Big Data Analytics In R

Big Data Analytics with RPackt PublishingBig Data Analytics with R and HadoopPackt Publishing Ltd

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

Over 80 recipes to help you breeze through your data analysis projects using RAbout This Book* Analyse your data using the popular R packages like ggplot2 with ready-to-use and customizable recipes* Find meaningful insights from your data and generate dynamic reports* A practical guide to help you put your data analysis skills in R to practical useWho This Book Is ForThis book is for data scientists, analysts and even enthusiasts who want to learn and implement the various data analysis techniques using R in a practical way. Those looking for quick, handy solutions to common tasks and challenges in data analysis will find this book to be very useful. Basic knowledge of statistics and R programming is assumed.What You Will Learn* Acquire, format and visualize your data using R* Using R to perform an Exploratory data analysis* Introduction to machine learning algorithms such as classification and regression* Get started with social network analysis* Generate dynamic reporting with Shiny* Get started with geospatial analysis* Handling large data with R using Spark and MongoDB* Build

Recommendation system- Collaborative Filtering, Content based and Hybrid* Learn real world dataset examples- Fraud Detection and Image RecognitionIn DetailData analytics with R has emerged as a very important focus for organizations of all kinds. R enables even those with only an intuitive grasp of the underlying concepts, without a deep mathematical background, to unleash powerful and detailed examinations of their data.This book will show you how you can put your data analysis skills in R to practical use, with recipes catering to the basic as well as advanced data analysis tasks. Right from acquiring your data and preparing it for analysis to the more complex data analysis techniques, the book will show you how you can implement each technique in the best possible manner. You will also visualize your data using the popular R packages like ggplot2 and gain hidden insights from it. Starting with implementing the basic data analysis concepts like handling your data to creating basic plots, you will master the more advanced data analysis techniques like performing cluster analysis, and generating effective analysis reports and visualizations. Throughout the book, you will get to know the common problems and obstacles you might encounter while implementing each of the data analysis techniques in R, with ways to overcoming them in the easiest possible way.By the end of this book, you will have all the knowledge you need to become an expert in data analysis with R, and put your skills to test in real-world scenarios.Style and Approach* Hands-on recipes to walk through data science challenges using R* Your one-stop solution for common and not-so-common pain points while performing real-world problems to execute a series of tasks.* Addressing your common and not-so-common pain points, this is a book that you must have on the shelf Cyber-physical systems (CPS) and the Internet of Things (IoT) are rapidly developing technologies that are transforming our society. The disruptive transformation of the economy and society is expected due to the data collected by these systems, rather than the technological aspects of such as networks, embedded systems, and cloud technology. However, to create value out of the data, it must be transformed into information and therefore, expertise in data analytics and machine learning is the key component of future smart systems in cities and other applications. Big Data Analytics in Cyber-Physical Systems examines sensor signal processing, IoT gateways, optimization and decision making, intelligent mobility, and implementation of machine learning algorithms in embedded systems. This book focuses on the interaction between IoT technology and the mathematical tools to evaluate the extracted data of those systems. Each chapter provides different tools and applications in order to present a broad list of data analytics and machine learning tools in multiple IoT applications. Additionally, this volume addresses the education transfer needed to incorporate these technologies into our society by examining new platforms for IoT in schools, new courses and concepts for universities and adult education on IoT and data science. Fills the gap between IoT, CPS, and mathematical modeling Numerous use cases that discuss how concepts are applied in different

domains and applications Provides "best practices," "real developments", and "winning stories" to complement technical information Uniquely covers contents within the context of mathematical foundations of signal processing and machine learning in CPS and IoT

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

This book is aimed at developers, designers, and architects who would like to build big data enterprise search solutions for their customers or organizations. No prior knowledge of Apache Hadoop and Apache Solr/Lucene technologies is required.

Master the craft of predictive modeling in R by developing strategy, intuition, and a solid foundation in essential concepts About This Book Grasping the major methods of predictive modeling and moving beyond black box thinking to a deeper level of understanding Leveraging the flexibility and modularity of R to experiment with a range of different techniques and data types Packed with practical advice and tips explaining important concepts and best practices to help you understand quickly and easily Who This Book Is For Although budding data scientists, predictive modelers, or quantitative analysts with only basic exposure to R and statistics will find this book to be useful, the experienced data scientist professional wishing to attain master level status , will also find this book extremely valuable.. This book assumes familiarity with the fundamentals of R, such as the main data types, simple functions, and how to move data around. Although no prior experience with machine learning or predictive modeling is required, there are some advanced topics provided that will require more than novice exposure. What You Will Learn Master the steps involved in the predictive modeling process Grow your expertise in using R and its diverse range of packages Learn how to classify predictive models and distinguish which models are suitable for a particular problem Understand steps for tidying data and

improving the performing metrics Recognize the assumptions, strengths, and weaknesses of a predictive model Understand how and why each predictive model works in R Select appropriate metrics to assess the performance of different types of predictive model Explore word embedding and recurrent neural networks in R Train models in R that can work on very large datasets In Detail R offers a free and open source environment that is perfect for both learning and deploying predictive modeling solutions. With its constantly growing community and plethora of packages, R offers the functionality to deal with a truly vast array of problems. The book begins with a dedicated chapter on the language of models and the predictive modeling process. You will understand the learning curve and the process of tidying data. Each subsequent chapter tackles a particular type of model, such as neural networks, and focuses on the three important questions of how the model works, how to use R to train it, and how to measure and assess its performance using real-world datasets. How do you train models that can handle really large datasets? This book will also show you just that. Finally, you will tackle the really important topic of deep learning by implementing applications on word embedding and recurrent neural networks. By the end of this book, you will have explored and tested the most popular modeling techniques in use on real- world datasets and mastered a diverse range of techniques in predictive analytics using R. Style and approach This book takes a step-by-step approach in explaining the intermediate to advanced concepts in predictive analytics. Every concept is explained in depth, supplemented with practical examples applicable in a real-world setting.
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 introduce 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.

R for Cloud Computing looks at some of the tasks performed by business analysts on the desktop (PC era) and helps the user navigate the wealth of information in R and its 4000 packages as well as transition the same analytics using the cloud. With this information the reader can select both cloud vendors and the sometimes confusing cloud ecosystem as well as the R packages that can help process the analytical tasks with minimum effort, cost and maximum usefulness and customization. The use of Graphical User Interfaces (GUI) and Step by Step screenshot tutorials is emphasized in this book to lessen the famous learning curve in learning R and some of the needless confusion created in cloud computing that hinders its widespread adoption. This will help you kick-start analytics on the cloud including chapters on both cloud computing, R, common tasks performed in analytics including the current focus and scrutiny of Big Data Analytics, setting up and navigating cloud providers. Readers are exposed to a breadth of cloud computing choices and analytics topics without being buried in needless depth. The included references and links allow the reader to pursue business analytics on the cloud easily. It is aimed at practical analytics and is easy to transition from existing analytical set up to the cloud on an open source system based primarily on R. This book is aimed at industry practitioners with basic programming skills and students who want to enter analytics as a profession. Note the scope of the book is neither statistical theory nor graduate level research for statistics, but rather it is for business analytics practitioners. It will also help researchers and academics but at a practical rather than conceptual level. The R statistical software is the fastest growing analytics platform in the world, and is established in both academia and corporations for robustness, reliability and accuracy. The cloud computing paradigm is firmly established as the next generation of computing from microprocessors to desktop PCs to cloud.

Harness the power of Scala to program Spark and analyze tonnes of data in the blink of an eye! About This Book Learn Scala's sophisticated type system that combines Functional Programming and object-oriented concepts Work on a wide array of applications, from simple batch jobs to stream processing and machine learning Explore the most common as well as some complex use-cases to perform large-scale

data analysis with Spark Who This Book Is For Anyone who wishes to learn how to perform data analysis by harnessing the power of Spark will find this book extremely useful. No knowledge of Spark or Scala is assumed, although prior programming experience (especially with other JVM languages) will be useful to pick up concepts quicker. What You Will Learn Understand object-oriented & functional programming concepts of Scala In-depth understanding of Scala collection APIs Work with RDD and DataFrame to learn Spark's core abstractions Analysing structured and unstructured data using SparkSQL and GraphX Scalable and fault-tolerant streaming application development using Spark structured streaming Learn machine-learning best practices for classification, regression, dimensionality reduction, and recommendation system to build predictive models with widely used algorithms in Spark MLlib & ML Build clustering models to cluster a vast amount of data Understand tuning, debugging, and monitoring Spark applications Deploy Spark applications on real clusters in Standalone, Mesos, and YARN In Detail Scala has been observing wide adoption over the past few years, especially in the field of data science and analytics. Spark, built on Scala, has gained a lot of recognition and is being used widely in productions. Thus, if you want to leverage the power of Scala and Spark to make sense of big data, this book is for you. The first part introduces you to Scala, helping you understand the object-oriented and functional programming concepts needed for Spark application development. It then moves on to Spark to cover the basic abstractions using RDD and DataFrame. This will help you develop scalable and fault-tolerant streaming applications by analyzing structured and unstructured data using SparkSQL, GraphX, and Spark structured streaming. Finally, the book moves on to some advanced topics, such as monitoring, configuration, debugging, testing, and deployment. You will also learn how to develop Spark applications using SparkR and PySpark APIs, interactive data analytics using Zeppelin, and in-memory data processing with Alluxio. By the end of this book, you will have a thorough understanding of Spark, and you will be able to perform full-stack data analytics with a feel that no amount of data is too big. Style and approach Filled with practical examples and use cases, this book will hot only help you get up and running with Spark, but will also take you farther down the road to becoming a data scientist. Learn the fundamental aspects of the business statistics, data mining, and machine learning techniques required to understand the huge amount of data generated by your organization. This book explains practical business analytics through examples, covers the steps involved in using it correctly, and shows you the context in which a particular technique does not make sense. Further, Practical Business Analytics using R helps you understand specific issues faced by organizations and how the solutions to these issues can be facilitated by business analytics. This book will discuss and explore the following through examples and case studies: An introduction to R: data management and R functions The architecture, framework, and life cycle of a business analytics project Descriptive analytics using R: descriptive statistics and data cleaning Data mining: classification, association rules, and clustering Predictive analytics: simple regression, multiple regression, and logistic regression This book includes case studies on important business analytic techniques, such as classification, association, clustering, and regression. The R language is the statistical tool used to demonstrate the concepts throughout the book. What You Will Learn • Write R programs to handle data • Build analytical models and draw useful inferences from them • Discover the

basic concepts of data mining and machine learning • Carry out predictive modeling • Define a business issue as an analytical problem Who This Book Is For Beginners who want to understand and learn the fundamentals of analytics using R. Students, managers, executives, strategy and planning professionals, software professionals, and BI/DW professionals.

This book has a collection of articles written by Big Data experts to describe some of the cutting-edge methods and applications from their respective areas of interest, and provides the reader with a detailed overview of the field of Big Data Analytics as it is practiced today. The chapters cover technical aspects of key areas that generate and use Big Data such as management and finance; medicine and healthcare; genome, cytome and microbiome; graphs and networks; Internet of Things; Big Data standards; bench-marking of systems; and others. In addition to different applications, key algorithmic approaches such as graph partitioning, clustering and finite mixture modelling of high-dimensional data are also covered. The varied collection of themes in this volume introduces the reader to the richness of the emerging field of Big Data Analytics.

R offers a free and open source environment that is perfect for both learning and deploying predictive modeling solutions in the real world. With its constantly growing community and plethora of packages, R offers the functionality to deal with a truly vast array of problems. This book is designed to be both a guide and a reference for moving beyond the basics of predictive modeling. The book begins with a dedicated chapter on the language of models and the predictive modeling process. Each subsequent chapter tackles a particular type of model, such as neural networks, and focuses on the three important questions of how the model works, how to use R to train it, and how to measure and assess its performance using real world data sets. By the end of this book, you will have explored and tested the most popular modeling techniques in use on real world data sets and mastered a diverse range of techniques in predictive analytics.

This book is aimed at students in communications and signal processing who want to extend their skills in the energy area. It describes power systems and why these backgrounds are so useful to smart grid, wireless communications being very different to traditional wireline communications.

"Big Data Analytics is a field that dissects, efficiently extricates data from, or in any case manages informational indexes that are excessively huge or complex to be managed by customary information preparing application programming. Information with numerous cases (lines) offers more noteworthy factual force, while information with higher multifaceted nature may prompt a higher bogus disclosure rate. Enormous information challenges incorporate catching information, information stockpiling, information investigation, search, sharing, move, representation, and questioning, refreshing, data security and data source. Large information was initially connected with three key ideas: volume, variety and velocity. Consequently, huge information regularly incorporates information with sizes that surpass the limit of conventional programming to measure inside a satisfactory time and worth. Current utilization of the term enormous information will in general allude to the utilization of predictive analytics, user behavior analytics, or certain other progressed information investigation techniques that concentrate an incentive from information, and sometimes to a specific size of

informational index. There is little uncertainty that the amounts of information now accessible are undoubtedly enormous, however that is not the most important quality of this new information biological system. Investigation of informational indexes can discover new relationships to spot business patterns or models. Researchers, business persons, clinical specialists, promoting and governments consistently meet challenges with huge informational collections in territories including Internet look, fintech, metropolitan informatics, and business informatics. Researchers experience constraints in e-Science work, including meteorology, genomics, connectomics, complex material science reproductions, science and ecological exploration. The main objective of this book is to write about issues, challenges, opportunities, and solutions in novel research projects about big data in various domains. The topics of interest include, but are not limited to: efficient storage, management and sharing large scale of data; novel approaches for analyzing data using big data technologies; implementation of high performance and/or scalable and/or real-time computation algorithms for analyzing big data; usage of various data sources like historical data, social networking media, machine data and crowd-sourcing data; using machine learning, visual analytics, data mining, spatio-temporal data analysis and statistical inference in different domains (with large scale datasets); Legal and ethical issues and solutions for using, sharing and publishing large datasets; and the results of data analytics, security and privacy issues"--

R for Business Analytics looks at some of the most common tasks performed by business analysts and helps the user navigate the wealth of information in R and its 4000 packages. With this information the reader can select the packages that can help process the analytical tasks with minimum effort and maximum usefulness. The use of Graphical User Interfaces (GUI) is emphasized in this book to further cut down and bend the famous learning curve in learning R. This book is aimed to help you kick-start with analytics including chapters on data visualization, code examples on web analytics and social media analytics, clustering, regression models, text mining, data mining models and forecasting. The book tries to expose the reader to a breadth of business analytics topics without burying the user in needless depth. The included references and links allow the reader to pursue business analytics topics. This book is aimed at business analysts with basic programming skills for using R for Business Analytics. Note the scope of the book is neither statistical theory nor graduate level research for statistics, but rather it is for business analytics practitioners. Business analytics (BA) refers to the field of exploration and investigation of data generated by businesses. Business Intelligence (BI) is the seamless dissemination of information through the organization, which primarily involves business metrics both past and current for the use of decision support in businesses. Data Mining (DM) is the process of discovering new patterns from large data using algorithms and statistical methods. To differentiate between the three, BI is mostly current reports, BA is models to predict and strategize and DM matches patterns in big data. The R statistical software is the fastest growing analytics platform in the world, and is established in both academia and corporations for robustness, reliability and accuracy. The book utilizes Albert Einstein’s famous remarks on making things as simple as possible, but no simpler. This book will blow the last remaining doubts in your mind about using R in your business environment. Even non-technical users will enjoy the easy-to-use examples. The interviews with creators and

corporate users of R make the book very readable. The author firmly believes Isaac Asimov was a better writer in spreading science than any textbook or journal author.

The Ultimate Guide to Data Science and Analytics This practical guide is accessible for the reader who is relatively new to the field of data analytics, while still remaining robust and detailed enough to function as a helpful guide to those already experienced in the field. Data science is expanding in breadth and growing rapidly in importance as technology rapidly integrates ever deeper into business and our daily lives. The need for a succinct and informal guide to this important field has never been greater. RIGHT NOW you can get ahead of the pack! This coherent guide covers everything you need to know on the subject of data science, with numerous concrete examples, and invites the reader to dive further into this exciting field. Students from a variety of academic backgrounds, including computer science, business, engineering, statistics, anyone interested in discovering new ideas and insights derived from data can use this as a textbook. At the same time, professionals such as managers, executives, professors, analysts, doctors, developers, computer scientists, accountants, and others can use this book to make a quantum leap in their knowledge of big data in a matter of only a few hours. Learn how to understand this field and uncover actionable insights from data through analytics. UNDERSTAND the following key insights when you grab your copy today: WHY DATA IS IMPORTANT TO YOUR BUSINESS DATA SOURCES HOW DATA CAN IMPROVE YOUR BUSINESS HOW BIG DATA CREATES VALUE DEVELOPMENT OF BIG DATA CONSIDERING THE PROS AND CONS OF BIG DATA BIG DATA FOR SMALL BUSINESSES THE COST EFFECTIVENESS OF DATA ANALYTICS WHAT TO CONSIDER WHEN PREPARING FOR A NEW BIG DATA SOLUTION DATA GATHERING DATA SCRUBBING DESCRIPTIVE ANALYTICS INFERENTIAL STATISTICS PREDICTIVE ANALYTICS PREDICTIVE MODELS DESCRIPTIVE MODELING DECISION MODELING PREDICTIVE ANALYSIS METHODS MACHINE LEARNING TECHNIQUES DATA ANALYSIS WITH "R" ANALYTICAL CUSTOMER RELATIONSHIP MANAGEMENT (CRM) THE USE OF PREDICTIVE ANALYTICS IN HEALTHCARE THE USE OF PREDICTIVE ANALYTICS IN THE FINANCIAL SECTOR PREDICTIVE ANALYTICS & BUSINESS MARKETING STRATEGIES FRAUD DETECTION SHIPPING BUSINESS CONTROLLING RISK FACTORS THE REVOLUTION OF PREDICTIVE ANALYSIS ACROSS A VARIETY OF INDUSTRIES DESCRIPTIVE AND PREDICTIVE ANALYSIS CRUCIAL FACTORS FOR DATA ANALYSIS RESOURCES AND FLEXIBLE TECHNICAL STRUCTURE BUSINESS INTELLIGENCE HYPER TARGETING WHAT IS DATA SCIENCE? DATA MUNGING DEMYSTIFYING DATA SCIENCE SECURITY RISKS TODAY BIG DATA AND IMPACTS ON EVERYDAY LIFE FINANCE AND BIG DATA APPLYING SENTIMENT ANALYSIS RISK EVALUATION AND THE DATA SCIENTIST THE FINANCE INDUSTRY AND REAL-TIME ANALYTICS HOW BIG DATA IS BENEFICIAL TO THE CUSTOMER CUSTOMER SEGMENTATION IS GOOD FOR BUSINESS USE OF BIG DATA BENEFITS IN MARKETING GOOGLE TRENDS THE PROFILE OF A PERFECT CUSTOMER LEAD SCORING IN PREDICTIVE ANALYSIS EVALUATING THE WORTH OF LIFETIME VALUE BIG DATA ADVANTAGES AND DISADVANTAGES MAKING COMPARISONS WITH COMPETITORS DATA SCIENCE IN THE TRAVEL SECTOR SAFETY ENHANCEMENTS THANKS TO BIG DATA BIG DATA AND AGRICULTURE BIG DATA AND LAW ENFORCEMENT THE USE OF BIG DATA IN THE PUBLIC SECTOR BIG DATA AND GAMING PRESCRIPTIVE ANALYTICS GOOGLE'S "SELF-DRIVING CAR" AND MUCH MORE! WANT MORE? Scroll up and grab this helpful guide toady!

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 at www.wiley.com/go/9781118876138. Get started discovering, analyzing, visualizing, and presenting data in a meaningful way today!

Review: "I would recommend this book to all prospective data scientists - as well as those software professionals who choose to transfer or migrate to the domain of data science. It is a useful addition to the body of work already available to guide project managers of data science projects." Lt Col (Dr) Rajesh Kapur (Retd), AI Investor, Asst. Prof. TIMSCDR, Hyderabad, India "It's a masterpiece of work for the aspiring leaders of data science and AI. It's also a guide for executives and investors to get maximum value from their investment in AI. Beginners in data science can also get the most out of this book.", Jay Ojha, Business intelligence and analytics manager, HCL Infosystem LtdWhy should you read this book? 87% of data science project fails to make to production in enterprises. Only 50% is the data leadership success rate. Is it not surprising to know when data science and AI are in the top trend? If you are looking for a career in data science or looking for leadership, these insights may disturb you. Don't worry, "Step up for Leadership in Enterprise Data Science & Artificial Intelligence with Big Data." will -Burst the myths around data science, AI & big data-Presents the real business scenarios -Take you on the journey of data science, AI & big data even if you are an ultimate beginner.-Introduce you to the essential skills of success in this field -Develop a leadership mindset by cutting edge methodologies & strategies-Make you aware of technical trends around it-Develop technical skills with R, Python, Machine learning with big data as well as business skills-Reduce failure possibility and increase the chance of success by covering the 360 degrees view of the field. Each day counts. So as your steps. Step up immediately and begin your journey to your dreams of data science and AI.

As the age of Big Data emerges, it becomes necessary to take the five dimensions of Big Data- volume, variety, velocity, volatility, and veracity- and focus these dimensions towards one critical emphasis - value. The Encyclopedia of Business Analytics and Optimization confronts the challenges of information retrieval in the age of Big Data by exploring recent advances in the areas of knowledge management, data visualization, interdisciplinary communication, and others. Through its critical approach and practical application, this book will be a must-have reference for any professional, leader, analyst, or manager interested in making the most of the knowledge resources at their disposal.

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.

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.

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.

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.
??????????????? ??????????????? ????????????????? ??? ??????? ??? ?Wired????????? ??? ???? ????????????????????????? ???????????? —— ????Lawrence Lessig??????????????? ????????????????????????? ????????????????????????? ?????????????????????? ????????????????????????? ?? ????????????????????????????????? ??????????????????????????? ?????????????????? ?????????????????????????? ??????????????…… ??????????????????????????? ??????????????????????? ????????????????????????????? ???????????????????? ????????????????????????
Leverage the capabilities of SAS to process and analyze Big Data About This Book Combine SAS with platforms such as Hadoop, SAP HANA, and Cloud Foundry-based platforms for effecient Big Data analytics Learn how to use the web browser-based SAS Studio and iPython Jupyter Notebook interfaces with SAS Practical, real-world examples on predictive modeling, forecasting, optimizing and reporting your Big Data analysis with SAS Who This Book Is For SAS professionals and data analysts who wish to perform analytics on Big Data using SAS to gain actionable insights will find this book to be very useful. If you are a data science professional looking to perform large-scale analytics with SAS, this book will also help you. A basic understanding of SAS will be helpful, but is not mandatory. What You Will Learn Configure a free version of SAS in order do hands-on exercises dealing with data management, analysis, and reporting. Understand the basic concepts of the SAS language which consists of the data step (for data preparation) and procedures (or PROCs) for analysis. Make use of the web browser based SAS Studio and iPython Jupyter Notebook interfaces for coding in the SAS, DS2, and FedSQL programming languages. Understand how the DS2 programming language plays an important role in Big Data preparation and analysis using SAS Integrate and work efficiently with Big Data platforms like Hadoop, SAP HANA, and cloud foundry based systems. In Detail SAS has been recognized by Money Magazine and Payscale as one of the top business skills to learn in order to advance one's career. Through innovative data management, analytics, and business intelligence software and services, SAS helps customers solve their business problems by allowing them to make better decisions faster. This book introduces the reader to the SAS and how they can use SAS to perform efficient analysis on any size data, including Big Data. The reader will learn how to prepare data for analysis, perform predictive, forecasting, and optimization analysis and then deploy or report on the results of these analyses. While performing the coding examples within this book the reader will learn how to use the web browser based SAS Studio and iPython Jupyter Notebook interfaces for working with SAS. Finally, the reader will learn how SAS's architecture is engineered and designed to scale up and/or out and be combined with the open source offerings such as Hadoop, Python, and R. By the end of this book, you will be able to clearly understand how you can efficiently analyze Big Data using SAS. Style and approach The book starts off by introducing the reader to SAS and the SAS programming language which provides data management, analytical, and reporting capabilities. Most chapters include hands on examples which highlights how SAS provides The Power to Know©. The reader will learn that if they are looking to perform large-scale data analysis that SAS provides an open platform engineered and designed to scale both up and out which allows the power of SAS to combine with open source offerings such as Hadoop, Python, and R.
??????????????????????? ?????2018????????? ????????????????????????????????????? ????????????Amazon????????? ?????TED?????????????????????? ?? ???????????????????????????????? ???? ??????????????????????? ??????????????????????????? ?????????????????????????????????? ??????????????????????????????

???????????????? ???????????????????????? ????????????????????????????????????
??
????????????????????????????????? ??
??? ????????????????????????????????????
??
??
???? ??
????????????????????????? ??????????? ?????????????????????????????? ?????????????
??????????? ?? ????????????????PaGamO/BoniO???? ????????????????????
????????????????? ????TEDxTaipei??????????? ??????
??? ???????
??
??
??? ???????
??
????????????????????????????????????……??TED???????????????????????????????????? ???
??……????????????????????????????????????
????????????????????????????????
??TED
????????????????????????????????Nature? ???? ?????? Hans Rosling ????????????????????
????????????????WHO???????????UNICEF????????????????????????????WEF????????
??? 2005????????????????????????????Gapminder Foundation????????????????????????????
????????????????????????????????????TED????????????3,500???
??Fast Company??????????? ???
??
2017???????????????????????? ?????? Ola Rosling??????? Anna Rosling Rönnlund ????????
????????????????????????????2010????????????????????????Trendalyzer?Google?????????Goog
le?? ???? ??? ????????????????????????????
??
???????????????????????????? ?? ???? ? ? ??????????? Chapter 1 ??????? ??????
?????????????????? ?????????? ???????? ??????????? ????????????????? ?????? ????????????????
Chapter 2 ??????? ????? ???????????????? ????????? ???????? ?????????? ??????
????????????? ????????? Chapter 3 ??????? ??????????? ????????????????????? ??????
???????? ????????????? ????????????????? Chapter 4 ??????? ?????? ???????? ??????
???????????????? ?????4000???? ?????? ??? ??? ??????? Chapter 5 ??????? ?????????
?????? ????????????? ?80/20?? ????? ?????? Chapter 6 ??????? ????? ?????? ?????
???????? ??????? Chapter 7 ??????? ?????? ?????? ??????? ????????????? ??????? Chapter
8 ???????? ???????? ??????? ???????? ?????? ??????????? Chapter 9 ??????? ??????
?????? ????? ??????? ????????? Chapter 10 ??????? ?????? ?????? ???????????
???????????????? Chapter 11 ????????? ????????????? ?????????? ????? ? ? ? ? ? ?

Standard tutorial-based approach."Getting Started with Greenplum for Big Data" Analytics is great for data scientists and data analysts with a basic knowledge of Data Warehousing and Business Intelligence platforms who are new to Big Data and who are looking to get a good grounding in how to use the Greenplum Platform. It’s assumed that you will have some experience with database design and programming as well as be familiar with analytics tools like R and Weka.

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.

Explore big data concepts, platforms, analytics, and their applications using the power of Hadoop 3 Key Features Learn Hadoop 3 to build effective big data analytics solutions on-premise and on cloud Integrate Hadoop with other big data tools such as R, Python, Apache Spark, and Apache Flink Exploit big data using Hadoop 3 with real-world examples Book Description Apache Hadoop is the most popular platform for big data processing, and can be combined with a host of other big data tools to build powerful analytics solutions. Big Data Analytics with Hadoop 3 shows you how to do just that, by providing insights into the software as well as its benefits with the help of practical examples. Once you have taken a tour of Hadoop 3's latest features, you will get an overview of HDFS, MapReduce, and YARN, and how they enable faster, more efficient big data processing. You will then move on to learning how to integrate Hadoop with the open source tools, such as Python and R, to analyze and visualize data and perform statistical computing on big data. As you get acquainted with all this, you will explore how to use Hadoop 3 with Apache Spark and Apache Flink for real-time data analytics and stream processing. In addition to this, you will understand how to use Hadoop to build analytics solutions on the cloud and an end-to-end pipeline to perform big data analysis using practical use cases. By the end of this book, you will be well-versed with the analytical capabilities of the Hadoop ecosystem. You will be able to build powerful solutions to perform big data analytics and get insight effortlessly. What you will learn Explore the new features of Hadoop 3 along with HDFS, YARN, and MapReduce Get well-versed with the analytical capabilities of Hadoop ecosystem using practical examples Integrate Hadoop with R and Python for more efficient big data processing Learn to use Hadoop with Apache Spark and Apache Flink for real-time data analytics Set up a Hadoop cluster on AWS cloud Perform big data analytics on AWS using Elastic Map Reduce Who this book is for Big Data Analytics with Hadoop 3 is for you if you are looking to build high-performance analytics solutions for your enterprise or business using Hadoop 3's powerful features, or you're new to big data analytics. A basic understanding of the Java programming language is required.

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.

Learn Big Data from the ground up with this complete and up-to-date resource from leaders in the field Big Data: Concepts, Technology, and Architecture delivers a comprehensive treatment of Big Data tools, terminology, and technology perfectly suited to a wide range of business professionals, academic researchers, and students. Beginning with a fulsome overview of what we mean when we say, “Big Data,” the book moves on to discuss every stage of the lifecycle of Big Data. You’ll learn about the creation of structured, unstructured, and semi-structured data, data storage solutions, traditional database solutions like SQL, data processing, data analytics, machine learning, and data mining. You’ll also discover how specific technologies like Apache Hadoop, SQOOP, and Flume work. Big Data also covers the central topic of big data visualization with Tableau, and you’ll learn how to create scatter plots,

histograms, bar, line, and pie charts with that software. Accessibly organized, Big Data includes illuminating case studies throughout the material, showing you how the included concepts have been applied in real-world settings. Some of those concepts include: The common challenges facing big data technology and technologists, like data heterogeneity and incompleteness, data volume and velocity, storage limitations, and privacy concerns Relational and non-relational databases, like RDBMS, NoSQL, and NewSQL databases Virtualizing Big Data through encapsulation, partitioning, and isolating, as well as big data server virtualization Apache software, including Hadoop, Cassandra, Avro, Pig, Mahout, Oozie, and Hive The Big Data analytics lifecycle, including business case evaluation, data preparation, extraction, transformation, analysis, and visualization Perfect for data scientists, data engineers, and database managers, Big Data also belongs on the bookshelves of business intelligence analysts who are required to make decisions based on large volumes of information. Executives and managers who lead teams responsible for keeping or understanding large datasets will also benefit from this book.

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.

This book discusses emerging technologies in the field of the Internet of Things and big data, an area that will be scaled in next two decades. Written by a team of leading experts, it is the only book focusing on the broad areas of both the Internet of things and big data. The thirteen chapters present real-time experimental methods and theoretical explanations, as well as the implementation of these technologies through various applications. Offering a blend of theory and hands-on practices, the book enables graduate, postgraduate and research students who are involved in real-time project scaling techniques to understand projects and their execution. It is also useful for senior computer students, researchers and industry workers who are involved in experimenting with the Internet of Things and big data technologies, helping them to solve the real-time problem. Moreover, the chapters covering cutting-edge technologies help multidisciplinary researchers who are bridging the gap of two different outset real-time problems.

Get your statistics basics right before diving into the world of data science About This Book No need to take a degree in statistics, read this book and get a strong statistics base for data science and real-world programs; Implement statistics in data science tasks such as data cleaning, mining, and analysis Learn all about probability, statistics, numerical computations, and more with the help of R programs Who This Book Is For This book is intended for those developers who are willing to enter the field of data science and are looking for concise information of statistics with the help of insightful programs and simple explanation. Some basic hands on R will be useful. What You Will Learn Analyze the transition from a data developer to a data scientist mindset Get acquainted with the R programs and the logic used for statistical computations Understand mathematical concepts such as variance, standard deviation, probability, matrix calculations, and more Learn to implement statistics in data science tasks such as data cleaning, mining, and analysis Learn the statistical techniques required to perform tasks such as linear regression, regularization, model assessment, boosting, SVMs, and working with neural networks Get comfortable with performing various statistical computations for data science programmatically In Detail Data science is an ever-evolving field, which is growing in popularity at an exponential rate. Data science includes techniques and theories extracted from the fields of statistics; computer science, and, most

importantly, machine learning, databases, data visualization, and so on. This book takes you through an entire journey of statistics, from knowing very little to becoming comfortable in using various statistical methods for data science tasks. It starts off with simple statistics and then move on to statistical methods that are used in data science algorithms. The R programs for statistical computation are clearly explained along with logic. You will come across various mathematical concepts, such as variance, standard deviation, probability, matrix calculations, and more. You will learn only what is required to implement statistics in data science tasks such as data cleaning, mining, and analysis. You will learn the statistical techniques required to perform tasks such as linear regression, regularization, model assessment, boosting, SVMs, and working with neural networks. By the end of the book, you will be comfortable with performing various statistical computations for data science programmatically. Style and approach Step by step comprehensive guide with real world examples

Tap into the realm of social media and unleash the power of analytics for data-driven insights using RAbout This Book* A practical guide written to help leverage the power of the R eco-system to extract, process, analyze, visualize and model social media data* Learn about data access, retrieval, cleaning, and curation methods for data originating from various social media platforms.* Visualize and analyze data from social media platforms to understand and model complex relationships using various concepts and techniques such as Sentiment Analysis, Topic Modeling, Text Summarization, Recommendation Systems, Social Network Analysis, Classification, and Clustering.Who This Book Is ForIt is targeted at IT professionals, Data Scientists, Analysts, Developers, Machine Learning Enthusiasts, social media marketers and anyone with a keen interest in data, analytics, and generating insights from social data. Some background experience in R would be helpful, but not necessary, since this book is written keeping in mind, that readers can have varying levels of expertise.What You Will Learn* Learn how to tap into data from diverse social media platforms using the R ecosystem* Use social media data to formulate and solve real-world problems* Analyze user social networks and communities using concepts from graph theory and network analysis* Learn to detect opinion and sentiment, extract themes, topics, and trends from unstructured noisy text data from diverse social media channels* Understand the art of representing actionable insights with effective visualizations* Analyze data from major social media channels such as Twitter, Facebook, Flickr, Foursquare, Github, StackExchange, and so on* Learn to leverage popular R packages such as ggplot2, topicmodels, caret, e1071, tm, wordcloud, twittR, Rfacebook, dplyr, reshape2, and many moreIn DetailThe Internet has truly become humongous, especially with the rise of various forms of social media in the last decade, which give users a platform to express themselves and also communicate and collaborate with each other. This book will help the reader to understand the current social media landscape and to learn how analytics can be leveraged to derive insights from it. This data can be analyzed to gain valuable insights into the behavior and engagement of users, organizations, businesses, and brands. It will help readers frame business problems and solve them using social data.The book will also cover several practical real-world use cases on social media using R and its advanced packages to utilize data science methodologies such as sentiment analysis, topic modeling, text summarization, recommendation systems, social network analysis, classification, and clustering. This will enable readers to learn different hands-on approaches to obtain data from diverse social media sources such as Twitter and Facebook. It will also show readers how to establish detailed workflows to process, visualize, and analyze data to transform social data into actionable insights.Style and approachThis book follows a step-by-step approach with detailed strategies for understanding, extracting, analyzing, visualizing, and modeling data from several major social network platforms such as Facebook, Twitter, Foursquare, Flickr, Github, and StackExchange. The chapters cover several real-world use cases and leverage data science, machine learning, network analysis, and graph theory concepts along with the R ecosystem,

including popular packages such as ggplot2, caret,dplyr, topicmodels, tm, and so on.
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.
This book is intended to present the state of the art in research on machine learning and big data analytics. The accepted chapters covered many themes including artificial intelligence and data mining applications, machine learning and applications, deep learning technology for big data analytics, and modeling, simulation, and security with big data. It is a valuable resource for researchers in the area of big data analytics and its applications.
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.