Exploitation of safety measures in Data Mining

Dr. Raman Chadha

Professor, Head(CSE)

CGC Technical Campus

Janjheri, Mohali

rc.cgctc@gmail.com

Harsimranjeet Kaur

Department of Computer Science

CGC Technical Campus

Janjheri, Punjab, India

harsimranjeet80@gmail.com

Er. Ramandeep Sandhu

Department of Computer Science

CGC Technical Campus

Janjheri, Punjab, India

rds.cgctc@gmail.com

Abstract
In this paper we have paying attention a variety of techniques, approaches and different areas of the research which are helpful and marked as the significant field of data mining Technologies. As we are aware that many MNC's and large organization are operate in different places of the different countries. Each place of operation may generate large volumes of data. Corporate decision makers require access from all such sources and take strategic decisions. Data Mining is a process of finding useful information from a set of relational databases and to use that information for our business or other purposes. Security is very important for data mining applications. We have also summarized the results in the tabular form. For Example Data Mining can be used to detect the buying pattern of the customers of the grocery stores and the stall of products can be arranged according to their buying style and pattern like on

people tends to buy more stuff in weekends so higher stocks then regular days should be maintained in the weekends.. In this paper we have discussed various security measure and performance of those measures on the data mining process.

Index Terms—Information, Security, Data Mining, Data mining task, Data mining life cycle , Visualization of the data mining model , Data mining Methods, Data mining applications.

I. INTRODUCTION

In the computer science world the biggest threat is the security of our information i.e. how to protect our information from outer threats. In IFIP conference the idea was given how to secure the data mining process and how to protect the future databases and legality of those databases. So, that large organization can securely extract their secret information from their large data warehouse and that help them to focus in achieving their goal [1].Data mining has many applications in security including in national security (e.g., surveillance) as well as in cyber security (e.g., virus detection)[2]. The threats to national security include attacking buildings and destroying critical infrastructures such as power grids and telecommunication systems. Data mining techniques are being used to identify suspicious individuals and groups, and to discover which individuals and groups are capable of carrying out terrorist activities. Cyber security is concerned with protecting computer and network systems from corruption due to malicious software including Trojan horse and viruses [3]. Fraud detection is important to the organization because companies and suppliers of services lose a significant proportion of their revenue as a result. Moreover, the modeling and characterization of users' behavior

in telecommunications can be used to improve security, improve services, provide personalized applications, and optimize the operation of electronic equipment and/or communication protocols. Another difficulty is that fraud detection problems involve huge data sets which are constantly evolving. Data sets can be as large as tenths of thousands of calls per weekday for a large organization with 3 or 4 thousand employees, to hundreds of millions of calls for national carriers. One should also consider the size of the related metadata [4]. The data mining tools can answer the business questions which are traditionally very complicated task and take too much time to analyze and produce the result. Data Mining is relatively young field with the existence of term Big Data it is now used extensively. Big Data means large amount of data which is caused by the integration of databases. When large amount of data sets are there no simple statistical technique can be used thus Data Mining comes into play. Data mining is the process of getting interesting patterns from large amounts of data, which could be stored in databases, data warehouses, or other information repositories.

II. SECURITY CONCERN IN DATA MINING

Data mining is used to extract the hidden information .Data mining, the discovery of new and interesting patterns in large datasets, is an exploding field. Recently there has been a realization that data mining has an impact on One aspect is the use of data mining to improve security, like in intrusion detection. A second aspect is the potential security hazards posed when an adversary has data mining capabilities. If we consider an example of a college database records are stored in regular as well as in secured database. Regular database can be viewed to faculty and students to check the data . Secured

database viewed only to admin and can modify the data when required.

The requirements of data mining security concerned with the following traits:

A. Physical Database Integrity

This physical database integrity concerned with the data is read from and write to the disk .Data can be lost due to some problems occur like system crash, power failure, etc .we cannot retrieve the information. If we go to example, Due to power failure of a university/college, server prevents to access the data. We cannot retrieve the information.

B. Logical Database Integrity

Managing the logical database integrity is essential in modern information system. It has been the responsibility of the database server to manage the logical integrity. Some rules are too complex and even impact on the performance. From the example in the class records two students with the name "Aman". This is most likely a logical duplication. For logical integrity, Setting up a new module or reconfiguring an existing module.

C. Auditability

The modification of records and fields of the database are taken with OLTP (Online transaction processing applications) and by the human operators or by database administrator. The date, time, fields, records and the previous value of the records should have to be recorded under log files [3]. Student data in the college database must be traceable and correct so if audited no error should be found.

D. User Authentication

Database management system requires the user authentication. Without valid user identification number and password the database does not allow the user to do anything on data items of database. Each user has its own user authentication and identification entity. The user has to keep its user ID and password secret and unauthorized user cannot access. Like in the example stated above, Individual student data in the college database can be accessed by only that student by login into database by its username and password. Classification is a technique in which a particular type of class of an item is identified by help of multiple attributes described about that class. We can build up an idea about the type of customer, item or object that is being described in those attributes. Like by the help of data about the people age, salary, gender etc you can easily classify those into different age groups and also can determine about the type of items they like to buy like teenagers tends to buy Jeans and T-Shirts while working age group buys formal wears that they could wear into their offices etc. This method can be applied with different attribute to any kind of product.

III. SECURITY MEASURE & PERFORMANCE

Data mining is associated term with database and data warehouse. A data warehouse is built by the help of relational database. The information extracted from the Data Warehouses is very confidential and important for the organization. Due to its importance it should be protected from outer threats that why we use security measures to protect our information. These type of security measures are based on the characteristics of data mining:

A. Privacy

Data Privacy is maintained by various integrity rules. It is mandatory to maintain the privacy of the organizations data so that it could not be used by other for their benefits. Large organizations have to train their employees time to time to make them understand the benefits of the employees like Student Data of CSE department should be accessed by only CSE department not by all other departments. This will be taken as departmental privacy. We can also take it to other levels. Like student level, faculty level and administration level which will give on the data department wise as well as user level wise.

B. Sensitivity

The Data Warehouse consist of all kind of information about the organization, in all those data some important data should be only accessible to few important people because of the highly confidentiality of the data. Various constraint measures and policies have to be made about the access of the data in the database. Like Sensitive data like contact number and address information should only be accessed by authenticated users who have the privileges to access that information.

C. Data Correctness

Data entered into the databases should be first ensured that it is correct. There are certain filters that ensues the correct entry of the database and correct entry of data ensures correct output. Input of the wrong data will lead to wrong output result that is why it is important that data should be entered correctly into database. Like if students in their college registration form will provide false details

about them then there will be false result about them in every database query.

D. Data Integrity

Integrity of data is important to remove redundant data and also from security point of view. Integrity is applied by applying various constraints on the database which helps to link the tables of databases and also to improve security in the database. It should be provisioned that once a constraint is applied anyone would not be able to remove that constraint. Like student data can be integrated in Library and Academic as well as Accounts department by giving its Unique Id as Foreign key in their respective databases.

E. Elimination of False Matches

False matches are the wrong output of a query entered in a database. False matches are prevented by the automatic filtering so that accidently any confidential information is not leaked. If the automated filtering is not applied it should be applied by manually and policies should be maintained for them. Like facility data, filters are applied on them students cannot see their data, if they want.

The mentioned security measures for databases of data warehouse for data mining applications for extraction useful information summarized in table. Summarization of different security measures for data mining is:

Security Levels Ranks			
Ranks	One	Two	Three
Level	Highest	Medium	Lowest

Table 1: Different Security Level Ranks

Security Measures	Requirements
Privacy	Two
Sensitivity	One
Correctness	Two
Integrity	One
False Matches	Three

Table 2: Security measures and their requirements.

We have assigned the Ranks of the security measures according to the security needed in those measures to protect the information from outer threats. In rank one the security is highest and mandatory for the measure, like in Sensitivity and Integrity security measures. In rank two security is high but not mandatory like in privacy and correctness of data. In third the security is no mandatory and is also low like in false matches.

F. Sequential Pattern technique

Sequential pattern technique is helpful in identifying trends in similar events. The trends are identified by analyzing the past data of those events. The long termed data of regular occurring similar events are saved and analyzed to determine a sequential pattern. Like in websites that sell products saves the data about the buying pattern of an individual customer and when next time they log in according to their past selection of items, similar kind items are displayed first to them which are according to their choice based on past events. This kind of display of product helps to increase sale on these website and also saves time of the customer.

IV. PROPOSED WORK:

We have studied the different papers about the different aspects of security measures and the level of security required on them.

Our proposed idea is that "Log Files" should be monitored daily. In a college database three modules are present Student, Admin and Faculty. Security constraints are least in student module and Admin has the highest security constraints. If a student is able to hack into an admin account that student can access all kind of data in the college database and the admin will able to find about the hacked account after a long time when information in huge amount get corrupt or lost from the database. The information either gets lost permanently or get corrupt for some time. Thus to prevent this kind of information lost admin should devise a process to monitor Log Files daily so that any unauthorized login can be detected and the hacked account can be recovered or deleted timely.

V. CONCLUSION

Security in Data Mining is very important because by the help of data mining the knowledge possessed from Data Warehouses is used to make strategic decisions for future of the organization. This makes that information very important because the future of the organization depends upon that and it should be protected at any cost. We have proposed an idea to help to protect that important information from outer threats and also from the treats within the organization in this paper.

REFERENCES

[1] Abdelsalam, H. et al., 2001. Drishti: An Integrated Navigation System for Visually Impaired and Disabled. IEEE Fifth International Symposium on Wearable Computers. P 149.

[2] PC World "http://www.pcworld.com/article/2083503/wearable-tech-at-ces-2014-prepare-your-body-parts-for-an-onslaught-of-options.html#tk.nl_today"

[3] IHS Electronics and Media "http://www.ihs.com/pdfs/Wearable-Technology"

[4] Wikipedia "http://en.wikipedia.org/wiki/Wearable_technology"

[5] Sung, M., DeVaul, R., Jimenez, S., Gips, J., and Pentland, A., "A Shiver Motion and Core Body Temperature Classification for Wearable Soldier Health Monitoring Systems" 9th IEEE International Symposium of Wearable Computers (ISWC'04),

[6] Murray, J., "Wearable computers in battle: recent advances in the Land Warrior system," Proceedings of Fourth International Symposium on Wearable Computers,

[7] TechCrunch"http://techcrunch.com/2014/02/14/state-of-the-early-wearables-world/"

[8] E. Rehmi Post, Margaret Orth "Smart Fabric, or "Wearable Clothing". First International Symposium on Wearable Computers. 13-14 Oct. pp 167-168. 1997

[9] Olalekan S. Akinola International Journal of Computer Science and Telecommunications [Volume 3, Issue 8, August 2012]

[10] Data Mining Tasks techniques and Applications by Youngjian Fu

[11] Data Mining Techniques: http://www.ibm.com/developerworks/library/ba-data-mining-techniques