Big Data Analysis New Algorithms For A New Society By Nathalie Japkowicz

This book reviews the theoretical concepts, leading-edge techniques and practical tools involved in the latest multi-disciplinary approaches addressing the challenges of big data. Illuminating perspectives from both academia and industry are presented by an international selection of experts in big data science. Topics and features: describes the innovative advances in theoretical aspects of big data, predictive analytics and cloud-based architectures; examines the applications and implementations that utilize big data in cloud architectures; surveys the state of the art in architectural approaches to the provision of cloud-based big data analytics functions; identifies potential research directions and technologies to facilitate the realization of emerging business models through big data approaches; provides relevant theoretical frameworks, empirical research findings, and numerous case studies; discusses real-world applications of algorithms and techniques to address the challenges of big datasets.
This textbook on practical data analytics unites fundamental principles, algorithms, and data. Algorithms are the keystone of data analytics and the focal point of this textbook. Clear and intuitive explanations of the mathematical and statistical foundations make the algorithms transparent. But practical data analytics requires more than just the foundations. Problems and data are enormously variable and only the most elementary

of algorithms can be used without modification. Programming fluency and experience with real and challenging data is indispensable and so the reader is immersed in Python and R and real data analysis. By the end of the book, the reader will have gained the ability to adapt algorithms to new problems and carry out innovative analyses. This book has three parts:(a) Data Reduction: Begins with the concepts of data reduction, data maps, and information extraction. The second chapter introduces associative statistics, the mathematical foundation of scalable algorithms and distributed computing. Practical aspects of distributed computing is the subject of the Hadoop and MapReduce chapter.(b) Extracting Information from Data: Linear regression and data visualization are the principal topics of Part II. The authors dedicate a chapter to the critical domain of Healthcare Analytics for an extended example of practical data analytics. The algorithms and analytics will be of much interest to practitioners interested in utilizing the large and unwieldly data sets of the Centers for Disease Control and Prevention's Behavioral Risk Factor Surveillance System.(c) Predictive Analytics Two foundational and widely used algorithms, k-nearest neighbors and naive Bayes, are developed in detail. A chapter is dedicated to forecasting. The last chapter focuses on streaming data and uses publicly accessible data streams originating from the Twitter API and the NASDAQ stock market in the tutorials. This book is intended for a one- or two-semester course in data analytics for upper-division undergraduate and graduate students in mathematics, statistics, and

computer science. The prerequisites are kept low, and students with one or two courses in probability or statistics, an exposure to vectors and matrices, and a programming course will have no difficulty. The core material of every chapter is accessible to all with these prerequisites. The chapters often expand at the close with innovations of interest to practitioners of data science. Each chapter includes exercises of varying levels of difficulty. The text is eminently suitable for self-study and an exceptional resource for practitioners.
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. Disk-Based Algorithms for Big Data is a product of recent advances in the areas of big data, data analytics, and the underlying file systems and data management algorithms used to support the storage and analysis of massive data collections. The book discusses hard disks and their impact on data management, since Hard Disk Drives continue to be common in large data clusters. It also explores ways to store and retrieve data though primary and secondary indices. This includes a review of different in-memory sorting and searching algorithms that build a foundation for more sophisticated on-disk approaches like mergesort, B-trees, and extendible hashing. Following this introduction, the book transitions to more recent topics, including advanced storage technologies like solid-state drives and holographic storage; peer-to-peer (P2P) communication; large file systems and query languages like Hadoop/HDFS,

Hive, Cassandra, and Presto; and NoSQL databases like Neo4j for graph structures and MongoDB for unstructured document data. Designed for senior undergraduate and graduate students, as well as professionals, this book is useful for anyone interested in understanding the foundations and advances in big data storage and management, and big data analytics. About the Author Dr. Christopher G. Healey is a tenured Professor in the Department of Computer Science and the Goodnight Distinguished Professor of Analytics in the Institute for Advanced Analytics, both at North Carolina State University in Raleigh, North Carolina. He has published over 50 articles in major journals and conferences in the areas of visualization, visual and data analytics, computer graphics, and artificial intelligence. He is a recipient of the National Science Foundation's CAREER Early Faculty Development Award and the North Carolina State University Outstanding Instructor Award. He is a Senior Member of the Association for Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE), and an Associate Editor of ACM Transaction on Applied Perception, the leading worldwide journal on the application of human perception to issues in computer science.
This book provides the reader with a basic understanding of the formal concepts of the cluster, clustering, partition, cluster analysis etc. The book explains feature-based, graph-based and spectral clustering methods and discusses their formal similarities and differences. Understanding the related formal concepts is particularly vital in the epoch of Big Data; due to the volume and characteristics of the data, it is no longer feasible to

predominantly rely on merely viewing the data when facing a clustering problem. Usually clustering involves choosing similar objects and grouping them together. To facilitate the choice of similarity measures for complex and big data, various measures of object similarity, based on quantitative (like numerical measurement results) and qualitative features (like text), as well as combinations of the two, are described, as well as graph-based similarity measures for (hyper) linked objects and measures for multilayered graphs. Numerous variants demonstrating how such similarity measures can be exploited when defining clustering cost functions are also presented. In addition, the book provides an overview of approaches to handling large collections of objects in a reasonable time. In particular, it addresses grid-based methods, sampling methods, parallelization via Map-Reduce, usage of tree-structures, random projections and various heuristic approaches, especially those used for community detection. Find the right big data solution for your business ororganization Big data management is one of the major challenges facingbusiness, industry, and not-for-profit organizations. Data setssuch as customer transactions for a mega-retailer, weather patternsmonitored by meteorologists, or social network activity can quicklyoutpace the capacity of traditional data management tools. If youneed to develop or manage big data solutions, you'll appreciate howthese four experts define, explain, and guide you through this newand often confusing concept. You'll learn what it is, why itmatters, and how to choose and implement solutions that work. Effectively managing big data is an issue of

growing importanceto businesses, not-for-profit organizations, government, and ITprofessionals Authors are experts in information management, big data, and avariety of solutions Explains big data in detail and discusses how to select andimplement a solution, security concerns to consider, data storageand presentation issues, analytics, and much more Provides essential information in a no-nonsense,easy-to-understand style that is empowering Big Data For Dummies cuts through the confusion and helpsyou take charge of big data solutions for your organization.

This volume explores the diverse applications of advanced tools and technologies of the emerging field of big data and their evidential value in business. It examines the role of analytics tools and methods of using big data in strengthening businesses to meet today's information challenges and shows how businesses can adapt big data for effective businesses practices. This volume shows how big data and the use of data analytics is being effectively adopted more frequently, especially in companies that are looking for new methods to develop smarter capabilities and tackle challenges in dynamic processes. Many illustrative case studies are presented that highlight how companies in every sector are now focusing on harnessing data to create a new way of doing business.

?This book includes 2 Manuscripts? Are you looking for new ways to grow your business, with resources you already have? Do you want to know how the big players like Netflix, Amazon, or Shopify use data analytics to MULTIPLY their growth? Keep

listening to learn how to use data analytics to maximize YOUR business.
Big Data Analytics in Cyber-Physical Systems: Machine Learning for the Internet of Things 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 used to evaluate the extracted data of those systems. Each chapter provides the reader with a broad list of data analytics and machine learning methods for multiple IoT applications. Additionally, this volume addresses the educational 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. . Bridges the gap between IoT, CPS, and mathematical modelling. Features numerous use cases that discuss how concepts are applied in different domains and applications. Provides "best practices", "winning stories" and "real-world examples" to complement innovation. Includes highlights of mathematical foundations of signal processing and machine learning in CPS and IoT.
As today's organizations are capturing exponentially larger amounts of data than ever, now is the time for organizations to rethink how they digest that data.

Through advanced algorithms and analytics techniques, organizations can harness this data, discover hidden patterns, and use the newly acquired knowledge to achieve competitive advantages.Pre "The chapters in this volume offer useful case studies, technical roadmaps, lessons learned, and a few prescriptions todo this, avoid that.'"-From the Foreword by Joe LaCugna, Ph.D., Enterprise Analytics and Business Intelligence, Starbucks Coffee CompanyWith the growing barrage of "big data," it becomes vitally important for organizations to mak This edited volume is devoted to Big Data Analysis from a Machine Learning standpoint as presented by some of the most eminent researchers in this area. It demonstrates that Big Data Analysis opens up new research problems which were either never considered before, or were only considered within a limited range. In addition to providing methodological discussions on the principles of mining Big Data and the difference between traditional statistical data analysis and newer computing frameworks, this book presents recently developed algorithms affecting such areas as business, financial forecasting, human mobility, the Internet of Things, information networks, bioinformatics, medical systems and life science. It explores, through a number of specific examples, how the study of Big Data Analysis has evolved and how it has started and will

most likely continue to affect society. While the benefits brought upon by Big
Data Analysis are underlined, the book also discusses some of the warnings that
have been issued concerning the potential dangers of Big Data Analysis along
with its pitfalls and challenges.

Big Data Analytics with Spark is a step-by-step guide for learning Spark, which is
an open-source fast and general-purpose cluster computing framework for large-
scale data analysis. You will learn how to use Spark for different types of big data
analytics projects, including batch, interactive, graph, and stream data analysis
as well as machine learning. In addition, this book will help you become a much
sought-after Spark expert. Spark is one of the hottest Big Data technologies. The
amount of data generated today by devices, applications and users is exploding.
Therefore, there is a critical need for tools that can analyze large-scale data and
unlock value from it. Spark is a powerful technology that meets that need. You
can, for example, use Spark to perform low latency computations through the use
of efficient caching and iterative algorithms; leverage the features of its shell for
easy and interactive Data analysis; employ its fast batch processing and low
latency features to process your real time data streams and so on. As a result,
adoption of Spark is rapidly growing and is replacing Hadoop MapReduce as the
technology of choice for big data analytics. This book provides an introduction to

Spark and related big-data technologies. It covers Spark core and its add-on libraries, including Spark SQL, Spark Streaming, GraphX, and MLlib. Big Data Analytics with Spark is therefore written for busy professionals who prefer learning a new technology from a consolidated source instead of spending countless hours on the Internet trying to pick bits and pieces from different sources. The book also provides a chapter on Scala, the hottest functional programming language, and the program that underlies Spark. You'll learn the basics of functional programming in Scala, so that you can write Spark applications in it. What's more, Big Data Analytics with Spark provides an introduction to other big data technologies that are commonly used along with Spark, like Hive, Avro, Kafka and so on. So the book is self-sufficient; all the technologies that you need to know to use Spark are covered. The only thing that you are expected to know is programming in any language. There is a critical shortage of people with big data expertise, so companies are willing to pay top dollar for people with skills in areas like Spark and Scala. So reading this book and absorbing its principles will provide a boost—possibly a big boost—to your career.

This book "Probabilistic Data Structures" is an Introduction to Probabilistic Data Structures and aims to introduce the readers to ideas of randomness in Data

Structure design. Contents of this book: • Preface • Introduction to Probabilistic Data Structures • List of Probabilistic Data Structures • Probabilistic Algorithms and Link with Data Structures • Basic Probabilistic Data Structures • Count Min Sketch • MinHash • LogLog • Bloom Filter • Skip List • Significance in Real Life/ Conclusion It is easier to understand randomness in algorithms with examples such as randomly splitting array in Quick Sort but most programmers fail to realize that Data Structures can be probabilistic as well. In this, not only the answer is probabilistic but also the structure. In fact, Google's Chrome browser uses a Probabilistic Data Structure within it. Read on to find out which data structure it is and how it is used. The ideas have been presented in a simple language (avoiding technical terms) with intuitive insights which will help anyone to go through this book and enjoy the knowledge. This knowledge will help you to design better systems suited for real use.
--- Authors: Aditya Chatterjee, Ethan Z. Booker Aditya is a Founding member at OpenGenus; Ethan has been an Intern at OpenGenus and a student at University of Wisconsin, La Crosse; In this book readers will find technological discussions on the existing and emerging technologies across the different stages of the big data value chain. They will learn about legal aspects of big data, the social impact, and about

education needs and requirements. And they will discover the business perspective and how big data technology can be exploited to deliver value within different sectors of the economy. The book is structured in four parts: Part I "The Big Data Opportunity" explores the value potential of big data with a particular focus on the European context. It also describes the legal, business and social dimensions that need to be addressed, and briefly introduces the European Commission's BIG project. Part II "The Big Data Value Chain" details the complete big data lifecycle from a technical point of view, ranging from data acquisition, analysis, curation and storage, to data usage and exploitation. Next, Part III "Usage and Exploitation of Big Data" illustrates the value creation possibilities of big data applications in various sectors, including industry, healthcare, finance, energy, media and public services. Finally, Part IV "A Roadmap for Big Data Research" identifies and prioritizes the cross-sectorial requirements for big data research, and outlines the most urgent and challenging technological, economic, political and societal issues for big data in Europe. This compendium summarizes more than two years of work performed by a leading group of major European research centers and industries in the context of the BIG project. It brings together research findings, forecasts and estimates related to this challenging technological context that is becoming the major axis of the

new digitally transformed business environment.

Currently many different application areas for Big Data (BD) and Machine Learning (ML) are being explored. These promising application areas for BD/ML are the social sites, search engines, multimedia sharing sites, various stock exchange sites, online gaming, online survey sites and various news sites, and so on. To date, various use-cases for this application area are being researched and developed. Software applications are already being published and used in various settings from education and training to discover useful hidden patterns and other information like customer choices and market trends that can help organizations make more informed and customer-oriented business decisions. Combining BD with ML will provide powerful, largely unexplored application areas that will revolutionize practice in Videos Surveillance, Social Media Services, Email Spam and Malware Filtering, Online Fraud Detection, and so on. It is very important to continuously monitor and understand these effects from safety and societal point of view. Hence, the main purpose of this book is for researchers, software developers and practitioners, academicians and students to showcase novel use-cases and applications, present empirical research results from user-centered qualitative and quantitative experiments of these new applications, and facilitate a discussion forum to explore the latest trends in big data and machine

learning by providing algorithms which can be trained to perform interdisciplinary techniques such as statistics, linear algebra, and optimization and also create automated systems that can sift through large volumes of data at high speed to make predictions or decisions without human intervention

As today's organizations are capturing exponentially larger amounts of data than ever, now is the time for organizations to rethink how they digest that data. Through advanced algorithms and analytics techniques, organizations can harness this data, discover hidden patterns, and use the newly acquired knowledge to achieve competitive advantages. Presenting the contributions of leading experts in their respective fields, Big Data: Algorithms, Analytics, and Applications bridges the gap between the vastness of Big Data and the appropriate computational methods for scientific and social discovery. It covers fundamental issues about Big Data, including efficient algorithmic methods to process data, better analytical strategies to digest data, and representative applications in diverse fields, such as medicine, science, and engineering. The book is organized into five main sections: Big Data Management—considers the research issues related to the management of Big Data, including indexing and scalability aspects Big Data Processing—addresses the problem of processing Big Data across a wide range of resource-intensive computational settings Big Data Stream Techniques and Algorithms—explores research issues regarding the management and mining of Big Data in streaming environments Big Data

Privacy—focuses on models, techniques, and algorithms for preserving Big Data privacy Big Data Applications—illustrates practical applications of Big Data across several domains, including finance, multimedia tools, biometrics, and satellite Big Data processing Overall, the book reports on state-of-the-art studies and achievements in algorithms, analytics, and applications of Big Data. It provides readers with the basis for further efforts in this challenging scientific field that will play a leading role in next-generation database, data warehousing, data mining, and cloud computing research. It also explores related applications in diverse sectors, covering technologies for media/data communication, elastic media/data storage, cross-network media/data fusion, and SaaS.

The first book to present the common mathematical foundations of big data analysis across a range of applications and technologies. Today, the volume, velocity, and variety of data are increasing rapidly across a range of fields, including Internet search, healthcare, finance, social media, wireless devices, and cybersecurity. Indeed, these data are growing at a rate beyond our capacity to analyze them. The tools—including spreadsheets, databases, matrices, and graphs—developed to address this challenge all reflect the need to store and operate on data as whole sets rather than as individual elements. This book presents the common mathematical foundations of these data sets that apply across many applications and technologies. Associative arrays unify and simplify data, allowing readers to look past the differences among the various tools and

leverage their mathematical similarities in order to solve the hardest big data challenges. The book first introduces the concept of the associative array in practical terms, presents the associative array manipulation system D4M (Dynamic Distributed Dimensional Data Model), and describes the application of associative arrays to graph analysis and machine learning. It provides a mathematically rigorous definition of associative arrays and describes the properties of associative arrays that arise from this definition. Finally, the book shows how concepts of linearity can be extended to encompass associative arrays. Mathematics of Big Data can be used as a textbook or reference by engineers, scientists, mathematicians, computer scientists, and software engineers who analyze big data.

Big Data Analytics and Intelligence is essential reading for researchers and experts working in the fields of health care, data science, analytics, the internet of things, and information retrieval.

This volume is aiming at a wide range of readers and researchers in the area of Big Data by presenting the recent advances in the fields of Big Data Analysis, as well as the techniques and tools used to analyze it. The book includes 10 distinct chapters providing a concise introduction to Big Data Analysis and recent Techniques and Environments for Big Data Analysis. It gives insight into how the expensive fitness evaluation of evolutionary learning can play a vital role in big data analysis by adopting Parallel, Grid, and Cloud computing environments.

A revelatory exploration of the hottest trend in technology and the dramatic impact it will have on the economy, science, and society at large. Which paint color is most likely to tell you that a used car is in good shape? How can officials identify the most dangerous New York City manholes before they explode? And how did Google searches predict the spread of the H1N1 flu outbreak? The key to answering these questions, and many more, is big data. "Big data" refers to our burgeoning ability to crunch vast collections of information, analyze it instantly, and draw sometimes profoundly surprising conclusions from it. This emerging science can translate myriad phenomena—from the price of airline tickets to the text of millions of books—into searchable form, and uses our increasing computing power to unearth epiphanies that we never could have seen before. A revolution on par with the Internet or perhaps even the printing press, big data will change the way we think about business, health, politics, education, and innovation in the years to come. It also poses fresh threats, from the inevitable end of privacy as we know it to the prospect of being penalized for things we haven't even done yet, based on big data's ability to predict our future behavior. In this brilliantly clear, often surprising work, two leading experts explain what big data is, how it will change our lives, and what we can do to protect ourselves from its hazards. Big Data is the first big book about the next big thing. www.big-data-book.com

We are living in the dawn of what has been termed as the "Fourth Industrial Revolution," which is marked through the emergence of "cyber-physical systems"

where software interfaces seamlessly over networks with physical systems, such as sensors, smartphones, vehicles, power grids or buildings, to create a new world of Internet of Things (IoT). Data and information are fuel of this new age where powerful analytics algorithms burn this fuel to generate decisions that are expected to create a smarter and more efficient world for all of us to live in. This new area of technology has been defined as Big Data Science and Analytics, and the industrial and academic communities are realizing this as a competitive technology that can generate significant new wealth and opportunity. Big data is defined as collections of datasets whose volume, velocity or variety is so large that it is difficult to store, manage, process and analyze the data using traditional databases and data processing tools. Big data science and analytics deals with collection, storage, processing and analysis of massive-scale data. Industry surveys, by Gartner and e-Skills, for instance, predict that there will be over 2 million job openings for engineers and scientists trained in the area of data science and analytics alone, and that the job market is in this area is growing at a 150 percent year-over-year growth rate. We have written this textbook, as part of our expanding "A Hands-On Approach"(TM) series, to meet this need at colleges and universities, and also for big data service providers who may be interested in offering a broader perspective of this emerging field to accompany their customer and developer training programs. The typical reader is expected to have completed a couple of courses in programming using traditional high-level languages at the college-level, and

is either a senior or a beginning graduate student in one of the science, technology, engineering or mathematics (STEM) fields. An accompanying website for this book contains additional support for instruction and learning (www.big-data-analytics-book.com) The book is organized into three main parts, comprising a total of twelve chapters. Part I provides an introduction to big data, applications of big data, and big data science and analytics patterns and architectures. A novel data science and analytics application system design methodology is proposed and its realization through use of open-source big data frameworks is described. This methodology describes big data analytics applications as realization of the proposed Alpha, Beta, Gamma and Delta models, that comprise tools and frameworks for collecting and ingesting data from various sources into the big data analytics infrastructure, distributed filesystems and non-relational (NoSQL) databases for data storage, and processing frameworks for batch and real-time analytics. This new methodology forms the pedagogical foundation of this book. Part II introduces the reader to various tools and frameworks for big data analytics, and the architectural and programming aspects of these frameworks, with examples in Python. We describe Publish-Subscribe messaging frameworks (Kafka & Kinesis), Source-Sink connectors (Flume), Database Connectors (Sqoop), Messaging Queues (RabbitMQ, ZeroMQ, RestMQ, Amazon SQS) and custom REST, WebSocket and MQTT-based connectors. The reader is introduced to data storage, batch and real-time analysis, and interactive querying frameworks including HDFS, Hadoop,

MapReduce, YARN, Pig, Oozie, Spark, Solr, HBase, Storm, Spark Streaming, Spark SQL, Hive, Amazon Redshift and Google BigQuery. Also described are serving databases (MySQL, Amazon DynamoDB, Cassandra, MongoDB) and the Django Python web framework. Part III introduces the reader to various machine learning algorithms with examples using the Spark MLlib and H2O frameworks, and visualizations using frameworks such as Lightning, Pygal and Seaborn. If you are ready to dive into the MapReduce framework for processing large datasets, this practical book takes you step by step through the algorithms and tools you need to build distributed MapReduce applications with Apache Hadoop or Apache Spark. Each chapter provides a recipe for solving a massive computational problem, such as building a recommendation system. You'll learn how to implement the appropriate MapReduce solution with code that you can use in your projects. Dr. Mahmoud Parsian covers basic design patterns, optimization techniques, and data mining and machine learning solutions for problems in bioinformatics, genomics, statistics, and social network analysis. This book also includes an overview of MapReduce, Hadoop, and Spark. Topics include: Market basket analysis for a large set of transactions Data mining algorithms (K-means, KNN, and Naive Bayes) Using huge genomic data to sequence DNA and RNA Naive Bayes theorem and Markov chains for data and market prediction Recommendation algorithms and pairwise document similarity Linear regression, Cox regression, and Pearson correlation Allelic frequency and mining DNA

Social network analysis (recommendation systems, counting triangles, sentiment analysis)

This book presents machine learning models and algorithms to address big data classification problems. Existing machine learning techniques like the decision tree (a hierarchical approach), random forest (an ensemble hierarchical approach), and deep learning (a layered approach) are highly suitable for the system that can handle such problems. This book helps readers, especially students and newcomers to the field of big data and machine learning, to gain a quick understanding of the techniques and technologies; therefore, the theory, examples, and programs (Matlab and R) presented in this book have been simplified, hardcoded, repeated, or spaced for improvements. They provide vehicles to test and understand the complicated concepts of various topics in the field. It is expected that the readers adopt these programs to experiment with the examples, and then modify or write their own programs toward advancing their knowledge for solving more complex and challenging problems. The presentation format of this book focuses on simplicity, readability, and dependability so that both undergraduate and graduate students as well as new researchers, developers, and practitioners in this field can easily trust and grasp the concepts, and learn them effectively. It has been written to reduce the mathematical complexity and help the vast majority of readers to understand the topics and get interested in the field. This book consists of four parts, with the total of 14 chapters. The first part mainly focuses on the

topics that are needed to help analyze and understand data and big data. The second part covers the topics that can explain the systems required for processing big data. The third part presents the topics required to understand and select machine learning techniques to classify big data. Finally, the fourth part concentrates on the topics that explain the scaling-up machine learning, an important solution for modern big data problems.

Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications covers timely topics, including the neural network (NN), particle swarm optimization (PSO), evolutionary algorithm (GA), fuzzy sets (FS) and rough sets (RS), etc. Furthermore, the book highlights recent research on representative techniques to elaborate how a data-centric system formed a powerful platform for the processing of cloud hosted multimedia big data and how it could be analyzed, processed and characterized by CI. The book also provides a view on how techniques in CI can offer solutions in modeling, relationship pattern recognition, clustering and other problems in bioengineering. It is written for domain experts and developers who want to understand and explore the application of computational intelligence aspects (opportunities and challenges) for design and development of a data-centric system in the context of multimedia cloud, big data era and its related applications, such as smarter healthcare,

homeland security, traffic control trading analysis and telecom, etc. Researchers
and PhD students exploring the significance of data centric systems in the next
paradigm of computing will find this book extremely useful. Presents a brief
overview of computational intelligence paradigms and its significant role in
application domains Illustrates the state-of-the-art and recent developments in
the new theories and applications of CI approaches Familiarizes the reader with
computational intelligence concepts and technologies that are successfully used
in the implementation of cloud-centric multimedia services in massive data
processing Provides new advances in the fields of CI for bio-engineering
application

Human attention is in the highest demand it has ever been. The drastic increase
in available information has compelled individuals to find a way to sift through the
media that is literally at their fingertips. Content recommendation systems have
emerged as the technological solution to this social and informational problem,
but they've also created a bigger crisis in confirming our biases by showing us
only, and exactly, what it predicts we want to see. Data versus Democracy
investigates and explores how, in the era of social media, human cognition,
algorithmic recommendation systems, and human psychology are all working
together to reinforce (and exaggerate) human bias. The dangerous confluence of

these factors is driving media narratives, influencing opinions, and possibly changing election results. In this book, algorithmic recommendations, clickbait, familiarity bias, propaganda, and other pivotal concepts are analyzed and then expanded upon via fascinating and timely case studies: the 2016 US presidential election, Ferguson, GamerGate, international political movements, and more events that come to affect every one of us. What are the implications of how we engage with information in the digital age? Data versus Democracy explores this topic and an abundance of related crucial questions. We live in a culture vastly different from any that has come before. In a society where engagement is currency, we are the product. Understanding the value of our attention, how organizations operate based on this concept, and how engagement can be used against our best interests is essential in responsibly equipping ourselves against the perils of disinformation. Who This Book Is For Individuals who are curious about how social media algorithms work and how they can be manipulated to influence culture. Social media managers, data scientists, data administrators, and educators will find this book particularly relevant to their work. This unique volume is an introduction for computer scientists, including a formal study of theoretical algorithms for Big Data applications, which allows them to work on such algorithms in the future. It also serves as a useful reference guide

for the general computer science population, providing a comprehensive overview of the fascinating world of such algorithms.To achieve these goals, the algorithmic results presented have been carefully chosen so that they demonstrate the important techniques and tools used in Big Data algorithms, and yet do not require tedious calculations or a very deep mathematical background. This volume comprises six well-versed contributed chapters devoted to report the latest fi ndings on the applications of machine learning for big data analytics. Big data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them. The possible challenges in this direction include capture, storage, analysis, data curation, search, sharing, transfer, visualization, querying, updating and information privacy. Big data analytics is the process of examining large and varied data sets - i.e., big data - to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful information that can help organizations make more-informed business decisions. This volume is intended to be used as a reference by undergraduate and post graduate students of the disciplines of computer science, electronics and telecommunication, information science and electrical engineering. THE SERIES: FRONTIERS IN COMPUTATIONAL INTELLIGENCE The series Frontiers In Computational Intelligence is envisioned

to provide comprehensive coverage and understanding of cutting edge research in computational intelligence. It intends to augment the scholarly discourse on all topics relating to the advances in artifi cial life and machine learning in the form of metaheuristics, approximate reasoning, and robotics. Latest research fi ndings are coupled with applications to varied domains of engineering and computer sciences. This field is steadily growing especially with the advent of novel machine learning algorithms being applied to different domains of engineering and technology. The series brings together leading researchers that intend to continue to advance the fi eld and create a broad knowledge about the most recent research.

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.

Serving as a flagship driver towards advance research in the area of Big Data platforms and applications, this book provides a platform for the dissemination of advanced topics of theory, research efforts and analysis, and implementation oriented on methods, techniques and performance evaluation. In 23 chapters, several important formulations of the architecture design, optimization techniques, advanced analytics methods, biological, medical and social media applications are presented. These chapters discuss the research of members from the ICT COST Action IC1406 High-Performance Modelling and Simulation

for Big Data Applications (cHiPSet). This volume is ideal as a reference for students, researchers and industry practitioners working in or interested in joining interdisciplinary works in the areas of intelligent decision systems using emergent distributed computing paradigms. It will also allow newcomers to grasp the key concerns and their potential solutions.

This book highlights major issues related to big data analysis using computational intelligence techniques, mostly interdisciplinary in nature. It comprises chapters on computational intelligence technologies, such as neural networks and learning algorithms, evolutionary computation, fuzzy systems and other emerging techniques in data science and big data, ranging from methodologies, theory and algorithms for handling big data, to their applications in bioinformatics and related disciplines. The book describes the latest solutions, scientific results and methods in solving intriguing problems in the fields of big data analytics, intelligent agents and computational intelligence. It reflects the state of the art research in the field and novel applications of new processing techniques in computer science. This book is useful to both doctoral students and researchers from computer science and engineering fields and bioinformatics related domains.

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

Master alternative Big Data technologies that can do what Hadoop can't: real-time analytics and iterative machine learning. When most technical professionals think of Big Data analytics today, they think of Hadoop. But there are many cutting-edge applications that Hadoop isn't well suited for, especially real-time analytics and contexts requiring the use of iterative machine learning algorithms. Fortunately, several powerful new technologies have been developed specifically for use cases such as these. Big Data Analytics Beyond Hadoop is the first guide specifically designed to help you take the next steps beyond Hadoop. Dr. Vijay Srinivas Agneeswaran introduces the breakthrough Berkeley Data Analysis Stack (BDAS) in detail, including its motivation, design, architecture, Mesos cluster management, performance, and more. He presents realistic use cases and up-to-date example code for: Spark, the next generation in-memory computing technology from UC Berkeley Storm, the parallel real-time Big Data analytics technology from Twitter GraphLab, the next-generation graph processing paradigm from CMU and the University of Washington (with comparisons to alternatives such as Pregel and Piccolo) Halo also offers architectural and design guidance and code sketches for scaling machine learning algorithms to Big Data, and then realizing them in real-time. He concludes by previewing emerging trends, including real-time video analytics, SDNs, and even Big Data governance, security, and privacy issues. He identifies intriguing startups and new research possibilities, including BDAS extensions and cutting-edge model-driven analytics. Big

Data Analytics Beyond Hadoop is an indispensable resource for everyone who wants to reach the cutting edge of Big Data analytics, and stay there: practitioners, architects, programmers, data scientists, researchers, startup entrepreneurs, and advanced students.

Unique prospective on the big data analytics phenomenon for both business and IT professionals The availability of Big Data, low-cost commodity hardware and new information management and analytics software has produced a unique moment in the history of business. The convergence of these trends means that we have the capabilities required to analyze astonishing data sets quickly and cost-effectively for the first time in history. These capabilities are neither theoretical nor trivial. They represent a genuine leap forward and a clear opportunity to realize enormous gains in terms of efficiency, productivity, revenue and profitability. The Age of Big Data is here, and these are truly revolutionary times. This timely book looks at cutting-edge companies supporting an exciting new generation of business analytics. Learn more about the trends in big data and how they are impacting the business world (Risk, Marketing, Healthcare, Financial Services, etc.) Explains this new technology and how companies can use them effectively to gather the data that they need and glean critical insights Explores relevant topics such as data privacy, data visualization, unstructured data, crowd sourcing data scientists, cloud computing for big data, and much more.

The second edition of a comprehensive introduction to machine learning approaches used in predictive data analytics, covering both theory and practice. Machine learning is often used to build predictive models by extracting patterns from large datasets. These models are used in predictive data analytics applications including price prediction, risk assessment, predicting customer behavior, and document classification. This introductory textbook offers a detailed

and focused treatment of the most important machine learning approaches used in predictive data analytics, covering both theoretical concepts and practical applications. Technical and mathematical material is augmented with explanatory worked examples, and case studies illustrate the application of these models in the broader business context. This second edition covers recent developments in machine learning, especially in a new chapter on deep learning, and two new chapters that go beyond predictive analytics to cover unsupervised learning and reinforcement learning.

This book provides a comprehensive survey of techniques, technologies and applications of Big Data and its analysis. The Big Data phenomenon is increasingly impacting all sectors of business and industry, producing an emerging new information ecosystem. On the applications front, the book offers detailed descriptions of various application areas for Big Data Analytics in the important domains of Social Semantic Web Mining, Banking and Financial Services, Capital Markets, Insurance, Advertisement, Recommendation Systems, Bio-Informatics, the IoT and Fog Computing, before delving into issues of security and privacy. With regard to machine learning techniques, the book presents all the standard algorithms for learning – including supervised, semi-supervised and unsupervised techniques such as clustering and reinforcement learning techniques to perform collective Deep Learning. Multi-layered and nonlinear learning for Big Data are also covered. In turn, the book highlights real-life case studies on successful implementations of Big Data Analytics at large IT companies such as Google, Facebook, LinkedIn and Microsoft. Multi-sectorial case studies on domain-based companies such as Deutsche Bank, the power provider Opower, Delta Airlines and a Chinese City Transportation application represent a valuable addition. Given its comprehensive

coverage of Big Data Analytics, the book offers a unique resource for undergraduate and graduate students, researchers, educators and IT professionals alike.

Now in its second edition, this book focuses on practical algorithms for mining data from even the largest datasets.

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!

This book aims to provide some insights into recently developed bio-inspired algorithms within recent emerging trends of fog computing, sentiment analysis, and data streaming as well as to provide a more comprehensive approach to the big data management from pre-processing to analytics to visualization phases. The subject area of this book is within the realm of computer science, notably algorithms (meta-heuristic and, more particularly, bio-inspired algorithms). Although application domains of these new algorithms may be mentioned, the scope of this book is not on the application of algorithms to specific or general domains but to provide an update on recent research trends for bio-inspired algorithms within a specific application domain or emerging area. These areas include data streaming, fog computing, and phases of big data management. One of the reasons for writing this book is that the bio-inspired approach does not receive much attention but shows considerable promise and diversity in terms of approach of many issues in big data and streaming. Some novel approaches of this book are the use of these algorithms to all phases of data management (not just a particular phase such as data mining or business intelligence as many books focus on); effective demonstration of the effectiveness of a selected algorithm within a chapter against comparative algorithms using the experimental method. Another novel approach is a brief overview and evaluation of traditional algorithms, both sequential and parallel, for use in data mining, in order to provide an overview of existing algorithms in use. This overview

complements a further chapter on bio-inspired algorithms for data mining to enable readers to make a more suitable choice of algorithm for data mining within a particular context. In all chapters, references for further reading are provided, and in selected chapters, the author also include ideas for future research.

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

Copyright: 4791d441bca2909696da8833a1687038