of commented R scripts and a practical set of data mining cheat sheets. Style and approach This book takes a practical, step-by-step approach to explain the concepts of data mining. Practical use-cases involving real-world datasets are used throughout the book to clearly explain theoretical concepts.

The development of business intelligence has enhanced the visualization of data to inform and facilitate business management and strategizing. By implementing effective data-driven techniques, this allows for advanced reporting tools to cater to company-specific issues and challenges. The Handbook of Research on Advanced Data Mining Techniques and Applications for Business Intelligence is a key resource on the latest advancements in business applications and the use of mining software solutions to achieve optimal decision-making and risk management results. Highlighting innovative studies on data warehousing, business activity monitoring, and text mining, this publication is an ideal reference source for research scholars, management faculty, and practitioners.

Huge amount of data is available in our society and the need for turning such data into useful information and knowledge is urgent. Data mining is an important field addressing that need and significant progress has been achieved in the last decade. In several important application areas, data arises in the format of Multiple Time Series Object (MTSO) data, where each data object consists of a set of time series over a large set of features. Very little research has been conducted towards this kind of data. Examples include computational toxicology, where each data object consists of a set of time series over thousands of genes, and operational stress management, where each data object consists of a set of time series over different measuring points on the human body. The purpose of this dissertation is to conduct a systematic data mining study over microarray time series data, with applications on computational toxicology. More specifically, we will consider several issues: feature selection algorithms for different classification cases, gene markers or feature set selection for toxic chemical exposure detection, toxic chemical exposure time prediction, wildness concept development and applications, and organizing diversified and parsimonious committee. We will formalize and analyze these research problems, design algorithms to address these problems, and perform experimental evaluations of the proposed algorithms. All these studies are based on microarray time series data set provided by Dr. McDougal.

This book is nothing less than a complete and comprehensive survey of the state-of-the-art of terrorism informatics. It covers the application of advanced methodologies and information fusion and analysis. It also lays out techniques to acquire, integrate, analyze, and present large amounts of terrorism-related information. Terrorism research is now conducted across many different fields, and a variety of techniques are being used in order to carry out the analysis. The data mining community is now struggling with mining distributed, interactive and heterogeneous data sources. Agents can be used to manage such data sources for data access, monitoring, integration, and pattern merging from different data mining models and algo, respectively; for instance, multi agent systems face the problem of enhancing agent learning capability, and avoiding the uncertainty of self organization and intelligence emergence. Data mining, if integrated into agent systems, can greatly enhance the learning skills of agents, and assist agents with prediction of future states, thus initiating follow up action or intervention. The data mining community is now struggling with mining distributed, interactive and heterogeneous data sources. Agents can be used to man age such data sources for data access, monitoring, integration, and pattern merging from the infrastructure, gateway, message passing and pattern delivery perspectives. These two examples illustrate the potential of agent mining in handling challenges in respective communities. There is an excellent opportunity to create innovative, dual agent mining interac tion and integration technology, tools and systems which will deliver results in one new technology.

Pattern recognition in data is a well known classical problem that falls under the ambit of data analysis. As we need to handle different data, the nature of patterns, their recognition and the types of data analyses are bound to change. Since the number of data collection channels increases in the recent time and becomes more diversified, many real-world data mining tasks can easily acquire data samples from various sources. In these cases, data mining becomes more challenging for several essential reasons. We may encounter sensitive data originating from different sources - those cannot be amalgamated. Even if we are allowed to place different data together, we are certainly not able to analyze them when local identities of patterns are required to be retained. Thus, pattern recognition in multiple databases gives rise to a suite of new, challenging problems different from those encountered before. Associated with this, the need for leveraging, global pattern discovery and mining patterns of select items provide different patterns discovery techniques in multiple data sources. Some interesting item-based data analyses are also

Download Free Research Issues In Mining Multiple Data Streams
covered in this book. Interesting patterns, such as exceptional patterns, icebergs and periodic patterns have been recently reported. The book presents a thorough influence analysis between items in time-stamped databases. The recent research on mining multiple related databases is covered while some previous contributions to the area are highlighted and contrasted with the most recent developments.

Handbook of Statistical Analysis and Data Mining Applications, Second Edition, is a comprehensive professional reference book that guides business analysts, scientists, engineers and researchers, both academic and industrial, through all stages of data analysis, model building and implementation. The handbook helps users discern technical and business problems, understand the strengths and weaknesses of modern data mining algorithms and employ the right statistical methods for practical application. This book is an ideal reference for users who want to address massive and complex datasets with novel statistical approaches and be able to objectively evaluate analyses and solutions. It has clear, intuitive explanations of the principles and tools for solving problems using modern analytic techniques and discusses their application to real problems in ways accessible and beneficial to practitioners across several areas—from science and engineering, to medicine, academia and commerce. Includes input by practitioners for practitioners Includes tutorials in numerous fields of study that provide step-by-step instruction on how to use supplied tools to build models Contains practical advice from successful real-world implementations Brings together, in a single resource, all the information a beginner needs to understand the tools and issues in data mining to build successful data mining solutions Features clear, intuitive explanations of novel analytical tools and techniques, and their practical applications

This two volume set LNCS 5981 and LNCS 5982 constitutes the refereed proceedings of the 15th International Conference on Database Systems for Advanced Applications, DASFAA 2010, held in Tsukuba, Japan, in April 2010. The 39 revised full papers and 16 revised short papers presented together with 3 invited keynote papers, 22 demonstration papers, 6 industrial papers, and 2 keynote talks were carefully reviewed and selected from 285 submissions. The papers of the first volume are organized in topical sections on P2P-based technologies, data mining technologies, XML search and matching, graphs, spatial databases, XML technologies, time series and streams, advanced data mining, query processing, Web, sensor networks and communications, information management, as well as communities and Web graphs. The second volume contains contributions related to trajectories and moving objects, skyline queries, privacy and security, data streams, similarity search and event processing, storage and advanced topics, industrial, demo papers, and tutorials and panels.

“This book provides a focal point for research and real-world data mining practitioners that advance knowledge discovery from low-quality data; It presents in-depth experiences and methodologies, providing theoretical and empirical guidance to users who have suffered from underlying low-quality data. Contributions also focus on interdisciplinary collaborations among data quality, data processing, data mining, data privacy, and data sharing”--Provided by publisher.

Presents an overview of the main issues of data mining, including its classification, regression, clustering, and ethical issues. Provides readers with knowledge enhancing processes as well as a wide spectrum of data mining applications.

Coal will continue to provide a major portion of energy requirements in the United States for at least the next several decades. It is imperative that accurate information describing the amount, location, and quality of the coal resources and reserves be available to fulfill energy needs. It is also important that the United States extract its coal resources efficiently, safely, and in an environmentally responsible manner. A renewed focus on federal support for coal-related research, coordinated across agencies and with the active participation of the states and industrial sector, is a critical element for each of these requirements. Coal focuses on the research and development needs and priorities in the areas of coal resource and reserve assessments, coal mining and processing, transportation of coal and coal products, and coal utilization.

This book constitutes the refereed proceedings at PAKDD Workshops 2014, held in conjunction with the 18th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) held in Tainan, Taiwan, in May 2014. The 73 revised papers presented were carefully reviewed and selected from 179 submissions. The workshops affiliated with PAKDD 2014 include: Data Analytics for Targeted Healthcare, DANTH; Data Mining and Decision Analytics for Public Health and Wellness, DMDA-Health; Biologically Inspired Data Mining Techniques, BDM; Mobile Data Management, Mining, and Computing on Social Networks, MobiSocial; Big Data Science and Engineering on E-Commerce, BigEC; Cloud Service Discovery, CloudSD; Mobile Sensing, Mining and Visualization for Human Behavior Inferences, MSMV-HB; Scalable Data Analytics: Theory and Algorithms, SDA; Algorithms for Large-Scale Information Processing in Knowledge Discovery, ALSIP; Data Mining in Social Networks, SocNet; Data Mining in Biomedical Informatics and Healthcare, DMBIH; and Pattern Mining and Application of Big Data, BigPMA.