Privacy in Data Mining Textbooks

James Lawler
Pace University, USA

John C. Molluzzo
Pace University, USA

INTRODUCTION

Many companies, such as Wal-Mart, store much of their business and customer data in large databases called data warehouses. Their customers are not told the extent of the information accumulated on them, how long it will be kept, nor the uses to which it will be put (Hays, 2004). This data is subsequently analyzed to produce new information to help the companies evaluate business processes and customer behavior. Data mining is usually used to do the analysis. Much of the mined data is public or semi-public—what we purchase at the supermarket, where we surf the Web, where we work, our salary.

The key ethical issues in mining personal data are that people: (1) are generally not aware their personal information is being gathered, (2) do not know to what use the data will be made, or (3) have not consented to such collecting or use.

In a survey of twenty Web data mining professionals, van Wel and Royakkers (2004) showed that the professionals prefer to focus on the advantages of Web data mining instead of discussing its possible dangers. These professionals argued that Web data mining does not threaten privacy.

One might wonder why professionals are not aware of or concerned over the possible misuse of their work, and the possible harm it might cause to individuals and society. Part of the reason might lie in the content of the data mining courses they have taken and in the textbooks they used to learn their craft. The purpose of this article is to analyze the content of contemporary data mining textbooks to determine the extent to which they introduce and discuss issues relating to privacy of consumer data, laws that govern the use of personal consumer data, and professional guidelines for the collection and use of consumer data.

BACKGROUND

Privacy

Privacy is not easily defined, perhaps because the notion of privacy has evolved over time and now means different things in different situations and in different cultures. This article focuses on the effects of data mining on *informational privacy* (Tavani, 2004), which is a person's ability to restrict access to and control the flow of his or her private information. Much of modern informational privacy theory is grounded on Moor's (1997) *control/restricted access* theory of privacy, in which a person has privacy in a situation if the person is protected from intrusion, interference, and information access by others.

Laws

There is no explicit right to privacy in the U.S. Constitution. However, legislation and court decisions on privacy are usually based on parts of the First, Fourth, Fifth, and Fourteenth Amendments. Most of the laws in the United States govern what the federal government can do with personal data. Except for healthcare and financial organizations, and data collected from children, there is no law that governs the collection and use of personal data by commercial enterprises. Therefore, each organization decides how it will use the personal data it has accumulated on its customers.

Privacy Guidelines

Although there are few laws in the United States governing the use of personal data, many businesses have used the Code of Fair Information practices of the Organization for Economic Cooperation and Development

Table 1. Number of books in each construct/rank

		Rank					
	Construct	5	4	3	2	1	0
Business and Consumer Ethics							
Ethics Codes	C1-1	2	6	1	1	1	18
Definitions of Privacy	C1-2	0	1	1	2	0	25
Functions of Privacy	C1-3	0	1	0	0	0	28
Personal vs. Group Privacy	C1-4	0	2	1	1	0	25
Studies of Privacy	C1-5	0	1	1	2	0	25
Subtotal		2	11	4	6	1	121
Government and Organizations							
Constitution	C2-1	0	0	0	0	0	29
Court Cases	C2-2	0	0	0	1	0	28
Federal Legislation	C2-3	2	0	1	2	3	21
State Legislation	C2-4	0	1	1	2	3	22
Authorities	C2-5	0	0	3	4	0	22
Organizations	C2-6	0	1	2	1	1	24
Subtotal		2	2	7	10	7	146
Managerial and Methodological							
Chief Privacy Officer	C3-1	0	1	2	2	4	20
Personal Privacy Policy Standards	C3-2	0	2	5	3	2	17
Personalization Techniques	C3-3	2	7	0	1	1	18
Privacy Systems	C3-4	0	3	0	1	1	24
Protection of Systems	C3-5	3	2	5	5	4	10
Subtotal		5	15	12	12	12	89
Pedagogical							
Privacy Studies	C4-1	0	0	0	0	0	29
Privacy Publications	C4-2	0	2	3	1	1	22
Privacy Conferences	C4-3	0	0	1	0	0	28
Scholarly Journals	C4-4	0	1	0	1	1	26
Privacy Groups	C4-5	0	0	0	0	4	25
Subtotal		0	3	4	2	6	130
Technological							
Digital Rights Management	C5-1	0	0	0	0	0	29
Platform for Privacy Preferences	C5-2	0	2	0	1	0	26
Privacy Aware Technology	C5-3	0	0	0	0	0	29
Privacy Invasive Technology	C5-4	0	0	0	0	0	29
Privacy Software Technology	C5-5	0	0	0	0	0	29
Subtotal		0	2	0	1	0	142
Total		9	33	27	31	26	628