Discovering Knowledge In Data An Introduction To Data Mining 2nd Edition

Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains introduces the reader to recent research activities in the field of data mining. This book covers association mining, classification, mobile marketing, opinion mining, microarray data mining, internet mining and applications of data mining on biological data, telecommunication and distributed databases, among others, while promoting understanding and implementation of data mining techniques in emerging domains.

Knowledge Discovery in the Social Sciences helps readers find valid, meaningful, and useful information. It is written for researchers and data analysts as well as students who have no prior experience in statistics or computer science. Suitable for a variety of classes—including upper-division courses for undergraduates, introductory courses for graduate students, and courses in data management and advanced statistical methods—the book guides readers in the application of data mining techniques and illustrates the significance of newly discovered knowledge. Readers will learn to: • appreciate the role of data mining in scientific research • develop an understanding of fundamental concepts of data mining and knowledge discovery • use software to carry out data mining tasks • select and assess appropriate models to ensure findings are valid and meaningful • develop basic skills in data preparation, data mining, model selection, and validation • apply concepts with end-of-chapter exercises and review summaries

This book presents a specific and unified approach to Knowledge Discovery and Data Mining, termed IFN for Information Fuzzy Network methodology. Data Mining (DM) is the science of modelling and generalizing common patterns from large sets of multi-type data. DM is a part of KDD, which is the overall process for Knowledge Discovery in Databases. The accessibility and abundance of information today makes this a topic of particular importance and need. The book has three main parts complemented by appendices as well as software and project data that are accessible from the book's web site (http://www.eng.tau.ac.il/maimonifn-kdg2). Part I (Chapters 1-4) starts with the topic of KDD and DM in general and makes reference to other works in the field, especially those related to the information theoretic approach. The remainder of the book presents our work, starting with the IFN theory and algorithms. Part II (Chapters 5-6) discusses the methodology of application and includes case studies. Then in Part III (Chapters 7-9) a comparative study is presented, concluding with some advanced
methods and open problems. The IFN, being a generic methodology, applies to a variety of fields, such as manufacturing, finance, health care, medicine, insurance, and human resources. The appendices expand on the relevant theoretical background and present descriptions of sample projects (including detailed results).

This text surveys research from the fields of data mining and information visualisation and presents a case for techniques by which information visualisation can be used to uncover real knowledge hidden away in large databases.

Data clustering is a highly interdisciplinary field, the goal of which is to divide a set of objects into homogeneous groups such that objects in the same group are similar and objects in different groups are quite distinct. Thousands of theoretical papers and a number of books on data clustering have been published over the past 50 years. However,

“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.

Clear and concise explanations to understand the learning paradigms. Chapters written by leading world experts.

Knowledge Discovery in Big Data from Astronomy and Earth Observation: Astrogateoinformatics bridges the gap between astronomy and geoscience in the context of applications, techniques and key principles of big data. Machine learning and parallel computing are increasingly becoming cross-disciplinary as the phenomena of Big Data is becoming common place. This book provides insight into the common workflows and data science tools used for big data in astronomy and geoscience. After establishing similarity in data gathering, pre-processing and handling, the data science aspects are illustrated in the context of both fields. Software, hardware and algorithms of big data are addressed. Finally, the book offers insight into the emerging science which combines data and expertise from both fields in studying the effect of cosmos on the earth and its inhabitants. Addresses both astronomy and geosciences in parallel, from a big data perspective. Includes introductory information, key principles, applications and the latest techniques. Well-supported by computing and information science-oriented chapters to introduce the necessary knowledge in these fields.

The field of data mining lies at the confluence of predictive analytics, statistical analysis, and business intelligence. Due to the ever-increasing complexity and size of data sets and the wide range of applications in computer science, business, and health care, the process of discovering knowledge in data is more relevant than ever before. This book provides the tools needed to thrive in today's big data world. The author demonstrates how to leverage a company's existing databases to increase profits and market share, and carefully explains the most current data science methods and techniques. The reader will “learn data mining by doing data mining” By adding chapters on data modelling preparation, imputation of missing data, and multivariate statistical analysis, Discovering Knowledge in Data, Second Edition remains the eminent reference on data mining. The second edition of a highly praised, successful reference on data mining, with thorough coverage of big data applications, predictive analytics, and statistical analysis. Includes new chapters on Multivariate Statistics, Preparing to Model the Data, and Imputation of Missing Data, and an Appendix on Data Summarization and Visualization. Offers extensive coverage of the R statistical programming language. Contains 280 end-of-chapter exercises. Includes a companion website for university instructors who adopt the book.

Mobile communications and ubiquitous computing generate large volumes of data. Mining this data can produce useful knowledge, yet individual privacy is at risk. This book investigates the various scientific and technological issues of mobility data, open problems, and roadmap. The editors manage a research project called GeoPKDD, Geographic Privacy-Aware Knowledge Discovery and Delivery, and this book relates their findings in 13 chapters covering all related subjects.

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.

This book provides a broad overview of the topic Bioinformatics with focus on data, information and
knowledge. From data acquisition and storage to visualization, ranging through privacy, regulatory and
other practical and theoretical topics, the author touches several fundamental aspects of the innovative
interface between Medical and Technology domains that is Biomedical Informatics. Each chapter starts by
providing a useful inventory of definitions and commonly used acronyms for each topic and throughout the
text, the reader finds several real-world examples, methodologies and ideas that complement the
technical and theoretical background. This new edition includes new sections at the end of each chapter,
called "future outlook and research avenues," providing pointers to future challenges. At the beginning of
each chapter a new section called "key problems", has been added, where the author discusses possible
traps and unsolvable or major problems.

Modern computer-based control systems are able to collect a large amount of information, display it to
operators and store it in databases but the interpretation of the data and the subsequent decision making
relies mainly on operators with little computer support. This book introduces developments in automatic
analysis and interpretation of process-operational data both in real-time and over the operational history,
and describes new concepts and methodologies for developing intelligent, state space-based systems for
process monitoring, control and diagnosis. The book brings together new methods and algorithms from
process monitoring and control, data mining and knowledge discovery, artificial intelligence, pattern
recognition, and causal relationship discovery, as well as signal processing. It also provides a framework
for integrating plant operators and supervisors into the design of process monitoring and control systems.

The field of data mining has made significant and far-reaching advances over the past three decades.
Because of its potential power for solving complex problems, data mining has been successfully applied to
diverse areas such as business, engineering, social media, and biological science. Many of these
applications search for patterns in complex structural information. In biomedicine for example, modeling
complex biological systems requires linking knowledge across many levels of science, from genes to
disease. Further, the data characteristics of the problems have also grown from static to dynamic and
spatiotemporal, complete to incomplete, and centralized to distributed, and grow in their scope and size
(this is known as big data). The effective integration of big data for decision-making also requires privacy
preservation. The contributions to this monograph summarize the advances of data mining in the
respective fields. This volume consists of nine chapters that address subjects ranging from mining data
from opinion, spatiotemporal databases, discriminative subgraph patterns, path knowledge discovery,
social media, and privacy issues to the subject of computation reduction via binary matrix factorization.

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

This comprehensive textbook on data mining details the unique steps of the knowledge discovery process
that prescribes the sequence in which data mining projects should be performed, from problem and data
understanding through data preprocessing to deployment of the results. This knowledge discovery
approach is what distinguishes Data Mining from other texts in this area. The book provides a suite of
exercises and includes links to instructional presentations. Furthermore, it contains appendices of relevant
mathematical material.
Data Mining is the science and technology of exploring large and complex bodies of data in order to discover useful patterns. It is extremely important because it enables modeling and knowledge extraction from abundant data availability. This book introduces soft computing methods extending the envelope of problems that data mining can solve efficiently. It presents practical soft-computing approaches in data mining and includes various real-world case studies with detailed results.

Data Mining is the science and technology of exploring data in order to discover previously unknown patterns. It is a part of the overall process of Knowledge Discovery in Databases (KDD). The accessibility and abundance of information today makes data mining a matter of considerable importance and necessity. This book provides an introduction to the field with an emphasis on advanced decomposition methods in general data mining tasks and for classification tasks in particular. The book presents a complete methodology for decomposing classification problems into smaller and more manageable sub-problems that are solvable by using existing tools. The various elements are then joined together to solve the initial problem. The benefits of decomposition methodology in data mining include: increased performance (classification accuracy); conceptual simplification of the problem; enhanced feasibility for huge databases; clearer and more comprehensible results; reduced runtime by solving smaller problems and by using parallel/distributed computation; and the opportunity of using different techniques for individual sub-problems.

This book integrates two areas of computer science, namely data mining and evolutionary algorithms. Both these areas have become increasingly popular in the last few years, and their integration is currently an active research area. In general, data mining consists of extracting knowledge from data. The motivation for applying evolutionary algorithms to data mining is that evolutionary algorithms are robust search methods which perform a global search in the space of candidate solutions. This book emphasizes the importance of discovering comprehensible, interesting knowledge, which is potentially useful for intelligent decision making. The text explains both basic concepts and advanced topics.

Modern medicine generates, almost daily, huge amounts of heterogeneous data. For example, medical data may contain SPECT images, signals like ECG, clinical information like temperature, cholesterol levels, etc., as well as the physician's interpretation. Those who deal with such data understand that there is a widening gap between data collection and data comprehension. Computerized techniques are needed to help humans address this problem. This volume is devoted to the relatively young and growing field of medical data mining and knowledge discovery. As more and more medical procedures employ imaging as a preferred diagnostic tool, there is a need to develop methods for efficient mining in databases of images. Other significant features are security and confidentiality concerns. Moreover, the physician's interpretation of images, signals, or other technical data, is written in unstructured English which is very difficult to mine. This book addresses all these specific features.

Apply powerful Data Mining Methods and Models to Leverage your Data for Actionable Results Data Mining Methods and Models provides: * The latest techniques for uncovering hidden nuggets of information * The insight into how the data mining algorithms actually work * The hands-on experience of performing data mining on large data sets Data Mining Methods and Models: * Applies a "white box" methodology, emphasizing an understanding of the model structures underlying the software * Walks the reader through the various algorithms and provides examples of the operation of the algorithms on actual large data sets, including a detailed case study, "Modeling Response to Direct-Mail Marketing" * Tests the reader's level of understanding of the concepts and methodologies, with over 110 chapter exercises * Demonstrates the Clementine data mining software suite, WEKA open source data mining software, SPSS statistical software, and Minitab statistical software * Includes a companion Web site, www.dataminingconsultant.com, where the data sets used in the book may be downloaded, along with a comprehensive set of data mining resources. Faculty adopters of the book have access to an array of helpful resources, including solutions to all exercises, a PowerPoint(r) presentation of each chapter, sample data mining course projects and accompanying data sets, and multiple-choice chapter quizzes. With its emphasis on learning by doing, this is an excellent textbook for students in business, computer science, and statistics, as well as a problem-solving reference for data analysts and professionals in the field. An Instructor's Manual presenting detailed solutions to all the problems in the book is available online.
Currently there are major challenges in data mining applications in the geosciences. This is due primarily to the fact that there is a wealth of available mining data amid an absence of the knowledge and expertise necessary to analyze and accurately interpret the same data. Most geoscientists have no practical knowledge or experience using data mining techniques. For the few that do, they typically lack expertise in using data mining software and in selecting the most appropriate algorithms for a given application. This leads to a paradoxical scenario of "rich data but poor knowledge". The true solution is to apply data mining techniques in geosciences databases and to modify these techniques for practical applications. Authored by a global thought leader in data mining, Data Mining and Knowledge Discovery for Geoscientists addresses these challenges by summarizing the latest developments in geosciences data mining and arming scientists with the ability to apply key concepts to effectively analyze and interpret vast amounts of critical information. Focuses on 22 of data mining's most practical algorithms and popular application samples Features 36 case studies and end-of-chapter exercises unique to the geosciences to underscore key data mining applications Presents a practical and integrated system of data mining and knowledge discovery for geoscientists Rigorous yet broadly accessible to geoscientists, engineers, researchers and programmers in data mining Introduces widely used algorithms, their basic principles and conditions of applications, diverse case studies, and suggests algorithms that may be suitable for specific applications
software engineering and knowledge engineering can easily understand and obtain useful information. Each chapter covers one topic and can be read independently of other chapters, providing both a general survey of the topic and an in-depth exposition of the state of the art. Practitioners will find this handbook useful when looking for solutions to practical problems. Researchers can use it for quick access to the background, current trends and most important references regarding a certain topic. The handbook consists of two volumes. Volume One covers the basic principles and applications of software engineering and knowledge engineering. Volume Two will cover the basic principles and applications of visual and multimedia software engineering, knowledge engineering, data mining for software knowledge, and emerging topics in software engineering and knowledge engineering.

Data Mining and Knowledge Discovery Handbook organizes all major concepts, theories, methodologies, trends, challenges and applications of data mining (DM) and knowledge discovery in databases (KDD) into a coherent and unified repository. This book first surveys, then provides comprehensive yet concise algorithmic descriptions of methods, including classic methods plus the extensions and novel methods developed recently. This volume concludes with in-depth descriptions of data mining applications in various interdisciplinary industries including finance, marketing, medicine, biology, engineering, telecommunications, software, and security. Data Mining and Knowledge Discovery Handbook is designed for research scientists and graduate-level students in computer science and engineering. This book is also suitable for professionals in fields such as computing applications, information systems management, and strategic research management.

Mohamed Medhat Gaber “It is not my aim to surprise or shock you – but the simplest way I can summarise is to say that there are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until – in a visible future – the range of problems they can handle will be coextensive with the range to which the human mind has been applied” by Herbert A. Simon (1916-2001)

Overview This book suits both graduate students and researchers with a focus on discovering knowledge from scientific data. The use of computational power for data analysis and knowledge discovery in scientific disciplines has found its roots with the resolution of high-performance computing systems. Computational science in physics, chemistry, and biology represents the first step towards automation of data analysis tasks. The rational behind the development of computational science in different areas was automating mathematical operations performed in those areas. There was no attention paid to the scientific discovery process. Automated Scientific Discovery (ASD) [1–3] represents the second natural step. ASD attempted to automate the process of theory discovery supported by studies in philosophy of science and cognitive sciences. Although early research articles have shown great successes, the area has not evolved due to many reasons. The most important reason was the lack of interaction between scientists and the automating systems.

“This is a new edition of a highly praised, successful reference on data mining, now more important than ever due to the growth of the field and wide range of applications. This edition features new chapters on multivariate statistical analysis, covering analysis of variance and chi-square procedures; cost-benefit analyses; and time-series data analysis. There is also extensive coverage of the R statistical programming language. Graduate and advanced undergraduate students of computer science and statistics, managers/CEOs/CFOs, marketing executives, market researchers and analysts, sales analysts, and medical professionals will want this comprehensive reference”--

As computer power grows and data collection technologies advance, a plethora of data is generated in almost every field where computers are used. The computer generated data should be analyzed by computers; without the aid of computing technologies, it is certain that huge amounts of data collected will not ever be examined, let alone be used to our advantages. Even with today’s advanced computer technologies (e. g., machine learning and data mining systems), discovering knowledge from data can still be fiendishly hard due to the characteristics of the computer generated data. Taking its simplest form, raw data are represented in feature-values. The size of a dataset can be measured in two dimensions, number of features (N) and number of instances (P). Both N and P can be enormously large. This enormity may cause serious problems to many data mining systems. Feature selection is one of the long existing methods that deal with these problems. Its objective is to select a minimal subset of features according to some reasonable criteria so that the original task can be achieved equally well, if not better. By choosing a minimal subset of features, irrelevant and redundant features are removed according to the criterion. When N is reduced, the data space shrinks and in a sense, the data set is now a better representative of the
The field of data mining lies at the confluence of predictive analytics, statistical analysis, and business intelligence. Due to the ever-increasing complexity and size of data sets and the wide range of applications in computer science, business, and health care, the process of discovering knowledge in data is more relevant than ever before. This book provides the tools needed to thrive in today's big data world. The author demonstrates how to leverage a company's existing databases to increase profits and market share, and carefully explains the most current data science methods and techniques. The reader will “learn data mining by doing data mining”. By adding chapters on data modelling preparation, imputation of missing data, and multivariate statistical analysis, Discovering Knowledge in Data, Second Edition remains the eminent reference on data mining. The second edition of a highly praised, successful reference on data mining, with thorough coverage of big data applications, predictive analytics, and statistical analysis. Includes new chapters on Multivariate Statistics, Preparing to Model the Data, and Imputation of Missing Data, and an Appendix on Data Summarization and Visualization Offers extensive coverage of the R statistical programming language Contains 280 end-of-chapter exercises Includes a companion website for university instructors who adopt the book

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.

Knowledge Discovery in the Social Sciences helps readers find valid, meaningful, and useful information. It is written for researchers and data analysts as well as students who have no prior experience in statistics or computer science. Suitable for a variety of classes—including upper-division courses for undergraduates, introductory courses for graduate students, and courses in data management and advanced statistical methods—the book guides readers in the application of data mining techniques and illustrates the significance of newly discovered knowledge. Readers will learn to: • appreciate the role of data mining in scientific research • develop an understanding of fundamental concepts of data mining and knowledge discovery • use software to carry out data mining tasks • select and assess appropriate models to ensure findings are valid and meaningful • develop basic skills in data preparation, data mining, model selection, and validation • apply concepts with end-of-chapter exercises and review summaries

Eight sections of this book span fundamental issues of knowledge discovery, classification and clustering, trend and deviation analysis, dependency derivation, integrated discovery systems, augmented database systems and application case studies. The appendices provide a list of terms used in the literature of the field of data mining and knowledge discovery in databases, and a list of online resources for the KDD researcher.

Data analysis, machine learning and knowledge discovery are research areas at the intersection of computer science, artificial intelligence, mathematics and statistics. They cover general methods and techniques that can be applied to a vast set of applications such as web and text mining, marketing, medicine, bioinformatics and business intelligence. This volume contains the revised versions of selected papers in the field of data analysis, machine learning and knowledge discovery presented during the 36th annual conference of the German Classification Society (GfKl). The conference was held at the University of Hildesheim (Germany) in August 2012. ?

Data mining is the art and science of intelligent data analysis. By building knowledge from information,
data mining adds considerable value to the ever increasing stores of electronic data that abound today. In performing data mining many decisions need to be made regarding the choice of methodology, the choice of data, the choice of tools, and the choice of algorithms. Throughout this book the reader is introduced to the basic concepts and some of the more popular algorithms of data mining. With a focus on the hands-on end-to-end process for data mining, Williams guides the reader through various capabilities of the easy to use, free, and open source Rattle Data Mining Software built on the sophisticated R Statistical Software. The focus on doing data mining rather than just reading about data mining is refreshing. The book covers data understanding, data preparation, data refinement, model building, model evaluation, and practical deployment. The reader will learn to rapidly deliver a data mining project using software easily installed for free from the Internet. Coupling Rattle with R delivers a very sophisticated data mining environment with all the power, and more, of the many commercial offerings.

"This book focuses on the mathematical models and methods that support most data mining applications and solution techniques, covering such topics as association rules; Bayesian methods; data visualization; kernel methods; neural networks; text, speech, and image recognition; an invaluable resource for scholars and practitioners in the fields of biomedicine, engineering, finance, manufacturing, marketing, performance measurement, and telecommunications"--Provided by publisher.

Data mining from traditional relational databases as well as from non-traditional ones such as semi-structured data, Web data, and scientific databases housing biological, linguistic, and sensor data has recently become a popular way of discovering hidden knowledge. This book on database support for data mining is developed to approaches exploiting the available database technology, declarative data mining, intelligent querying, and associated issues, such as optimization, indexing, query processing, languages, and constraints. Attention is also paid to the solution of data preprocessing problems, such as data cleaning, discretization, and sampling. The 16 reviewed full papers presented were carefully selected from various workshops and conferences to provide complete and competent coverage of the core issues. Some papers were developed within an EC funded project on discovering knowledge with inductive queries.

Copyright code : b72bbeb3c133c2617d8bc2da417a80d3