DATA MINING FOR INTELLIGENCE, FRAUD, & CRIMINAL DETECTION

Advanced Analytics & Information Sharing Technologies

CHRISTOPHER WESTPHAL

National Law Enforcement Officers
MEMORIAL FUND

CRC Press
Taylor & Francis Group

DATA MINING FOR INTELLIGENCE, FRAUD, & CRIMINAL DETECTION

Advanced Analytics &
Information Sharing Technologies

CHRISTOPHER WESTPHAL

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

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

International Standard Book Number-13: 978-1-4200-6723-1 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Westphal, Christopher R. (Christopher Ralph), 1965-
Data mining for intelligence, fraud & criminal detection : advanced analytics & information sharing technologies / Christopher Westphal.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-4200-6723-1 (alk. paper)
1. Law enforcement--United States. 2. Data mining--United States. I. Title.

HV7921.W47 2008
363.25'6--dc22 2008021209

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

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

Dedication

This book is dedicated to the analysts around the world who work diligently to secure the borders, critical infrastructure, and integrity of their homelands. It is for all those people who support law enforcement, the intelligence community, and corporate security. It is for the police officers, special agents, and criminal investigators who ensure our safety every day. It is for all the heroes who have given their lives to uphold our laws, protect our rights, and guarantee our freedoms.

All royalty proceeds from this book are being donated to the National Law Enforcement Officers Memorial Fund (NLEOMF) in honor of those individuals who have made the ultimate sacrifice to the service, protection, and security of others. More than 18,200 names, representing law enforcement officers who died in the line of duty, are engraved on the National Law Enforcement Memorial located in Washington, D.C. To learn more about NLEOMF and to further contribute to the fund, please visit their site at www.nleomf.com.

National Law Enforcement Officers
MEMORIAL FUND

Foreword

A lot has occurred in the world during the 10 years since I wrote my last book, *Data Mining Solutions,*[1] with Teresa Blaxton: Google is formally incorporated and Viagra is approved for prescription sale (1998); the euro currency is introduced into Europe and Y2K software concerns loom (1999); America Online buys Time Warner for $162 billion and Bon Jovi is still topping the music charts (2000); 9/11 shakes the world and *Shrek* is released into movie theaters (2001); the United States invades Afghanistan and Kelly Clarkson wins on the first season of *American Idol* (2002); the United States declares war with Iraq and Arnold Schwarzenegger gets elected the governor of California (2003); a massive tsunami in Southeast Asia kills more than 200,000 people and the Boston Red Sox win the World Series after 86 years (2004); Hurricane Katrina devastates New Orleans and gas prices in the United States inflate to more than $3 a gallon (2005); Saddam Hussein is hanged for his crimes against humanity and Microsoft formally releases the Vista operating system (2006); the iPhone is brought to market and Evel Knievel finally meets his maker (2007). In 2008 and beyond, we now have global warming concerns, the emergence of China as an economic powerhouse, and ever-expanding terrorist threats and incidents.

So, when Taylor & Francis Group approached me about doing another book, I had to ask myself, what has *really* changed in this field and is it worth writing about? There are already a number of data-mining books in the marketplace that briefly touch on a few of the topics that I would want to cover in a new book. However, most of the coverage is "simple" at best

[1] Christopher Westphal and Teresa Blaxton, *Data Mining Solutions: Methods and Tools for Solving Real-World Problems* (New York: John Wiley & Sons, 1998).

and there is little discussion of the real-world detail required to understand and implement the concepts presented. Additionally, many of these books are geared toward a more generalized audience and I wanted to focus on homeland security professionals and consultants, law enforcement officials, the intelligence community, corporate security personnel, intelligence analysts, special agents, special investigative units, private investigators, financial-crimes units, and broadly to corporate information technology (IT) professionals.

To write another book I would have to draw on my experience from a "real world" perspective—as someone who has been in the trenches implementing and structuring the analytical and information-sharing systems in use across a number of government programs and commercial industries. There would have to be little hype or dramatization with respect to how the systems are described and, if anything, I would have to err on the side of being too honest about the positive and negative aspects of what is really being done behind the closed doors of our intelligence and law enforcement agencies.

I thought about all the systems I have been involved with implementing, the different technology companies I have worked with over the years, and the numerous types of requirements defined by the user communities, and determined that there was enough advancement in the market to create a publication. Thus, I agreed to write this book, and after a number of iterations with the publisher, we decided to title it *Data Mining for Intelligence, Fraud, & Criminal Detection: Advanced Analytics & Information Sharing Technologies.*

Even though there have been many changes in the world, a lot has stayed the same, specifically in the context of information sharing and data analytics. The post-9/11 era has brought about many promises of sharing information, performing better analysis, and generally making the world a safer place for everyone. Every organization, bureau, agency, and corporation has fundamental analytical needs that traditionally require a significant amount of data integration and resources to best understand the data. Whether trying to identify money laundering, insider trading, insurance fraud, terrorist behavior, or other forms of criminal activity, the analytical processes and system architectures are very similar to each other. In fact, the types of patterns exposed in one domain can frequently be used in another, and it is often not necessary to reinvest and re-create these capabilities across different industries when a common approach can be used. This book will address these topics in depth and review the commonalities, framework, and infrastructures necessary to implement and deploy complex analytical systems.

In 2004, the Government Accountability Office (GAO) provided a report[2] detailing approximately 200 government-based, data-mining projects. In 2005, they issued a follow-up report[3] discussing privacy protections. These and other reports[4] show that there are many controls in place to ensure the systems are documented, audited, and accountable for the types of analytics they are delivering. What they do not state is the overall effectiveness of these systems—successes or pitfalls. This book will review several such systems and explain both how they function and how they produce results, and will provide an overall review of their capabilities and relative limitations (data, representation, and structure).

In addition to analytical approaches (technologies and methodologies), this book will also cover the topic of information sharing. Law enforcement agencies are always looking for better ways to conduct their investigations. On TV, shows like *CSI* (Crime Scene Investigation) and *NCIS* (Naval Criminal Investigative Service) depict elite teams of special investigators quickly resolving cases by accessing different high-tech resources to analyze the evidence. With a few clicks of a button, they search through their data archives to find the smoking gun—case solved. Traditionally, law enforcement agencies have not been as proficient with advanced technologies and although intriguing, these TV shows do not reflect what occurs in the mainstream community. This book will shed light on the current state of affairs within law enforcement, as well as within the intelligence and commercial communities.

A significant gap exists between local- and state-level investigative efforts of counterdrug, financial crimes, terrorism, and fraud. While sharing a common and collective goal of combating crime, there is currently little, if any, analytical collaboration and minimal data sharing among state and local law enforcement agencies because each organization operates independently. Although politics, jurisdictional boundaries, and other factors all play into how much one agency is willing to support the sharing of its resources, many agencies embrace the ability to make effective use of their data resources. This book will address a number of information-sharing issues and why no large-scale capabilities are currently deployed throughout the government. It will also review several commercial efforts that have had limited success.

[2] "Data Mining: Federal Efforts Cover a Wide Range of Uses," U.S. Government Accountability Office, GAO-04-548, May 2004, http://www.gao.gov/new.items/d04548.pdf.

[3] "Data Mining: Agencies Have Taken Key Steps to Protect Privacy in Selected Efforts, but Significant Compliance Issues Remain," U.S. Government Accountability Office, GAO-05-866, August 2005, http://www.gao.gov/new.items/d05866.pdf.

[4] "Data Mining Report," Office of the Director of National Intelligence, February 15, 2008.

In the rapid pace of our changing world, it is difficult to keep up-to-date with industry trends in complex fields, such as data mining, text processing, crime mapping, link analysis, and other forms of advanced analytics. Many investigators are not adequately trained in the IT field—although this is changing as more advanced training is being provided to investigators coming up through the ranks. To better foster cooperation and data sharing among different agencies, and to alleviate the current noncollaborative investigative situation, fusion centers and programs have been proposed, are under development, or are actively operating to address these issues. This book will dedicate a fair amount of time to discussing how current fusion centers are really being designed and will review their Achilles' heel in terms of being able to meet their stated objectives.

Currently, there is very little in published literature that truly defines real-world systems, how they are deployed, and the positive and negative aspects of their operations. Other books only briefly touch the surface of what is possible, or potentially can be done, leaving the reader wondering what the true status and capabilities are in today's high-end analytical systems. Most importantly, this book provides a significant number of examples based on real-world data, systems, and operations. Specifically, the analytical approaches presented throughout this book are heavily based on graph theory (e.g., connect the dots) because it holds the most promise for understanding large quantities of discrete-valued information.

The book is organized into three parts: Part 1 provides an overview of the main topics involved with understanding the types of data that can be used in current analytical and information-sharing systems. This section covers the fundamental approaches to analyzing data and clearly delineates how to connect the dots among different data elements. Part 2 is exclusively focused on providing real-world examples of how data is used, manipulated, integrated, and interpreted. All scenarios presented in this section are derived from operational systems. Finally, Part 3 provides an overview of many information-sharing systems, organizations, and task forces as well as data interchange formats. It also discusses more ideal information-sharing and analytical architectures for use across a broad spectrum of applications.

I feel it is important to stress that the content, opinions, explanations, discussions, and materials presented in this manuscript do not necessarily reflect the official views of, or make endorsements for, any government or private organization or product. The interpretations of the data, patterns, and results presented herein are entirely based on my personal observations and opinions and alternative interpretations are certainly encouraged. Reasonable efforts have been made to present the material in the most objective fashion possible; however, it is still derived from a subjective understanding and viewpoint. The accuracy of this content is made according to the best materials publicly and readily available at the time of research. There may be omissions or errors in the descriptions of some systems, laws, or processes, but they do not materially affect the concepts being conveyed to the readership. Additionally, this field is rapidly changing and new or updated statistics, numbers, or laws and regulations may be introduced after the period of research and writing of this book has been concluded; therefore all information should be revalidated if it will be used for more in-depth discussion or related research.

Acknowledgments

There are a number of people I would like to personally thank for helping to support the content, writings, background, research, review, and comments on this book. The book is a culmination of my experiences and exposure to different environments, situations, cultures, and scenarios. Many people have provided me with invaluable support throughout my career and helped define the approaches, methodologies, and techniques presented in this book. I am certainly indebted to all who have contributed to this material. Specifically, I would like to thank the following:

Mark Listewnik, my editor from Taylor & Francis, who was the catalyst in identifying the need for this book and pulling together the resources necessary to turn it into a reality. I appreciate the encouragement, support, and guidance he provided throughout this endeavor. Additionally, I would like to acknowledge my project editor, Ari Silver, for his coordination and our interactions throughout this process. Also, thanks to Susan Lagerstrom-Fife and Sharon Palleschi from Springer for their quick responses and permission to use important content from a previous publication.

Cameron "Kip" Holmes, chief of the Financial Remedies Section of the Arizona Attorney General's Office (AZAG), for his years of dedicated service combating financial crimes, and his innovative approaches to prosecuting money laundering activities. I would also like to thank Kip for supplying the breadth of materials documenting a number of the anti–money laundering operations on the Southwest Border. Detective John Shallue of the Phoenix Police Department for his service to law enforcement and for the case reference materials generously provided. Also, thanks to Hal

White from the AZAG for his insightful research, behind-the-scenes work, and for his dedication to law enforcement.

David R. Dugas, United States Attorney's Office for the Middle District of Louisiana, for his insight into several cases and prosecutions related to Hurricane Katrina. Michael Delafosse, director of the Office of Unemployment Insurance Administration, Louisiana Department of Labor, for his inputs on how the automated safeguards implemented in their systems help flag anomalies in their claims data.

Vicente S. Aquino, executive director, Anti–Money Laundering Council (AMLC) Secretariat (Philippines), for his diligence in combating financial crimes and for his input regarding the operations at AMLC. Agus "Gooseman" Surachman from Pelaporan dan Analisis Transaksi: Keuangan (PPATK) (Indonesia) for his energy and enthusiasm in wanting to expand his knowledge of the world and for contributing to this manuscript. Alejandro Rocchetti, Pablo Aliaga, and Danielle Iriarte of Systech S.A. (Argentina) for their support with Financial Intelligence Units (FIUs) throughout Latin America including Conselho de Controle de Atividades Financeiras (COAF) (Brazil), Secretaría de Prevención de Lavado de Dinero o Bienes (SEPRELAD) (Paraguay), and Unidad de Investigaćion Financiera (UIF) (El Salvador). Pavel Cherkashin from Security Problems Institute (SPI2) for his comments, inputs, and background on the Russian Financial Monitoring Service (FMS). Police Col. Seehanat Prayoonrat, the deputy secretary-general at the Anti–Money Laundering Office (AMLO) in Thailand, for his dedication and pursuit of prosecuting criminal enterprises and for positioning his country in a leadership role.

Andrew Shankman, Financial Crimes Enforcement Network (FinCEN), for his stimulating conversations, his realistic viewpoints on all matters large and small, and his voice of reason. Christina Klinger, FinCEN, for her steadfastness, realism, and unwavering support to do what is right. Thank you for the many years of support and for all the discussions and conversations we have had throughout those years. Shawn Polonet, acting Assistant Special Agent in Charge (ASAC), Immigrations and Custom Enforcement, New York, and Gary Murray, former director, New York High Intensity Financial Crime Area (HIFCA) for their devoted support to law enforcement and for their inputs and feedback on the El Dorado Task Force. Also, thanks to Mark A. Marshall, chief of police, Smithfield, Virginia, for his edits/comments as well as his involvement in creating information-sharing systems and for his perseverance in ensuring cooperation among law enforcement agencies.

Special thanks to Debbie Tyler from Visual Analytics Inc. (VAI) for her technical editing, insightful inputs, direct comments, and overall support during this process. David O'Connor (VAI) for his brilliance, his stellar programming abilities, and his ingenuity in creating VisuaLinks®—by far the best visualization and link discovery tool available in the open marketplace. Bennett McPhatter (VAI) for his determination and resolve to implement DIG® (Digital Information Gateway), perhaps one of the most successful real-world and operationally proven information-sharing tools in existence. Also, in memory of Carl Antonelli, a member of my staff at VAI, for his positive attitude, creative personality, vast intellect, and off-sense of humor; he is missed by many people.

The following also deserve my thanks for their continued support over the years: Dick Baer, Tom Carr, Leo Urbaniak, Kevin Hohn, Ken Middleton, Phil Canter, Robert Dintino, Brad King, Johan du Plooy, Peter Fryer, K. J. Min, Andrea Garavaglia, Paolo Consonni, Karl Eiselsberg, John Carbaugh, Jesus Cruz, Al Brandenstein, Mustafa Cem Arpaci, Frank Doe, Paul Bernier, Len Starling, as well as the entire staff at VAI.

To my son, Fletcher, who made sure that I had enough breaks and playtime to keep things properly balanced in the house. Finally, my biggest "thank you" is to my beautiful wife, Libby, for the support she provided throughout this whole undertaking. I appreciate her inputs and comments, her guidance, and most of all, her patience and endurance to ensure I had the time to dedicate to completing this book. I love you.

—Chris Westphal

The Author

Christopher Westphal is cofounder and CEO of Visual Analytics, Inc. (VAI—http://www.visualanalytics.com). Since its inception in 1998, he has guided the growth of the company from a fledgling start-up into a world-class provider of visualization software, information-sharing systems, and advanced analytical training. His clients include federal, state, and local law enforcement, all major intelligence agencies, the Department of Defense, civilian agencies, international Financial Intelligence Units (FIUs), and large corporations. VAI is a recognized leader in the intelligence and law enforcement industries and has garnered industry recognition and accolades, including Deloitte's Technology Fast 50 (Maryland), Maryland's International Leadership Award, and Maryland Muscle Award, and was named a finalist for the Ernst & Young Entrepreneur of the Year Award.

Prior to starting VAI, Mr. Westphal held key roles at the BDM Corporation, the Institute for Defense Analyses (IDA), Syscon (Logicon), and several other high-tech companies. During his collective tenures, he designed analytical systems and provided management expertise to critical government programs addressing organized crime, narcotics trafficking, money laundering, terrorism, tax evasion, insider trading, border crossings, smuggling, and criminal enterprises. He has supported a number of government offices and programs including the Office of National Drug Control Policy (ONDCP), Financial Crimes Enforcement Network (FinCEN), Internal Revenue Service (IRS) Criminal Investigations Division (CID), Drug Enforcement Administration (DEA), U.S. Army, Department of Justice (DOJ), Defense Intelligence Agency (DIA), Federal Bureau of Investigation (FBI), and Bureau of Alcohol, Tobacco and Firearms (ATF), and has consulted with government agencies in more than thirty countries on six continents.

Mr. Westphal has defined and created a number of unique and innovative approaches for performing visual data mining on large volumes of disparate data acquired from multiple sources. He is a recognized authority in the detection and exposure of complex patterns. Some of his investigative background stems from his early operational experience implementing anti–money laundering and fraud detection systems. He has also applied his expertise to help fight white-collar crimes and various forms of corruption, which has resulted in criminal convictions.

Mr. Westphal has worked personally with the banking and financial ministries throughout Europe, Asia, the Middle East, and South America in relation to fraud, anti–money laundering, embezzlement, asset forfeiture investigations, and various regulatory matters. He has worked to coordinate foreign regional practice areas across industry segments to aid enforcement and compliance. Fortune 1,000 companies, government agencies, high-tech companies, and consulting firms have sought advice from Westphal relating to compliance, risk management, data governance issues, and systems modernization.

He has authored numerous publications and several books including *Data Mining Solutions: Methods and Tools for Solving Real-World Problems* (Westphal/Blaxton, John & Wiley Sons, 1998) and *Readings in Knowledge Acquisition: Current Practices and Trends* (McGraw/Westphal, Ellis Horwood Limited, 1990), and a chapter in *Net-Centric Approaches to Intelligence and National Security* (Ladner/Petry, Springer, 2005). He has served as a referee, editor, and reviewer for large international journals and conferences on topics, such as data mining, expert systems, decision support, and data visualization.

Addressing data visualization, data mining, and the associated challenges and benefits of analyzing complex data sets, Mr. Westphal speaks frequently on these matters before government, banking, legal, compliance, and academic audiences. He has lectured internationally to thousands of business and information technology professionals. He is routinely asked to speak and participate on technology panels and sought out for his expert opinion on advanced analytical systems and data-mining methodologies.

Mr. Westphal was born and raised in New York. He received his B.S. in computer science from the School of Engineering at Tulane University. He can be reached at westphal@visualanalytics.com or chris_westphal@yahoo.com.

Contents

PART 2 REAL-WORLD EXAMPLES AND OPERATIONS