

## Data Mining Case Studies

The field of data mining provides techniques for automated discovery of valuable information from the accumulated data of computerized operations of enterprises. This book offers a clear and comprehensive introduction to both data mining theory and practice. It is written primarily as a textbook for the students of computer science, management, computer applications, and information technology. The book ensures that the students learn the major data mining techniques even if they do not have a strong mathematical background. The techniques include data pre-processing, association rule mining, supervised classification, cluster analysis, web data mining, search engine query mining, data warehousing and OLAP. To enhance the understanding of the concepts introduced, and to show how the techniques described in the book are used in practice, each chapter is followed by one or two case studies that have been published in scholarly journals. Most case studies deal with real business problems (for example, marketing, e-commerce, CRM). Studying the case studies provides the reader with a greater insight into the data mining techniques. The book also provides many examples, review questions, multiple choice questions, chapter-end exercises and a good list of references and Web resources especially those which are easy to understand and useful for students. A number of class projects have also been included.

**Powerful, Flexible Tools for a Data-Driven World** As the data deluge continues in today's world, the need to master data mining, predictive analytics, and business analytics has never been greater. These techniques and tools provide unprecedented insights into data, enabling better decision making and forecasting, and ultimately the solution of increasingly complex problems. **Learn from the Creators of the RapidMiner Software** Written by leaders in the data mining community, including the developers of the RapidMiner software, **RapidMiner: Data Mining Use Cases and Business Analytics Applications** provides an in-depth introduction to the application of data mining and business analytics techniques and tools in scientific research, medicine, industry, commerce, and diverse other sectors. It presents the most powerful and flexible open source software solutions: RapidMiner and RapidAnalytics. The software and their extensions can be freely downloaded at [www.RapidMiner.com](http://www.RapidMiner.com). **Understand Each Stage of the Data Mining Process** The book and software tools cover all relevant steps of the data mining process, from data loading, transformation, integration, aggregation, and visualization to automated feature selection, automated parameter and process optimization, and integration with other tools, such as R packages or your IT infrastructure via web services. The book and software also extensively discuss the analysis of unstructured data, including text and image mining. **Easily Implement Analytics Approaches Using RapidMiner and RapidAnalytics** Each chapter describes an application, how to approach it with data mining methods, and how to implement it with RapidMiner and RapidAnalytics. These application-oriented chapters give you not only the necessary analytics to solve problems and tasks, but also reproducible, step-by-step descriptions of using RapidMiner and RapidAnalytics. The case studies serve as blueprints for your own data mining applications, enabling you to effectively solve similar problems.

With the rise in data science development, we now have many remarkable techniques and tools to extend data analysis from numeric and categorical data to textual data. Sifting through the open-ended responses from a survey, for example, was an arduous process when performed by hand. Using a case study approach, this book was written for business analysts who wish to increase their skills in extracting answers for text data in order to support business decision making. Most of the exercises use Excel, today's most common analysis tool, and R, a popular analytic computer environment. The techniques covered range from the most basic text analytics, such as key word analysis, to more sophisticated techniques, such as topic extraction and text similarity scoring. Companion files with numerous datasets are included for use with case studies and exercises. **FEATURES:** Organized by tool or technique, with the basic techniques presented first and the more sophisticated techniques presented later **Uses Excel and R for datasets in case studies and exercises** Features the CRISP-DM data mining standard with early chapters for conducting the preparatory steps in data mining **Companion files with numerous datasets and figures from the text.**

This volume introduces data mining through case studies of enrollment management. Six case studies employed data mining for solving real-life issues in enrollment yield, retention, transfer-outs, utilization of advanced-placement scores, and predicting graduation rates, among others. The authors furnish a tangible sense of data mining at work. The volume also demonstrates that data mining bears great potential to enhance institutional research. The opening chapter deciphers the similarities and differences between data mining and statistics, debunks the myths surrounding both data mining and traditional statistics, and points out the intrinsic conflict between statistical inference and the emerging need for individual pattern recognition and resulting customized treatment of students - the so-called new reality in applied institutional research. This is the 131st volume of *New Directions for Institutional Research*, a quarterly journal published by Jossey-Bass. [Click here to see the entire list of titles for New Directions for Institutional Research.](#)

Whether you are brand new to data mining or working on your tenth predictive analytics project, *Commercial Data Mining* will be there for you as an accessible reference outlining the entire process and related themes. In this book, you'll learn that your organization does not need a huge volume of data or a Fortune 500 budget to generate business using existing information assets. Expert author David Nettleton guides you through the process from beginning to end and covers everything from business objectives to data sources, and selection to analysis and predictive modeling. *Commercial Data Mining* includes case studies and practical examples from Nettleton's more than 20 years of commercial experience. Real-world cases covering customer loyalty, cross-selling, and audience prediction in industries including insurance, banking, and media illustrate the concepts and techniques explained throughout the book. Illustrates cost-benefit evaluation of potential projects **Includes vendor-agnostic advice on what to look for in off-the-shelf solutions as well as tips on building your own data mining tools** **Approachable reference can be read from cover to cover by readers of all experience levels** **Includes practical examples and case studies as well as actionable business insights from author's own experience**

This manager's guide to customer-centric business intelligence teaches data mining in an accessible way, explaining how it drives next-generation customer relationship strategies. Readers learn how to find patterns such as customer buying habits within their huge stores of data.

*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

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.

Text Mining and Visualization: Case Studies Using Open-Source Tools provides an introduction to text mining using some of the most popular and powerful open-source tools: KNIME, RapidMiner, Weka, R, and Python. The contributors—all highly experienced with text mining and open-source software—explain how text data are gathered and processed from a w R and Data Mining introduces researchers, post-graduate students, and analysts to data mining using R, a free software environment for statistical computing and graphics. The book provides practical methods for using R in applications from academia to industry to extract knowledge from vast amounts of data. Readers will find this book a valuable guide to the use of R in tasks such as classification and prediction, clustering, outlier detection, association rules, sequence analysis, text mining, social network analysis, sentiment analysis, and more. Data mining techniques are growing in popularity in a broad range of areas, from banking to insurance, retail, telecom, medicine, research, and government. This book focuses on the modeling phase of the data mining process, also addressing data exploration and model evaluation. With three in-depth case studies, a quick reference guide, bibliography, and links to a wealth of online resources, R and Data Mining is a valuable, practical guide to a powerful method of analysis. Presents an introduction into using R for data mining applications, covering most popular data mining techniques Provides code examples and data so that readers can easily learn the techniques Features case studies in real-world applications to help readers apply the techniques in their work

A step-by-step guide to data mining applications in CRM. Following a handbook approach, this book bridges the gap between analytics and their use in everyday marketing, providing guidance on solving real business problems using data mining techniques. The book is organized into three parts. Part one provides a methodological roadmap, covering both the business and the technical aspects. The data mining process is presented in detail along with specific guidelines for the development of optimized acquisition, cross/ deep/ up selling and retention campaigns, as well as effective customer segmentation schemes. In part two, some of the most useful data mining algorithms are explained in a simple and comprehensive way for business users with no technical expertise. Part three is packed with real world case studies which employ the use of three leading data mining tools: IBM SPSS Modeler, RapidMiner and Data Mining for Excel. Case studies from industries including banking, retail and telecommunications are presented in detail so as to serve as templates for developing similar applications. Key Features: Includes numerous real-world case studies which are presented step by step, demystifying the usage of data mining models and clarifying all the methodological issues. Topics are presented with the use of three leading data mining tools: IBM SPSS Modeler, RapidMiner and Data Mining for Excel. Accompanied by a website featuring material from each case study, including datasets and relevant code. Combining data mining and business knowledge, this practical book provides all the necessary information for designing, setting up, executing and deploying data mining techniques in CRM. Effective CRM using Predictive Analytics will benefit data mining practitioners and consultants, data analysts, statisticians, and CRM officers. The book will also be useful to academics and students interested in applied data mining.

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.

"Because so much data is now becoming readily available to investigate health outcomes, it is important to examine just how statistical models are used to do this. This book studies health outcomes research using data mining techniques"--Provided by publisher.

Data mining is well on its way to becoming a recognized discipline in the overlapping areas of IT, statistics, machine learning, and AI. Practical Data Mining for Business presents a user-friendly approach to data mining methods, covering the typical uses to which it is applied. The methodology is complemented by case studies to create a versatile reference book, allowing readers to look for specific methods as well as for specific applications. The book is formatted to allow statisticians, computer scientists, and economists to cross-reference from a particular application or method to sectors of interest.

Machine learning allows for non-conventional and productive answers for issues within various fields, including problems related to visually perceptive computers. Applying these strategies and algorithms to the area of computer vision allows for higher achievement in tasks such as spatial recognition, big data collection, and image processing. There is a need for research that seeks to understand the development and efficiency of current methods that enable machines to see. Challenges and Applications for Implementing Machine Learning in Computer Vision is a collection of innovative research that combines theory and practice on adopting the latest deep learning advancements for machines capable of visual processing. Highlighting a wide range of topics such as video segmentation, object recognition, and 3D modelling, this publication is ideally designed for computer scientists, medical professionals, computer engineers, information technology practitioners, industry experts, scholars, researchers, and students seeking current research on the utilization of evolving computer vision techniques.

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.

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

The versatile capabilities and large set of add-on packages make R an excellent alternative to many existing and often expensive data mining tools. Exploring this area from the perspective of a practitioner, Data Mining with R: Learning with Case Studies uses practical examples to illustrate the power of R and data mining. Assuming no prior knowledge of R or data mining/statistical techniques, the book covers a diverse set of problems that pose different challenges in terms of size, type of data, goals of analysis, and analytical tools. To present the main data mining processes and techniques, the author takes a hands-on approach that utilizes a series of detailed, real-world case studies: Predicting algae blooms Predicting stock market returns Detecting fraudulent transactions Classifying microarray samples With these case studies, the author supplies all necessary steps, code, and data. Web ResourceA supporting website mirrors the do-it-yourself approach of the text. It offers a collection of freely available R source files that encompass all the code used in the case studies. The site also provides the data sets from the case studies as well as an R package of several functions.

The healthcare industry produces a constant flow of data, creating a need for deep analysis of databases through data mining tools and techniques resulting in expanded medical research, diagnosis, and treatment. Data Mining and Medical Knowledge Management: Cases and Applications presents case studies on applications of various modern data mining methods in several important areas of medicine, covering classical data mining methods, elaborated approaches related to mining in electroencephalogram and electrocardiogram data, and methods related to mining in genetic data. A premier resource for those involved in data mining and medical knowledge management, this book tackles ethical issues related to cost-sensitive learning in medicine and produces theoretical contributions concerning general problems of data, information, knowledge, and ontologies.

KDD '09: The 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Jun 28, 2009-Jul 01, 2009 Paris, France. You can view more information about this proceeding and all of ACMs other published conference proceedings from the ACM Digital Library: <http://www.acm.org/dl>.

This book explores a proactive and domain-driven method to classification tasks. This novel proactive approach to data mining not only induces a model for predicting or explaining a phenomenon, but also utilizes specific problem/domain knowledge to suggest specific actions to achieve optimal changes in the value of the target attribute. In particular, the authors suggest a specific implementation of the domain-driven proactive approach for classification trees. The book centers on the core idea of moving observations from one branch of the tree to another. It introduces a novel splitting criterion for decision trees, termed maximal-utility, which maximizes the potential for enhancing profitability in the output tree. Two real-world case studies, one of a leading wireless operator and the other of a major security company, are also included and demonstrate how applying the proactive approach to classification tasks can solve business problems. Proactive Data Mining with Decision Trees is intended for researchers, practitioners and advanced-level students.

Big data: It's unstructured, it's coming at you fast, and there's lots of it. In fact, the majority of big data is text-oriented, thanks to the proliferation of online sources such as blogs, emails, and social media. However, having big data means little if you can't leverage it with analytics. Now you can explore the large volumes of unstructured text data that your organization has collected with Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS. This hands-on guide to text analytics using SAS provides detailed, step-by-step instructions and explanations on how to mine your text data for valuable insight. Through its comprehensive approach, you'll learn not just how to analyze your data, but how to collect, cleanse, organize, categorize, explore, and interpret it as well. Text Mining and Analysis also features an extensive set of case studies, so you can see examples of how the applications work with real-world data from a variety of industries. Text analytics enables you to gain insights about your customers' behaviors and sentiments. Leverage your organization's text data, and use those insights for making better business decisions with Text Mining and Analysis.

Shows how marketing research and data mining techniques will boost return on investment.

Data Mining with R Learning with Case Studies, Second Edition CRC Press

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

Big Data is a growing business trend, but there little advice available on how to use it practically. Written by a data mining expert with over 30 years of experience, this book uses case studies to help marketers, brand managers and IT professionals understand how to capture and measure data for marketing purposes.

Embrace machine learning approaches and Python to enable automatic rendering of rich insights and solve business problems. The book uses a hands-on case study-based approach to crack real-world applications to which machine learning concepts can be applied. These smarter machines will enable your business processes to achieve efficiencies on minimal time and resources. Python Machine Learning Case Studies takes you through the steps to improve business processes and determine the pivotal points that frame strategies. You'll see machine learning techniques that you can use to support your products and services. Moreover you'll learn the pros and cons of each of the machine learning concepts to help you decide which one best suits your needs. By taking a step-by-step approach to coding in Python you'll be able to understand the rationale behind model selection and decisions within the machine learning process. The book is equipped with practical examples along with code snippets to ensure that you understand the data science approach to solving real-world problems. What You Will Learn Gain insights into machine learning concepts Work on real-world applications of machine learning Learn concepts of model selection and optimization Get a hands-on overview of Python from a machine learning point of view Who This Book Is For Data scientists, data analysts, artificial intelligence engineers, big data enthusiasts, computer scientists, computer sciences students, and capital market analysts.

Data Analytics Applied to the Mining Industry describes the key challenges facing the mining sector as it transforms into a digital industry able to fully exploit process automation, remote operation centers, autonomous equipment and the opportunities offered by the industrial internet of things. It provides guidelines on how data needs to be collected, stored and managed to enable the different advanced data analytics methods to be applied effectively in practice, through use of case studies, and worked examples. Aimed at graduate students, researchers, and professionals in the industry of mining engineering, this book: Explains how to implement advanced data analytics through case studies and examples in mining engineering Provides approaches and methods to improve data-driven decision making Explains a concise overview of the state of the art for Mining Executives and Managers Highlights and describes critical opportunity areas for mining optimization Brings experience and learning in digital transformation from adjacent sectors

This book is the first technical guide to provide a complete, generalized road map for developing data-mining applications, together with advice on performing these large-scale, open-ended analyses for real-world data warehouses.

What does the Web look like? How can we find patterns, communities, outliers, in a social network? Which are the most central nodes in a network? These are the questions that motivate this work. Networks and graphs appear in many diverse settings, for example in social networks, computer-communication networks (intrusion detection, traffic management), protein-protein interaction networks in biology, document-text bipartite graphs in text retrieval, person-account graphs in financial fraud detection, and others. In this work, first we list several surprising patterns that real graphs tend to follow. Then we give a detailed list of generators that try to mirror these patterns. Generators are important, because they can help with "what if" scenarios, extrapolations, and anonymization. Then we provide a list of powerful tools for graph analysis, and specifically spectral methods (Singular Value Decomposition (SVD)), tensors, and case studies like the famous "pageRank" algorithm and the "HITS" algorithm for ranking web search results. Finally, we conclude with a survey of tools and observations from related fields like sociology, which provide complementary viewpoints. Table of Contents: Introduction / Patterns in Static Graphs / Patterns in Evolving Graphs / Patterns in Weighted Graphs / Discussion: The Structure of Specific Graphs / Discussion: Power Laws and Deviations / Summary of Patterns / Graph Generators / Preferential Attachment and Variants / Incorporating Geographical Information / The RMat / Graph Generation by Kronecker Multiplication / Summary and Practitioner's Guide / SVD, Random Walks, and Tensors / Tensors / Community Detection / Influence/Virus Propagation and Immunization / Case Studies / Social Networks / Other Related Work / Conclusions

Data Mining Applications with R is a great resource for researchers and professionals to understand the wide use of R, a free software environment for statistical computing and graphics, in solving different problems in industry. R is widely used in leveraging data mining techniques across many different industries, including government, finance, insurance, medicine, scientific research and more. This book presents 15 different real-world case studies illustrating various techniques in rapidly growing areas. It is an ideal companion for data mining researchers in academia and industry looking for ways to turn this versatile software into a powerful analytic tool. R code, Data and color figures for the book are provided at the RDataMining.com website. Helps data miners to learn to use R in their specific area of work and see how R can apply in different industries Presents various case studies in real-world applications, which will help readers to apply the techniques in their work Provides code examples and sample data for readers to easily learn the techniques by running the code by themselves

The second edition of a comprehensive introduction to machine learning approaches used in predictive data analytics, covering both theory and practice. Machine learning is often used to build predictive models by extracting patterns from large datasets. These models are used in predictive data analytics applications including price prediction, risk assessment, predicting customer behavior, and document classification. This introductory textbook offers a detailed and focused treatment of the most important machine learning approaches used in predictive data analytics, covering both theoretical concepts and practical applications. Technical and mathematical material is augmented with explanatory worked examples, and case studies illustrate the application of these models in the broader business context. This second edition covers recent developments in machine learning, especially in a new chapter on deep learning, and two new chapters that go beyond predictive analytics to cover unsupervised learning and reinforcement learning.

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.

Data mining can be defined as the process of selection, exploration and modelling of large databases, in order to discover models and patterns. The increasing availability of data in the current information society has led to the need for valid tools for its modelling and analysis. Data mining and applied statistical methods are the appropriate tools to extract such knowledge from data. Applications occur in many different fields, including statistics, computer science, machine learning, economics, marketing and finance. This book is the first to describe applied data mining methods in a consistent statistical framework, and then show how they can be applied in practice. All the methods described are either computational, or of a statistical modelling nature. Complex probabilistic models and mathematical tools are not used, so the book is accessible to a wide audience of students and industry professionals. The second half of the book consists of nine case studies, taken from the author's own work in industry, that demonstrate how the methods described can be applied to real problems. Provides a solid introduction to applied data mining methods in a consistent statistical framework Includes coverage of classical, multivariate and Bayesian statistical methodology Includes many recent developments such as web mining, sequential Bayesian analysis and memory based reasoning Each statistical method described is illustrated with real life applications Features a number of detailed case studies based on applied projects within industry Incorporates discussion on software used in data mining, with particular emphasis on SAS Supported by a website featuring data sets, software and additional material Includes an extensive bibliography and pointers to further reading within the text Author has many years experience teaching introductory and multivariate statistics and data mining, and working on applied projects within industry A valuable resource for advanced undergraduate and graduate students of applied statistics, data mining, computer science and economics, as well as for professionals working in industry on projects involving large volumes of data - such as in marketing or financial risk management.

Effectively Access, Transform, Manipulate, Visualize, and Reason about Data and Computation Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving illustrates the details involved in solving real computational problems encountered in data analysis. It reveals the dynamic and iterative process by which data analysts approach a problem and reason about different ways of implementing solutions. The book's collection of projects, comprehensive sample solutions, and follow-up exercises encompass practical topics pertaining to data processing, including: Non-standard, complex data formats, such as robot logs and email messages Text processing and regular expressions Newer technologies, such as Web scraping, Web services, Keyhole Markup Language (KML), and Google Earth Statistical methods, such as classification trees, k-nearest neighbors, and naïve Bayes Visualization and exploratory data analysis Relational databases and Structured Query Language (SQL) Simulation Algorithm implementation Large data and efficiency Suitable for self-study or as supplementary reading in a statistical computing course, the book enables instructors to incorporate interesting problems into their courses so that students gain valuable experience and data science skills. Students learn how to acquire and work with unstructured or semistructured data as well as how to narrow down and carefully frame the questions of interest about the data. Blending computational details with statistical and data analysis concepts, this book provides readers with an understanding of how professional data scientists think about daily computational tasks. It will improve readers' computational reasoning of real-world data analyses.

The increasing availability of data in our current, information overloaded society has led to the need for valid tools for its modelling and analysis. Data mining and applied statistical methods are the appropriate tools to extract knowledge from such data. This book provides an accessible introduction to data mining methods in a consistent and application oriented statistical framework, using case studies drawn from real industry projects and highlighting the use of data mining methods in a variety of business applications. Introduces data mining methods and applications. Covers classical and Bayesian multivariate statistical methodology as well as machine learning and computational data mining methods. Includes many recent developments such as association and sequence rules, graphical Markov models, lifetime value modelling, credit risk, operational risk and web mining. Features detailed case studies based on applied projects within industry. Incorporates discussion of data mining software, with case studies analysed using R. Is accessible to anyone with a basic knowledge of statistics or data analysis. Includes an extensive bibliography and pointers to further reading within the text. Applied Data Mining for Business and Industry, 2nd edition is aimed at advanced undergraduate and graduate students of data mining, applied statistics, database management, computer science and economics. The case studies will provide guidance to professionals working in industry on projects involving large volumes of data, such as customer relationship management, web design, risk management, marketing, economics and finance.

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.

Big data: It's unstructured, it's coming at you fast, and there's lots of it. In fact, the majority of big data is text-oriented, thanks to the proliferation of online sources such as blogs, emails, and social media. However, having big data means little if you can't leverage it with analytics. Now you can explore the large volumes of unstructured text data that your organization has collected with *Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS*. This hands-on guide to text analytics using SAS provides detailed, step-by-step instructions and explanations on how to mine your text data for valuable insight. Through its comprehensive approach, you'll learn not just how to analyze your data, but how to collect, cleanse, organize, categorize, explore, and interpret it as well. *Text Mining and Analysis* also features an extensive set of case studies, so you can see examples of how the applications work with real-world data from a variety of industries. Text analytics enables you to gain insights about your customers' behaviors and sentiments. Leverage your organization's text data, and use those insights for making better business decisions with *Text Mining and Analysis*. This book is part of the SAS Press program.

[Copyright: c64ca9d05e64c7d33182292b9fdcf44](#)