# Proceedings Of Workshop Big Data Analytic Technology For

This book addresses the usefulness of knowledge discovery through data mining. With this aim, contributors from different fields propose concrete problems and applications showing how data mining and discovering embedded knowledge from raw data can be beneficial to social organizations, domestic spheres, and ICT markets. Data mining or knowledge discovery in databases (KDD) has received increasing interest due to its focus on transforming large amounts of data into novel, valid, useful, and structured knowledge by detecting concealed patterns and relationships. The concept of knowledge is broad and speculative and has promoted epistemological debates in western philosophies. The intensified interest in knowledge management and data mining stems from the difficulty in identifying computational models able to approximate human behaviors and abilities in resolving organizational, social, and physical problems. Current ICT interfaces are not yet adequately advanced to support and simulate the abilities of physicians, teachers, assistants or housekeepers in domestic spheres. And unlike in industrial contexts where abilities are routinely applied, the domestic world is continuously changing and unpredictable. There are challenging questions in this field: Can knowledge locked in conventions, rules of conduct, common sense, ethics, emotions, laws, cultures, and experiences be mined from data? Is it acceptable for automatic systems displaying emotional behaviors to govern complex interactions based solely on the mining of large volumes of data? Discussing multidisciplinary themes, the book proposes computational models able to approximate, to a certain degree, human behaviors and abilities in resolving organizational, social, and

physical problems. The innovations presented are of primary importance for: a. The academic research community b. The ICT market c. Ph.D. students and early stage researchers d. Schools, hospitals, rehabilitation and assisted-living centers e. Representatives from multimedia industries and standardization bodies

This book provides a comprehensive picture of fog computing technology, including of fog architectures, latency aware application management issues with real time requirements, security and privacy issues and fog analytics, in wide ranging application scenarios such as M2M device communication, smart homes, smart vehicles, augmented reality and transportation management. This book explores the research issues involved in the application of traditional shallow machine learning and deep learning techniques to big data analytics. It surveys global research advances in extending the conventional unsupervised or clustering algorithms, extending supervised and semi-supervised algorithms and association rule mining algorithms to big data Scenarios. Further it discusses the deep learning applications of big data analytics to fields of computer vision and speech processing, and describes applications such as semantic indexing and data tagging. Lastly it identifies 25 unsolved research problems and research directions in fog computing, as well as in the context of applying deep learning techniques to big data analytics, such as dimensionality reduction in high-dimensional data and improved formulation of data abstractions along with possible directions for their solutions.

ICSE '16: 38th International Conference on Software Engineering May 14, 2016-May 22, 2016 Austin, USA. You can view more information about this proceeding and all of ACMs other published conference proceedings from the ACM Digital Library: http://www.acm.org/dl.

Emerging techniques in data analytics, including machine

learning and artificial intelligence, offer exciting opportunities for advancing scientific discovery and innovation in materials science. Vast repositories of experimental data and sophisticated simulations are being utilized to predict material properties, design and test new compositions, and accelerate nearly every facet of traditional materials science. How can the materials science community take advantage of these opportunities while avoiding potential pitfalls? What roadblocks may impede progress in the coming years, and how might they be addressed? To explore these issues, the Workshop on Data Analytics and What It Means to the Materials Community was organized as part of a workshop series on Defense Materials, Manufacturing, and Its Infrastructure. Hosted by the National Academies of Sciences, Engineering, and Medicine, the 2-day workshop was organized around three main topics: materials design, data curation, and emerging applications. Speakers identified promising data analytics tools and their achievements to date, as well as key challenges related to dealing with sparse data and filling data gaps; decisions around data storage, retention, and sharing; and the need to access, combine, and use data from disparate sources. Participants discussed the complementary roles of simulation and experimentation and explored the many opportunities for data informatics to increase the efficiency of materials discovery, design, and testing by reducing the amount of experimentation required. With an eye toward the ultimate goal of enabling applications, attendees considered how to ensure that the benefits of data analytics tools carry through the entire materials development process, from exploration to validation, manufacturing, and use. This publication summarizes the presentations and discussion of the workshop.

The concept of utilizing big data to enable scientific discovery has generated tremendous excitement and investment from

both private and public sectors over the past decade, and expectations continue to grow. Using big data analytics to identify complex patterns hidden inside volumes of data that have never been combined could accelerate the rate of scientific discovery and lead to the development of beneficial technologies and products. However, producing actionable scientific knowledge from such large, complex data sets requires statistical models that produce reliable inferences (NRC, 2013). Without careful consideration of the suitability of both available data and the statistical models applied, analysis of big data may result in misleading correlations and false discoveries, which can potentially undermine confidence in scientific research if the results are not reproducible. In June 2016 the National Academies of Sciences, Engineering, and Medicine convened a workshop to examine critical challenges and opportunities in performing scientific inference reliably when working with big data. Participants explored new methodologic developments that hold significant promise and potential research program areas for the future. This publication summarizes the presentations and discussions from the workshop.

This book constitutes the thoroughly revised selected papers of the 4th and 5th workshops on Big Data Benchmarks, Performance Optimization, and Emerging Hardware, BPOE 4 and BPOE 5, held respectively in Salt Lake City, in March 2014, and in Hangzhou, in September 2014. The 16 papers presented were carefully reviewed and selected from 30 submissions. Both workshops focus on architecture and system support for big data systems, such as benchmarking; workload characterization; performance optimization and evaluation; emerging hardware.

As the availability of high-throughput data-collection technologies, such as information-sensing mobile devices, remote sensing, internet log records, and wireless sensor

networks has grown, science, engineering, and business have rapidly transitioned from striving to develop information from scant data to a situation in which the challenge is now that the amount of information exceeds a human's ability to examine, let alone absorb, it. Data sets are increasingly complex, and this potentially increases the problems associated with such concerns as missing information and other quality concerns, data heterogeneity, and differing data formats. The nation's ability to make use of data depends heavily on the availability of a workforce that is properly trained and ready to tackle high-need areas. Training students to be capable in exploiting big data requires experience with statistical analysis, machine learning, and computational infrastructure that permits the real problems associated with massive data to be revealed and, ultimately, addressed. Analysis of big data requires cross-disciplinary skills, including the ability to make modeling decisions while balancing trade-offs between optimization and approximation, all while being attentive to useful metrics and system robustness. To develop those skills in students, it is important to identify whom to teach, that is, the educational background, experience, and characteristics of a prospective data-science student; what to teach, that is, the technical and practical content that should be taught to the student; and how to teach, that is, the structure and organization of a data-science program. Training Students to Extract Value from Big Data summarizes a workshop convened in April 2014 by the National Research Council's Committee on Applied and Theoretical Statistics to explore how best to train students to use big data. The workshop explored the need for training and curricula and coursework that should be included. One impetus for the workshop was the current fragmented view of what is meant by analysis of big data, data analytics, or data science. New graduate programs are introduced regularly,

and they have their own notions of what is meant by those terms and, most important, of what students need to know to be proficient in data-intensive work. This report provides a variety of perspectives about those elements and about their integration into courses and curricula.

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

business analysts exploiting data science technologies, and it will also be a useful resource to research students in disciplines and courses related to these topics. Overall, readers will learn modern and effective data science solutions to create tangible innovations for economic and financial applications.

This book constitutes the refereed post-conference proceedings of the 7th International Conference on Big data Technologies and Applications, BDTA 2016, held in Seoul, South Korea, in November 2016. BDTA 2016 was collocated with the First International Workshop on Internet of Things, Social Network, and Security in Big Data, ISSB 2016 and the First International Workshop on Digital Humanity with Big Data, DiHuBiDa 2016. The 17 revised full papers were carefully reviewed and selected from 25 submissions and handle theoretical foundations and practical applications which premise the new generation of data analytics and engineering.

This handbook offers comprehensive coverage of recent advancements in Big Data technologies and related paradigms. Chapters are authored by international leading experts in the field, and have been reviewed and revised for maximum reader value. The volume consists of twenty-five chapters organized into four main parts. Part one covers the fundamental concepts of Big Data technologies including data curation mechanisms, data models, storage models, programming models and programming platforms. It also dives into the details of implementing Big SQL query engines and big stream processing systems. Part Two focuses on the semantic aspects of Big Data management including data

integration and exploratory ad hoc analysis in addition to structured querying and pattern matching techniques. Part Three presents a comprehensive overview of large scale graph processing. It covers the most recent research in large scale graph processing platforms, introducing several scalable graph querying and mining mechanisms in domains such as social networks. Part Four details novel applications that have been made possible by the rapid emergence of Big Data technologies such as Internet-of-Things (IOT), Cognitive Computing and SCADA Systems. All parts of the book discuss open research problems, including potential opportunities, that have arisen from the rapid progress of Big Data technologies and the associated increasing requirements of application domains. Designed for researchers, IT professionals and graduate students, this book is a timely contribution to the growing Big Data field. Big Data has been recognized as one of leading emerging technologies that will have a major contribution and impact on the various fields of science and varies aspect of the human society over the coming decades. Therefore, the content in this book will be an essential tool to help readers understand the development and future of the field.

With the amount of data in the world exploding, big data could generate significant value in the field of infectious disease. The increased use of social media provides an opportunity to improve public health surveillance systems and to develop predictive models. Advances in machine learning and crowdsourcing may also offer the possibility to gather information about disease dynamics, such as

contact patterns and the impact of the social environment. New, rapid, point-of-care diagnostics may make it possible to capture not only diagnostic information but also other potentially epidemiologically relevant information in real time. With a wide range of data available for analysis, decision-making and policy-making processes could be improved. While there are many opportunities for big data to be used for infectious disease research, operations, and policy, many challenges remain before it is possible to capture the full potential of big data. In order to explore some of the opportunities and issues associated with the scientific, policy, and operational aspects of big data in relation to microbial threats and public health, the National Academies of Sciences, Engineering, and Medicine convened a workshop in May 2016. Participants discussed a range of topics including preventing, detecting, and responding to infectious disease threats using big data and related analytics; varieties of data (including demographic, geospatial, behavioral, syndromic, and laboratory) and their broader applications; means to improve their collection, processing, utility, and validation; and approaches that can be learned from other sectors to inform big data strategies for infectious disease research, operations, and policy. This publication summarizes the presentations and discussions from the workshop.
"This book discusses the exponential growth of information size and the innovative methods for data capture, storage, sharing, and analysis for big data"--Provided by publisher.

Big Data in Materials Research and Development is the summary of a workshop convened by the National Research Council Standing Committee on Defense Materials Manufacturing and Infrastructure in February 2014 to discuss the impact of big data on materials and manufacturing. The materials science community would benefit from appropriate access to data and metadata for materials development, processing, application development, and application life cycles. Currently, that access does not appear to be sufficiently widespread, and many workshop participants captured the constraints and identified potential improvements to enable broader access to materials and manufacturing data and metadata. This report discusses issues in defense materials, manufacturing and infrastructure, including data ownership and access; collaboration and exploitation of big data's capabilities; and maintenance of data.

This book constitutes the thoroughly refereed post-workshop proceedings of the 6th International Workshop on Big Data Benchmarking, WBDB 2015, held in Toronto, ON, Canada, in June 2015 and the 7th International Workshop, WBDB 2015, held in New Delhi, India, in December 2015. The 8 full papers presented in this book were carefully reviewed and selected from 22 submissions. They deal with recent trends in big data and HPC convergence, new proposals for big data benchmarking, as well as tooling and performance results.

Refining the Concept of Scientific Inference When Working with Big DataProceedings of a WorkshopNational Academies

Press

The National Academies of Sciences, Engineering, and Medicine held the workshop Applying Big Data to Address the Social Determinants of Health in Oncology on October 28â€"29, 2019, in Washington, DC. This workshop examined social determinants of health (SDOH) in the context of cancer, and considered opportunities to effectively leverage big data to improve health equity and reduce disparities. The workshop featured presentations and discussion by experts in technology, oncology, and SDOH, as well as representatives from government, industry, academia, and health care systems. This publication summarizes the presentations and discussions from the workshop.

This book constitutes the thoroughly refereed proceedings of the Third International Conference on Big Data, Cloud and Applications, BDCA 2018, held in Kenitra, Morocco, in April 2018.The 45 revised full papers presented in this book were carefully selected from 99 submissions with a thorough double-blind review process. They focus on the following topics: big data, cloud computing, machine learning, deep learning, data analysis, neural networks, information system and social media, image processing and applications, and natural language processing.

Big Data is a large-sized and complex dataset, which cannot be managed using traditional data processing tools. Mining process of big data is the ability to extract valuable information from these large datasets.

This book constitutes the thoroughly refereed joint proceedings of the Third and Fourth Workshop on Big Data Benchmarking. The third WBDB was held in Xi'an, China, in July 2013 and the Fourth WBDB was held in San José, CA, USA, in October, 2013. The 15 papers presented in this book were carefully reviewed and selected from 33 presentations. They focus on big data benchmarks; applications and

scenarios; tools, systems and surveys.

The objective of this book is to introduce the basic concepts of big data computing and then to describe the total solution of big data problems using HPCC, an open-source computing platform. The book comprises 15 chapters broken into three parts. The first part, Big Data Technologies, includes introductions to big data concepts and techniques; big data analytics; and visualization and learning techniques. The second part, LexisNexis Risk Solution to Big Data, focuses on specific technologies and techniques developed at LexisNexis to solve critical problems that use big data analytics. It covers the open source High Performance Computing Cluster (HPCC Systems®) platform and its architecture, as well as parallel data languages ECL and KEL, developed to effectively solve big data problems. The third part, Big Data Applications, describes various data intensive applications solved on HPCC Systems. It includes applications such as cyber security, social network analytics including fraud, Ebola spread modeling using big data analytics, unsupervised learning, and image classification. The book is intended for a wide variety of people including researchers, scientists, programmers, engineers, designers, developers, educators, and students. This book can also be beneficial for business managers, entrepreneurs, and investors.

Big data has always been a major challenge in geoinformatics as geospatial data come in various types and formats, new geospatial data are acquired very fast, and geospatial databases are inherently very large. And while there have been advances in hardware and software for handling big data, they often fall short of handling geospatial big data ef

Data management and analysis is one of the fastest growing and most challenging areas of research and

development in both academia and industry. Numerous types of applications and services have been studied and re-examined in this field resulting in this edited volume which includes chapters on effective approaches for dealing with the inherent complexity within data management and analysis. This edited volume contains practical case studies, and will appeal to students, researchers and professionals working in data management and analysis in the business, education, healthcare, and bioinformatics areas.

This book provides a comprehensive picture of mobile big data starting from data sources to mobile data driven applications. Mobile Big Data comprises two main components: an overview of mobile big data, and the case studies based on real-world data recently collected by one of the largest mobile network carriers in China. In the first component, four areas of mobile big data life cycle are surveyed: data source and collection, transmission, computing platform and applications. In the second component, two case studies are provided, based on the signaling data collected in the cellular core network in terms of subscriber privacy evaluation and demand forecasting for network management. These cases respectively give a vivid demonstration of what mobile big data looks like, and how it can be analyzed and mined to generate useful and meaningful information and knowledge. This book targets researchers, practitioners and professors relevant to this field. Advanced-level students studying computer science and electrical engineering will also be interested in this book as supplemental reading.

Today, cloud computing, big data, and the internet of things (IoT) are becoming indubitable parts of modern information and communication systems. They cover not only information and communication technology but also all types of systems in society including within the realms of business, finance, industry, manufacturing, and management. Therefore, it is critical to remain up-to-date on the latest advancements and applications, as well as current issues and challenges. The Handbook of Research on Cloud Computing and Big Data Applications in IoT is a pivotal reference source that provides relevant theoretical frameworks and the latest empirical research findings on principles, challenges, and applications of cloud computing, big data, and IoT. While highlighting topics such as fog computing, language interaction, and scheduling algorithms, this publication is ideally designed for software developers, computer engineers, scientists, professionals, academicians, researchers, and students.

Big data is presenting challenges to cybersecurity. For an example, the Internet of Things (IoT) will reportedly soon generate a staggering 400 zettabytes (ZB) of data a year. Self-driving cars are predicted to churn out 4000 GB of data per hour of driving. Big data analytics, as an emerging analytical technology, offers the capability to collect, store, process, and visualize these vast amounts of data. Big Data Analytics in Cybersecurity examines security challenges surrounding big data and provides actionable insights that can be used to improve the current practices of network operators and administrators. Applying big data analytics in

cybersecurity is critical. By exploiting data from the networks and computers, analysts can discover useful network information from data. Decision makers can make more informative decisions by using this analysis, including what actions need to be performed, and improvement recommendations to policies, guidelines, procedures, tools, and other aspects of the network processes. Bringing together experts from academia, government laboratories, and industry, the book provides insight to both new and more experienced security professionals, as well as data analytics professionals who have varying levels of cybersecurity expertise. It covers a wide range of topics in cybersecurity, which include: Network forensics Threat analysis Vulnerability assessment Visualization Cyber training. In addition, emerging security domains such as the IoT, cloud computing, fog computing, mobile computing, and cyber-social networks are examined. The book first focuses on how big data analytics can be used in different aspects of cybersecurity including network forensics, root-cause analysis, and security training. Next it discusses big data challenges and solutions in such emerging cybersecurity domains as fog computing, IoT, and mobile app security. The book concludes by presenting the tools and datasets for future cybersecurity research.

On January 30-31, 2019 the Board on Mathematical Sciences and Analytics, in collaboration with the Board on Energy and Environmental Systems and the Computer Science and Telecommunications Board, convened a workshop in Washington, D.C. to explore the frontiers of mathematics and data science needs for

sustainable urban communities. The workshop strengthened the emerging interdisciplinary network of practitioners, business leaders, government officials, nonprofit stakeholders, academics, and policy makers using data, modeling, and simulation for urban and community sustainability, and addressed common challenges that the community faces. Presentations highlighted urban sustainability research efforts and programs under way, including research into air quality, water management, waste disposal, and social equity and discussed promising urban sustainability research questions that improved use of big data, modeling, and simulation can help address. This publication summarizes the presentation and discussion of the workshop.

This book provides an overview of the resources and research projects that are bringing Big Data and High Performance Computing (HPC) on converging tracks. It demystifies Big Data and HPC for the reader by covering the primary resources, middleware, applications, and tools that enable the usage of HPC platforms for Big Data management and processing.Through interesting use-cases from traditional and non-traditional HPC domains, the book highlights the most critical challenges related to Big Data processing and management, and shows ways to mitigate them using HPC resources. Unlike most books on Big Data, it covers a variety of alternatives to Hadoop, and explains the differences between HPC platforms and Hadoop.Written by professionals and researchers in a range of departments and fields, this book is designed for anyone studying Big

Data and its future directions. Those studying HPC will also find the content valuable.

This book constitutes revised papers from the eight International Workshops held at the 16th International Conference on Business Process Management, BPM 2018, in Sydney, Australia, in September 2018: BPI 2018: 14th International Workshop on Business Process Intelligence; BPMS2 2018: 11th Workshop on Social and Human Aspects of Business Process Management;? PODS4H 2018: 1st International Workshop on Process-Oriented Data Science for Healthcare; AI4BPM 2018: 1st International Workshop on Artificial Intelligence for Business Process Management; CCBPM 2018: 1st International Workshop on Emerging Computing Paradigms and Context in Business Process Management; BP-Meet-IoT / PQ 2018: Joint Business Processes Meet the Internet-of-Things and Process Querying Workshop; DeHMiMoP 2018: 1st Declarative/Decision/Hybrid Mining and Modelling for Business Processes Workshop; REBM /EdForum 2018: Joint Requirements Engineering and Business Process Management Workshop and Education Forum The 45 full papers presented in this volume were carefully reviewed and selected from 90 submissions.

This book constitutes the thoroughly refereed post-workshop proceedings of the AVI 2020 Workshop on Road Mapping Infrastructures for Artificial Intelligence Supporting Advanced Visual Big Data Analysis, AVI-BDA 2020, held in Ischia, Italy, in June 2020, and the Second Italian Workshop on Visualization and Visual Analytics, held in Ischia, Italy, in September 2020. The 14 regular

papers in this volume present topics such as big data collection, management and curation; big data analytics; big data interaction and perception; big data insight and effectuation; configuration and management of big data storage and compute infrastructures, services, and tools; advanced visual interaction in big data applications; user empowerment and meta design in big data applications; prediction and automation of big data analysis workflows; as well as data visualization; information visualization; visual analytics; infographics; and design.

This book constitutes the thoroughly refereed prost-workshop proceedings of the International Workshop on Medical Computer Vision: Algorithms for Big Data, MCS 2015, held in Munich, Germany, in October 2015, held in conjunction with the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015. The workshop shows well the current trends and tendencies in medical computer vision and how the techniques can be used in clinical work and on large data sets. It is organized in the following sections: predicting disease; atlas exploitation and avoidance; machine learning based analyses; advanced methods for image analysis; poster sessions. The 10 full, 5 short, 1 invited papers and one overview paper presented in this volume were carefully reviewed and selected from 22 submissions.

In September 2018, researchers from Armenia, Chile, Germany and Japan met in Yerevan to discuss technologies with applications in Smart Cities, Data Science and Information-Theoretic Approaches for Smart Systems, Technical Challenges for Smart Environments,

and Smart Human Centered Computing. This book presents their contributions to the CODASSCA 2018 workshop on Collaborative Technologies and Data Science in Smart City Applications, a cutting-edge topic in Computer Science today.