

Optimization Based Data Mining Theory And Applications Advanced Information And Knowledge Processing

Introduction to Algorithms for Data Mining and Machine Learning introduces the essential ideas behind all key algorithms and techniques for data mining and machine learning, along with optimization techniques. Its strong formal mathematical approach, well selected examples, and practical software recommendations help readers develop confidence in their data modeling skills so they can process and interpret data for classification, clustering, curve-fitting and predictions. Masterfully balancing theory and practice, it is especially useful for those who need relevant, well explained, but not rigorous (proofs based) background theory and clear guidelines for working with big data. Presents an informal, theorem-free approach with concise, compact coverage of all fundamental topics Includes worked examples that help users increase confidence in their understanding of key algorithms, thus encouraging self-study Provides algorithms and techniques that can be implemented in any programming language, with each chapter including notes about relevant software packages

ICDM / MLDM Medaille (limited edition) Meissner Porcellan, the "White Gold" of King August the Strongest of Saxonia ICDM 2008 was the eighth event of the Industrial Conference on Data Mining held in Leipzig (www.data-mining-forum.de). For this edition the Program Committee received 116 submissions from 20 countries. After the peer-review process, we accepted 36 high-quality papers for oral presentation, which are included in these proceedings. The topics range from aspects of classification and prediction, clustering, Web mining, data mining in medicine, applications of data mining, time series and frequent pattern mining, and association rule mining. Thirteen papers were selected for poster presentations that are published in the ICDM Poster Proceeding Volume. In conjunction with ICDM there were three workshops focusing on special hot application-oriented topics in data mining. The workshop Data Mining in Life Science DMLS 2008 was held the third time this year and the workshop Data Mining in Marketing DMM 2008 ran for the second time this year. Additionally, we introduced an International Workshop on Case-Based Reasoning for Multimedia Data CBR-MD.

Business intelligence is a broad category of applications and technologies for gathering, providing access to, and analyzing data for the purpose of helping enterprise users make better business decisions. The term implies having a comprehensive knowledge of all factors that affect a business, such as customers, competitors, business partners, economic environment, and internal operations, therefore enabling optimal decisions to be made. Business Intelligence provides readers with an introduction and practical guide to the mathematical models and analysis methodologies vital to business intelligence. This book: Combines detailed coverage with a practical guide to the mathematical models and analysis methodologies of business intelligence. Covers all the hot topics such as data warehousing, data mining and its applications, machine learning, classification, supply optimization models, decision support systems, and analytical methods for performance evaluation. Is made accessible to readers through the careful definition and introduction of each concept, followed by the extensive use of examples and numerous real-life case studies. Explains how to utilise mathematical models and analysis models to make effective and good quality business decisions. This book is aimed at postgraduate students following data analysis and data mining courses. Researchers looking for a systematic and broad coverage of topics in operations research and mathematical models for decision-making will find this an invaluable guide.

A guide to modern optimization applications and techniques in newly emerging areas spanning optimization, data science, machine intelligence, engineering, and computer sciences Optimization Techniques and Applications with Examples introduces the fundamentals of all the commonly used techniques in optimization that encompass the broadness and diversity of the methods (traditional and new) and algorithms. The author—a noted expert in the field—covers a wide range of topics including mathematical foundations, optimization formulation, optimality conditions, algorithmic complexity, linear programming, convex optimization, and integer programming. In addition, the book discusses artificial neural network, clustering and classifications, constraint-handling, queueing theory, support vector machine and multi-objective optimization, evolutionary computation, nature-inspired algorithms and many other topics. Designed as a practical resource, all topics are explained in detail with step-by-step examples to show how each method works. The book's exercises test the acquired knowledge that can be potentially applied to real problem solving. By taking an informal approach to the subject, the author helps readers to rapidly acquire the basic knowledge in optimization, operational research, and applied data mining. This important resource: Offers an accessible and state-of-the-art introduction to the main optimization techniques Contains both traditional optimization techniques and the most current algorithms and swarm intelligence-based techniques Presents a balance of theory, algorithms, and implementation Includes more than 100 worked examples with step-by-step explanations Written for upper undergraduates and graduates in a standard course on optimization, operations research and data mining, Optimization Techniques and Applications with Examples is a highly accessible guide to understanding the fundamentals of all the commonly used techniques in optimization.

Optimization Based Data Mining: Theory and Applications Springer Science & Business Media

This three volume set (CCIS 853-855) constitutes the proceedings of the 17th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2017, held in Cádiz, Spain, in June 2018. The 193 revised full papers were carefully reviewed and selected from 383 submissions. The papers are organized in topical sections on advances on explainable artificial intelligence; aggregation operators, fuzzy metrics and applications; belief function theory and its applications; current techniques to model, process and describe time series; discrete models and computational intelligence; formal concept analysis and uncertainty; fuzzy implication functions; fuzzy logic and artificial intelligence problems; fuzzy mathematical analysis and applications; fuzzy methods in data mining and knowledge discovery; fuzzy transforms: theory and applications to data analysis and image processing; imprecise probabilities: foundations and applications; mathematical fuzzy logic, mathematical morphology; measures of comparison and entropies for fuzzy sets and their extensions; new trends in data aggregation; pre-aggregation functions and generalized forms of monotonicity; rough and fuzzy similarity modelling tools; soft computing for decision making in uncertainty; soft computing in information retrieval and sentiment analysis; tri-partitions and uncertainty; decision making modeling and applications; logical methods in mining knowledge from big data; metaheuristics and machine learning; optimization models for modern analytics; uncertainty in medicine; uncertainty in Video/Image Processing (UVIP).

This book is mainly about an innovative and fundamental method called "intelligent knowledge" to bridge the gap between data mining and knowledge management, two important fields

recognized by the information technology (IT) community and business analytics (BA) community respectively. The book includes definitions of the “first-order” analytic process, “second-order” analytic process and intelligent knowledge, which have not formally been addressed by either data mining or knowledge management. Based on these concepts, which are especially important in connection with the current Big Data movement, the book describes a framework of domain-driven intelligent knowledge discovery. To illustrate its technical advantages for large-scale data, the book employs established approaches, such as Multiple Criteria Programming, Support Vector Machine and Decision Tree to identify intelligent knowledge incorporated with human knowledge. The book further shows its applicability by means of real-life data analyses in the contexts of internet business and traditional Chinese medicines.

This book constitutes the refereed proceedings of the 6th KES International Conference on Agent and Multi-Agent Systems, KES-AMSTA 2012, held in Dubrovnik, Croatia, in June 2012. The conference attracted a substantial number of researchers and practitioners from all over the world who submitted their papers for ten main tracks covering the methodology and applications of agent and multi-agent systems, one workshop (TRUMAS 2012) and five special sessions on specific topics within the field. The 66 revised papers presented were carefully reviewed and selected for inclusion in the book. The papers are organized in topical sections on virtual organizations, knowledge and learning agents, intelligent workflow, cloud computing and intelligent systems, self-organization, ICT-based alternative and augmentative communication, multi-agent systems, mental and holonic models, assessment methodologies in multi-agent and other paradigms, business processing agents, Trumas 2012 (first international workshop), conversational agents and agent teams, digital economy, and multi-agent systems in distributed environments.

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.

Artificial Intelligence in Data Mining: Theories and Applications offers a comprehensive introduction to data mining theories, relevant AI techniques, and their many real-world applications. This book is written by experienced engineers for engineers, biomedical engineers, and researchers in neural networks, as well as computer scientists with an interest in the area. Provides coverage of the fundamentals of Artificial Intelligence as applied to data mining, including computational intelligence and unsupervised learning methods for data clustering Presents coverage of key topics such as heuristic methods for data clustering, deep learning methods for data classification, and neural networks Includes case studies and real-world applications of AI techniques in data mining, for improved outcomes in clinical diagnosis, satellite data extraction, agriculture, security and defense

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.

" This volume, edited as a Festschrift in honor of Prof. Milan Zeleny, reflects and emulates his unmistakable legacy: the essential multidimensionality of human and social affairs. There are many levels of this multidimensionality presented in this volume: 1. Multidisciplinarity of contributed papers 2. Multinationality of their authors, extending even to the editors and the publisher and 3. Multicultural and multilevel exposition, ranging from empirical studies to philosophical foundations. Generally, these papers can be divided into three parts: Multiple Criteria Decision Making; Social and Human System Management; and Information, Knowledge and Wisdom Management. It is the recognition of multidimensionality in decision making, economics, optimization, systems, cybernetics and the pursuit of knowledge that bear the stamp of specific Zelenys contributions. His life-long dedication to multidimensionality has produced an ultimate multidimensional being, living in academic multiverse, functioning in a boundaryless world of all continents,

cultures and countries. This book is as diverse and as multidimensional as the man and his work. "

"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.

This book constitutes the refereed proceedings of the 8th International Conference on Electronic Government and the Information Systems Perspective, EGOVIS 2019, held in Linz, Austria, in August 2019. The 17 full papers presented were carefully reviewed and selected from 25 submissions. The papers are organized in the following topical sections: open data and open innovation; data-driven approaches in e-government; e-government cases – data and knowledge management; e-government theoretical background; and digitalization and transparency.

Coverage in this proceedings volume includes data mining and knowledge discovery, wireless, sensor networks and grid, XML and query processing and optimization, security, information extraction, semantic Web and Web applications, and workflow and middleware.

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

The papers in this volume represent the proceedings of the 7th Industrial Conference on Data Mining. They are organized into topical sections on aspects of classification and prediction, clustering, web mining, data mining in medicine, applications of data mining, time series and frequent pattern mining, and association rule mining. Readers gain new insights into theories underlying data mining and discover state-of-the-technology applications.

Dynamic programming is an efficient technique for solving optimization problems. It is based on breaking the initial problem down into simpler ones and solving these sub-problems, beginning with the simplest ones. A conventional dynamic programming algorithm returns an optimal object from a given set of objects. This book develops extensions of dynamic programming, enabling us to (i) describe the set of objects under consideration; (ii) perform a multi-stage optimization of objects relative to different criteria; (iii) count the number of optimal objects; (iv) find the set of Pareto optimal points for bi-criteria optimization problems; and (v) to study relationships between two criteria. It considers various applications, including optimization of decision trees and decision rule systems as algorithms for problem solving, as ways for knowledge representation, and as classifiers; optimization of element partition trees for rectangular meshes, which are used in finite element methods for solving PDEs; and multi-stage optimization for such classic combinatorial optimization problems as matrix chain multiplication, binary search trees, global sequence alignment, and shortest paths. The results presented are useful for researchers in combinatorial optimization, data mining, knowledge discovery, machine learning, and finite element methods, especially those working in rough set theory, test theory, logical analysis of data, and PDE solvers. This book can be used as the basis for graduate courses.

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

For organizations operating in a modern business environment, adopting the latest information technologies (IT) is of paramount importance. Organizational decision makers are increasingly interested in IT acquisition, constantly seeking the most advanced solutions in order to give their constituents a distinct competitive advantage. *Managing Enterprise Information Technology Acquisitions: Assessing Organizational Preparedness* provides leaders and innovators with research and strategies to make the most of their options involving IT and organizational management approaches. This book will serve as a critical resource for leaders, managers, strategists, and other industry professionals who must be prepared to meet the constant changes in the field of information technologies in order to effectively guide their organizations and achieve their respective goals.

This book constitutes the refereed proceedings of the 15th Industrial Conference on Advances in Data Mining, ICDM 2015, held in Hamburg, Germany, in July 2015. The 16 revised full papers presented were carefully reviewed and selected from numerous submissions. The topics range from theoretical aspects of data mining to applications of data mining, such as in multimedia data, in marketing, in medicine and agriculture, and in process control, industry and society.

Optimization techniques have been widely adopted to implement various data mining algorithms. In addition to well-known Support Vector Machines (SVMs) (which are based on quadratic programming), different versions of Multiple Criteria Programming (MCP) have been extensively used in data separations. Since optimization based data mining methods differ from statistics, decision tree induction, and neural networks, their theoretical inspiration has attracted many researchers who are interested in algorithm development of data mining. *Optimization based Data Mining: Theory and Applications*, mainly focuses on MCP and SVM especially their recent theoretical progress and real-life applications in various fields. These include finance, web services, bio-informatics and petroleum engineering, which has triggered the interest of practitioners who look for new methods to improve the results of data mining for knowledge discovery. Most of the material in this book is directly from the research and application activities that the authors' research group has conducted over the last ten years. Aimed at practitioners and graduates who have a fundamental knowledge in data mining, it demonstrates the basic concepts

and foundations on how to use optimization techniques to deal with data mining problems.

Focused on the mathematical foundations of social media analysis, Graph-Based Social Media Analysis provides a comprehensive introduction to the use of graph analysis in the study of social and digital media. It addresses an important scientific and technological challenge, namely the confluence of graph analysis and network theory with linear algebra, digital media, machine learning, big data analysis, and signal processing. Supplying an overview of graph-based social media analysis, the book provides readers with a clear understanding of social media structure. It uses graph theory, particularly the algebraic description and analysis of graphs, in social media studies. The book emphasizes the big data aspects of social and digital media. It presents various approaches to storing vast amounts of data online and retrieving that data in real-time. It demystifies complex social media phenomena, such as information diffusion, marketing and recommendation systems in social media, and evolving systems. It also covers emerging trends, such as big data analysis and social media evolution. Describing how to conduct proper analysis of the social and digital media markets, the book provides insights into processing, storing, and visualizing big social media data and social graphs. It includes coverage of graphs in social and digital media, graph and hyper-graph fundamentals, mathematical foundations coming from linear algebra, algebraic graph analysis, graph clustering, community detection, graph matching, web search based on ranking, label propagation and diffusion in social media, graph-based pattern recognition and machine learning, graph-based pattern classification and dimensionality reduction, and much more. This book is an ideal reference for scientists and engineers working in social media and digital media production and distribution. It is also suitable for use as a textbook in undergraduate or graduate courses on digital media, social media, or social networks.

This dissertation addresses important problems in decision theory and data mining. In particular, we focus on problems of the form: Each of several information sources provides evaluations or measurements of the objects in a universal set, and the objective is to aggregate these, possibly conflicting, evaluations into a consensus evaluation of each object in the universal set. In addition, we concentrate on the scenario where each source provides evaluations of only a strict subset of the objects; that is, each source provides an incomplete evaluation. In order to define the consensus evaluation from a given set of incomplete evaluations, two distances are developed: the first is a distance between incomplete rankings (ordinal evaluations) and the second is a distance between incomplete ratings (cardinal evaluations). These two distances generalize Kemeny and Snell's distance between complete rankings and Cook and Kress' distance between complete ratings, respectively. Specifically, we introduce a set of natural axioms that must be satisfied by a distance between two incomplete rankings (ratings) and prove the uniqueness and existence of a distance satisfying such axioms. Given a set of incomplete rankings (ratings), the consensus ranking (rating) is defined as the complete ranking (rating) that minimizes the sum of distances to each of the given rankings (ratings). We provide several examples that show that the consensus ranking (rating) obtained by this approach is more intuitive than that obtained by other approaches. Finding the consensus ranking is NP-hard, thus we develop two optimization methodologies to find the consensus ranking: one efficient approximation algorithm based on the separation-deviation model and one exact algorithm based on the implicit hitting set approach. In addition, we show that the optimization problem that needs to be solved in order to find the consensus rating is a special case of the separation-deviation model (hereafter SD model), which is solvable in polynomial time. In this sense, the herein developed theory (described in the previous paragraph) can be thought of an axiomatization of the SD model. Three applications of the SD model are presented: rating the credit-risk of countries; customer segmentation; and ranking the participants in a student paper competition. In the credit-risk rating study, it is shown that the SD model leads to an improved aggregate rating with respect to several criteria. We compare the SD model with other aggregation methods and show the following: Although the SD model is a method to aggregate cardinal evaluations, the aggregate credit-risk ratings obtained by the SD model are also good with respect to "ordinal criteria". Several properties of the SD model are proven, including the property that the aggregate rating obtained by the SD model agrees with the majority of agencies or reviewers, regardless of the scale used. The customer segmentation study shows how to use the SD model to process data on customer purchasing timing. The outcome of the SD model provides insights on the rate of new product adoption by the company's consumers. In particular, the SD model is used as follows: given the purchase dates for each customer of several products, this information is aggregated in order to rate the customers with regard to their promptness to adopt new technology. We show that this approach outperforms unidimensional scaling--a widely used data mining methodology. We analyze the results with respect to various dimensions of the customer base and report on the generated insights. The last presented application illustrates our aggregation methods in the context of the 2007 MSOM's student paper competition. The aggregation problem in this competition poses two challenges. First, each paper was reviewed only by a very small fraction of the judges; thus the aggregate evaluation is highly sensitive to the subjective scales chosen by the judges. Second, the judges provided both cardinal and ordinal evaluations (ratings and rankings) of the papers they reviewed. This chapter develops the first known methodology to simultaneously aggregate ordinal and cardinal evaluations into a consensus evaluation. Although the content of this dissertation is framed in terms of decision theory, Hochbaum showed that data mining problems can be viewed as special cases of decision theory problems. In particular, the customer segmentation study is a classic data mining problem. This book constitutes revised papers from the 12th International Conference on Large-Scale Scientific Computing, LSSC 2019, held in Sozopol, Bulgaria, in June 2019. The 70 papers presented in this volume were carefully reviewed and selected from 81 submissions. The book also contains two invited talks. The papers were organized in topical sections named as follows: control and optimization of dynamical systems; meshfree and particle methods; fractional diffusion problems: numerical methods, algorithms and applications; pore scale flow and transport simulation; tensors based algorithms and structures in optimization and applications; HPC and big data: algorithms and applications; large-scale models: numerical methods, parallel computations and applications; monte carlo algorithms: innovative applications in conjunctions with other methods; application of metaheuristics to large-scale problems; large scale machine learning: multiscale algorithms and performance guarantees; and contributed papers.

Data Mining is an emerging technology that has made its way into science, engineering, commerce and industry as many existing inference methods are obsolete for dealing with massive datasets that get accumulated in data warehouses. This comprehensive and up-to-date text aims at providing the reader with sufficient information about data mining methods and algorithms so that they can make use of these methods for solving real-world problems. The authors have taken care to include most of the widely used methods in data mining with simple examples so as to make the text ideal for classroom learning. To make the theory more comprehensible to the students, many illustrations have been used, and this in turn explains how certain parameters of interest change as the algorithm proceeds. Designed as a textbook for the undergraduate and postgraduate students of computer science, information technology, and master of computer applications, the book can also be used for MBA courses in Data Mining in Business, Business Intelligence, Marketing Research, and Health Care Management. Students of Bioinformatics will also find the text extremely useful. CD-ROM INCLUDE' The accompanying CD contains Large collection of datasets. Animation on how to use WEKA and ExcelMiner to do data mining. The microelectronics market, with special emphasis to the production of complex mixed-signal systems-on-chip (SoC), is driven by three main dynamics, time-- market, productivity and managing complexity. Pushed by the progress in na- meter technology, the design teams are facing a curve of complexity that grows exponentially, thereby slowing down the productivity design rate. Analog design automation tools are not developing at the same pace of technology, once custom design, characterized by decisions taken at each step of the analog design flow, -

lies most of the time on designer knowledge and expertise. Actually, the use of design management platforms, like the Cadences Virtuoso platform, with a set of integrated CAD tools and database facilities to deal with the design transformations from the system level to the physical implementation, can significantly speed-up the design process and enhance the productivity of analog/mixed-signal integrated circuit (IC) design teams. These design management platforms are a valuable help in analog IC design but they are still far behind the development stage of design automation tools already available for digital design. Therefore, the development of new CAD tools and design methodologies for analog and mixed-signal ICs is essential to increase the designer's productivity and reduce design productivity gap. The work presented in this book describes a new design automation approach to the problem of sizing analog ICs.

From the Foreword: "While large-scale machine learning and data mining have greatly impacted a range of commercial applications, their use in the field of Earth sciences is still in the early stages. This book, edited by Ashok Srivastava, Ramakrishna Nemani, and Karsten Steinhaeuser, serves as an outstanding resource for anyone interested in the opportunities and challenges for the machine learning community in analyzing these data sets to answer questions of urgent societal interest...I hope that this book will inspire more computer scientists to focus on environmental applications, and Earth scientists to seek collaborations with researchers in machine learning and data mining to advance the frontiers in Earth sciences." --Vipin Kumar, University of Minnesota Large-Scale Machine Learning in the Earth Sciences provides researchers and practitioners with a broad overview of some of the key challenges in the intersection of Earth science, computer science, statistics, and related fields. It explores a wide range of topics and provides a compilation of recent research in the application of machine learning in the field of Earth Science. Making predictions based on observational data is a theme of the book, and the book includes chapters on the use of network science to understand and discover teleconnections in extreme climate and weather events, as well as using structured estimation in high dimensions. The use of ensemble machine learning models to combine predictions of global climate models using information from spatial and temporal patterns is also explored. The second part of the book features a discussion on statistical downscaling in climate with state-of-the-art scalable machine learning, as well as an overview of methods to understand and predict the proliferation of biological species due to changes in environmental conditions. The problem of using large-scale machine learning to study the formation of tornadoes is also explored in depth. The last part of the book covers the use of deep learning algorithms to classify images that have very high resolution, as well as the unmixing of spectral signals in remote sensing images of land cover. The authors also apply long-tail distributions to geoscience resources, in the final chapter of the book.

The book consists of 29 extended chapters which have been selected and invited from the submissions to the 1st International Conference on Computer Science, Applied Mathematics and Applications (ICCSAMA 2013) held on 9-10 May, 2013 in Warsaw, Poland. The book is organized into five parts, which are: Advanced Optimization Methods and Their Applications, Queuing Theory and Applications, Computational Methods for Knowledge Engineering, Knowledge Engineering with Cloud and Grid Computing, and Logic Based Methods for Decision Making and Data Mining, respectively. All chapters in the book discuss theoretical and practical issues connected with computational methods and optimization methods for knowledge engineering. Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions presents an accessible treatment of the two main components of support vector machines (SVMs)-classification problems and regression problems. The book emphasizes the close connection between optimization theory and SVMs since optimization is one of the pillars on which 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 Data Science and Analytics with Python is designed for practitioners in data science and data analytics in both academic and business environments. The aim is to present the reader with the main concepts used in data science using tools developed in Python, such as SciKit-learn, Pandas, Numpy, and others. The use of Python is of particular interest, given its recent popularity in the data science community. The book can be used by seasoned programmers and newcomers alike. The book is organized in a way that individual chapters are sufficiently independent from each other so that the reader is comfortable using the contents as a reference. The book discusses what data science and analytics are, from the point of view of the process and results obtained. Important features of Python are also covered, including a Python primer. The basic elements of machine learning, pattern recognition, and artificial intelligence that underpin the algorithms and implementations used in the rest of the book also appear in the first part of the book. Regression analysis using Python, clustering techniques, and classification algorithms are covered in the second part of the book. Hierarchical clustering, decision trees, and ensemble techniques are also explored, along with dimensionality reduction techniques and recommendation systems. The support vector machine algorithm and the Kernel trick are discussed in the last part of the book. About the Author Dr. Jesús Rogel-Salazar is a Lead Data scientist with experience in the field working for companies such as AKQA, IBM Data Science Studio, Dow Jones and others. He is a visiting researcher at the Department of Physics at Imperial College London, UK and a member of the School of Physics, Astronomy and Mathematics at the University of Hertfordshire, UK, He obtained his doctorate in physics at Imperial College London for work on quantum atom optics and ultra-cold matter. He has held a position as senior lecturer in mathematics as well as a consultant in the financial industry since 2006. He is the author of the book Essential Matlab and Octave, also published by CRC Press. His interests include mathematical modelling, data science, and optimization in a wide range of applications including optics, quantum mechanics, data journalism, and finance.

Data mining aims at finding interesting, useful or profitable information in very large databases. The enormous increase in the size of available scientific and commercial databases (data avalanche) as well as the continuing and exponential growth in performance of present day computers make data mining a very active field. In many cases, the burgeoning volume of data sets has grown so large that it threatens to overwhelm rather than enlighten scientists. Therefore, traditional methods are revised and streamlined, complemented by many new methods to address challenging new problems. Mathematical Programming plays a key role in this endeavor. It helps us to formulate precise objectives (e.g., a clustering criterion or a measure of discrimination) as well as the constraints imposed on the solution (e.g., find a partition, a covering or a hierarchy in clustering). It also provides powerful mathematical tools to build highly performing exact or approximate algorithms. This book is based on lectures presented at the workshop on "Data Mining and Mathematical Programming" (October 10-13, 2006, Montreal) and will be a valuable scientific source of information to faculty, students, and researchers in

optimization, data analysis and data mining, as well as people working in computer science, engineering and applied mathematics.

The importance of having efficient and effective methods for data mining and knowledge discovery (DM&KD), to which the present book is devoted, grows every day and numerous such methods have been developed in recent decades. There exists a great variety of different settings for the main problem studied by data mining and knowledge discovery, and it seems that a very popular one is formulated in terms of binary attributes. In this setting, states of nature of the application area under consideration are described by Boolean vectors defined on some attributes. That is, by data points defined in the Boolean space of the attributes. It is postulated that there exists a partition of this space into two classes, which should be inferred as patterns on the attributes when only several data points are known, the so-called positive and negative training examples. The main problem in DM&KD is defined as finding rules for recognizing (classifying) new data points of unknown class, i. e. , deciding which of them are positive and which are negative. In other words, to infer the binary value of one more attribute, called the goal or class attribute. To solve this problem, some methods have been suggested which construct a Boolean function separating the two given sets of positive and negative training data points.

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

This book reviews the latest developments in nature-inspired computation, with a focus on the cross-disciplinary applications in data mining and machine learning. Data mining, machine learning and nature-inspired computation are current hot research topics due to their importance in both theory and practical applications. Adopting an application-focused approach, each chapter introduces a specific topic, with detailed descriptions of relevant algorithms, extensive literature reviews and implementation details. Covering topics such as nature-inspired algorithms, swarm intelligence, classification, clustering, feature selection, cybersecurity, learning algorithms over cloud, extreme learning machines, object categorization, particle swarm optimization, flower pollination and firefly algorithms, and neural networks, it also presents case studies and applications, including classifications of crisis-related tweets, extraction of named entities in the Tamil language, performance-based prediction of diseases, and healthcare services. This book is both a valuable reference resource and a practical guide for students, researchers and professionals in computer science, data and management sciences, artificial intelligence and machine learning.

This book constitutes the thoroughly refereed post-proceedings of the 4th International Conference on Machine Learning and Cybernetics, ICMLC 2005, held in Guangzhou, China in August 2005. The 114 revised full papers of this volume are organized in topical sections on agents and distributed artificial intelligence, control, data mining and knowledge discovery, fuzzy information processing, learning and reasoning, machine learning applications, neural networks and statistical learning methods, pattern recognition, vision and image processing.

[Copyright: c5a93ea5c01924df71ac54bdd9dc7263](https://doi.org/10.1007/978-1-4419-2471-1)