A Data Mining Approach To Discover Genetic And

This book presents innovative work in Climate Informatics, a new field that reflects the application of data mining methods to climate science, and shows where this new and fast growing field is headed. Given its interdisciplinary nature, Climate Informatics offers insights, tools and methods that are increasingly needed in order to understand the climate system, an aspect which in turn has become crucial because of the threat of climate change. There has been a veritable explosion in the amount of data produced by satellites, environmental sensors and climate models that monitor, measure and forecast the earth system. In order to meaningfully pursue knowledge discovery on the basis of such voluminous and diverse datasets, it is necessary to apply machine learning methods, and Climate Informatics lies at the intersection of machine learning and climate science. This book grew out of the fourth workshop on Climate Informatics held in Boulder, Colorado in Sep. 2014.

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

As it becomes easier and less expensive for service providers to store huge amounts of data, the information collected about individuals is growing rapidly. At the same time, individuals' concerns about their privacy is increasing. Although most service providers use privacy policies to explain what information they are collecting, who will access it, and for what purpose, existing research shows that users often do not read privacy policies or they find privacy policies difficult to understand. Thus, users may not make the right regarding securing their privacy. Some studies have proposed tools to enhance the effectiveness of privacy policies and thus facilitate decision making, whereas others have introduced visualization models to increase privacy usability and effectiveness. Despite all of this, there has been relatively little work done on providing a comparison model to assist users when comparing the privacy practices of different companies in an effort to make informed decisions. In this study, we first analyze users' awareness of privacy policies and the privacy practices described in them, their privacy concerns, and their privacy needs. Next, we use text mining techniques to extract information users care about such as collected information, shared information, and provided controls. Unlike existing techniques, our approach attempts to avoid the use of patterns or rules as much as possible because the format of privacy policies often changes over time, and therefore, patterns and rules often become obsolete. We then develop a comparison tool to show the extracted information side by side. Then we conduct a survey to validate our comparison tool and gather users' privacy preferences. Because a side-by-side comparison may not work well when users are comparing a large number of policies, we propose a data mining based method to rank privacy policies. Unlike existing techniques that rely on user ratings, which are often not reliable, our approach relies on pair-wise preferences given by users, which are often a lot more reliable.

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.

"This book addresses existing solutions for data mining, with particular emphasis on potential real-world applications. It captures defining research on topics such as fuzzy set theory, clustering algorithms, semi-supervised clustering, modeling and managing data mining patterns, and sequence motif mining"--Provided by publisher.

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

Stressed about statistics? Overcome your anxieties and build your confidence! Develop the data-gathering, analysis, and application skills at the heart of health care today and see how statistics apply to the real world in which you will work. Step by step, the authors, noted educators and practitioners, make the often intimidating field of statistics approachable and illustrate its relevance to the future of health care from personal, institutional, regional, national, and international perspectives. Clear, easy-to-follow explanations of statistical methods and basic calculations along with hands-on practice with actual data from recognized sources point the way toward mastery of this growing field.

Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python presents an applied approach to data mining concepts and methods, using Python software for illustration Readers will learn how to implement a variety of popular data mining algorithms in Python (a free and open-source software) to tackle business problems and opportunities. This is the sixth version of this successful text, and the first using Python. It covers both statistical and machine learning algorithms for prediction, classification, visualization, dimension reduction, recommender systems, clustering, text mining and network analysis. It also includes: A new co-author, Peter Gedeck, who brings both experience teaching business analytics courses using Python, and expertise in the application of machine learning methods to the drug-discovery process A new section on ethical issues in data mining Updates and new material based on feedback from instructors teaching MBA, undergraduate, diploma and executive courses, and from their students More than a dozen case studies demonstrating applications for the data mining techniques described End-of-chapter exercises that help readers gauge and expand their comprehension and competency of the material presented A companion website with more than two dozen data sets, and instructor materials including exercise solutions, PowerPoint slides, and case solutions Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python is an ideal textbook for graduate and upper-undergraduate level courses in data mining, predictive analytics, and business

analytics. This new edition is also an excellent reference for analysts, researchers, and practitioners working with quantitative methods in the fields of business, finance, marketing, computer science, and information technology. “This book has by far the most comprehensive review of business analytics methods that I have ever seen, covering everything from classical approaches such as linear and logistic regression, through to modern methods like neural networks, bagging and boosting, and even much more business specific procedures such as social network analysis and text mining. If not the bible, it is at the least a definitive manual on the subject.” —Gareth M. James, University of Southern California and co-author (with Witten, Hastie and Tibshirani) of the best-selling book An Introduction to Statistical Learning, with Applications in R

"We live, today, in world of big data. The amount of information collected on human behavior every day is staggering, and exponentially greater than at any time in the past. At the same time, we are inundated by stories of powerful algorithms capable of churning through this sea of data and uncovering patterns. These techniques go by many names - data mining, predictive analytics, machine learning - and they are being used by governments as they spy on citizens and by huge corporations are they fine-tune their advertising strategies. And yet social scientists continue mainly to employ a set of analytical tools developed in an earlier era when data was sparse and difficult to come by. In this timely book, Paul Attewell and David Monaghan provide a simple and accessible introduction to Data Mining geared towards social scientists. They discuss how the data mining approach differs substantially, and in some ways radically, from that of conventional statistical modeling familiar to most social scientists. They demystify data mining, describing the diverse set of techniques that the term covers and discussing the strengths and weaknesses of the various approaches. Finally they give practical demonstrations of how to carry out analyses using data mining tools in a number of statistical software packages. It is the hope of the authors that this book will empower social scientists to consider incorporating data mining methodologies in their analytical toolkits"--Provided by publisher.

Data mining and social network analysis have been widely used in law enforcement to solve crimes. Research questions such as strength of ties in social networks, crime pattern discovery and prioritizing offenders have been studied in this area. However, most of those studies failed to consider the noisy nature of the data. The techniques they proposed only have been applied to small scale data sets. Therefore, it is an important task to design a framework that can work on large scale data sets and tolerance noisy data. In this dissertation, we built an integrated crime detection framework that combined two data mining techniques: decision tree and genetic algorithm and graph theories to solve the problems we pointed out. Our crime pattern analysis is based on all offenders of the state of Alabama in the past 50 years. Our constructed social network contains all Alabama residents. It allows us to fully evaluate the proposed models. Two case studies have been conducted to evaluate the framework. One is based on 625 inmates released from Madison county jail in 2004. Our experimental results show that our recommended risk level has strong correlation in predicting future offense. Another case study is based on the 100 real police reports. The experimental results show that the median ranking of arrestees remains at the top 3% of the return list.

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

Ontologies are now increasingly used to integrate, and organize data and knowledge, particularly in data and knowledge-intensive applications in both research and industry. The book is devoted to semantic data mining – a data mining approach where domain ontologies are used as background knowledge, and where the new challenge is to mine knowledge encoded in domain ontologies and knowledge graphs, rather than only purely empirical data. The introductory chapters of the book provide theoretical foundations of both data mining and ontology representation. Taking a unified perspective, the book then covers several methods for semantic data mining, addressing tasks such as pattern mining, classification and similarity-based approaches. It attempts to provide state-of-the-art answers to specific challenges and peculiarities of data mining with use of ontologies, in particular: How to deal with incompleteness of knowledge and the so-called Open World Assumption? What is a truly “semantic” similarity measure? The book contains several chapters with examples of applications of semantic data mining. The examples start from a scenario with moderate use of lightweight ontologies for knowledge graph enrichment and end with a full-fledged scenario of an intelligent knowledge discovery assistant using complex domain ontologies for meta-mining, i.e., an ontology-based meta-learning approach to full data mining processes. The book is intended for researchers in the fields of semantic technologies, knowledge engineering, data science, and data mining, and developers of knowledge-based systems and applications.

The menace of illegal access to data resources is a growing concern of researchers in the field of computer science. A significant amount of effort is required to monitor the activities in a computer network with a view to detect any attempt for intrusion. From this perspective, the main motivation behind this research is to design an efficient intrusion detection system using some novel data mining approaches that have the capability to detect intrusions with high detection rate with low false positive rate. In this work, we take multiple supports Apriori algorithm with various interestiness measures to obtain the most significant rules in detecting network intrusions. Further, we propose some novel ensemble of classifiers in order to enhance the detection rate of network attacks. Some unsupervised clustering algorithms have been proposed to further increase the detection rate of new or unseen attacks that fall under rare attacks categories. Finally, certain hybrid data mining approaches have been employed in order to design an efficient anomaly based network intrusion detection system that can achieve high detection rate and low false positive rate.

This book presents state-of-the-art research on intrusion detection using reinforcement learning, fuzzy and rough set theories, and genetic algorithm. Reinforcement learning is employed to incrementally learn the computer network behavior, while rough and fuzzy sets are utilized

to handle the uncertainty involved in the detection of traffic anomaly to secure data resources from possible attack. Genetic algorithms make it possible to optimally select the network traffic parameters to reduce the risk of network intrusion. The book is unique in terms of its content, organization, and writing style. Primarily intended for graduate electrical and computer engineering students, it is also useful for doctoral students pursuing research in intrusion detection and practitioners interested in network security and administration. The book covers a wide range of applications, from general computer security to server, network, and cloud security.

Data Mining: Practical Machine Learning Tools and Techniques, Fourth Edition, offers a thorough grounding in machine learning concepts, along with practical advice on applying these tools and techniques in real-world data mining situations. This highly anticipated fourth edition of the most acclaimed work on data mining and machine learning teaches readers everything they need to know to get going, from preparing inputs, interpreting outputs, evaluating results, to the algorithmic methods at the heart of successful data mining approaches. Extensive updates reflect the technical changes and modernizations that have taken place in the field since the last edition, including substantial new chapters on probabilistic methods and on deep learning. Accompanying the book is a new version of the popular WEKA machine learning software from the University of Waikato. Authors Witten, Frank, Hall, and Pal include today's techniques coupled with the methods at the leading edge of contemporary research. Please visit the book companion website at http://www.cs.waikato.ac.nz/ml/weka/book.html It contains Powerpoint slides for Chapters 1-12. This is a very comprehensive teaching resource, with many PPT slides covering each chapter of the book Online Appendix on the Weka workbench; again a very comprehensive learning aid for the open source software that goes with the book Table of contents, highlighting the many new sections in the 4th edition, along with reviews of the 1st edition, errata, etc. Provides a thorough grounding in machine learning concepts, as well as practical advice on applying the tools and techniques to data mining projects Presents concrete tips and techniques for performance improvement that work by transforming the input or output in machine learning methods Includes a downloadable WEKA software toolkit, a comprehensive collection of machine learning algorithms for data mining tasks-in an easy-to-use interactive interface Includes open-access online courses that introduce practical applications of the material in the book

"This book addresses the computations that are needed in order to help a student with the RHIT/RHIA certifications. It is a complete statistics textbook which also covers medical ethical considerations. This book introduces new features on how to visualize health care statistics using MS Excel and R-Project statistical software (open source) and more hands-on examples using real-world data from websites provided throughout the chapters. The books covers classical statistics in a general way such that many fields would benefit from using it. "Big Data" aka data-mining is a real market mover these days. With the vast amount of data available in the healthcare sector like financial, clinical, R&D, administration and operational data, big data can derive meaningful insights to improve the operational efficiency of the industry"--

Modern education has increased its reach through ICT tools and techniques. To manage educational data with the help of modern artificial intelligence, data and web mining techniques on dedicated cloud or grid platforms for educational institutes can be used. By utilizing data science techniques to manage educational data, the safekeeping, delivery, and use of knowledge can be increased for better quality education. Utilizing Educational Data Mining Techniques for Improved Learning: Emerging Research and Opportunities is a critical scholarly resource that explores data mining and management techniques that promote the improvement and optimization of educational data systems. The book intends to provide new models, platforms, tools, and protocols in data science for educational data analysis and introduces innovative hybrid system models dedicated to data science. Including topics such as automatic assessment, educational analytics, and machine learning, this book is essential for IT specialists, data analysts, computer engineers, education professionals, administrators, policymakers, researchers, academicians, and technology experts.

The leading introductory book on data mining, fully updated andrevised! When Berry and Linoff wrote the first edition of Data MiningTechniques in the late 1990s, data mining was just starting tomove out of the lab and into the office and has since grown tobecome an indispensable tool of modern business. This newedition—more than 50% new and revised— is asignificant update from the previous one, and shows you how toharness the newest data mining methods and techniques to solvecommon business problems. The duo of unparalleled authors shareinvaluable advice for improving response rates to direct marketingcampaigns, identifying new customer segments, and estimating creditrisk. In addition, they cover more advanced topics such aspreparing data for analysis and creating the necessaryinfrastructure for data mining at your company. Features significant updates since the previous edition andupdates you on best practices for using data mining methods andtechniques for solving common business problems Covers a new data mining technique in every chapter along withclear, concise explanations on how to apply each techniqueimmediately Touches on core data mining techniques, including decisiontrees, neural networks, collaborative filtering, association rules,link analysis, survival analysis, and more Provides best practices for performing data mining using simpletools such as Excel Data Mining Techniques, Third Edition covers a new datamining technique with each successive chapter and then demonstrateshow you can apply that technique for improved marketing, sales, andcustomer support to get immediate results.

The NASA Technical Reports Servcr (NTRS) houses half a million publications that are a valuable means of information to researchers, teachers, students, and the general public. These documents are all aerospace related with much scientific and technical information created or funded by NASA. Some types of documents include conference papers, research reports, meeting papers, journal articles and more. This is one of those documents.

Intrusion DetectionA Data Mining ApproachSpringer Nature

The Laser Interferometer Gravitational-Wave Observatory (LIGO) detectors are designed to capture and record gravitational wave strain signals from astrophysical sources. The Fscan family of algorithms is a key signal processing tool for analyzing the spectral content of LIGO data. An area of ongoing investigation are features of the LIGO Hanford strain data in the sub-100Hz range that are not fully understood. This work utilizes a statistically-oriented, data-mining approach to analyze three related but distinct subjects in LIGO Hanford data from the O3-era observing run: the coherence of auxiliary data channels with the strain, the potential for coherence between the Hanford and Livingston interferometers not caused by astrophysical sources, and the shifting density of high-power spectral lines in strain data.

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

"This book explores the key concepts of data mining and utilizing them on online social media platforms, offering valuable insight into data mining approaches for big data and sentiment analysis in online social media and covering many important security and other aspects and current trends"--

Data analysis and machine learning are research areas at the intersection of computer science, artificial intelligence, mathematics and statistics. They cover general methods and techniques that can be applied to a vast set of applications such as web and text mining, marketing, medical science, bioinformatics and business intelligence. This volume contains the revised versions of selected papers in the field of data analysis, machine learning and applications presented during the 31st Annual Conference of the German Classification Society (Gesellschaft für Klassifikation - GfKl). The conference was held at the Albert-Ludwigs-University in Freiburg, Germany, in March 2007.

Visualizing Health Care Statistics: A Data-Mining Approach is an introductory statistics text that demonstrates how to visualize health care statistics using Microsoft Excel and R-Project (open source statistical software) and hands-on examples using real-world data. In each chapter, students are encouraged to apply statistical knowledge to real-world health care situations. Through this approach, students develop data gathering and analysis skills all while preparing for the national Registered Health Information Technician (RHIT) exam.