Towards Privacy-Preserving Data Mining in Law Enforcement

Stijn Vanderlooy, Joop Verbeek, and Jaap van den Herik
Maastricht ICT Competence Centre – Institute for Knowledge and Agent Technology
Maastricht University, Maastricht, The Netherlands
{s.vanderlooy, verbeek, herik}@micc.unimaas.nl

Abstract. For law enforcement to be effective, it needs to extract previously unknown knowledge from large amounts of different types of data. Data mining is the most compelling tool for this task as it is motivated by successful applications in numerous domains. Therefore, many believe that data mining can significantly improve the execution of law enforcement. However, a severe problem occurs when data mining is applied: many inevitable mistakes result in privacy violations. Recently, we developed a new approach to data mining, called the ROC isometrics approach, which is proven to produce reliable outputs in the sense that we can set the number of mistakes before the data mining is actually applied. In the paper, we determine the implications of the approach to law enforcement and we propose several recommendations for legislations that try to deal with data mining. As a result, we may conclude that the ROC isometrics approach allows for privacy-preserving data mining so that law enforcement becomes more effectively and efficiently than so far.

1. Introduction

Recent advances in information and communication technology have made data easy to use and cheap to store and exchange. Databases across the world contain large amounts of data of various types. We mention digitized text documents, video and audio files, and financial transactions. The so-called data explosion is also apparent in the domain of law enforcement. After new knowledge has been extracted from data, an information position can be established by which a more effective and efficient execution of law enforcement is possible than before. The new form of law enforcement guided by means of data analysis is known as intelligence led policing (Cope, 2004; Tilley, 2005).

It is non-trivial to extract knowledge from data. Yet, many believe that data mining will enable us to obtain new knowledge in a reasonable amount of time, so that law enforcement can be adequately executed on a tactical, strategical, and operational level (Yen & Popp, 2005). Obviously, data mining provides a variety of tools designed to analyze the available data automatically, i.e. by computer. However, we argue that a serious problem arises when data mining is applied in law enforcement: the output of data mining is not reliable in the sense that mistakes are made. These mistakes may have serious consequences, e.g., violations of privacy when personal data are involved. Clearly, the legal consequences of mistakes are severe and therefore not affordable. To make the problem even worse, it is unknown how many mistakes will be made before the data mining is actually applied. Since the output of data mining cannot be shown to be reliable, there is no ground for making legally correct decisions (Groothuis & Svensson, 2000). It follows that it is important to consider the reliability of the data-mining output in legislation.

The remainder of this paper is organized as follows. In Section 2 we discuss the necessity of knowledge in law enforcement and we introduce two types of investigations that we use throughout the paper. In Section 3 we focus on data mining and we comment on the problem of inevitable mistakes. In Section 4 we review a recently introduced approach for data mining, called the ROC isometrics approach, that is proven to produce reliable outputs in the sense that we can set the number of mistakes before the data mining is actually applied (Vanderlooy *et al*., 2006). In Section 5, we use a juridical framework to discuss the implications of the approach. Consequently, we provide several recommendations for upcoming legislations that try to reach a balance between new possibilities of automatic data analysis and the protection of civilians' privacy. In Section 6 we give our conclusions.

2. Knowledge in Law Enforcement

Law-enforcement agencies store large amounts of data that need to be analyzed in order to find previously unknown and relevant knowledge. This knowledge is used to establish and maintain an information position. However, obvious legal questions arise concerning (1) which data may be analyzed, (2) in which situations, and (3) for which goals.

In Subsection 2.1 we show that knowledge is a necessity for law enforcement in order to fulfil its tasks appropriately. In Subsection 2.2 we distinguish two types of investigation by means of data analysis. We provide a

systematic comparison of the investigation types, relate them to different law-enforcement tasks, and focus on some legal questions that will naturally arise.

2.1. Necessity of Knowledge

The main task of law enforcement is to secure legal order and to provide assistance to civilians in need (Shavell, 2003; Newburn, 2005). Securing legal order implies two subtasks. First, public safety should be established and maintained to reduce the growing unsafe feelings of civilians. Second, crimes should be tracked down and the offenders should be prosecuted subsequently. Executing the second subtask significantly contributes to executing the first subtask, e.g., terrorism aims at causing fear and disruption, and therefore unravelling upcoming terrorist plots strongly increases public safety.

One of the most important activities of law-enforcement agencies is the investigation of suspicions and clues about persons and events. Investigation can result in the prevention of crime in two distinct ways. First, the suspicions and / or clues are correct and the law-enforcement agency responded sufficiently fast to stop the crime before it has been committed. Second, the investigation did not lead to an early termination of the specific crime, but it resulted in new knowledge that is used to manage activities more efficiently and effectively than was previously possible. The latter is called proactive investigation and assumes that sufficient data is available to be analyzed (Thibault *et al*., 2006). Analysis has to be done automatically due to the large amounts of data. The obtained knowledge is necessary to establish and maintain a good information position. This is the motivation behind intelligence led policing: a good information position enables law enforcement to prevent crimes and reduce risks of potential dangers (Cope, 2004; Tilley, 2005).

2.2. Investigation by means of Data Analysis

Knowledge extraction by means of automatic data analysis can be performed (1) in various ways, (2) using various types of data, and (3) for various investigation goals. With respect to the investigation goals, we distinguish between investigation aimed to obtain knowledge for solving specific cases and investigation aimed to obtain any knowledge that leads to new investigations or contributes to the execution of current specific investigations. We call the former goal-oriented investigation and the latter global-oriented investigation.Table 1 summarizes the differences between both types of investigation.

Table 1: a comparison of global-oriented investigation and goal-oriented investigation.

	Goal-Oriented	*Global-Oriented*
Goal	Specific knowledge	General knowledge
Cause	Strong suspicions or clues	No specific cause
Endurance	Temporary	Permanent
Type of Data	Detailed data concerning the specific persons and / or events	Variety of data, possibly concerning persons not related to the obtained knowledge

A goal-oriented investigation takes place when there are strong suspicions or clues concerning a specific event, person, or small groups of events and / or persons. Hence, this type of investigation has a temporary character. It aims at improving a specific information position and it is clearly most useful on the operational level of law enforcement. An example is the investigation of a murder case. The available data to be analyzed naturally consists out of relevant facts for the specific investigation and therefore disproportional privacy violations are not likely to occur. This is in contrast to the global-oriented investigation where such privacy violations are likely to occur. The global-oriented investigation has a permanent character since it aims at improving and maintaining a general information position. The investigation is not oriented towards a specific case or person, and therefore there are no strong suspicions and clues that can be verified. Consequently, the available data to be analyzed contain facts about persons and events that are not related to known crimes. For example, assume for a moment that the global-oriented investigation resulted in knowledge about an organized network of human smugglers. Persons involved in such networks operate across the borders of nations and take advantage of economic corruptibility and conflicts in certain regions. There are almost no reports about human smuggling filed in law-enforcement agencies due to the secrecy of human smuggling and the intimidation of victims (Dutch Upper Chamber, 2007c). So, specific data is not available and knowledge about human smuggling networks should therefore be extracted by analyzing a wider range of data. Care is nonetheless necessary since the legal question arises whether it is allowed to analyze the data for general law-enforcement purposes. This is a difficult but important question to answer since without global-oriented investigation it is virtually impossible to establish and

maintain a good information position. Finally, we note that the results of a global-oriented investigation can lead to a goal-oriented investigation.

It follows that it is challenging to protect the civilians' privacy right, in particular if a global-oriented investigation takes place. Hence, legislation should find the ideal balance between (1) an effective automatic data analysis and (2) the protection of the privacy of civilians. In the next section, we focus on data mining to analyze data and emphasize that the reliability of the data-mining output is most important to consider in legislation.

3. Data Mining

The collection of data and the extraction of new knowledge from data is a significant economic and political activity. The most compelling and promising tools for knowledge extraction are data-mining tools. However, up to now the use of data mining is scarce in the domain of law enforcement.

In Subsection 3.1 we propose a definition of data mining that is suitable for legislation. In Subsection 3.2 we give some applications of data mining and in Subsection 3.3 we analyze why data mining is not used much in the real-life practice of law enforcement.

3.1. Definition

Data mining is an interdisciplinary field mainly consisting of research in applied mathematics and computer science (Han & Kamber, 2006). Various definitions of data mining have been proposed. With respect to legislation it is important to use the most general definition so that (1) unwanted exceptions and backdoors are prevented, and (2) legislation will also apply to future data-mining tools. Brevity and clarity should also be promoted by focussing on the goal of data mining and not on how the data analysis is performed. After all, various data-mining tools often use different analysis methods.

For the aforementioned reasons, we propose to define data mining as follows: "data mining is the analysis of data by automatic means in order to discover previously unknown knowledge in data". The proposed definition is general in the following three ways. First, the data can be of any kind and can be stored in one or more databases in any format of choice. Second, automatic data analysis can be performed in various ways, e.g., by only comparing the data as in a hit / no hit situation or by performing complex mathematical operations on the data. Third, the output of data mining can be anything as long as it provides new insights into executing law enforcement more efficiently and effectively than was previously possible.

3.2. Applications

The number of potential data-mining applications in law enforcement is growing and researchers are publishing empirical results[1] . In this subsection we give a brief review of five interesting applications that cover a wide range of law-enforcement investigations.

First, data mining can be used to analyze cases of armed robberies such that ten times as many links between the cases are found in comparison with a team of human law-enforcement officials (Dahbur & Muscarello, 2003). Useful information to find links includes physical characteristics of victims and offenders, type of weapon used, and location. Second, in a similar vein, data mining can analyze data of crime offenders in order to construct a typical profile of repeated offenders (Blokland *et al*., 2005; De Bruin *et al*., 2006). Useful information here includes marital status, conviction history, and drugs and alcohol addiction. Third, data mining to detect credit card fraud using financial transaction data is gaining interest because a high level of organized crime activity is involved (Kingston *et al*., 2004). Fourth, data mining to reveal links between crimes and offenders has shown to be promising (Goldberg & Wong, 1998; Oatley *et al*., 2004). Fifth, grouping sex offenders that have common characteristics is also an interesting application (Adderley & Musgrove, 2001).

The aforementioned applications illustrate that data mining can be used to find links and patterns in the available data. Investigating these links and patterns results in a better understanding of a group of persons and / or events. In addition to this use of data mining, data mining can be used to construct a classifier. The data to be analyzed then consist of objects and corresponding labels, e.g., objects are persons and the label indicates if the person is a human smuggler or not. The classifier is then a profile used to predict whether persons are human smugglers, e.g., a (completely hypothetical) classifier can be: "If $23 < \text{age} < 27$ and number of prior convictions = 10 and involved in prostitution = yes, then human smuggler = yes". We note that data mining to construct classifiers is considered as a further analysis of data mining to find links and patterns.

[1] Conferences that publish data-mining applications in law enforcement are JURIX, AAAI Symposium, and industrial tracks of data-mining conferences such as ICDM, KDD, and ACM SIGKDD. Two interesting journals are journal of artificial intelligence and law, and journal of information, law and technology.

In the remainder of the paper, we focus on data mining to construct classifiers because many applications can be written as problems for which one has to predict labels for objects.

3.3. Problem of Inevitable Mistakes

Despite the numerous data-mining possibilities in law enforcement, almost no data mining is applied in real-life practice (Oskamp & Lauritsen, 2002). We illustrate the main problem by the following example. Assume for a moment that data mining is used to construct a classifier that reveals the profile of persons who are connected to a network of human smugglers. The available data that are analyzed consist of characteristics of persons, their conviction history, and other personal and legal information that is available. The amount of data is high and it is unknown whether persons involved in human smuggling share a well-defined profile.

Data mining will fail due to an inevitable large number of false positives and false negatives. A false positive is a person incorrectly predicted as a human smuggler. A false negative is a person incorrectly predicted as not a human smuggler. Both types of incorrect predictions have undesired legal consequences. The main undesired legal consequences of a false positive are a privacy violation and a waste of limited human and financial resources. An undesired legal consequence of a false negative is not reducing crime and therefore a failure of the law-enforcement task. In addition, the number of false positives and false negatives is not known before the data mining is actually applied. It is thus not possible to obtain a classifier that is reliable in the sense that we can trust its predictions.

In the next section we review a new approach to construct classifiers that guarantee a preset number of incorrect predictions. For example, if we preset that at most five percent of all predictions may be incorrect, then the classifier will produce at least ninety-five percent correct predictions. This implies that we can preset the reliability of the data-mining output in such a way that (1) the number of undesired legal consequences is still acceptable and (2) a more efficient execution of the law-enforcement task is possible (Vanderlooy *et al*., 2006).

4. Privacy-Preserving Data Mining by the ROC Isometrics Approach

A recently introduced data-mining approach, called the ROC isometrics approach, makes it possible to preset the number of incorrect predictions that we allow a classifier to make. Hence, we know the number of false positives and false negatives before the classifier is actually applied. The abbreviation ROC stands for Receiver Operator Characteristic and it refers to a collection of tools that are used in data mining for various purposes. As is the case with most data-mining approaches, the ROC isometrics approach is defined and proven to work by means of mathematics. A definition of ROC isometrics and consequently an in-depth discussion of the ROC isometrics approach are beyond the scope of this paper. Therefore, we illustrate the idea and working of the approach on a conceptual level that is comprehensible for law enforcement officials, legislators, and privacy watchers. We again consider our example of predicting persons as human smugglers or not.

We slightly enhance our idea about classifiers such that they do not predict the label of a person, but they assign a number to each person. This number is called the score of a person and the higher the score is, the more likely it is that the person is a human smuggler [2]. The classifier is applied to eight test persons for whom we want to know whether they are human smugglers. The result of applying the classifier to the eight test persons is given in Table 2.

Person	***Score***	***Label***
P1	10	Positive
P2	8	Positive
P3	8	Positive
P4	6	Negative
P5	6	Positive
P6	5	Positive
P7	2	Negative
P8	1	Negative

Table 2: the result of applying the data-mining classifier to eight test persons. *The output of the classifier is the score and we ranked persons by their scores. The label of a person is said to be positive if the person is a human smuggler, and negative otherwise. This is analogue to the definition of a false positive and a false negative.*

[2] Assigning scores to persons instead of labels can be seen as being more careful. The score reflects the degree of what is believed to be true. How a score is computed is technical and beyond the scope of this paper.

In an ideal setting,. we have that all test persons who are indeed human smugglers have a higher score than the other test persons. However, this is not the case in our example; and not in real-life applications.

The combination of scores and labels provides us with useful information on how to find incorrect predictions. More specifically, a threshold on the scores is used to predict labels for persons. For example, we can say that persons with a score higher than five are human smugglers. This gives us the result as given in Table 3, which shows that persons P4 and P6 are predicted incorrectly as being a human smuggler and not being a human smuggler, respectively.

Person	*Score*	*Label*	*Prediction*
P1	10	Positive	Positive
P2	8	Positive	Positive
P3	8	Positive	Positive
P4	**6**	**Negative**	**Positive**
P5	6	Positive	Positive
P6	**5**	**Positive**	**Negative**
P7	2	Negative	Negative
P8	1	Negative	Negative

Table 3: the result of applying a threshold of value five to the eight test persons. Incorrect predictions are given in boldface. Clearly, person P4 is a false positive and person P6 is a false negative.

Note that a different threshold leads to incorrect predictions as well. However, if we construct a classifier that uses two thresholds on the scores, then we can significantly reduce the number of mistakes. In our example we can construct such as a classifier as follows: "if score > 6 then human smuggler = yes and if score < 5 then human smuggler = no". It follows that we do not make incorrect predictions by refusing to say something about persons P4, P5, and P6. Refusing to say something about persons does not imply that we cannot learn anything about these persons. A refusal to produce output in fact indicates that (1) the amount of data is not sufficient to construct a classifier that is able to predict labels for all persons with a high reliability, and / or (2) the reliability of the available data is questionable, e.g., too many data are left unspecified, some data contradict each other, or some data are simply incorrect. Obviously, in these cases of uncertainty, the output of data mining cannot be considered as reliable. Therefore, the option to refuse to predict some labels is considered as an advantage of the ROC isometrics approach.

The approach is able to find automatically the thresholds that are needed in such a way that the number of incorrect predictions equals a preset value. Thus if we preset zero incorrect predictions, then the approach will find the classifier in the previous paragraph. If we preset that one incorrect prediction is allowed, then the found classifier will be: "if score > 4 then human smuggler = yes and otherwise human smuggler = no". It is easily verified that this classifier only predicts person P6 incorrectly. Since the number of incorrect predictions is preset, we also preset the number of privacy violations as well as human and financial resources that were lost by chasing dead-ends. Hence, dependent on the necessity of the data-mining application and the societal and legal context, we have to determine the number of mistakes we are willing to allow for a reliable detection of persons involved in human smuggling. We note that the lower the preset number of mistakes is, the more likely it is that the classifier refuses to say something.

5. Juridical Embedding

In the previous section, we showed that the ROC isometrics approach allows presetting the reliability of the data-mining output in such a way that (1) the number of mistakes is acceptable and (2) law enforcement is executed efficiently. In this section we determine the implications of the approach to legislation and we discuss these implications with respect to the Dutch Police Data Bill.

The Police Data Bill was received by the Dutch Lower Chamber on October 17, 2005 and is now examined by the Dutch Upper Chamber (Dutch Upper Chamber, 2007a). The goal of the proposed legislation is to allow for

automatic and non-automatic data analysis while trying to protect the civilians' privacy. We restrict ourselves to automatic data analysis and consequently to the use of data mining .[3]

In Subsection 5.1 we examine how data mining for goal-oriented investigation is regulated and we determine which privacy problems may arise. In Subsection 5.2 we focus on global-oriented investigation and in Subsection 5.3 we provide insights into how new legislations can deal with data mining as best as possible. The insights are accompanied with five recommendations for new legislations that try to deal with automatic data analysis.

5.1. Goal-Oriented Investigation

The use of data mining to maintain the legal order in specific cases is allowed on the basis of article 9 juncto 11 Police Data Bill. The provisions allow goal-oriented investigation by analyzing large amounts of police data for a specific case, even if personal data are involved of persons for whom the exact involvement is not yet known. This is in correspondence with Recommendation R (87) 15 of the Council of Europe which states that there are no limitations concerning the status of the persons involved provided that the purpose of the investigation is served (purpose-binding principle). The data that was analyzed as well as the data-mining result may, after the purpose is served and after approval of an authorized official, be put at disposal for further analysis. Disposal of the data for further analysis occurs when this is believed to be necessary with respect to four goals of which the following two are of interest in this paper: (1) the goal to execute another goal-oriented investigation and (2) the goal to develop insights into the involvement of persons who cause serious threats of the legal order. Article 9 juncto 11 mentions privacy safeguards of which the principle of purpose binding is most convincing. However, this principle alone is not sufficient to prevent privacy violations since the provisions allow personal data to be analyzed and they seem to believe that the result of data mining is mostly correct; otherwise it is unsafe to use the new knowledge in other investigations. However, we argued in Subsection 3.3 that data mining fails due to an unknown number of inevitable mistakes. The ROC isometrics approach is thus needed.

Next to article 9, section 1 of article 8 regulates goal-oriented investigation in order to execute the daily law-enforcement task. The type of data mining that is considered here is rather primitive: data analysis is restricted to a simple search in databases such that a hit / no hit situation occurs. In case of a hit it is allowed to analyze the data further when the data are of a type as defined in the Police Data Decree (in preparation). The proposed hit / no hit situation can be seen as a privacy safeguard.

5.2. Global-Oriented Investigation

Global-oriented investigation by means of data mining for the execution of the daily law-enforcement task is regulated by sections 2 and 3 of article 8 Police Data Bill. The article allows analyzing data for a permanent execution of the daily law-enforcement task in order to establish a global information position. Data mining may be applied on data one year after registration and this for a maximum of five years. In Subsection 2.2 we argued that the data to be analyzed for a global-oriented investigation typically need to contain facts about persons for whom there are no strong suspicions or clues of involvement in crime. Article 8 acknowledges our argument since no distinction is made with respect to the status of the persons concerned, including non-suspects. The legislator claims that article 8 provides sufficient privacy safeguards because it is in accordance with the principles of the European legal framework, in particular article 8 of the European Convention on Human Rights, the Data Protection Convention, and Recommendation R (87) 15 of the Council of Europe (Council of Europe, 1950; 1981; 1987). These legal instruments provide four principles. The four principles are as follows:

- necessity principle: data analysis should be necessary to execute the law-enforcement task;
- purpose-binding principle: data analysis should occur with respect to the purpose of the investigation;
- proportionality principle: data analysis should be in balance with the possibility of privacy violations;
- subsidiarity principle: data analysis should only be used when less intrusive methods are not effective.

Despite the accordance of article 8 with these principles, it remains that data mining makes mistakes. Without knowing the mistakes, it is impossible to judge whether the data-mining output can be used to make legally correct decisions. It follows that the legislator has to consider the reliability of the data-mining output as a major concern in automatic data analysis. Without taking reliability into account there is no legally correct execution of

[3] The Dutch Police Data Bill does not mention data mining. It uses the term automatically processing data which is defined in article 1 as any act or series of acts with regard to police data. This includes among others gathering, registering, updating, improving, comparing, and joining data. Our notion of data analysis and data mining clearly falls within the broad scope of the definition.

global-oriented investigation, although this type of investigation is necessary to establish and maintain a global information position. The ROC isometrics approach is clearly needed to secure civilians' privacy.

In addition to article 8, article 10 juncto 11 Police Data Bill also regulates data mining for global-oriented investigation. The distinction between articles 8 and 10 is that article 10 emphasizes on analyzing links between crimes and offenders in order to obtain new insights into anything that may be of interest. The provisions allow data mining for any purpose as long as a serious threat of the legal order exists. The Minister of Justice remarks with our consent that analyzing data about persons with no clear connection with the crimes is needed to find the desired links (Dutch Upper Chamber, 2007c). In Subsection 2.2 we already gave the example of unravelling organized networks of human smugglers. Two similar examples are unravelling terrorist plots and tracking down persons located in some region or building using the log data of mobile phone operators. Unfortunately, privacy violations are now far more likely to occur than before. This means that data mining becomes impossible, unless the ROC isometrics approach is used.

5.3. Five Recommendations

Subsections 5.1 and 5.2 show that it is difficult to regulate data mining in such a way that (1) the execution of law enforcement becomes more effectively and efficiently than before, and (2) privacy violations are reduced to a minimum. The ROC isometrics approach makes it possible to preset the number of mistakes and therefore also the number of privacy violations. The approach facilitates the use of data mining and it helps to formulate clear legislation. In addition, and most importantly, the approach enables the safe use of automatic data analysis for global-oriented investigation. This makes it possible to establish a good information position. So, we recommend that the reliability of the data-mining output needs to become an important issue in future legislation.

The number of allowed mistakes should be preset in such a way that the number of privacy violations is in accordance with the principles of necessity, proportionality, and subsidiarity. For example, in case of serious threats of the legal order, we may believe that the number of allowed mistakes becomes higher than usual since law-enforcement officials are eager to have much data-mining output at their disposal. In case of non-serious threats and a low crime rate for a sustained time period, we may assume that securing privacy will become more important. The number of allowed mistakes will therefore become lower than before and consequently the ROC isometrics approach is restricted to produce outputs that are correct with a higher probability. So, we recommend that the reliability of the data-mining output is preset as the result of evaluating the principles of necessity, proportionality, and subsidiarity.

Once privacy-preserving data mining is implemented in law-enforcement practice, it might be necessary to adapt the European and national data protection legislation with a view to the privacy-preserving safeguards provided by the data-mining approach. Therefore, we recommend the data protection working groups within the Council of Europe and the European Union as well as the national legislators to take this into account.

Moreover, we recommend taking notice of the provisions of the Police Data Bill with respect to the sensitivity of data, the availability of data, and the duration of the data analysis. A discussion of these provisions is outside the scope of this paper, although they can clearly serve as an example for other European countries.

Finally, we recommend more interdisciplinary research between the field of law and that of computer science. Only then it is possible to formulate legislation that (1) fits the requirements and possibilities of data mining to improve law enforcement and (2) protects civilians' privacy.

6. Conclusions

To establish and maintain a good information position we argued that goal-oriented investigations and global-oriented investigations need to be performed successfully. Goal-oriented investigations aim to obtain knowledge for solving specific cases and global-oriented investigations aim to obtain any knowledge that improves the information position in general. Due to the large amounts of available data, law-enforcement officials are eager to apply data mining in order to find new knowledge. We gave several example applications of data mining.

Conventional data-mining approaches make inevitable mistakes. These mistakes are unaffordable in law enforcement, e.g., because they may result in privacy violations. The ROC isometrics approach makes it possible to set the number of allowed mistakes before the data mining is actually applied. Hence, we can also preset the number of allowed privacy violations. If this preset number is chosen to be in accordance with the principles of necessity, proportionality, and subsidiarity, then we may conclude that the ROC isometrics approach enables a safe application of data mining in law enforcement. The approach guarantees that we can trust the data-mining output and therefore legally correct decisions are made. This is even the case for the important global-oriented investigations. So, we may conclude that a major achievement is presented towards the use of privacy-preserving data mining.

Finally, we have argued that the ROC isometrics approach cannot be applied in law enforcement without clear legislation that considers the ideal balance between (1) an effective automatic data analysis and (2) the

number of legally allowed privacy violations. For this argument we have determined the implications of the data-mining approach to legislation. Here, we may conclude that legislation has to pay much attention to the reliability of the data-mining output. This is currently not the case in the Dutch Police Data Bill that thus falsely claims to provide sufficient privacy safeguards to account for more data-mining possibilities. For a proper treatment of the claim we refer to our five recommendations.

Acknowledgements
The first author is supported by the Dutch Organization for Scientific Research (NWO); the research is part of the ToKeN programme, viz. the IPOL project, grant no. 634.000.435.

References

1. Adderley, A., & Musgrove, P. (2001). Data Mining Case Study: Modelling the Behaviour of Offenders who Commit Serious Sexual Assaults. In F. Provost, & R. Srikant (Eds.), 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 215-220). San Francisco, CA: Association for Computer Machinery.
2. Blokland, A., Nagin, D., & Nieuwbeerta, P. (2005). Life Span Offending Trajectories of a Dutch Conviction Cohort. Criminology Volume 43 (Issue 4), 919-954.
3. Cope, N. (2004). Intelligence Led Policing or Policing Led Intelligence? Integrating Volume Crime Analysis into Policing. The British Journal of Criminology Volume 44 (Issue 2), 188-203.
4. Council of Europe (1950). Convention for the Protection of Human Rights and Fundamental Freedoms, SETS No. 005, from http://conventions.coe.int/.
5. Council of Europe (1981). Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, SETS No. 108, from http://conventions.coe.int/.
6. Council of Europe (1987). Recommendation No. R (87) 15 of the Committee of Ministers to Member States Regulating the Use of Personal Data in Law Enforcement, from http://coe.int/legal/.
7. Dahbur, K., & Muscarello, T. (2003). Classification System for Serial Criminal Patterns. Artificial Intelligence and Law Volume 11 (Issue 4), 251-269.
8. De Bruin, J., Cocx, T., Kosters, W., Laros, J., & Kok, J. (2006). Data Mining Approaches to Criminal Career Analysis. In C. Clifton, & N. Zhong (Eds.), 6th IEEE International Conference on Data Mining (pp. 171-177). Hong Kong: IEEE Computer Society.
9. Goldberg, H., & Wong, R. (1998). Restructuring Transactional Data for Link Analysis in the FinCEN AI System. In D. Jensen, & H. Goldberg (Eds.), AAAI Fall Symposium (pp. 38-46). Orlando, FL: AAAI Press.
10. Groothuis, M., & Svensson, J. (2000). Expert System Support and Juridical Quality. In A. Breuker, R. Leenes, & R. Winkels (Eds.), 13th Foundation for Legal Knowledge Based Systems Conference (pp. 1-11). Enschede: IOS Press.
11. Han, J., & Kamber, M. (2006). Data Mining: Concepts and Techniques, Second Edition. Oxford: Morgan Kaufmann.
12. Kingston, J., Burkhard, S., & Vandenberghe, W. (2004). Towards a Financial Fraud Ontology: a Legal Modelling Approach. Artificial Intelligence and Law Volume 12 (Issue 28), 419-446.
13. Dutch Upper Chamber (2007a). Dutch Police Data Bill, from http://www.eerstekamer.nl/, in Dutch. Den Haag: Sdu Uitgevers.
14. Dutch Upper Chamber (2007b). Dutch Police Data Bill (Explanatory Memorandum), from http://www.eerstekamer.nl/, in Dutch. Den Haag: Sdu Uitgevers.
15. Dutch Upper Chamber (2007c). Dutch Police Data Bill (Memorandum of Reply), from http://www.eerstekamer.nl/, in Dutch. Den Haag: Sdu Uitgevers.
16. Newburn, T. (Ed.) (2005). Policing: Key Readings (pp. 129-260). Pleymouth Devon: Willian Publishing.
17. Oatley, G., Zeleznikow, J., & Ewart, B. (2004). Matching and Predicting Crimes. In A. Macintosh, R. Ellis, & T. Allen (Eds.), 24th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence (pp. 19-32). Cambridge: Springer.
18. Oskamp, A., & Lauritsen, M. (2002). AI in Law Practice? So Far, Not Much. Artificial Intelligence and Law Volume 10 (Issue 4), 227-236.
19. Shavell, S. (1993). The Optimal Structure of Law Enforcement. Journal of Law and Economics Volume 36 (Issue 1), 255-287.
20. Thibault, E., Lynch, L., & McBride, R. (2006). Proactive Police Management, Seventh Edition. London: Prentice Hall.
21. Tilley, N. (2005). Community Policing, Problem-Oriented Policing and Intelligence-Led Policing. In T. Newburn (Ed.), Handbook of Policing (pp. 311-339). Pleymouth Devon: Willian Publishing.

22. Vanderlooy, S., Postma, E., Tuyls, K., & Sprinkhuizen-Kuyper, I. (2006). Reliable Instance Classifications in Law Enforcement. In P.-Y. Schobbens, W. Vanhoof, & G. Schwanen (Eds.), 18th Benelux Conference on Artificial Intelligence (pp. 323-330). Namur: University of Namur.
23. Yen, J., & Popp, R. (2005). AI Technologies for Homeland Security (Tech. Rep. No. 1). AAAI Spring Symposium: AAAI Press.