Big Data, Mining, and Analytics

Components of Strategic Decision Making

Stephan Kudyba

Foreword by Thomas H. Davenport

CRC Press
Taylor & Francis Group
AN AUERBACH BOOK

Big Data, Mining, and Analytics

Components of Strategic Decision Making

Stephan Kudyba

Foreword by Thomas H. Davenport

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **Informa** business
AN AUERBACH BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper
Version Date: 20140203

International Standard Book Number-13: 978-1-4665-6870-9 (Hardback)

Library of Congress Cataloging-in-Publication Data

Big data, mining, and analytics : components of strategic decision making / editor, Stephan Kudyba.
pages cm
Includes bibliographical references and index.
ISBN 978-1-4665-6870-9 (hardback)
1. Strategic planning--Data processing. 2. Data mining. 3. Big data. 4. Business planning--Data processing. 5. Webometrics. 6. Data loggers. I. Kudyba, Stephan, 1963- editor of compilation.

HD30.28.B544 2014
658.4'012--dc23 2013049469

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Big Data, Mining, and Analytics

Components of Strategic Decision Making

To my family, for their consistent support to pursue and complete these types of projects. And to two new and very special family members, Lauren and Kirsten, who through their evolving curiosity have reminded me that you never stop learning, no matter what age you are. Perhaps they will grow up to become analysts . . . perhaps not. Wherever their passion takes them, they will be supported.

To the contributors to this work, sincere gratitude for taking the time to share their expertise to enlighten the marketplace of an evolving era, and to Tom Davenport for his constant leadership in promoting the importance of analytics as a critical strategy for success.

Foreword

Big data and analytics promise to change virtually every industry and business function over the next decade. Any organization that gets started early with big data can gain a significant competitive edge. Just as early analytical competitors in the "small data" era (including Capital One bank, Progressive Insurance, and Marriott hotels) moved out ahead of their competitors and built a sizable competitive edge, the time is now for firms to seize the big data opportunity.

As this book describes, the potential of big data is enabled by ubiquitous computing and data gathering devices; sensors and microprocessors will soon be everywhere. Virtually every mechanical or electronic device can leave a trail that describes its performance, location, or state. These devices, and the people who use them, communicate through the Internet—which leads to another vast data source. When all these bits are combined with those from other media—wireless and wired telephony, cable, satellite, and so forth—the future of data appears even bigger.

The availability of all this data means that virtually every business or organizational activity can be viewed as a big data problem or initiative. Manufacturing, in which most machines already have one or more microprocessors, is increasingly becoming a big data environment. Consumer marketing, with myriad customer touchpoints and clickstreams, is already a big data problem. Google has even described its self-driving car as a big data project. Big data is undeniably a big deal, but it needs to be put in context.

Although it may seem that the big data topic sprang full blown from the heads of IT and management gurus a couple of years ago, the concept actually has a long history. As Stephan Kudyba explains clearly in this book, it is the result of multiple efforts throughout several decades to make sense of data, be it big or small, structured or unstructured, fast moving or quite still. Kudyba and his collaborators in this volume have the knowledge and experience to put big data in the broader context of business and organizational intelligence.

If you are thinking, "I only want the new stuff on big data," that would be a mistake. My own research suggests that within both large non-online businesses (including GE, UPS, Wells Fargo Bank, and many other leading firms) and online firms such as Google, LinkedIn, and Amazon, big

data is not being treated separately from the more traditional forms of analytics. Instead, it is being combined with traditional approaches into a hybrid capability within organizations.

There is, of course, considerable information in the book about big data alone. Kudyba and his fellow experts have included content here about the most exciting and current technologies for big data—and Hadoop is only the beginning of them. If it's your goal to learn about all the technologies you will need to establish a platform for processing big data in your organization, you've come to the right place.

These technologies—and the subject of big data in general—are exciting and new, and there is no shortage of hype about them. I may have contributed to the hype with a coauthored article in the *Harvard Business Review* called "Data Scientist: The Sexiest Job of the 21st Century" (although I credit the title to my editors). However, not all aspects of big data are sexy. I remember thinking when I interviewed data scientists that it was not a job I would want; there is just too much wrestling with recalcitrant data for my skills and tastes.

Kudyba and his collaborators have done a good job of balancing the sexy (Chapter 1, for example) and the realistic (Chapter 5, for example). The latter chapter reminds us that—as with traditional analytics—we may have to spend more time cleaning, integrating, and otherwise preparing data for analysis than we do actually analyzing it. A major part of the appeal of big data is in combining diverse data types and formats. With the new tools we can do more of this combining than ever before, but it's still not easy.

Many of the applications discussed in this book deal with marketing—using Internet data for marketing, enhancing e-commerce marketing with analytics, and analyzing text for information about customer sentiments. I believe that marketing, more than any other business function, will be reshaped dramatically by big data and analytics. Already there is very strong demand for people who understand both the creative side of marketing and the digital, analytical side—an uncommon combination. Reading and learning from Chapters 6, 7, 10, and others will help to prepare anyone for the big data marketing jobs of the future.

Other functional domains are not slighted, however. For example, there are brief discussions in the book of the massive amounts of sensor data that will drive advances in supply chains, transportation routings, and the monitoring and servicing of industrial equipment. In Chapter 8, the role of streaming data is discussed in such diverse contexts as healthcare equipment and radio astronomy.

The discussions and examples in the book are spread across different industries, such as Chapter 12 on evolving data sources in healthcare. We can now begin to combine structured information about patients and treatments in electronic medical record systems with big data from medical equipment and sensors. This unprecedented amount of information about patients and treatments should eventually pay off in better care at lower cost, which is desperately needed in the United States and elsewhere. However, as with other industry and functional transformations, it will take considerable work and progress with big data before such benefits can be achieved.

In fact, the combination of hope and challenge is the core message of this book. Chapters 10 and 11, which focus on the mining and automated interpretation of textual data, provide an exemplary illustration of both the benefits from this particular form of big data analytics and the hard work involved in making it happen. There are many examples in these two chapters of the potential value in mining unstructured text: customer sentiment from open-ended surveys and social media, customer service requests, news content analysis, text search, and even patent analysis. There is little doubt that successfully analyzing text could make our lives and our businesses easier and more successful.

However, this field, like others in big data, is nothing if not challenging. Meta Brown, a consultant with considerable expertise in text mining, notes in Chapter 10, "Deriving meaning from language is no simple task," and then provides a description of the challenges. It is easy to suggest that a firm should analyze all the text in its customers' blogs and tweets, or that it should mine its competitors' patents. But there are many difficulties involved in disambiguating text and dealing with quintessentially human expressions like sarcasm and slang. As Brown notes, even the best automated text analysis will be only somewhat correct.

As we move into the age of big data, we'll be wrestling with these implementation challenges for many years. The book you're about to read is an excellent review of the opportunities involved in this revolution, but also a sobering reminder that no revolution happens without considerable effort, money, and false starts. The road to the Big Data Emerald City is paved with many potholes. Reading this book can help you avoid many of them, and avoid surprise when your trip is still a bit bumpy.

Thomas H. Davenport
Distinguished Professor, Babson College
Fellow, MIT Center for Digital Business
Co-Founder, International Institute for Analytics

About the Author

Stephan Kudyba, MBA, PhD, is a faculty member in the school of management at New Jersey Institute of Technology (NJIT), where he teaches courses in the graduate and executive MBA curriculum addressing the utilization of information technologies, business intelligence, and information and knowledge management to enhance organizational efficiency and innovation. He has published numerous books, journal articles, and magazine articles on strategic utilization of data, information, and technologies to enhance organizational and macro productivity. Dr. Kudyba has been interviewed by prominent magazines and speaks at university symposiums, academic conferences, and corporate events. He has over 20 years of private sector experience in the United States and Europe, having held management and executive positions at prominent companies. He maintains consulting relations with organizations across industry sectors with his company Null Sigma Inc. Dr. Kudyba earned an MBA from Lehigh University and a PhD in economics with a focus on the information economy from Rensselaer Polytechnic Institute.

Contributors

Billie Anderson
Bryant University
Smithfield, Rhode Island

Steven Barber
TIBCO StreamBase, Inc.
New York, New York

Jerry Baulier
SAS Institute
Cary, North Carolina

Meta S. Brown
Business consultant
Chicago, Illinois

Thomas H. Davenport
Babson College
Wellesley, Massachusetts

J. Michael Hardin
University of Alabama
Tuscaloosa, Alabama

Ioannis Korkontzelos
University of Manchester
Manchester, United Kingdom

Stephan Kudyba
New Jersey Institute of Technology
Newark, New Jersey

Matthew Kwatinetz
QBL Partners
New York, New York

David Lubliner
New Jersey Institute of Technology
Newark, New Jersey

Viju Raghupathi
Brooklyn College
City University of New York
New York, New York

Wullianallur Raghupathi
Fordham University
New York, New York

Wayne Thompson
SAS Institute
Cary, North Carolina

Robert Young
PHD, Inc.
Toronto, Ontario, Canada

Contents