DATA SCIENCE AND ANALYTICS

EDITORS

DR. SNEHA KUMARI
DR. K. K. TRIPATHY
DR. VIDYA KUMBHAR

Data Science and Analytics

This page intentionally left blank

Data Science and Analytics

EDITORS

DR SNEHA KUMARI
Assistant Professor, Vaikunth Mehta National Institute of Cooperative Management, Maharashtra, India

DR K. K. TRIPATHY
I.E.S. Director, Vaikunth Mehta National Institute of Cooperative Management, Maharashtra, India

DR VIDYA KUMBHAR
Deputy Director, Symbiosis Institute of Geoinformatics, Symbiosis International (Deemed University), Maharashtra, India

United Kingdom – North America – Japan – India – Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2021

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-80043-877-4 (Print)
ISBN: 978-1-80043-876-7 (Online)
ISBN: 978-1-80043-878-1 (Epub)

ISOQAR certified
Management System,
awarded to Emerald
for adherence to
Environmental
standard
ISO 14001:2004.

Certificate Number 1985
ISO 14001

INVESTOR IN PEOPLE

Acknowledgments

We would like to express our gratitude to each and everyone who directly or indirectly supported us during this project. We would like to thank all the authors for their contribution in the intellectual capital of this book. Our extended thanks to all the authors for their hard work, diligence and dedication in serving the readers. Their contribution is appreciable.

Words fail to express our heartiest thanks and profound sense of gratitude to Dr Rameshwar Dubey for his commendable guidance, motivation and support. Our sincere thanks to Dr Urvashi Rathod and Dr Yogesh Patil for providing the support and motivation to carry out this project work. We would also like to thank the reviewers Dr Manisha Dinesh Bhosle, Dr Nisha Bharti and others for their constant support and help during this work.

We would like to express our gratitude to the editorial advisory board members who helped us with their timely reviews and the team of Emerald who supported us constantly at every phase of this book.

We would like to thank the Almighty for guiding us with mature thoughts and decision-making for the success of this book. The support of family remains unparalleled in any success and the same goes here as well. We would like to thank our parents, family members and friends who helped us stay motivated during the entire course of this book. Everyone has been a great source of inspiration, motivation and support for this work.

This page intentionally left blank

Contents

Preface

According to Kitchin (2014), big data is defined as huge volume of structured and unstructured data. Boyd and Crawford (2012) have defined big data as cultural, technological and scholarly phenomenon while Fan, Han and Liu (2014) have defined big data as the ocean of information.

Five Vs of Big Data

While the term "big data" is relatively new, the act of gathering and storing large amounts of data is characterized by five Vs (Jeble, Kumari, & Patil, 2016):

Volume – Organizations collect data from a variety of sources, including business transactions, social media and information from sensor or machine-to-machine data. In the past, storing it would've been a problem – but new technologies (such as Hadoop) have eased the burden.
Velocity – Data stream in at an unprecedented speed and must be dealt with in a timely manner. Radio Frequency Identification (RFID) tags, sensors and smart metering are driving the need to deal with torrents of data in near real time.
Variety – Data come in all types of formats – from structured, numeric data in traditional databases to unstructured text documents, email, video, audio, stock ticker data and financial transactions.
At Statistical Analysis System (SAS), two additional dimensions are considered when it comes to big data:
Veracity – In addition to the increasing velocities and varieties of data, data flows can be highly inconsistent with periodic peaks. Is something trending in social media? Daily, seasonal and event-triggered peak data loads can be challenging to manage. Even more so with unstructured data.
Value – Today's data come from multiple sources, which make them difficult to link, match, cleanse and transform data across systems. However, it's necessary to connect and correlate relationships, hierarchies and multiple data linkages or data can quickly spiral out of control for social and economic outcomes.

Big data and business analytics are conceived to provide a platform for academicians and practitioners to identify and explore the solutions to various problems in society, environment and industry using advance analytic tools. Business

analytics is the process of converting data into insights (Xavier, Srinivasan, & Thamizhvanan, 2011). It is "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions." With the increase in the availability of data, analytics has now become a major element in both the top line and the bottom line of any organization. However, the rate of absorption of analytics in decision-making is slow. This is due to the fact that there are several ambiguities in the definition and scope of analytics (Jeble, Kumari, Venkatesh, & Singh, 2019). An effective use of analytics must grow with time and experience in most individuals. There is much more in analytics besides descriptive data collection and reporting. By 2025, there will be an increasing need for more data analytics to be involved in business. Effective performance management analytics is an integration of IT-based solutions, management accounting applications and analytical methods. Therefore, descriptive, predictive and prescriptive analytics are essential for any business.

Use of analytics in better decision-making has evolved since past. In the late 1960s, technology-based analytics had been the base of the decision support system (Jeble, Kumari, & Patil, 2018). Later, in 1987, scanner panel data were used to analyze decision-making in retail shops which was further followed by OLAP, a software analytical tool. In 1990, Enterprise Resource Planning system became the prime use for analytics in company. This led to the evolution of analytics using internet, e-commerce, mobiles, sensors and software analytical tools. With time big data predictive analytics has been used in decision-making in different streams. Analytics in big data has been useful in improving the visibility and coordination (Dubey et al., 2018). Businesses can have better decision-making capability with the use and better understanding of data analytics (Agrawal, 2014).

Why is Big Data Important?

The importance of big data doesn't revolve around the availability of data but the purpose of data. One can take data from any source and analyze them to find answers that enable (1) cost reductions, (2) time reductions, (3) new product development and optimized offerings and (4) smart decision-making. When big data are combined with high-powered analytics, managerial decisions can be performed such as

- Determining root causes of failures, issues and defects in near real time.
- Generating sustainable solutions for any stream.
- Recalculating entire risk portfolios in minutes.
- Detecting fraudulent behavior before it affects any organization.

Application of Big Data and Business Analytics

Big data have been in use by government institutes for forecasting weather patterns, discovering seismic activities that predict earthquakes and preparing descriptive reports. It has been in use by the economists to stimulate economic growth. Big data mining is the patterns in the data that are normally not looked

by the users. These unlooked data also lead to several important information which can make decision-making smooth.

Big data are used in health insurance for predicting customer dissatisfaction through speech to text data from call center recordings (Devenport & Dyche, 2013). Several retail banks have focused on exploiting big data at times of financial crisis for doing a better job.

LinkedIn has used big data and data scientists to develop product features and product offerings. This has helped the consumers as well as the companies to make decisions about the product.

Google has constantly developed new products and services that have big data algorithms for search. Most of the companies are master's in developing standard reports and multidimensional reports through big data analytics. In many companies, big data are directly focused on products, services and customers. Senior managers have used predictive analytics as the next step in data analytics.

Big data not only allow knowledge discovery efforts but also need to promote them. The sooner the business executives understand the value of knowledge discovery the better competitor they become. This can lead to high-level innovations and high rewards.

Big Data and Business Analytics for Decision-Making

In the current era, world is challenged with demanding customers, high competition, short product life cycles, rising costs of labor and materials, unemployment and unsustainability. Globalization is making it even more challenging as blurring boundaries among countries create level playing field for selling products and services across the globe. Firms need to make operational, tactical and strategic decisions based on available information. In addition to traditional decision support systems, big data provide additional tools to arrive at decisions.

Big data can provide valuable competitive intelligence (Jeong et al., 2016), help in dramatic cost reductions, substantial improvements and development of sustainable goals for the world.

Objectives

Data science and business analytics will bring together researchers, engineers and practitioners and encompass wide and diverse topics of application in almost every field. It will also invite the participation of scholars, analysts and data scientists to present their ideas, concepts and proof of works indicating application of big data and business analytics.

Target Audience

The primary target audience of this book includes researchers, academicians and data scientist from a variety of disciplines interested in analyzing and application of big data analytics. A secondary target audience consists of data analysts, students and scholars pursuing advanced study in big data.

Organization of this Book

This book is organized into eight chapters. A brief description of each of the chapters follows:

Chapter 1 authored by Aarti Mehta Sharma identifies that analytics is the science of examining raw data with the purpose of drawing conclusions about that information and using it for decision-making. It also looks at the pairing of visualization tools with different measurements of data. Before the formal written language, there were pictures which shared ideas, plans and history. Visualizations in the form of bar charts, scatter plots or dashboards are essential tools in business intelligence as they help managers to absorb information and take apt decisions quickly. Dashboards in particular are very helpful for managers as multiple charts and graphs giving the latest information about sales, returns, market share, etc. keep them up to date on the latest developments in the company. There are a number of visualization software in the market which are easy to learn and communicate the analyzed data in an easily understood form; the leading ones being Tableau, QlikView, etc. with each one having its positives.

Chapter 2 authored by Hiral R. Patel, Ajay M. Patel and Satyen M. Parikh takes the fundamentals for multimedia big data computing, feasibility study and salient features of multimedia big data and explore the technical problems and challenges to be addressed. It focuses on methodologies and approaches that are available from the perspectives of multimedia big data computing life cycle. Multimedia data are real-time unstructured, heterogeneous and multimodal as per the qualitative requirement. It has vast scope to mine model, learn and analyze the service provided by multimedia. An advance-level storage-related mechanism is also needed for efficient parallel processing, transmission and presentation. The multimedia data in form of videos are easily understood by human compared to textual data, but it's more complex task to make it understandable to machines.

Chapter 3 authored by Gauri Rajendra Virkar and Supriya Sunil Shinde reviews that predictive analytics is the science of decision-making that eliminates guesswork out of the decision-making process and applies proven scientific procedures to find right solutions. Predictive analytics provides ideas on the occurrences of future downtimes and rejections, thereby aids in taking preventive actions before abnormalities occur. Considering these advantages, predictive analytics is adopted in various diverse fields such as health care, finance, education, marketing, automotive, etc. Predictive analytics tools can be used to predict various behaviors and patterns, thereby saving the time and money of its users. Many open-source predictive analysis tools namely R, scikit-learn, KNIME, Orange, RapidMiner, WEKA, etc. are freely available for the users. This chapter aims to reveal the best accurate tools and techniques for the classification task that aid in decision-making.

Chapter 4 authored by Shakti Ranjan Panigrahy reviews the information of 300 customers in Anand districts of Gujarat, India. Cluster analysis techniques were presented to analyze the data. At the end, it was found high quality, better service, convenient location, presentation of food in parlor and restaurants and zero time delivery are playing key roles for getting customers for the food.

Chapter 5 authored by K.S.S. Iyer and Madhavi Damle presents an analysis of the method in advance predictive analytics. Varied or different domains are put together to establish a fine-tuned technique of "Random Point Process" (RPP) and "Product Density" (PD), two techniques in stochastic modeling, which can be used for advanced predictive analytics. This formulation arises from these techniques being used in different fields like energy requirement in IoT devices, growth of cancer cells, cosmic rays' study, to customer equity and many more approaches.

Chapter 6 authored by Samir Yerpude addresses a paradigm shift observed in the last decade where transactional marketing is taken over by relationship marketing. Customer relationship management (CRM) has been an integral part of a business strategy in the current era. CRM integrates product sales, product marketing and, most importantly, customer service in a seamless manner to generate value for the organization as well as for its customers in short a win–win situation. Profoundly, CRM needs to be a part of the top management agenda and driven top-down instead of an IT initiative. Industrial revolution 4.0 is characterized by cyber-physical systems. Internet of Things (IoT) is the digital technology for the present and future. IoT primarily aids in gathering real-time data and transmitting the same over the internet to a central repository for consuming the same in business models. Real-time customer data analytics can be performed by customer-centric organizations to enhance CRM.

Chapter 7 authored by K. K. Tripathy and Sneha Kumari addresses that Rural Development constitutes of lot of big data related to rural employment with special reference to the world's largest public works and wage employment generating poverty alleviation programme – Mahatma Gandhi National Rural Employment Guarantee Act (MNREGA)The concepts of MGNREGA are novel and innovative though the programme continues to suffer from various rigidities depicted from the data. This drives us to the objectives of our research. The study has explored literature and big data on rural development with special reference to MNREGA, the upcoming challenges in rural employment with special reference to MGNREGA, identify gaps in existing literature and pave out future research direction for academicians, researchers and policy maker.

Chapter 8 authored by Jorge Tarifa-Fernandez, Almudena Martínez Aguilera and José Felipe Jiménez-Guerrero concludes and presents the critical importance and value of digital technologies which have led companies to remain competitive. Despite the benefits, the transformation of a firm into a digital-based one requires it to choose a set of digital technologies. In this sense, there may be multiple combinations of digital technologies by means of which firms can obtain their digitalization. Thus, each firm should decide which combination best fits its needs.

Dr Sneha Kumari
Assistant Professor, Vaikunth Mehta National Institute of Cooperative Management, Pune, India

Dr K. K. Tripathy
I.E.S., Director, Vaikunth Mehta National Institute of Cooperative Management, Pune, India

Dr Vidya Kumbhar
Assistant Professor, Symbiosis Institute of Geoinformatics, Symbiosis International (Deemed University), Pune, India

Foreword

The world has generated data in every field giving rise to big data. There is plenty of data everywhere and data storage is becoming critical at present. The importance of big data doesn't revolve around the availability of data, but the purpose of data. Researchers, academicians, policy makers and practitioners are consistently driving ways to find out the application of big data. With so much of data, it is time to understand the big data and how analytics can help in better decision making and manage things. Application of Big Data and Business Analytics at present needs to be explored among academicians, practitioners, policy makers and researchers. The book will have academic and managerial implications to manage the decision-making process. This book explores a number of perspectives on how big data and business analytics can help in better decision making. The book can be an asset for the readers at the present time. The authors from different countries and universities have made a contribution in organizing their research ideas into research chapter meeting the scope of the book. The chapters in the book have been selected carefully, providing a fine balance between trends in big data analytics and it's application in different streams. The chapters have diverse themes in application of big data analytics.

The chapters selected have been classified into Themes like data visualization, multiple aspects of data analytics, predictive analytics, application of data analytics in industry, agriculture and service sector followed by the challenges in digital technologies. The chapter on Customer Segmentation Using RFM Analysis: Real Case Application on a Fuel Company establishes and presents a real case application of RFM analysis on customer segmentation for a Fuel company with possible strategies for the company are generated. The chapter on Applications of Big Data Analytics: A Boon for the Food Industry analyses big data with technologies like machine learning and artificial intelligence to get faster and more personalized experiences generating an opportunity for the food industries to reduce food loss and gain better returns on investment by going for a digital transformation. The chapter on Big Data for Sustainable Rural Development with special reference to MGNREGA addresses the application of big data related to rural employment with special reference to the world's largest public works and wage employment generating poverty alleviation programme – Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) on Rural Development. The authors have done an appreciable work in presenting the applications of big data and business analytics in different managerial decision making. The application of big data can be seen in descriptive, predictive and prescriptive analytics.

All the chapters are topical. The chapters are well-balanced covering the application of big data and business analytics by academicians, researchers, industrial experts, policy makers and practitioners. This book will bring together researchers, engineers and practitioners and encompass wide and diverse topics of application in almost every field. It will also invite the participation of scholars, analysts and data scientists to analyze the application of Big Data and Business Analytics by the contributors from different countries. The book paves a way for the readers to understand how big data can be efficiently utilized in better managerial applications. Dr. Sneha Kumari, Dr. K. K. Tripathy; Vaikunth Mehta National Institute of Cooperative Management and Dr. Vidya Kumbhar, Symbiosis International (Deemed University) have done a commendable job as book editors in making the application of big data analytics research available for a wide audience.

Dr. Lt. Col. Anupama Munshi
Veteran, Faculty and Researcher, India

Dr. Lt. Col. Anupama Munshi (retd) is a doctorate in Management with 17 years of experience in Indian Army in handling big data of entire gamut of Human Resource Development functions , Human Resource Management and Industrial Relations. She has dealt with the application of big data in decision making in the Army. She is an expert in application of big data in designing & implementing training programs to enhance efficiency & motivation levels. She has also applied the big data for imparting teaching and training to officers of Indian Army as well as officers of other armies in subjects like Quality control, Logistics & Supply chain management, Transport management, Tendering and procurement for Defense supplies. She is the First Lady Officer of Indian Army to command an Independent Food Inspection Unit for providing logistics support to a specialized brigade and was first officer of Army Service Corps to be awarded General Officer Commanding in Chief's Commendation card for outstanding service. She has also worked as a Professor in Symbiosis International Deemed University, Pune and is the visiting faculty for many renowned Management Institutes in Pune. She is consistently involved in researches of application of big data in the Indian Army.

.

Editorial Advisory Board

List of Contributors

Madhavi Damle	Symbiosis International (Deemed University), India
K. S. S. Iyer	Retired Sr. Scientist, Defence Research and Development Organisation (DRDO), India
José Felipe Jiménez-Guerrero	University of Almería, Spain
Sneha Kumari	Vaikunth Mehta National Institute of Cooperative Management, India
Almudena Martínez Aguilera	University of Almería, Spain
Shakti Ranjan Panigrahy	Anand Agricultural University, India
Satyen M. Parikh	Ganpat University, India
Ajay M. Patel	Ganpat University, India
Hiral R. Patel	Ganpat University, India
Aarti Mehta Sharma	Symbiosis International (Deemed University), India
Supriya Sunil Shinde	JSPM's Bhivarabai Sawant Institute of Technology and Research, India
Jorge Tarifa-Fernandez	University of Almería, Spain
K. K. Tripathy	Vaikunth Mehta National Institute of Cooperative Management, India
Gauri Rajendra Virkar	JSPM's Bhivarabai Sawant Institute of Technology and Research, India
Samir Yerpude	Symbiosis International (Deemed University), India

This page intentionally left blank

Editor Biographies

Dr Sneha Kumari, PhD, is a full-time Junior Research Fellow in the Symbiosis International (Deemed University). She has completed her undergraduate in Agriculture at the Indian Council of Agriculture Research Fellowship and her master's in Agribusiness Management under Indian Council of Agriculture Research. She has also completed her PGDBA in Human Resource Management and Marketing Management. She is an Assistant Professor at Vaikunth Mehta National Institute of Cooperative Management – a National Institute of Ministry of Agriculture and Farmers Welfare, Government of India. She has a rich experience as statistical officer, researcher and assistant professor. She is also associated with different institutes for various educational- and research-related project assignments. She has published several research papers in the area of big data, agriculture, sustainability in ABDC and Scopus journals and has attended several national and international conferences.

Dr K. K. Tripathy has a PhD from Department of Management Studies of Indian Institute of Technology Delhi on "Micro-finance Management and Its Impact on Rural Livelihoods"; master's in Economics and Bachelor of Law (LLB). He is an Officer of Indian Economic Service (IES), Government of India. He joined the IES in 1999. Prior to joining the IES, he also worked as an Executive Magistrate in the Government of Odisha after joining the Odisha Administrative Service in 1998. He is presently the Director of the Vaikunth Mehta National Institute of Cooperative Management – a National Institute of Ministry of Agriculture and Farmers Welfare, Government of India. During his 21 years of public service, he has served in the Ministry of Human Resource Development, Ministry of Agriculture and Farmers Welfare, Ministry of Rural Development, Ministry of Food Processing Industries, Planning Commission and United Nations Development Programme. He has published around 70 research articles and has attended several international and national conferences. His areas of research are rural development, microfinance and agriculture.

Dr Vidya Kumbhar has a PhD in Computer Studies from Symbiosis International (Deemed University). Her area of specialization includes data science and analytics, software development, big data handling, spatial data mining, spatial decision support system and agro-informatics. She is also associated with different

institutes for various educational-related assignments. She has worked on software development projects in National Informatics Centre, Pune, and currently working on different research projects funded by government and private organizations. She has published various research papers in national and international journals and conferences. She is also a reviewer for national and international journals indexed in Scopus and Web of Science.

References

Agrawal, D. (2014). Analytics based decision making. *Journal of Indian Business Research*, *6*(4), 332–340.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679.

Devenport, T. H., & Dyché, J. (2013). Big data in companies. *International Institute for Analytics*. Retrieved from https://www. sas. com/content/dam/SAS/en_us/doc/whitepaper2/bigdatabigcompanies-106461. Accessed on March 12, 2020.

Dubey, R., Luo, Z., Gunasekaran, A., Akter, S., Hazen, B. T., & Douglas, M. A. (2018). Big data and predictive analytics in humanitarian supply chains: Enabling visibility and coordination in the presence of swift trust. *International Journal of Logistics Management*. *29*(2), 485–512.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review*, *1*(2), 293–314.

Jeble, S., Kumari, S., & Patil, Y. (2016). Role of big data and predictive analytics. *International Journal of Automation and Logistics*, *2*(4), 307–331.

Jeble, S., Kumari, S., & Patil, Y. (2018). Role of big data in decision making. *Operations and Supply Chain Management: An International Journal*, *11*(1), 36–44.

Jeble, S., Kumari, S., Venkatesh, V. G., & Singh, M. (2019). Influence of big data and predictive analytics and social capital on performance of humanitarian supply chain. *Benchmarking: An International Journal*. *27*(2), 606–633.

Jeong, Y. S., Kim, B., Choi, H. J., & Lee, J. S. (2016, January). Issues about combination of temporal information and spatial information. 2016 International conference on big data and smart computing (BigComp) (pp. 462–465). Retrieved from https://ieeexplore.ieee.org/abstract/document/7425971. Accessed on June 12, 2020.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1). doi:10.1177/2053951714528481

Xavier, M. J., Srinivasan, A., & Thamizhvanan, A. (2011). Use of analytics in Indian enterprises: An exploratory study. *Journal of Indian Business Research*, *3*(3), 168–179