Download File PDF Series Discovery Knowledge And Mining Data Hallcrc Chapman Applications Ytics Business And Cases Use Mining Data Rapidminer

This is likewise one of the factors by obtaining the soft documents of this **Series Discovery Knowledge And Mining Data Hallcrc Chapman Applications Ytics Business And Cases Use Mining Data Rapidminer** by online. You might not require more epoch to spend to go to the books commencement as without difficulty as search for them. In some cases, you likewise attain not discover the notice Series Discovery Knowledge And Mining Data Hallcrc Chapman Applications Ytics Business And Cases Use Mining Data Rapidminer that you are looking for. It will entirely squander the time.

However below, in imitation of you visit this web page, it will be suitably extremely easy to get as competently as download guide Series Discovery Knowledge And Mining Data Hallcrc Chapman Applications Ytics Business And Cases Use Mining Data Rapidminer

It will not say yes many time as we run by before. You can complete it even if operate something else at home and even in your workplace. therefore easy! So, are you question? Just exercise just what we manage to pay for below as competently as evaluation **Series Discovery Knowledge And Mining Data Hallcrc Chapman Applications Ytics Business And Cases Use Mining Data Rapidminer** what you afterward to read!

KEY=BUSINESS - MCLEAN MILLS

KNOWLEDGE DISCOVERY FROM DATA STREAMS

CRC Press *Since the beginning of the Internet age and the increased use of ubiquitous computing devices, the large volume and continuous flow of distributed data have imposed new constraints on the design of learning algorithms. Exploring how to extract knowledge structures from evolving and time-changing data, Knowledge Discovery from Data Streams presents a coherent overview of state-of-the-art research in learning from data streams. The book covers the fundamentals that are imperative to understanding data streams and describes important applications, such as TCP/IP traffic, GPS data, sensor networks, and customer click streams. It also addresses several challenges of data mining in the future, when stream mining will be at the core of many applications. These challenges involve designing useful and efficient data mining solutions applicable to real-world problems. In the appendix, the author includes examples of publicly available software and online data sets. This practical, up-to-date book focuses on the new requirements of the next generation of data mining. Although the concepts presented in the text are mainly about data streams, they also are valid for different areas of machine learning and data mining.*

DATA MINING WITH R

LEARNING WITH CASE STUDIES, SECOND EDITION

CRC Press *Data Mining with R: Learning with Case Studies, Second Edition uses practical examples to illustrate the power of R and data mining. Providing an extensive update to the best-selling first edition, this new edition is divided into two parts. The first part will feature introductory material, including a new chapter that provides an introduction to data mining, to complement the already existing introduction to R. The second part includes case studies, and the new edition strongly revises the R code of the case studies making it more up-to-date with recent packages that have emerged in R. The book does not assume any prior knowledge about R. Readers who are new to R and data mining should be able to follow the case studies, and they are designed to be self-contained so the reader can start anywhere in the document. The book is accompanied by a set of freely available R source files that can be obtained at the book's web site. These files include all the code used in the case studies, and they facilitate the "do-it-yourself" approach followed in the book. Designed for users of data analysis tools, as well as researchers and developers, the book should be useful for anyone interested in entering the "world" of R and data mining. About the Author Luís Torgo is an associate professor in the Department of Computer Science at the University of Porto in Portugal. He teaches Data Mining in R in the NYU Stern School of Business' MS in Business Analytics program. An active researcher in machine learning and data mining for more than 20 years, Dr. Torgo is also a researcher in the Laboratory of Artificial Intelligence and Data Analysis (LIAAD) of INESC Porto LA.*

THE TOP TEN ALGORITHMS IN DATA MINING

CRC Press *Identifying some of the most influential algorithms that are widely used in the data mining community, The Top Ten Algorithms in Data Mining provides a description of each algorithm, discusses its impact, and reviews current and future research. Thoroughly evaluated by independent reviewers, each chapter focuses on a particular algorithm and is written by either the original authors of the algorithm or world-class researchers who have extensively studied the respective algorithm. The book concentrates on the following important algorithms: C4.5, k-Means, SVM, Apriori, EM, PageRank, AdaBoost, kNN, Naive Bayes, and CART. Examples illustrate how each algorithm works and highlight its overall performance in a real-world application. The text covers key topics—including classification, clustering, statistical learning, association analysis, and link mining—in data mining research and development as well as in data mining, machine learning, and artificial intelligence courses. By naming the leading algorithms in this field, this book encourages the use of data mining techniques in a broader realm of real-world applications. It should inspire more data mining researchers to further explore the impact and novel research issues of these algorithms.*

BIOLOGICAL DATA MINING

CRC Press *Like a data-guzzling turbo engine, advanced data mining has been powering post-genome biological studies for two decades. Reflecting this growth, Biological Data Mining presents comprehensive data mining concepts, theories, and applications in current biological and medical research. Each chapter is written by a distinguished team of interdisciplinary data mining researchers who cover state-of-the-art biological topics. The first section of the book discusses challenges and opportunities in analyzing and mining biological sequences and structures to gain insight into molecular functions. The second section addresses emerging computational challenges in interpreting high-throughput Omics data. The book then describes the relationships between data mining and related areas of computing, including knowledge representation, information retrieval, and data integration for structured and unstructured biological data. The last part explores emerging data mining opportunities for biomedical applications. This volume examines the concepts, problems, progress, and trends in developing and applying new data mining techniques to the rapidly growing field of genome biology. By studying the concepts and case studies presented, readers will gain significant insight and develop practical solutions for similar biological data mining projects in the future.*

TEMPORAL DATA MINING

CRC Press *Temporal data mining deals with the harvesting of useful information from temporal data. New initiatives in health care and business organizations have increased the importance of temporal information in data today. From basic data mining concepts to state-of-the-art advances, Temporal Data Mining covers the theory of this subject as well as its application in a variety of fields. It discusses the incorporation of temporality in databases as well as temporal data representation, similarity computation, data classification, clustering, pattern discovery, and prediction. The book also explores the use of temporal data mining in medicine and biomedical informatics, business and industrial applications, web usage mining, and spatiotemporal data mining. Along with various state-of-the-art algorithms, each chapter includes detailed references and short descriptions of relevant algorithms and techniques described in other references. In the appendices, the author explains how data mining fits the overall goal of an organization and how these data can be interpreted for the purpose of characterizing a population. She also provides programs written in the Java language that implement some of the algorithms presented in the first chapter. Check out the author's blog at http://theophanomitsa.wordpress.com/*

DATA CLUSTERING

ALGORITHMS AND APPLICATIONS

CRC Press *Research on the problem of clustering tends to be fragmented across the pattern recognition, database, data mining, and machine learning communities. Addressing this problem in a unified way, Data Clustering: Algorithms and Applications provides complete coverage of the entire area of clustering, from basic methods to more refined and complex data clustering approaches. It pays special attention to recent issues in graphs, social networks, and other domains. The book focuses on three primary aspects of data clustering: Methods, describing key techniques commonly used for clustering, such as feature selection, agglomerative clustering, partitional clustering, density-based clustering, probabilistic clustering, grid-based clustering, spectral clustering, and nonnegative matrix factorization Domains, covering methods used for different domains of data, such as categorical data, text data, multimedia data, graph data, biological data, stream data, uncertain data, time series clustering, high-dimensional clustering, and big data Variations and Insights, discussing important variations of the clustering process, such as semisupervised clustering, interactive clustering, multiview clustering, cluster ensembles, and cluster validation In this book, top researchers from around the world explore the characteristics of clustering problems in a variety of application areas. They also explain how to glean detailed insight from the clustering process—including how to verify the quality of the underlying clusters—through supervision, human intervention, or the automated generation of alternative clusters.*

MACHINE LEARNING AND KNOWLEDGE DISCOVERY FOR ENGINEERING SYSTEMS HEALTH MANAGEMENT

CRC Press *This volume presents state-of-the-art tools and techniques for automatically detecting, diagnosing, and predicting the effects of adverse events in an engineered system. It emphasizes the importance of these techniques in managing the intricate interactions within and between engineering systems to maintain a high degree of reliability. Reflecting the interdisciplinary nature of the field, the book explains how the fundamental algorithms and methods of both physics-based and data-driven approaches effectively address systems health management in application areas such as data centers, aircraft, and software systems.*

SPECTRAL FEATURE SELECTION FOR DATA MINING (OPEN ACCESS)

CRC Press *Spectral Feature Selection for Data Mining introduces a novel feature selection technique that establishes a general platform for studying existing feature selection algorithms and developing new algorithms for emerging problems in real-world applications. This technique represents a unified framework for supervised, unsupervised, and semisupervise*

DATA MINING

A TUTORIAL-BASED PRIMER, SECOND EDITION

CRC Press *Data Mining: A Tutorial-Based Primer, Second Edition provides a comprehensive introduction to data mining with a focus on model building and testing, as well as on interpreting and validating results. The text guides students to understand how data mining can be employed to solve real problems and recognize whether a data mining solution is a feasible alternative for a specific problem. Fundamental data mining strategies, techniques, and evaluation methods are presented and implemented with the help of two well-known software tools. Several new topics have been added to the second edition including an introduction to Big Data and data analytics, ROC curves, Pareto lift charts, methods for handling large-sized, streaming and imbalanced data, support vector*

machines, and extended coverage of textual data mining. The second edition contains tutorials for attribute selection, dealing with imbalanced data, outlier analysis, time series analysis, mining textual data, and more. The text provides in-depth coverage of RapidMiner Studio and Weka's Explorer interface. Both software tools are used for stepping students through the tutorials depicting the knowledge discovery process. This allows the reader maximum flexibility for their hands-on data mining experience.

GEOGRAPHIC DATA MINING AND KNOWLEDGE DISCOVERY

CRC Press *The Definitive Volume on Cutting-Edge Exploratory Analysis of Massive Spatial and Spatiotemporal DatabasesSince the publication of the first edition of Geographic Data Mining and Knowledge Discovery, new techniques for geographic data warehousing (GDW), spatial data mining, and geovisualization (GVis) have been developed. In addition, there has bee*

HANDBOOK OF EDUCATIONAL DATA MINING

CRC Press *Handbook of Educational Data Mining (EDM) provides a thorough overview of the current state of knowledge in this area. The first part of the book includes nine surveys and tutorials on the principal data mining techniques that have been applied in education. The second part presents a set of 25 case studies that give a rich overview of the problems that EDM has addressed. Researchers at the Forefront of the Field Discuss Essential Topics and the Latest Advances With contributions by well-known researchers from a variety of fields, the book reflects the multidisciplinary nature of the EDM community. It brings the educational and data mining communities together, helping education experts understand what types of questions EDM can address and helping data miners understand what types of questions are important to educational design and educational decision making. Encouraging readers to integrate EDM into their research and practice, this timely handbook offers a broad, accessible treatment of essential EDM techniques and applications. It provides an excellent first step for newcomers to the EDM community and for active researchers to keep abreast of recent developments in the field.*

CONTRAST DATA MINING

CONCEPTS, ALGORITHMS, AND APPLICATIONS

CRC Press *A Fruitful Field for Researching Data Mining Methodology and for Solving Real-Life Problems Contrast Data Mining: Concepts, Algorithms, and Applications collects recent results from this specialized area of data mining that have previously been scattered in the literature, making them more accessible to researchers and developers in data mining and other fields. The book not only presents concepts and techniques for contrast data mining, but also explores the use of contrast mining to solve challenging problems in various scientific, medical, and business domains. Learn from Real Case Studies of Contrast Mining Applications In this volume, researchers from around the world specializing in architecture engineering, bioinformatics, computer science, medicine, and systems engineering focus on the mining and use of contrast patterns. They demonstrate many useful and powerful capabilities of a variety of contrast mining techniques and algorithms, including tree-based structures, zero-suppressed binary decision diagrams, data cube representations, and clustering algorithms. They also examine how contrast mining is used in leukemia characterization, discriminative gene transfer and microarray analysis, computational toxicology, spatial and image data classification, voting analysis, heart disease prediction, crime analysis, understanding customer behavior, genetic algorithms, and network security.*

MUSIC DATA MINING

CRC Press *The research area of music information retrieval has gradually evolved to address the challenges of effectively accessing and interacting large collections of music and associated data, such as styles, artists, lyrics, and reviews. Bringing together an interdisciplinary array of top researchers, Music Data Mining presents a variety of approaches to successfully employ data mining techniques for the purpose of music processing. The book first covers music data mining tasks and algorithms and audio feature extraction, providing a framework for subsequent chapters. With a focus on data classification, it then describes a computational approach inspired by human auditory perception and examines instrument recognition, the effects of music on moods and emotions, and the connections between power laws and music aesthetics. Given the importance of social aspects in understanding music, the text addresses the use of the Web and peer-to-peer networks for both music data mining and evaluating music mining tasks and algorithms. It also discusses indexing with tags and explains how data can be collected using online human computation games. The final chapters offer a balanced exploration of hit song science as well as a look at symbolic musicology and data mining. The multifaceted nature of music information often requires algorithms and systems using sophisticated signal processing and machine learning techniques to better extract useful information. An excellent introduction to the field, this volume presents state-of-the-art techniques in music data mining and information retrieval to create novel ways of interacting with large music collections.*

ADVANCED DATA SCIENCE AND ANALYTICS WITH PYTHON

CRC Press *Advanced Data Science and Analytics with Python enables data scientists to continue developing their skills and apply them in business as well as academic settings. The subjects discussed in this book are complementary and a follow-up to the topics discussed in Data Science and Analytics with Python. The aim is to cover important advanced areas in data science using tools developed in Python such as SciKit-learn, Pandas, Numpy, Beautiful Soup, NLTK, NetworkX and others. The model development is supported by the use of frameworks such as Keras, TensorFlow and Core ML, as well as Swift for the development of iOS and MacOS applications. Features: Targets readers with a background in programming, who are interested in the tools used in data analytics and data science Uses Python throughout Presents tools, alongside solved examples, with steps that the reader can easily reproduce and adapt to their needs Focuses on the practical use of the tools rather than on lengthy explanations Provides the reader with the opportunity to use the book whenever needed rather than following a sequential path The book can be read independently from the previous volume and each of the chapters in this volume is sufficiently independent from the others, providing flexibility for the reader. Each of the topics addressed in the book tackles the data science workflow from a practical perspective, concentrating on the*

process and results obtained. The implementation and deployment of trained models are central to the book. Time series analysis, natural language processing, topic modelling, social network analysis, neural networks and deep learning are comprehensively covered. The book discusses the need to develop data products and addresses the subject of bringing models to their intended audiences – in this case, literally to the users' fingertips in the form of an iPhone app. About the Author Dr. Jesús Rogel-Salazar is a lead data scientist in the field, working for companies such as Tympa Health Technologies, Barclays, AKQA, IBM Data Science Studio and Dow Jones. He is a visiting researcher at the Department of Physics at Imperial College London, UK and a member of the School of Physics, Astronomy and Mathematics at the University of Hertfordshire, UK.

INTELLIGENT TECHNOLOGIES FOR WEB APPLICATIONS

CRC Press *The Internet has become an integral part of human life, yet the web still utilizes mundane interfaces to the physical world, which makes Internet operations somewhat mechanical, tedious, and less human-oriented. Filling a large void in the literature, Intelligent Technologies for Web Applications is one of the first books to focus on providing vital fundamental and advanced guidance in the area of Web intelligence for beginners and researchers. The book covers techniques from diverse areas of research, including: Natural language processing Information extraction, retrieval, and filtering Knowledge representation and management Machine learning Databases Data, web, and text mining Human–computer interaction Semantic web technologies To develop effective and intelligent web applications and services, it is critical to discover useful knowledge through analyzing large amounts of content, hidden content structures, or usage patterns of web data resources. Intended to improve and reinforce problem-solving methods in this area, this book delves into the hybridization of artificial intelligence (AI) and web technologies to help simplify complex Web operations. It introduces readers to the state-of-the art development of web intelligence techniques and teaches how to apply these techniques to develop the next generation of intelligent Web applications. The book lays out presented projects, case studies, and innovative ideas, which readers can explore independently as standalone research projects. This material facilitates experimentation with the book's content by including fundamental tools, research directions, practice questions, and additional reading.*

PRACTICAL GRAPH MINING WITH R

CRC Press *Discover Novel and Insightful Knowledge from Data Represented as a GraphPractical Graph Mining with R presents a "do-it-yourself" approach to extracting interesting patterns from graph data. It covers many basic and advanced techniques for the identification of anomalous or frequently recurring patterns in a graph, the discovery of groups or cluste*

COMPUTATIONAL INTELLIGENT DATA ANALYSIS FOR SUSTAINABLE DEVELOPMENT

CRC Press *Going beyond performing simple analyses, researchers involved in the highly dynamic field of computational intelligent data analysis design algorithms that solve increasingly complex data problems in changing environments, including economic, environmental, and social data. Computational Intelligent Data Analysis for Sustainable Development present*

COMPUTATIONAL METHODS OF FEATURE SELECTION

CRC Press *Due to increasing demands for dimensionality reduction, research on feature selection has deeply and widely expanded into many fields, including computational statistics, pattern recognition, machine learning, data mining, and knowledge discovery. Highlighting current research issues, Computational Methods of Feature Selection introduces the basic concepts and principles, state-of-the-art algorithms, and novel applications of this tool. The book begins by exploring unsupervised, randomized, and causal feature selection. It then reports on some recent results of empowering feature selection, including active feature selection, decision-border estimate, the use of ensembles with independent probes, and incremental feature selection. This is followed by discussions of weighting and local methods, such as the ReliefF family, k-means clustering, local feature relevance, and a new interpretation of Relief. The book subsequently covers text classification, a new feature selection score, and both constraint-guided and aggressive feature selection. The final section examines applications of feature selection in bioinformatics, including feature construction as well as redundancy-, ensemble-, and penalty-based feature selection. Through a clear, concise, and coherent presentation of topics, this volume systematically covers the key concepts, underlying principles, and inventive applications of feature selection, illustrating how this powerful tool can efficiently harness massive, high-dimensional data and turn it into valuable, reliable information.*

DATA CLASSIFICATION

ALGORITHMS AND APPLICATIONS

CRC Press *Comprehensive Coverage of the Entire Area of Classification Research on the problem of classification tends to be fragmented across such areas as pattern recognition, database, data mining, and machine learning. Addressing the work of these different communities in a unified way, Data Classification: Algorithms and Applications explores the underlying algorithms of classification as well as applications of classification in a variety of problem domains, including text, multimedia, social network, and biological data. This comprehensive book focuses on three primary aspects of data classification: Methods-The book first describes common techniques used for classification, including probabilistic methods, decision trees, rule-based methods, instance-based methods, support vector machine methods, and neural networks. Domains-The book then examines specific methods used for data domains such as multimedia, text, time-series, network, discrete sequence, and uncertain data. It also covers large data sets and data streams due to the recent importance of the big data paradigm. Variations-The book concludes with insight on variations of the classification process. It discusses ensembles, rare-class learning, distance function learning, active learning, visual learning, transfer learning, and semi-supervised learning as well as evaluation aspects of classifiers.*

PRIVACY-AWARE KNOWLEDGE DISCOVERY

NOVEL APPLICATIONS AND NEW TECHNIQUES

CRC Press *Covering research at the frontier of this field, Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques presents state-of-the-art privacy-preserving data mining techniques for application domains, such as medicine and social networks, that face the increasing heterogeneity and complexity of new forms of data. Renowned authorities from prominent organizations not only cover well-established results—they also explore complex domains where privacy issues are generally clear and well defined, but the solutions are still preliminary and in continuous development. Divided into seven parts, the book provides in-depth coverage of the most novel reference scenarios for privacy-preserving techniques. The first part gives general techniques that can be applied to various applications discussed in the rest of the book. The second section focuses on the sanitization of network traces and privacy in data stream mining. After the third part on privacy in spatio-temporal data mining and mobility data analysis, the book examines time series analysis in the fourth section, explaining how a perturbation method and a segment-based method can tackle privacy issues of time series data. The fifth section on biomedical data addresses genomic data as well as the problem of privacy-aware information sharing of health data. In the sixth section on web applications, the book deals with query log mining and web recommender systems. The final part on social networks analyzes privacy issues related to the management of social network data under different perspectives. While several new results have recently occurred in the privacy, database, and data mining research communities, a uniform presentation of up-to-date techniques and applications is lacking. Filling this void, Privacy-Aware Knowledge Discovery presents novel algorithms, patterns, and models, along with a significant collection of open problems for future investigation.*

MINING SOFTWARE SPECIFICATIONS

METHODOLOGIES AND APPLICATIONS

CRC Press *An emerging topic in software engineering and data mining, specification mining tackles software maintenance and reliability issues that cost economies billions of dollars each year. The first unified reference on the subject, Mining Software Specifications: Methodologies and Applications describes recent approaches for mining specifications of sof*

DATA MINING FOR DESIGN AND MARKETING

CRC Press *Data Mining for Design and Marketing shows how to design and integrate data mining tools into human thinking processes in order to make better business decisions, especially in designing and marketing products and systems. The expert contributors discuss how data mining can identify valuable consumer patterns, which aid marketers and designers in detecting consumers' needs. They also explore visualization tools based on the computational methods of data mining. Discourse analysis, chance discovery, knowledge discovery, formal concept analysis, and an adjacency matrix are just some of the novel approaches covered. The book explains how these methods can be applied to website design, the retrieval of scientific articles from a database, personalized e-commerce support tools, and more. Through the techniques of data mining, this book demonstrates how to effectively design business processes and develop competitive products and services. By embracing data mining tools, businesses can better understand the behavior and needs of their customers.*

CONSTRAINED CLUSTERING

ADVANCES IN ALGORITHMS, THEORY, AND APPLICATIONS

CRC Press *Since the initial work on constrained clustering, there have been numerous advances in methods, applications, and our understanding of the theoretical properties of constraints and constrained clustering algorithms. Bringing these developments together, Constrained Clustering: Advances in Algorithms, Theory, and Applications presents an extensive collection of the latest innovations in clustering data analysis methods that use background knowledge encoded as constraints. Algorithms The first five chapters of this volume investigate advances in the use of instance-level, pairwise constraints for partitional and hierarchical clustering. The book then explores other types of constraints for clustering, including cluster size balancing, minimum cluster size,and cluster-level relational constraints. Theory It also describes variations of the traditional clustering under constraints problem as well as approximation algorithms with helpful performance guarantees. Applications The book ends by applying clustering with constraints to relational data, privacy-preserving data publishing, and video surveillance data. It discusses an interactive visual clustering approach, a distance metric learning approach, existential constraints, and automatically generated constraints. With contributions from industrial researchers and leading academic experts who pioneered the field, this volume delivers thorough coverage of the capabilities and limitations of constrained clustering methods as well as introduces new types of constraints and clustering algorithms.*

TEXT MINING

CLASSIFICATION, CLUSTERING, AND APPLICATIONS

CRC Press *The Definitive Resource on Text Mining Theory and Applications from Foremost Researchers in the Field Giving a broad perspective of the field from numerous vantage points, Text Mining: Classification, Clustering, and Applications focuses on statistical methods for text mining and analysis. It examines methods to automatically cluster and classify text documents and applies these methods in a variety of areas, including adaptive information filtering, information distillation, and text search. The book begins with chapters on the classification of documents into predefined categories. It presents state-of-the-art algorithms and their use in practice. The next chapters describe novel methods for clustering documents into groups that are not predefined. These methods seek to automatically determine topical structures that may exist in a document corpus. The book concludes by discussing various text mining applications that have significant implications for future research and industrial use. There is no doubt that text mining will continue to play a critical role in the development of future information systems and advances in research will be instrumental to their success.*

This book captures the technical depth and immense practical potential of text mining, guiding readers to a sound appreciation of this burgeoning field.

HEALTHCARE DATA ANALYTICS

CRC Press *At the intersection of computer science and healthcare, data analytics has emerged as a promising tool for solving problems across many healthcare-related disciplines. Supplying a comprehensive overview of recent healthcare analytics research, Healthcare Data Analytics provides a clear understanding of the analytical techniques currently available to solve healthcare problems. The book details novel techniques for acquiring, handling, retrieving, and making best use of healthcare data. It analyzes recent developments in healthcare computing and discusses emerging technologies that can help improve the health and well-being of patients. Written by prominent researchers and experts working in the healthcare domain, the book sheds light on many of the computational challenges in the field of medical informatics. Each chapter in the book is structured as a "survey-style" article discussing the prominent research issues and the advances made on that research topic. The book is divided into three major categories: Healthcare Data Sources and Basic Analytics - details the various healthcare data sources and analytical techniques used in the processing and analysis of such data Advanced Data Analytics for Healthcare - covers advanced analytical methods, including clinical prediction models, temporal pattern mining methods, and visual analytics Applications and Practical Systems for Healthcare - covers the applications of data analytics to pervasive healthcare, fraud detection, and drug discovery along with systems for medical imaging and decision support Computer scientists are usually not trained in domain-specific medical concepts, whereas medical practitioners and researchers have limited exposure to the data analytics area. The contents of this book will help to bring together these diverse communities by carefully and comprehensively discussing the most relevant contributions from each domain.*

DATA CLUSTERING

ALGORITHMS AND APPLICATIONS

CRC Press *Research on the problem of clustering tends to be fragmented across the pattern recognition, database, data mining, and machine learning communities. Addressing this problem in a unified way, Data Clustering: Algorithms and Applications provides complete coverage of the entire area of clustering, from basic methods to more refined and complex data clustering approaches. It pays special attention to recent issues in graphs, social networks, and other domains. The book focuses on three primary aspects of data clustering: Methods, describing key techniques commonly used for clustering, such as feature selection, agglomerative clustering, partitional clustering, density-based clustering, probabilistic clustering, grid-based clustering, spectral clustering, and nonnegative matrix factorization Domains, covering methods used for different domains of data, such as categorical data, text data, multimedia data, graph data, biological data, stream data, uncertain data, time series clustering, high-dimensional clustering, and big data Variations and Insights, discussing important variations of the clustering process, such as semisupervised clustering, interactive clustering, multiview clustering, cluster ensembles, and cluster validation In this book, top researchers from around the world explore the characteristics of clustering problems in a variety of application areas. They also explain how to glean detailed insight from the clustering process—including how to verify the quality of the underlying clusters—through supervision, human intervention, or the automated generation of alternative clusters.*

KNOWLEDGE GUIDED MACHINE LEARNING

ACCELERATING DISCOVERY USING SCIENTIFIC KNOWLEDGE AND DATA

CRC Press *Given their tremendous success in commercial applications, machine learning (ML) models are increasingly being considered as alternatives to science-based models in many disciplines. Yet, these "black-box" ML models have found limited success due to their inability to work well in the presence of limited training data and generalize to unseen scenarios. As a result, there is a growing interest in the scientific community on creating a new generation of methods that integrate scientific knowledge in ML frameworks. This emerging field, called scientific knowledge-guided ML (KGML), seeks a distinct departure from existing "data-only" or "scientific knowledge-only" methods to use knowledge and data at an equal footing. Indeed, KGML involves diverse scientific and ML communities, where researchers and practitioners from various backgrounds and application domains are continually adding richness to the problem formulations and research methods in this emerging field. Knowledge Guided Machine Learning: Accelerating Discovery using Scientific Knowledge and Data provides an introduction to this rapidly growing field by discussing some of the common themes of research in KGML using illustrative examples, case studies, and reviews from diverse application domains and research communities as book chapters by leading researchers. KEY FEATURES First-of-its-kind book in an emerging area of research that is gaining widespread attention in the scientific and data science fields Accessible to a broad audience in data science and scientific and engineering fields Provides a coherent organizational structure to the problem formulations and research methods in the emerging field of KGML using illustrative examples from diverse application domains Contains chapters by leading researchers, which illustrate the cutting-edge research trends, opportunities, and challenges in KGML research from multiple perspectives Enables cross-pollination of KGML problem formulations and research methods across disciplines Highlights critical gaps that require further investigation by the broader community of researchers and practitioners to realize the full potential of KGML*

PATTERN RECOGNITION ALGORITHMS FOR DATA MINING

CRC Press *Pattern Recognition Algorithms for Data Mining addresses different pattern recognition (PR) tasks in a unified framework with both theoretical and experimental results. Tasks covered include data condensation, feature selection, case generation, clustering/classification, and rule generation and evaluation. This volume presents various theories, methodologies, and algorithms, using both classical approaches and hybrid paradigms. The authors emphasize large datasets with overlapping, intractable, or nonlinear boundary classes, and datasets that demonstrate granular computing in soft frameworks. Organized into eight chapters, the*

book begins with an introduction to PR, data mining, and knowledge discovery concepts. The authors analyze the tasks of multi-scale data condensation and dimensionality reduction, then explore the problem of learning with support vector machine (SVM). They conclude by highlighting the significance of granular computing for different mining tasks in a soft paradigm.

EVENT MINING

ALGORITHMS AND APPLICATIONS

CRC Press *Event mining encompasses techniques for automatically and efficiently extracting valuable knowledge from historical event/log data. The field, therefore, plays an important role in data-driven system management. Event Mining: Algorithms and Applications presents state-of-the-art event mining approaches and applications with a focus on computing system management. The book first explains how to transform log data in disparate formats and contents into a canonical form as well as how to optimize system monitoring. It then shows how to extract useful knowledge from data. It describes intelligent and efficient methods and algorithms to perform data-driven pattern discovery and problem determination for managing complex systems. The book also discusses data-driven approaches for the detailed diagnosis of a system issue and addresses the application of event summarization in Twitter messages (tweets). Understanding the interdisciplinary field of event mining can be challenging as it requires familiarity with several research areas and the relevant literature is scattered in diverse publications. This book makes it easier to explore the field by providing both a good starting point for readers not familiar with the topics and a comprehensive reference for those already working in this area.*

ACCELERATING DISCOVERY

MINING UNSTRUCTURED INFORMATION FOR HYPOTHESIS GENERATION

CRC Press *Unstructured Mining Approaches to Solve Complex Scientific Problems As the volume of scientific data and literature increases exponentially, scientists need more powerful tools and methods to process and synthesize information and to formulate new hypotheses that are most likely to be both true and important. Accelerating Discovery: Mining Unstructured Information for Hypothesis Generation describes a novel approach to scientific research that uses unstructured data analysis as a generative tool for new hypotheses. The author develops a systematic process for leveraging heterogeneous structured and unstructured data sources, data mining, and computational architectures to make the discovery process faster and more effective. This process accelerates human creativity by allowing scientists and inventors to more readily analyze and comprehend the space of possibilities, compare alternatives, and discover entirely new approaches. Encompassing systematic and practical perspectives, the book provides the necessary motivation and strategies as well as a heterogeneous set of comprehensive, illustrative examples. It reveals the importance of heterogeneous data analytics in aiding scientific discoveries and furthers data science as a discipline.*

NEXT GENERATION OF DATA MINING

CRC Press *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 potential challenges faced by the field. Gathering perspectives from top experts across different disciplines, the book 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.*

ADVANCES IN MACHINE LEARNING AND DATA MINING FOR ASTRONOMY

CRC Press *Advances in Machine Learning and Data Mining for Astronomy documents numerous successful collaborations among computer scientists, statisticians, and astronomers who illustrate the application of state-of-the-art machine learning and data mining techniques in astronomy. Due to the massive amount and complexity of data in most scientific disciplines*

THE ESSENTIALS OF DATA SCIENCE: KNOWLEDGE DISCOVERY USING R

CRC Press *The Essentials of Data Science: Knowledge Discovery Using R presents the concepts of data science through a hands-on approach using free and open source software. It systematically drives an accessible journey through data analysis and machine learning to discover and share knowledge from data. Building on over thirty years' experience in teaching and practising data science, the author encourages a programming-by-example approach to ensure students and practitioners attune to the practise of data science while building their data skills. Proven frameworks are provided as reusable templates. Real world case studies then provide insight for the data scientist to swiftly adapt the templates to new tasks and datasets. The book begins by introducing data science. It then reviews R's capabilities for analysing data by writing computer programs. These programs are developed and explained step by step. From analysing and visualising data, the framework moves on to tried and tested machine learning techniques for predictive modelling and knowledge discovery. Literate programming and a consistent style are a focus throughout the book.*

FOUNDATIONS OF PREDICTIVE ANALYTICS

CRC Press *Drawing on the authors' two decades of experience in applied modeling and data mining, Foundations of Predictive*

Analytics presents the fundamental background required for analyzing data and building models for many practical applications, such as consumer behavior modeling, risk and marketing analytics, and other areas. It also discusses a variety of practical topics that are frequently missing from similar texts. The book begins with the statistical and linear algebra/matrix foundation of modeling methods, from distributions to cumulant and copula functions to Cornish–Fisher expansion and other useful but hard-to-find statistical techniques. It then describes common and unusual linear methods as well as popular nonlinear modeling approaches, including additive models, trees, support vector machine, fuzzy systems, clustering, naïve Bayes, and neural nets. The authors go on to cover methodologies used in time series and forecasting, such as ARIMA, GARCH, and survival analysis. They also present a range of optimization techniques and explore several special topics, such as Dempster–Shafer theory. An in-depth collection of the most important fundamental material on predictive analytics, this self-contained book provides the necessary information for understanding various techniques for exploratory data analysis and modeling. It explains the algorithmic details behind each technique (including underlying assumptions and mathematical formulations) and shows how to prepare and encode data, select variables, use model goodness measures, normalize odds, and perform reject inference. Web Resource The book's website at www.DataMinerXL.com offers the DataMinerXL software for building predictive models. The site also includes more examples and information on modeling.

DATA SCIENCE AND ANALYTICS WITH PYTHON

CRC Press *Data Science and Analytics with Python is designed for practitioners in data science and data analytics in both academic and business environments. The aim is to present the reader with the main concepts used in data science using tools developed in Python, such as SciKit-learn, Pandas, Numpy, and others. The use of Python is of particular interest, given its recent popularity in the data science community. The book can be used by seasoned programmers and newcomers alike. The book is organized in a way that individual chapters are sufficiently independent from each other so that the reader is comfortable using the contents as a reference. The book discusses what data science and analytics are, from the point of view of the process and results obtained. Important features of Python are also covered, including a Python primer. The basic elements of machine learning, pattern recognition, and artificial intelligence that underpin the algorithms and implementations used in the rest of the book also appear in the first part of the book. Regression analysis using Python, clustering techniques, and classification algorithms are covered in the second part of the book. Hierarchical clustering, decision trees, and ensemble techniques are also explored, along with dimensionality reduction techniques and recommendation systems. The support vector machine algorithm and the Kernel trick are discussed in the last part of the book. About the Author Dr. Jesús Rogel-Salazar is a Lead Data scientist with experience in the field working for companies such as AKQA, IBM Data Science Studio, Dow Jones and others. He is a visiting researcher at the Department of Physics at Imperial College London, UK and a member of the School of Physics, Astronomy and Mathematics at the University of Hertfordshire, UK, He obtained his doctorate in physics at Imperial College London for work on quantum atom optics and ultra-cold matter. He has held a position as senior lecturer in mathematics as well as a consultant in the financial industry since 2006. He is the author of the book Essential Matlab and Octave, also published by CRC Press. His interests include mathematical modelling, data science, and optimization in a wide range of applications including optics, quantum mechanics, data journalism, and finance.*

DATA CLUSTERING

THEORY, ALGORITHMS, AND APPLICATIONS

SIAM *Cluster analysis is an unsupervised process that divides a set of objects into homogeneous groups. This book starts with basic information on cluster analysis, including the classification of data and the corresponding similarity measures, followed by the presentation of over 50 clustering algorithms in groups according to some specific baseline methodologies such as hierarchical, center-based, and search-based methods. As a result, readers and users can easily identify an appropriate algorithm for their applications and compare novel ideas with existing results. The book also provides examples of clustering applications to illustrate the advantages and shortcomings of different clustering architectures and algorithms. Application areas include pattern recognition, artificial intelligence, information technology, image processing, biology, psychology, and marketing. Readers also learn how to perform cluster analysis with the C/C++ and MATLAB programming languages.*

SERVICE-ORIENTED DISTRIBUTED KNOWLEDGE DISCOVERY

CRC Press *A new approach to distributed large-scale data mining, service-oriented knowledge discovery extracts useful knowledge from today's often unmanageable volumes of data by exploiting data mining and machine learning distributed models and techniques in service-oriented infrastructures. Service-Oriented Distributed Knowledge Discovery presents techniques, algorithms, and systems based on the service-oriented paradigm. Through detailed descriptions of real software systems, it shows how the techniques, models, and architectures can be implemented. The book covers key areas in data mining and service-oriented computing. It presents the concepts and principles of distributed knowledge discovery and service-oriented data mining. The authors illustrate how to design services for data analytics, describe real systems for implementing distributed knowledge discovery applications, and explore mobile data mining models. They also discuss the future role of service-oriented knowledge discovery in ubiquitous discovery processes and large-scale data analytics. Highlighting the latest achievements in the field, the book gives many examples of the state of the art in service-oriented knowledge discovery. Both novices and more seasoned researchers will learn useful concepts related to distributed data mining and service-oriented data analysis. Developers will also gain insight on how to successfully use service-oriented knowledge discovery in databases (KDD) frameworks.*

MULTIMEDIA DATA MINING

A SYSTEMATIC INTRODUCTION TO CONCEPTS AND THEORY

CRC Press *Collecting the latest developments in the field, Multimedia Data Mining: A Systematic Introduction to Concepts and Theory*

defines multimedia data mining, its theory, and its applications. Two of the most active researchers in multimedia data mining explore how this young area has rapidly developed in recent years. The book first discusses the theoretical foundations of multimedia data mining, presenting commonly used feature representation, knowledge representation, statistical learning, and soft computing techniques. It then provides application examples that showcase the great potential of multimedia data mining technologies. In this part, the authors show how to develop a semantic repository training method and a concept discovery method in an imagery database. They demonstrate how knowledge discovery helps achieve the goal of imagery annotation. The authors also describe an effective solution to large-scale video search, along with an application of audio data classification and categorization. This novel, self-contained book examines how the merging of multimedia and data mining research can promote the understanding and advance the development of knowledge discovery in multimedia data.

INTRODUCTION TO PRIVACY-PRESERVING DATA PUBLISHING

CONCEPTS AND TECHNIQUES

CRC Press *Gaining access to high-quality data is a vital necessity in knowledge-based decision making. But data in its raw form often contains sensitive information about individuals. Providing solutions to this problem, the methods and tools of privacy-preserving data publishing enable the publication of useful information while protecting data privacy. Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques presents state-of-the-art information sharing and data integration methods that take into account privacy and data mining requirements. The first part of the book discusses the fundamentals of the field. In the second part, the authors present anonymization methods for preserving information utility for specific data mining tasks. The third part examines the privacy issues, privacy models, and anonymization methods for realistic and challenging data publishing scenarios. While the first three parts focus on anonymizing relational data, the last part studies the privacy threats, privacy models, and anonymization methods for complex data, including transaction, trajectory, social network, and textual data. This book not only explores privacy and information utility issues but also efficiency and scalability challenges. In many chapters, the authors highlight efficient and scalable methods and provide an analytical discussion to compare the strengths and weaknesses of different solutions.*

FEATURE ENGINEERING FOR MACHINE LEARNING AND DATA ANALYTICS

CRC Press *Feature engineering plays a vital role in big data analytics. Machine learning and data mining algorithms cannot work without data. Little can be achieved if there are few features to represent the underlying data objects, and the quality of results of those algorithms largely depends on the quality of the available features. Feature Engineering for Machine Learning and Data Analytics provides a comprehensive introduction to feature engineering, including feature generation, feature extraction, feature transformation, feature selection, and feature analysis and evaluation. The book presents key concepts, methods, examples, and applications, as well as chapters on feature engineering for major data types such as texts, images, sequences, time series, graphs, streaming data, software engineering data, Twitter data, and social media data. It also contains generic feature generation approaches, as well as methods for generating tried-and-tested, hand-crafted, domain-specific features. The first chapter defines the concepts of features and feature engineering, offers an overview of the book, and provides pointers to topics not covered in this book. The next six chapters are devoted to feature engineering, including feature generation for specific data types. The subsequent four chapters cover generic approaches for feature engineering, namely feature selection, feature transformation based feature engineering, deep learning based feature engineering, and pattern based feature generation and engineering. The last three chapters discuss feature engineering for social bot detection, software management, and Twitter-based applications respectively. This book can be used as a reference for data analysts, big data scientists, data preprocessing workers, project managers, project developers, prediction modelers, professors, researchers, graduate students, and upper level undergraduate students. It can also be used as the primary text for courses on feature engineering, or as a supplement for courses on machine learning, data mining, and big data analytics.*