Ethical Data Mining
Applications for Socio-Economic Development

Hakikur Rahman
University of Minho, Portugal

Isabel Ramos
University of Minho, Portugal

A volume in the Advances in Data Mining and Database Management (ADMDM) Book Series
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Chapter Title</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preface</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Acknowledgment</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Section 1</strong></td>
<td><strong>Theoretical and Conceptual Approaches</strong></td>
</tr>
<tr>
<td>Chapter 1</td>
<td>Ethical Issues of 'Morality Mining': Moral Identity as a Focus of Data Mining</td>
</tr>
<tr>
<td></td>
<td>Markus Christen, University of Zurich, Switzerland &amp; University of Notre Dame, USA</td>
</tr>
<tr>
<td></td>
<td>Mark Alfano, Princeton University, USA &amp; University of Oregon, USA</td>
</tr>
<tr>
<td></td>
<td>Endre Bangerter, Bern University of Applied Sciences, Switzerland</td>
</tr>
<tr>
<td></td>
<td>Daniel Lapsley, University of Notre Dame, USA</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>Ethical Data Mining and Social Science Data Exploration and Description: Scope and Limitations in Social Science Research</td>
</tr>
<tr>
<td></td>
<td>Arabi U., Mangalore University, India</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Reuse of Excess Research Data for New Research</td>
</tr>
<tr>
<td></td>
<td>Amiram Porath, College for Academic Studies, Israel</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Data Mining Techniques to Improve Early Warning Systems across the Bay of Bengal: A Bangladesh Perspective</td>
</tr>
<tr>
<td></td>
<td>Hakikur Rahman, University of Minho, Portugal</td>
</tr>
<tr>
<td><strong>Section 2</strong></td>
<td><strong>Socio-Economic Development and Implications</strong></td>
</tr>
<tr>
<td>Chapter 5</td>
<td>Spidering Scripts for Opinion Monitoring</td>
</tr>
<tr>
<td></td>
<td>Antonella Capriello, University of Eastern Piedmont, Italy</td>
</tr>
<tr>
<td></td>
<td>Piercarlo Rossi, University of Eastern Piedmont, Italy</td>
</tr>
</tbody>
</table>
Chapter 6  
Big Data Dilemmas: The Theory and Practice of Ethical Big Data Mining for Socio-Economic Development  
*Debashis "Deb" Aikat, University of North Carolina at Chapel Hill, USA*

Chapter 7  
Social Science Data Analysis: The Ethical Imperative  
*Anthony Scime, The College at Brockport - State University of New York, USA*  
*Gregg R. Murray, Texas Tech University, USA*

Chapter 8  
Data Mining Algorithms for Measuring Performance Impact of Social Development Processes: Ethical Implications  
*Hakikur Rahman, University of Minho, Portugal*

Section 3  
Governance and Applications

Chapter 9  
A Customised Dataset to Assist Legal and Ethical Governance of Seaports  
*AnaXimena Halabi Echeverry, Macquarie University, Australia & Universidad de La Sabana, Colombia*  
*Deborah Richards, Macquarie University, Australia*

Chapter 10  
Ethical and Legal Data Mining: Paradigms of Organizational Development Practitioners and Organizational Psychologists  
*Ben Tran, Alliant International University, USA*

Chapter 11  
Measuring Student Learning Responsibly: A Learning Analytics Perspective with Web 2.0  
*Kam Hou Vat, University of Macau, Macau*

Chapter 12  
Improved Decision Support System to Develop a Public Policy to Reduce Dropout Rates for Four Minorities in a Society  
*Alberto Ochoa-Zezzatti, Juarez City University, Mexico*  
*Saul Gonzalez, Juarez City University, Mexico*  
*Fernando Monies, Juarez City University, Mexico*  
*Seyed Amin, Technological Teheran University, Iran*  
*Lourdes Margin, Polytechnic University of Aguascalientes, Mexico*  
*Guadalupe Gutierrez, Polytechnic University of Aguascalientes, Mexico*
Chapter 13
Digital Rights Management and Corporate Hegemony: Avenues for Reform
Nikhil Moro, University of North Texas, USA

Compilation of References

About the Contributors

Index
When data mining aims to disclose information about the moral competences and values of individuals or groups — an undertaking we call 'morality mining’ — novel ethical problems emerge. These are only partially covered by the current debate on ethical data mining focusing on privacy with respect to discrimination, threats to autonomy, misuse of data, and the consequences of erroneous information. An ethics of morality mining is of particular relevance for research in social science and psychology that increasingly relies on data emerging from social networks, media portals, etc., where people act from or at least in accordance with their own values. In this conceptual contribution, we outline the basic idea of morality mining, explain why we believe that morality mining is associated with novel ethical problems, and suggest ways to address these problems that could potentially help to resolve various socio-economic problems a society or community faces.

Data mining allows one to make crucial business or other similar decisions by extracting valid, previously unknown, comprehensible and actionable information from extremely large data bases. Such data analyses identify interesting trends, patterns automatically. As it uses mainly statistical techniques, under this process, it helps in identifying valid, non-trivial, novel, potentially useful data patterns. In this aspect, data mining is the process of discovering significant, valuable and interesting structures and relationships in large and complex volumes of data, especially in data-enriched areas of socio-economic domains and in this socio-economic aspect of a society, and thereby data mining applications essentially act as effective instruments for providing support for measuring socio-economic pattern in a society.
Although social and ethical matters are nowadays concerns to the society of which people are the only elements, in the days of technology innovations, computers are being manipulated with programs to act more like people, and eventually several social and ethical matters come into focus related to computer programming, or artificial intelligence. Hence, learning about ethical issues on data mining applications formed an important area of research interest. Researchers from almost all corners of social science discipline have found to be evaluating many questions, testing many hypotheses, or comparing many point estimates! In program evaluation, this arises, when comparing the impact of several different policy interventions; comparing the status of social indicators like test scores, poverty rates, teen pregnancy rates etc. across multiple schools, states, or countries; examining whether treatment effects vary meaningfully across different sub groups of the population; or investigating the impact of a program on many different outcomes. This chapter fits well in social science research as such information is highly useful in testing many of the hypotheses of economic or socio-economic in nature.

Chapter 3
Reuse of Excess Research Data for New Researches

Amiram Porath, College for Academic Studies, Israel

The chapter presents a specific niche dilemma regarding the ethical aspects of utilization of data gathered for one specific research for another research, which was not specified to the data suppliers (sample) at the time of gathering, or that the data suppliers were not even informed of the possibility of such occurrence. That dilemma of the reuse of research data stems from motives that are also (among others) rooted in the aim to increase the public good, and the dilemma is between potential benefits and potential harm. As the forces that create the dilemma are growing in concurrence with current trends in research and research financing, the dilemma commands some attention, even if it seems at first glance to be minor compared to issues related to business data mining and governmental data bases. The discussion ends with a possible solution, but the reader is encouraged to think about the dilemma and understand it rather than "solve" it. The novice can regard this as an introduction to the dilemma while the experienced researcher will view it as a summary. However, this sort of research need to be supported to uphold the ethical aspects of data mining and their various applications in the socio-economic development processes of a country, such as generic or specific researches, entrepreneurship development through innovation, and trading, commerce or e-governance through the utilization of innovative technologies.

Chapter 4
Data Mining Techniques to Improve Early Warning Systems across the Bay of Bengal: A Bangladesh Perspective

Hakikur Rahman, University of Minho, Portugal

This chapter is a conceptual contribution to this book on data mining applications upholding ethical issues related to two extremely important aspects of the Bangladeshi population: the early warning system and the disaster management system. The chapter tries to provide a few conceptual ideas to introduce ethical data mining application in these systems to support the agencies that are involved for an improved, efficient, and transparent support system in the country, especially across the Bay of Bengal. Resembling a triangular shape (deltaic), a major portion of the bay touches the southern portion of Bangladesh. Sediments from rivers have made the bay a shallow sea. Due to its shallowness and shape, monsoon rains and cyclone storms become destructive, causing great loss of life along the southern part of the country. Moreover, the three mighty rivers (Padma, Jamuna, and Meghna) form one of the largest river systems in the world. They have a large number of distributaries and tributaries, which cause a major portion of the country to be inundated by monsoon rain. In addition, being the lowest landing zone of the Himalayan
water, Bangladesh becomes victim to floods almost every year. Loss of lives, destruction of properties, suffering of numerous people and hampering of economic development have become part and parcel of Bangladeshi communities. This chapter suggests that the newly emerged data mining techniques can be introduced to collect, synthesize, analyze, archive, disseminate, and even make future forecasts forming a reliable early warning system across the Bay of Bengal.

Section 2
Socio-Economic Development and Implications

Chapter 5
Spidering Scripts for Opinion Monitoring

Antonella Capriello, University of Eastern Piedmont, Italy
Piercarlo Rossi, University of Eastern Piedmont, Italy

With the advent of Web 2.0 technologies, online forms of communication are rich sources of data to study socio-economic growth patterns and consumer behaviours. In this research field, the more robust development of data mining and opinion monitoring depends on fully automating data collection to monitor the evolution of customer opinions and preferences in real time. Although web crawlers or spiders can assist researchers in an innovative and effective way, this data collection approach could give rise to ethical concerns on the cost of web crawling processes and on data protection and privacy.

Chapter 6
Big Data Dilemmas: The Theory and Practice of Ethical Big Data Mining for Socio-Economic Development

Debashis "Deb" Aikat, University of North Carolina at Chapel Hill, USA

This chapter delineates the theory and practice of ethical big data mining for socio-economic development in four parts. The first part enunciates the ethical role of big data mining for socio-economic development by theorizing big data as a 20th Century phenomenon and its surging significance in the 21st Century digital era. The second part elucidates ethical values relating to big data mining with particular emphasis on the interplay of theory and practice. The third part connects classical theories of ethics to propose a code of conduct that relates to core ethical values such as privacy, confidentiality, objectivity, transparency, conflict of interest, and common good. The fourth and final part identifies privacy as a major challenge of ethical big data mining and postulates needed research directions. This chapter also features a list of additional reading and big data terms with concise definitions explicating their relevance to big data mining for socio-economic development.

Chapter 7
Social Science Data Analysis: The Ethical Imperative

Anthony Scime, The College at Brockport - State University of New York, USA
Gregg R. Murray, Texas Tech University, USA

Social scientists address some of the most pressing issues of society such as health and wellness, government processes and citizen reactions, individual and collective knowledge, working conditions and socio-economic processes, and societal peace and violence. In an effort to understand these and many other consequential issues, social scientists invest substantial resources to collect large quantities of data, much of which are not fully explored. This chapter proffers the argument that privacy protection and responsible use are not the only ethical considerations related to data mining social data. Given (1) the
substantial resources allocated and (2) the leverage these "big data" give on such weighty issues, this chapter suggests social scientists are ethically obligated to conduct comprehensive analysis of their data. Data mining techniques provide pertinent tools that are valuable for identifying attributes in large data sets that may be useful for addressing important issues in the social sciences. By using these comprehensive analytical processes, a researcher may discover a set of attributes that is useful for making behavioral predictions, validating social science theories, and creating rules for understanding behavior in social domains. Taken together, these attributes and values often present previously unknown knowledge that may have important applied and theoretical consequences for a domain, social scientific or otherwise. This chapter concludes with examples of important social problems studied using various data mining methodologies including ethical concerns.

Chapter 8
Data Mining Algorithms for Measuring Performance Impact of Social Development Processes: Ethical Implications

Hakikur Rahman, University of Minho, Portugal

This chapter has given focus on data mining applications and their utilizations in formulating performance measuring tools for social development initiatives upholding ethical issues. In this context, this chapter has provided justifications to include data mining algorithm to establish monitoring and evaluation tools for various social development applications. Specifically, this chapter gave in-depth analytical observations to establish knowledge centers (has been found successful in many regions as a catalyst of socio-economic growth) with various approaches and finally it put forward a few research issues and challenges to transform the contemporary human society into a knowledge society.

Section 3
Governance and Applications

Chapter 9
A Customised Dataset to Assist Legal and Ethical Governance of Seaports

Ana Ximena Halabi Echeverry, Macquarie University, Australia & Universidad de La Sabana, Colombia
Deborah Richards, Macquarie University, Australia

Attention to the legal and ethical principles of governance of seaport authorities (PAs) can enhance the future possibility of sustainable development of a port. This chapter presents a customised dataset and accompanying descriptions compiled from multiple sources and repositories that can be mined to provide adequate understanding over key decisional variables to assist the implementation of three Port State Control (PSC) mechanisms. Considerable care is given to the selection and combination of variables which may identify potentially serious accidents and the port's legal and ethical liabilities. The authors seek to clarify the relationship between the Corporate Social Responsibility (CSR) of PAs and, what is possibly the most important issue facing PAs nowadays, the issue of security. In order to validate the relationship between PSC and CSR, the authors suggest the use of the Regression Approach in Time Series Analysis (RATS) method that offers an assessment of mutual impacts of the PSC variables and a forecast of future values of CSR. RATS would enable PAs to be aware of the CSR challenges occurring among partner ports at least one time-step ahead. This may represent an important advance in using decision support systems to assist managers in performing complex analyses and making strategic choices.
Chapter 10
Ethical and Legal Data Mining: Paradigms of Organizational Development Practitioners and Organizational Psychologists

Ben Tran, AUiant International University, USA

In the process of data mining, techniques used, and results gathered, became ethically or even legally questionable due to the concern of the violation of individuals' privacy when specific information is obtained, manipulated, and disseminated by other entities without an individual's knowledge or consent. In other words, it is not the concern of what data is being mined, but more so regarding who is the miner of these data. This chapter is based on a review of the existing studies, which shows that not enough attention has been paid to the study of the "miner," as well as the ethical and legal qualifications of the miner, from the perspectives of organizational development practitioners (OD) and organizational psychologists (OP) in approaching for certifying or licensing. During the review, several cases are being included to justify the certifying or licensing approach, thus upholding ethical data mining, especially in relation to the socio-economic development of a community, or county. This research concludes with a few practical recommendations for both data miner(s) and entities that are involved with data mining.

Chapter 11
Measuring Student Learning Responsibly: A Learning Analytics Perspective with Web 2.0

Kam Hon Vat, University of Macau, Macau

This chapter investigates an ethical mechanism of organizational measurement for student learning that is based on the learning analytics gathered from various learning-related activities over an extended period of time. In the context of today's Web 2.0, such learning analytics are often collected from an electronic learning environment, such as a Web-based course management system (CMS), providing various tools of interest in scaffolding student learning: blogs, wikis, online forums, RSS, and many other innovative resources to facilitate learning online. This mechanism, intended to be ethically sound, could be considered as an instance of an accountability system typically installed in institutions of higher education and/or secondary schools, serving to gather evidence of student learning in a virtual learning environment involving electronic presence from both teachers and students in the context of learning development. It is understood that today's university as a higher education institution (HEI) must put in place such an accountability system to measure student college experience, as her sustained commitment to continuous improvement in the quality of student learning; yet, without the context of data analysis, the transformation of any existing accountability infrastructure in support of assessment for student learning could hardly be innovated effectively, especially regarding the productivity and coordination of its staff, both academic and administrative. The question is how innovatively a HEI could establish such an accountability system to measure and assess student learning responsibly by collecting, analyzing, and interpreting student learning analytics designed into their various learning activities.
Chapter 12
Improved Decision Support System to Develop a Public Policy to Reduce Dropout Rates for Four Minorities in a Society

Alberto Ochoa-Zezzatti, Juarez City University, Mexico
Saul Gonzalez, Juarez City University, Mexico
Fernando Montes, Juarez City University, Mexico
Seyed Amin, Technological Teheran University, Iran
Lourdes Margain, Polytechnic University of Aguascalientes, Mexico
Guadalupe Gutierrez, Polytechnic University of Aguascalientes, Mexico

This chapter proposes a decision support system applied to public schooling, especially for reducing dropout rates for minorities. That is relevant enough to enable understanding of ethical data mining in a strategic planning context. This understanding explains the importance of adequate different aspects related with Strategic Planning. The authors focus their analysis on a specific problem related with reducing dropout rates based on decision support systems under uncertainty. To this end, surveys are performed to gather information about this problem using Data Mining techniques to profile a number of behavioral patterns and choices that describe social behaviors. Ethical Data Mining is used for reasons of culture to improve the socio economic development. In addition, the chapter describes innovative models that capture salient variables of modernization, and how these variables give raise to intervening aspects that end up shaping behavioral patterns in ethical and social aspects. Finally, the chapter remarks and extends discussions of the authors' approach and will provide general guidelines for future work in diverse application domains, including further analysis on how those public politics organize and operate.

Chapter 13
Digital Rights Management and Corporate Hegemony: Avenues for Reform

Nikhil Moro, University of North Texas, USA

To the extent that data constitutes creative work in a tangible form, it represents intellectual property and implicates copyright. Understanding data mining, thus, invokes an appreciation of the principle and law of copyright. With the advent of digital rights management (DRM), the discourse of the U.S. law of copyright has shifted from fair use to circumventions of copyright-protecting software by hackers. In light of the ever-rising market power of transnational media corporations, this essay engages the political philosophy of Antonio Gramsci to address a normative question of who ought to control protected content and to what extent. It identifies six inflection points for any effort to divest copyright law of the disproportionate influence of corporate lobbying. Underlying the discussion is a presumption that reform of the law of copyright through dialogue among key stakeholders – creative individuals, technology innovators, legislators, policy makers, international mediators, data miners and other scholars, and transnational corporations – would help preempt any Gramsci-prescribed socialist reaction to the exclusive copyright regime enabled by DRM.