



Chapter XVIII

Social, Ethical and Legal Issues of Data Mining

Jack S. Cook

Rochester Institute of Technology, USA

Laura L. Cook

State University of New York at Geneseo, USA

ABSTRACT

This chapter highlights both the positive and negative aspects of Data Mining (DM). Specifically, the social, ethical, and legal implications of DM are examined through recent case law, current public opinion, and small industry-specific examples. There are many issues concerning this topic. Therefore, the purpose of this chapter is to expose the reader to some of the more interesting ones and provide insight into how information systems (IS) professionals and businesses may protect themselves from the negative ramifications associated with improper use of data. The more experience with and exposure to social, ethical, and legal concerns with respect to DM, the better prepared you will be to prevent trouble down the road.

INTRODUCTION

What price are you willing to pay for convenience and personalization? Customers expect businesses to not only meet their needs but also anticipate them. Companies entice consumers to sing an elaborate song detailing their personal lives with offers of giveaways, discounts, and better service, while companies gladly listen and record the tune. With technological advances, companies regularly employ data-mining techniques to explore the contents of data warehouses looking for trends, relationships, and outcomes to enhance their overall operations and discover new patterns that theoretically allow companies to better serve their customers. The uses of data mining are

numerous and implemented by many organizations including government agencies and not-for-profit organizations. However, with this great ability, concerns exist regarding social, ethical, and legal issues associated with data mining. Unfortunately, these issues are not fully defined, and the future of data mining is uncertain with the threat of regulation looming.

Data has been gathered and recorded for thousands of years and used to manage day-to-day operations. For a long time, financial and insurance companies have mined their data to detect patterns of fraudulent credit card usage, find hidden correlations between financial indicators, identify behavior patterns of risky customers, and analyze claims. Utility companies have used it for decades to predict when a generator might fail. Data mining is not new, but widespread interest in it is a recent phenomenon.

Until recently, data sets were small in size, typically containing fewer than ten variables (Fayyad, 2001). Data analysis traditionally revolved around graphs, charts, and tables. But the real-time collection of data, based on thousands of variables, is practically impossible for anyone to analyze today without the aid of information systems. With such aid however, the amount of information you can “mine” is astonishing. Organizations have only recently begun to examine how their vast collections of data can be mined strategically. Data mining extracts value from volume.

Numerous definitions of data mining exist. One definition is that it is the “mechanized process of identifying and discovering useful structures in data” (Fayyad, 2001, p. 62). In this context, “structure” refers to patterns, models, and relationships in the data. Data-mining techniques draw upon such diverse areas as probability theory, information theory, estimation, uncertainty, graph theory, and database techniques (Fayyad, 2001). In addition, artificial intelligence techniques such as neural networks, expert systems, and classification algorithms are used. Another definition states that data mining involves extracting hidden predictive information from databases to solve business problems (Brandel, 2001). Often the term is misused to describe new ways to present data. Data mining does more than just present existing data in new ways—it facilitates the discovery of previously unknown relationships among the data. To further illustrate the point, consider most standard database operations. These operations present results that users already intuitively knew existed in the database. Data mining extracts information from the database the user did not anticipate. Data mining creates information that can be leveraged by the organization to create a competitive advantage. However, it is just as likely to identify meaningless patterns or trends, wasting time and resources.

This chapter highlights both the positive and negative aspects of data mining. Specifically, the social, ethical, and legal implications of data mining are examined through recent case law, current public opinion, and small industry-specific examples. There are many issues concerning this topic. Therefore, the purpose of this chapter is to expose the reader to some of the more interesting ones and provide insight into how information systems (IS) professionals and businesses may protect themselves from the negative ramifications associated with improper use of data. The more experience with and exposure to social, ethical, and legal concerns with respect to data mining, the better prepared you will be to prevent trouble down the road. Before examining these issues, it is helpful to understand how data is gathered and what can be inferred from it.

24 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/social-ethical-legal-issues-data/7611

Related Content

Towards a Sentiment Analysis Model Based on Semantic Relation Analysis

Thien Khai Tran and Tuoi Thi Phan (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 290-313).

www.irma-international.org/chapter/towards-a-sentiment-analysis-model-based-on-semantic-relation-analysis/308493

A Taxonomy for Distance-Based Spatial Join Queries

Lingxiao Li and David Tanir (2017). *International Journal of Data Warehousing and Mining* (pp. 1-24).

www.irma-international.org/article/a-taxonomy-for-distance-based-spatial-join-queries/185656

Clustering of COVID-19 Multi-Time Series-Based K-Means and PCA With Forecasting

Sundus Naji Alaziz, Bakr Albayati, Abd al-Aziz H. El-Bagoury and Wasswa Shafik (2023). *International Journal of Data Warehousing and Mining* (pp. 1-25).

www.irma-international.org/article/clustering-of-covid-19-multi-time-series-based-k-means-and-pca-with-forecasting/317374

Big Data – Small World: Materializing Digital Information for Discourse and Cognition

Ian Gwilt (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 288-301).

www.irma-international.org/chapter/big-data--small-world/150170

Routing Attribute Data Mining Based on Rough Set Theory

Yanbing Liu, Shixin Sun, Menghao Wang and Hong Tang (2006). *International Journal of Data Warehousing and Mining* (pp. 27-41).

www.irma-international.org/article/routing-attribute-data-mining-based/1769