Sibi.R *et al*, International Journal of Computer Science and Mobile Applications,
Vol.5 Issue. 12, December- 2017, pg. 30-35 **ISSN: 2321-8363**
Impact Factor: 5.515

A SURVEY ON VARIOUS TECHNIQUES IN DATA MINING

Sibi.R[1], Valarmathi.V[2]
[1]*Research Scholar, Sri Krishna Arts and Science College, rsibi.rks@gmail.com*
[2]*Asst.Professor, Department of IT, Sri Krishna Arts and Science College, valarskasc1@gmail.com*

ABSTRACT

In data mining various techniques are supported to form the strong security. The security is major purpose for data now-a-days. Data mining is extensively used for knowledge discovery from large databases. A fundamental challenge is to develop privacy and security methods appropriate for data mining. It is a large dataset. Privacy is one of the most important properties of an information system must satisfy, in which systems the need to share information among different, not trusted entities, the protection of sensible information has a relevant role. The various techniques that used for privacy purpose are privacy preserving data mining, cryptography, sensors in DM, cloud computing. This paper addresses the privacy problem survey by considering the privacy and algorithmic requirements simultaneously. In this survey, how the data is secured and sensing in various technologies.

Keywords: Privacy Preserving, Security, Sensing, Cryptography, Cloud.

1. INTRODUCTION

Data mining technology is born out of these requirements of handling avalanche knowledge automatically to urge perceptive pattern that will be useful to the organisation doing this technique or person whose experiments generate such information. Processing is presently wide utilized in many areas additionally as science, business, politics, nuclear and astrophysics[6]. Privacy protective processing could be a crucial property that any mining system ought to satisfy. The common definition of privacy at intervals the crypto logic community limits the information that is leaked by the distributed computation to be the information that will be learned from the chosen output of the computation. Above all, tho' the parties perceive that combining their info has some mutual profit, none of them is willing to reveal its information to the opposite party[1].

Sensors notice the data regarding an object with a sensing element and regarding the environment of the thing, and report it to a number. As for knowledge stream mining, frequency analysis and association rule mining have in the main been conducted[3].Cloud computing may be a quite Internet-based computing that uses the net and central remote servers to take care of knowledge and applications. Cloud computing permits customers and businesses to use applications while not installation and access their personal files at any pc with web access [4]. The cryptography protocols would modify secure communications by addressing the authentication [5].

2. PRIVACY PRESERVING DATA MINING

Explosive progress in networking, storage and processor technologies has junction rectifier to the creation of extremist massive info that record unexampled quantity of transactional data. There are, however, completely different levels of adversarial behaviour. Privacy conserving protocols area unit designed so as to preserve privacy even within the presence of adversarial participants that arrange to gather data concerning the inputs of their peers. Privacy conserving protocols area unit designed so as to preserve privacy even within the presence of adversarial participants that arrange to gather data concerning the inputs of their peers. If the party is semi-honest then we are able to assume that this range is so random [1].

Privacy-Preserving data processing of Association Rules from Outsourced group action Databases technique is developed with associate degree secret writing theme. Attack ready to establish the intricacies of the rule preservation and information item property supports aren't true supports. To Secure Association Rules, Secure Multi-party Computation (SMC) rule is introduced to cover the association rules in an exceedingly horizontally distributed database [5].

2.1. PPDM COMPUTATION

Here, we discuss the various computation techniques in data mining which is using for data:

Classification:

John has a private database D1 and Steve has private database D2. How can John and Steve build a decision tree based on D1-D2 without disclosing the contents of their private database to each other? Several algorithms like ID3, Gain Ratio, Gini Index and many other can be used for Decision Tree[1].

Data Clustering:

John has a private database D1 and Steve has private database D2. John and Steve want to jointly perform data clustering on D1-D2. This is primarily based on data clustering principle that tries to increase intra class similarity and minimize interclass similarity[1].

Mining Association Rules:

Let John has a private database D1 and Steve has private database D2. If John and Steve wish to jointly find the association rules from D1-D2 without revealing the information from individual databases[1].

Data Generalization, Summarization and Characterization:

Let John has a private database D1 and Steve has private database D2. If they wish to jointly perform data generalization, summarization or characterization on their combined database D1-D2, then this problem becomes an Secure Multiparty Communication problem[1].

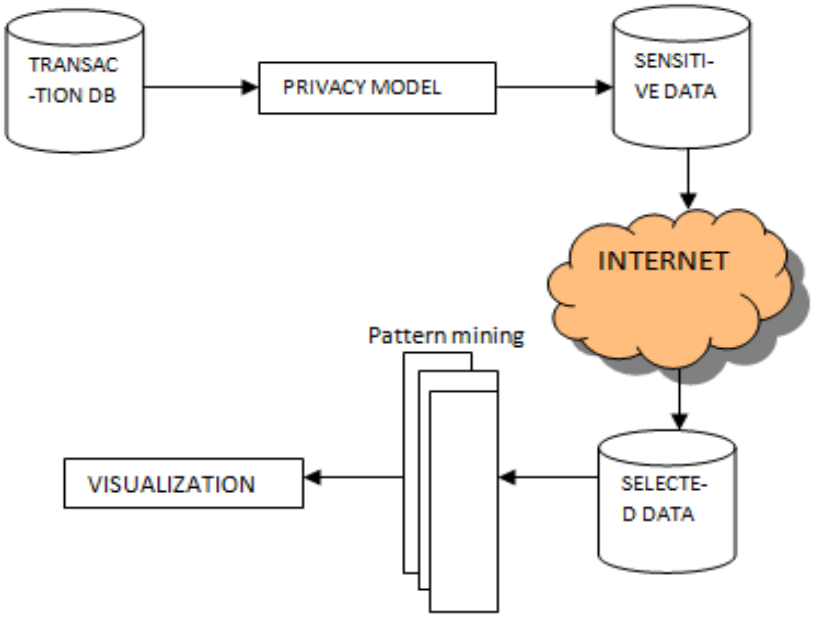
Profile Matching:

John has a database of hacker's profile. Steve has recently traced a behaviour of a person, whom he suspects a hacker. Now, if Steve wants to check whether his doubt is correct, he needs to check John's database. John's database needs to be protected because it contains hacker's related sensitive information. Therefore, when Steve enters the hacker's behaviour and searches the John's database, he can't view his whole database, but instead, only gets the comparison results of the matching behaviour[1].

Fraud Detection:

Two major financial organizations want to cooperate in preventing fraudulent intrusions into their computing system, without sharing their data patterns, since their individual private database contains sensitive data [1].

2.2. SYSTEM ARCHITECHTURE:



3. SECURITY

Data mining technology is born out of those needs of handling avalanche of information mechanically to get perceptive pattern that may be helpful to the organisation doing this method or person whose experiments generate such data. A pattern may be an easy knowledge outline, an information segmentation, or a model of dependencies inside the info. As a information discovery process is meant to steer all means from information to 'documented knowledge'. Data mining could be a powerful suggests that of extracting helpful info from knowledge. With the rise of simple availableness of digital knowledge, the potential for misuse of raw additionally as deep-mined knowledge grows. A basic challenge is to develop privacy and security ways applicable for data processing [6].

4. SENSING

Scheme of sensing information for economical information stream mining the setting that takes under consideration the frequency analysis and association rule mining of information stream. Sensors do sensing work a set interval and report the sensing data to a bunch. During this case, they need the factors on whether or not signals are reportable or not, and filter information in accordance with the factors. The method is termed information filtering. during this analysis, sensing information was encoded within the ways in which of presenting the options of the events detected by sensors and of as well as the changes of the sensing information stream[3].The planned theme during this thesis reflects solely the modification of the worth detected by a detector, however presents the worth as well as the kind and alter of a detector. To method the info from multiple sensors, it's necessary to convert them into symbols or numbers in an exceedingly totally different system reflective the options and properties of the measured values.

4.1 SENSOR DATA MINING MODEL:

In this section, we first define various definitions for sensor data mining model and then discuss sensor data mining model for the sensor data. Let denote the sensor S and sensor node SN. The sensor is defined like following Definitions [9].

[Definition 1]

S= s Є sensor, SN= n Є sensor Node for the sensor data mining, we have a tendency to outline the device category to pick out the devices of specific sort like temperature sensor, humidness device, etc. The device category is consisted of location and sensing sort. Let denote sensing sort ST. The ST is outlined like following Definition a pair of.

[Definition 2]

ST = t Є sensingTypeSpec Let denote a specific sensor type St. the St is outlined like following Definition three.

[Definition 3]

St = s Є S∩ST ^ t.sensingTypeName = {a,b, …, t Є ST} Let denote sensing area Sl. The foreign terrorist organization is outlined like following Definition four.

[Definition 4]

Sl = s Є S∩SN ^ l.sensorNodeLocation = {a, b, …, l Є SN} Let denote a specific sensing area and sensor type S l , t. The S l, t is outlined like Definition five.

[Definition 5]

Sl .t = s□ Є S l ∩St Time interval specifies specific time point recorded in sensor database or specific time point which will be recorded in the future. Let be denote the amount T. The T is outlined like Definition vi.

0, snapshot

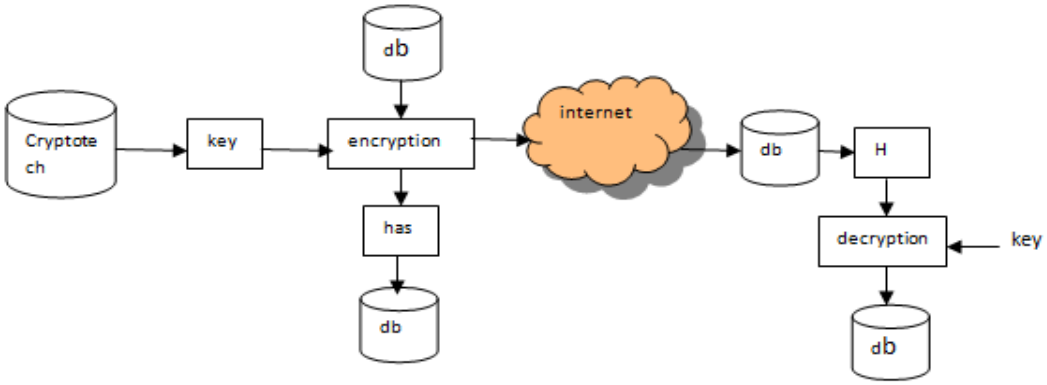
[Definition 6]

a>a,0 // Time window

5. CRYPTOGRAPHY

The cryptanalytic system principally supported the privacy conserving data processing. The encryption/decryption theme is that the method of reworking dealing information TD into its sensitive information TD'. It's been encrypted by victimization cryptography techniques for every plain item. The privacy conserving module, there's hash functions are used. It's accustomed economical storage and quick retrieval of things.

The main use of hash operate is maps n keys to n integers[5].



5.1 ASSOCIATION RULE MINING:

The association rule mining finds attention-grabbing association or correlation relationships among an oversized set of information things. The association rules area unit thought of attention-grabbing if they safety each a minimum support and a minimum confidence. Let I= be a collection of things. Let R be the task-relevant information, be a collection of database transaction where each transaction T could be a set of things such T⊆I. every group action T is related to associate degree symbol TID. Let A be the set of things. An association rule is implication of the shape A⇒B, where A⊆I and B⊆I and A∩B=ϕ. The rule A⇒B holds in the transaction set R with support S, where S is the percentage of transactions in R that contain A∪B. This is taken to be the probability P(A∪B). The rule A⇒B has confidence C in the transaction set R if C is the percentage of transactions in R containing A that also contain B. this can be taken to be the contingent probability P (A/B). That is,

$$\text{Support } (A \Rightarrow B) = P\ (A \cup B)$$

$$\text{Confidence } (A \Rightarrow B) = P\ (A \cap B)$$

A set of things is remarked as associate degree item set (pattern). Associate degree item set that contains k-items could be a k-item set. As an instance the set could be a 2-itemset. Associate degree item set satisfies minimum support if the prevalence frequency of the item set is larger than or adequate to the merchandise of minimum support and therefore the total variety of transactions in R. the quantity of transactions needed for the item set to satisfy the minimum support count. If associate degree item set satisfies the minimum support, then it's aforementioned to be frequent item set[5].

6. CLOUD

As cloud computing is penetrating additional and additional all told ranges of business and scientific computing, it becomes a good space to be centered by data processing. Cloud Computing denotes the new trend in web services that believe clouds of servers to handle tasks. In cloud computing is that the process of extracting structured info from unstructured or semi-structured net information sources. The information mining in Cloud Computing permits organizations to centralise the management of code and data storage, with assurance of economical, reliable and secure services for his or her users.

The information mining in Cloud Computing permits organizations to centralise the management of code and data storage, with assurance of economical, reliable and secure services for his or her users. Data mining techniques and applications square measure a great deal required within the cloud computing paradigm. As Cloud computing refers to code and hardware delivered as services over the net, in Cloud computing data processing code is additionally provided during this way[8].

7. LITERATURE WORK

This chapter offers an introduction to the assorted data processing techniques popularly used. Models like Privacy conserving, Security, Sensing, Cryptography, and Cloud area unit delineate well. necessary analysis works distributed victimization these models area unit reviewed during this survey. Privacy conserving area unit terribly straightforward to interpret and work quicker. Cloud computing manage missing values and categorical values terribly with efficiency. Cryptography insists that their numeric knowledge ought to be commonly distributed. From the discussions, it may be complete that for data processing one will not notice one classifier that outperforms each alternative however rather one can notice a classifier that performs well for a specific domain.

This section presents the comparative analysis of various data processing techniques and algorithms that are utilized by most of the researchers in data processing. a quick outline of those data processing algorithms with their deserves and demerits are mentioned. The comparative study of classification algorithms like privacy conserving, cryptography and cloud.

8. CONCLUSION

This paper provides a more current evaluation and updates of various techniques analysis research available. Literatures have been reviewed based on the different aspects of various techniques analysis. This paper studies the application of techniques and concepts of data mining for various techniques analysis, and reviews the related literature about data mining. The various techniques include privacy preserving data mining, security, sensing, cryptography, cloud computing. Above all the techniques privacy preserving is considered most powerful than others, which is "securing with the multi-party computation". However, there are many challenging in this research field to be resolve with privacy preserving data mining.

REFERENCES:

1. D. Beaver, S. Micali and P. Rogaway, The round complexity of secure protocols, Proc. of 22nd ACM Symposium on Theory of Computing (STOC), pp. 503-513, 1990.

2. M. Bellare and S. Micali, Non-Interactive Oblivious Transfer and Applications, Advances in Cryptology - CRYPTO '89. Lecture Notes in Computer Science, Vol. 435, Springer-Verlag, 1997, pp. 547-557.

3. Deepti Mittal, DamandeepKaur, AshishAggarwal, "Secure Data Mining in Cloud using Homomorphic Encryption" IEEE 2014 Cloud Security.

4. S.M. Mahajan and A. K. Reshamwala," Data Mining Ethics in Privacy Preservation - A Survey", in International Journal of Computer Theory and Engineering, Vol. 3, No. 4, August 2011.

5. SunandaRavindran , ParsiKalpana, "Data storage security using partially Homomorphic Encryption in cloud", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, April 2013.

6. Teo S, Cao J., and Lee, DAG: A Model for Privacy Preserving Computation, IEEE, International Conference on Web Services, (2015).

7. DimitriosKarapiperis and Vassilios S. Verykios, Member, IEEE, "An LSH-Based Blocking Approach with a Homomorphic Matching Technique for Privacy-Preserving Record Linkage", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 4, APRIL 2015.

8. S. SelvaRatna , Dr. T. Karthikeyan, "Survey on recent algorithms for privacy preserving data mining", S.SelvaRathna et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (2) , 2015, 1835-1840.

9. Chun-Wei Lin, Tzung-Pei Hong and Hung-Chuan Hsu, " Reducing Side Effects of Hiding Sensitive Itemsets in Privacy Preserving Data Mining", Hindawi Publishing Corporation, the Scientific World Journal, Volume 2014, Article ID 235837 April 2014.

