Data Mining and Predictive Analysis

Praise for Data Mining and Predictive Analysis

"Dr. Colleen McCue pairs an educational background in neuroscience and psychology with extensive experience in the fields of behavioral science, cirme analysis, and intelligence gathering to create *Data Mining and Predictive Analysis*, a must-read for all law enforcement professionals. Within the ever-growing fields of criminal justice and crime analysis, Dr. McCue combines all facets of the public safety community, effortlessly examining techniques in which law enforcement, analysts, and researchers are able to delve deeper through her accessible explanations of relative degrees of data quality, validity and reliability; all essential tools in this modern, technological era."

Arthur E. Westveer (Associate Professor, L. Douglas Wilder School of Government and Public Affairs, Virginia Commonwealth University)

"[Data Mining and Predictive Analysis] is a must-read . . ., blending analytical horsepower with real-life operational examples. Operators owe it to themselves to dig in and make tactical decisions more efficiently, and learn the language that sells good tactics to leadership. Analysts, intell support, and leaders owe it to themselves to learn a new way to attack the problem in support of law enforcement, security, and intelligence operations. Not just a dilettante academic, Dr. McCue is passionate about getting the best tactical solution in the most efficient way—and she uses data mining to do it. Understandable yet detailed, [*Data Mining and Predictive Analysis*] puts forth a solid argument for integrating predictive analytics into action. Not just for analysts!"

Tim King (Director, Special Programs and Global Business Development, ArmorGroup International Training)

"Dr. McCue's clear and brilliant guide to attacking society's greatest threats reveals how to best combine the powers of statistical computation and the experience of domain experts. Her emphasis on understanding the essential data through fieldwork and close partnership with the end users of the information is vital to making the discovered patterns "actionable". Anyone seeking to harness the power of data mining to "connect the dots" or "find needles in a haystack" will benefit from this lively and reliable book packed with practical techniques proven effective on tough real-world problems."

Dr. John Elder (Chief Scientist of Elder Research, Inc., www.datamininglab.com)

"[Data mining] is a hot area—not just for Hollywood any more—but real people and real situations are benefiting from these analytical investigations. "

Mary Grace Crissey (Technology Marketing Manager, SAS Institute)

Data Mining and Predictive Analysis

Intelligence Gathering and Crime Analysis

Colleen McCue

∞ Recognizing the importance of preserving what has been written, Elsevier prints its books on
acid-free paper whenever possible.

This book is dedicated to Patrick Michael McLaughlin,
the first miner in our family.

Contents

Foreword

We all know crime doesn't pay. But did you know there is "prophet" in policing? Thanks to the fine work of Dr. Colleen McCue of the Richmond Police Department, Crime Analysis Unit, it is now possible to predict the future when it comes to crime, such as identifying crime trends, anticipating hotspots in the community, refining resource deployment decisions, and ensuring the greatest protection for citizens in the most efficient manner.

A number of years ago, the United States Attorney's Office for the Eastern District of Virginia formed a partnership with the Richmond Police Department to address the pressing problem of gun violence in the city. In 2002, we renewed that relationship and formed a new commitment as part of President George W. Bush's antigun crime initiative, Project Safe Neighborhoods (PSN). At that time, Dr. McCue was selected as our research partner to assist our efforts in evaluating the outcomes of our districtwide PSN initiatives. In light of the work Dr. McCue was already doing for the Richmond Police Department, we wanted to apply the innovative tools she had used so effectively in Richmond to support our efforts targeting gun crime in other hot spots around eastern Virginia.

Dr. McCue has done pioneering work in the practical application of datamining techniques to the administration of a police department. In this book, she describes her use of "off-the-shelf" software to correlate data on gun violence with data on other violent crimes in order to graphically depict crime trends in a most compelling way and to predict where future crimes are likely to occur. Armed with such analyses, the police executive is thus enabled to develop "risk-based deployment strategies," permitting the executive to make informed and cost-efficient staffing decisions based on the likelihood of specific criminal activity.

The application of Dr. McCue's techniques has paid off in Richmond, where the police department used them to deploy resources during the period surrounding the New Year's Eve holiday—December 31, 2003, through January 1, 2004. The results of that effort were dramatic. Not only were gunfire complaints reduced by almost 50% on New Year's Eve, but the number of seized

illegal weapons increased by an impressive 246% from the previous year. These statistics represent compelling evidence that these techniques are adding value to the work of fighting gun crime. But there is more. This accomplishment was realized using fewer street officers than originally planned. In other words, risk-based deployment enabled the Richmond Police Department to deploy fewer officers strategically, while at the same time obtaining better results.

In writing this book, Dr. McCue was mindful of the need to convey sophisticated analyses in practical terms and, accordingly, she prepared her text in a very user-friendly manner. As United States Attorney, I am proud to be associated with such a dedicated partner in our shared mission. I am confident that you, too, will benefit from Dr. McCue's exceptional contribution to the field of police science.

Paul J. McNulty

Preface

Like many kids growing up in America, I always had a love of science. I also happened to be blessed with two incredibly supportive and involved parents. My mother was always there with words of encouragement. Her typing skills got me through high school and most of college. She also led by example, balancing her work as a probation and parole officer with her role as wife and mother. My father, on the other hand, would try to learn as much as he could about what we were interested in so that he could participate in the activities with us. When I started graduate school, however, there was something of a dilemma. What do an engineer and a budding neuroscientist have in common, particularly when the engineer is not big on things like rats and brains? Fortunately, it was during this time that cognitive neuroscience and artificial intelligence systems started becoming accessible to the mainstream. So, throughout graduate school and my subsequent career, my father would send me books and articles on topics such as neural nets, case-based reasoning, machine learning, and cognitive neuroscience. It provided for interesting conversation and some common ground for two professionals in relatively disparate fields.

As time went on and life changed, I found myself working as a behavioral scientist in the criminal justice field. In this environment, I was able to bring my training as a scientist to the study of human criminal behavior. I found that I was able to apply much of what I had learned about psychology, behavioral science, and, perhaps most importantly, multivariate statistics and computer modeling to my new field. I was in an interesting position, working in a local police department and receiving first-hand training in a variety of topics, from death investigation to CompStat. While I did not realize it at the time, I also was acquiring a tremendous amount of domain expertise, something absolutely essential to competent data mining, which would distinguish my work from many others trying to gain entry into a rather closed professional world. I also began to understand the relative degrees of data quality, validity, and reliability associated with law enforcement and intelligence data. Although I was familiar with the work regarding the often questionable reliability of eyewitness

testimony, it was not until I had read many offense reports that trends and patterns to the witness statements began to emerge and make sense.

I became profoundly intrigued by how many of the seasoned detectives who I worked with were often able to generate quick yet accurate hypotheses about their cases, sometimes only moments after they had arrived at the scene. Like the "profilers" on television and in the movies, many of them seemed to have an uncanny ability to accurately describe a likely motive and related suspect based merely on a review of the crime scene and some preliminary knowledge regarding the victim's lifestyle and related risk factors. Over time, I started to acquire this ability as well, although to a lesser degree. It became much easier to read a report and link a specific incident to others, predict future related crimes, or even calculate the likelihood that a particular case would be solved based on the nature of the incident. Drawing on my training as a scientist, I frequently found myself looking for some order in the chaos of crime, trying to generate testable hypotheses regarding emerging trends and patterns, as well as investigative outcomes. Sometimes I was correct. However, even when I was not, I was able to include the information in my ever-expanding internal rule sets regarding crime and criminal behavior.

Prior to working for the Richmond Police Department, I spent several years working with that organization. Perhaps one of the most interesting aspects of this early relationship with the Department was my weekly meetings with the Officer in Charge of Violent Crimes. Each week we would discuss the homicides from the previous week, particularly any unique or unusual behavioral characteristics. Over time, we began to generate casual predictions of violent crime trends and patterns that proved to be surprisingly accurate. During this same time period, I began to examine intentional injuries among incarcerated offenders. As I probed the data and drilled down in an effort to identify potentially actionable patterns of risk, it became apparent that many of the individuals I looked at were not just in the wrong place at the wrong time, as they frequently indicated. Rather, they were in the wrong place at the wrong time *doing the wrong things with the wrong people* and were assaulted as a result of their involvement in these high-risk activities. As I explored the data further, I found that different patterns of offending were associated with different patterns of risk. This work had immediate implications for violence reduction, something that I continue to be involved in. Similarly, it had implications for the analysis of crime and intelligence data. Fortunately, the field of data mining and predictive analytics had evolved to the point that many of the most sophisticated algorithms were available in a PC environment, so that everyone from a software-challenged psychologist like myself to a beat cop could begin to not only understand but also use these incredibly powerful tools. Unfortunately, the

transfer of this powerful technology to the public safety arena has not advanced nearly as quickly.

While I did not realize it at the time, a relatively new approach to marketing and business was emerging at the same time we were engaging in this lively speculation about crime and criminals at the police department. Professionals in the business community were exploiting artificial intelligence and machine learning to characterize and retain customers, increase sales, focus marketing campaigns, and perform a variety of other business-related tasks. For example, each time I went through the checkout counter at my local supermarket, my purchasing habits were coded, collected, and analyzed. This information was aggregated with data from other shoppers and employed in the creation of models about purchasing behavior and how to turn a shopper into a buyer. These models were then used to gently mold my future behavior through everything from direct marketing based on my existing preferences to the strategic stocking of shelves in an effort to encourage me to make additional purchases during my next trip down the aisle. Similarly, data and information were collected and analyzed each time I perused the Internet. As I skipped through web pages, I left cookies, letting the analysts behind the scenes know where I went and when and in what sequence I moved through their sites. All of this information was analyzed and used to make their sites more friendly and easier to navigate or to subtly guide my behavior in a manner that would benefit the online businesses that I visited. The examples of data mining and predictive analytics in our lives are almost endless, but the contrast between my professional and personal lives was profound. Contrasting the state of public safety analytical capacity to that of the business community only serves to underscore this shortcoming. Throughout almost every aspect of my life, data and information were being collected on me and analyzed using sophisticated data mining algorithms; however, the use of these very powerful tools was severely limited or nonexistent in the public safety arena in which I worked. With very few exceptions, data mining and predictive analytics were not readily available for the analysis of crime or intelligence data, particularly at the state and local levels.

Like most Americans, I was profoundly affected by the events of September 11th. The week of September 10th, 2001, I was attending a specialized course in intelligence analysis in northern Virginia. Like many, I can remember exactly what I was doing that Tuesday morning when I saw the first plane hit the World Trade Center and how I felt as the horror continued to unfold throughout the day. As I drove back to Richmond, Virginia, that afternoon (the training had been postponed indefinitely), I saw the smoke rise up over the Beltway from the fire at the Pentagon, which was still burning. Those of us working in the public safety community were inundated with information

over the next several days, some of it reliable, much of it not. Like many agencies, we were swamped with the intelligence reports and BOLOs (be on the lookout reports) that came in over the teletype, many of which were duplicative or contradictory. Added to that were the numerous suspicious situation reports from concerned citizens and requests for assistance from the other agencies pursuing the most promising leads. Described as the "volume challenge" by former CIA director George Tenent, the amount of information almost continuously threatened to overwhelm us. Because of this, it lost its value. There was no way to effectively manage the information, let alone analyze it. In many cases, the only viable option was to catalog the reports in three-ring binders, with the hope that it could be reviewed thoroughly at some later date. Like others in law enforcement, our lives as analysts changed dramatically that day. Our professional work would never again be the same. In addition to violent crimes and vice, we now have the added responsibility of analyzing data related to the war on terrorism and the protection of homeland security, regardless of whether we work at the state, local, or federal level. Moreover, if there was one take-home message from that day as an analyst, particularly in Virginia, it was that the terrorists had been hiding in plain sight among us, sometimes for years, and they had been engaging in a variety of other crimes in an effort to further their terrorist agenda, including identity theft, forgery, and smuggling, not to mention the various immigration laws they violated. Many of these crimes fall within the purview of local law enforcement.

As we moved through the days and weeks following the attacks, I realized that we could do much better as analysts. The subsequent discussions regarding "connecting the dots" highlighted the sad fact that quite a bit of information had been available before the attacks; however, flaws in the analysis and sharing of information resulted in tragic consequences. While information sharing will require culture change and a paradigm shift in the larger public safety community, advanced analytical techniques are available now. The same tools that were being used to prevent people from switching their cellular telephone service provider and to stock shelves at our local supermarkets on September 10th can be used to create safer, healthier communities and enhance homeland security. The good news is that these techniques and tools are used widely in the business community. The key is to apply them to questions or challenges in public safety, law enforcement, and intelligence analysis. Adapting existing technologies and analytics to the public safety domain will keep many of us busy for years to come. If the past is any indicator, however, by the time we have completed this initial technology transfer and have caught up to where the business community is today, there should be other new and exciting technologies to appropriate from the private sector. In all seriousness, the public safety

community has become extremely adept at developing and adapting new and advanced technologies for operational capacity and support. The battlefields have changed, though. To achieve dominance in the war on terrorism, the war on drugs, and the war on crime, we need to devote additional attention to our ability to manage, analyze, and utilize the incredible amounts of information available. Ultimately, data mining and predictive analytics offer the promise of allowing data and information to serve as a transparent, fluid interface between analytical and operational personnel, rather than the vast ideological divide that frequently is encountered today.

Although I say "I" quite a bit in this book, the book certainly was not created in a vacuum. Countless individuals have helped me throughout my career, and a few have truly inspired me. What follows is a very brief list of those that contributed directly to this effort in some way.

I would like to thank Dave Dunn from Advizor Solutions, Inc. Dave first suggested that I write this book, and it never would have occurred to me that this was possible without his feedback and support. Mark Listewnik at Elsevier has the patience of a saint. His ongoing support and encouragement, not to mention the very nice Christmas cards that I continued to receive despite the fact that I was horrendously late on my rewrite and edits, kept me going if for no other reason than I felt very guilty putting things off even further in the face of his ongoing kindness. Finally, Kayla Gray at RTI International edited the manuscript and helped create something far more readable than what I originally wrote. Her attention to detail and thoughtful comments are reflected throughout the text.

Most of the early work referenced came out of some very lively discussions that began several years ago with my colleagues at the Federal Bureau of Investigation. In particular, Supervisory Special Agents Charlie Dorsey and Dr. Wayne Lord provided considerable guidance to my early research. Over time, they have become both colleagues and friends, and my work definitely reflects a level of quality that is attributable directly to their input. Also with the FBI, Mr. Art Westveer taught me almost everything that I know about death investigation. I have learned a tremendous amount from his lectures, which are punctuated with his dry sense of humor and wonderful anecdotes from a very successful career with the Baltimore Police Department. Rich Weaver and Tim King, president and vice president, respectively, at International Training, Inc. graciously allowed me to attend their lectures and training on surveillance detection in support of my research. They also provided some very unique opportunities for field testing many of my ideas in this area to see how well they would play in the real world.

While many of my former employers merely tolerated my analytical proclivities, the Project Safe Neighborhoods folks provided funding, as well as ongoing support and encouragement for much of the recent work outlined in this book. In particular, Paul McNulty, the United States Attorney for the Eastern District of Virginia, carried the message of our success far beyond the audience that I could reach alone. I also want to recognize Frank Shults and Brian Whisler, who are blessed with both excellent writing skills and unbridled humility. To them, I am grateful.

I also would like to thank Dr. Harvey Sugerman. I still remember the day when he called me out of the blue and told me that he thought that I should be paid for the work I had been doing. A single mother, I had been responding to homicide calls on my own time in the evenings in an effort to gain additional knowledge and insight into violent crime and the investigative process. That particular act made a tremendous positive impact in my life. I gained invaluable experience through my affiliation with the University, but his gentle mentoring and decision to offer me compensation for my work only begins to underscore the kindness in his heart.

I owe a tremendous debt of gratitude to the software companies that have provided me with some wonderful toys—I mean software. Without their support, I would still be performing unnatural acts with multivariate statistics and trying to convey the results to operational personnel with a lot of hand waving. In particular, Dr. Tom Khabaza and Bill Haffey with SPSS, Tracye Giles at SAS, and Dave Dunn with Advizor Solutions trusted me enough to give me the tools to do a lot of the work outlined in this book.

My family at the Richmond, Virginia Police Department has taught me almost everything that I know about police work and law enforcement. To name every individual that has contributed to my training and life would resemble a roll call of the current and previous command, as well as the line staff, who frequently know as much if not more than their supervisors. In particular, I would like to thank Colonel Andre Parker and Lieutenant Colonel Teresa Gooch for their ongoing support of my work. I also would like to thank Jerry Oliver, the former Chief of the Richmond Police Department, who, with Teresa Gooch, recruited me for the most rewarding yet challenging position I have ever enjoyed. Other colleagues in the Department include Captain David Martin, Majors Peggy Horn and Dave McCoy, the late Major Rick Hicks and Captain Donnie Robinson, and my friend Alicia Zatcoff, Esq. I also owe a tremendous debt to the Virginia Homicide Investigators Association, where I have received some outstanding training in death investigation and was fortunate enough to

meet my husband, who is a member of their board of directors. My colleagues in law enforcement have taught me as much, if not more, about life in the many years that I worked with them.

Underscoring the length of time that it took me to complete the text, I changed employment during the writing of this book. After several years in the applied setting, I joined RTI International, a nonprofit research organization with an international reputation for excellence in criminal justice research. The ability to work with other like-minded researchers in an effort to advance the science and practice of public safety and security has been energizing. In particular, Dr. Victoria Franchetti Haynes, president and CEO of RTI International, has created an environment that fosters creativity and the opportunity to improve the human condition by turning knowledge into practice. Adam Saffer and Brent Ward have helped me translate my work into something tangible that can be shared with other public safety and security organizations through the creation of technology and the provision of professional services. Other colleagues at RTI include Drs. Al Miedema and Jim Trudeau, and MG (Ret) Lon "Bert" Maggart, as well as the other members of my research team, which includes Dr. Kevin Strom and Mark Pope. Confucius said that if you love your job you will never work a day in your life, something that I am blessed to live.

I also would like to thank Mike Sullivan, USMC Staff Sergeant Tom Ferguson, and Special Agent BJ Kang for giving me permission to use their photographs throughout this book. Their photographs graphically illustrate our recent history as a nation and serve to further underscore the importance of fighting the good fight, and doing so with honor. Joey Vail from SAS, Bill Haffey from SPSS, Eric Greisdorf from Information Builders, and Kurt Rivard from Advizor Solutions all provided screen shots that illustrated the value that their software can bring to applied public safety and security analysis.

Perhaps most importantly, I would like to acknowledge my family. My parents, Phil and Lucy McLaughlin, always expected the best from me and my siblings, Michele and Tim, giving us the tools necessary to achieve that and more. This included loving words and kind gestures, as well as giving us permission to find our own way in life. My path has not always been direct or easy, but they always loved me enough to allow me to find my own way, having faith in me even when I did not. Some of the most challenging lectures that I have ever given professionally were the ones where they were in the audience. To look out and see their faces filled with pride was at once humbling, heart-warming, and also terrifying. Who would have known that the girl from

Downers Grove, Illinois, who started out as an aerospace engineering student would have taken the career path that I did? It still seems amazing to me at times, but I know that I am a far more successful person because of it. Unfortunately, I think that my parents went prematurely grey in the process. Hopefully, it was worth it.

In many ways, my husband, Special Agent Rick McCue, has contributed more than enough to have earned the right to be a coauthor. Through him, I have first-hand insight into the needs of operational personnel and the importance of making analytical products accessible to the folks that need them the most: those on the front lines. Whether with outright encouragement or a vacant stare when I became long-winded or obtuse, he has provided invaluable guidance to my skills as an analyst. I also would like to thank the United States government for sending him out of the country so much during the writing of this book. I always looked for projects to occupy my time when he was out of pocket; we could not afford any more redecorating, so this book seemed like a good alternative. In all seriousness, though, I am forever grateful for the experiences that I have had vicariously through my husband. As one of the team assigned to the Pentagon recovery immediately after September 11th, my husband saw first-hand the devastation that the terrorist agenda can rain down on innocent lives. I know that neither of us will ever be the same. In his subsequent missions with Operations Noble Eagle and Iraqi Freedom, I began to truly understand the value that good intelligence and analysis will bring to the war on terrorism.

Our children, Paul, Alexandra, Elaine, Patraic, and Gabriel, keep me humble. Although Rick and I lead very exciting lives professionally, our kids still think that we are the biggest dorks in the whole world, clueless and goofy. That fact alone keeps me anchored in reality and reminds me daily what is most important in life. Like many folks in public safety, there have been more than a few times that I have come home and hugged my children a little bit harder because of what I have seen or done at work. I am so grateful to be blessed with such a wonderful life and family, which makes me work that much harder for those who are not. I believe that other women love their children just as I do. Unfortunately, too many of their children will not be coming home again. Whether it is the result of drugs, gang violence, or the war on terrorism, there is too much pain and suffering in our world, too much killing. For that reason, as a homicide researcher, it always has been important for me to remember that every one of the "subjects" in my studies is a lost life, a devastated family, and a loss to our community. In all humility, it is my sincere wish that the techniques and approaches outlined in this book will help us increase the health

and well-being of our communities and create safer neighborhoods for all of our children.

"If there must be trouble, let it be in my day, that my child may have peace."
Thomas Paine

Colleen McLaughlin McCue, PhD
Senior Research Scientist
RTI International

Introduction

Good analysts are like sculptors. They can look at a data set and see underlying form and structure. Data mining tools can function as the chisels and hammer, allowing the analysts to expose the hidden patterns and reveal meaning in a data set so that others can enjoy its composition and beauty.

Whether it is called data mining, predictive analytics, sense making, or knowledge discovery, the rapid development and increased availability of advanced computational techniques have changed our world in many ways. There are very few, if any, electronic transactions that are not monitored, collected, aggregated, analyzed, and modeled. Data are collected about everything, from our financial activities to our shopping habits. Even casino gambling is being analyzed and modeled in an effort to characterize, predict, or modify behavior.

One area that has been somewhat limited in its acceptance and use of these powerful new techniques is the public safety community, particularly in security, crime prevention, and crime analysis. This is somewhat surprising because in many ways analysts, detectives, agents, professionals in the intelligence community, and other operational personnel embody many of the principles of data mining or knowledge discovery. For example, the process of training detectives in investigative techniques and practices bears a strong resemblance to case-based reasoning.[i] In addition, the characterization, modeling, and prediction associated with the behavioral analysis of violent crime are very similar to some of the categorization, linking, and predictive analytics associated with data mining and predictive analytics.

While the relationship between the two areas seems to be natural, the law enforcement community in particular has not enjoyed many of the analytical benefits coming from these powerful new tools. It is unclear whether this is due to cost, training, or just a lack of knowledge of the existence and availability of these tools, but when they are adopted, the increased quality of life for law enforcement personnel, as well as the communities that they serve, is remarkable. In these times of dwindling economic and personnel resources,

no agency can afford to deploy carelessly. As organizations compete for qualified personnel, a candidate's final decision often comes down to quality of life and job satisfaction issues. Just a few of the questions potential employees ask themselves before making a final decision are: Will I have a reasonable work schedule? Will I be able to manage my workload effectively? Will my time be used productively? Can I make a difference in my community? Similar decision processes are associated with maintaining a satisfied work force and long-term retention—something that is increasingly difficult, given the rapidly emerging employment opportunities for law enforcement personnel.

At the same time, requirements for accountability and outcome studies are coming from funding agencies and constituents alike. It is no longer acceptable to run programs without the outcome indicators and metrics necessary to demonstrate their efficacy. The emphasis on these measures of accountability highlights the need for new methodologies to document progress and change in response to new initiatives and strategies.

Given the infinitely increasing amounts of information, "connecting the dots" will be possible only with automated systems. Perhaps more important than trying to create these associations, though, will be addressing gaps in information and information sharing. Only after these challenges have been addressed will we be able to identify and characterize trends and patterns so that future events can be predicted, anticipated, and perhaps even prevented. The emphasis needs to shift from describing the past to predicting the future. Only then will we have the possibility to enhance public safety and create safe neighborhoods for all.

Skill Set

Analysts are deluged with information on a daily basis. The ability to bring some order into this informational chaos can have a huge impact on public safety and the quality of life in the communities that they serve. On the other hand, the opportunity to bring analytical and predictive models directly into the operational environment holds the promise of giving public safety and intelligence professionals the ability to maneuver within the decision and execution cycles of their opponent. Whether it is the war on terrorism, the war on drugs, or the war on crime, enhanced knowledge and the ability to anticipate future actions can afford operational personnel essential situational awareness.

Knowledge of advanced statistics is not a prerequisite for using predictive analytics. In fact, the discovery process associated with data mining also could be viewed as after-the-fact explanations for unpredicted outcomes, something

somewhat distasteful in inferential statistics. When examined under the intense scrutiny of the analyst's domain knowledge, however, these unanticipated or surprising findings can have significant value and greatly enhance our understanding of crime and intelligence data. For those who are analytically inclined, it can be a wonderful and exciting process of data exploration and discovery. Those with a strong background in statistics, though, might be somewhat handicapped by the comparatively rigid nature of inferential statistics, with all of its associated rules and assumptions. With a little confidence and practice, even statisticians will be able to overcome their previous training and perform what they once considered to be unnatural acts with data and information.

On the other hand, data mining brings powerful analytics to those who really need them, including operational personnel. In my experience, it is far easier to teach someone with interest who knows something about crime and criminals how to effectively use these tools. With some guidance regarding a few "rules of the road" for data mining, and the application of off-the-shelf software tools, data mining is well within the reach of any organization with an interest and willingness to put more science and less fiction into crime and intelligence analysis. Moreover, many of the new tools have been adapted to run in a web-based environment and are no more difficult than making a purchase or completing a survey over the Internet. These advancements have created the opportunity for "24/7" analytical capacity,[ii] even within smaller agencies with comparatively limited personnel resources.

The more that operational personnel, managers, and command staff understand the information requirements and possible outcomes from analytical products; the more likely they will be to contribute data that is meaningful, detailed, and valuable. They also will be in a better position to work with the analyst and participate in the analytical process, requesting output that has increased value for them as they acquire a better understanding of what is available. By understanding the importance of the data inputs and the potential range of outputs, operational personnel, managers, and command staff alike can become informed information consumers and increase the likelihood of identifying actionable output from the analytical process. This subtle change in relationships and understanding can greatly enhance analysts' ability to gather the necessary data and information, ultimately increasing their ability to support operational personnel, policy decisions, managers, and command staff.

At a recent security expo, Tom Clancy advised the security and intelligence professionals in the audience to seek out the "smart people," observing that, "[t]he best guys are the ones who can cross disciplines . . . [t]he smartest ones look at other fields and apply them to their own."[iii] In my opinion, many of

the "smart people" Clancy refers to will rise out of the operational ranks, given the intuitive nature and relative ease of use associated with the new generation of data mining and predictive analytics software tools. While most analysts probably do not need to fear for their jobs just yet, increasingly friendly and intuitive computer systems will allow data and information to serve as a fluid interface between analytical and operational personnel. At some point in the future, that distinction will become almost meaningless with the emergence of increasingly powerful software tools and systems and the "agent/analysts" that employ them.

"Agent/Analysts" and Future Trends

I see a day in the not-too-distant future when analysis will be available without immediate access to an analyst. Information from operations will feed analysis, while the analysis will concomitantly drive the operations, thereby creating a feedback loop of ever-increasing information and actionable intelligence. I see a day when a patrol officer will come back to work after several days off and, at the beginning of the tour, will be able to review recent patterns and trends within the context of historical data and accumulated knowledge from the mobile data terminal in his cruiser. After responding to his first call, he will be able to enter the incident information directly into the department's computerized records management system (RMS) using direct voice commands. This information then will be used to create the computerized offense report. Any digital images captured from the incident will be quickly uploaded and linked directly to the offense report, as well as any associated or linked information already stored in the RMS. During the data entry process, this new information will pass through an analytical filter prepared earlier in the week by the analytical staff, who are home asleep at this hour. The algorithm running in the background will quickly link this most recent incident to a recent series and prompt the patrol officer to consider several possible alternatives. With this real time, value-added analysis, the officer can make quick, information-based operational decisions that result in a rapid apprehension of the criminal.

This is handled similarly when an agent in a remote location is debriefing a suspected terrorist. The verbal information is recorded and transcribed directly into a free format text file using voice recognition software. The file is then uploaded to an analytical fusion center a thousand miles away. An analyst there uses sophisticated text mining technology to probe and characterize the results of the interview. Several key phrases are identified and compared to an existing database generated from earlier interviews with members of the same terrorist

cell being held in other locations around the world. Based on the analysis of the current interview and its comparison to the existing models, areas of possible deception and truth are identified and highlighted, as are promising interviewing strategies. This information, including the interviewing strategies and approaches, is sent back to the agent in the field, further informing and guiding the ongoing interview process, while concomitantly enhancing the existing intelligence on the operations, practices, and strategies of this particular terrorist group.

Are these extravagant predictions? Absolutely not. Both scenarios outlined above are based on existing technologies and resources. In many ways, approaches and methodologies similar to information management have been used in the business community for years. All that is required to implement these strategies is a commitment to take advantage of the currently existing analytical tools and incorporate them into our world. Unfortunately, a paradigm shift in how we view information, analysis, and the relationship between analytical and operational personnel also will be required. That probably will be the most difficult task. Once we overcome that hurdle, however, adapting these new technologies promises to be one of the most exciting adventures in public safety in our lifetime.

How To Use This Book

All of the examples included in this book come from real experience. In some cases, though, the specifics have been changed to protect ongoing investigations, sensitive data, or methods. Whenever possible, I have tried to distinguish between real cases, particularly those taken from published work, and those generated specifically as examples. Given the nature of some topics covered in this book, however, it would be inappropriate to provide too much specific detail and compromise methods. To be sure, though, while the names might have been changed to protect the "not so innocent," the examples are based on real experiences.

This book is divided into five main sections: "Introduction," "Methods," "Applications," "Case Examples," and "Advanced Concepts/ Future Trends." The third and fourth sections include annotated examples focusing on the why and how, as well as the limitless possibilities for data mining and predictive analytics in crime and intelligence analysis. While this organization is relatively logical for training purposes, many readers will choose to read the book out of sequence. In particular, managers, command staff, supervisors, policy makers, and operational personnel interested in learning more about data mining and

predictive analytics but not expecting to use these tools first hand will have neither an interest in nor a need for detailed information on specific methods and algorithms. These readers could benefit from reading and understanding the annotated examples if they make acquisition and purchasing decisions for analytical products and determine the focus of their analytical personnel. Moreover, operational personnel can more fully exploit the new technology and work more effectively with analytical personnel if they understand the vast array of possibilities available with these new tools. With the opportunity to deploy data mining and predictive analytics directly into the field, an increasing number of operational personnel will be using data mining products. While they might not be generating the specific algorithms or models, a general understanding of data mining and predictive analytics will certainly enhance their ability to exploit these new opportunities.

Similarly, many analysts will use this book to explore the possibilities for data mining in their environment; identifying ideas and strategies from the annotated examples in the third section, and then returning to the methods section for specific information regarding the use and implementation of these approaches. This book is not intended to provide detailed information about specific software packages or analytical tools, but merely provides an overview of them. It should serve as a starting point, using terminology, concepts, practical application of these concepts, and examples to highlight specific techniques and approaches in crime and intelligence analysis using data mining and predictive analytics, which each law enforcement or intelligence professional can tailor to their own unique situation and responsibilities. While the basic approaches will be similar, the available data, specific questions, and access to technology will differ for each analyst and agency, requiring unique solutions and strategies in almost every setting.

Perhaps one of the most challenging aspects of writing this book was keeping abreast of the new developments and data mining applications that now appear on an almost daily basis. It is both frustrating and exciting to consider how much this field is likely to change even in the short time between completion of the manuscript and actual publication of the text. Therefore, the final section, "Advanced Concepts/ Future Trends," should not be viewed as inclusive. Rather, this particular section is intended to serve as a beginning for ascending to the next level of training for those interested in this field. This rapid pace of innovation, however, is what keeps the field of analysis fresh and exciting, particularly for those with the interest and creativity to define the cutting edge of this new and evolving field.

Bibliography

i. Casey, E. (2002). Using case-based reasoning and cognitive apprenticeship to teach criminal profiling and internet crime investigation. Knowledge Solutions. www.corpus-delicti.com/case_based.html

ii. McCue, C. and Parker, A. (2004). Web-based data mining and predictive analytics: 24/7 crime analysis. *Law Enforcement Technology*, **31**: 92–99.

iii. Fisher, D. (2003). Clancy urges CIOs: seek out the "smart people." eWeek, www.eweek.com.