

THE IDENTIFICATION OF ETHICAL FOCUS AREAS: A LITERATURE STUDY INTO DATA MINING ETHICAL FOCUS AREAS

ANNEMAE VAN DE HOEF, SAM LEEWIS, MATTHIJS BERKHOUT & KOEN SMIT

HU University of Applied Sciences, Digital Ethics, Utrecht, The Netherlands.
E-mail: annemae.vandehoef@hu.nl, sam.leewis@hu.nl, matthijs.berkhout@hu.nl, koen.smit@hu.nl

Abstract Improper use of data must be avoided, as the consequences of improper use of data can be catastrophic. In the design of information systems, ethical focus areas could help combat improper use of data. Currently, more research is available on ethical focus areas in Data Mining compared to related research fields of Data Mining, such as Decision Mining and Process Mining. For this paper, a theoretical review was conducted to identify ethical focus areas of Data Mining and their possible solutions. Seven ethical focus areas were identified focussing on privacy, collection of personal information, consent, unpredictability and inaccuracy, group profiling and biased data. Future research is needed on the ethical focus areas, to validate the possible solutions related to these ethical focus areas in the context of related research fields of Data Mining.

Keywords:

data mining, ethics, ethical focus areas, decision mining.

1 Introduction

The value of data is determined by the techniques used to understand the data based on the context the data comes from or is placed in (Calders & Custers, 2013). Credit card fraud (Bhattacharyya et al., 2011), heart disease (Purusothaman & Krishnakumari, 2015) and terrorism (Thuraisingham, 2020) are all examples of a variety of serious (societal) problems for which Data Mining offers a solution. Data Mining is a technique that consists of machine learning, statistics, visualization (Calders & Custers, 2013; Wahlstrom et al., 2006), databases (Wahlstrom et al., 2006) and artificial intelligence (Calders & Custers, 2013). Data Mining is defined as “*the automated process of analyzing large sets of data and then extracting patterns in the data.*” (Kulkarni et al., 2010, p. 1). It is known for the benefits it offers to public and private sectors (Schermer, 2011). For example, Data Mining can benefit consumers when their information is used to take up less of the consumer’s time by rearranging a store based on products purchased together (Payne & Trumbach, 2009). In addition to benefits, Data Mining also has drawbacks that raise ethical focus areas regarding privacy through the improper use of its data processing methods (Schermer, 2011). For example, serious problems such as identity theft and discrimination occur when consumers provide personal information (Payne & Trumbach, 2009). This paper discusses the ethical focus areas present in the Data Mining research field.

Multiple research fields address ethical focus areas. For example, AI (Brendal et al., 2021) and databases (Chadwick & Berg, 2001). AI and AI ethical focus areas can be considered too broad for the specific application of Data Mining, where each research field has their overlapping, but also their specific solutions. To reveal these specific ethical focus areas and solutions, we, therefore, solely focus on Data Mining in this study. With the goal to ultimately design and deploy ethical Data Mining techniques. This raises the following research question: “*What ethical focus areas could be retrieved from Data Mining research that can be used in sub-research fields of Data Mining?*” Sub-research fields of Data Mining include Decision Mining and Process Mining. Identifying focus areas may also be relevant to these fields. The remainder of the paper is structured as follows: first, the background and related work that provides further insight into the topic is being discussed. Second, the research method used to collect and analyze data is presented. Third, the use of a theoretical review as a data collection technique is detailed. Fourth, the results of the theoretical review are

defined. Fifth and lastly, the research is concluded, followed by a discussion and possible future research directions.

2 Background and Related Work

The growing world of technology offers many opportunities when it comes to processing data with Data Mining solutions (Pal, 2011). Data Mining is a method aimed at discovering useful information and data patterns that emerge from a collection of analyzed data (Kulkarni et al., 2010; Pal, 2011; Payne & Trumbach, 2009). It analyzes information from a database (Payne & Landry, 2012) to make discoveries about this information (Cary et al., 2003). Data Mining is also referred to as knowledge discovery because it is not always apparent what information is being discovered (Cary et al., 2003). The analyzing process is performed using analytical tools and techniques (Pal, 2011). The most common applications that Data Mining offers are: security, efficiency, product innovation and customer service (Payne & Trumbach, 2009). For example, it can help with marketing that is more targeted to specific consumers (Cary et al., 2003) or cleaning data before using it for a different purpose (Payne & Landry, 2012). The use of Data Mining can create ethical problems around, for example, the violation of consumers' privacy since personal information is involved (Cary et al., 2003). Thus, despite the benefits of Data Mining, it certainly presents ethical challenges (Johnson, 2014, 2018).

When research extends on existing solutions to tackle new problems, according to research by Hevner and Gregor (2013), these solutions can be adopted from any related research field. Related fields to Data Mining are e.g. Process Mining (van der Aalst, 2011) or Decision Mining (Leewis, Smit, et al., 2020; Rozinat & Aalst, 2006). Process Mining focuses on the discovery of processes from event logs, conformance checking of process models, and the enhancement of process models (van der Aalst, 2011). Decision Mining focuses on the discovery of decisions from decisions logs, conformance checking of decision models, and the improvement of decision models (De Smedt et al., 2017; Leewis, Berkhout, et al., 2020; Leewis, Smit, et al., 2020). Both Process Mining and Decision Mining utilize Data Mining techniques (Leewis, Smit, et al., 2020). Therefore, the authors argue that any ethical focus area identified with regards to Data Mining can be adopted in the related fields of Data Mining and vice versa.

3 Research Methods

The purpose of this paper is to explore what ethical focus areas of Data Mining can be identified in literature. Therefore, the focus of this research should be on existing research related to this topic. To achieve this, a theoretical review is conducted (Paré et al., 2015). A theoretical review looks at an emerging problem and that the research in question should address this problem or support it with new research based on existing research (Paré et al., 2015). This will be done based on other scientific papers that contain ethical focus areas about Data Mining. The collection and analysis of these scientific papers is done using a protocol, as shown in figure 1.

4 Data Collection and Analysis

Over a period of four months, between November 2021 and February 2022, a theoretical review was conducted where scientific papers were found using Google Scholar in line with the systematic review approach and phases described by Snyder (2019). Franceschet (2010) recommends using Google Scholar when it comes to topics related to computer science. Google Scholar has a broader selection of cited papers on a wider range of topics than, for example, Web of Science has (Franceschet, 2010; Harzing & Alakangas, 2016). Two search queries are utilized for retrieving articles related to ethical focus areas in data mining. The first search query is that of *ethics in Data Mining*. This includes the possible publications concerning ethical issues in Data Mining. The second search query is that of *ethical Data Mining*. This covers the possible publications related to the field of ethical Data Mining, which is a slight nuance compared to the first query. Applying these queries resulted in the following number of results using Google Scholar collected within the first two weeks of November 2021: 45 papers were found with the search term *ethics in Data Mining* and 104 papers were found with the search term *ethical Data Mining*. Papers were excluded on being non-English, duplicate papers, and being anything other than a book, conference proceeding or journal paper. This resulted in a set of 64 papers (21 *ethics in Data Mining* and 43 *ethical Data Mining*), as shown in Figure 1. The papers were then reviewed by three reviewers based on title, abstract, and full paper review rounds. The reviewers did this separately for all three rounds. After a review, the results were discussed to decide for each paper whether it will be included or not. This resulted in 27 relevant papers after the title review, 20 relevant papers after the abstract review and 19 relevant papers after the full paper review.

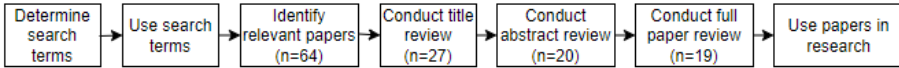


Figure 1: Review Protocol

All resulting papers were reviewed against the selection criteria using two questions: 1) The paper is published in a conference proceedings, book or journal, is English and not duplicate. 2) The paper addresses the combination of ethics and Data Mining. The first question was asked during the identification of papers and the second during all review rounds. The coding of the ethical focus areas was performed by one researcher and by fully analyzing each paper, a general topic (ethical focus area) could be identified, thus coding the article under that specific topic. An additional researcher was consulted when there was doubt about which ethical focus area could be identified. Each article was also analyzed for possible solutions to combat problems related to a specific ethical focus area.

5 Results

The theoretical review resulted in 19 papers that were deemed relevant for this research dating from 2002 to 2021. After analyzing the papers, a number of ethical focus areas could be identified. These ethical focus areas are: 1) privacy, 2) collection of personal information, 3) consent, 4) Data Mining unpredictability, 5) Data Mining inaccuracy, 6) group profiling, and 7) biased data. In addition, most of the identified ethical focus areas are presented with possible solutions, which were also identified from the analyzed 19 papers.

5.1 Area 1: Privacy

The first ethical focus area is that of privacy. Through the use of the internet and the digital economy, we create digital data structures that include information about our work, education, health, and entertainment, and these data structures are stored in various places such as hotels, airlines, and banks. However, these data structures can also be used to find out more personal information about an individual. The method that makes this possible is Data Mining and this raises a number of ethical privacy concerns when used by business organizations and government agencies (Shahidullah & Hossain, 2009). Moreover, within the U.S., for example, there is not

yet a legal framework available for protecting the privacy of individuals when Data Mining is used by the businesses or governments (Sprague, 2009).

The use of Data Mining processes in organizations often seems to benefit an individual, but the individual's personal information is not always kept private (Robinson, 2015). Although individuals fear potentially losing their privacy, they still often want the benefits Data Mining offers (Dean et al., 2016). For example, benefits such as targeted marketing, discounts, money and time savings. However, if a consumer wants to gain benefits within an organization, he or she must often give away personal information (Payne & Trumbach, 2009). In addition, organizations and websites can track a lot of an individual's online behavior (Cook & Cook, 2002; Robinson, 2015). This could be dangerous and harmful when this information is disclosed to unauthorized individuals. For example, it could be harmful when you or someone close to you has a terminal illness and you start to receive information about questionable cures (Cook & Cook, 2002).

There are also concerns about privacy violations within U.S. Data Mining programs. After the terrorist attack on the New York World Trade Center on September 11 2001, there has been an increase in the establishment of federal Data Mining programs. Privacy concerns also increased with the use of personal information within these Data Mining programs, such as emails, bank accounts, or social security numbers (Shahidullah & Hossain, 2009).

5.2 Area 1: Possible Solutions

Individuals will continue to use the internet even though there is a threat of organizations misusing personal information through Data Mining processes. Therefore, it is important that there are more built-in privacy protections for the internet as this is not available now. This will ultimately benefit both individuals and organizations (Robinson, 2015). Unfortunately, an individuals' behavior on websites will not always be completely anonymous, but it is important that organizations provide options, such as Virtual Private Networks (VPNs), not being tracked and blocking cookies with the use of a program to allow an individual to act as anonymously as possible (Robinson, 2015). After all, it is in the interest of an organization to engage in proper privacy processing. If they make a mistake in this regard, it is not only bad for the individuals involved, but also for the image of the

organization (Cook & Cook, 2002). Robinson (2015) used a Rawlsian ethical approach to analyze the ethical implications of Data Mining of three major codes of ethics used by American marketing firms. Rawls's ethics focus on among others the protection of the least advantaged in society and that anonymity should be a primary good for individuals as a form of self-protection on the internet. Robinson concluded that the code of ethics needs to be reviewed and updated and that there needs to be more clarity on what digital ownership means within the codes of ethics. A definition like this can then be used by organizations when making choices about the use of an individual's personal information. In addition, consumers should be allowed to make more informed choices as organizations are more open about the use of personal information through clear policies on how an individual's personal information is used. Despite the difficulty of achieving this, the goal is for consumers to have more control over their personal information (Robinson, 2015). Dean et al. (2016) also indicate that it is good for a business to establish a code of ethics because, among other things, it positively influences the relationship between the business and its external stakeholders. Lastly, Oliveira and Zaïane (2009) propose a privacy-preserving clustering method called Dimensionality Reduction-Based Transformation (DRBT). This method transforms a database so that the objects within the dataset retain the similarity but no longer resemble the original database. This allows the database to be used for clustering while maintaining privacy, accuracy of the process, and also allows a dataset to be securely shared between organizations at a reduced cost. Violating an individual's privacy stems from multiple problems, such as the unethical collection of personal information and the deprivation of consent.

5.3 Area 2 and 3: Collection of Personal Information and Consent

The second and third ethical focus areas are the collection of personal information and consent. All stages of Data Mining raise ethical concerns related to personal information, especially the collection of the data prior to the Data Mining process (Cary et al., 2003). In criminological articles, for example, ethical considerations around data collection are often not addressed by the researcher (Brewer et al., 2021). Organizations are increasingly collecting personal information without the average individual being aware and this is becoming more common. The collection of this personal information is done in a variety of ways, such as through cookies or registration forms (Cary et al., 2003; Payne & Landry, 2012). Information can also

be collected by using web crawlers, which are programs that download web pages from the internet. This brings problems, such as higher costs for website owners due to the bandwidth used for websites during web crawling or the creation of lists that contain emails for spam (Thelwall & Stuart, 2006). Individuals often fail to see the risk they put themselves at when they choose to share personal information for benefits within organizations. These can be risks related to identity theft, phishing schemes, mail fraud, or discrimination (Robinson, 2015). Furthermore, the danger of collecting personal information is that the individuals about whom the information is about have no control over their information, but they are the ones who are disadvantaged by the inaccuracy of Data Mining (Payne & Landry, 2012; Payne & Trumbach, 2009). However, an economic justification for collecting personal information would be that it allows organizations to communicate with individuals in a more targeted way (Fogel, 2006).

Personal information may be used without informing the individual concerned (Cary et al., 2003; Payne & Landry, 2012; Robinson, 2015). Not to mention, information someone puts online can be used for something completely different than its original purpose (Cook & Cook, 2002; Danielson, 2009; Payne & Landry, 2012). The fact that individuals are not always well informed by organizations about the use or consequences of their personal information deprives them of the opportunity to give informed consent (Payne & Landry, 2012). It cannot be assumed that an individual knows exactly what they are consenting to when they give an organization permission to collect personal information (Fogel, 2006). Additionally, it is a major problem when an individual's personal information, which he or she has chosen to share online on a particular organization's website, is used by multiple organizations without their knowledge or consent (Robinson, 2015). Also, organizations often assume that an individual gives consent when he or she uses the organization's services (Cary et al., 2003). In the U.S., for example, consent is deemed to have been given for linking an individual's profile to the telephone number disclosed by a toll-free service when individuals use such a toll-free number (Danielson, 2009). Lastly, when an individual does give consent, organizations do not give an individual the opportunity to withdraw that consent (Cary et al., 2003). Another problem are the privacy policies that individuals have to agree to before they can make use of a service. These are difficult to read, understand or not even followed by the organization (Cary et al., 2003). Implementing a proper privacy policy is already a

challenge because the concepts surrounding privacy are complex in themselves (Dean et al., 2016).

5.4 Area 2 and 3: Possible Solutions

Using Data Mining in an ethical manner works by collecting data in an honest way and informing the individual that he or she may, or may not, be monitored (Wowczko, 2013). When data is collected through the use of web crawlers, it is important that privacy is considered without making assumptions that web page owners will consider this in their interest (Thelwall & Stuart, 2006). In addition, it is important for organizations to be more open about collecting personal information and how they use that information to benefit individuals (Fogel, 2006). An organization could implement a Data Mining policy that focuses on collecting information, but there are often drawbacks to this. For example, a policy that optimizes consequences for those who are affected by Data Mining is hard to achieve because it is difficult to map out all the consequences. Further, if you are dealing with a policy that prohibits Data Mining processes if they violate privacy you are again dealing with drawbacks such as very clear communication and constantly asking individuals for permission (Cary et al., 2003). Dean et al. (2016) suggest a few practices for establishing online privacy policies for businesses, such as letting an individual know that their data is being collected or what Data Mining practices are being used within the business. Furthermore, Payne and Landry (2012) mention that personal information that is collected should be secured by the individual that collects it. Organizations should use security measures, such as data encryption, that prevent the use of personal information without consent or anonymity tools. Finally, if researchers do not wish to seek consent from individuals, it is important that they can prove that the research, including the collection of data and reporting of results, will have virtually no negative impact on these individuals (Brewer et al., 2021). Problems surrounding Data Mining do not only affect the privacy violation of an individual, but the Data Mining process itself also plays a role in this.

5.5 Area 4 and 5: The Unpredictability and Inaccuracy of Data Mining

The fourth and fifth ethical focus areas are unpredictability and inaccuracy of Data Mining. Decisions in our society that are based on Data Mining do not sufficiently consider the privacy of an individual. When making a decision, more thought is given

to avoiding costs and making a profit. This can be problematic when Data Mining uses sensitive datasets, such as medical datasets (Roddick & Fule, 2004). A counterargument to this may be that Data Mining is unpredictable (Roddick & Fule, 2004). It is known for discovering unknown relationships between data (Cary et al., 2003; Cook & Cook, 2002; Johnson, 2018; Payne & Landry, 2012; Shahidullah & Hossain, 2009). Therefore, it is not known in advance what kind of information will be found in the Data Mining process (Cary et al., 2003; Roddick & Fule, 2004). But due to the unpredictability of Data Mining, problems still arise such as that it is difficult for an organization to properly provide the opportunity for informed consent. As well as, that the unpredictability of Data Mining also reveals information about an individual that he or she does not want to disclose and when this information is spread it could have harmful consequences if it is revealed to the wrong kind of people (Payne & Landry, 2012; Payne & Trumbach, 2009). When it comes to sensitive data about an individual in health care databases, such as life expectancy or drug use, these types of people may base decisions about mortgages, for example, on this type of information (Gál et al., 2014). When criminals can obtain more personal information about a person by using Data Mining. This can lead to unpleasant consequences, such as identity theft (Christen et al., 2014).

In addition, the Data Mining process is not perfect which can cause mistakes (Cook & Cook, 2002). These mistakes happen because e.g., sensitive information often has an expiration date or the method of data collection is not accurate (Cary et al., 2003). For example, information about the same person is considered two different people or information is linked to the wrong person, which makes the information used in a Data Mining process inaccurate (Cook & Cook, 2002). Individuals may be judged based on incorrect information and they do not have the option to change their personal information (Cary et al., 2003). Another problem is that the relationships that emerge from Data Mining are often perceived for the fact that they are probably correct (Johnson, 2014, 2018) or that organizations assume Data Mining is very accurate when this is not the case (Payne & Trumbach, 2009). This can lead to unnecessary actions that have no consequences (Johnson, 2014, 2018). Lastly, it is important that the results of Data Mining are not influenced by the individual who allows the Data Mining process to take place or the individual who provides the data (Wowczko, 2013).

5.6 Area 4 and 5: Possible Solutions

Roddick and Fule (2004) present a system that is used to filter the results of Data Mining for sensitivity before these results are used in a decision-making process. The system takes into account data sensitivity by checking the rules of Data Mining. They consider this to be a partial solution for addressing privacy protection, stereotyping and the use of sensitive data. Data anonymization methods can also be used to prevent re-identification of an individual when sensitive data is involved, for example linking a piece of data to multiple individuals or adding noise to data, but this leads to a significant reduction in the quality of the data. Unfortunately, the validity of the results of Data Mining depends on the quality of the data (Gál et al., 2014). However, Cary et al. (2003) recommend ten practices for developing Data Mining systems that help maintain healthy relationships between organizations and individuals. These include evaluating the quality of the data used for Data Mining. Lastly, Wowczko (2013) says that to prevent the influences on the Data Mining results an expert should check the results to ensure they are correct and legitimate, taking into account any assumptions, limitations and faults made in the research. Finally, among the problems already mentioned is the categorization of individuals and having biased data which causes the violation of the privacy of an individual.

5.7 Area 6 and 7: Group Profiling and Biased Data

The sixth and seventh ethical focus areas are group profiling and biased data. group profiling is the categorization of individuals into groups based on associations about these individuals (Christen et al., 2014). For example, students are placed in a group during the Data Mining process based on a set of characteristics. These groups will encourage students to do what fits their profile or matches them with an outcome that best fits their skills and not necessarily the student's interests. A counter-argument may be that students do not always know what they want because they are young and the consequences of making a big choice can be costly. However, this should still not undermine a student's individuality (Johnson, 2014, 2018). Cary et al. (2003) give the hypothetical example of a bank asking an individual to fill out a form that is about interests and lifestyle factors when they open a bank account. This individual is then categorized into a bad credit risk group after the data from the form is mined and correlation takes place. As a result, an individual is then not kept informed of certain services or even denied a loan based on the group he or she is

in. This makes the Data Mining process biased, which brings us to the topic of biased data.

In some cases, the construction of data is negatively influenced from the outside either intentionally or unintentionally, and these negative influences will be reflected in the data. Often this type of data comes into contact with individuals who utilize the services of bureaucratic organizations, but not every individual comes into contact with such an organization to the same degree, so privilege can seep into the data. The data begins to reflect the social privileges of these frequently encountered individuals and underrepresents those individuals who do not have much contact with the data. For example, minorities are reluctant to provide health data because they fear the government will use it to their detriment or, if they eventually do choose to share their data, they are excluded from the dataset. In addition, for example, the wrong assumption is made that individuals are adequately informed about the topic to which the data relates during data collection (Johnson, 2018). If this type of biased data is used in Data Mining, the results will also contain biased data.

No possible solutions were identified for group profiling and biased data.

6 Conclusion, Discussion and Future Research

As machine learning techniques, such as Data Mining, evolve, so do the ethical implications. Therefore, it is important to take the opportunity to address these ethical focus areas continuously. These ethical focus areas help to ultimately design and deploy ethical Data Mining techniques. The following research question is addressed: “*What ethical focus areas could be retrieved from Data Mining research that can be used in sub-research fields of Data Mining?*” A theoretical review was conducted that resulted in 19 papers that were analyzed from which seven ethical focus areas in Data Mining were identified. Seven ethical focus areas were identified together with possible solutions. From a theoretical viewpoint, this research contributes to the body of knowledge of Data Mining. It also provides a starting point of ethical focus areas that can be adopted in sub-research fields of Data Mining and serve as a future research direction. This has the potential to support, for example, ethical Decision Mining and Process Mining by design, also avoiding common ethical issues existing in Data Mining. From a practical viewpoint, this research contributes to practice because it makes researchers and practitioners alike think about the problems

surrounding these ethical focus areas and what solutions, if any, there are or should be. The results and especially the possible solutions could guide the design and offers a way to structure the thinking of professionals designing and applying solutions for Data Mining and sub-research fields of Data Mining. Furthermore, because these ethical focus areas are made explicit, their impact on the design can be measured and evaluated explicitly. This has high potential because doing so in the design phase of Data Mining solutions decreases the risks of improper use of its underlying and resulting data. The necessity of this is confirmed frequently considering the scandals at different organizations getting negative attention from the press as well as from regulatory authorities.

The following limitations should be taken into account when utilizing the seven focus areas. The oldest paper dates from 2002, which is almost twenty years ago. It could be said that the information in these papers is outdated. However, more recent papers used in this study cite these older papers. Moreover, the information in these older papers is consistent with more recent papers and there was no impression that it was outdated. In addition, the coding of the ethical focus areas was done by a single researcher. This may compromise validity. Any future research should focus on involving additional researchers in coding the ethical focus areas. Furthermore, to the researchers' knowledge, no solution exists to avoid unwanted group profiling and biased data. Therefore, future research on this topic is needed. Additionally, the identified ethical focus areas should be validated in the respective sub-research fields of Data Mining, such as Decision Mining and Process Mining, in order to establish the ethical focus areas and should not be limited to only literature studies. As Hevner and Gregor (2013) pointed out, related research fields can adopt solutions from each other. With this rationale, the identified ethical focus areas are also relevant to related and sub-research fields of Data Mining. Future research should focus on empirical validation and the implementation of the ethical focus areas in order to impact the way people and organizations handle ethically.

References

- Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3), 602–613. <https://doi.org/10.1016/j.dss.2010.08.008>
- Brendel, A. B., Mirbabaie, M., Lembcke, T. B., & Hofeditz, L. (2021). Ethical management of artificial intelligence. *Sustainability*, 13(4), 1974.

- Brewer, R., Westlake, B., Hart, t. & Arauza, O. (2021). The Ethics of Web Crawling and Web Scraping in Cybercrime Research: Navigating Issues of Consent, Privacy, and Other Potential Harms Associated with Automated Data Collection. In A. Lavorgna & T. J. Holt (Eds.), *Researching Cybercrimes* (pp. 435-456). Springer International Publishing. <https://doi.org/10.1007/978-3-030-74837-1>
- Calders, T., & Custers, B. (2013). What is data mining and how does it work? *Studies in Applied Philosophy, Epistemology and Rational Ethics*, 3, 27–42. https://doi.org/10.1007/978-3-642-30487-3_2
- Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3), 602–613. <https://doi.org/10.1016/j.dss.2010.08.008>
- Calders, T., & Custers, B. (2013). What is data mining and how does it work? *Studies in Applied Philosophy, Epistemology and Rational Ethics*, 3, 27–42. https://doi.org/10.1007/978-3-642-30487-3_2
- Cary, C., Wen, H. J., & Mahatanankoon, P. (2003). Data mining: Consumer privacy, ethical policy, and systems development practices. *Human Systems Management*, 22(4), 157–168. <https://doi.org/10.3233/hsm-2003-22402>
- Chadwick, R., & Berg, K. (2001). Solidarity and equity: new ethical frameworks for genetic databases. *Nature Reviews Genetics*, 2(4), 318-321.
- Christen, M., Alfano, M., Bangerter, E., & Lapsley, D. (2014). Ethical issues of “Morality Mining”: Moral identity as a focus of data mining. *Human Rights and Ethics: Concepts, Methodologies, Tools, and Applications*, 3, 1146–1166. <https://doi.org/10.4018/978-1-4666-6433-3.ch062>
- Cook, L. L., & Cook, J. S. (2002). *Ethical Data Mining*. Proceedings - Annual Meeting of the Decision Sciences Institute, 725–729.
- Danielson, P. (2009). Metaphors and models for data mining ethics. *Social Implications of Data Mining and Information Privacy: Interdisciplinary Frameworks and Solutions*, 33–47. <https://doi.org/10.4018/978-1-60566-196-4.ch003>
- De Smedt, J., vanden Broucke, S. K. L. M., Obregon, J., Kim, A., Jung, J.-Y., Vanthienen, J., B, J. D. S., Broucke, S. K. L. M., & Obregon, J. (2017). Decision Mining in a Broader Context: An Overview of the Current Landscape and Future Directions. In *Lecture Notes in Business Information Processing* (Vol. 281, pp. 197–207). Springer International Publishing. https://doi.org/10.1007/978-3-319-58457-7_15
- Dean, M. D., Payne, D. M., & Landry, B. J. L. (2016). Data mining: an ethical baseline for online privacy policies. *Journal of Enterprise Information Management*, 29(4), 482–504. <https://doi.org/10.1108/JEIM-04-2014-0040>
- Fogel, G. K. (2006). Ethical Concerns of Web Marketing. 73–80.
- Franceschet, M. (2010). A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics*, 83(1), 243–258. <https://doi.org/10.1007/s11192-009-0021-2>
- Gál, T. Z., Kovács, G., & Kardkovács, Z. T. (2014). Survey on privacy preserving data mining techniques in health care databases. *Acta Universitatis Sapientiae, Informatica*, 6(1), 33–55. <https://doi.org/10.2478/ausi-2014-0017>
- Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804. <https://doi.org/10.1007/s11192-015-1798-9>
- Johnson, J. A. (2014). The Ethics of Big Data in Higher Education. *The International Review of Information Ethics*, 21(April 2013), 3–10. <https://doi.org/10.29173/irie365>
- Johnson, J. A. (2018). Open data, big data, and just data. *Public Administration and Information Technology*, 33, 23–49. https://doi.org/10.1007/978-3-319-70894-2_2
- Kulkarni, D., Wang, Y., & Sridhar, B. (2010). DATA MINING FOR UNDERSTANDING AND IMPROVING DECISION- MAKING AFFECTING GROUND DELAY PROGRAMS Section 1 . Introduction Section 2 . Statistics of Ground Delay Programs. 1–8.
- Leevis, S., Berkhout, M., & Smit, K. (2020). Future Challenges in Decision Mining at Governmental Institutions. *AMCIS 2020 Proceedings.*, 6.

- https://doi.org/https://aisel.aisnet.org/amcis2020/adv_info_systems_research/adv_info_systems_research/6
- Leevis, S., Smit, K., & Zoet, M. (2020). Putting Decision Mining into Context: A Literature Study. *Lecture Notes in Information Systems and Organisation*, 38(September), 31–46. https://doi.org/10.1007/978-3-030-47355-6_3
- Oliveira, S. R. M., & Zaiane, O. R. (2009). Business collaboration by privacy-preserving clustering. *Social Implications of Data Mining and Information Privacy: Interdisciplinary Frameworks and Solutions*, 113–133. <https://doi.org/10.4018/978-1-60566-196-4.ch007>
- Pal, J. K. (2011). Usefulness and applications of data mining in extracting information from different perspectives. *Annals of Library and Information Studies*, 58(1), 7–16.
- Paré, G., Trudel, M. C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information and Management*, 52(2), 183–199. <https://doi.org/10.1016/j.im.2014.08.008>
- Payne, D., & Landry, B. J. L. (2012). A Composite Strategy for the Legal and Ethical Use of Data Mining. *International Journal of Management, Knowledge and Learning*, 1(1), 27–43.
- Payne, D., & Trumbach, C. C. (2009). Data mining: proprietary rights, people and proposals. *Business Ethics: A European Review*, 18(3), 241–252. <https://doi.org/10.1111/j.1467-8608.2009.01560.x>
- Purusothaman, G., & Krishnakumari, P. (2015). A Survey of Data Mining Techniques on Risk Prediction: Heart Disease. *Indian Journal of Science and Technology*, 8(12). <https://doi.org/10.17485/ijst/2015/v8i12/58385>
- Robinson, S. C. (2015). The good, the bad, and the ugly: Applying rawlsian ethics in data mining marketing. *Journal of Media Ethics: Exploring Questions of Media Morality*, 30(1), 19–30. <https://doi.org/10.1080/08900523.2014.985297>
- Roddick, J., Fule, P., & Ethica, A. (2004). A System for Detecting Privacy and Ethical Sensitivity in Data Mining Results. *Conferences in Research and Practice in Information Technology*, 26, 159–166. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.3.4686&rep=rep1&type=pdf> <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.3.4686>
- Rozinat, A., & Aalst, W. M. P. Van Der. (2006). LNCS 4102 - Decision Mining in ProM. 420–425.
- Schermer, B. W. (2011). The limits of privacy in automated profiling and data mining. *Computer Law and Security Review*, 27(1), 45–52. <https://doi.org/10.1016/j.clsr.2010.11.009>
- Shahidullah, S. M., & Hossain, M. (2009). U.S. federal data mining programs in the context of the war on terror: The congress, court, and concerns for privacy protection. In *Social Implications of Data Mining and Information Privacy: Interdisciplinary Frameworks and Solutions* (pp. 151–180). IGI Global. <https://doi.org/10.4018/978-1-60566-196-4.ch009>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104(August), 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Sprague, R. (2009). Legal frameworks for data mining and privacy. In *Social Implications of Data Mining and Information Privacy: Interdisciplinary Frameworks and Solutions* (pp. 181–198). IGI Global. <https://doi.org/10.4018/978-1-60566-196-4.ch010>
- Thelwall, M., & Stuart, D. (2006). Web crawling ethics revisited: Cost, privacy, and denial of service. *Journal of the American Society for Information Science and Technology*, 57(13), 1771–1779. <https://doi.org/10.1002/asi.20388>
- Thuraisingham, B. (2020). Web Data Mining for Counter-Terrorism. *Web Data Mining and Applications in Business Intelligence and Counter-Terrorism*, 337–350. <https://doi.org/10.1201/9780203499511-22>
- van der Aalst, W. M. P. (2011). *Process Mining*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-19345-3>

- Wahlstrom, K., Roddick, J. F., Sarre, R., Estivill-Castro, V., & Devries, D. (2006). On the Ethical and Legal Implications of Data Mining. School of Informatics and Engineering Flinders University , Adelaide, Australia, February.
- Wowczko, I. A. (2013). Ethical Issues and Guidelines for Data Mining Policy in SOLAS 2 . Ethical Issues in Data Mining. Business Intelligence and Data Mining.