



IGNITED MINDS
Journals

*Journal of Advances in
Science and Technology*

*Vol. VIII, Issue No. XVI,
February-2015, ISSN 2230-
9659*

**DATA SECURITY AND PRIVACY IN DATA MINING:
ISSUES AND CHALLENGES**

AN
INTERNATIONALLY
INDEXED PEER
REVIEWED &
REFEREED JOURNAL

Data Security and Privacy in Data Mining: Issues and Challenges

Shivali Yadav

Research Scholar, Jodhpur National University, Rajasthan

Abstract – Database mining can be defined as the process of mining for implicit, formerly unidentified, and potentially essential information from awfully huge databases by efficient knowledge discovery techniques. The privacy and security of user information have become significant public policy anxieties and these anxieties are receiving increased interest by the both public and government lawmaker and controller, privacy advocates, and the media. In this paper we focuses on key online privacy and security issues and concerns, the role of self-regulation and the user on privacy and security protections, data protection laws, regulatory trends, and the outlook for privacy and security legislation.

Keywords: Database Mining, Database Security, Data Privacy, Inferences, Intrusion Detection, Law

INTRODUCTION

Security and Privacy protection have been a public policy concern for decades. However, rapid technological changes, the rapid growth of the internet and electronic commerce, and the development of more sophisticated methods of collecting, analyzing, and using personal information have made privacy a major public and government issues. The field of data mining is gaining significance recognition to the availability of large amounts of data, easily collected and stored via computer systems. Recently, the large amount of data, gathered from various channels, contains much personal information. When personal and sensitive data are published and/or analyzed, one important question to take into account is whether the analysis violates the privacy of individuals whose data is referred to. The importance of information that can be used to increase revenue cuts costs or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data privacy is growing constantly. For this reason, many research works have focused on privacy-preserving data mining, proposing novel techniques that allow extracting knowledge while trying to protect the privacy of users. Some of these approaches aim at individual privacy while others aim at corporate privacy.

REVIEW OF LITERATURE:

Data mining, popularly known as Knowledge Discovery in Databases (KDD), it is the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. Knowledge discovery is needed to make sense and use of data. Though, data mining and knowledge

discovery in databases (or KDD) are frequently treated as synonyms, data mining is actually part of the knowledge discovery process. [1, 2, 3]

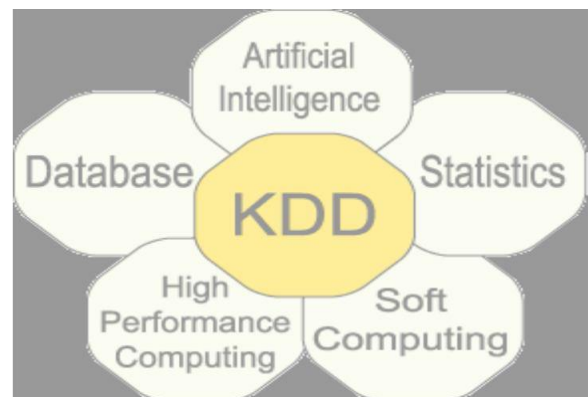


Figure -1

Usually, data mining e.g. data or knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful information from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases [4]. Although data mining is a comparatively new term but the technology is not. Companies have used powerful computers to filter through volumes of superstore scanner data and analyze market research reports for many years. Naturally such a process may open up new assumption dimensions, detect new invasion patterns, and raises new data security problems. Recent developments in information technology have

enabled collection and processing of enormous amount of personal data, such as criminal records, online shopping habits, online banking, credit and medical history, and driving records and almost important government the concerned data. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost.[5] Data mining, the discovery of new and interesting patterns in large datasets, is an exploding field. One aspect is the use of data mining to improve security, e.g., for intrusion detection. A second aspect is the potential security hazards posed when an adversary has data mining capabilities. Privacy issues have attracted the attention of the media, politicians, government agencies, businesses, and privacy advocates. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line. The databases and data warehouses become more and more popular and imply huge amount of data which need to be efficiently analyzed. Knowledge Discovery in Databases can be defined as the discovery of interesting, implicit, and previously unknown knowledge from large databases.[6,7] The data mining database may be a logical rather than a physical subset of your data warehouse, provided that the data warehouse DBMS can support the additional resource demands of data mining. If it cannot, then you will be better off with a separate data mining database. [8]

1. DATA MINING:

Data mining is an iterative and interactive process of discovering something innovative. The same as Novel something we are not aware, Valid- generalize the future,

Useful- some reaction is possible, Understandable- leading to insight, many step and process. Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques. Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner".[9]

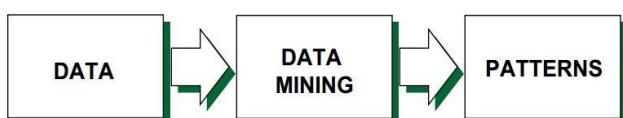


Figure -2

Data mining is an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to

address the issue of information extraction from large data bases".[10]

2. PRIVACY-PRESERVING DATA MINING ISSUES:

One of the key issues raised by data mining technology is not a business or technological one, but a social one. It is the issue of individual privacy. Data mining makes it possible to analyze routine business transactions and glean a significant amount of information about individuals buying habits and preferences. This particular research paper is considered the seminal work in PPDM research. Their research laid the foundation for future research that addresses privacy issues within a data mining context. They explain that the Internet has made data collection and data storage much easier, but the potential for misuse has also risen significantly. Data mining results can show models of aggregate data, but the model's accuracy depends on the quality of data. The authors raise the concern that any changes to data affect the accuracy and output of data mining models. Their approach to this problem allows the consumer to provide a perturbed value for sensitive attributes. This allows consumers to participate in the process and hopefully gives the consumer a sense of control over his or her own information. A major drawback of this approach is that output accuracy is lost during data mining activities. However, the authors maintain that small drops in accuracy are an acceptable trade-off for privacy. Finally, there is the issue of cost. While system hardware costs have dropped dramatically within the past five years, data mining and data warehousing tend to be self-reinforcing. The more powerful the data mining queries, the greater the utility of the information being gleaned from the data, and the greater the pressure to increase the amount of data being collected and maintained, which increases the pressure for faster, more powerful data mining queries. This increases pressure for larger, faster systems, which are more expensive.

3. DATA MINING FUNCTIONALITIES:

Data mining functionalities are used to specify the kind of patterns to be found in data mining tasks. In general, data mining tasks can be classified into two categories: descriptive and predictive. Descriptive mining tasks characterize the general properties of the data in the database. Predictive mining tasks perform inference on the current data in order to make predictions. In some cases, users may have no idea of which kinds of patterns in their data may be interesting, and hence may like to search for several different kinds of patterns in parallel.

- Concept description – characterization and discrimination
- Association – correlation and causality

- Classification and Prediction
- Cluster Analysis
- Outlier Analysis
- Trend and Evolution Analysis
- Other Pattern – direct or statistical analysis

In above first two functionalities involves first generalize, summarize and contrast data characteristics second association, multi-dimensional vs. single-dimensional association. Next two functionalities that is classification and prediction finding models that describe and distinguish classes or concepts for future prediction i.e. classify countries based on climate or classify cars based on gas mileage, presentation means decision-tree, classification rule, neural network and cluster analysis like class label is unknown – group data to form new classes, clustering based on the principle i.e. maximizing the intra-class similarity and minimizing the interclass similarity. Last three functionalities one is outlier analysis i.e. a data object that does not comply with the general behavior of the data, It can be considered as noise or exception but is quite useful in fraud detection, rare events analysis second is trend and evolution analysis i.e. trend and deviation by regression analysis, sequential pattern mining, periodicity analysis and similarity based analysis and last includes all other type of pattern-directed or statistical analysis [11]. Data mining represents a significant advance in the type of analytical tools currently available; there are limitations to its capability. One limitation is that although data mining can help reveal patterns and relationships, it does not tell the user the value or significance of these patterns. These types of determinations must be made by the user. A second limitation is that while data mining can identify connections between behaviors and/or variables, it does not necessarily identify a causal relationship. Successful data mining still requires skilled technical and analytical specialists who can structure the analysis and interpret the output. Data mining is becoming increasingly common in both the private and public sectors. Industries such as banking, insurance, medicine, and retailing commonly use data mining to reduce costs, enhance research, and increase sales. In the public sector, data mining applications initially were used as a means to detect fraud and waste, but have grown to also be used for purposes such as measuring and improving program performance.

CONCLUSION:

Data mining has become one of the key features of many homeland security initiatives. Often used as a

means for detecting fraud, assessing risk, and product retailing, data mining involves the use of data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. In the context of homeland security, data mining can be a potential means to identify terrorist activities, such as money transfers and communications, and to identify and track individual terrorists themselves, such as through travel and immigration records.

REFERENCES:

- [1] Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999.
- [2] Dunham, M. H., Sridhar S., "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, ISBN: 81-7758-785-4, 1st Edition, 2006
- [3] Fayyad, U., Piatetsky-Shapiro, G., and Smyth P., "From Data Mining to Knowledge Discovery in Databases," AI Magazine, American Association for Artificial Intelligence, 1996.
- [4] Larose, D. T., "Discovering Knowledge in Data: An Introduction to Data Mining", ISBN 0-471-66657-2, ohn Wiley & Sons, Inc, 2005.
- [5] L. Getoor, C. P. Diehl. "Link mining: a survey", ACM SIGKDD Explorations, vol. 7, pp. 3-12, 2005.
- [6] Fayyad U.M., Piatetsky-Shapiro G., Smyth P. "From Data Mining to KDD : An Overview", AAAI/MIT Press, 1996.
- [7] Han J. et Kamber M., "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, Canada, 2002.
- [8] Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999
- [9] David Hand, Heikki Mannila, and Padhraic Smyth," Principles of Data Mining", MIT Press, Cambridge, MA, 2001.
- [10] Peter Cabena, Pablo Hadjinian, Rolf Stadler, JaapVerhees, and Alessandro Zanasi, Discovering Data Mining: From Concept to Implementation, Prentice Hall, Upper Saddle River, NJ, 1998.

- [11] Clifton, C. and D. Marks, "Security and Privacy Implications of Data Mining", Proceedings of the ACM SIGMOD Conference Workshop on Research Issues in Data Mining and Knowledge Discovery, Montreal, June 1996.