HOW DATA MINING AFFECTS CONFIDENTIAL DATA

[1]Aashita Chhabra, [2]Teena K Bhatia
[1]Assistant Professor, [2]Assistant Professor
[1]Information Technology,
[1]Northern India Engineering College affiliated to GGSIPU, New Delhi, India

Abstract: In this paper we first look at data mining applications in safety measures and their suggestions for privacy. After that we inspect the idea of privacy and give a synopsis of the developments specifically on preserving private data after applying data mining on it. Later we examine some terms data mining, privacy and then study the relationship between them and its effect on the security of sensitive data. Further we then present an outline for research on confidentiality and data mining.

***Index Terms* – Data Mining, Inference Rules, Privacy Data, Fuzzy, Confidentiality**

INTRODUCTION

Data Mining is a technique for extracting data from the database and setting out the patterns as per requirements. Data Mining has got many applications like protecting data, analysing data etc. Now here comes up a question, whether to apply data mining on sensitive or private data or not? Will it affect privacy or not? Let us discuss about privacy first, everyone has their own perception about privacy, it's like hiding the information or keeping the confidential information at one place for example a Doctor never disclose complete information to the patient for his/her own benefits. Similarly if we need to apply data mining for counter terrorism occurring, one would like to gather data from various resources as data on revolutionary attack can be demanded at very least: who, what, when and where questions are being found. Other personal and business related information like place of birth, religion, education, ethnic origin, work history etc. All these data has to be included, warehoused and mined to get the complete information. It is being done to take pattern of possible terrorist and forecast their further activities and goods. But every coin has got two sides for e.g. If we are working as a Director of Big Bazaar Pvt. Ltd. and a company offers us the product at low price in return of our customer's database, obviously we would agree but if they would come to know about our competent products sale, It would cut down their product's sale which is ethically wrong.
Now a question arises what is the relationship between data mining and data privacy-Are they friends or rivals? It totally depends upon the use of the data or the person who is applying several data mining techniques over it. Even if the sensitive data is being mined it should be used for beneficial purpose and making a better co-operation between organizations. So in our paper we will discuss majorly three sections- first section contains the problems which arises regarding privacy; second section contains the outcomes of data mining over sensitive data and third section contains the solution of all the problems stated.

OBJECTIVES OF THE STUDY

- o To discuss various privacy problems related to security issues.
- o To find the major outcomes of the data mining over private data.
- o To discuss various solutions to the privacy problems related to security issues in various distributed environments.

PRIVACY PROBLEMS

There are many types of issues which comes up while applying data mining tools over private data. Some of them are discussed below

a. Some of the data is being used for malicious purpose, apart from the better outcomes of the organization.

b. Here the experts are using "Inference Rules" to predict the further proceedings of any execution. Suppose we have an information A implies B and another is B implies C so after applying inference rules, one can say that A implies C and can use it for their own benefit. It could be used in an illegal manner.

c. As association among the data is un-encrypted so it becomes easier for the tracker to make way out from the associated data.

d. Usually sensitive data is being linked with other source of data or are having references with many different database, which leads to loss of credibility of the organization.

OUTCOME OF DATA MINING OVER PRIVATE DATA

a. The contention is upcoming cyber-attacks in our nation, cyber security is concerned with protecting the nodes and the network systems against fraudulent viruses. Data mining fits into it for identifying these hacking activities. After analysing the data one will be able to find the "the needle in the haystack" or probably finding some needles from the millions of needles. For data mining to be affectively applicable in this scenario, it must be insured that the quality of technique applied must be of the highest standard.

b. To understand another application of data mining technique, let us assume that we are Walmart and another wholesale company offers us a product at a low price in return for our customers' database. In this case both the parties would obviously agree to the mentioned arrangement. But if the wholesaler was to find out about the competitors' products/database/strategies then it would affect the competitors sales which is ethically wrong.

c. The most crucial scenario to be discussed is about Finance i.e. majorly in banking sector. Suppose they provide data to insurance policy makers, they may have an access to their online banking transactions, which they can track out by using some basic tools and hence leads to fraud.

Hence both positive and negative aspects are there in the outcomes. It's about protecting the private data from the world of wrong people.

VARIOUS KEYS TO PRIVACY PROBLEMS

a. Limiting the access: If we apply control to access the data by the users. It will maintain the confidentiality of data and there will not be any misuse of the mined data. Hence the integrity of the data remains constant.

b. Rules applied on the inferred data should be kept in the fuzzy form. For example someone applies aggregation method or tries to cluster the data values one must not get the exact figure. It will be helpful for us by altering the data and keeping the update lock on the database.

c. We may apply different encryption techniques like AES, DES and tiny algorithm in order to represent the data in a more secure form.

d. Data can be kept in a distributed form and further applying various constraint rules to restrict the access to the sensitive data. Suppose a component of a document is private so after applying constraints if the document contain the information X which is private the document will also be private whereas association constraint leads to individual public documents used together privately only.

CONCLUSION

Privacy, in data mining, has no specific definition. What is private and what is not, is completely at the discretion of the organization and the nature of its activities. What is private for an organization A need not be private for an organization B. Also the degree to which privacy is applied in data mining depends upon the provisions given by various experts such as legal experts, social experts, social scientists etc. Here we conclude that privacy must be applied on data to some extent so as to avoid discrepancy among organizations.

REFERENCES

[1] Agrawal, R., Srikant, R.: Privacy-Preserving Data Mining. In: SIGMOD Conference, p.439–450 (2000)

[2] Agrawal, R.: Data Mining and Privacy: Friends or Foes. In: SIGKDD Panel (2003)

[3] Kantarcioglu, M., Clifton, C.: Privately Computing a Distributed k-nn Classifier. In: Bou-licaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) PKDD 2004. LNCS, vol. 3202,279–290. Springer, Heidelberg (2004)

[4] Kantarcioglu, M., Kardes, O.: Privacy-Preserving Data Mining Applications in the Mali-cious Model. In: ICDM Workshops, pp. 717–722 (2007)

[5] Liu, L., Kantarcioglu, M., Thuraisingham, B.M.: The applicability of the perturbation based privacy preserving data mining for real-world data. Data Knowl. Eng. 65(1), 5–21 (2008)

[6] Liu, L., Kantarcioglu, M., Thuraisingham, B.M.: A Novel Privacy Preserving Decision Tree. In: Proceedings Hawaii International Conf. on Systems Sciences (2009)

[7] Thuraisingham, B.: One the Complexity of the Inference Problem. In: IEEE Computer Se-curity Foundations Workshop (1990) (also available as MITRE Report, MTP-291)

[8] Thuraisingham, B.M.: Privacy constraint processing in a privacy-enhanced database management system. Data Knowl. Eng. 55(2), 159–188 (2005)

[9] Clifton, C.: Using Sample Size to Limit Exposure to Data Mining. Journal of Computer Security 8(4) (2000)

[10] Khan, S., Sharma, A., Zamani, A. S., &Akhtar, A. (2012). Data Mining for Security Purpose & its Solitude Suggestions.International Journal of Scientific & Technology Research,1(7), 1-4.

[11] Clifton, C., & Marks, D. (1996, May). Security and privacy implications of data mining. In ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery(pp. 15-19).