International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 09, September 2016
www.ijates.com
ijates
ISSN 2348 - 7550

# STUDY ON APPLICATION OF DATA MINING AND SECURITY MEASURES

**[1]D.Gandhimathi, [2] Al-Rehnaz and [3]Anupama Sudeep**

*[1] Assistant Professor, Department of Computer Science and Applications and Software Systems, Sri Krishna Arts And Science College Sugunapuram, Kuniyamuthur, Coimbatore, Tamil Nadu, (India)*

*[2,3] BSc students , Department of Computer Science and Applications and Software Systems, Sri Krishna Arts And Science College Sugunapuram, Kuniyamuthur, Coimbatore, Tamil Nadu, (India)*

## ABSTRACT

*In this huge world of computer and technologies. There is major threat on protection of our data. In order to protect this data we use the technique called data mining. Data mining is a process to search the data from a huge data base for different work. It helps in the protection of data like the passwords and all of social websites online banking etc. In this paper we have discourse on various technique, approaches, application architecture and security measures for the data mining.*

***Keywords: Cyber Security, Data Exploration, Data Security, Data Stockpiles, Rule Indication.***

## I. INTRODUCTION

Assuring the integrity of computer networks, that are both in relation to security and also with regard to the institutional life of ones domain in common, is really a growing concern. Security and defense networks, proprietary research, intellectual property, and market mechanisms based on data that depend on unhampered and undistorted access, can all be severely compromised by malicious intrusion. Data mining is an interdisciplinary subfield of science. It is the computational process of exploring patterns in large sets of data involving methods at the intersection of artificial_intelligence, machine study, statistics, and database_systems. [1]

Data mining, *the extraction of hidden predictive information from ample databases*, is an assertive new technology with great realm of possibility to help companies focus on the most significant enlightenment in their data stockpile. Data mining tools foresee future trends and behaviours, allowing pursuit to make proactive, knowledge-driven decisions. The overall goal of the data mining process is to extract knowledge from a data set and revolutionize it into a comprehensible structure for further use. Aside from the raw examination step, it involves database and data management forms, data pre-processing, model and inference considerations, interestingness metrics, complexity deliberations, post-processing of invented structures, visualization, and online updating. Data mining is a very relevant process where potentially useful and previously unexplored enlightenment is derived from huge volumes of data. There are a number of components committed in the data

mining process. These factors constitute the architecture of a data mining system The development of enlightenment Technology has generated large amount of databases and huge data in various areas. The exploration in databases and enlightenment technology has given rise to an approach to store and mould this precious data for further decision making. [2]

Data mining is a process of extraction of useful enlightenment and patterns from huge data. It is also called as knowledge revelation process, knowledge mining from data, knowledge extraction or data /pattern examination. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by deliberate tools typical of decision support systems. Data mining tools can answer pursuit questions that traditionally were too time consuming to intention. They burnish databases for hidden patterns, finding predictive enlightenment that experts may miss because it lies outside their prospects. Most companies already collect and refine huge quantities of data. Data mining techniques can be carried out rapidly on existing software and hardware podium to enhance the value of current information resources, and can be combined with new products and systems as they are brought on-line.[3][4].

## II. DATA MINING ARCHITECTURE

The considerable factors of any data mining system are data source, data stockpiles server, data mining engine, pattern assessment broadness, graphical user interface and knowledge base.

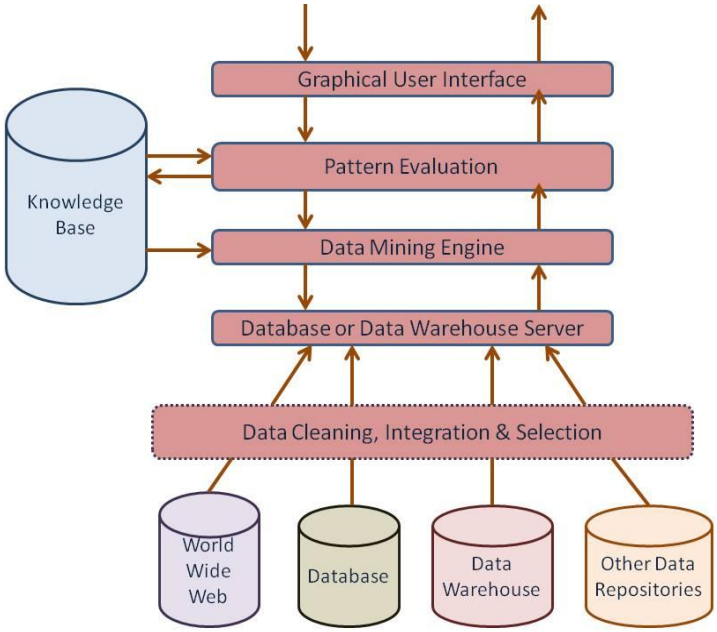


**Fig 1: Architecture of data mining**

### 2.1 Data Sources

Database, data stockpile, World Wide Web (WWW), text files and other documents are the actual sources of data. You need enormous amount of historical data for data mining to be successful. Organizations usually store

data in databases or data stockpile. Data stockpile may contain one or more databases, text files, spreadsheets or other kinds of information archive. Sometimes, data may dwell even in plain text files or spreadsheets. World Wide Web or the Internet is another big source of data.

## 2.2 Contrasting Processes

The data needs to be cleaned, integrated and selected before passing it to the database or data stockpile server. As the data is from different sources and in different formats, it cannot be used precisely for the data mining mechanism because the data might not be complete and reliable. So, first data needs to be cleaned and unified. Again, more data than required will be collected from different data sources and only the data of concern needs to be preferred and passed to the server. These processes are not as simple as we think. A number of techniques may be predetermined on the data as part of cleaning, combination and selection. [5]

### 2.2 Database or Data Warehouse Server

The database or data stockpile server contains the actual data that is ready to be processed. Hence, the server is liable for restoring the relevant data based on the data mining request of the user.

### 2.3 Data Mining Engine

The data mining engine is the core element of any data mining system. It consists of a number of modules for carrying out data mining tasks along with association, classification, characterization, clustering, prediction, time-series analysis etc.

### 2.4 Pattern Evaluation Modules

The pattern evaluation module is mainly responsible for the measure of pleasantness of the pattern by using a brink value. It interacts with the data mining engine to focus the search towards fascinating patterns.

### 2.5 Graphical User Interface

The graphical user interface module communicates betwixt the user and the data mining structure. This module helps the user use the system easily and efficiently without knowing the real intricacy behind the process. When the user define a query or a task, this module interacts with the data mining system and displays the result in an easily understandable manner.[6]

### 2.6 Knowledge Base

The knowledge base is advantageous in the whole data mining process. It might be useful for guiding the exploration or evaluating the interestingness of the result patterns. The proficiency base might even contain user acceptance and data from user experiences that can be profitable in the course of data mining. The data mining engine might get inputs from the knowledge abominable to make the result more authentic and reliable. The pattern evaluation module interacts with the knowledge base on a regular basis to get inputs and also to update it.

To best apply these extreme techniques, they must be fully integrated with a data stockpile as well as flexible interactive business study tools. Many data mining tools currently operate outside of the stockpile, requiring

extra steps for extracting, importing, and analysing the data. Furthermore, when new insights require operational implementation, integration with the stockpiles simplifies the application of results from data mining. The following analytic data stockpile can be applied to improve pursuit processes throughout the organization, in areas such as advertising campaign management, fraud detection, new product rollout, and so on. The ideal starting point is a data stockpile containing a combination of internal data tracking all customer contact coupled with external market data about competitor activity. Background information on ability of customers also provides an excellent basis for prospecting. This stockpiles can be implemented in a variety of relational database systems: Sybase, Oracle, Redbrick, and so on, and should be improved for flexible and fast data access. An OLAP (On-Line Analytical Processing) server enables a more sophisticated end-user pursuit model to be applied when navigating the data stockpiles. The multidimensional structures allow the user to evaluate the data as they want to view their pursuit – summarizing by product line, region, and other key perspectives of their pursuit. The Data Mining Server must be integrated with the data stockpiles and the OLAP server to embed ROI-focused pursuit analysis directly into this infrastructure. An advanced, process-centric metadata template defines the data mining targets for specific pursuit issues like campaign management, prospecting, and promotion optimization. Integration with the data stockpiles enables operational decisions to be directly implemented and tracked. As the stockpiles grows with new decisions and results, the organization can continually mine the best practices and apply them to coming decisions.

This design represents a fundamental shift from conventional decision support systems. Rather than simply distributing data to the end user through query and reporting software, the Advanced Analysis Server applies users' pursuit models directly to the warehouse and returns a proactive analysis of the most relevant information. These denouements enhance the metadata in the OLAP Server by providing a dynamic metadata layer that represents a clear view of the data. Reporting, visualization, and other examination tools can then be applied to plan future actions and confirm the impact of those plans. [7][8]

## III. DATA MINING APPLICATIONS

Here is the ballot of areas where data mining is enormously used −

- Financial Data Analysis
- Retail Industry
- Telecommunication Industry
- Biological Data Analysis
- Other Scientific functions.
- Intrusion Detection

### 3.1 Financial Data Analysis

The financial data in banking and financial industry is broadly reliable and of high quality which facilitates systematic data analysis and data mining. Some of the classical cases are as follows

- Designing and construction of data stockpiles for multidimensional data analysis and data mining.

- Loan payment cost and customer credit policy analysis.
- Classification and clustering of customers for aspire marketing.
- Detection of money laundering and other financial delinquency.

**3.2 Retail Industry**

Data Mining has its great Appositeness in Retail Industry because it collects large amount of data from on sales, customer purchasing history, goods transportation, consumption and services. It is natural that the number of data collected will continue to expand briskly because of the increasing ease, availability and popularity of the web.

Data mining in retail industry assist in identifying customer buying patterns and bias that lead to improved quality of customer service and good customer retention and satisfaction. Here is the list of examples of data mining in the retail corporation

- Design and Construction of data stockpiles based on the advantages of data mining.
- Multidimensional analysis of sales, customers, products, time and region.
- Determination of efficiency of sales campaigns.
- Customer Retention.
- Product recommendation and cross-referencing of items.

**3.3 Telecommunication Industry**

At present the telecommunication corporation is one of the most emerging industries providing different services such as fax, pager, cellular phone, internet messenger, images, e-mail, web data transmission, etc. Due to the development of modern computer and communication technologies, the telecommunication industry is briskly expanding. This is the reason behind data mining becoming very essential to help and understand the business.

Data mining in telecommunication corporation helps in identifying the telecommunication patterns, catch fraudulent activities, make better use of resource, and amend quality of service. Here is the arrangement of examples for which data mining improves telecommunication aids −

- Multidimensional Analysis of Telecommunication data.
- Fraudulent pattern analysis.
- Identification of uncommon patterns.
- Multidimensional association and sequential patterns analysis.
- Mobile Telecommunication aids.
- Use of visualization tools in telecommunication data analysis.

### 3.4 Biological Data Analysis

In current times, we have seen a tremendous augmentation in the field of biology such as genomics, proteomics, functional Genomics and biomedical research. Biological data mining is a very essential part of Bioinformatics. Following are the arrangements in which data mining shares for biological data analysis −

- Semantic integration of heterogeneous, distributed genomic and proteomic databases.
- Adjustment, indexing, similarity search and comparative analysis multiple nucleotide sequences.
- Exploration of structural patterns and evaluation of genetic networks and protein pathways.
- Association and path analysis.
- Visualization tools in genetic data analysis.

### 3.5 Other Scientific Applications

The practice discussed above tends to handle comparatively small and homogeneous data sets for which the sophistically techniques are adapted. Huge amount of data have been collected from scientific authority such as geosciences, astrographs, etc. An ample pile of data sets is being accomplished because of the fast numerical refreshments in various fields such as climate and ecosystem modelling, chemical engineering, fluid dynamics, etc. These are the appositeness of data mining in the field of scientific operation −

- Data stockpile and data pre-processing.
- Graph-based excavating.
- Perception and domain specific knowledge.

### 3.6 Intrusion Detection

Intrusion credits to any kind of action that abuse integrity, confidentiality, or the availability of network assets. In this world of connectivity, safety has become the dominant issue. With elevated usage of internet and possibility of the tools and tricks for encroach and attacking network prompted intrusion detection to become a cavillous component of network administration. Here is the list of field in which data mining technology may be adapted for intrusion detection −

- Development of data mining process for intrusion detection.
- Association and alternation analysis, collection to help select and build discriminating aspect.
- Analysis of Stream data.
- Distributed data mining.
- Visualization and query tools.[9]

### 3.7 Trends in Data Mining

Data mining approach are still evolving and here are the current trends that we get to see in this field

- Appositeness Exploration
- Scalable and interactive data mining approach
- Integration of data mining with database systems, data stockpiles systems and web database systems.
- Evenness of data mining query language.

- Visual data mining.
- New mechanism for mining elaborate types of data.
- Biological data mining.
- Data mining and software engineering.
- Web mining.
- Distributed data mining.
- Actual time data mining.
- Multi database data mining.
- Privacy protection and information protection in data mining.

## IV. CONCLUSION

This paper has discourse about data mining for security applications. We first started with a discussion of data mining architecture then we moved onto application of data mining in several real life entities. Information security (sometimes shortened as InfoSec) is the conveyance of protecting information from unauthorized user, disclosure, disruption, modification or abolition. Computer and communication systems again and again suffer requirements Information security technology is an imperative component for protecting public and private computing infrastructures. Betterment in technology is making people more determine towards frequent use of information technology resulting in more usage of online assets which in turn is giving rise to a large number of security bluff to these resources [10]

## REFERENCES

[1] Soumen Bhowmik, Rabisankar Chattopadhyay& Uddalok Chatterjee had discussed on the topic review on data mining 1,3Assistant Professor, Department of Computer Science & Engineering, Bengal Institute of M.Tech Student, Department of Computer Science & Engineering, Bengal Institute of Technology & Management, Santiniktan .

[2] Dileep Kumar Singh, Vishnu Swaroop IT Resource Centre Madan Mohan Malaviya Engineering College, Gorakhpur, India .Dept. of Computer Science & Engineering, Madan Mohan Malaviya Engineering College IT Resource Centre Gorakhpur, India

[3] Bhavani Thuraisingham, Latifur Khan, Mohammad M. Masud, Kevin W. Hamlen The University of Texas at Dallas

[4] Waldemar Wójcik and Konrad Gromaszek Lublin University of Technology Poland

[5] Osmar R. Zaïane, 1999 Department of Computing Science University of Alberta

[6] Shakir Khan, Dr. Arun Sharma, Abu Sarwar Zamani, Ali Akhtar INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 1, ISSUE 7 AUGUST 2012

[7] Mrs. Bharati M. Ramageri, Lecturer Modern Institute of Information Technology and Research, Department of Computer Application, Yamunanagar, Nigdi Pune, Maharashtra, India-411044. Bharati M. Ramageri / Indian Journal of Computer Science and Engineering

[8] Aarti Sharma, Rahul Sharma,Vivek Kr. Sharma,Vishal Shrivatava Aarti Sharma et al, / (IJCSIT) International Journal of Computer Science and information Technologies, Vol. 5 (2) 2014, 2023-2025

[9] Peter Kramme Martin Šeleng Ondrej Habala and Ladislav Hluch Institute of Informatics, Slovak Academy of Sciences, Dubravska cesta 9845 07 Bratislava, Slovakia

[10] Preeti Aggarwal CS/IT, KIIT College of Engineering Gurgaon, India

[11] M. M. Chaturvedi SET, Ansal UniversitySector-55, Gurgaon