

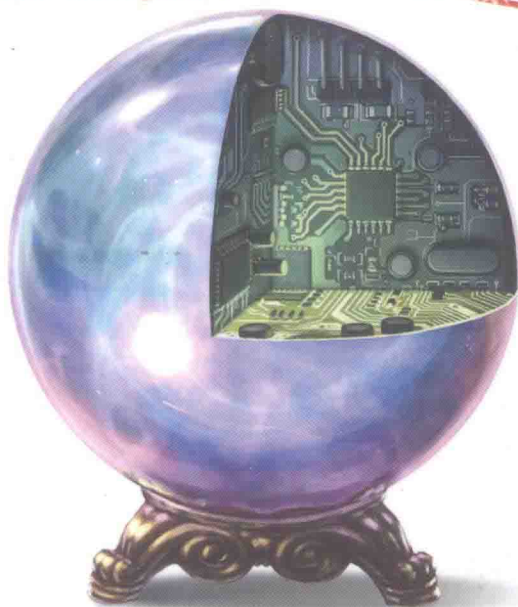
REVISED AND UPDATED

# PREDICTIVE ANALYTICS

"Mesmerizing & fascinating..."

—*The Seattle Post-Intelligencer*

AN INTRODUCTION  
FOR EVERYONE

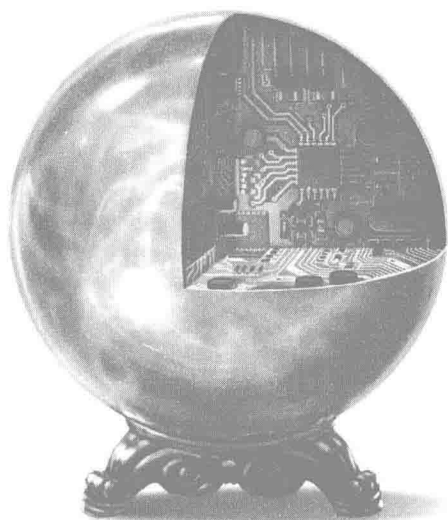


THE POWER TO PREDICT WHO WILL  
CLICK, BUY, LIE, OR DIE

ERIC SIEGEL

WILEY

# PREDICTIVE ANALYTICS



THE POWER TO PREDICT WHO WILL  
CLICK, BUY, LIE, OR DIE

ERIC SIEGEL

WILEY

Cover image: Winona Nelson  
Cover design: Wiley  
Interior image design: Matt Kornhaas

Copyright © 2016 by Eric Siegel. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.  
Published simultaneously in Canada.

*Jeopardy!*® is a registered trademark of Jeopardy Productions, Inc.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600, or on the Web at [www.copyright.com](http://www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at [www.wiley.com/go/permissions](http://www.wiley.com/go/permissions).

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with the respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor the author shall be liable for damages arising herefrom.

For general information about our other products and services, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit [www.wiley.com](http://www.wiley.com).

### ***Library of Congress Cataloging-in-Publication Data:***

Names: Siegel, Eric, 1968-

Title: Predictive analytics : the power to predict who will click, buy, lie,  
or die / Eric Siegel.

Description: Revised and Updated Edition. | Hoboken : Wiley, 2016. | Revised  
edition of the author's Predictive analytics, 2013. | Includes index.

Identifiers: LCCN 2015031895 (print) | LCCN 2015039877 (ebook) |  
ISBN 9781119145677 (paperback) | ISBN 9781119145684 (pdf) |  
ISBN 9781119153658 (epub)

Subjects: LCSH: Social sciences—Forecasting. | Economic forecasting |  
Prediction (Psychology) | Social prediction. | Human behavior. | BISAC:  
BUSINESS & ECONOMICS / Consumer Behavior. | BUSINESS & ECONOMICS /  
Econometrics. | BUSINESS & ECONOMICS / Marketing / General.

Classification: LCC H61.4 .S54 2016 (print) | LCC H61.4 (ebook) | DDC  
303.49—dc23

LC record available at <http://lcn.loc.gov/2015031895>

Printed in the United States of America

10 9 8 7 6 5 4 3 2

## Praise for *Predictive Analytics*

“Littered with lively examples . . .”

—*The Financial Times*

“Readers will find this a mesmerizing and fascinating study. I know I did! . . . I was entranced by the book.”

—*The Seattle Post-Intelligencer*

“Siegel is a capable and passionate spokesman with a compelling vision.”

—*Analytics Magazine*

“A must-read for the normal layperson.”

—*Journal of Marketing Analytics*

“This book is an operating manual for twenty-first-century life. Drawing predictions from big data is at the heart of nearly everything, whether it’s in science, business, finance, sports, or politics. And Eric Siegel is the ideal guide.”

—**Stephen Baker, author, *The Numerati and Final Jeopardy: The Story of Watson, the Computer That Will Transform Our World***

“Simultaneously entertaining, informative, and nuanced. Siegel goes behind the hype and makes the science exciting.”

—**Rayid Ghani, Chief Data Scientist,  
Obama for America 2012 Campaign**

“The most readable (for we laymen) ‘big data’ book I’ve come across. By far. Great vignettes/stories.”

—**Tom Peters, coauthor, *In Search of Excellence***

“The future is right now—you’re living in it. Read this book to gain understanding of where we are and where we’re headed.”

—**Roger Craig, record-breaking analytical *Jeopardy!*  
champion; Data Scientist, Digital Reasoning**

“A clear and compelling explanation of the power of predictive analytics and how it can transform companies and even industries.”

—**Anthony Goldbloom, founder and CEO, Kaggle.com**

“The definitive book of this industry has arrived. Dr. Siegel has achieved what few have even attempted: an accessible, captivating tome on predictive analytics that is a must-read for all interested in its potential—and peril.”

—**Mark Berry, VP, People Insights, ConAgra Foods**

“I’ve always been a passionate data geek, but I never thought it might be possible to convey the excitement of data mining to a lay audience. That is what Eric Siegel does in this book. The stories range from inspiring to downright scary—read them and find out what we’ve been up to while you weren’t paying attention.”

—**Michael J. A. Berry, author of *Data Mining Techniques, Third Edition***

“Eric Siegel is the Kevin Bacon of the predictive analytics world, organizing conferences where insiders trade knowledge and share recipes. Now, he has thrown the doors open for you. Step in and explore how data scientists are rewriting the rules of business.”

—**Kaiser Fung, VP, Vimeo; author of *Numbers Rule Your World***

“Written in a lively language, full of great quotes, real-world examples, and case studies, it is a pleasure to read. The more technical audience will enjoy chapters on The Ensemble Effect and uplift modeling—both very hot trends. I highly recommend this book!”

—**Gregory Piatetsky-Shapiro, Editor, KDnuggets;  
founder, KDD Conferences**

“Exciting and engaging—reads like a thriller! Predictive analytics has its roots in people’s daily activities and, if successful, affects people’s actions. By way of examples, Siegel describes both the opportunities and the threats predictive analytics brings to the real world.”

—**Marianna Dizik, Statistician, Google**

“A fascinating page-turner about the most important new form of information technology.”

—**Emiliano Pasqualetti, CEO, DomainsBot Inc.**

“Succeeds where others have failed—by demystifying big data and providing real-world examples of how organizations are leveraging the power of predictive analytics to drive measurable change.”

—**Jon Francis, Senior Data Scientist, Nike**

“In a fascinating series of examples, Siegel shows how companies have made money predicting what customers will do. Once you start reading, you will not be able to put it down.”

—**Arthur Middleton Hughes, VP, Database Marketing Institute;  
author of *Strategic Database Marketing, Fourth Edition***

“Excellent. Each chapter makes the complex comprehensible, making heavy use of graphics to give depth and clarity. It gets you thinking about what else might be done with predictive analytics.”

—**Edward Nazarko, Client Technical Advisor, IBM**

“What is predictive analytics? This book gives a practical and up-to-date answer, adding new dimension to the topic and serving as an excellent reference.”

—**Ramendra K. Sahoo, Senior VP,  
Risk Management and Analytics, Citibank**

“Competing on information is no longer a luxury—it’s a matter of survival. Despite its successes, predictive analytics has penetrated only so far, relative to its potential. As a result, lessons and case studies such as those provided in Siegel’s book are in great demand.”

—**Boris Evelson, VP and Principal Analyst, Forrester Research**

“Fascinating and beautifully conveyed. Siegel is a leading thought leader in the space—a must-have for your bookshelf!”

—**Sameer Chopra, Chief Analytics Officer, Orbitz Worldwide**

“A brilliant overview—strongly recommended to everyone curious about the analytics field and its impact on our modern lives.”

—**Kerem Tomak, VP of Marketing Analytics, Macys.com**

“Eric explains the science behind predictive analytics, covering both the advantages and the limitations of prediction. A must-read for everyone!”

—**Azhar Iqbal, VP and Econometrician,  
Wells Fargo Securities, LLC**

“*Predictive Analytics* delivers a ton of great examples across business sectors of how companies extract actionable, impactful insights from data. Both the novice and the expert will find interest and learn something new.”

—**Chris Pouliot, Director, Algorithms and Analytics, Netflix**

“In this new world of big data, machine learning, and data scientists, Eric Siegel brings deep understanding to deep analytics.”

—**Marc Parrish, VP, Membership, Barnes & Noble**

“A detailed outline for how we might tame the world’s unpredictability. Eric advocates quite clearly how some choices are predictably more profitable than others—and I agree!”

—**Dennis R. Mortensen, CEO of Visual Revenue,  
former Director of Data Insights at Yahoo!**

“This book is an invaluable contribution to predictive analytics. Eric’s explanation of how to anticipate future events is thought provoking and a great read for everyone.”

—**Jean Paul Isson, Global VP Business Intelligence and Predictive Analytics, Monster Worldwide; coauthor, *Win with Advanced Business Analytics: Creating Business Value from Your Data***

“Predictive analytics is the key to unlocking new value at a previously unimaginable economic scale. In this book, Siegel explains how, doing an excellent job to bridge theory and practice.”

—**Sergo Grigalashvili, VP of Information Technology,  
Crawford & Company**

“Predictive analytics has been steeped in fear of the unknown. Eric Siegel distinctively clarifies, removing the mystery and exposing its many benefits.”

—**Jane Kuberski, Engineering and Analytics,  
Nationwide Insurance**

“As predictive analytics moves from fashionable to mainstream, Siegel removes the complexity and shows its power.”

—**Rajeev Kaul, Senior VP, OfficeMax**

“Dr. Siegel humanizes predictive analytics. He blends analytical rigor with real-life examples with an ease that is remarkable in his field. The book is informative, fun, and easy to understand. I finished reading it in one sitting. A must-read . . . not just for data scientists!”

—**Madhu Iyer, Marketing Statistician, Intuit**

“An engaging encyclopedia filled with real-world applications that should motivate anyone still sitting on the sidelines to jump into predictive analytics with both feet.”

—**Jared Waxman, Web Marketer at LegalZoom,  
previously at Adobe, Amazon, and Intuit**

“Siegel covers predictive analytics from start to finish, bringing it to life and leaving you wanting more.”

—**Brian Seeley, Manager, Risk Analytics, Paychex, Inc.**

“A wonderful look into the world of predictive analytics from the perspective of a true practitioner.”

—**Shawn Hushman, VP, Analytic Insights,  
Kelley Blue Book**

“A must—*Predictive Analytics* provides an amazing view of the analytical models that predict and influence our lives on a daily basis. Siegel makes it a breeze to understand, for all readers.”

—**Zhou Yu, Online-to-Store Analyst, Google**



“As our ability to collect and analyze information improves, experts like Eric Siegel are our guides to the mysteries unlocked and the moral questions that arise.”

**—Jules Polonetsky, Co-Chair and Director, Future of Privacy Forum; former Chief Privacy Officer, AOL and DoubleClick**

“Highly recommended. As Siegel shows in his very readable new book, the results achieved by those adopting predictive analytics to improve decision making are game changing.”

**—James Taylor, CEO, Decision Management Solutions**

“An engaging, humorous introduction to the world of the data scientist. Dr. Siegel demonstrates with many real-life examples how predictive analytics makes big data valuable.”

**—David McMichael, VP, Advanced Business Analytics**

“An excellent exposition on the next generation of business intelligence—it’s really mankind’s latest quest for artificial intelligence.”

**—Christopher Hornick, President and CEO,  
HBSC Strategic Services**

# PREDICTIVE ANALYTICS

*This book is dedicated with all my heart to my mother,  
Lisa Schamberg, and my father, Andrew Siegel.*

# Foreword

This book deals with quantitative efforts to predict human behavior. One of the earliest efforts to do that was in World War II. Norbert Wiener, the father of “cybernetics,” began trying to predict the behavior of German airplane pilots in 1940—with the goal of shooting them from the sky. His method was to take as input the trajectory of the plane from its observed motion, consider the pilot’s most likely evasive maneuvers, and predict where the plane would be in the near future so that a fired shell could hit it. Unfortunately, Wiener could predict only one second ahead of a plane’s motion, but 20 seconds of future trajectory were necessary to shoot down a plane.

In Eric Siegel’s book, however, you will learn about a large number of prediction efforts that are much more successful. Computers have gotten a lot faster since Wiener’s day, and we have a lot more data. As a result, banks, retailers, political campaigns, doctors and hospitals, and many more organizations have been quite successful of late at predicting the behavior of particular humans. Their efforts have been helpful at winning customers, elections, and battles with disease.

My view—and Siegel’s, I would guess—is that this predictive activity has generally been good for humankind. In the context of healthcare, crime, and terrorism, it can save lives. In the context of advertising, using predictions is more efficient and could conceivably save both trees (for direct mail and catalogs) and the time and attention of the recipient. In politics, it seems to reward those candidates who respect the scientific method (some might disagree, but I see that as a positive).

However, as Siegel points out—early in the book, which is admirable—these approaches can also be used in somewhat harmful ways. “With great power comes great responsibility,” he notes in quoting *Spider-Man*. The implication is that we must be careful as a society about how we use predictive models, or we may be restricted from using and benefiting from them. Like other powerful technologies or disruptive human innovations, predictive analytics is essentially amoral and can be used for good or evil. To avoid the evil applications, however, it is certainly important to understand what is possible with predictive analytics, and you will certainly learn that if you keep reading.

This book is focused on predictive analytics, which is not the only type of analytics, but the most interesting and important type. I don’t think we need more books anyway on purely descriptive analytics, which only describe the past and don’t provide any insight as to why it happened. I also often refer in my own writing to a third type of analytics—“prescriptive”—that tells its users what to do through controlled experiments or optimization. Those quantitative methods are much less popular, however, than predictive analytics.

This book and the ideas behind it are a good counterpoint to the work of Nassim Nicholas Taleb. His books, including *The Black Swan*, suggest that many efforts at prediction are doomed to fail because of randomness and the inherent unpredictability of complex events. Taleb is no doubt correct that some events are black swans that are beyond prediction, but the fact is that most human behavior is quite regular and predictable. The many examples that Siegel provides of successful prediction remind us that most swans are white.

Siegel also resists the blandishments of the “big data” movement. Certainly some of the examples he mentions fall into this category—data that is too large or unstructured to be easily managed by conventional relational databases. But the point of predictive analytics is not the relative size or unruliness of your data, but what you do with it. I have found that “big data often equals small math,” and many big data practitioners are content just to use their data to create some appealing visual analytics. That’s not nearly as valuable as creating a predictive model.

Siegel has fashioned a book that is both sophisticated and fully accessible to the non-quantitative reader. It's got great stories, great illustrations, and an entertaining tone. Such non-quants should definitely read this book, because there is little doubt that their behavior will be analyzed and predicted throughout their lives. It's also quite likely that most non-quants will increasingly have to consider, evaluate, and act on predictive models at work.

In short, we live in a predictive society. The best way to prosper in it is to understand the objectives, techniques, and limits of predictive models. And the best way to do that is simply to keep reading this book.

—**Thomas H. Davenport**

Thomas H. Davenport is the President's  
Distinguished Professor at Babson College,  
a fellow of the MIT Center for Digital Business,  
Senior Advisor to Deloitte Analytics,  
and cofounder of the International Institute for Analytics.  
He is the coauthor of *Competing on Analytics*,  
*Big Data @ Work*, and several other books on analytics.



# Preface to the Revised and Updated Edition

## ***What's New and Who's This Book for— The Predictive Analytics FAQ***

*Data Scientist: The Sexiest Job of the Twenty-first Century*

—Title of a *Harvard Business Review* article by  
Thomas Davenport and DJ Patil, who in 2015  
became the first U.S. Chief Data Scientist

Prediction is booming. It reinvents industries and runs the world.

More and more, predictive analytics (PA) drives commerce, manufacturing, healthcare, government, and law enforcement. In these spheres, organizations operate more effectively by way of predicting behavior—i.e., the outcome for each individual customer, employee, patient, voter, and suspect.

Everyone's doing it. Accenture and Forrester both report that PA's adoption has more than doubled in recent years. Transparency Market Research projects the PA market will reach \$6.5 billion within a few years. A Gartner survey ranked business intelligence and analytics as the current number one investment priority of chief information officers. And in a Salesforce.com study, PA showed the highest growth rate of all sales tech trends, more than doubling its adoption in the next 18 months. High-performance sales teams are four times more likely to already be using PA than underperformers.



I am a witness to PA's expanding deployment across industries. Predictive Analytics World (PAW), the conference series I founded, has hosted over 10,000 attendees since its launch in 2009 and is expanding well beyond its original PAW Business events. With the expert assistance of industry partners, we've launched the industry-focused events PAW Government, PAW Healthcare, PAW Financial, PAW Workforce, and PAW Manufacturing, events for senior executives, and the news site *The Predictive Analytics Times*.

Since the publication of this book's first edition in 2013, I have been commissioned to deliver keynote addresses in each of these industries: marketing, market research, e-commerce, financial services, insurance, news media, healthcare, pharmaceuticals, government, human resources, travel, real estate, construction, and law, plus executive summits and university conferences.

Want a future career in futurology? The demand is blowing up. McKinsey forecasts a near-term U.S. shortage of 140,000 analytics experts and 1.5 million managers "with the skills to understand and make decisions based on analysis of big data." LinkedIn's number one "Hottest Skills That Got People Hired" is "statistical analysis and data mining."

PA is like *Moneyball* for . . . money.

## FREQUENTLY ASKED QUESTIONS ABOUT *PREDICTIVE ANALYTICS*

### **Who is this book for?**

Everyone. It's easily understood by all readers. Rather than a how-to for hands-on techies, the book serves lay readers, technology enthusiasts, executives, and analytics experts alike by covering new case studies and the latest state-of-the-art techniques.

### **Is the idea of predictive analytics hard to understand?**

Not at all. The heady, sophisticated notion of *learning from data to predict* may sound beyond reach, but breeze through the short Introduction chapter and you'll see: The basic idea is clear, accessible, and undeniably far-reaching.