HCTL Open Int. J. of Technology Innovations and Research
HCTL Open IJTIR, Volume 8, March 2014
e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

A Survey of Privacy-Handling Techniques and Algorithms for Data Mining

***Vivek Uniyal*[*], *Sudeep Panchpuri*[†], *Govind Kamboj*[‡]**
vivek.akshay@gmail.com

Abstract

Data mining is nothing but all about to analysis step of KDD which is Knowledge Discovery from Data. Data mining concept is all about the computational process of patterns discovery from a large data sets including methods with intersection of database system, machine learning and artificial intelligence. Privacy-handling is one of the most important concern in data mining these days. In this paper, we have presented a survey various methods, techniques used for privacy-handling in data mining.

[*]M.Tech Student, Department of Computer Science & Engineering, Graphic Era University, Dehradun, India

[†]M.Tech Student, Department of Computer Science & Engineering, Graphic Era University, Dehradun, India. Email: sudeepbeit@gmail.com

[‡]Assistant Professor, Department of Computer Science & Engineering, Graphic Era University, Dehradun, India. Email: govind.kamboj@gmail.com

HCTL Open Int. J. of Technology Innovations and Research
HCTL Open IJTIR, Volume 8, March 2014
e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

Keywords

Data Mining, Privacy-Handling Issues, Privacy-Handling Algorithms.

Introduction

Data mining is nothing but all about to analysis step of KDD which is Knowledge Discovery from Data. Data mining concept is all about the computational process of patterns discovery from a large data sets including methods with intersection of database system, machine learning and artificial intelligence. Data mining process has a main goal of extracting information from data set to transform and build an understandable structure for future use. Data mining is an essential step within the knowledge discovery process.

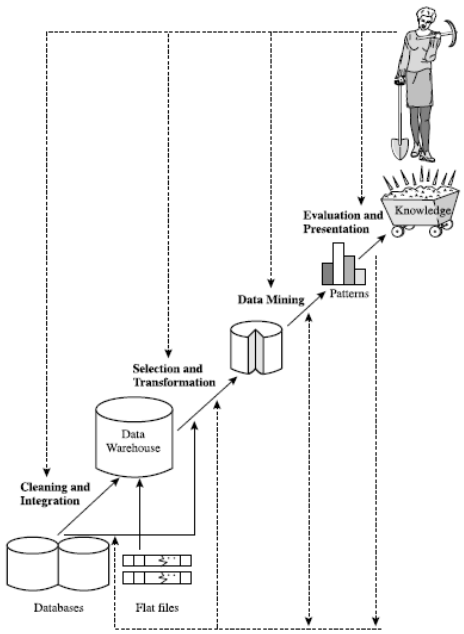

Figure 1: *Data mining a step included in the process of knowledge discovery [18]*

Data mining architecture have included 6 major components:

1. World Wide Web, data base, data warehouse or other information repository which are one or set of spreadsheets, database, data warehouses or other types of information repositories. Data cleaning (remove inconsistent and noise data), data integration (combination of multiple data source)

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

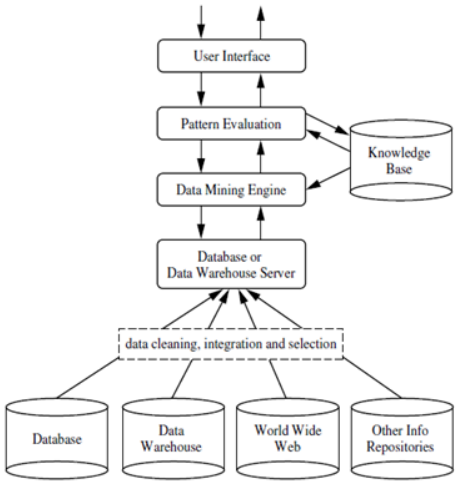

Figure 2: *Typical data mining system Architecture [18]*

and data selection (relevant data retrieval from database by analysis task) techniques may be performed on data.

2. Database or data warehouse servers for fetching relevant data according to user's data mining request.

3. Knowledge base is domain knowledge to guide evaluation or search interestingness of resulting patterns. It includes concept hierarchies, to organize attributes or attribute values into a different level of abstraction. Examples: pattern interest based on unexpectedness, metadata and additional interestingness constraints or thresholds.

4. Data mining engine which is set of functional modules for tasks. Characterization, classification, association and correlation analysis, cluster analysis, prediction, outlier and evolution analysis.

5. Pattern evaluation module focuses the search toward interesting patterns. Pattern evaluation modules integrated with mining modules, relying on implementation of data mining method used.

6. User interface to communicate between user and data mining system.

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

Data mining can involves 6 common classes of tasks:

1. Anomaly, outlier, deviation or change detection is for identifying unusual data records and data errors which would be interesting or required for future investigation.

2. Dependency modelling or Association rule learning is for searching relationships between variables.

3. Clustering is to discovering structures and groups in data which are similar in some way or another way without using known structures in data.

4. Classification is to generalizing known structure to apply for making new data.

5. Regression is to find function for modelling the data with least error.

6. Summarization is to provide more compact dataset representation including report generalization and visualization.

Advantages of Data Mining [19]

1. For marketing and retailing to provide useful and accurate trends about the customer behavior of purchasing. By which they can predict the purchasing interest of consumer or customer. According to the history records of shopping and purchasing trends of customer's company will introduce new products in market to surprise the customer. Also in retailing at similar way through trend the store manager arrange the products and discount to attract his customers.

2. For banking and crediting data mining will reduce the risk of fraud. Data mining can help to bank for estimate risk which is associated with each given loan by examining the previous customers with similar attributes. Data mining can also helps on detection of potentially fraudulent credit card transaction.

3. For law enforcement to identifying criminal suspects also to arrest them by examining trends in crime type, location, habit and other patterns of behaviors.

4. For researchers to speeding up the process of data analyzing by which they can utilize more time to work on other project.

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

Disadvantages of Data Mining [19]

1. `Privacy issues`, because there is no way to protect the personal information from others in data mining. Companies can share their customer's personal information to another company which will not be known to customers. And there are some peoples afraid of purchasing things through online because they think that the details of their personal information can be hacked.

2. `Security issues`, because of the companies do not have sufficient security system to protect the information or their databases. And there are lots of hackers who want to crash or hack that database. Companies have to secure their mining process from the intruders. They have to develop some security over the accessing of database. Hackers can hack the database and find out all the personal information like address, account number, social security number, payment history of all the consumers, it will become a big problem for company and consumers. So identity theft could be a real problem on data mining.

3. `Misuse of information and inaccurate information`, which is intended to be used for marketing or for some ethical purpose from data mining, it may be misused.

Why Privacy-Handling is Required in Data-Mining

There are so many places where the data mining with privacy handling will required. We can see many popular places where this combination is being applied.

1. In India we all are identified by our voter id card, pan card, aadhar card and so on. These provide us a unique code for individual identification. The whole data is recorded officially in government's server. Data collected in servers in a large manner it may be in terabytes or petabytes. The data is managed in the server according to a data mining algorithm. And if all the records of the transactions and personal information are hacked by the unauthorized persons or intruders for some terrorist activities or harm the data. More than this they can also change the records which may bring that person or the government into severe problem. Hence for this authentication and privacy is important.

2. Nowadays everyone is having one or more contact numbers which can uniquely identify him. Whole personal information (name, address, identification etc.) regarding the person can be stored in the service provider's

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

server. Sometimes companies deal to share contacts or detail of their customers for their own profits which is strictly prohibited. Companies should not disclose the detail about their customers. Hence privacy is required.

3. In banking sector huge amount of transactions are performed. Data about the details of customer and banks are stored in servers which can be misused by the hackers. Bankers do not provide any personal details about their customers to other.

4. There are three main approaches for privacy in data mining: heuristic-, reconstruction- and cryptography-based.

5. Heuristic-algorithms for hiding knowledge which is not need to reveal by an organization. This algorithm like as adaptive modification which can modify only selected values to minimize the loss of utility.

6. Reconstruction-based techniques by which original distribution of data is reconstructed regarding of randomized data.

7. Cryptography-based techniques where multiple parties computed the data in a secure manner. The computation is secured between the multiple users, none of them known about anything except their own input and result.

Literature Review

Qiang Yang and Xindong Wu (2006) in their research paper **10 Challenging Problems in Data Mining Research** [1] have proposed their research in which they have explained 10 challenging problems over data mining research. They concerned their research after consulting most active researches over machine learning and data mining. They provide 10 challenges : (1) develop unifying theory, (2) scaling the highly speed data stream and highly dimensional data, (3) time series and sequential data mining, (4) mining complex interesting knowledge from complex data, (5) data mining inside network setting, (6) mining multi-agent and distributed data, (7) data mining for environment and biological problems, (8) process-related problems of data mining, (9) privacy, security and data integrity, (10) dealing the unbalanced, non-static and cost-sensitive data.

Godswill Chukwugozie Nsofor (2006) in his research thesis **A Comparative**

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

Analysis of Predictive Data-Mining Techniques [2] has proposed a research for comparative analysis of predictive data mining techniques. He present five different predictive techniques of data mining in his thesis paper in which he proposed four linear and one nonlinear technique. These five techniques are compared on four unique and different data sets having combination of few predictor, many predictor, highly collinear and very redundant variables also presence of outliers. Five techniques are: Multiple Linear Regression MLR, Principal Component Regression PCR, Ridge Regression, Partial Least Squares PLS and Nonlinear Partial Least Squares NLPLS. The datasets are Boston Housing, Collinear (COL), Airlinear and Simulated data sets. He found that PLS performance is better than other four techniques to build linear models, also dealt with COL dataset and gave best predictions by simplest model. PLS is reduced data dimensionality. He described in his research that supervised are better than unsupervised techniques to demonstrate better predictive ability.

Charu C.Aggarwal and Philip S.Yu in their research paper **Privacy-Preserving Data Mining: Models and Algorithms** [3] have provided a review of privacy preserving data mining algorithms and models. They found that privacy-preserving data mining would take importance because of rapidly increment of sensitive information over the internet. In this paper they described state-of-art methods for privacy. They also define methods for k-anonymization, randomization and distributed privacy-preserving data mining. They have present that there are some cases where output of data mining applications would be disinfected for privacy-preservation purpose over high dimensional data set. They also discussed methods for vertically and horizontally partitioned data. They review some issue over data mining and data management applications of their downgrading effectiveness like as query processing, classification and rule mining.

Korosh Golnabi et al. (2006) in their research paper **Analysis of Firewall Policy Rules using Data Mining Techniques** [4] have analysis firewall policy rules using Data Mining Techniques. Main purpose of firewall technology is network security and defense. They addressed the main problem that how much firewall rules are used, well-organized, up-to-dated or efficient to reflect network traffic's current characteristics. They have presented a set of techniques and algorithms by which they analysis and manage policy rules about firewall: (1) mining frequency of its log based network traffic to deduce effective firewall policy rules by data mining techniques, (2) reduce policy rule generalization by Filtering-Rule Generalization and (3) to generate effective firewall policy rules set by a technique to identify some dominant rules and any decaying rule. They developed a prototype system also demonstrated usefulness of their approaches.

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

They have described a process of managing firewall policy with generalization, anomaly detection and policy update using Association rules.

Benny Pinkas (2002) in his research paper **Cryptographic Techniques for Privacy-Preserving Data Mining** [5] presented his research about some implementations over privacy-preserving data mining by cryptographic techniques. He described that for achieving a remarkable results in secure distributed computation research should be done by research in theory of cryptography. It is known that non-trusted parties will compute functions of their different inputs jointly while they ensued that no one learns anything but defined output of the function. Results are shown using generic construction and applied for any functions which have efficient representation as a circuit. In this paper he described those results and their efficiency, Also discussed their relevance with privacy-preserving computation of algorithms of data mining. This paper defines security and generic construction for two or multi party scenarios.

Wenke Lee and Salvatore J. Stolfo (1998) in their research paper **Data Mining Approaches for Intrusion Detection** [6] have defined approaches of data mining for intrusion detection. In this research they develop the general and systematic intrusion detection methods. They used data mining techniques to discover useful and consistent system features patterns that describe program and user behaviour. They used relevant system features set to compute (inductively learned) classifiers to recognize known intrusion and anomalies. To detect anomalies they experiments on network tcp dump data and send mail system call data, and demonstrate to construct concise and accurate classifiers. They provide overview and implement two general data mining algorithms: association rules and frequent episodes algorithms to compute inter- and intra-audit record patterns for user behaviour or describing program. Discovered patterns guides the process of audit data gathering and selection of facilitate feature. They propose architecture of agent based for intrusion detection system to meet both algorithms challenges, here learning agent continuously computed and gave updated detection models to agents of detection.

Yehuda Lindell and Benny Pinkas (2000) in their research paper **Privacy Preserving Data Mining** [7] have provided an introductive concept about privacy preserving data mining. In their model, two parties are going to union their own confidential databases and run data mining algorithm over it but showing any unnecessary information. It is a solution for generic secure computation for multi-party, based on circuit computing algorithm evaluation on entire input. They focus on the decision tree learning problem with using ID3

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

algorithm. They have provided a better efficient solution than generic solution.

Sheng Zhong (2004) in his research paper **Privacy Preserving Data Mining** [8] has discussed about the computations with untrusted parties over privacy, integrity and incentive-compatibility. He presented mixed network tailored to build substantially speedup election system. At the very first chapter of his dissertation he mentioned that privacy and integrity problems belong to secured multi-party computation researching traditionally but naturally extension between economically rational or selfish parties with multi-party computation securely is incentive compatibility. He divided his dissertation in three parts: chapter 2 presents all techniques which are frequently used; chapter 3 to 6 he provide practical solution for integrity and privacy problems in various scenarios; chapter 7 is for adding incentive consideration to problems of multi-party computation.

Aman Jain and Bikash Sharma (2007) in their research dissertation paper **Privacy, Integrity, and Incentive-Compatibility in Computations with Untrusted Parties** [9] have presented the thesis for the purpose of privacy preserving with data mining. They proposed privacy over data mining in reconstruction and randomization. Data mining service needs accurate data input for meaningful results, but privacy may effects users to provide fake information. They defined that for client privacy techniques based on random agitated data records used in data mining process. In randomization, client will protect the data privacy by disordering or perturbing with randomization algorithm and after that submitting the randomized version. In the other side of this proposed thesis they provide reconstruction of randomized data set by using algorithm to get approximate original data set. They have calculated all the performance metrics like accuracy, privacy and percentage deviation branches.

Rakesh Agrawal and Ramakrishnan Srikant (2000) in their research paper **Privacy-Preserving Data Mining** [10] have proposed some methods in privacy preserving Data mining. They took a concrete case of creating decision-tree classifier from training data where individual records are in disorder. They have proposed a novel reconstruction procedure for accurate estimating the distribution of original data values. They proposed a plan for effectiveness of randomization with categorical attributes by reconstruction.

Alexandre Evfimievski and Tyrone Grandison (2009) in their research paper **Privacy-Preserving Data Mining** [11] have introduced the privacy-preserving data mining in his research paper. They have identified that a nave

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

approach for privacy-preserving data mining is not proven privacy guarantees which is `security by obscurity`. They have researched over the algorithm and found that this algorithm is not claimed privacy preservation over attacks and all datasets of a certain class according to its nature. They have listed some principle approaches to enable privacy preserving data mining also defines their enforcing privacy and methods. The two important emerged needs for privacy-preserving data mining are deliver better services in data analysis and ensuring data owner's privacy rights. They have presented privacy preserving data mining approaches namely randomization, suppression, summarization and cryptography. They have also describes advantages, disadvantages and privacy guarantees of each approach stated to draw state of art in balanced view.

Murat Kantarcioglu and Chris Clifton (2003) in their research paper **Privacy-preserving Distributed Mining of Association Rules on Horizontally Partitioned Data** [12] have given the method for privacy-preserving distributed mining of horizontally partitioned data with association rules. They developed methods with cryptographic techniques to information shared minimization.

Yehuda Lindell and Benny pinkas (2008) in their research paper **Secure Multiparty Computation for Privacy-Preserving Data Mining** [13] have described some notions and paradigms for Secure Multiparty Computation for Privacy-Preserving Data Mining. They surveyed about the efficiency and demonstration of difficulties in construction of highly efficient protocols. There is also presented the common errors of secure multiparty computation techniques applied on privacy-preserving data mining, then they discussed about the relationship between privacy-preserving and secure multiparty computation on data mining.

T.Y.Lin et al. (1996) in their research paper **Security and Data Mining** [14] have defined about Security and Data mining. They have discussed well developed theory and rough set theory. Also they have illustrated some potential applications to security problems. The research paper has mainly four sections: in first section they included the data mining as a security concern by their non-security research results, in second section identifying classical inference problem and data mining, in third section two possible views over security problems of data mining and last fourth section rough sets and data mining.

Chris Clifton and Don Marks (1996) in their research paper **Security and**

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

Privacy Implications of Data Mining [15] have presented in their research paper the data mining privacy and security implications. In this research paper they have discussed all the problems and their solutions also outlines ideas for privacy and security in data mining. They discussed about the possible solutions of the problems where the possible solutions are divided into: Limited Access, Fuzz the data, Eliminates unnecessary grouping, Augment the data and Audit.

Vassilios S. Verykios et al. (2004) in their research paper **State-of-the-art in Privacy Preserving Data Mining** [16] have provided privacy preserving data mining with state of art. They provide an overview of data mining algorithms with classification, clustering, extend description of various algorithms of privacy preserving data mining. This research is all about for securing sensitive data and knowledge from intruders or malicious users.

Chris Clifton et al. in their research paper **Tools for Privacy Preserving Distributed Data** [17] Mining have presented privacy preserving tool in distributed data mining. They have defined that there are numerous applications of privacy preserving distributed data mining. They provide a solution for applications of privacy preserving data mining specifically combined in toolkit of components. They provide some components of toolkit and showed how to solve privacy preserving data mining problems. They defined algorithms for techniques and applications with numerical equation in their research paper. The resulting data mining technique some time not satisfied the secure multiparty computation definition so they provide a solution which is intermediate information as part of result, by which enabling secured multiparty computation proof. This technique provides guarantee controlled disclosure ability. But for iterative techniques intermediate results may revealed because of lot of information by several iterations, so there is a solution split the intermediate results into shared randomly-determined, the shares will combined only after the end of computation.

Methods of Privacy Handling

Privacy-preserving data mining have numerous applications which are supposed to be `privacy-violating` applications. Methods designing will continue to be effective without compromising in security. Most of the methods do the transformation on the data for privacy in order to perform privacy-preservation. Such methods reduce the granularity of representation of data in order to

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

reduce the privacy. This reduction can generate some loss of effectiveness of data management shown in resulting data. This is the trade-off between information privacy and loss. Some techniques are shown below:

The Randomization Method

Introduced as randomization method for privacy-preserving data mining in which distorting the data by probability distribution. In this technique noise is added to mask the attribute values of data records. The noise added will have to be sufficiently large enough by which individual record values can't be recovered. Let a set of data record which is denoted by $A = a_1...a_N$. For record $a_i \in A$, add the noise component which is included from probability distribution fB(b). The noise component drawn independently denoted as $b_1...b_N$. now resultant new set of distorted records denoted by $a_1 + b_1...a_N + b_N$. This new set of record is denoted by $c_1...c_N$. Generally we can say that variance of adding noise in large enough size will make difficult to easily guess the original record from distorted data. Original record can't be recovered but the original record distribution can be recovered.

We can say that if A be the random variable indicating the original record data distribution, B be indicated as noise distribution random variable and C be indicated as the final record random variable. Then,

$C = A + B$
$A = C - B$

Here, N instantiations about probability distribution C are known, whereas publicly known is B distribution. For N which is large enough number value, distribution of C can be approximate close enough using various methods like kernel density estimation. There are iteration methods to found the approximated distribution of C with subtraction result by $C - B$ to found approximate original probability distribution A.

Group Based Anonymization Method

In this method, there are many methods for privacy transformations by constructing anonymous record in groups which will be transformed in a group-specific way.

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

k-Anonymity Framework

This method can reduce the granularity of pseudo-identifiers representation with the use of generalization and suppression techniques.

In generalization, generalize the attribute values to a range for reducing representation granularity as like as date of birth generalized to a range like as year of birth, this can reduce the identification risk. In suppression, remove completely the value of attribute. Such method reduces the identification risk by using public records, while accuracy-reduction of application on transformed record.

To reduce the risk of identification, k-anonymity approach has to be required that each data release must be in the way that every quasi-identifiers combination value can be matched indistinguishably to at least k respondents. k-Anonymity algorithm, first approach which uses domain generalization hierarchies of quasi-identifiers to built k-anonymous tables. For limit the level of generalization by k-minimal generalization for maintaining data precision as much as possible for anonymity as a given level.

Incognito method is for k-minimal generalization using bottom-up aggregation along with domain generalization hierarchy. Incognito method uses the bottom-up-breadth-first-search of hierarchy of domain generalization, to generate all possible minimal k-anonymous tables for known private table. Firstly checking the k-anonymity to each single attribute, remove all generalizations attributes which do not satisfy the k-anonymity. After that computation of generalizations in paring, again check k-anonymity for each pair and remove all generalization which can not satisfy k-anonymity.

Top-down specialization and bottom-up generalization are two main methods for k-anonymity. In top-down heuristic starts with general solution, after that specialize some attributes of current solution to increase information but anonymity reduction. The k-anonymity never violated because anonymity reduction is always controlled. Complementary method of top-down specialization is bottom-up generalization.

Personalized Privacy-preservation

If the value of k for anonymization may vary with the record and not fixed value, then use the personalized privacy-preservation method. One approach is

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

condensation-based which is used in the variable constraints presence on the data records privacy. In this technique constructs non-homogeneous size based groups from data, such that each record lies in a group and its size is at least equals to the anonymity level of group.

In another model a person specifies privacy level for his sensitive values. In this technique the assumption is that an individual can specify generalization hierarchy node to decide anonymity level.

Utility based Privacy-preservation

This method is leads to loss of utility (information) for data mining purpose. There may need to be suppressed many attributes in order to preserve anonymity and utility.

A method for utility based for data mining with local recording which is based on a fact that is an application point of view in which different attributes have different utility. Mostly the anonymization methods are global, where particular tuple value maps globally with same generalize value. Data space is divided into a number of regions in local recoding, and tuple maps with generalized value which is local to that region.

Sequential Releases

This type of method is for dynamic application that is like data streams, where data releases sequentially. Different type of data in table are sequentially released, then join is used to sharpen the ability to particular distinguish records in the data.

The l-diversity Method

This technique is used for the main attacks where background knowledge is available to attacker. Name of attacks are homogeneity attack and background knowledge attack. L-diversity is for maintaining minimum group size of k and focuses the sensitive attribute diversity maintenance.

Distributed Privacy-preserving Data Mining

Main goal for this type of methods is allow useful aggregate statistical computation over the entire data set, there is no compromise with the privacy of different participant individual data set. The participant collaborate in obtaining an

aggregate results, but they will not fully trust each other while the distribution is occur on their own data set. So according to this purpose data sets may either be vertically or be horizontally partitioned. In vertically partitioning, there may be different attributes or views of same set of records in individual entities over data set. In horizontally partitioning, there may be individual record which is spread out across multiple entities, where each entity has same set of attributes. The distributed privacy-preserving method is overlaps with the field of cryptography which is determining the secure multi-party computation.

References

[1] Qiang Yang and Xindong Wu, **10 Challenging Problems in Data Mining Research**, International Journal of Information Technology & Decision Making, World Scientific Publishing Company, Vol. 5, No. 4 (2006) 597-604.

[2] Godswill Chukwugozie Nsofor, **A Comparative Analysis Of Predictive Data-Mining Techniques**, University of Tennessee, Knoxville, August, 2006.

[3] Charu C. Aggarwal and Philip S. Yu, **Privacy-preserving Data Mining: Models and Algorithms**, Kluwer Academic Publishers, Boston, Dordrecht, London, ISBN: 978-0-387-70991-8, 2008, XXII, 514 p. 60 illus., Hardcover.

[4] Korosh Golnabi, Richard K. Min, Latifur Khan and Ehab Al-Shaer, **Analysis of Firewall Policy Rules Using Data Mining Techniques**, 1-4244-0143-7/06, IEEE 2006, USA.

[5] Benny Pinkas, **Cryptographic Techniques for Privacy-preserving Data Mining**, HP Labs, Volume 4, Issue 2, SIGKDD Explorations.

[6] Wenke Lee and Salvatore J. Stolfo, **Data Mining Approaches for Intrusion Detection**, Computer Science Department Columbia University 500 West 120th Street, New York, NY 10027, DARPA (F30602-96-1-0311) and NSF (IRI-96-32225 and CDA-96-25374).

[7] Yehuda Lindell and Benny Pinkas, **Privacy Preserving Data Mining**, Eshkol grant of the Israel Ministry of Science, M. Bellare (Ed.): CRYPTO 2000, LNCS 1880, pp. 3654, 2000, Israel,Springer-Verlag Berlin Heidelberg 2000.

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4

[8] Sheng Zhong, **Privacy, Integrity, and Incentive-Compatibility in Computations with Untrusted Parties**, Yale University, December 2004.

[9] Aman Jain and Bikash Sharma, **Privacy Preserving Data Mining**, Department of Computer Science and Engineering, National Institute of Technology, Rourkela, 10 May 2007.

[10] Rakesh Agrawal and Ramakrishnan Srikant, **Privacy-Preserving Data Mining**, IBM Almaden Research Center 650 Harry Road, San Jose, CA 95120, M SIGMOD , TX, USA ACM 1-58113-218-2/00/0005, 2000.

[11] Alexandre Evfimievski and Tyrone Grandison, **Privacy-Preserving Data Mining**, IBM Almaden Research Center, USA, 2009, IGI Global.

[12] Murat Kantarcioglu and Chris Clifton, **Privacy-preserving Distributed Mining of Association Rules on Horizontally Partitioned Data**, Senior Member, IEEE, 2003.

[13] Yehuda Lindell and Benny Pinkas, **Secure Multiparty Computation for Privacy-Preserving Data Mining**, Ministry of Science, Israel, 860/06, May 6, 2008.

[14] T. Y. Lin, T. H. Hinke, D. G. Marks and B. Thuraisingham, **Security and Data Mining**, USA, ISBN 0 412 72920 2, 1996, pp. 391-399.

[15] Chris Clifton and Don Marks, **Security and Privacy Implication of Data Mining**, 1996 ACM SIGMOD Workshop on Data Mining and Knowledge Discovery, MITRE Corporation, Department of Defense contract number DAAB07-96-C-E601.

[16] Vassilios S. Verykios, Elisa Bertino, Igor Nai Fovino Loredana Parasiliti Provenza, Yucel Saygin and Yannis Theodoridis, **State-of-the-art in Privacy Preserving Data Mining**, CODMINE IST FET Project IST-2001-39151, SIGMOD Record, Vol. 33, No. 1, March 2004.

[17] Chris Clifton, Murat Kantarcioglu, Jaideep Vaidya, Xiaodong Lin and Michael Y. Zhu, **Tools for Privacy Preserving Distributed Data Mining**, SIGKDD Explorations, Volume 4, Issue 2.

[18] Jiawei Han and Micheline Kamber, **Data Mining: Concepts and Techniques**, Second Edition, Elsevier Inc. 2006. [BOOK]

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-499-4 *References*

[19] Advantages and Disadvantages of Data Mining - xiangyun86.wordpress.com/2006/12/05/advantages-disadvantages-of-data-mining/

[20] Aman Sagar, Sanjeev Kumar, Palladium in Cryptography: The Advancement in Data Security, HCTL Open International Journal of Technology Innovations and Research, Volume 7, January 2014, ISSN: 2321-1814, ISBN: 978-1-62951-250-1.