

Data Science And Simulation In Transportation Research

This book offers the proceedings of the Second International Data Science Conference (iDSC2019), organized by Salzburg University of Applied Sciences, Austria. The Conference brought together researchers, scientists, and business experts to discuss new ways of embracing agile approaches to various facets of data science, including machine learning and artificial intelligence, data mining, data visualization, and communication. The papers gathered here include case studies of applied techniques, and theoretical papers that push the field into the future. The full-length scientific-track papers on Data Analytics are broadly grouped by category, including Complexity; NLP and Semantics; Modelling; and Comprehensibility. Included among real-world applications of data science are papers on Exploring insider trading using hypernetworks Data-driven approach to detection of autism spectrum disorder Anonymization and sentiment analysis of Twitter posts Theoretical papers in the book cover such topics as Optimal Regression Tree Models Through Mixed Integer Programming; Chance Influence in Datasets with Large Number of Features; Adversarial Networks — A Technology for Image Augmentation; and Optimal Regression Tree Models Through Mixed Integer Programming. Five shorter student-track papers are also published here, on topics such as State-of-the-art Deep Learning Methods to effect Neural Machine Translation from Natural Language into SQL A Smart Recommendation System to Simplify Projecting for a HMI/SCADA Platform Use of Adversarial Networks as a Technology for Image Augmentation Using Supervised Learning to Predict the Reliability of a Welding Process The work collected in this volume of proceedings will provide researchers and practitioners with a detailed snapshot of current progress in the field of data science. Moreover, it will stimulate new study, research, and the development of new applications.

"Jobs in data science abound, but few people have the data science skills needed to fill these increasingly important roles in organizations. Data Science For Dummies is the perfect starting point for IT professionals and students interested in making sense of their organization's massive data sets and applying their findings to real-world business scenarios. From uncovering rich data sources to managing large amounts of data within hardware and software limitations, ensuring consistency in reporting, merging various data sources, and beyond, you'll develop the know-how you need to effectively interpret data and tell a story that can be understood by anyone in your organization."--Provided by publisher.

This book encompasses a systematic exploration of Cybersecurity Data Science (CSDS) as an emerging profession, focusing on current versus idealized practice. This book also analyzes challenges facing the emerging CSDS profession, diagnoses key gaps, and prescribes treatments to facilitate advancement. Grounded in the management of information systems (MIS) discipline, insights derive from literature analysis and interviews with 50 global CSDS practitioners. CSDS as a diagnostic process grounded in the scientific method is emphasized throughout Cybersecurity Data Science (CSDS) is a rapidly evolving discipline which applies data science methods to cybersecurity challenges. CSDS reflects the rising interest in applying data-focused statistical, analytical, and machine learning-driven methods to address growing security gaps. This book offers a systematic assessment of the developing domain. Advocacy is provided to strengthen professional rigor and best practices in the emerging CSDS profession. This book will be of interest to a range of professionals associated with cybersecurity and data science, spanning practitioner, commercial, public sector, and academic domains. Best practices framed will be of interest to CSDS practitioners, security professionals, risk management stewards, and institutional stakeholders. Organizational and industry perspectives will be of interest to cybersecurity analysts, managers, planners, strategists, and regulators. Research professionals and academics are presented with a systematic analysis of the CSDS field, including an overview of the state of the art, a structured evaluation of key challenges, recommended best practices, and an extensive bibliography.

This book constitutes the refereed proceedings of the Second International Conference on Data Science, ICDS 2015, held in Sydney, Australia, during August 8-9, 2015. The 19 revised full papers and 5 short papers presented were carefully reviewed and selected from 31 submissions. The papers focus on the following topics: mathematical issues in data science; big data issues and applications; data quality and data preparation; data-driven scientific research; evaluation and measurement in data service; big data mining and knowledge management; case study of data science; social impacts of data science.

Given its effective techniques and theories from various sources and fields, data science is playing a vital role in transportation research and the consequences of the inevitable switch to electronic vehicles. This fundamental insight provides a step towards the solution of this important challenge. Data Science and Simulation in Transportation Research highlights entirely new and detailed spatial-temporal micro-simulation methodologies for human mobility and the emerging dynamics of our society. Bringing together novel ideas grounded in big data from various data mining and transportation science sources, this book is an essential tool for professionals, students, and researchers in the fields of transportation research and data mining.

We are living at the dawn of what has been termed 'the fourth paradigm of science,' a scientific revolution that is marked by both the emergence of big data science and analytics, and by the increasing adoption of the underlying technologies in scientific and scholarly research practices. Everything about science development or knowledge production is fundamentally changing thanks to the ever-increasing deluge of data. This is the primary fuel of the new age, which powerful computational processes or analytics algorithms are using to generate valuable knowledge for enhanced decision-making, and deep insights pertaining to a wide variety of practical uses and applications. This book addresses the complex interplay of the scientific, technological, and social dimensions of the city, and what it entails in terms of the systemic implications for smart sustainable urbanism. In concrete terms, it explores the interdisciplinary and transdisciplinary field of smart sustainable urbanism and the unprecedented paradigmatic shifts and practical advances it is undergoing in light of big data science and analytics. This new era of science and technology embodies an unprecedentedly transformative and constitutive power—manifested not only in the form of revolutionizing science and transforming knowledge, but also in advancing social practices, producing new discourses, catalyzing major shifts, and fostering societal transitions. Of particular relevance, it is instigating a massive change in the way both smart cities and sustainable cities are studied and understood, and in how they are planned, designed, operated, managed, and governed in the face of urbanization. This relates to what has been dubbed data-driven smart sustainable urbanism, an emerging approach based on a computational understanding of city systems and processes that reduces urban life to logical and algorithmic rules and procedures, while also harnessing urban big data to provide a more holistic and integrated view or synoptic intelligence of the city. This is increasingly being directed towards improving,

advancing, and maintaining the contribution of both sustainable cities and smart cities to the goals of sustainable development. This timely and multifaceted book is aimed at a broad readership. As such, it will appeal to urban scientists, data scientists, urbanists, planners, engineers, designers, policymakers, philosophers of science, and futurists, as well as all readers interested in an overview of the pivotal role of big data science and analytics in advancing every academic discipline and social practice concerned with data-intensive science and its application, particularly in relation to sustainability.

This two volume set (CCIS 1451 and 1452) constitutes the refereed proceedings of the 7th International Conference of Pioneering Computer Scientists, Engineers and Educators, ICPCSEE 2021 held in Taiyuan, China, in September 2021. The 81 papers presented in these two volumes were carefully reviewed and selected from 256 submissions. The papers are organized in topical sections on big data management and applications; social media and recommendation systems; infrastructure for data science; basic theory and techniques for data science; machine learning for data science; multimedia data management and analysis; social media and recommendation systems; data security and privacy; applications of data science; education research, methods and materials for data science and engineering; research demo.

This two volume set (CCIS 1058 and 1059) constitutes the refereed proceedings of the 5th International Conference of Pioneering Computer Scientists, Engineers and Educators, ICPCSEE 2019 held in Guilin, China, in September 2019. The 104 revised full papers presented in these two volumes were carefully reviewed and selected from 395 submissions. The papers cover a wide range of topics related to basic theory and techniques for data science including data mining; data base; net work; security; machine learning; bioinformatics; natural language processing; software engineering; graphic images; system; education; application.

Cyber-security is a matter of rapidly growing importance in industry and government. This book provides insight into a range of data science techniques for addressing these pressing concerns. The application of statistical and broader data science techniques provides an exciting growth area in the design of cyber defences. Networks of connected devices, such as enterprise computer networks or the wider so-called Internet of Things, are all vulnerable to misuse and attack, and data science methods offer the promise to detect such behaviours from the vast collections of cyber traffic data sources that can be obtained. In many cases, this is achieved through anomaly detection of unusual behaviour against understood statistical models of normality. This volume presents contributed papers from an international conference of the same name held at Imperial College. Experts from the field have provided their latest discoveries and review state of the art technologies.

"This book highlights entirely new and detailed spatial-temporal micro-simulation methodologies for human mobility and the emerging dynamics of our society, offering novel ideas grounded in big data from various data mining and transportation science sources"--

This two-volume set constitutes the proceedings of the Third Conference on Creativity in Intellectual Technologies and Data Science, CIT&DS 2019, held in Volgograd, Russia, in September 2019. The 67 full papers, 1 short paper and 3 keynote papers presented were carefully reviewed and selected from 231 submissions. The papers are organized in topical sections in the two volumes. Part I: cyber-physical systems and Big Data-driven world. Part II: artificial intelligence and deep learning technologies for creative tasks; intelligent technologies in social engineering.

Marketing text: This book covers the overlap between informatics, computer science, philosophy of causation, and causal inference in epidemiology and population health research. Key concepts covered include how data are generated and interpreted, and how and why concepts in health informatics and the philosophy of science should be integrated in a systems-thinking approach. Furthermore, a formal epistemology for the health sciences and public health is suggested. Causation in Population Health Informatics and Data Science provides a detailed guide of the latest thinking on causal inference in population health informatics. It is therefore a critical resource for all informaticians and epidemiologists interested in the potential benefits of utilising a systems-based approach to causal inference in health informatics.

Convergence has been a key topic of discussion about the future of cyberinfrastructure for science and engineering research. Convergence refers both to the combined use of simulation and data-centric techniques in science and engineering research and the possibilities for a single type of cyberinfrastructure to support both techniques. The National Academies of Science, Engineering, and Medicine convened a Workshop on Converging Simulation and Data-Driven Science on May 10, 2018, in Washington, D.C. The workshop featured speakers from universities, national laboratories, technology companies, and federal agencies who addressed the potential benefits and limitations of convergence as they relate to scientific needs, technological capabilities, funding structures, and system design requirements. This publication summarizes the presentations and discussions from the workshop.

This book constitutes selected papers of the Third International Conference on Data Science, Medicine and Bioinformatics, IDMB 2019, held in Nanning, China, in June 2019. The 19 full papers and 1 short paper were carefully reviewed and selected from 93 submissions. The papers are organized according to the following topical sections: business data science: fintech, management, and analytics.- health and biological data science.- novel data science theory and applications.

The internet of things (IoT) has emerged to address the need for connectivity and seamless integration with other devices as well as big data platforms for analytics. However, there are challenges that IoT-based applications face including design and implementation issues; connectivity problems; data gathering, storing, and analyzing in cloud-based environments; and IoT security and privacy issues. Emerging Trends in IoT and Integration with Data Science, Cloud Computing, and Big Data Analytics is a critical reference source that provides theoretical frameworks and research findings on IoT and big data integration. Highlighting topics that include wearable sensors, machine learning, machine intelligence, and mobile computing, this book serves professionals who want to improve their understanding of the strategic role of trust at different levels of the information and

knowledge society. It is therefore of most value to data scientists, computer scientists, data analysts, IT specialists, academicians, professionals, researchers, and students working in the field of information and knowledge management in various disciplines that include but are not limited to information and communication sciences, administrative sciences and management, education, sociology, computer science, etc. Moreover, the book provides insights and supports executives concerned with the management of expertise, knowledge, information, and organizational development in different types of work communities and environments.

This book gathers invited presentations from the 2nd Symposium of the ICOSA- CANADA Chapter held at the University of Calgary from August 4-6, 2015. The aim of this Symposium was to promote advanced statistical methods in big-data sciences and to allow researchers to exchange ideas on statistics and data science and to embrace the challenges and opportunities of statistics and data science in the modern world. It addresses diverse themes in advanced statistical analysis in big-data sciences, including methods for administrative data analysis, survival data analysis, missing data analysis, high-dimensional and genetic data analysis, longitudinal and functional data analysis, the design and analysis of studies with response-dependent and multi-phase designs, time series and robust statistics, statistical inference based on likelihood, empirical likelihood and estimating functions. The editorial group selected 14 high-quality presentations from this successful symposium and invited the presenters to prepare a full chapter for this book in order to disseminate the findings and promote further research collaborations in this area. This timely book offers new methods that impact advanced statistical model development in big-data sciences.

This edited book first consolidates the results of the EU-funded EDISON project (Education for Data Intensive Science to Open New science frontiers), which developed training material and information to assist educators, trainers, employers, and research infrastructure managers in identifying, recruiting and inspiring the data science professionals of the future. It then deepens the presentation of the information and knowledge gained to allow for easier assimilation by the reader. The contributed chapters are presented in sequence, each chapter picking up from the end point of the previous one. After the initial book and project overview, the chapters present the relevant data science competencies and body of knowledge, the model curriculum required to teach the required foundations, profiles of professionals in this domain, and use cases and applications. The text is supported with appendices on related process models. The book can be used to develop new courses in data science, evaluate existing modules and courses, draft job descriptions, and plan and design efficient data-intensive research teams across scientific disciplines.

Data Science Applications using Python and R is the second book in a series that began in 2018. This volume is dedicated to text analytics and natural language processing. Using real data, the author leads the reader through the analysis of Tweet sentiment analysis, banking product-group complaint analysis, presidential debate analysis, and more. The book covers text mining, natural language processing (NLP), vectorizing text data, discrete classifiers, bag-of-words (BOW) models, sentiment analysis, and Latent Dirichlet Allocation (LDA). The book offers complete Python and R code with detail explanations. It is designed for use with Jupyter Notebook and R Studio. It also includes notes on Python and R markdown and features full color graphics and text on heavy paper. All data sets used in the book are downloadable from GitHub. Some data can also be customized and downloaded from the Federal Consumer Complaint Data Catalog. Finally, each chapter contains practice exercises.

Data Science and Analytics for Ordinary People is a collection of blogs I have written on LinkedIn over the past year. As I continue to perform big data analytics, I continue to discover, not only my weaknesses in communicating the information, but new insights into using the information obtained from analytics and communicating it. These are the kinds of things I blog about and are contained herein. Data science and analytics have been used as synonyms on occasion. In reality data science includes data modeling, data mining, data analysis, database architecture and so on. Analytics is what we do to make sense of the data. That is, we take data and turn it into information for business decision makers. This our course implies that we translate our data science jargon into English.

Now , a leader of Northwestern University's prestigious analytics program presents a fully-integrated treatment of both the business and academic elements of marketing applications in predictive analytics. Writing for both managers and students, Thomas W. Miller explains essential concepts, principles, and theory in the context of real-world applications. Building on Miller's pioneering program, Marketing Data Science thoroughly addresses segmentation, target marketing, brand and product positioning, new product development, choice modeling, recommender systems, pricing research, retail site selection, demand estimation, sales forecasting, customer retention, and lifetime value analysis. Starting where Miller's widely-praised Modeling Techniques in Predictive Analytics left off, he integrates crucial information and insights that were previously segregated in texts on web analytics, network science, information technology, and programming. Coverage includes: The role of analytics in delivering effective messages on the web Understanding the web by understanding its hidden structures Being recognized on the web – and watching your own competitors Visualizing networks and understanding communities within them Measuring sentiment and making recommendations Leveraging key data science methods: databases/data preparation, classical/Bayesian statistics, regression/classification, machine learning, and text analytics Six complete case studies address exceptionally relevant issues such as: separating legitimate email from spam; identifying legally-relevant information for lawsuit discovery; gleaning insights from anonymous web surfing data, and more. This text's extensive set of web and network problems draw on rich public-domain data sources; many are accompanied by solutions in Python and/or R. Marketing Data Science will be an invaluable resource for all students, faculty, and professional marketers who want to use business analytics to improve marketing performance.

Data science has always been an effective way of extracting knowledge and insights from information in various forms. One industry that can utilize the benefits from the

advances in data science is the healthcare field. The Handbook of Research on Data Science for Effective Healthcare Practice and Administration is a critical reference source that overviews the state of data analysis as it relates to current practices in the health sciences field. Covering innovative topics such as linear programming, simulation modeling, network theory, and predictive analytics, this publication is recommended for all healthcare professionals, graduate students, engineers, and researchers that are seeking to expand their knowledge of efficient techniques for information analysis in the healthcare professions.

The 5-volume proceedings, LNAI 12457 until 12461 constitutes the refereed proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases, ECML PKDD 2020, which was held during September 14-18, 2020. The conference was planned to take place in Ghent, Belgium, but had to change to an online format due to the COVID-19 pandemic. The 232 full papers and 10 demo papers presented in this volume were carefully reviewed and selected for inclusion in the proceedings. The volumes are organized in topical sections as follows: Part I: Pattern Mining; clustering; privacy and fairness; (social) network analysis and computational social science; dimensionality reduction and autoencoders; domain adaptation; sketching, sampling, and binary projections; graphical models and causality; (spatio-) temporal data and recurrent neural networks; collaborative filtering and matrix completion. Part II: deep learning optimization and theory; active learning; adversarial learning; federated learning; Kernel methods and online learning; partial label learning; reinforcement learning; transfer and multi-task learning; Bayesian optimization and few-shot learning. Part III: Combinatorial optimization; large-scale optimization and differential privacy; boosting and ensemble methods; Bayesian methods; architecture of neural networks; graph neural networks; Gaussian processes; computer vision and image processing; natural language processing; bioinformatics. Part IV: applied data science: recommendation; applied data science: anomaly detection; applied data science: Web mining; applied data science: transportation; applied data science: activity recognition; applied data science: hardware and manufacturing; applied data science: spatiotemporal data. Part V: applied data science: social good; applied data science: healthcare; applied data science: e-commerce and finance; applied data science: computational social science; applied data science: sports; demo track. .

This is the second edition of Wil van der Aalst's seminal book on process mining, which now discusses the field also in the broader context of data science and big data approaches. It includes several additions and updates, e.g. on inductive mining techniques, the notion of alignments, a considerably expanded section on software tools and a completely new chapter of process mining in the large. It is self-contained, while at the same time covering the entire process-mining spectrum from process discovery to predictive analytics. After a general introduction to data science and process mining in Part I, Part II provides the basics of business process modeling and data mining necessary to understand the remainder of the book. Next, Part III focuses on process discovery as the most important process mining task, while Part IV moves beyond discovering the control flow of processes, highlighting conformance checking, and organizational and time perspectives. Part V offers a guide to successfully applying process mining in practice, including an introduction to the widely used open-source tool ProM and several commercial products. Lastly, Part VI takes a step back, reflecting on the material presented and the key open challenges. Overall, this book provides a comprehensive overview of the state of the art in process mining. It is intended for business process analysts, business consultants, process managers, graduate students, and BPM researchers.

Applied Statistical Modeling and Data Analytics: A Practical Guide for the Petroleum Geosciences provides a practical guide to many of the classical and modern statistical techniques that have become established for oil and gas professionals in recent years. It serves as a "how to" reference volume for the practicing petroleum engineer or geoscientist interested in applying statistical methods in formation evaluation, reservoir characterization, reservoir modeling and management, and uncertainty quantification. Beginning with a foundational discussion of exploratory data analysis, probability distributions and linear regression modeling, the book focuses on fundamentals and practical examples of such key topics as multivariate analysis, uncertainty quantification, data-driven modeling, and experimental design and response surface analysis. Data sets from the petroleum geosciences are extensively used to demonstrate the applicability of these techniques. The book will also be useful for professionals dealing with subsurface flow problems in hydrogeology, geologic carbon sequestration, and nuclear waste disposal. Authored by an internationally renowned scientist and a leader in the field of spatial statistics Presents an easy to follow narrative which progresses from simple concepts to more challenging ones Includes practical examples for the most relevant and popular statistical methods, using data sets from the petroleum geosciences Addresses the theory and practice of statistical modeling and analysis from the perspective of geoscience applications

This book constitutes the post-conference proceedings of the 4th International Conference on Machine Learning, Optimization, and Data Science, LOD 2018, held in Volterra, Italy, in September 2018. The 46 full papers presented were carefully reviewed and selected from 126 submissions. The papers cover topics in the field of machine learning, artificial intelligence, reinforcement learning, computational optimization and data science presenting a substantial array of ideas, technologies, algorithms, methods and applications.

This is the eBook of the printed book and may not include any media, website access codes, or print supplements that may come packaged with the bound book. This up-to-the-minute reference will help you master all three facets of sports analytics — and use it to win! Sports Analytics and Data Science is the most accessible and practical guide to sports analytics for everyone who cares about winning and everyone who is interested in data science. You'll discover how successful sports analytics blends business and sports savvy, modern information technology, and sophisticated modeling techniques. You'll master the discipline through realistic sports vignettes and intuitive data

visualizations—not complex math. Every chapter focuses on one key sports analytics application. Miller guides you through assessing players and teams, predicting scores and making game-day decisions, crafting brands and marketing messages, increasing revenue and profitability, and much more. Step by step, you'll learn how analysts transform raw data and analytical models into wins: both on the field and in any sports business.

Data Science gets thrown around in the press like it's magic. Major retailers are predicting everything from when their customers are pregnant to when they want a new pair of Chuck Taylors. It's a brave new world where seemingly meaningless data can be transformed into valuable insight to drive smart business decisions. But how does one exactly do data science? Do you have to hire one of these priests of the dark arts, the "data scientist," to extract this gold from your data? Nope. Data science is little more than using straightforward steps to process raw data into actionable insight. And in *Data Smart*, author and data scientist John Foreman will show you how that's done within the familiar environment of a spreadsheet. Why a spreadsheet? It's comfortable! You get to look at the data every step of the way, building confidence as you learn the tricks of the trade. Plus, spreadsheets are a vendor-neutral place to learn data science without the hype. But don't let the Excel sheets fool you. This is a book for those serious about learning the analytic techniques, the math and the magic, behind big data. Each chapter will cover a different technique in a spreadsheet so you can follow along: Mathematical optimization, including non-linear programming and genetic algorithms Clustering via k-means, spherical k-means, and graph modularity Data mining in graphs, such as outlier detection Supervised AI through logistic regression, ensemble models, and bag-of-words models Forecasting, seasonal adjustments, and prediction intervals through monte carlo simulation Moving from spreadsheets into the R programming language You get your hands dirty as you work alongside John through each technique. But never fear, the topics are readily applicable and the author laces humor throughout. You'll even learn what a dead squirrel has to do with optimization modeling, which you no doubt are dying to know.

This book is based on the "Summer Simulation Multi-Conference" (SCSC), which has been a prominent platform for the dissemination of scholarly research in the M&S community for the last 50 years. In keeping with the conference's seasonal title, the authors have called this half-century "the summer of simulation," and it has led not only to simulation-based disciplines but also simulation as a discipline. This book discusses contributions from the SCSC in four sections. The first section is an introduction to the work. The second section is devoted to contributions from simulation research fellows who were associated with the SCSC, while the third section features the SCSC's most influential contributions. Lastly, the fourth section includes contributions from the best papers in the last five years. Features: • A comprehensive volume dedicated to one of the simulation domain's major conferences: the SCSC • Offers a scientometric analysis of the SCSC • Revisits high-impact topics from 50 years of the SCSC • Includes chapters by simulation research fellows associated with the SCSC • Presents updated best-paper contributions from the recent conference This work will be of value to anyone interested in the evolution of modeling and simulation over the last fifty years. Readers will gain a perspective on what drove this evolution, and develop an understanding of the key contributions that allowed this technology to grow into its own academic discipline and profession.

Learn how to build a data science technology stack and perform good data science with repeatable methods. You will learn how to turn data lakes into business assets. The data science technology stack demonstrated in *Practical Data Science* is built from components in general use in the industry. Data scientist Andreas Vermeulen demonstrates in detail how to build and provision a technology stack to yield repeatable results. He shows you how to apply practical methods to extract actionable business knowledge from data lakes consisting of data from a polyglot of data types and dimensions. What You'll Learn Become fluent in the essential concepts and terminology of data science and data engineering Build and use a technology stack that meets industry criteria Master the methods for retrieving actionable business knowledge Coordinate the handling of polyglot data types in a data lake for repeatable results Who This Book Is For Data scientists and data engineers who are required to convert data from a data lake into actionable knowledge for their business, and students who aspire to be data scientists and data engineers

This book constitutes the proceedings of the Research School on Statistics and Data Science, RSSDS 2019, held in Melbourne, VIC, Australia, in July 2019. The 11 papers presented in this book were carefully reviewed and selected from 23 submissions. The volume also contains 7 invited talks. The workshop brought together academics, researchers, and industry practitioners of statistics and data science, to discuss numerous advances in the disciplines and their impact on the sciences and society. The topics covered are data analysis, data science, data mining, data visualization, bioinformatics, machine learning, neural networks, statistics, and probability. --

This two volume set (CCIS 727 and 728) constitutes the refereed proceedings of the Third International Conference of Pioneering Computer Scientists, Engineers and Educators, ICPCSEE 2017 (originally ICYCSEE) held in Changsha, China, in September 2017. The 112 revised full papers presented in these two volumes were carefully reviewed and selected from 987 submissions. The papers cover a wide range of topics related to Basic Theory and Techniques for Data Science including Mathematical Issues in Data Science, Computational Theory for Data Science, Big Data Management and Applications, Data Quality and Data Preparation, Evaluation and Measurement in Data Science, Data Visualization, Big Data Mining and Knowledge Management, Infrastructure for Data Science, Machine Learning for Data Science, Data Security and Privacy, Applications of Data Science, Case Study of Data Science, Multimedia Data Management and Analysis, Data-driven Scientific Research, Data-driven Bioinformatics, Data-driven Healthcare, Data-driven Management, Data-driven eGovernment, Data-driven Smart City/Planet, Data Marketing and Economics, Social Media and Recommendation Systems, Data-driven Security, Data-driven Business Model Innovation, Social and/or organizational impacts of Data Science.

Harness actionable insights from your data with computational statistics and simulations using R About This Book Learn five different simulation techniques (Monte Carlo,

Discrete Event Simulation, System Dynamics, Agent-Based Modeling, and Resampling) in-depth using real-world case studies A unique book that teaches you the essential and fundamental concepts in statistical modeling and simulation Who This Book Is For This book is for users who are familiar with computational methods. If you want to learn about the advanced features of R, including the computer-intense Monte-Carlo methods as well as computational tools for statistical simulation, then this book is for you. Good knowledge of R programming is assumed/required. What You Will Learn The book aims to explore advanced R features to simulate data to extract insights from your data. Get to know the advanced features of R including high-performance computing and advanced data manipulation See random number simulation used to simulate distributions, data sets, and populations Simulate close-to-reality populations as the basis for agent-based micro-, model- and design-based simulations Applications to design statistical solutions with R for solving scientific and real world problems Comprehensive coverage of several R statistical packages like boot, simPop, VIM, data.table, dplyr, parallel, StatDA, simecol, simecolModels, deSolve and many more. In Detail Data Science with R aims to teach you how to begin performing data science tasks by taking advantage of Rs powerful ecosystem of packages. R being the most widely used programming language when used with data science can be a powerful combination to solve complexities involved with varied data sets in the real world. The book will provide a computational and methodological framework for statistical simulation to the users. Through this book, you will get in grips with the software environment R. After getting to know the background of popular methods in the area of computational statistics, you will see some applications in R to better understand the methods as well as gaining experience of working with real-world data and real-world problems. This book helps uncover the large-scale patterns in complex systems where interdependencies and variation are critical. An effective simulation is driven by data generating processes that accurately reflect real physical populations. You will learn how to plan and structure a simulation project to aid in the decision-making process as well as the presentation of results. By the end of this book, you reader will get in touch with the software environment R. After getting background on popular methods in the area, you will see applications in R to better understand the methods as well as to gain experience when working on real-world data and real-world problems. Style and approach This book takes a practical, hands-on approach to explain the statistical computing methods, gives advice on the usage of these methods, and provides computational tools to help you solve common problems in statistical simulation and computer-intense methods.

Master predictive analytics, from start to finish Start with strategy and management Master methods and build models Transform your models into highly-effective code—in both Python and R This one-of-a-kind book will help you use predictive analytics, Python, and R to solve real business problems and drive real competitive advantage. You'll master predictive analytics through realistic case studies, intuitive data visualizations, and up-to-date code for both Python and R—not complex math. Step by step, you'll walk through defining problems, identifying data, crafting and optimizing models, writing effective Python and R code, interpreting results, and more. Each chapter focuses on one of today's key applications for predictive analytics, delivering skills and knowledge to put models to work—and maximize their value. Thomas W. Miller, leader of Northwestern University's pioneering program in predictive analytics, addresses everything you need to succeed: strategy and management, methods and models, and technology and code. If you're new to predictive analytics, you'll gain a strong foundation for achieving accurate, actionable results. If you're already working in the field, you'll master powerful new skills. If you're familiar with either Python or R, you'll discover how these languages complement each other, enabling you to do even more. All data sets, extensive Python and R code, and additional examples available for download at <http://www.ftpress.com/miller/> Python and R offer immense power in predictive analytics, data science, and big data. This book will help you leverage that power to solve real business problems, and drive real competitive advantage. Thomas W. Miller's unique balanced approach combines business context and quantitative tools, illuminating each technique with carefully explained code for the latest versions of Python and R. If you're new to predictive analytics, Miller gives you a strong foundation for achieving accurate, actionable results. If you're already a modeler, programmer, or manager, you'll learn crucial skills you don't already have. Using Python and R, Miller addresses multiple business challenges, including segmentation, brand positioning, product choice modeling, pricing research, finance, sports, text analytics, sentiment analysis, and social network analysis. He illuminates the use of cross-sectional data, time series, spatial, and spatio-temporal data. You'll learn why each problem matters, what data are relevant, and how to explore the data you've identified. Miller guides you through conceptually modeling each data set with words and figures; and then modeling it again with realistic code that delivers actionable insights. You'll walk through model construction, explanatory variable subset selection, and validation, mastering best practices for improving out-of-sample predictive performance. Miller employs data visualization and statistical graphics to help you explore data, present models, and evaluate performance. Appendices include five complete case studies, and a detailed primer on modern data science methods. Use Python and R to gain powerful, actionable, profitable insights about: Advertising and promotion Consumer preference and choice Market baskets and related purchases Economic forecasting Operations management Unstructured text and language Customer sentiment Brand and price Sports team performance And much more

As data holdings get bigger and questions get harder, data scientists and analysts must focus on the systems, the tools and techniques, and the disciplined process to get the correct answer, quickly! Whether you work within industry or government, this book will provide you with a foundation to successfully and confidently process large amounts of quantitative data. Here are just a dozen of the many questions answered within these pages: What does quantitative analysis of a system really mean? What is a system? What are big data and analytics? How do you know your numbers are good? What will the future data science environment look like? How do you determine data provenance? How do you gather and process information, and then organize, store, and synthesize it? How does an organization implement data analytics? Do you really need to think like a Chief Information Officer? What is the best way to protect data? What makes a good dashboard? What is the relationship between eating ice cream and getting attacked by a shark? The nine chapters in this book are arranged in three parts that address systems concepts in general, tools and techniques, and future trend topics. Systems concepts include contrasting open and closed systems, performing data mining and big data analysis, and gauging data quality. Tools and techniques include analyzing both continuous and discrete data, applying probability basics, and practicing quantitative analysis such as descriptive and inferential statistics.

Future trends include leveraging the Internet of Everything, modeling Artificial Intelligence, and establishing a Data Analytics Support Office (DASO). Many examples are included that were generated using common software, such as Excel, Minitab, Tableau, SAS, and Crystal Ball. While words are good, examples can sometimes be a better teaching tool. For each example included, data files can be found on the companion website. Many of the data sets are tied to the global economy because they use data from shipping ports, air freight hubs, largest cities, and soccer teams. The appendices contain more detailed analysis including the 10 T's for Data Mining, Million Row Data Audit (MRDA) Processes, Analysis of Rainfall, and Simulation Models for Evaluating Traffic Flow.

Simulation for Data Science with R Packt Publishing Ltd

This edited volume is brought out from the contributions of the research papers presented in the International Conference on Data Science and Business Analytics (ICDSBA- 2017), which was held during September 23-25 2017 in ChangSha, China. As we all know, the field of data science and business analytics is emerging at the intersection of the fields of mathematics, statistics, operations research, information systems, computer science and engineering. Data science and business analytics is an interdisciplinary field about processes and systems to extract knowledge or insights from data. Data science and business analytics employ techniques and theories drawn from many fields including signal processing, probability models, machine learning, statistical learning, data mining, database, data engineering, pattern recognition, visualization, descriptive analytics, predictive analytics, prescriptive analytics, uncertainty modeling, big data, data warehousing, data compression, computer programming, business intelligence, computational intelligence, and high performance computing among others. The volume contains 55 contributions from diverse areas of Data Science and Business Analytics, which has been categorized into five sections, namely: i) Marketing and Supply Chain Analytics; ii) Logistics and Operations Analytics; iii) Financial Analytics. iv) Predictive Modeling and Data Analytics; v) Communications and Information Systems Analytics. The readers shall not only receive the theoretical knowledge about this upcoming area but also cutting edge applications of this domains.

Data Science for Business and Decision Making covers both statistics and operations research while most competing textbooks focus on one or the other. As a result, the book more clearly defines the principles of business analytics for those who want to apply quantitative methods in their work. Its emphasis reflects the importance of regression, optimization and simulation for practitioners of business analytics. Each chapter uses a didactic format that is followed by exercises and answers. Freely-accessible datasets enable students and professionals to work with Excel, Stata Statistical Software®, and IBM SPSS Statistics Software®. Combines statistics and operations research modeling to teach the principles of business analytics Written for students who want to apply statistics, optimization and multivariate modeling to gain competitive advantages in business Shows how powerful software packages, such as SPSS and Stata, can create graphical and numerical outputs

See how data science can answer the questions your business faces! Applying Data Science: Business Case Studies Using SAS, by Gerhard Svolba, shows you the benefits of analytics, how to gain more insight into your data, and how to make better decisions. In eight entertaining and real-world case studies, Svolba combines data science and advanced analytics with business questions, illustrating them with data and SAS code. The case studies range from a variety of fields, including performing headcount survival analysis for employee retention, forecasting the demand for new projects, using Monte Carlo simulation to understand outcome distribution, among other topics. The data science methods covered include Kaplan-Meier estimates, Cox Proportional Hazard Regression, ARIMA models, Poisson regression, imputation of missing values, variable clustering, and much more! Written for business analysts, statisticians, data miners, data scientists, and SAS programmers, Applying Data Science bridges the gap between high-level, business-focused books that skimp on the details and technical books that only show SAS code with no business context.

Clouds are being positioned as the next-generation consolidated, centralized, yet federated IT infrastructure for hosting all kinds of IT platforms and for deploying, maintaining, and managing a wider variety of personal, as well as professional applications and services. Handbook of Research on Cloud Infrastructures for Big Data Analytics focuses exclusively on the topic of cloud-sponsored big data analytics for creating flexible and futuristic organizations. This book helps researchers and practitioners, as well as business entrepreneurs, to make informed decisions and consider appropriate action to simplify and streamline the arduous journey towards smarter enterprises.

Over 85 recipes to help you complete real-world data science projects in R and Python About This Book Tackle every step in the data science pipeline and use it to acquire, clean, analyze, and visualize your data Get beyond the theory and implement real-world projects in data science using R and Python Easy-to-follow recipes will help you understand and implement the numerical computing concepts Who This Book Is For If you are an aspiring data scientist who wants to learn data science and numerical programming concepts through hands-on, real-world project examples, this is the book for you. Whether you are brand new to data science or you are a seasoned expert, you will benefit from learning about the structure of real-world data science projects and the programming examples in R and Python. What You Will Learn Learn and understand the installation procedure and environment required for R and Python on various platforms Prepare data for analysis by implement various data science concepts such as acquisition, cleaning and munging through R and Python Build a predictive model and an exploratory model Analyze the results of your model and create reports on the acquired data Build various tree-based methods and Build random forest In Detail As increasing amounts of data are generated each year, the need to analyze and create value out of it is more important than ever. Companies that know what to do with their data and how to do it well will have a competitive advantage over companies that don't. Because of this, there will be an increasing demand for people that possess both the analytical and technical abilities to extract valuable insights from data and create valuable solutions that put those insights to use. Starting with the basics, this book covers how to set up your numerical programming environment, introduces you to the data science pipeline, and guides you through several data projects in a step-by-step format. By sequentially working through the steps in each chapter, you will quickly familiarize yourself with the process and learn how to apply it to a variety of situations with examples using the two most popular programming languages for data analysis—R and Python. Style and approach This step-by-step guide to data science is full of hands-on examples of real-world data science tasks. Each recipe focuses on a particular task involved in the data science pipeline, ranging from readying the dataset to analytics and visualization

[Copyright: 7cff4c7a73ab79a614ddcd2cae8e75e0](#)