

Big Data Analytics In R

Big Data Analytics with R and Hadoop is a tutorial style book that focuses on all the powerful big data tasks that can be achieved by integrating R and Hadoop. This book is ideal for R developers who are looking for a way to perform big data analytics with Hadoop. This book is also aimed at those who know Hadoop and want to build some intelligent applications over Big data with R packages. It would be helpful if readers have basic knowledge of R.

Overview: This book is aimed at undergraduate students of computer science and engineering. The book will be useful companion for IT professionals to data analysts and decision makers responsible for driving strategic initiatives, and management graduates and business analysts, engaged in self-study. This book by Acharya unleashes the power of R as a statistical data analytics and visualization tool and introduces the learners to several data mining algorithms and chart forms / visualizations. It has good emphasis on 'asking the right questions'. Salient Features: - Exhaustive coverage includes installation of R and its package, getting accustomed to R interface and R commands, working with data from disparate data sources (.csv, JSON, XML, RDBMS etc.), getting conversant with classification, clustering, association rule mining, regression, text mining etc. - 12 Case studies namely Insurance Fraud Detection, Customer Insights Analysis, Sales Forecasting, Credit Card Spending by Customer Groups and Helping Retailers Predict In-store Customer Traffic - Pedagogical features: * 300+ chapter-end and check your progress questions for self-assessment * 200 Multiple-choice questions * 10+ hands-on practical exercises * Exhaustive illustrations

Data Science in Education Using R is the go-to reference for learning data science in the education field. The book answers questions like: What does a data scientist in education do? How do I get started learning R, the popular open-source statistical programming language? And what does a data analysis project in education look like? If you're just getting started with R in an education job, this is the book you'll want with you. This book gets you started with R by teaching the building blocks of programming that you'll use many times in your career. The book takes a "learn by doing" approach and offers eight analysis walkthroughs that show you a data analysis from start to finish, complete with code for you to practice with. The book finishes with how to get involved in the data science community and how to integrate data science in your education job. This book will be an essential resource for education professionals and researchers looking to increase their data analysis skills as part of their professional and academic development.

Load, wrangle, and analyze your data using the world's most powerful statistical programming language About This Book Load, manipulate and analyze data from different sources Gain a deeper understanding of fundamentals of applied statistics A practical guide to performing data analysis in practice Who This Book Is For Whether you are learning data analysis for the first time, or you want to deepen the understanding you already have, this book will prove to an invaluable resource. If you are looking for a book to bring you all the way through the fundamentals to the application of advanced and effective analytics methodologies, and have some prior programming experience and a mathematical background, then this is for you. What You Will Learn Navigate the R environment Describe and visualize the behavior of data and relationships between data Gain a thorough understanding of statistical reasoning and sampling Employ hypothesis tests to draw inferences from your data Learn Bayesian methods for estimating parameters Perform regression to predict continuous variables Apply powerful classification methods to predict categorical data Handle missing data gracefully using multiple imputation Identify and manage problematic data points Employ parallelization and Rcpp to scale your analyses to larger data Put best practices into effect to make your job easier and facilitate reproducibility In Detail Frequently the tool of choice for academics, R has spread deep into the private sector and can be found in the production pipelines at some of the most

advanced and successful enterprises. The power and domain-specificity of R allows the user to express complex analytics easily, quickly, and succinctly. With over 7,000 user contributed packages, it's easy to find support for the latest and greatest algorithms and techniques. Starting with the basics of R and statistical reasoning, *Data Analysis with R* dives into advanced predictive analytics, showing how to apply those techniques to real-world data though with real-world examples. Packed with engaging problems and exercises, this book begins with a review of R and its syntax. From there, get to grips with the fundamentals of applied statistics and build on this knowledge to perform sophisticated and powerful analytics. Solve the difficulties relating to performing data analysis in practice and find solutions to working with "messy data", large data, communicating results, and facilitating reproducibility. This book is engineered to be an invaluable resource through many stages of anyone's career as a data analyst. Style and approach Learn data analysis using engaging examples and fun exercises, and with a gentle and friendly but comprehensive "learn-by-doing" approach. 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 Addressing a broad range of big data analytics in cross-disciplinary applications, this essential handbook focuses on the statistical prospects offered by recent developments in this field. To do so, it covers statistical methods for high-dimensional problems, algorithmic designs, computation tools, analysis flows and the software-hardware co-designs that are needed to support insightful discoveries from big data. The book is primarily intended for statisticians, computer experts, engineers and application developers interested in using big data analytics with statistics. Readers should have a solid background in statistics and computer science. This book examines the Internet of Things (IoT) and Data Analytics from a technical, application, and business point of view. Internet of Things and Data Analytics Handbook describes essential technical knowledge, building blocks, processes, design principles, implementation, and marketing for IoT projects. It provides readers with knowledge in planning, designing, and implementing IoT projects. The book is written by experts on the subject matter, including international experts from nine countries in the consumer and enterprise fields of IoT. The text starts with an overview and anatomy of IoT, ecosystem of IoT, communication protocols, networking, and available hardware, both present and future applications and transformations, and business models. The text also addresses big data analytics, machine learning, cloud computing, and consideration of sustainability that are essential to be both

socially responsible and successful. Design and implementation processes are illustrated with best practices and case studies in action. In addition, the book: Examines cloud computing, data analytics, and sustainability and how they relate to IoT overs the scope of consumer, government, and enterprise applications Includes best practices, business model, and real-world case studies Hwaiyu Geng, P.E., is a consultant with Amica Research (www.AmicaResearch.org, Palo Alto, California), promoting green planning, design, and construction projects. He has had over 40 years of manufacturing and management experience, working with Westinghouse, Applied Materials, Hewlett Packard, and Intel on multi-million high-tech projects. He has written and presented numerous technical papers at international conferences. Mr. Geng, a patent holder, is also the editor/author of *Data Center Handbook* (Wiley, 2015).

This invaluable addition to any data scientist's library shows you how to apply the R programming language and useful statistical techniques to everyday business situations as well as how to effectively present results to audiences of all levels. To answer the ever-increasing demand for machine learning and analysis, this new edition boasts additional R tools, modeling techniques, and more. *Practical Data Science with R, Second Edition* takes a practice-oriented approach to explaining basic principles in the ever-expanding field of data science. You'll jump right to real-world use cases as you apply the R programming language and statistical analysis techniques to carefully explained examples based in marketing, business intelligence, and decision support. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.

If you're like most R users, you have deep knowledge and love for statistics. But as your organization continues to collect huge amounts of data, adding tools such as Apache Spark makes a lot of sense. With this practical book, data scientists and professionals working with large-scale data applications will learn how to use Spark from R to tackle big data and big compute problems. Authors Javier Luraschi, Kevin Kuo, and Edgar Ruiz show you how to use R with Spark to solve different data analysis problems. This book covers relevant data science topics, cluster computing, and issues that should interest even the most advanced users. Analyze, explore, transform, and visualize data in Apache Spark with R Create statistical models to extract information and predict outcomes; automate the process in production-ready workflows Perform analysis and modeling across many machines using distributed computing techniques Use large-scale data from multiple sources and different formats with ease from within Spark Learn about alternative modeling frameworks for graph processing, geospatial analysis, and genomics at scale Dive into advanced topics including custom transformations, real-time data processing, and creating custom Spark extensions

Learn how to program by diving into the R language, and then use your newfound skills to solve practical data science problems. With this book, you'll learn how to load data, assemble and disassemble data objects, navigate R's environment system, write your own functions, and use all of R's programming tools. RStudio Master Instructor Garrett Golemund not only teaches you how to program, but also shows you how to get more from R than just visualizing and modeling data. You'll gain valuable programming skills and support your work as a data scientist at the same time. Work hands-on with three practical data analysis projects based on casino games Store, retrieve, and change data values in your computer's memory Write programs and simulations that outperform those written by typical R users Use R programming tools such as if else statements, for loops, and S3 classes Learn how to write lightning-fast vectorized R code Take advantage of R's package system and debugging tools Practice and apply R programming concepts as you learn them

Gain the key language concepts and programming techniques of Scala in the context of big data analytics and Apache Spark. The book begins by introducing you to Scala and establishes a firm contextual understanding of why you should learn this language, how it

stands in comparison to Java, and how Scala is related to Apache Spark for big data analytics. Next, you'll set up the Scala environment ready for examining your first Scala programs. This is followed by sections on Scala fundamentals including mutable/immutable variables, the type hierarchy system, control flow expressions and code blocks. The author discusses functions at length and highlights a number of associated concepts such as functional programming and anonymous functions. The book then delves deeper into Scala's powerful collections system because many of Apache Spark's APIs bear a strong resemblance to Scala collections. Along the way you'll see the development life cycle of a Scala program. This involves compiling and building programs using the industry-standard Scala Build Tool (SBT). You'll cover guidelines related to dependency management using SBT as this is critical for building large Apache Spark applications. Scala Programming for Big Data Analytics concludes by demonstrating how you can make use of the concepts to write programs that run on the Apache Spark framework. These programs will provide distributed and parallel computing, which is critical for big data analytics. What You Will Learn See the fundamentals of Scala as a general-purpose programming language Understand functional programming and object-oriented programming constructs in Scala Use Scala collections and functions Develop, package and run Apache Spark applications for big data analytics Who This Book Is For Data scientists, data analysts and data engineers who intend to use Apache Spark for large-scale analytics. /div

Get to grips with key data visualization and predictive analytic skills using R About This Book Acquire predictive analytic skills using various tools of R Make predictions about future events by discovering valuable information from data using R Comprehensible guidelines that focus on predictive model design with real-world data Who This Book Is For If you are a statistician, chief information officer, data scientist, ML engineer, ML practitioner, quantitative analyst, and student of machine learning, this is the book for you. You should have basic knowledge of the use of R. Readers without previous experience of programming in R will also be able to use the tools in the book. What You Will Learn Customize R by installing and loading new packages Explore the structure of data using clustering algorithms Turn unstructured text into ordered data, and acquire knowledge from the data Classify your observations using Naive Bayes, k-NN, and decision trees Reduce the dimensionality of your data using principal component analysis Discover association rules using Apriori Understand how statistical distributions can help retrieve information from data using correlations, linear regression, and multilevel regression Use PMML to deploy the models generated in R In Detail R is statistical software that is used for data analysis. There are two main types of learning from data: unsupervised learning, where the structure of data is extracted automatically; and supervised learning, where a labeled part of the data is used to learn the relationship or scores in a target attribute. As important information is often hidden in a lot of data, R helps to extract that information with its many standard and cutting-edge statistical functions. This book is packed with easy-to-follow guidelines that explain the workings of the many key data mining tools of R, which are used to discover knowledge from your data. You will learn how to perform key predictive analytics tasks using R, such as train and test predictive models for classification and regression tasks, score new data sets and so on. All chapters will guide you in acquiring the skills in a practical way. Most chapters also include a theoretical introduction that will sharpen your understanding of the subject matter and invite you to go further. The book familiarizes you with the most common data mining tools of R, such as k-means, hierarchical regression, linear regression, association rules, principal component analysis, multilevel modeling, k-NN, Naive Bayes, decision trees, and text mining. It also provides a description of visualization techniques using the basic visualization tools of R as well as lattice for visualizing patterns in data organized in groups. This book is invaluable for anyone fascinated by the data mining opportunities offered by GNU R and its packages. Style and approach This is a practical book, which analyzes compelling data about life, health, and death with the help of

tutorials. It offers you a useful way of interpreting the data that's specific to this book, but that can also be applied to any other data.

Collecting, analyzing, and extracting valuable information from a large amount of data requires easily accessible, robust, computational and analytical tools. *Data Mining and Business Analytics with R* utilizes the open source software R for the analysis, exploration, and simplification of large high-dimensional data sets. As a result, readers are provided with the needed guidance to model and interpret complicated data and become adept at building powerful models for prediction and classification. Highlighting both underlying concepts and practical computational skills, *Data Mining and Business Analytics with R* begins with coverage of standard linear regression and the importance of parsimony in statistical modeling. The book includes important topics such as penalty-based variable selection (LASSO); logistic regression; regression and classification trees; clustering; principal components and partial least squares; and the analysis of text and network data. In addition, the book presents:

- A thorough discussion and extensive demonstration of the theory behind the most useful data mining tools
- Illustrations of how to use the outlined concepts in real-world situations
- Readily available additional data sets and related R code allowing readers to apply their own analyses to the discussed materials
- Numerous exercises to help readers with computing skills and deepen their understanding of the material

Data Mining and Business Analytics with R is an excellent graduate-level textbook for courses on data mining and business analytics. The book is also a valuable reference for practitioners who collect and analyze data in the fields of finance, operations management, marketing, and the information sciences.

Utilize R to uncover hidden patterns in your Big Data About This Book Perform computational analyses on Big Data to generate meaningful results Get a practical knowledge of R programming language while working on Big Data platforms like Hadoop, Spark, H2O and SQL/NoSQL databases, Explore fast, streaming, and scalable data analysis with the most cutting-edge technologies in the market Who This Book Is For This book is intended for Data Analysts, Scientists, Data Engineers, Statisticians, Researchers, who want to integrate R with their current or future Big Data workflows. It is assumed that readers have some experience in data analysis and understanding of data management and algorithmic processing of large quantities of data, however they may lack specific skills related to R. What You Will Learn Learn about current state of Big Data processing using R programming language and its powerful statistical capabilities Deploy Big Data analytics platforms with selected Big Data tools supported by R in a cost-effective and time-saving manner Apply the R language to real-world Big Data problems on a multi-node Hadoop cluster, e.g. electricity consumption across various socio-demographic indicators and bike share scheme usage Explore the compatibility of R with Hadoop, Spark, SQL and NoSQL databases, and H2O platform In Detail Big Data analytics is the process of examining large and complex data sets that often exceed the computational capabilities. R is a leading programming language of data science, consisting of powerful functions to tackle all problems related to Big Data processing. The book will begin with a brief introduction to the Big Data world and its current industry standards. With introduction to the R language and presenting its development, structure, applications in real world, and its shortcomings. Book will progress towards revision of major R functions for data management and transformations. Readers will be introduced to Cloud based Big Data solutions (e.g. Amazon EC2 instances and Amazon RDS, Microsoft Azure and its HDInsight clusters) and also provide guidance on R connectivity with relational and non-relational databases such as MongoDB and HBase etc. It will further expand to include Big Data tools such as Apache Hadoop ecosystem, HDFS and MapReduce frameworks. Also other R compatible tools such as Apache Spark, its machine learning library Spark MLlib, as well as H2O. Style and approach This book will serve as a practical guide to tackling Big Data problems using R programming language and its statistical environment. Each section of the book will present

you with concise and easy-to-follow steps on how to process, transform and analyse large data sets.

Big Data Analytics with R Packt Publishing Ltd

Using data from one season of NBA games, *Basketball Data Science: With Applications in R* is the perfect book for anyone interested in learning and applying data analytics in basketball.

Whether assessing the spatial performance of an NBA player's shots or doing an analysis of the impact of high pressure game situations on the probability of scoring, this book discusses a variety of case studies and hands-on examples using a custom R package. The codes are supplied so readers can reproduce the analyses themselves or create their own. Assuming a basic statistical knowledge, *Basketball Data Science with R* is suitable for students, technicians, coaches, data analysts and applied researchers. Features:

- One of the first books to provide statistical and data mining methods for the growing field of analytics in basketball.
- Presents tools for modelling graphs and figures to visualize the data.
- Includes real world case studies and examples, such as estimations of scoring probability using the Golden State Warriors as a test case.
- Provides the source code and data so readers can do their own analyses on NBA teams and players.

Gain sharp insights into your data and solve real-world data science problems with R—from data munging to modeling and visualization

About This Book Handle your data with precision and care for optimal business intelligence Restructure and transform your data to inform decision-making Packed with practical advice and tips to help you get to grips with data mining

Who This Book Is For If you are a data scientist or R developer who wants to explore and optimize your use of R's advanced features and tools, this is the book for you. A basic knowledge of R is required, along with an understanding of database logic.

What You Will Learn

- Connect to and load data from R's range of powerful databases
- Successfully fetch and parse structured and unstructured data
- Transform and restructure your data with efficient R packages
- Define and build complex statistical models with glm
- Develop and train machine learning algorithms
- Visualize social networks and graph data
- Deploy supervised and unsupervised classification algorithms

Discover how to visualize spatial data with R

In Detail R is an essential language for sharp and successful data analysis. Its numerous features and ease of use make it a powerful way of mining, managing, and interpreting large sets of data. In a world where understanding big data has become key, by mastering R you will be able to deal with your data effectively and efficiently. This book will give you the guidance you need to build and develop your knowledge and expertise. Bridging the gap between theory and practice, this book will help you to understand and use data for a competitive advantage. Beginning with taking you through essential data mining and management tasks such as munging, fetching, cleaning, and restructuring, the book then explores different model designs and the core components of effective analysis. You will then discover how to optimize your use of machine learning algorithms for classification and recommendation systems beside the traditional and more recent statistical methods.

Style and approach Covering the essential tasks and skills within data science, *Mastering Data Analysis* provides you with solutions to the challenges of data science. Each section gives you a theoretical overview before demonstrating how to put the theory to work with real-world use cases and hands-on examples.

Leverage the power of advanced analytics and predictive modeling in Tableau using the statistical powers of R

About This Book A comprehensive guide that will bring out the creativity in you to visualize the results of complex calculations using Tableau and R

Combine Tableau analytics and visualization with the power of R using this step-by-step guide

Wondering how R can be used with Tableau? This book is your one-stop solution.

Who This Book Is For This book will appeal to Tableau users who want to go beyond the Tableau interface and deploy the full potential of Tableau, by using R to perform advanced analytics with Tableau. A basic familiarity with R is useful but not compulsory, as the book will start off with concrete examples

of R and will move quickly into more advanced spheres of analytics using online data sources to support hands-on learning. Those R developers who want to integrate R in Tableau will also benefit from this book. What You Will Learn Integrate Tableau's analytics with the industry-standard, statistical prowess of R. Make R function calls in Tableau, and visualize R functions with Tableau using RServe. Use the CRISP-DM methodology to create a roadmap for analytics investigations. Implement various supervised and unsupervised learning algorithms in R to return values to Tableau. Make quick, cogent, and data-driven decisions for your business using advanced analytical techniques such as forecasting, predictions, association rules, clustering, classification, and other advanced Tableau/R calculated field functions. In Detail Tableau and R offer accessible analytics by allowing a combination of easy-to-use data visualization along with industry-standard, robust statistical computation. Moving from data visualization into deeper, more advanced analytics? This book will intensify data skills for data viz-savvy users who want to move into analytics and data science in order to enhance their businesses by harnessing the analytical power of R and the stunning visualization capabilities of Tableau. Readers will come across a wide range of machine learning algorithms and learn how descriptive, prescriptive, predictive, and visually appealing analytical solutions can be designed with R and Tableau. In order to maximize learning, hands-on examples will ease the transition from being a data-savvy user to a data analyst using sound statistical tools to perform advanced analytics. By the end of this book, you will get to grips with advanced calculations in R and Tableau for analytics and prediction with the help of use cases and hands-on examples. Style and approach Tableau (uniquely) offers excellent visualization combined with advanced analytics; R is at the pinnacle of statistical computational languages. When you want to move from one view of data to another, backed up by complex computations, the combination of R and Tableau makes the perfect solution. This example-rich guide will teach you how to combine these two to perform advanced analytics by integrating Tableau with R and create beautiful data visualizations.

The Oracle Press Guide to Big Data Analytics using R Cowritten by members of the Big Data team at Oracle, this Oracle Press book focuses on analyzing data with R while making it scalable using Oracle's R technologies. Using R to Unlock the Value of Big Data provides an introduction to open source R and describes issues with traditional R and database interaction. The book then offers in-depth coverage of Oracle's strategic R offerings: Oracle R Enterprise, Oracle R Distribution, ROracle, and Oracle R Connector for Hadoop. You can practice your new skills using the end-of-chapter exercises.

Big data, analytics, and artificial intelligence are revolutionizing work, management, and lifestyles and are becoming disruptive technologies for healthcare, e-commerce, and web services. However, many fundamental, technological, and managerial issues for developing and applying intelligent big data analytics in these fields have yet to be addressed. Managerial Perspectives on Intelligent Big Data Analytics is a collection of innovative research that discusses the integration and application of artificial intelligence, business intelligence, digital transformation, and intelligent big data analytics from a perspective of computing, service, and management. While highlighting topics including e-commerce, machine learning, and fuzzy logic, this book is ideally designed for students, government officials, data scientists, managers, consultants, analysts, IT specialists, academicians, researchers, and industry professionals in fields that include big data, artificial intelligence, computing, and commerce. A handy reference guide for data analysts and data scientists to help to obtain value from big data analytics using Spark on Hadoop clusters About This Book This book is based on the latest 2.0 version of Apache Spark and 2.7 version of Hadoop integrated with most commonly used tools. Learn all Spark stack components including latest topics such as DataFrames, DataSets, GraphFrames, Structured Streaming, DataFrame based ML Pipelines and SparkR. Integrations with frameworks such as HDFS, YARN and tools such as Jupyter, Zeppelin, NiFi,

Mahout, HBase Spark Connector, GraphFrames, H2O and Hivemall. Who This Book Is For
Though this book is primarily aimed at data analysts and data scientists, it will also help architects, programmers, and practitioners. Knowledge of either Spark or Hadoop would be beneficial. It is assumed that you have basic programming background in Scala, Python, SQL, or R programming with basic Linux experience. Working experience within big data environments is not mandatory. What You Will Learn Find out and implement the tools and techniques of big data analytics using Spark on Hadoop clusters with wide variety of tools used with Spark and Hadoop Understand all the Hadoop and Spark ecosystem components Get to know all the Spark components: Spark Core, Spark SQL, DataFrames, DataSets, Conventional and Structured Streaming, MLlib, ML Pipelines and Graphx See batch and real-time data analytics using Spark Core, Spark SQL, and Conventional and Structured Streaming Get to grips with data science and machine learning using MLlib, ML Pipelines, H2O, Hivemall, Graphx, SparkR and Hivemall. In Detail Big Data Analytics book aims at providing the fundamentals of Apache Spark and Hadoop. All Spark components – Spark Core, Spark SQL, DataFrames, Data sets, Conventional Streaming, Structured Streaming, MLlib, Graphx and Hadoop core components – HDFS, MapReduce and Yarn are explored in greater depth with implementation examples on Spark + Hadoop clusters. It is moving away from MapReduce to Spark. So, advantages of Spark over MapReduce are explained at great depth to reap benefits of in-memory speeds. DataFrames API, Data Sources API and new Data set API are explained for building Big Data analytical applications. Real-time data analytics using Spark Streaming with Apache Kafka and HBase is covered to help building streaming applications. New Structured streaming concept is explained with an IOT (Internet of Things) use case. Machine learning techniques are covered using MLlib, ML Pipelines and SparkR and Graph Analytics are covered with GraphX and GraphFrames components of Spark. Readers will also get an opportunity to get started with web based notebooks such as Jupyter, Apache Zeppelin and data flow tool Apache NiFi to analyze and visualize data. Style and approach This step-by-step pragmatic guide will make life easy no matter what your level of experience. You will deep dive into Apache Spark on Hadoop clusters through ample exciting real-life examples. Practical tutorial explains data science in simple terms to help programmers and data analysts get started with Data Science

We are in an age of big data where all of our everyday interactions and transactions generate data. Much of this data is spatial – it is collected some-where – and identifying analytical insight from trends and patterns in these increasing rich digital footprints presents a number of challenges. Whilst other books describe different flavours of Data Analytics in R and other programming languages, there are none that consider Spatial Data (ie the location attached to data), or that consider issues of inference, linking Big Data, Geography, GIS, Mapping and Spatial Analytics. This is a 'learning by doing' text book, building on the previous book by the same authors, An Introduction to R for Spatial Analysis and Mapping. It details the theoretical issues in analyses of Big Spatial Data and developing practical skills in the reader for addressing these with confidence.

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.

"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"--

Chapter 1 - Basics of R, Chapter 2 - Data Types in R , Chapter 3 - Data Preparation. Chapter 4 - Graphics using R, Chapter 5 - Statistical Analysis Using R, Chapter 6 - Data Mining Using R, Chapter 7 - Case Studies. Huge volumes of data are being generated by many sources like commercial enterprises, scientific domains and general public daily. According to a recent research, data production will be 44 times greater in 2020 than it was in 2010. Data being a vital resource for business organizations and other domains like education, health, manufacturing etc., its management and analysis is becoming increasingly important. This data, due to its volume, variety and velocity, often referred to as Big Data, also includes highly unstructured data in the form of textual documents, web pages, graphical information and social media comments. Since Big Data is characterised by massive sample sizes, high dimensionality and intrinsic heterogeneity, traditional approaches to data management, visualisation and analytics are no longer satisfactorily applicable. There is therefore an urgent need for newer tools, better frameworks and workable methodologies for such data to be appropriately categorised, logically segmented, efficiently analysed and securely managed. This requirement has resulted in an emerging new discipline of Data Science that is now gaining much attention with researchers and practitioners in the field of Data Analytics. Data science revolves around two giants: Big Data analytics and Deep Learning. It is becoming challenging to handle and retrieve useful information due to how fast data is expanding. This book presents the technologies and tools to simplify and streamline the formation of Big Data as well as Deep Learning systems. This book discusses how Big Data and Deep Learning hold the potential to significantly increase data understanding and decision-making. It also covers numerous applications in healthcare, education, communication, media, and entertainment. Integrating Deep Learning Algorithms to Overcome Challenges in Big Data Analytics offers innovative platforms for integrating Big Data and Deep Learning and presents issues related to adequate data storage, semantic indexing, data tagging, and fast information retrieval. FEATURES Provides insight into the skill set that leverages one's strength to act as a good data analyst Discusses how Big Data and Deep Learning hold the potential to significantly increase data understanding and help in decision-making Covers numerous potential applications in healthcare, education, communication, media, and entertainment Offers innovative platforms for integrating Big Data and Deep Learning Presents issues related to adequate data storage, semantic indexing, data tagging, and fast information retrieval from Big

Data This book is aimed at industry professionals, academics, research scholars, system modelers, and simulation experts.

Over 100 hands-on recipes to effectively solve real-world data problems using the most popular R packages and techniques About This Book Gain insight into how data scientists collect, process, analyze, and visualize data using some of the most popular R packages Understand how to apply useful data analysis techniques in R for real-world applications An easy-to-follow guide to make the life of data scientist easier with the problems faced while performing data analysis Who This Book Is For This book is for those who are already familiar with the basic operation of R, but want to learn how to efficiently and effectively analyze real-world data problems using practical R packages. What You Will Learn Get to know the functional characteristics of R language Extract, transform, and load data from heterogeneous sources Understand how easily R can confront probability and statistics problems Get simple R instructions to quickly organize and manipulate large datasets Create professional data visualizations and interactive reports Predict user purchase behavior by adopting a classification approach Implement data mining techniques to discover items that are frequently purchased together Group similar text documents by using various clustering methods In Detail This cookbook offers a range of data analysis samples in simple and straightforward R code, providing step-by-step resources and time-saving methods to help you solve data problems efficiently. The first section deals with how to create R functions to avoid the unnecessary duplication of code. You will learn how to prepare, process, and perform sophisticated ETL for heterogeneous data sources with R packages. An example of data manipulation is provided, illustrating how to use the “dplyr” and “data.table” packages to efficiently process larger data structures. We also focus on “ggplot2” and show you how to create advanced figures for data exploration. In addition, you will learn how to build an interactive report using the “ggvis” package. Later chapters offer insight into time series analysis on financial data, while there is detailed information on the hot topic of machine learning, including data classification, regression, clustering, association rule mining, and dimension reduction. By the end of this book, you will understand how to resolve issues and will be able to comfortably offer solutions to problems encountered while performing data analysis. Style and approach This easy-to-follow guide is full of hands-on examples of data analysis with R. Each topic is fully explained beginning with the core concept, followed by step-by-step practical examples, and concluding with detailed explanations of each concept used.

Big Data Analytics Made Easy is a must-read for everybody as it explains the power of Analytics in a simple and logical way along with an end to end code in R. Even if you are a novice in Big Data Analytics, you will still be able to understand the concepts explained in this book. If you are already working in Analytics and dealing with Big Data, you will still find this book useful, as it covers exhaustive Data Mining Techniques, which are considered to be Advanced topics. It covers Machine Learning concepts and provides in-depth knowledge on unsupervised as well as supervised Learning, which is very important for decision-making. The toughest Data Analytics concepts are made simpler, It features examples from all the domains so that the reader gets connected to the book easily. This book is like a personal trainer that will help you master the Art of Data Science.

Get command of your organizational Big Data using the power of data science and analytics Key Features A perfect companion to boost your Big Data storing, processing, analyzing skills to help you take informed business decisions Work with the best tools such as Apache Hadoop, R, Python, and Spark for NoSQL platforms to perform massive online analyses Get expert tips on statistical inference, machine learning, mathematical modeling, and data visualization for Big Data Book Description Big Data analytics relates to the strategies used by organizations to collect, organize and analyze large amounts of data to uncover valuable business insights that otherwise cannot be analyzed through traditional systems. Crafting an

enterprise-scale cost-efficient Big Data and machine learning solution to uncover insights and value from your organization's data is a challenge. Today, with hundreds of new Big Data systems, machine learning packages and BI Tools, selecting the right combination of technologies is an even greater challenge. This book will help you do that. With the help of this guide, you will be able to bridge the gap between the theoretical world of technology with the practical ground reality of building corporate Big Data and data science platforms. You will get hands-on exposure to Hadoop and Spark, build machine learning dashboards using R and R Shiny, create web-based apps using NoSQL databases such as MongoDB and even learn how to write R code for neural networks. By the end of the book, you will have a very clear and concrete understanding of what Big Data analytics means, how it drives revenues for organizations, and how you can develop your own Big Data analytics solution using different tools and methods articulated in this book. What you will learn - Get a 360-degree view into the world of Big Data, data science and machine learning - Broad range of technical and business Big Data analytics topics that caters to the interests of the technical experts as well as corporate IT executives - Get hands-on experience with industry-standard Big Data and machine learning tools such as Hadoop, Spark, MongoDB, KDB+ and R - Create production-grade machine learning BI Dashboards using R and R Shiny with step-by-step instructions - Learn how to combine open-source Big Data, machine learning and BI Tools to create low-cost business analytics applications - Understand corporate strategies for successful Big Data and data science projects - Go beyond general-purpose analytics to develop cutting-edge Big Data applications using emerging technologies Who this book is for The book is intended for existing and aspiring Big Data professionals who wish to become the go-to person in their organization when it comes to Big Data architecture, analytics, and governance. While no prior knowledge of Big Data or related technologies is assumed, it will be helpful to have some programming experience.

Welcome to this book on "Big Data Analytics - Derive Insights" is a beginners guide to big data, mining, manipulation, integration, processing, visualisation and interpretation based on actual case studies in implementing Data Management frameworks across diverse organisations. Big data is a term used collectively for data sets that are voluminous, large and complicated that cannot be processed by traditional databases and applications. Some of the major challenges are the collection, extraction, mining, integration, analysis, sharing, storage, transfer, loading, querying, updating and visualisation of these data sets from multiple streams and formats. An effective and accurate visualisation of big data will lead to a more informed decision which in turn will result in greater operational efficiency, cost reduction, increased revenues and reduction in risk. Big Data analysis is used to identify business trends, predict accurate supply and demand, prevent diseases, combat crime and growth in revenues. Businessmen, organisations, scientists, hospitals, social media, marketing and governments face difficulties with large data sets in areas including internet search, finance and business trends. This book will give you a basic understanding of how big data works and the process used to get insights from such data.

Over the past decade, Big Data have become ubiquitous in all economic sectors, scientific disciplines, and human activities. They have led to striking technological advances, affecting all human experiences. Our ability to manage, understand, interrogate, and interpret such extremely large, multisource, heterogeneous, incomplete, multiscale, and incongruent data has not kept pace with the rapid increase of the volume, complexity and proliferation of the deluge of digital information. There are three reasons for this shortfall. First, the volume of data is increasing much faster than the corresponding rise of our computational processing power (Kryder's law > Moore's law). Second, traditional discipline-bounds inhibit expeditious progress. Third, our education and training activities have fallen behind the accelerated trend of scientific, information, and communication advances. There are very few rigorous

instructional resources, interactive learning materials, and dynamic training environments that support active data science learning. The textbook balances the mathematical foundations with dexterous demonstrations and examples of data, tools, modules and workflows that serve as pillars for the urgently needed bridge to close that supply and demand predictive analytic skills gap. Exposing the enormous opportunities presented by the tsunami of Big data, this textbook aims to identify specific knowledge gaps, educational barriers, and workforce readiness deficiencies. Specifically, it focuses on the development of a transdisciplinary curriculum integrating modern computational methods, advanced data science techniques, innovative biomedical applications, and impactful health analytics. The content of this graduate-level textbook fills a substantial gap in integrating modern engineering concepts, computational algorithms, mathematical optimization, statistical computing and biomedical inference. Big data analytic techniques and predictive scientific methods demand broad transdisciplinary knowledge, appeal to an extremely wide spectrum of readers/learners, and provide incredible opportunities for engagement throughout the academy, industry, regulatory and funding agencies. The two examples below demonstrate the powerful need for scientific knowledge, computational abilities, interdisciplinary expertise, and modern technologies necessary to achieve desired outcomes (improving human health and optimizing future return on investment). This can only be achieved by appropriately trained teams of researchers who can develop robust decision support systems using modern techniques and effective end-to-end protocols, like the ones described in this textbook.

- A geriatric neurologist is examining a patient complaining of gait imbalance and posture instability. To determine if the patient may suffer from Parkinson's disease, the physician acquires clinical, cognitive, phenotypic, imaging, and genetics data (Big Data). Most clinics and healthcare centers are not equipped with skilled data analytic teams that can wrangle, harmonize and interpret such complex datasets. A learner that completes a course of study using this textbook will have the competency and ability to manage the data, generate a protocol for deriving biomarkers, and provide an actionable decision support system. The results of this protocol will help the physician understand the entire patient dataset and assist in making a holistic evidence-based, data-driven, clinical diagnosis.
- To improve the return on investment for their shareholders, a healthcare manufacturer needs to forecast the demand for their product subject to environmental, demographic, economic, and bio-social sentiment data (Big Data). The organization's data-analytics team is tasked with developing a protocol that identifies, aggregates, harmonizes, models and analyzes these heterogeneous data elements to generate a trend forecast. This system needs to provide an automated, adaptive, scalable, and reliable prediction of the optimal investment, e.g., R&D allocation, that maximizes the company's bottom line. A reader that complete a course of study using this textbook will be able to ingest the observed structured and unstructured data, mathematically represent the data as a computable object, apply appropriate model-based and model-free prediction techniques. The results of these techniques may be used to forecast the expected relation between the company's investment, product supply, general demand of healthcare (providers and patients), and estimate the return on initial investments.

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

Data Science for Business with R, written by Jeffrey S. Saltz and Jeffrey M. Stanton, focuses on the concepts foundational for students starting a business analytics or data science degree program. To keep the book practical and applied, the authors feature a running case using a global airline business's customer survey dataset to illustrate how to turn data in business decisions, in addition to numerous examples throughout. To aid in usability beyond the classroom, the text features full integration of freely-available R and RStudio software, one of the most popular data science tools available. Designed for students with little to no experience in related areas like computer science, the book chapters follow a logical order from introduction and installation of R and RStudio, working with data architecture, undertaking data collection, performing data analysis, and transitioning to data archiving and presentation. Each chapter follows a familiar structure, starting with learning objectives and background, following the basic steps of functions alongside simple examples, applying these functions to the case study, and ending with chapter challenge questions, sources, and a list of R functions so students know what to expect in each step of their data science course. Data Science for Business with R provides readers with a straightforward and applied guide to this new and evolving field.

Introduction to Data Science: Data Analysis and Prediction Algorithms with R introduces concepts and skills that can help you tackle real-world data analysis challenges. It covers concepts from probability, statistical inference, linear regression, and machine learning. It also helps you develop skills such as R programming, data wrangling, data visualization, predictive algorithm building, file organization with UNIX/Linux shell, version control with Git and GitHub, and reproducible document preparation. This book is a textbook for a first course in data science. No previous knowledge of R is necessary, although some experience with programming may be helpful. The book is divided into six parts: R, data visualization, statistics with R, data wrangling, machine learning, and productivity tools. Each part has several chapters meant to be presented as one lecture. The author uses motivating case studies that realistically mimic a data scientist's experience. He starts by asking specific questions and answers these through data analysis so concepts are learned as a means to answering the questions. Examples of the case studies included are: US murder rates by state, self-reported student heights, trends in world health and economics, the impact of vaccines on infectious disease rates, the financial crisis of 2007-2008, election forecasting, building a baseball team, image processing of hand-written digits, and movie recommendation systems. The statistical concepts used to answer the case

study questions are only briefly introduced, so complementing with a probability and statistics textbook is highly recommended for in-depth understanding of these concepts. If you read and understand the chapters and complete the exercises, you will be prepared to learn the more advanced concepts and skills needed to become an expert. Learn, by example, the fundamentals of data analysis as well as several intermediate to advanced methods and techniques ranging from classification and regression to Bayesian methods and MCMC, which can be put to immediate use. Key Features Analyze your data using R – the most powerful statistical programming language Learn how to implement applied statistics using practical use-cases Use popular R packages to work with unstructured and structured data Book Description Frequently the tool of choice for academics, R has spread deep into the private sector and can be found in the production pipelines at some of the most advanced and successful enterprises. The power and domain-specificity of R allows the user to express complex analytics easily, quickly, and succinctly. Starting with the basics of R and statistical reasoning, this book dives into advanced predictive analytics, showing how to apply those techniques to real-world data though with real-world examples. Packed with engaging problems and exercises, this book begins with a review of R and its syntax with packages like Rcpp, ggplot2, and dplyr. From there, get to grips with the fundamentals of applied statistics and build on this knowledge to perform sophisticated and powerful analytics. Solve the difficulties relating to performing data analysis in practice and find solutions to working with messy data, large data, communicating results, and facilitating reproducibility. This book is engineered to be an invaluable resource through many stages of anyone's career as a data analyst. What you will learn Gain a thorough understanding of statistical reasoning and sampling theory Employ hypothesis testing to draw inferences from your data Learn Bayesian methods for estimating parameters Train regression, classification, and time series models Handle missing data gracefully using multiple imputation Identify and manage problematic data points Learn how to scale your analyses to larger data with Rcpp, data.table, dplyr, and parallelization Put best practices into effect to make your job easier and facilitate reproducibility Who this book is for Budding data scientists and data analysts who are new to the concept of data analysis, or who want to build efficient analytical models in R will find this book to be useful. No prior exposure to data analysis is needed, although a fundamental understanding of the R programming language is required to get the best out of this book.

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!

Use big data analytics to efficiently drive oil and gas exploration and production
Harness Oil and Gas Big Data with Analytics provides a complete view of big data and analytics techniques as they are applied to the oil and gas industry. Including a compendium of specific case studies, the book underscores the acute need for optimization in the oil and gas exploration and production stages and shows how data analytics can provide such optimization. This spans exploration, development, production and rejuvenation of oil and gas assets. The book serves as a guide for fully leveraging data, statistical, and quantitative analysis, exploratory and predictive modeling, and fact-based management to drive decision making in oil and gas operations. This comprehensive resource delves into the three major issues that face the oil and gas industry during the exploration and production stages: Data management, including storing massive quantities of data in a manner conducive to analysis and effectively retrieving, backing up, and purging data Quantification of uncertainty, including a look at the statistical and data analytics methods for making predictions and determining the certainty of those predictions Risk assessment, including predictive analysis of the likelihood that known risks are realized and how to properly deal with unknown risks Covering the major issues facing the oil and gas industry in the exploration and production stages, Harness Big Data with Analytics reveals how to model big data to realize efficiencies and business benefits.

Analytical tools and algorithms are essential in business data and information systems. Efficient economic and financial forecasting in machine learning techniques increases gains while reducing risks. Providing research on predictive models with high accuracy, stability, and ease of interpretation is important in improving data preparation, analysis, and implementation processes in business organizations. Machine Learning Techniques for Improved Business Analytics is a collection of innovative research on the methods and applications of artificial intelligence in strategic business decisions and management. Featuring coverage on a broad range of topics such as data mining, portfolio optimization, and social network analysis, this book is ideally designed for business managers and practitioners, upper-level business students, and researchers seeking current research on large-scale information control and evaluation technologies that exceed the functionality of conventional data processing techniques.

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, HTML, 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

Data science and analytics have emerged as the most desired fields in driving business decisions. Using the techniques and methods of data science, decision makers can uncover hidden patterns in their data, develop algorithms and models that help improve processes and make key business decisions. Data science is a data driven decision making approach that uses several different areas and disciplines with a purpose of extracting insights and knowledge from structured and unstructured data. The algorithms and models of data science along with machine learning and predictive modeling are widely used in solving business problems and predicting future outcomes. This book combines the key concepts of data science and analytics to help you gain a practical understanding of these fields. The four different sections of the book are divided into chapters that explain the core of data science. Given the booming interest in data science, this book is timely and informative.

[Copyright: 70e96bf780c4c3985713504a84bcafc4](#)