Data Mining Theory Methodology Techniques And Applications Lecture Notes In Computer Science Lecture Notes In Artificial Intelligence

Exotic methods refer to specific functions within general soft computing methods such as genetic algorithms, neural networks and rough sets theory. They are applied to ordinary shares for a variety of financial purposes, such as portfolio selection and optimization, classification of market states, forecasting of market states and data mining. This is in contrast to the wide spectrum of work done on exotic financial instruments, wherein advanced mathematics is used to construct financial instruments for hedging risks and for investment.In this book, particular aspects of the general method are used to create interesting applications. For instance, genetic niching produces a family of portfolios for the trader to choose from. Support vector machines, a special form of neural networks, forecast the financial markets; such a forecast is on market states, of which there are three OCo uptrending, mean reverting and downtrending. A self-organizing map displays in a vivid manner the states of the market. Rough sets with a new discretization method extract information from stock prices."

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

Decision trees have become one of the most powerful and popular approaches in knowledge discovery and data mining; it is the science of exploring large and complex bodies of data in order to discover useful patterns. Decision tree learning continues to evolve over time. Existing methods are constantly being improved and new methods introduced. This 2nd Edition is dedicated entirely to the field of decision trees in data mining; to cover all aspects of this important technique, as well as improved or new methods and techniques developed after the publication of our first edition. In this new

edition, all chapters have been revised and new topics brought in. New topics include Cost-Sensitive Active Learning, Learning with Uncertain and Imbalanced Data, Using Decision Trees beyond Classification Tasks, Privacy Preserving Decision Tree Learning, Lessons Learned from Comparative Studies, and Learning Decision Trees for Big Data. A walk-through guide to existing open-source data mining software is also included in this edition. This book invites readers to explore the many benefits in data mining that decision trees offer: Self-explanatory and easy to follow when compacted Able to handle a variety of input data: nominal, numeric and textual Scales well to big data Able to process datasets that may have errors or missing values High predictive performance for a relatively small computational effort Available in many open source data mining packages over a variety of platforms Useful for various tasks, such as classification, regression, clustering and feature selection

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

Extensive treatment of the most up-to-date topics Provides the theory and concepts behind popular and emerging methods Range of topics drawn from Statistics, Computer Science, and Electrical Engineering

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.

Data Mining: Practical Machine Learning Tools and Techniques, Third Edition, offers a thorough grounding in machine learning concepts as well as practical advice on applying machine learning tools and techniques in real-world data mining situations. This highly anticipated third edition of the most acclaimed work on data mining and machine learning will teach

you everything you need to know about preparing inputs, interpreting outputs, evaluating results, and the algorithmic methods at the heart of successful data mining. Thorough updates reflect the technical changes and modernizations that have taken place in the field since the last edition, including new material on Data Transformations, Ensemble Learning, Massive Data Sets, Multi-instance Learning, plus a new version of the popular Weka machine learning software developed by the authors. Witten, Frank, and Hall include both tried-and-true techniques of today as well as methods at the leading edge of contemporary research. The book is targeted at information systems practitioners, programmers, consultants, developers, information technology managers, specification writers, data analysts, data modelers, database R&D professionals, data warehouse engineers, data mining professionals. The book will also be useful for professors and students of upper-level undergraduate and graduate-level data mining and machine learning courses who want to incorporate data mining as part of their data management knowledge base and expertise. Provides a thorough grounding in machine learning concepts as well as practical advice on applying the tools and techniques to your data mining projects Offers concrete tips and techniques for performance improvement that work by transforming the input or output in machine learning methods Includes downloadable Weka software toolkit, a collection of machine learning algorithms for data mining tasks—in an updated, interactive interface. Algorithms in toolkit cover: data pre-processing, classification, regression, clustering, association rules, visualization

Advances in Data Mining Knowledge Discovery and Applications aims to help data miners, researchers, scholars, and PhD students who wish to apply data mining techniques. The primary contribution of this book is highlighting frontier fields and implementations of the knowledge discovery and data mining. It seems to be same things are repeated again. But in general, same approach and techniques may help us in different fields and expertise areas. This book presents knowledge discovery and data mining applications in two different sections. As known that, data mining covers areas of statistics, machine learning, data management and databases, pattern recognition, artificial intelligence, and other areas. In this book, most of the areas are covered with different data mining applications. The eighteen chapters have been classified in two parts: Knowledge Discovery and Data Mining Applications.

Visual Data Mining—Opening the Black Box Knowledge discovery holds the promise of insight into large, otherwise opaque datasets. Thenatureofwhatmakesaruleinterestingtoauserhasbeendiscussed 1 widely but most agree that it is a subjective quality based on the practical u- fulness of the information. Being subjective, the user needs to provide feedback to the system and, as is the case for all systems, the sooner the feedback is given the quicker it can in?uence the behavior of the system. There have been some impressive research activities over the past few years but the question to be asked is why is visual data mining only now being - vestigated commercially? Certainly, there have been

arguments for visual data 2 mining for a number of years – Ankerst and others argued in 2002 that current (autonomous and opaque) analysis techniques are ine?cient, as they fail to - rectly embed the user in dataset exploration and that a better solution involves the user and algorithm being more tightly coupled. Grinstein stated that the "current state of the art data mining tools are automated, but the perfect data mining tool is interactive and highly participatory," while Han has suggested that the "data selection and viewing of mining results should be fully inter- tive, the mining process should be more interactive than the current state of the 2 art and embedded applications should be fairly automated . " A good survey on 3 techniques until 2003 was published by de Oliveira and Levkowitz .

"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.

Advances in social science research methodologies and data analytic methods are changing the way research in information systems is conducted. New developments in statistical software technologies for data mining (DM) such as regression splines or decision tree induction can be used to assist researchers in systematic post-positivist theory testing and development. Established management science techniques like data envelopment analysis (DEA), and value focused thinking (VFT) can be used in combination with traditional statistical analysis and data mining techniques to more effectively explore behavioral questions in information systems research. As adoption and use of these research methods expand, there is growing need for a resource book to assist doctoral students and advanced researchers in understanding their potential to contribute to a broad range of research problems. Advances in Research Methods for Information Systems Research: Data Mining, Data Envelopment Analysis, Value Focused Thinking focuses on bridging and unifying these three different methodologies in order to bring them together in a unified volume for the information systems community. This book serves as a resource that provides overviews on each method, as well as applications on how they can be employed to address IS research problems. Its goal is to help researchers in their continuous efforts to set the pace for having an appropriate interplay between behavioral research and design science.

This book reviews state-of-the-art methodologies and techniques for analyzing enormous quantities of raw data in high-dimensional data spaces, to extract new information for decision making. The goal of this book is to provide a single introductory source, organized in a systematic way, in which we could direct the readers in analysis of large data sets, through the explanation of basic concepts, models and methodologies developed in recent decades. If you are an

instructor or professor and would like to obtain instructor's materials, please visit http://booksupport.wiley.com If you are an instructor or professor and would like to obtain a solutions manual, please send an email to: pressbooks@ieee.org Data Mining Applications in Engineering and Medicine targets to help data miners who wish to apply different data mining techniques. Data mining generally covers areas of statistics, machine learning, data management and databases, pattern recognition, artificial intelligence, etc. In this book, most of the areas are covered by describing different applications. This is why you will find here why and how Data Mining can also be applied to the improvement of project management. Since Data Mining has been widely used in a medical field, this book contains different chapters reffering to some aspects and importance of its use in the mentioned field: Incorporating Domain Knowledge into Medical Image Mining, Data Mining Techniques in Pharmacovigilance, Electronic Documentation of Clinical Pharmacy Interventions in Hospitals etc. We hope that this book will inspire readers to pursue education and research in this emerging field.

Interest in predictive analytics of big data has grown exponentially in the four years since the publication of Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data, Second Edition. In the third edition of this bestseller, the author has completely revised, reorganized, and repositioned the original chapters and produced 13 new chapters of creative and useful machine-learning data mining techniques. In sum, the 43 chapters of simple yet insightful quantitative techniques make this book unique in the field of data mining literature. What is new in the Third Edition: The current chapters have been completely rewritten. The core content has been extended with strategies and methods for problems drawn from the top predictive analytics conference and statistical modeling workshops. Adds thirteen new chapters including coverage of data science and its rise, market share estimation, share of wallet modeling without survey data, latent market segmentation, statistical regression modeling that deals with incomplete data, decile analysis assessment in terms of the predictive power of the data, and a user-friendly version of text mining, not requiring an advanced background in natural language processing (NLP). Includes SAS subroutines which can be easily converted to other languages. As in the previous edition, this book offers detailed background, discussion, and illustration of specific methods for solving the most commonly experienced problems in predictive modeling and analysis of big data. The author addresses each methodology and assigns its application to a specific type of problem. To better ground readers, the book provides an in-depth discussion of the basic methodologies of predictive modeling and analysis. While this type of overview has been attempted before, this approach offers a truly nitty-gritty, step-by-step method that both tyros and experts in the field can enjoy playing with.

· This book is an updated version of a well-received book previously published in Chinese by Science Press of China (the first edition in 2006 and the second in 2013). It offers a systematic and practical overview of spatial data mining, which

combines computer science and geo-spatial information science, allowing each field to profit from the knowledge and techniques of the other. To address the spatiotemporal specialties of spatial data, the authors introduce the key concepts and algorithms of the data field, cloud model, mining view, and Deren Li methods. The data field method captures the interactions between spatial objects by diffusing the data contribution from a universe of samples to a universe of population, thereby bridging the gap between the data model and the recognition model. The cloud model is a qualitative method that utilizes quantitative numerical characters to bridge the gap between pure data and linguistic concepts. The mining view method discriminates the different requirements by using scale, hierarchy, and granularity in order to uncover the anisotropy of spatial data mining. The Deren Li method performs data preprocessing to prepare it for further knowledge discovery by selecting a weight for iteration in order to clean the observed spatial data as much as possible. In addition to the essential algorithms and techniques, the book provides application examples of spatial data mining in geographic information science and remote sensing. The practical projects include spatiotemporal video data mining for protecting public security, serial image mining on nighttime lights for assessing the severity of the Syrian Crisis, and the applications in the government project 'the Belt and Road Initiatives'.

Data mining is the process of automatically searching large volumes of data for models and patterns using computational techniques from statistics, machine learning and information theory; it is the ideal tool for such an extraction of knowledge. Data mining is usually associated with a business or an organization's need to identify trends and profiles, allowing, for example, retailers to discover patterns on which to base marketing objectives. This book looks at both classical and recent techniques of data mining, such as clustering, discriminant analysis, logistic regression, generalized linear models, regularized regression, PLS regression, decision trees, neural networks, support vector machines, Vapnik theory, naive Bayesian classifier, ensemble learning and detection of association rules. They are discussed along with illustrative examples throughout the book to explain the theory of these methods, as well as their strengths and limitations. Key Features: Presents a comprehensive introduction to all techniques used in data mining and statistical learning, from classical to latest techniques. Starts from basic principles up to advanced concepts. Includes many step-by-step examples with the main software (R, SAS, IBM SPSS) as well as a thorough discussion and comparison of those software. Gives practical tips for data mining implementation to solve real world problems. Looks at a range of tools and applications, such as association rules, web mining and text mining, with a special focus on credit scoring. Supported by an accompanying website hosting datasets and user analysis. Statisticians and business intelligence analysts, students as well as computer science, biology, marketing and financial risk professionals in both commercial and government organizations across all business and industry sectors will benefit from this book.

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.

Decision trees have become one of the most powerful and popular approaches in knowledge discovery and data mining; it is the science of exploring large and complex bodies of data in order to discover useful patterns. Decision tree learning continues to evolve over time. Existing methods are constantly being improved and new methods introduced.This 2nd Edition is dedicated entirely to the field of decision trees in data mining; to cover all aspects of this important technique, as well as improved or new methods and techniques developed after the publication of our first edition. In this new edition, all chapters have been revised and new topics brought in. New topics include Cost-Sensitive Active Learning, Learning with Uncertain and Imbalanced Data, Using Decision Trees beyond Classification Tasks, Privacy Preserving Decision Tree Learning, Lessons Learned from Comparative Studies, and Learning Decision Trees for Big Data. A walk-through guide to existing open-source data mining software is also included in this edition.This book invites readers to explore the many benefits in data mining that decision trees offer:

Many companies have invested in building large databases and data warehouses capable of storing vast amounts of information. This book offers business, sales and marketing managers a practical guide to accessing such information. The book presents some of the most efficient statistical and deterministic methods for information processing and applications in order to extract targeted information and find hidden patterns. The techniques presented range from Bayesian approaches and their variations such as sequential Monte Carlo methods, Markov Chain Monte Carlo filters, Rao Blackwellization, to the biologically inspired paradigm of Neural Networks and decomposition techniques such as Empirical Mode Decomposition, Independent Component Analysis and Singular Spectrum Analysis. The book is directed

to the research students, professors, researchers and practitioners interested in exploring the advanced techniques in intelligent signal processing and data mining paradigms.

"This study develops a systematic design methodology based on data mining theory for decision-making in the development of crashworthy vehicles. The new data mining methodology allows the exploration of a large crash simulation dataset to discover the underlying relationships among vehicle crash responses and design variables at multiple levels and to derive design rules based on the whole-vehicle safety requirements to make decisions about component-level and subcomponent-level design. The method can resolve a major issue with existing design approaches related to vehicle crashworthiness: that is, limited abilities to explore information from large datasets, which may hamper decision-making in the design processes. At the component level, two structural design approaches were implemented for detailed component design with the data mining method: namely, a dimension-based approach and a node-based approach to handle structures with regular and irregular shapes, respectively. These two approaches were used to design a thin-walled vehicular structure, the S-shaped beam, against crash loading. A large number of design alternatives were created, and their responses under loading were evaluated by finite element simulations. The design variables and computed responses formed a large design dataset. This dataset was then mined to build a decision tree. Based on the decision tree, the interrelationships among the design parameters were revealed, and design rules were generated to produce a set of good designs. After the data mining, the critical design parameters were identified and the design space was reduced, which can simplify the design process. To partially replace the expensive finite element simulations, a surrogate model was used to model the relationships between design variables and response. Four machine learning algorithms, which can be used for surrogate model development, were compared. Based on the results, Gaussian process regression was determined to be the most suitable technique in the present scenario, and an optimization process was developed to tune the algorithm's hyperparameters, which govern the model structure and training process. To account for engineering uncertainty in the data mining method, a new decision tree for uncertain data was proposed based on the joint probability in uncertain spaces, and it was implemented to again design the S-beam structure. The findings show that the new decision tree can produce effective decision-making rules for engineering design under uncertainty. To evaluate the new approaches developed in this work, a comprehensive case study was conducted by designing a vehicle system against the frontal crash. A publicly available vehicle model was simplified and validated. Using the newly developed approaches, new component designs in this vehicle were generated and integrated back into the vehicle model so their crash behavior could be simulated. Based on the simulation results, one can conclude that the designs with the new method can outperform the original design in terms of measures of mass,

intrusion and peak acceleration. Therefore, the performance of the new design methodology has been confirmed. The current study demonstrates that the new data mining method can be used in vehicle crashworthiness design, and it has the potential to be applied to other complex engineering systems with a large amount of design data."--Abstract.
New technologies have enabled us to collect massive amounts of data in many fields. However, our pace of discovering useful information and knowledge from these data falls far behind our pace of collecting the data. Data Mining: Theories, Algorithms, and Examples introduces and explains a comprehensive set of data mining algorithms from various dat
Ensemble methods have been called the most influential development in Data Mining and Machine Learning in the past decade. They combine multiple models into one usually more accurate than the best of its components. Ensembles can provide a critical boost to industrial challenges -- from investment timing to drug discovery, and fraud detection to recommendation systems -- where predictive accuracy is more vital than model interpretability. Ensembles are useful with all modeling algorithms, but this book focuses on decision trees to explain them most clearly. After describing trees and their strengths and weaknesses, the authors provide an overview of regularization -- today understood to be a key reason for the superior performance of modern ensembling algorithms. The book continues with a clear description of two recent developments: Importance Sampling (IS) and Rule Ensembles (RE). IS reveals classic ensemble methods -- bagging, random forests, and boosting -- to be special cases of a single algorithm, thereby showing how to improve their accuracy and speed. REs are linear rule models derived from decision tree ensembles. They are the most interpretable version of ensembles, which is essential to applications such as credit scoring and fault diagnosis. Lastly, the authors explain the paradox of how ensembles achieve greater accuracy on new data despite their (apparently much greater) complexity. This book is aimed at novice and advanced analytic researchers and practitioners -- especially in Engineering, Statistics, and Computer Science. Those with little exposure to ensembles will learn why and how to employ this breakthrough method, and advanced practitioners will gain insight into building even more powerful models. Throughout, snippets of code in R are provided to illustrate the algorithms described and to encourage the reader to try the techniques. The authors are industry experts in data mining and machine learning who are also adjunct professors and popular speakers. Although early pioneers in discovering and using ensembles, they here distill and clarify the recent groundbreaking work of leading academics (such as Jerome Friedman) to bring the benefits of ensembles to practitioners. Table of Contents: Ensembles Discovered / Predictive Learning and Decision Trees / Model Complexity, Model Selection and Regularization / Importance Sampling and the Classic Ensemble Methods / Rule Ensembles and Interpretation Statistics / Ensemble Complexity
"Comprising more than 500 entries, the Encyclopedia of Research Design explains how to make decisions about research design,

undertake research projects in an ethical manner, interpret and draw valid inferences from data, and evaluate experiment design strategies and results. Two additional features carry this encyclopedia far above other works in the field: bibliographic entries devoted to significant articles in the history of research design and reviews of contemporary tools, such as software and statistical procedures, used to analyze results. It covers the spectrum of research design strategies, from material presented in introductory classes to topics necessary in graduate research; it addresses cross- and multidisciplinary research needs, with many examples drawn from the social and behavioral sciences, neurosciences, and biomedical and life sciences; it provides summaries of advantages and disadvantages of often-used strategies; and it uses hundreds of sample tables, figures, and equations based on real-life cases."--Publisher's description.

This volume provides a snapshot of the current state of the art in data mining, presenting it both in terms of technical developments and industrial applications. The collection of chapters is based on works presented at the Australasian Data Mining conferences and industrial forums. Authors include some of Australia's leading researchers and practitioners in data mining. The volume also contains chapters by regional and international authors.

Managing IT in Construction/Managing Construction for Tomorrow presents new developments in:- Managing IT strategies - Model based management tools including building information modeling- Information and knowledge management- Communication and collaboration - Data acquisition and storage- Visualization and simulation- Architectural design and

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

This is an applied handbook for the application of data mining techniques in the CRM framework. It combines a technical and a business perspective to cover the needs of business users who are looking for a practical guide on data mining. It focuses on Customer Segmentation and presents guidelines for the development of actionable segmentation schemes. By using non-technical language it guides readers through all the phases of the data mining process.

Exotic methods refer to a particular function within a general soft computing method such as genetic algorithms, neural networks and rough sets theory. They are applied to ordinary shares for a variety of financial purposes, such as portfolio selection and optimization, classification of market states, forecasting of market states and data mining. This is in contrast to the wide spectrum of work done on exotic financial instruments, wherein advanced mathematics is used to construct financial instruments for hedging risks and for investment. In this book, particular aspects of the general method are used to create interesting applications. For instance, genetic niching produces a family of portfolios for the trader to choose from. Support vector machines, a special form of neural networks, forecast the financial markets; such a forecast is on market states, of which there are three -- uptrending, mean reverting and downtrending. A self-organizing map displays in a vivid manner the states of the market. Rough sets with a new discretization method extract information from stock prices.

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

This book 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. Technological advancements have extracted a vast amount of useful knowledge and information for applications and services. These developments have evoked intelligent solutions that have been utilized in efforts to secure this data and avoid potential complex problems. Advances in Secure Computing, Internet Services, and Applications presents current research on the applications of computational intelligence in order to focus on the challenge humans face when securing knowledge and data. This book is a vital reference source for researchers, lecturers, professors, students, and developers, who have interest in secure computing and recent advanced in real life applications.

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.

The importance of having ef cient and effective methods for data mining and kn- ledge 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 de ned on some attributes. That is, by data points de ned 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 de ned as nding rules for recognizing (cl- sifying) 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. 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.

With the growing use of information technology and the recent advances in web systems, the amount of data available to users has increased exponentially. Thus, there is a critical need to understand the content of the data. As a result, data-mining has become a popular research topic in recent years for the treatment of the "data rich and information poor" syndrome. In this carefully edited volume a theoretical foundation as well as important new directions for data-mining research are presented. It brings together a set of well respected data mining theoreticians and researchers with practical data mining experiences. The presented theories will give data mining practitioners a scientific perspective in data mining and thus provide more insight into their problems, and the provided new data mining topics can be expected to stimulate further research in these important directions.

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.

Copyright: bc367d65504021ab8b1eaad0d7165d6a