Data Science And Big Data Analytics Discovering Analyzing Visualizing And Presenting Data

A concise introduction to the emerging field of data science, explaining its evolution, relation to machine learning, current uses, data infrastructure issues, and ethical challenges. The goal of data science is to improve decision making through the analysis of data. Today data science determines the ads we see online, the books and movies that are recommended to us online, which emails are filtered into our spam folders, and even how much we pay for health insurance. This volume in the MIT Press Essential Knowledge series offers a concise introduction to the emerging field of data science, explaining its evolution, current uses, data infrastructure issues, and ethical challenges. It has never been easier for organizations to gather, store, and process data. Use of data science is driven by the rise of big data and social media, the development of high-performance computing, and the emergence of such powerful methods for data analysis and modeling as deep learning. Data science encompasses a set of principles, problem definitions, algorithms, and processes for extracting non-obvious and useful patterns from large datasets. It is closely related to the fields of data mining and machine learning, but broader in scope. This book offers a brief history of the field, introduces fundamental data concepts, and describes the stages in a data science project. It considers data infrastructure and the challenges posed by integrating data from multiple sources, introduces the basics of machine learning, and discusses how to link machine learning expertise with real-world problems. The book also reviews ethical and legal issues, developments in data regulation, and computational approaches to preserving privacy. Finally, it considers the future impact of data science and offers principles for success in data science projects.

Do you want to understand what big data is all about, but you don't get all the hype? Are you intrigued by the idea of building a career around big data and data science, but you just don't understand it? If so, this book could be what you are looking for.

Recent technological advancements and other related factors and trends are contributing to the production of an astoundingly large and rapidly accelerating collection of data, or 'Big Data'. This data now allows us to examine urban and regional phenomena in ways that were previously not possible. Despite the tremendous potential 0f big data for regional science, its use and application in this context is fraught with issues and challenges. This book brings together leading contributors to present an interdisciplinary, agenda-setting and action-oriented platform for research and practice in the urban and regional community. This book provides a comprehensive, multidisciplinary and cutting-edge perspective on big data for regional science. Chapters contain a collection of research notes contributed by experts from all over the world with a wide array of disciplinary backgrounds. The content is organized along four themes: sources of big data; integration, processing and management of big data; analytics for big data; and, higher level policy and programmatic considerations. As well as concisely and comprehensively synthesising work done to date, the book also considers future challenges and prospects for the use of big data in regional science. Big Data for Regional Science provides a seminal contribution to the field of regional science and will appeal to a broad audience, including those at all levels of academia, industry, and government.

We are living in the dawn of what has been termed as the "Fourth Industrial Revolution," which is marked through the emergence of "cyber-physical systems" where software interfaces seamlessly over networks with physical systems, such as sensors, smartphones, vehicles, power grids or buildings, to create a new world of Internet of Things (IoT). Data and information are fuel of this new age where powerful analytics algorithms burn this fuel to generate decisions that are expected to create a smarter and more efficient world for all of us to live in. This new area of technology has been defined as Big Data Science and Analytics, and the industrial and academic communities are realizing this as a competitive technology that can generate significant new wealth and opportunity. Big data is defined as collections of datasets whose volume, velocity or variety is so large that it is difficult to store, manage, process and analyze the data using traditional databases and data processing tools. Big data science and analytics deals with collection, storage, processing and analysis of massive-scale data. Industry surveys, by Gartner and e-Skills, for instance, predict that there will be over 2 million job openings for engineers and scientists trained in the area of data science and analytics alone, and that the job market is in this area is growing at a 150 percent year-over-year growth rate. We have written this textbook, as part of our expanding "A Hands-On Approach"(TM) series, to meet this need at colleges and universities, and also for big data service providers who may be interested in offering a broader perspective of this emerging field to accompany their customer and developer training programs. The typical reader is expected to have completed a couple of courses in programming using traditional high-level languages at the college-level, and is either a senior or a beginning graduate student in one of the science, technology, engineering or mathematics (STEM) fields. An accompanying website for this book contains additional support for instruction and learning (www.big-data-analytics-book.com) The book is organized into three main parts, comprising a total of twelve chapters. Part I provides an introduction to big data, applications of big data, and big data science and analytics patterns and architectures. A novel data science and analytics application system design methodology is proposed and its realization through use of open-source big data frameworks is described. This methodology describes big data analytics applications as realization of the proposed Alpha, Beta, Gamma and Delta models, that comprise tools and frameworks for collecting and ingesting data from various sources into the big data analytics infrastructure, distributed filesystems and non-relational (NoSQL) databases for data storage, and processing frameworks for batch and real-time analytics. This new methodology forms the pedagogical foundation 0f this book. Part II introduces the reader to various tools and frameworks for big data analytics, and the architectural and programming aspects of these frameworks, with examples in Python. We describe Publish-Subscribe messaging frameworks (Kafka & Kinesis), Source-Sink connectors (Flume), Database Connectors (Sqoop), Messaging Queues (RabbitMQ, ZeroMQ,

RestMQ, Amazon SQS) and custom REST, WebSocket and MQTT-based connectors. The reader is introduced to data storage, batch and real-time analysis, and interactive querying frameworks including HDFS, Hadoop, MapReduce, YARN, Pig, Oozie, Spark, Solr, HBase, Storm, Spark Streaming, Spark SQL, Hive, Amazon Redshift and Google BigQuery. Also described are serving databases (MySQL, Amazon DynamoDB, Cassandra, MongoDB) and the Django Python web framework. Part III introduces the reader to various machine learning algorithms with examples using the Spark MLlib and H2O frameworks, and visualizations using frameworks such as Lightning, Pygal and Seaborn. While the term Big Data is open to varying interpretation, it is quite clear that the Volume, Velocity, and Variety (3Vs) of data have impacted every aspect of computational science and its applications. The volume of data is increasing at a phenomenal rate and a majority of it is unstructured. With big data, the volume is so large that processing it using traditional database and software techniques is difficult, if not impossible. The drivers are the ubiquitous sensors, devices, social networks and the all-pervasive web. Scientists are increasingly looking to derive insights from the massive quantity of data to create new knowledge. In common usage, Big Data has come to refer simply to the use of predictive analytics or other certain advanced methods to extract value from data, without any required magnitude thereon. Challenges include analysis, capture, curation, search, sharing, storage, transfer, visualization, and information privacy. While there are challenges, there are huge opportunities emerging in the fields of Machine Learning, Data Mining, Statistics, Human-Computer Interfaces and Distributed Systems to address ways to analyze and reason with this data. The edited volume focuses on the challenges and opportunities posed by "Big Data" in a variety of domains and how statistical techniques and innovative algorithms can help glean insights and accelerate discovery. Big data has the potential to help companies improve operations and make faster, more intelligent decisions. Review of big data research challenges from diverse areas of scientific endeavor Rich perspective on a range of data science issues from leading researchers Insight into the mathematical and statistical theory underlying the computational methods used to address big data analytics problems in a variety of domains

Learn how to embed data science, Big Data and AI in your organization's decision-making process and make your organization more data-driven, profitable, and intelligent in 10 steps. Book description This book covers every aspect of the implementation of data science, from the algorithms that make your decisions more refined, effective and faster to the people, skills, culture, and mindset required to make it happen. How do you set the right KPIs and targets? How are the best data-driven organizations structured? Why do you need a data warehouse or data lake? How do you manage a data science project? This book tackles every question relevant to implementing data science. Many organizations start by collecting data without a goal, but that data science approach is doomed to fail. This book takes you through the process of implementing data science from the ground floor all the way to the top. It all starts with the question: what do we want to achieve? It covers all the subsequent steps on a macro and micro level, from the process of registering data, to processing it, to the organization's response. All the relevant data science techniques and technologies are discussed, from algorithms and AI to the right management strategies. Based on many practical case studies and best practices, this book reveals what works and what doesn't. Benefit from the author's many years of experience in making organizations more intelligent and data-driven as a consultant and an educator. What you will learn - The most important benefits of data science. - The essential aspects of decision making and the role of data science. - How to determine the right KPIs and use them to manage effectively. - How to turn data into knowledge and information. - How to make your organization more agile. - The many types of algorithms that can be used to make more effective decisions on every level. - How to manage data science projects - who and what do you need to effectively implement data science? - How to design a data science roadmap. - And much, much more. Who is this book for This book is for every manager or professional, and all those who want to learn how to embed the effective use of data science in every facet of the organization. This comprehensive management handbook is a must-read for (business) consultants, business managers, Chief Data Officers (CDOs), CIOs, and other executives, project managers, Data Science consultants, Data Scientists, AI consultants, (business) controllers, quality managers, and BI consultants.

The amount of data in our world has been exploding, and analyzing large data sets—so called big data—will become a key basis of competition in business. Statisticians and researchers will be updating their analytic approaches, methods and research to meet the demands created by the availability of big data. The goal of this book is to show how advances in data science have the ability to fundamentally influence and improve organizational science and practice. This book is primarily designed for researchers and advanced undergraduate and graduate students in psychology, management and statistics.

This book presents a comprehensive and up-to-date treatise of a range of methodological and algorithmic issues. It also discusses implementations and case studies, identifies the best design practices, and assesses data analytics business models and practices in industry, health care, administration and business.Data science and big data go hand in hand and constitute a rapidly growing area of research and have attracted the attention of industry and business alike. The area itself has opened up promising new directions of fundamental and applied research and has led to interesting applications, especially those addressing the immediate need to deal with large repositories of data and building tangible, user-centric models of relationships in data. Data is the lifeblood of today's knowledge-driven economy.Numerous data science models are oriented towards end users and along with the regular requirements for accuracy (which are present in any modeling), come the requirements for ability to process huge and varying data sets as well as robustness, interpretability, and simplicity (transparency). Computational intelligence with its underlying methodologies and tools helps address data analytics needs.The book is of interest to those researchers and practitioners involved in data science, Internet engineering, computational intelligence, management, operations research, and knowledge-based systems.

The contemporary world lives on the data produced at an unprecedented speed through social networks and the internet of things (IoT). Data has been called the new global currency, and its rise is transforming entire industries, providing a wealth of opportunities. Applied data science research is necessary to derive useful information from big data for the effective and efficient utilization to solve real-world problems. A broad analytical set allied with strong business logic is fundamental in today's corporations. Organizations work to obtain competitive advantage by analyzing the data produced within and outside their organizational limits to support their decision-making processes. This book aims to provide an overview of the concepts, tools, and techniques behind the fields of data science and artificial intelligence (AI) applied to business and industries. The Handbook of Research on Applied Data Science and Artificial Intelligence in Business and Industry discusses all stages of data science to AI and their application to real problems across industries—from science and engineering to academia and commerce. This book brings together practice and science to build successful data solutions, showing how to uncover hidden patterns and leverage them to improve all aspects of business performance by making sense of data from both web and offline environments. Covering topics including applied AI, consumer behavior analytics, and machine learning, this text is essential for data scientists, IT specialists, managers, executives, software and computer engineers, researchers, practitioners, academicians, and students.

Written by renowned data science experts Foster Provost and Tom Fawcett, Data Science for Business introduces the fundamental principles of data science, and walks you through the "data-analytic thinking" necessary for extracting useful knowledge and business value from the data you collect. This guide also helps you understand the many data-mining techniques in use today. Based on an MBA course Provost has taught at New York University over the past ten years, Data Science for Business provides examples of real-world business problems to illustrate these principles. You'll not only learn how to improve communication between business stakeholders and data scientists, but also how participate intelligently in your company's data science projects. You'll also discover how to think data-analytically, and fully appreciate how data science methods can support business decision-making. Understand how data science fits in your organization—and how you can use it for competitive advantage Treat data as a business asset that requires careful investment if you're to gain real value Approach business problems data-analytically, using the data-mining process to gather good data in the most appropriate way Learn general concepts for actually extracting knowledge from data Apply data science principles when interviewing data science job candidates

Emerging Spatial Big Data (SBD) has transformative potential in solving many grand societal challenges such as water resource management, food security, disaster response, and transportation. However, significant computational challenges exist in analyzing SBD due to the unique spatial characteristics including spatial autocorrelation, anisotropy, heterogeneity, multiple scales and resolutions which is illustrated in this book. This book also discusses current techniques for, spatial big data science with a particular focus on classification techniques for earth observation imagery big data. Specifically, the authors introduce several recent spatial classification techniques, such as spatial decision trees and spatial ensemble learning. Several potential future research directions are also discussed. This book targets an interdisciplinary audience including computer scientists, practitioners and researchers working in the field of data mining, big data, as well as domain scientists working in earth science (e.g., hydrology, disaster), public safety and public health. Advanced level students in computer science will also find this book useful as a reference.

Perspectives on the varied challenges posed by big data for health, science, law, commerce, and politics. Big data is ubiquitous but heterogeneous. Big data can be used to tally clicks and traffic on web pages, find patterns in stock trades, track consumer preferences, identify linguistic correlations in large corpuses of texts. This book examines big data not as an undifferentiated whole but contextually, investigating the varied challenges posed by big data for health, science, law, commerce, and politics. Taken together, the chapters reveal a complex set of problems, practices, and policies. The advent of big data methodologies has challenged the theory-driven approach to scientific knowledge in favor of a data-driven one. Social media platforms and self-tracking tools change the way we see ourselves and others. The collection of data by corporations and government threatens privacy while promoting transparency. Meanwhile, politicians, policy makers, and ethicists are ill-prepared to deal with big data's ramifications. The contributors look at big data's effect on individuals as it exerts social control through monitoring, mining, and manipulation; big data and society, examining both its empowering and its constraining effects; big data and science, considering issues of data governance, provenance, reuse, and trust; and big data and organizations, discussing data responsibility, "data harm," and decision making. Contributors Ryan Abbott, Cristina Alaimo, Kent R. Anderson, Mark Andrejevic, Diane E. Bailey, Mike Bailey, Mark Burdon, Fred H. Cate, Jorge L. Contreras, Simon DeDeo, Hamid R. Ekbia, Allison Goodwell, Jannis Kallinikos, Inna Kouper, M. Lynne Markus, Michael Mattioli, Paul Ohm, Scott Peppet, Beth Plale, Jason Portenoy, Julie Rennecker, Katie Shilton, Dan Sholler, Cassidy R. Sugimoto, Isuru Suriarachchi, Jevin D. West

Big data is defined as collections of datasets whose volume, velocity or variety is so large that it is difficult to store, manage, process and analyze the data using traditional databases and data processing tools. We have written this textbook to meet this need at colleges and universities, and also for big data service providers.

Data science has never had more influence on the world. Large companies are now seeing the benefit of employing data scientists to interpret the vast amounts of data that now exists. However, the field is so new and is evolving so rapidly that the analysis produced can be haphazard at best. The 9 Pitfalls of Data Science shows us real-world examples of what can go wrong. Written to be an entertaining read, this invaluable guide investigates the all too common mistakes of data scientists - who can be plagued by lazy thinking, whims, hunches, and prejudices - and indicates how they have been at the root of many disasters, including the Great Recession. Gary Smith and Jay Cordes emphasise how scientific rigor and critical thinking skills are indispensable in this age of Big Data, as machines often find meaningless patterns that can lead to dangerous false conclusions. The 9 Pitfalls of Data Science is loaded with entertaining tales of both successful and misguided approaches to interpreting data, both grand successes and epic failures. These cautionary tales will not only help data scientists be more effective, but also help the public distinguish between good and bad data science.

Big Data and Social Science: Data Science Methods and Tools for Research and Practice, Second Edition shows how to apply data science to real-world problems, covering all stages of a data-intensive social science or policy project. Prominent leaders in the social sciences, statistics, and computer science as well as the field of data science provide a unique perspective on how to apply modern social science research principles and current analytical and computational tools. The text teaches you how to identify and collect appropriate data, apply data science

methods and tools to the data, and recognize and respond to data errors, biases, and limitations. Features Takes an accessible, hands-on approach to handling new types of data in the social sciences Presents the key data science tools in a non-intimidating way to both social and data scientists while keeping the focus on research questions and purposes Illustrates social science and data science principles through real-world problems Links computer science concepts to practical social science research Promotes good scientific practice Provides freely available data and code as well as practical programming exercises through Binder and GitHub New to the Second Edition Increased use of examples from different areas of social sciences New chapter on dealing with Bias and Fairness in Machine Learning models Expanded chapters focusing on Machine Learning and Text Analysis Revamped hands-on Jupyter notebooks to reinforce concepts covered in each chapter This classroom-tested book fills a major gap in graduate- and professional-level data science and social science education. It can be used to train a new generation of social data scientists to tackle real-world problems and improve the skills and competencies of applied social scientists and public policy practitioners. It empowers you to use the massive and rapidly growing amounts of available data to interpret economic and social activities in a scientific and rigorous manner.

This thoroughly revised guide demonstrates how the flexibility of the command line can help you become a more efficient and productive data scientist. You'll learn how to combine small yet powerful command-line tools to quickly obtain, scrub, explore, and model your data. To get you started, author Jeroen Janssens provides a Docker image packed with over 80 tools--useful whether you work with Windows, macOS, or Linux. You'll quickly discover why the command line is an agile, scalable, and extensible technology. Even if you're comfortable processing data with Python or R, you'll learn how to greatly improve your data science workflow by leveraging the command line's power. This book is ideal for data scientists, analysts, and engineers; software and machine learning engineers; and system administrators. Obtain data from websites, APIs, databases, and spreadsheets Perform scrub operations on text, CSV, HTM, XML, and JSON files Explore data, compute descriptive statistics, and create visualizations Manage your data science workflow Create reusable command-line tools from one-liners and existing Python or R code Parallelize and distribute data-intensive pipelines Model data with dimensionality reduction, clustering, regression, and classification algorithms

The guide to targeting and leveraging business opportunities using big data & analytics By leveraging big data & analytics, businesses create the potential to better understand, manage, and strategically exploiting the complex dynamics of customer behavior. Analytics in a Big Data World reveals how to tap into the powerful tool of data analytics to create a strategic advantage and identify new business opportunities. Designed to be an accessible resource, this essential book does not include exhaustive coverage of all analytical techniques, instead focusing on analytics techniques that really provide added value in business environments. The book draws on author Bart Baesens' expertise on the topics of big data, analytics and its applications in e.g. credit risk, marketing, and fraud to provide a clear roadmap for organizations that want to use data analytics to their advantage, but need a good starting point. Baesens has conducted extensive research on big data, analytics, customer relationship management, web analytics, fraud detection, and credit risk management, and uses this experience to bring clarity to a complex topic. Includes numerous case studies on risk management, fraud detection, customer relationship management, and web analytics Offers the results of research and the author's personal experience in banking, retail, and government Contains an overview of the visionary ideas and current developments on the strategic use of analytics for business Covers the topic of data analytics in easy-to-understand terms without an undo emphasis on mathematics and the minutiae of statistical analysis For organizations looking to enhance their capabilities via data analytics, this resource is the go-to reference for leveraging data to enhance business capabilities.

Statistical Foundations of Data Science gives a thorough introduction to commonly used statistical models, contemporary statistical machine learning techniques and algorithms, along with their mathematical insights and statistical theories. It aims to serve as a graduate-level textbook and a research monograph on high-dimensional statistics, sparsity and covariance learning, machine learning, and statistical inference. It includes ample exercises that involve both theoretical studies as well as empirical applications. The book begins with an introduction to the stylized features of big data and their impacts on statistical analysis. It then introduces multiple linear regression and expands the techniques of model building via nonparametric regression and kernel tricks. It provides a comprehensive account on sparsity explorations and model selections for multiple regression, generalized linear models, quantile regression, robust regression, hazards regression, among others. High-dimensional inference is also thoroughly addressed and so is feature screening. The book also provides a comprehensive account on high-dimensional covariance estimation, learning latent factors and hidden structures, as well as their applications to statistical estimation, inference, prediction and machine learning problems. It also introduces thoroughly statistical machine learning theory and methods for classification, clustering, and prediction. These include CART, random forests, boosting, support vector machines, clustering algorithms, sparse PCA, and deep learning.

Are you interested to get into the programming world? Do you want to learn and understand Python and Machine Learning? Python Machine Learning for Beginners is the guide for you. Python Machine Learning for Beginners is the ultimate guide for beginners looking to learn and understand how Python programming works. Python Machine Learning for Beginners is split up into easy to learn chapters that will help guide the readers through the early stages of Python programming. It's this thought out and systematic approach to learning which makes Python Machine Learning for Beginners such a sought-after resource for those that want to learn about Python programming and about Machine Learning using an object-oriented programming approach. Inside Python Machine Learning for Beginners you will discover: An introduction to Machine Learning The main concepts of Machine Learning The basics of Python for beginners Machine Learning with Python Data Processing, Analysis, and Visualizations Case studies and much more! Throughout the book, you will learn the basic concepts behind Python programming which is designed to introduce you to Python programming. You will learn about getting started, the keywords and statements, data types and type conversion. Along with different examples, there are also exercises to help ensure that the information sinks in. You will find this book an invaluable tool for starting and mastering Machine Learning using Python. Once you complete Python Machine Learning for Beginners, you will be more than prepared to take on any Python programming. Scroll back up to the top of this page and hit BUY IT NOW to get your copy of Python Machine Learning for Beginners! You won't regret it!

Data Science and Big Data Analytics is about harnessing the power of data for new insights. The book covers the breadth of activities and methods and tools that Data Scientists use. The content focuses on concepts, principles and practical applications that are applicable to any industry and technology environment, and the learning is supported and explained with examples

that you can replicate using open-source software. This book will help you: Become a contributor on a data science team Deploy a structured lifecycle approach to data analytics problems Apply appropriate analytic techniques and tools to analyzing big data Learn how to tell a compelling story with data to drive business action Prepare for EMC Proven Professional Data Science Certification Corresponding data sets are available from the book's page at Wiley which you can find on the Wiley site by searching for the ISBN 9781118876138. Get started discovering, analyzing, visualizing, and presenting data in a meaningful way today!

Both Traditional Students and Working Professionals Acquire the Skills to Analyze Social Problems. Big Data and Social Science: A Practical Guide to Methods and Tools shows how to apply data science to real-world problems in both research and the practice. The book provides practical guidance on combining methods and tools from computer science, statistics, and social science. This concrete approach is illustrated throughout using an important national problem, the quantitative study of innovation. The text draws on the expertise of prominent leaders in statistics, the social sciences, data science, and computer science to teach students how to use modern social science research principles as well as the best analytical and computational tools. It uses a real-world challenge to introduce how these tools are used to identify and capture appropriate data, apply data science models and tools to that data, and recognize and respond to data errors and limitations. For more information, including sample chapters and news, please visit the author's website.

The big data era is upon us: data are being generated, analyzed, and used at an unprecedented scale, and data-driven decision making is sweeping through all aspects of society. Since the value of data explodes when it can be linked and fused with other data, addressing the big data integration (BDI) challenge is critical to realizing the promise of big data. BDI differs from traditional data integration along the dimensions of volume, velocity, variety, and veracity. First, not only can data sources contain a huge volume of data, but also the number of data sources is now in the millions. Second, because of the rate at which newly collected data are made available, many of the data sources are very dynamic, and the number of data sources is also rapidly exploding. Third, data sources are extremely heterogeneous in their structure and content, exhibiting considerable variety even for substantially similar entities. Fourth, the data sources are of widely differing qualities, with significant differences in the coverage, accuracy and timeliness of data provided. This book explores the progress that has been made by the data integration community on the topics of schema alignment, record linkage and data fusion in addressing these novel challenges faced by big data integration. Each of these topics is covered in a systematic way: first starting with a quick tour of the topic in the context of traditional data integration, followed by a detailed, example-driven exposition of recent innovative techniques that have been proposed to address the BDI challenges of volume, velocity, variety, and veracity. Finally, it presents merging topics and opportunities that are specific to BDI, identifying promising directions for the data integration community.

Summary Introducing Data Science teaches you how to accomplish the fundamental tasks that occupy data scientists. Using the Python language and common Python libraries, you'll experience firsthand the challenges of dealing with data at scale and gain a solid foundation in data science. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Many companies need developers with data science skills to work on projects ranging from social media marketing to machine learning. Discovering what you need to learn to begin a career as a data scientist can seem bewildering. This book is designed to help you get started. About the Book Introducing Data ScienceIntroducing Data Science explains vital data science concepts and teaches you how to accomplish the fundamental tasks that occupy data scientists. You'll explore data visualization, graph databases, the use of NoSQL, and the data science process. You'll use the Python language and common Python libraries as you experience firsthand the challenges of dealing with data at scale. Discover how Python allows you to gain insights from data sets so big that they need to be stored on multiple machines, or from data moving so quickly that no single machine can handle it. This book gives you hands-on experience with the most popular Python data science libraries, Scikit-learn and StatsModels. After reading this book, you'll have the solid foundation you need to start a career in data science. What's Inside Handling large data Introduction to machine learning Using Python to work with data Writing data science algorithms About the Reader This book assumes you're comfortable reading code in Python or a similar language, such as C, Ruby, or JavaScript. No prior experience with data science is required. About the Authors Davy Cielen, Arno D. B. Meysman, and Mohamed Ali are the founders and managing partners of Optimately and Maiton, where they focus on developing data science projects and solutions in various sectors. Table of Contents Data science in a big data world The data science process Machine learning Handling large data on a single computer First steps in big data Join the NoSQL movement The rise of graph databases Text mining and text analytics Data visualization to the end user

This illuminating text/reference surveys the state of the art in data science, and provides practical guidance on big data analytics. Expert perspectives are provided by authoritative researchers and practitioners from around the world, discussing research developments and emerging trends, presenting case studies on helpful frameworks and innovative methodologies, and suggesting best practices for efficient and effective data analytics. Features: reviews a framework for fast data applications, a technique for complex event processing, and agglomerative approaches for the partitioning of networks; introduces a unified approach to data modeling and management, and a distributed computing perspective on interfacing physical and cyber worlds; presents techniques for machine learning for big data, and identifying duplicate records in data repositories; examines enabling technologies and tools for data mining; proposes frameworks for data extraction, and adaptive decision making and social media analysis.

Introduces professionals and scientists to statistics and machine learning using the programming language R Written by and for practitioners, this book provides an overall introduction to R, focusing on tools and methods commonly used in data science, and placing emphasis on practice and business use. It covers a wide range of topics in a single volume, including big data, databases, statistical machine learning, data wrangling, data visualization, and the reporting of results. The topics covered are all important for someone with a science/math background that is looking to quickly learn several practical technologies to enter or transition to the growing field of data science. The Big R-Book for Professionals: From Data Science to Learning Machines and Reporting with R includes nine parts, starting with an introduction to the subject and followed by an overview of R and elements of statistics. The third part revolves around data, while the fourth focuses on data wrangling. Part 5 teaches readers about exploring data. In Part 6 we learn to build models, Part 7 introduces the reader to the reality in companies, Part 8 covers reports and interactive applications and finally Part 9 introduces the reader to big data and performance computing. It also includes some helpful appendices. Provides a practical guide for non-

experts with a focus on business users Contains a unique combination of topics including an introduction to R, machine learning, mathematical models, data wrangling, and reporting Uses a practical tone and integrates multiple topics in a coherent framework Demystifies the hype around machine learning and AI by enabling readers to understand the provided models and program them in R Shows readers how to visualize results in static and interactive reports Supplementary materials includes PDF slides based on the book's content, as well as all the extracted R-code and is available to everyone on a Wiley Book Companion Site The Big R-Book is an excellent guide for science technology, engineering, or mathematics students who wish to make a successful transition from the academic world to the professional. It will also appeal to all young data scientists, quantitative analysts, and analytics professionals, as well as those who make mathematical models.

Big Data is a big topic, based on simple principles. Guided by leading expert in the field, David Stephenson, you will be amazed at how you can transform your company, and significantly improve KPIs across a broad range of business units and applications. Find out how an ecommerce company avoided two million product returns per year, how a newspaper saw triple-digit annual growth in digital subscriptions, how researchers in England learned to better detect pending cardiovascular problems, and how AI programs taught themselves to win games using techniques that even their human programmers didn't understand, all thanks to big data. Find out also how one company realized it could swap a million dollar hardware system with a twenty thousand dollar replacement. With simple and straightforward chapters that allow you to map examples onto your own business, Big Data Demystified will help you: · Know which data is most useful to collect now and why it's important to start collecting that data as soon as possible.· Understand big data and data science and how they can help you reach your business goals and gain competitive advantage.· Use big data to understand where you are now and how you can improve in the future. · Understand factors in choosing a big data system, including whether to go with cloud-based solutions.· Construct your big data team in a way that supports an effective strategy and helps make your business more data-driven. BIG DATA MAKES A BIG DIFFERENCE "Read this book! It is an essential guide to using data in a practical way that drives results." Ian McHenry, CEO Beyond Pricing "This is the book we've been missing: big data explained without the complexity." Marc Salomon, Professor in Decision Sciences and Dean at University of Amsterdam Business School "Big Data for the rest of us! I have never come across a book that is so full of practical advice, actionable examples and helpful explanations. Read this one book and start executing Big Data at your workplace tomorrow!"Tobias Wann CEO at @Leisure Group

Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured (labeled) and unstructured (unlabeled) data. It is the future of Artificial Intelligence (AI) and a necessity of the future to make things easier and more productive. In simple terms, data science is the discovery of data or uncovering hidden patterns (such as complex behaviors, trends, and inferences) from data. Moreover, Big Data analytics/data analytics are the analysis mechanisms used in data science by data scientists. Several tools, such as Hadoop, R, etc., are used to analyze this large amount of data to predict valuable information and for decision-making. Note that structured data can be easily analyzed by efficient (available) business intelligence tools, while most of the data (80% of data by 2020) is in an unstructured form that requires advanced analytics tools. But while analyzing this data, we face several concerns, such as complexity, scalability, privacy leaks, and trust issues. Data science helps us to extract meaningful information or insights from unstructured or complex or large amounts of data (available or stored virtually in the cloud). Data Science and Data Analytics: Opportunities and Challenges covers all possible areas, applications with arising serious concerns, and challenges in this emerging field in detail with a comparative analysis/taxonomy. FEATURES Gives the concept of data science, tools, and algorithms that exist for many useful applications Provides many challenges and opportunities in data science and data analytics that help researchers to identify research gaps or problems Identifies many areas and uses of data science in the smart era Applies data science to agriculture, healthcare, graph mining, education, security, etc. Academicians, data scientists, and stockbrokers from industry/business will find this book useful for designing optimal strategies to enhance their firm's productivity.

This book presents conjectural advances in big data analysis, machine learning and computational intelligence, as well as their potential applications in scientific computing. It discusses major issues pertaining to big data analysis using computational intelligence techniques, and the conjectural elements are supported by simulation and modelling applications to help address real-world problems. An extensive bibliography is provided at the end of each chapter. Further, the main content is supplemented by a wealth of figures, graphs, and tables, offering a valuable guide for researchers in the field of big data analytics and computational intelligence.

This book presents the outcomes of the 3rd IEEE/ACIS International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD 2018), which was held on July 10–12, 2018 in Kanazawa. The aim of the conference was to bring together researchers and scientists, businesspeople and entrepreneurs, teachers, engineers, computer users, and students to discuss the various fields of computer science, to share their experiences, and to exchange new ideas and information in a meaningful way. All aspects (theory, applications and tools) of computer and information science, the practical challenges encountered along the way, and the solutions adopted to solve them are all explored here. The conference organizers selected the best papers from among those accepted for presentation. The papers were chosen on the basis of review scores submitted by members of the program committee and subsequently underwent further rigorous review. Following this second round of review, 13 of the conference's most promising papers were selected for this Springer (SCI) book. We eagerly await the important contributions that we know these authors will make to the field of computer and information science.

Big Data: Principles and Paradigms captures the state-of-the-art research on the architectural aspects, technologies, and applications of Big Data. The book identifies potential future directions and technologies that facilitate insight into numerous scientific, business, and consumer applications. To help realize Big Data's full potential, the book addresses numerous challenges, offering the conceptual and technological solutions for tackling them. These challenges include life-cycle data management, large-scale storage, flexible processing infrastructure, data modeling, scalable machine learning, data analysis algorithms, sampling techniques, and privacy and ethical issues. Covers computational platforms supporting Big Data applications Addresses key principles underlying Big Data computing Examines key developments supporting next generation Big Data platforms Explores the challenges in Big Data computing and ways to overcome them Contains expert contributors from both academia and industry

Explains the mathematics, theory, and methods of Big Data as applied to finance and investing Data science has fundamentally changed Wall Street—applied mathematics and software code are increasingly driving finance and investment-decision tools. Big Data Science in Finance examines the mathematics, theory, and practical use of the revolutionary techniques that are transforming the industry. Designed for mathematically-advanced students and discerning financial practitioners alike, this energizing book presents new, cutting-edge content based on world-class research taught in the leading Financial Mathematics and Engineering programs in the world. Marco Avellaneda, a leader in quantitative finance, and quantitative methodology author Irene Aldridge help readers harness the power of Big Data. Comprehensive in scope, this book offers in-depth instruction on how to separate signal from noise, how to deal with missing data values, and how to utilize Big Data techniques in decision-making. Key topics include data clustering, data storage optimization, Big Data dynamics, Monte Carlo methods and their applications in Big Data analysis, and more. This valuable book: Provides a complete account of Big Data that includes proofs, step-by-step applications, and code samples Explains the difference between Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) Covers vital topics in the field in a clear, straightforward manner Compares, contrasts, and discusses Big Data and Small Data Includes Cornell University-tested educational materials such as lesson plans, end-of-chapter questions, and downloadable lecture slides Big Data Science in Finance: Mathematics and Applications is an important, up-to-date resource for students in economics, econometrics, finance, applied mathematics, industrial engineering, and business courses, and for investment managers, quantitative traders, risk and portfolio managers, and other financial practitioners.

"This book introduces you to R, RStudio, and the tidyverse, a collection of R packages designed to work together to make data science fast, fluent, and fun. Suitable for readers with no previous programming experience"--

Summary Introducing Data Science teaches you how to accomplish the fundamental tasks that occupy data scientists. Using the Python language and common Python libraries, you'll experience firsthand the challenges of dealing with data at scale and gain a solid foundation in data science. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Many companies need developers with data science skills to work on projects ranging from social media marketing to machine learning. Discovering what you need to learn to begin a career as a data scientist can seem bewildering. This book is designed to help you get started. About the Book Introducing Data ScienceIntroducing Data Science explains vital data science concepts and teaches you how to accomplish the fundamental tasks that occupy data scientists. You'll explore data visualization, graph databases, the use of NoSQL, and the data science process. You'll use the Python language and common Python libraries as you experience firsthand the challenges of dealing with data at scale. Discover how Python allows you to gain insights from data sets so big that they need to be stored on multiple machines, or from data moving so quickly that no single machine can handle it. This book gives you hands-on experience with the most popular Python data science libraries, Scikit-learn and StatsModels. After reading this book, you'll have the solid foundation you need to start a career in data science. What's Inside Handling large data Introduction to machine learning Using Python to work with data Writing data science algorithms About the Reader This book assumes you're comfortable reading code in Python or a similar language, such as C, Ruby, or JavaScript. No prior experience with data science is required. About the Authors Davy Cielen, Arno D. B. Meysman, and Mohamed Ali are the founders and managing partners of Optimately and Maiton, where they focus on developing data science projects and solutions in various sectors. Table of Contents Data science in a big data world The data science process Machine learning Handling large data on a single computer First steps in big data Join the NoSQL movement The rise of graph databases Text mining and text analytics Data visualization to the end user

This open access book covers the use of data science, including advanced machine learning, big data analytics, Semantic Web technologies, natural language processing, social media analysis, time series analysis, among others, for applications in economics and finance. In addition, it shows some successful applications of advanced data science solutions used to extract new knowledge from data in order to improve economic forecasting models. The book starts with an introduction on the use of data science technologies in economics and finance and is followed by thirteen chapters showing success stories of the application of specific data science methodologies, touching on particular topics related to novel big data sources and technologies for economic analysis (e.g. social media and news); big data models leveraging on supervised/unsupervised (deep) machine learning; natural language processing to build economic and financial indicators; and forecasting and nowcasting of economic variables through time series analysis. This book is relevant to all stakeholders involved in digital and data-intensive research in economics and finance, helping them to understand the main opportunities and challenges, become familiar with the latest methodological findings, and learn how to use and evaluate the performances of novel tools and frameworks. It primarily targets data scientists and business analysts exploiting data science technologies, and it will also be a useful resource to research students in disciplines and courses related to these topics. Overall, readers will learn modern and effective data science solutions to create tangible innovations for economic and financial applications.

Frontiers in Data Science deals with philosophical and practical results in Data Science. A broad definition of Data Science describes the process of analyzing data to transform data into insights. This also involves asking philosophical, legal and social questions in the context of data generation and analysis. In fact, Big Data also belongs to this universe as it comprises data gathering, data fusion and analysis when it comes to manage big data sets. A major goal of this book is to understand data science as a new scientific discipline rather than the practical aspects of data analysis alone.

We've all heard it: according to Hal Varian, statistics is the next sexy job. Five years ago, in What is Web 2.0, Tim O'Reilly said that "data is the next Intel Inside." But what does that statement mean? Why do we suddenly care about statistics and about data? This report examines the many sides of data science -- the technologies, the companies and the unique skill sets.The web is full of "data-driven apps." Almost any e-commerce application is a data-driven application. There's a database behind a web front end, and middleware that talks to a number of other databases and data services (credit card processing companies, banks, and so on). But merely using data isn't really what we mean by "data science." A data application acquires its value from the data itself, and creates more data as a result. It's not just an application with data; it's a data product. Data science enables the creation of data products.

This book is a refreshingly practical yet theoretically sound roadmap to leveraging data analytics and data science. The vast amount of data generated about us and our world is useless without plans and strategies that are designed to cope with its size and complexity, and which enable organizations to leverage the information to create value in marketing. Creating Value with Data Analytics in Marketing provides a nuanced view of big data developments and data science, arguing that big data is not a revolution but an evolution of the increasing availability of data that has been observed in recent times. Building on the authors' extensive academic and practical knowledge, this book aims to provide managers and analysts with strategic directions and practical analytical solutions on how to create value from existing and new big data. The second edition of this bestselling text has been fully updated in line with developments in the field and includes a selection of new, international cases and examples, exercises, techniques and methodologies. Tying data and analytics to specific goals and processes for implementation makes this essential reading for advanced undergraduate and postgraduate students and specialists of data analytics, marketing research, marketing management and customer relationship management. Online resources include chapter-by-chapter lecture slides and data sets and corresponding

R code for selected chapters.

Integrate big data into business to drive competitive advantage and sustainable success Big Data MBA brings insight and expertise to leveraging big data in business so you can harness the power of analytics and gain a true business advantage. Based on a practical framework with supporting methodology and hands-on exercises, this book helps identify where and how big data can help you transform your business. You'll learn how to exploit new sources of customer, product, and operational data, coupled with advanced analytics and data science, to optimize key processes, uncover monetization opportunities, and create new sources of competitive differentiation. The discussion includes guidelines for operationalizing analytics, optimal organizational structure, and using analytic insights throughout your organization's user experience to customers and front-end employees alike. You'll learn to "think like a data scientist" as you build upon the decisions your business is trying to make, the hypotheses you need to test, and the predictions you need to produce. Business stakeholders no longer need to relinquish control of data and analytics to IT. In fact, they must champion the organization's data collection and analysis efforts. This book is a primer on the business approach to analytics, providing the practical understanding you need to convert data into opportunity. Understand where and how to leverage big data Integrate analytics into everyday operations Structure your organization to drive analytic insights Optimize processes, uncover opportunities, and stand out from the rest Help business stakeholders to "think like a data scientist" Understand appropriate business application of different analytic techniques If you want data to transform your business, you need to know how to put it to use. Big Data MBA shows you how to implement big data and analytics to make better decisions.

Now that people are aware that data can make the difference in an election or a business model, data science as an occupation is gaining ground. But how can you get started working in a wide-ranging, interdisciplinary field that's so clouded in hype? This insightful book, based on Columbia University's Introduction to Data Science class, tells you what you need to know. In many of these chapter-long lectures, data scientists from companies such as Google, Microsoft, and eBay share new algorithms, methods, and models by presenting case studies and the code they use. If you're familiar with linear algebra, probability, and statistics, and have programming experience, this book is an ideal introduction to data science. Topics include: Statistical inference, exploratory data analysis, and the data science process Algorithms Spam filters, Naive Bayes, and data wrangling Logistic regression Financial modeling Recommendation engines and causality Data visualization Social networks and data journalism Data engineering, MapReduce, Pregel, and Hadoop Doing Data Science is collaboration between course instructor Rachel Schutt, Senior VP of Data Science at News Corp, and data science consultant Cathy O'Neil, a senior data scientist at Johnson Research Labs, who attended and blogged about the course.

"Many applications generate large datasets, like social networking and social influence programs, smart cities applications, smart house environments, Cloud applications, public web sites, scientific experiments and simulations, data warehouse, monitoring platforms, and e-government services. This book will primarily encompass practical approaches that advance research in all aspects of data processing, data analytics, data processing in Cloud/Edge/Fog systems, having a large variety of tools and software to manage them. The book focuses on focuses on topics covering algorithms, architectures, management models, high performance computing techniques and large-scale distributed systems"--

Copyright: acef2627ad5af5f43e4f2fa5bde3f24c