Preface

There is not a crime, there is not a dodge, there is not a trick, there is not a swindle, there is not a vice which does not live by secrecy. – Joseph Pulitzer

The data mining field is receiving a significant attention in today’s information rich society, where data is available from different sources and formats, in large volumes, and no longer constitutes a bottleneck for knowledge acquisition. This rich information has paved the way for novel areas of research, particularly in the crime data analysis realm. Further, the potential usefulness of data mining in various domains has resulted in a holistic development of useful analytics for knowledge support in the crime data analysis domain. Data mining approaches, involving techniques at the intersection of machine learning, artificial intelligence, statistics and database systems, have proven to be valuable tools in discovering interesting, non-trivial and potentially useful information, which can assist in decision support, crime investigation, prediction and control. The importance of data mining is clearly evident in the plethora of least utilised crime incident reports archived by the law enforcement agencies and the emerging dynamics of offenders behaviour in today’s world, across different domains of life. However, with the increasing nature of crime data on a daily basis, due to the high crime committing trend, historical data available on repositories and heterogeneous sources of crime information by voluntary witnesses must be efficiently and maximally utilised for knowledge discovery in combating crime. Therefore, it makes sense to explore existing and unconventional approaches to analyzing crime data across different domains in knowledge-driven decision support.

This edited book, titled *Data Mining Trends and Applications in Criminal Science and Investigation*, emerges as a result of the crucial need for improving public safety in different domains and parts of the world. It is particularly aimed at resource constrained environments such as in developing nations, where crime is increasing at an alarming rate across different domains of life and impeding economic growth. By resource constrained environments, we mean scenarios where crime intelligence experts are limited and not enough technological solutions are put in
place to meet up with operational safety needs and citizens’ security. For example, where the quantity of assets or population of citizens’ to be protected significantly outweighs the available public safety or security personnel, it becomes very difficult to achieve and sustain a safe society or community. However, it is profoundly liberating to discover that the use of “smart statistics” (Milgram, 2013), which can be derived through crime data mining approaches could assist in making optimal use of these limited resources and consequently counter some of the safety challenges faced in resource constrained environments. In particular, of interest is the quest to understand the nature, scope and level of impact of existing crime mining solutions across different domains and develop new paradigms for a more inclusive solution. This will provide new insights that will assist to derive interventions to tackle emerging dynamic scenarios of criminal activities. Further, this book presents emerging issues, challenges and management strategies in public safety and crime control development across different domains. These domains encompass health-care, Internet-of-things and public safety.

**CHALLENGES AND OPPORTUNITIES**

Crime is an anti-social behaviour which costs our society greatly in several ways. Therefore, deterring crime is a top priority for decision makers, security and public agencies, in realising a sustainable “safe and smart” city. While in recent times there have been a vast amount of techniques and solutions emerging in a bid to tackle crime, these appear to be insufficient because despite the vast resources allocated to crime, people still fall victim of crime. Moreover, offenders and crime perpetrators are devising various means to subdue existing solutions, and as such constantly inventing dynamic ways to perpetuate crime without being suspected, apprehended or convicted of any crime. In addition, these crimes are committed across different domains of life, such as in health-care, public safety, business, information and communication technology, and needs to be tackled using a more proactive (tactical and strategic) approach.

The International Association of Crime Analysts (IACA) presents various ways by which crime patterns are conceptualized for tactical and/or strategic analysis (IACA, 2011), such as crime spree, crime series, and hot-spots, to mention a few. In the field of crime data analysis, the trend in crime spree, crime series, co-location or similarity crimes, where a set of offenders repeatedly and dynamically commit related crimes is still under exploration, among others. It is important to note that while these patterns are explained individually, they sometimes interconnect as clearly expressed and/or implied by the various content of this book. More generally, crimes are committed within a space-time relation. A proper understanding of
these crime patterns and criminal activities is crucial as it will help public safety and security agencies focus resources to achieve crime reduction and criminal apprehension targets. There are several advantages in using data mining techniques in criminal science and investigation and understanding crime situation using “smart statistics”. Some of the advantages include the fact that:

- Deriving reliable crime-situation information, through crime data mining approaches, will assist decision makers, public safety, and security agencies in knowledge driven decision support. For example, such information will assist them in distinguishing between normal and crime prone locations, identifying patrol boundaries, better attribution of past crime, apprehension of prime suspects and determination of mitigation priorities. A direct consequence of this would be the realisation of the optimal utilization of the limited available resources, where it matters most, across different domains to achieve safety and crime reduction targets.
- Timely and accurate knowledge of crime situation creates new insights into true patterns. This will help to reduce the prevalence of criminal activities and inform effective safety policies that need to be implemented, in order to subsequently deter crime.

Therefore, crime analysts, decision makers, public safety and security agencies alike owe it to themselves to understand the nature, scope and limitations of existing crime solutions and learn new ways to confront crime challenges in support of a more proactive law enforcement, citizen security, and intelligence operation. This book provides the opportunity to understand timely and effective intervention strategies and the deep applicability of data mining techniques in criminal sciences and investigation across different domains.

**ORGANIZATION OF THE BOOK**

This book presents 10 chapters contributed from 30 authors and co-authors from 6 countries and region: Canada, India, Iran, South Africa, Tunisia and USA. It includes quality chapters that presents scientific concepts, framework and ideas on data mining trends and applications in criminal science and investigations across different domains, such as public safety, health-care, networking, Internet-of-things, intrusion and fraud detection. The Editorial Advisory Board and expert reviewers have confirmed the high caliber of chapters through careful refereeing of the submitted papers. For the purpose of coherence, we have organised the chapters with respect to similarity of topic addressed. The topics addressed range from crime min-
ing issues pertaining to public safety improvements, hotspots and visual analytics, cyber-crime, intrusion detection, denial of service, juvenile abuse identification, suspect modeling and analysis of crime scenarios.

Section 1: Challenges and Existing Strategies in Public Safety and Crime Mining

Chapter 1, “On the Advancement of Using Data Mining for Crime Situation Recognition: A Comparative Review” by O. Isafiade, A. Bagula and S. Berman, presents details on existing crime data mining techniques and solutions, associated challenges and the exciting avenues that could be explored for future research. It presents the three most common strategies for exploring crime data mining, and discusses ways of integrating crime data sources for tactical crime analysis. This chapter pave the way for the discussion in Chapter 2 on the different research work that have been carried out in the field of criminal science and the kinds of knowledge support discovered from the various studies.

Chapter 2, by M. Goyal, V. Bhatnagar and A. Jain, presents a classification framework for criminal science and investigation and further presents a road map regarding the usage of various data mining tools and techniques. This chapter further elaborates on the application of data mining techniques in criminal science and investigation across different domains.

Section 2: HotSpot, Spatial, and Visual Analytics

In Chapter 3, C. Ku, A. Iriberri and G. K. Jena present a detailed information on visual analytics for crime analysis and decision support. They provide various visual representations that have been used to reveal the meaning of crime data, and visualization tools that have been used in the analysis of crime data, which allows four types of data analysis; these are statistical, textual, multimedia, and spatio-temporal. Example cases where these types of analysis have been used and the toolkits that support these analyses are also reported. Furthermore, the chapter reports on how the combination of visualization techniques with text and data mining seems to offer added benefits in crime data analysis.

E. Eftelioglu, S. Shekhar, and X. Tang present an important discussion and analysis in Chapter 4 on a computational perspective to crime hotspot analysis and detection. They focus on statistically significant hotspot detection techniques that aim to remove chance patterns, using spatial scan statistics. Three representative techniques for discovering significant crime hotspots are presented and each technique is illustrated using sample case studies on real crime datasets.
In Chapter 5, M. Ghaziasgar, N. De La Cruz, A. Bagula, and J. Connan present visual data mining as a great opportunity for criminal investigation. The chapter investigates the automatic detection of micro-expressions from a video feed using Local Binary Patterns (LBPs) as a facial feature descriptor and Support Vector Machines (SVMs) as a classifier into one of two categories, namely: “micro-expression” and “neutral expression”. Facial micro-expressions have been reported to be strong indicators of deception, giving rise to the possibility of detecting deception automatically. The chapter revisits the issue of deception detection by proposing visual data mining (VDM) as a non-invasive alternative to deception detection in next generation criminal justice.

Section 3: Forensics, Suspect Modeling, and Intelligence Gathering

O. Isafiade, A. Bagula and S. Berman provide an insightful discussion and analysis in Chapter 6 on the use of Bayesian Network (BN) in Crime Suspect Modeling and Legal Decision Support. The chapter describes the value of reasoning under uncertainty in legal decision making and evidence gathering. It describes how BN based inference could help to collect useful evidence about a crime scenario in the absence of domain experts and the process of sentiment-free candidate generation and rules extraction for offenders. It further reports on the performance of empirical analysis in the legal decision support process, in order to elucidate the practical relevance and challenges of using BN in the crime domain.

In Chapter 7, N. Ellouze, S. Rekhis and N. Boudriga present the issue of Forensic Investigation of digital crimes in the healthcare domain. This domain is of particular interest due to the fact that advances in Information and Communications Technologies have led to the release of a wide set of health-care applications and these applications exhibit significant security weaknesses, which include the use of weak authentication techniques and the lack of appropriate security mechanisms. The chapter focuses on postmortem investigation of crimes on healthcare applications. After classifying crimes targeting health-care applications, the requirements for the design of an appropriate postmortem investigation system are discussed. A literature review of proposals related to the investigation of crimes in healthcare applications together with a discussion of the advanced issues in this domain then follows.

Section 4: Denial of Service, Cyber-Crime, and Intrusion Detection Management

In Chapter 8, B. Fessi, Y. Djemaiel and N. Boudriga provides a useful discussion on data mining techniques to detect intrusion within dynamic environments and its
contribution in digital investigation. The chapter also presents different requirements that should be fulfilled to efficiently perform cyber-crime investigation based on data mining analytics.

M. Keyvanpour, M. Ebrahimi and N. G Nayebi, Olga Ormandjieva and Y. Suen, in Chapter 9, have capably considered improving public safety by providing a safe environment for juveniles and children in online social networks through the automatic identification of online predators. The prevalence of the online conversations have motivated the need for mitigating the undesirable effects of child abuse in cyber space. Two major on-line predator identification problems in which data mining plays an important role are:

1. Detecting predators, and
2. Visualizing and analyzing predator criminal networks.

The problem is casted as a combination of textual preprocessing in data or text mining and pattern classification in machine learning. The chapter further presents a discussion on classification algorithms in the domain of automated predator identification.

In Chapter 10, P. Machaka, A. McDonald and F. Nelwamondo have effectively perused state-of-the-art discussion on data mining techniques for distributed denial of service (DDoS) attack detection. The chapter elaborates on various techniques by which DoS attack is perpetuated and reports on the data mining techniques that are used to combat and detect these attacks, their advantages and disadvantages are explored. Furthermore, the characteristics and pervasiveness of DDoS attacks, alongside motives, mechanisms and techniques used to execute a DDoS attack, are discussed.

Omowunmi E. Isafiade
University of Cape Town, South Africa

Antoine B. Bagula
University of the Western Cape, South Africa
REFERENCES
