

Ethical Dilemmas in Data Mining and Warehousing

Joseph A. Cazier

Appalachian State University, USA

Ryan C. LaBrie

Seattle Pacific University, USA

INTRODUCTION

As we move into the twenty-first century, marketing is increasingly becoming a one-to-one affair. No longer are marketers satisfied with pools of potential customers extracted from mailing lists or government records. Instead they're aggressively seeking personalized information and creating computing systems that categorize individual consumers. (Garfinkel, 2000, p. 158)

Information has become a commodity, moving from place to place with little thought given to the ultimate value provided to consumers, business, and society. With advanced innovations in e-business, data warehousing, and data mining, it is becoming increasingly important to look at the impact of information use on individuals. This free flow of information within an organization and between different organizations can be both an asset and liability for consumers, businesses, and society.

As we have increasing privacy and risk concerns in the world today with identity theft, questionable marketing, data mining, and profiling, it is becoming increasingly important to explore how consumers feel and react to the use of their data. This study makes an important contribution to the literature by presenting common positive and negative myths surrounding these issues and exploring how ethical or unethical consumers believe these practices are by looking at the myths and their reaction to them. We focus on consumers' perceptions because at the end of the day it is what the consumers perceive to be happening that will determine their reaction. An ethical data practice is one that is believed to increase consumer, business, or societal value, and an unethical data practice is one which causes harm to these groups.

BACKGROUND

Ethical use of data within information systems has been investigated for a number of years. Mason (1986) suggested that there are four ethical issues in the information age, readily identified through the acronym PAPA. PAPA stands for privacy, accuracy, property, and accessibility. While in 1986, we did not believe that the sheer amount of data produced through the Internet and transactional e-business systems, including consumer personalization and the use of radio frequency identification (RFIDs), could have been foreseen, these four issues still remain the center of focus for computing ethics.

With respect to the introduction of new information technology and its affect on privacy, Tavani (1999) proposes two possibilities. Use of technology to collect information without the knowledge of the individual and use of technology for unintended purposes even when the individual has consented for other purposes both can have serious consequences to privacy. Tavani goes on to provide an excellent example of how a bank might have a privacy policy that protects an individual's information from being sold or transferred to other companies, but then abuses that information within the bank itself to the detriment of the customer.

In an *Information Week* research study on business ethics, Wilder and Soat (2001) discuss the ethics of data management. Among their findings are: (1) information technology professionals can no longer abstain from the ethics and privacy debate, and thus should undergo ethics and data security training; and (2) customer profiling is being led by the healthcare and financial services industries, with more than 80% of the companies surveyed in those fields participating in profiling, followed closely by business services,

Table 1. Data mining and warehousing practices myths (adapted from Cazier & LaBrie, 2003)

Value Level	Myth	Counter Myth
Consumer	<i>The merging of current customer data with secondary sources ultimately increases value for the consumer.</i>	<i>The merging of current customer data with secondary sources ultimately hurts the consumer.</i>
	<i>Customer profiling, leading to more customized service, creates consumer value.</i>	<i>Customer profiling, leading to more customized service, reduces consumer value.</i>
	<i>Using persuasive marketing techniques increases consumer value.</i>	<i>Using persuasive marketing techniques reduces consumer value.</i>
Business	<i>Data warehousing improves organizational productivity.</i>	<i>Data warehousing reduces organizational productivity.</i>
	<i>Data warehousing improves your organizational image.</i>	<i>Data warehousing hurts your organizational image.</i>
Societal	<i>Data warehousing reduces waste and helps the environment.</i>	<i>Data warehousing increases waste and harms the environment.</i>
	<i>Governmental use of data warehousing technologies is good for society.</i>	<i>Governmental use of data warehousing technologies is not good for society.</i>

manufacturing, and retailing, which average more than 70% of the companies participating in customer profiling.

This article explores the myths and perceptions that consumers have regarding the ethics and usefulness of data mining and warehousing activities. These myth/counter myth pairs were first introduced in Cazier and LaBrie (2003). This study extends that research by attempting to quantify whether or not what was proposed in Cazier and LaBrie (2003) is valid. That is, are these myths truly perceived by the consumer or not? These myths are broken up into three value classifications: consumer, business, and societal. Table 1 presents a summation of those myths.

To better understand public reaction to the increase in data analysis and information flow, this research presents findings from a survey (N=121) that presented the arguments for both sides of seven misconceptions of the use (or misuse) of data in data warehouses in a myth/counter myth fashion. Respondents are asked how they believe they are affected in value propositions for consumers, business, and society. Because there are elements of truth in both sides of the arguments, we present the results of how much users agree or disagree with each myth and then ask them to choose which they believe is more likely. It is possible to agree or disagree with both myth and counter myth by examining different aspects of the impact, hence we also ask respondents to choose which they believe is more likely to be true. Respondents were chosen from

a cross-section of the population of three universities in different regions to maximize generalizability.

Investigation of Consumer Value

It is generally accepted that successful data warehousing projects can benefit the businesses that implements them. Less recognized however are the benefits to the consumer. This section describes three myths on how data warehousing may positively or negatively affect the consumer. Results from the survey follow each myth.

Myth 1: *The merging of current customer data with secondary sources ultimately increases value for the consumer.* Myth 1 suggests that by acquiring data from secondary sources about customers, and adding it to their existing data, businesses can better understand their customers and more accurately target their needs. The collection and merging of this data is a good business practice that helps the customer find products and services they desire. Customer satisfaction is increased because their service is more customized and tailored to their needs and desires. Offers that they receive from companies are more likely to be things they are interested in. Our respondents were split on this issue, with 45% agreeing or strongly agreeing with this myth, 39% disagreeing, and 16% undecided.

6 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/ethical-dilemmas-data-mining-warehousing/13476

Related Content

Privacy, Security, and Identity Theft Protection: Advances and Trends

Guillermo A. Francia III, Frances Shannon Hutchinson and Xavier Paris Francia (2017). *Identity Theft: Breakthroughs in Research and Practice* (pp. 1-12).

www.irma-international.org/chapter/privacy-security-and-identity-theft-protection/167217

Authentication Through Elliptic Curve Cryptography (ECC) Technique in WMN

Geetanjali Rathee and Hemraj Saini (2018). *International Journal of Information Security and Privacy* (pp. 42-52).

www.irma-international.org/article/authentication-through-elliptic-curve-cryptography-ecc-technique-in-wmn/190855

Secure and Private Service Discovery in Pervasive Computing Environments

Feng Zhu and Wei Zhu (2011). *Applied Cryptography for Cyber Security and Defense: Information Encryption and Cyphering* (pp. 229-243).

www.irma-international.org/chapter/secure-private-service-discovery-pervasive/46245

An Ensemble Approach for Feature Selection and Classification in Intrusion Detection Using Extra-Tree Algorithm

Ankit Rajeshkumar Kharwar and Devendra V. Thakor (2022). *International Journal of Information Security and Privacy* (pp. 1-21).

www.irma-international.org/article/an-ensemble-approach-for-feature-selection-and-classification-in-intrusion-detection-using-extra-tree-algorithm/285019

Privacy and Security in E-Learning

Khalil El-Khatib, Larry Korba, Yuefei Xu and George Yee (2008). *Information Security and Ethics: Concepts, Methodologies, Tools, and Applications* (pp. 299-315).

www.irma-international.org/chapter/privacy-security-learning/23094