

Data Mining Orange Documentation

This book constitutes the refereed proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases, PKDD 2004, held in Pisa, Italy, in September 2004 jointly with ECML 2004. The 39 revised full papers and 9 revised short papers presented together with abstracts of 5 invited talks were carefully reviewed and selected from 194 papers submitted to PKDD and 107 papers submitted to both, PKDD and ECML. The papers present a wealth of new results in knowledge discovery in databases and address all current issues in the area.

Data Science & Business Analytics explores the application of big data and business analytics by academics, researchers, industrial experts, policy makers and practitioners, helping the reader to understand how big data can be efficiently utilized in better managerial applications.

This book constitutes the refereed proceedings of the 15th International Conference on Text, Speech and Dialogue, TSD 2012, held in Brno, Czech Republic, in September 2012. The 82 papers presented together with 2 invited talks were carefully reviewed and selected from 173 submissions. The papers are organized in topical sections on corpora and language resources, speech recognition, tagging, classification and parsing of text and speech, speech and spoken language generation, semantic processing of text and speech, integrating applications of text and speech processing, machine translation, automatic dialogue systems, multimodal techniques and modeling.

A lot of digital ink has been spilled on "big data" over the past few years. Most of this surge owes its origin to the various types of unstructured data in the wild, among which the proliferation of text-heavy data is particularly overwhelming, attributed to the daily use of web documents, business reviews, news, social posts, etc., by so many people worldwide. A core challenge presents itself: How can one efficiently and effectively turn massive, unstructured text into structured representation so as to further lay the foundation for many other downstream text mining applications? In this book, we investigated one promising paradigm for representing unstructured text, that is, through automatically identifying high-quality phrases from innumerable documents. In contrast to a list of frequent n-grams without proper filtering, users are often more interested in results based on variable-length phrases with certain semantics such as scientific concepts, organizations, slogans, and so on. We propose new principles and powerful methodologies to achieve this goal, from the scenario where a user can provide meaningful guidance to a fully automated setting through distant learning. This book also introduces applications enabled by the mined phrases and points out some promising research directions.

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.

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

Amity University Dubai is organizing 2nd International conference on Computational, Automation and Knowledge Management on 19th 21st Jan 2021 This International Conference will bring together the scholars, scientists and industrialists from across the world to the wide spectrum of engineering fields to a common platform and achieve the following To present the ongoing researches in different fields and hence to foster research relations between the Universities and the industry Give participants a review of the latest and upcoming trends in the next few years Exposing the audience to the need for more development and research for innovation Provide the delegates to share their new ideas and the application experiences face to face

Some say we live in the Information Age; others, the Social Age; and still others, the Big Data Age. Regardless of what name we give it, we live in an age that generates monumental amounts of data-in all different kinds of formats. In business, and in our personal lives, we use smartphones and tablets, web sites and watches; with apps and interfaces to shop, learn, entertain and inform. Businesses increasingly use technology to interact with consumers to provide marketing, customer service, product information and more. All of this technological activity generates data, and we're increasingly good at gathering, storing and analyzing it. Data mining can help to identify interesting patterns and messages that exist in data, often hidden beneath the surface. In this modern age of information systems, it is easier than ever before to extract meaning from data. From classification to prediction, data mining can help. In Data Mining for the Masses, Third Edition, professor Matt North-a former risk analyst and software engineer at eBay-uses simple examples and clear explanations with free, powerful software tools to teach you the basics of data mining. In this Third Edition, implementations of these examples are offered in current versions of the RapidMiner software, and in the increasingly popular R Statistical Package. You've got more data than ever before and you know it's got value, if only you can figure out how to get to it. This book can show you how. Let's start digging!

Open Source Data Warehousing and Business Intelligence is an all-in-one reference for developing open source based data warehousing (DW) and business intelligence (BI) solutions that are business-centric, cross-customer viable, cross-functional, cross-technology based, and enterprise-wide. Considering the entire lifecycle of an open source DW &

This book constitutes the refereed proceedings of the 19th EPIA Conference on Artificial Intelligence, EPIA 2019, held in Funchal, Madeira, Portugal, in September 2019. The 119 revised full papers and 6 short papers presented were carefully reviewed and selected from a total of 252 submissions. The papers are organized in 18 tracks devoted to the following topics: AIEd - Artificial Intelligence in Education, AI4G - Artificial Intelligence for Games, AIoTA - Artificial Intelligence and IoT in Agriculture, AIL - Artificial Intelligence and Law, AIM - Artificial Intelligence in Medicine, AICPDES - Artificial Intelligence in Cyber-Physical and Distributed Embedded Systems, AIPES - Artificial Intelligence in Power and Energy Systems, AITS - Artificial Intelligence in Transportation Systems, ALEA - Artificial Life and Evolutionary Algorithms, AmlA - Ambient Intelligence and Affective Environments, BAAI - Business Applications of Artificial Intelligence, GAI- General AI, IROBOT - Intelligent Robotics, KDBI - Knowledge Discovery and Business Intelligence, KRR - Knowledge Representation and Reasoning, MASTA - Multi-Agent Systems: Theory and Applications, SSM - Social Simulation and Modelling, TeMA - Text Mining and Applications.

This manual has been designed and written with the purpose of introducing key concepts and areas of debate around the "creative economy", a valuable development opportunity that Latin America, the Caribbean and the world at large cannot afford to miss. The creative economy, which we call the "Orange Economy" in this book (you'll see why), encompasses the immense wealth of talent, intellectual property, interconnectedness, and, of course, cultural heritage of the Latin American and Caribbean region (and indeed, every region). At the end of this manual, you will have the knowledge base necessary to understand and explain what the Orange Economy is and why it is so important. You will also acquire the analytical tools needed to take better advantage of opportunities across the arts, heritage, media, and creative services.

Data Mining and Data Visualization focuses on dealing with large-scale data, a field commonly referred to as data mining. The book is divided into three sections. The first deals with an introduction to statistical aspects of data mining and machine learning and includes applications to text analysis, computer intrusion detection, and hiding of information in digital files. The second section focuses on a variety of statistical methodologies that have proven to be effective in data mining applications. These include clustering, classification, multivariate density estimation, tree-based methods, pattern recognition, outlier detection, genetic algorithms, and dimensionality reduction. The third section focuses on data visualization and covers issues of visualization of high-dimensional data, novel graphical techniques with a focus on human factors, interactive graphics, and data visualization using virtual reality. This book represents a thorough cross section of internationally renowned thinkers who are inventing methods for dealing with a new data paradigm. Distinguished contributors who are international experts in aspects of data mining Includes data mining approaches to non-numerical data mining including text data, Internet traffic data, and geographic data Highly topical discussions reflecting current thinking on contemporary technical issues, e.g. streaming data Discusses taxonomy of dataset sizes, computational complexity, and scalability usually ignored in most discussions Thorough discussion of data visualization issues blending statistical, human factors, and computational insights

Data Mining and Analytics provides a broad and interactive overview of a rapidly growing field. The exponentially increasing rate at which data is generated creates a corresponding need for professionals who can effectively handle its storage, analysis, and translation.

Get complete instructions for manipulating, processing, cleaning, and crunching datasets in Python. Updated for Python 3.6, the second edition of this hands-on guide is packed with practical case studies that show you how to solve a broad set of data analysis problems effectively. You'll learn the latest versions of pandas, NumPy, IPython, and Jupyter in the process. Written by Wes McKinney, the creator of the Python pandas project, this book is a practical, modern introduction to data science tools in Python. It's ideal for analysts new to Python and for Python programmers new to data science and scientific computing. Data files and related material are available on GitHub. Use the IPython shell and Jupyter notebook for exploratory computing Learn basic and advanced features in NumPy (Numerical Python) Get started with data analysis tools in the pandas library Use flexible tools to load, clean, transform, merge, and reshape data Create informative visualizations with matplotlib Apply the pandas groupby facility to slice, dice, and summarize datasets Analyze and manipulate regular and irregular time series data Learn how to solve real-world data analysis problems with thorough, detailed examples

Much of the data available today is unstructured and text-heavy, making it challenging for analysts to apply their usual data wrangling and visualization tools. With this practical book, you'll explore text-mining techniques with tidytext, a package that authors Julia Silge and David Robinson developed using the tidy principles behind R packages like ggraph and dplyr. You'll learn how tidytext and other tidy tools in R can make text analysis easier and more effective. The authors demonstrate how treating text as data frames enables you to manipulate, summarize, and visualize characteristics of text. You'll also learn how to integrate natural language processing (NLP) into effective workflows.

Practical code examples and data explorations will help you generate real insights from literature, news, and social media. Learn how to apply the tidy text format to NLP Use sentiment analysis to mine the emotional content of text Identify a document's most important terms with frequency measurements Explore relationships and connections between words with the ggraph and widyr packages Convert back and forth between R's tidy and non-tidy text formats Use topic modeling to classify document collections into natural groups Examine case studies that compare Twitter archives, dig into NASA metadata, and analyze thousands of Usenet messages

The next step in the information age is to gain insights from the deluge of data coming our way. Data mining provides a way of finding this insight, and Python is one of the most popular languages for data mining, providing both power and flexibility in analysis. This book teaches you to design and develop data mining applications using a variety of datasets, starting with basic classification and affinity analysis. Next, we move on to more complex data types including text, images, and graphs. In every chapter, we create

models that solve real-world problems. There is a rich and varied set of libraries available in Python for data mining. This book covers a large number, including the IPython Notebook, pandas, scikit-learn and NLTK. Each chapter of this book introduces you to new algorithms and techniques. By the end of the book, you will gain a large insight into using Python for data mining, with a good knowledge and understanding of the algorithms and implementations.

"This book provides an overall view of recent solutions for mining, and explores new patterns, offering theoretical frameworks and presenting challenges and possible solutions concerning pattern extractions, emphasizing research techniques and real-world applications. It portrays research applications in data models, methodologies for mining patterns, multi-relational and multidimensional pattern mining, fuzzy data mining, data streaming and incremental mining"--Provided by publisher.

Data mining is often referred to by real-time users and software solutions providers as knowledge discovery in databases (KDD). Good data mining practice for business intelligence (the art of turning raw software into meaningful information) is demonstrated by the many new techniques and developments in the conversion of fresh scientific discovery into widely accessible software solutions. This book has been written as an introduction to the main issues associated with the basics of machine learning and the algorithms used in data mining. Suitable for advanced undergraduates and their tutors at postgraduate level in a wide area of computer science and technology topics as well as researchers looking to adapt various algorithms for particular data mining tasks. A valuable addition to the libraries and bookshelves of the many companies who are using the principles of data mining (or KDD) to effectively deliver solid business and industry solutions. Provides an introduction to the main issues associated with the basics of machine learning and the algorithms used in data mining A valuable addition to the libraries and bookshelves of companies using the principles of data mining (or KDD) to effectively deliver solid business and industry solutions

This book presents Proceedings of the 2021 Intelligent Systems Conference which is a remarkable collection of chapters covering a wider range of topics in areas of intelligent systems and artificial intelligence and their applications to the real world. The conference attracted a total of 496 submissions from many academic pioneering researchers, scientists, industrial engineers, and students from all around the world. These submissions underwent a double-blind peer-review process. Of the total submissions, 180 submissions have been selected to be included in these proceedings. As we witness exponential growth of computational intelligence in several directions and use of intelligent systems in everyday applications, this book is an ideal resource for reporting latest innovations and future of AI. The chapters include theory and application on all aspects of artificial intelligence, from classical to intelligent scope. We hope that readers find the book interesting and valuable; it provides the state-of-the-art intelligent methods and techniques for solving real-world problems along with a vision of the future research. .

This streamlined "essential" version of the Molecular Pathology (2009) textbook extracts key information, illustrations and photographs from the main textbook in the same number and organization of chapters. It is aimed at teaching students in courses where the full textbook is not needed, but the concepts included are desirable (such as graduate students in allied health programs or undergraduates). It is also aimed at students who are enrolled in courses that primarily use a traditional pathology textbook, but need the complementary concepts of molecular pathology (such as medical students). Further, the textbook will be valuable for pathology residents and other postdoctoral fellows who desire to advance their understanding of molecular mechanisms of disease beyond what they learned in medical/graduate school. Offers an essential introduction to molecular genetics and the "molecular" aspects of human disease Teaches from the perspective of "integrative systems biology," which encompasses the intersection of all molecular aspects of biology, as applied to understanding human disease In-depth presentation of the principles and practice of molecular pathology: molecular pathogenesis, molecular mechanisms of disease, and how the molecular pathogenesis of disease parallels the evolution of the disease using histopathology. "Traditional" pathology section provides state-of-the-art information on the major forms of disease, their pathologies, and the molecular mechanisms that drive these diseases. Explains the practice of "molecular medicine" and the translational aspects of molecular pathology: molecular diagnostics, molecular assessment, and personalized medicine Each chapter ends with Key Summary Points and Suggested Readings

With continuous advancements and an increase in user popularity, data mining technologies serve as an invaluable resource for researchers across a wide range of disciplines in the humanities and social sciences. In this comprehensive guide, author and research scientist Kalev Leetaru introduces the approaches, strategies, and methodologies of current data mining techniques, offering insights for new and experienced users alike. Designed as an instructive reference to computer-based analysis approaches, each chapter of this resource explains a set of core concepts and analytical data mining strategies, along with detailed examples and steps relating to current data mining practices. Every technique is considered with regard to context, theory of operation and methodological concerns, and focuses on the capabilities and strengths relating to these technologies. In addressing critical methodologies and approaches to automated analytical techniques, this work provides an essential overview to a broad innovative field.

This book examines the impact of digital innovation on organizations. It reveals how the digital revolution is redefining traditional levels of analysis while at the same time blurring the internal and external boundaries of the organizational environment. It presents a collection of research papers that examine the interaction between Information and Communication Technology (ICT) and behavior from a threefold perspective: First, they analyze individual behavior in terms of specific organizational practices like learning, collaboration and knowledge transfer, as well as the use of ICT within the organization. Second, they explore the dynamics at work on the border between the internal and the external environments by analyzing the organizational impact of ICT usage outside the company, as can be seen in employer branding, consumer behavior and organizational image. Third, they investigate how ICT is being adopted to help face societal challenges outside the company like waste and pollution, smart cities, and e-government. The diversity of views offered makes this book particularly relevant for users, companies, scientists and governments. The content is based on a selection of the best papers – original double-blind peer reviewed contributions –presented at the annual conference of the Italian chapter of AIS, which took place in Rome, Italy, in October 2015, or in other reputable international peer-refereed Information Systems conferences.

Big Data is a new field, with many technological challenges to be understood in order to use it to its full potential. These challenges arise at all stages of working with Big Data, beginning with data generation and acquisition. The storage and management phase presents two critical challenges: infrastructure, for storage and transportation, and conceptual models. Finally, to extract meaning from Big Data requires complex analysis. Here the authors propose using metaheuristics as a solution to these challenges; they are first able to deal with large size problems and

secondly flexible and therefore easily adaptable to different types of data and different contexts. The use of metaheuristics to overcome some of these data mining challenges is introduced and justified in the first part of the book, alongside a specific protocol for the performance evaluation of algorithms. An introduction to metaheuristics follows. The second part of the book details a number of data mining tasks, including clustering, association rules, supervised classification and feature selection, before explaining how metaheuristics can be used to deal with them. This book is designed to be self-contained, so that readers can understand all of the concepts discussed within it, and to provide an overview of recent applications of metaheuristics to knowledge discovery problems in the context of Big Data.

As telescopes, detectors, and computers grow ever more powerful, the volume of data at the disposal of astronomers and astrophysicists will enter the petabyte domain, providing accurate measurements for billions of celestial objects. This book provides a comprehensive and accessible introduction to the cutting-edge statistical methods needed to efficiently analyze complex data sets from astronomical surveys such as the Panoramic Survey Telescope and Rapid Response System, the Dark Energy Survey, and the upcoming Large Synoptic Survey Telescope. It serves as a practical handbook for graduate students and advanced undergraduates in physics and astronomy, and as an indispensable reference for researchers. Statistics, Data Mining, and Machine Learning in Astronomy presents a wealth of practical analysis problems, evaluates techniques for solving them, and explains how to use various approaches for different types and sizes of data sets. For all applications described in the book, Python code and example data sets are provided. The supporting data sets have been carefully selected from contemporary astronomical surveys (for example, the Sloan Digital Sky Survey) and are easy to download and use. The accompanying Python code is publicly available, well documented, and follows uniform coding standards. Together, the data sets and code enable readers to reproduce all the figures and examples, evaluate the methods, and adapt them to their own fields of interest. Describes the most useful statistical and data-mining methods for extracting knowledge from huge and complex astronomical data sets Features real-world data sets from contemporary astronomical surveys Uses a freely available Python codebase throughout Ideal for students and working astronomers

The progress of data mining technology and large public popularity establish a need for a comprehensive text on the subject. The series of books entitled by "Data Mining" address the need by presenting in-depth description of novel mining algorithms and many useful applications. In addition to understanding each section deeply, the two books present useful hints and strategies to solving problems in the following chapters. The contributing authors have highlighted many future research directions that will foster multi-disciplinary collaborations and hence will lead to significant development in the field of data mining.

Integrates social media, social network analysis, and data mining to provide an understanding of the potentials of social media mining.

This book A Beginner's Guide to Learning Analytics is designed to meet modern educational trends' needs. It is addressed to readers who have no prior knowledge of learning analytics and functions as an introductory text to learning analytics for those who want to do more with evaluation/assessment in their organizations. The book is useful to all who need to evaluate their learning and teaching strategies. It aims to bring greater efficiency and deeper engagement to individual students, learning communities, and educators. Covered here are the key concepts linked to learning analytics for researchers and practitioners interested in learning analytics. This book helps those who want to apply analytics to learning and development programs and helps educational institutions to identify learners who require support and provide a more personalized learning experience. Like chapters show diverse uses of learning analytics to enhance student and faculty performance. It presents a coherent framework for the effective translation of learning analytics research for educational practice to its practical application in different educational domains. This book provides educators and researchers with the tools and frameworks to effectively make sense of and use data and analytics in their everyday practice. This book will be a valuable addition to researchers' bookshelves.

This book constitutes the refereed proceedings of the Third International Workshop on Mining Complex Data, MCD 2007, held in Warsaw, Poland, in September 2007, co-located with ECML and PKDD 2007. The 20 revised full papers presented were carefully reviewed and selected; they present original results on knowledge discovery from complex data. In contrast to the typical tabular data, complex data can consist of heterogeneous data types, can come from different sources, or live in high dimensional spaces. All these specificities call for new data mining strategies.

SAS Visual Data Mining and Machine Learning, powered by SAS Viya, means that users of all skill levels can visually explore data on their own while drawing on powerful in-memory technologies for faster analytic computations and discoveries. You can manually program with custom code or use the features in SAS Studio, Model Studio, and SAS Visual Analytics to automate your data manipulation and modeling. These programs offer a flexible, easy-to-use, self-service environment that can scale on an enterprise-wide level. In this book, we will explore some of the many features of SAS Visual Data Mining and Machine Learning including: programming in the Python interface; new, advanced data mining and machine learning procedures; pipeline building in Model Studio, and model building and comparison in SAS Visual Analytics.

Molecular Pathology: The Molecular Basis of Human Disease provides a current and comprehensive view of the molecular basis and mechanisms of human disease. Combining accepted principles with broader theoretical concepts and with contributions from a group of experts, the book looks into disease processes in the context of traditional pathology and their implications for translational molecular medicine. It also discusses concepts in molecular biology and genetics, recent scientific and technological advances in modern pathology, the concept of "molecular pathogenesis" of disease, and how disease evolves from normal cells and tissues due to perturbations in molecular pathways. The book describes the integration of molecular and cellular pathogenesis using a bioinformatics approach and a systems biology approach to disease pathogenesis. It also discusses current and future strategies in molecular diagnosis of human disease, and the impact of molecular diagnosis on treatment decisions and the practice of personalized medicine. This book is a valuable resource for students, biomedical researchers, practicing physician-scientists who undertake disease-related basic science and translational research, and pathology residents and other postdoctoral fellows. * Exam Master® web site will host "Self-assessment" questions that students can use to study for the molecular section of the board exam * Companion Web Site – will host a complete set of PowerPoint slides: to include images from the book and additional images for teaching; course materials; lecture materials * Teaches from the perspective of "integrative systems biology, which encompasses the

intersection of all molecular aspects of biology, as applied to understanding human disease * Outlines the principles and practice of molecular pathology * Explains the practice of “molecular medicine and the translational aspects of molecular pathology

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 constitutes the refereed proceedings of the Second International Conference on Data Mining and Big Data, DMBD 2017, held in Fukuoka, Japan, in July/August 2017. The 53 papers presented in this volume were carefully reviewed and selected from 96 submissions. They were organized in topical sections named: association analysis; clustering; prediction; classification; schedule and sequence analysis; big data; data analysis; data mining; text mining; deep learning; high performance computing; knowledge base and its framework; and fuzzy control.

This book discusses the evolution of security and privacy issues and brings related technological tools, techniques, and solutions into one single source. The book will take readers on a journey to understanding the security issues and possible solutions involving various threats, attacks, and defense mechanisms, which include IoT, cloud computing, Big Data, lightweight cryptography for blockchain, and data-intensive techniques, and how it can be applied to various applications for general and specific use. Graduate and postgraduate students, researchers, and those working in this industry will find this book easy to understand and use for security applications and privacy issues.

Knowledge Discovery in Databases: PKDD 2004 8th European Conference on Principles and Practice of Knowledge Discovery in Databases, Pisa, Italy, September 20-24, 2004, Proceedings Springer Science & Business Media

Artificial Intelligence and Innovations (AIAI) will interest researchers, IT professionals and consultants by examining technologies and applications of demonstrable value. The conference focused on profitable intelligent systems and technologies. AIAI focuses on real world applications; therefore authors should highlight the benefits of AI technology for industry and services. Novel approaches solving business and industrial problems, using AI, will emerge from this conference.

What is text mining, and how can it be used? What relevance do these methods have to everyday work in information science and the digital humanities? How does one develop competences in text mining? Working with Text provides a series of cross-disciplinary perspectives on text mining and its applications. As text mining raises legal and ethical issues, the legal background of text mining and the responsibilities of the engineer are discussed in this book. Chapters provide an introduction to the use of the popular GATE text mining package with data drawn from social media, the use of text mining to support semantic search, the development of an authority system to support content tagging, and recent techniques in automatic language evaluation. Focused studies describe text mining on historical texts, automated indexing using constrained vocabularies, and the use of natural language processing to explore the climate science literature. Interviews are included that offer a glimpse into the real-life experience of working within commercial and academic text mining. Introduces text analysis and text mining tools Provides a comprehensive overview of costs and benefits Introduces the topic, making it accessible to a general audience in a variety of fields, including examples from biology, chemistry, sociology, and criminology

This volume showcases the presentations and discussions delivered at the 2018 POMS International Conference in Rio. Through a collection of selected papers, it is possible to review the impact and application of operations management for social good, with contributions across a wide range of topics, including: humanitarian operations and crisis management, healthcare operations management, sustainable operations, artificial intelligence and data analytics in operations, product innovation and technology in operations management, marketing and operations management, service operations and servitization, logistics and supply chain management, resilience and risk in operations, defense, and tourism among other emerging Operations Management issues. The Production and Operations Management Society (POMS) is one of the most important and influential societies in the subject of Production Engineering and, as an international professional and academic organization, represents the interests of professionals and academics in production management and operations around the world.

During the past decade there has been an explosion in computation and information technology. With it have come vast amounts of data in a variety of fields such as medicine, biology, finance, and marketing. The challenge of understanding these data has led to the development of new tools in the field of statistics, and spawned new areas such as data mining, machine learning, and bioinformatics. Many of these tools have common underpinnings but are often expressed with different terminology. This book describes the important ideas in these areas in a common conceptual framework. While the approach is statistical, the emphasis is on concepts rather than mathematics. Many examples are given, with a liberal use of color graphics. It should be a valuable resource for statisticians and anyone interested in data mining in science or industry. The book's coverage is broad, from supervised learning (prediction) to unsupervised learning. The many topics include neural networks, support vector machines, classification trees and boosting--the first comprehensive treatment of this topic in any book. This major new edition features many topics not covered in the original, including graphical models, random forests, ensemble methods, least angle regression & path algorithms for the lasso, non-negative matrix factorization, and spectral clustering. There is also a chapter on methods for “wide” data (p bigger than n), including multiple testing and false discovery rates. Trevor Hastie, Robert Tibshirani, and Jerome Friedman are professors of statistics at Stanford University. They are prominent researchers in this area: Hastie and Tibshirani developed generalized additive models and wrote a popular book of that title. Hastie co-developed much of the statistical modeling software and environment in R/S-PLUS and invented principal curves and surfaces. Tibshirani proposed the lasso and is co-author of the very successful An Introduction to the Bootstrap. Friedman is the co-inventor of many data-mining tools including CART, MARS, projection pursuit and gradient boosting.

Data Preparation for Data Mining addresses an issue unfortunately ignored by most authorities on data mining: data preparation. Thanks largely to its perceived difficulty, data preparation has traditionally taken a backseat to the more alluring question of how best to extract meaningful knowledge. But without adequate preparation of your data, the return on the resources invested in mining is certain to be disappointing. Dorian Pyle corrects this imbalance. A twenty-five-year veteran of what has become the data mining industry, Pyle shares his own successful data preparation methodology, offering both a

conceptual overview for managers and complete technical details for IT professionals. Apply his techniques and watch your mining efforts pay off-in the form of improved performance, reduced distortion, and more valuable results. On the enclosed CD-ROM, you'll find a suite of programs as C source code and compiled into a command-line-driven toolkit. This code illustrates how the author's techniques can be applied to arrive at an automated preparation solution that works for you. Also included are demonstration versions of three commercial products that help with data preparation, along with sample data with which you can practice and experiment. * Offers in-depth coverage of an essential but largely ignored subject. * Goes far beyond theory, leading you-step by step-through the author's own data preparation techniques. * Provides practical illustrations of the author's methodology using realistic sample data sets. * Includes algorithms you can apply directly to your own project, along with instructions for understanding when automation is possible and when greater intervention is required. * Explains how to identify and correct data problems that may be present in your application. * Prepares miners, helping them head into preparation with a better understanding of data sets and their limitations.

[Copyright: de4e2a0130eb047f97ee895d5d672ac0](#)