Design Engineering

ISSN: 0011-9342 | Year 2021
Issue: 8 | Pages: 15852-15862

Implementation of Novel Association Rule Hiding Algorithm Using FLA with Privacy Preserving in Big Data Mining

[1]M. Kiran Kumar, [2]Dr. Pankaj Kawad Kar

[1]Research Scholar, Dept. of CSE, Sri Satya Sai University of Technology and Medical Sciences.
ORCID: 0000-0001-9604-6799, Mail- kirann.intell@gmail.com

[2]Associate Professor, Sri Satya Sai University of Technology and Medical Sciences.

Article Received: April 2021; Accepted: Aug 2021; Published online: Sep2021.

ABSTRACT: Association rule mining is the identification of unclear knowledge links, but it is also devastating defense. To solve this problem, we should easily cover the rules of associations in order to protect fine-tuned rules of association. Different procedures are suggested to mask the rules on associations, but several procedures reduce the trust values below the threshold values specified. In addition, there is no current system for processing large data in the parallel world. In addition to removing a collection of objects, the coming data items have a significant issue. We also used the fluid logic approach in the present work to hide mining practice from the great conditions of mining. This will attempt to decrease a sensitive rule's unwanted effect on sensitive data sets laws. The architecture suggested contains characteristics such as parallelism and scalability, which enables large data processing. Implementation of following:

- To protecting the sensitive principles of affiliation by changing the data collection
- A creative way of hiding the sensitive laws of association is suggested for studying.
- The HSARH analysis hides all the sensitive rules for those objects (the "Heuristic Sensitive Association Rule Hiding").
- The chosen sensitive rules are preserved to investigate GBSARH ("Genetic Based Sensitive Association Rule Hiding"). Per repetition, it hides a single law. A new fitness feature is established to reduce the effect of the association law.

Keywords: Association Rule Hiding, Fuzzy Logic Approach, Big Data.

INTRODUCTION

In our daily lives, various new technologies, such as mobile phones, social networking, and the Internet of Things (IoT), combine intelligent world practices such as smart terminals, trustworthy transportation, vibrant cities, and others, which generate massive knowledge. The various types of electronic devices constantly generate vast details on each character and location. Single, total and complicated details, in particular huge material, then become a great deal of value. In addition, the potential benefits of the large-scale awareness generated are growing dramatically including improving information interpretation provided by artificial intelligence and information processing techniques, which includes assessing the ability to calculate aid across the Internet. Big data is also aimed at increasing fertility flows in this conference. In addition, existing safety tests must lead to

ISSN: 0011-9342 | Year 2021
Issue: 8 | Pages: 15852-15862

large data collection and the square measure of the need for simultaneous victimization for a comprehensive examination of facts. Secrecy problems then Aggravated square behavior will only be restored instead of a mass-type as a function of dispersed results. In one of the most important data processing methods, opening of group work is harmony. However, abuse of this approach could cause sensitive details about people to be revealed. Most relevant of these shared activities are often worn out by various forms of science, which ensures that they distinguish things from sensible legislation. Unfortunately, the effect of the features available is obvious. People function and take dynamic forms to justify this drawback. Although such plans should not ensure that the right response is not just found and the power improved. During this study, anonymization techniques are used to shield fragile controls rather than to mask fine group practicalities in large information collection. The unwanted feature of deleting several item sets (ISs) towards new immigration details should be disconnected by rendering the motion sensor control information. Parallels also quantifiability alternatives square measurement of thinking to help shape this direction suited as a broad knowledge analysis. The sensitive line in an organization's legislation decides on appropriate uses of victimization which can be facilitated by anonymization.

Association rule mining is the (and most promising) method used in the area of data mining. It has provided a number of possibilities for mechanical data since its launch. It enables citizens to relate to shared goals and offers new possibilities for improvement in public health and medicine; one example is the estimation of recovery after transplantation. This approach lets consumers locate search records and contributes to useful programs. Privacy, though, remains a key topic.

Data sanitization strategy for privacy protection is divided into four categories: boundary, accurate, evolutionary and heuristic. The frontier strategy establishes the updated positive and negative limits of all popular products. It just depends on the weight of the maximum or supportive border in order to reduce the support of the new negative border. By forming a hybrid algorithm named the decrease of the trust rule (DCR) based on the Maximin method, the performance of the border-based algorithm had been improved. As the name implies, the maximin method utilizes two heuristics to conceal the association principle in order to monitor the sanitizing process by identifying the victim objects identified with the maximin solution for results on subsequent output. This is done by removing the shortest duration victim piece. Hai at el. was largely faced with the intersection of a frequent method for hiding laws of associations. These algorithms are intended to disguise in three phases a particular collection of sensitive rules. The first phase sets a set of articles which fulfill three conditions: I containing the sensitive rule on the right, (ii) a maximum sub-item set of a maximum set and (iii) having the minimum support of those articles listed in (ii). The object is marked as a victim item on the right side of the responsive law related to the listed maximum assistance item. The second is to calculate the amount of confidential transactions. Thirdly, the victims' objects shall be removed from the transactions to a minimum level of trust under the rules. Finally, both transactions are given a number to measure the impact of the hiding data mechanism on the non-sensitive association rules (NSARs).

PRIVACY PRESERVING DATA MINING

Matwin (2013) has recently carefully studied and explored the importance of privacy-preserving data management strategies. The usage of specific techniques has shown that they are able to discourage the unfair use of data mining. Any approaches indicated that every stigmatized community could not

ISSN: 0011-9342 | Year 2021
Issue: 8 | Pages: 15852-15862

be more concerned about generalizing data than the population as a whole. Vatsalan et al. (2013) examined the 'PRRL' methodology for the linking of datasets to organizations through the safeguarding of privacy. In order to analyze them in 15 dimensions, a taxonomy focused on PPRL methods is therefore suggested. Qi and Zong (2012) have overviewed many available privacy mining strategies based on data sharing, manipulation, mining algorithms and hidden data or regulations. With respect to the dissemination of data, a few algorithms are currently employed on centralized and distributed data for privacy security. In order to acquire joint data mining while maintaining intact private data between shared partners, Raju et al. (2009) recognized the need to incorporate or multiply protocol dependent, homomorphic encryption along with the current definition of the Digital Envelope technique. In various implementations, the methodology suggested showed significant impact.

The latest cloud services privacy protecting approach, focused on advanced cryptographic elements, was analyzed by Malina and Hajnye (2013) and Sachan et al. (2013). The solution included anonymous entry, the right to unlink and the secrecy of data transferred. Finally, this solution is used, experimental findings are collected and efficiency comparisons are carried out. In the sense of privacy preservation features and the right to preserve the same relation in other areas, Mukkamala and Ashok (2011) contrasted a series of fuzzy mapping procedures. This comparison shall be subject to: (1) the four front changes in the fuzzy function definition, (2) the introduction of seven ways of integrating different functional values of a specific data object to a single value, (3) the use of many similarity metrics to compare the initial data and mapped data.

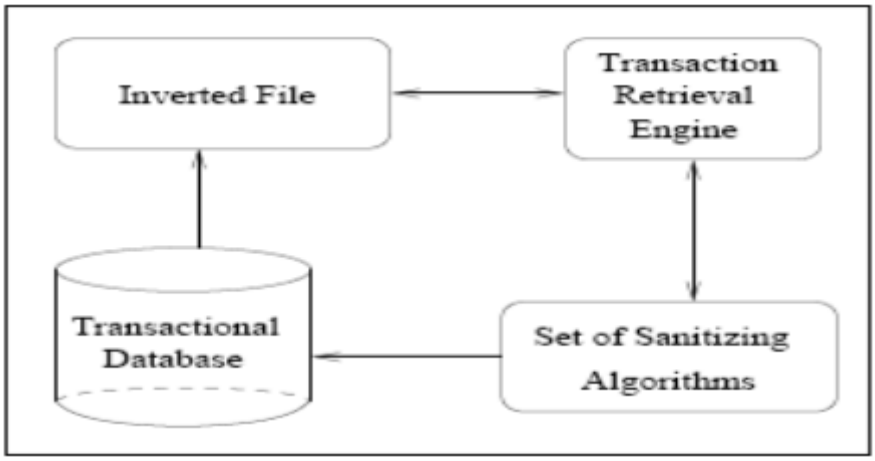

Figure 1 Privacy Preservation Classification Techniques

ASSOCIATION RULE MINING

Data mining is an advanced, common way of discovering the fascinating relationships between variables in broad datasets. Different measures of interest are used to analysise and display the laws contained in databases. Authors in alliance rules adopted for the detection of high-scale transaction data between items registered in supermarket point-of-sale (POS) systems.

Apriori is a classic learning algorithm for data mining. It is intended for work with transaction databases, such as collections of products purchased by clients and information on the frequency of the website. The issue is described as follows of association rule mining:

ISSN: 0011-9342 | Year 2021
Issue: 8 | Pages: 15852-15862

Let $I = \{i1, i2, ..., in\}$ be a set of n features named items. Let $D = \{t1, t2, ..., tn\}$ Be a database-listed series of transactions. Each transaction in D comprises a single ID and a subset of I elements.

An indication of the shape is known as a rule $X \rightarrow Y$ where $X, Y \subseteq I$. The collections of objects (in short items) X and Y are referred to as the antecedent (left hand or LHS) and the resulting (right or RHS) of the law.

The definition is shown by a brief illustration from the mobile store. The set of items is **I= {JIO SIM, LYF, Mobile case, 16GB Memory Card}** And there is a limited database of objects (1 is current, and 0 represents absence of item) as seen in the following Table1.2. The following is a small database. A mobile shop example rule could be

{JIO SIM, LYF} => {Mobile case}

The aim of the rule is to buy a customer mobile case if JIOSIM and LYF are purchased.

This is a really limited case. A law requires several hundred transactions in realistic implementations to be called statistics and sometimes thousands or millions of transactions can be included in data sets.

PRIVACY PRESERVING ALGORITHMS

HEURISTIC-BASED TECHNIQUES

A variety of techniques for many techniques such as sorting, relationship rule discovery and clustering have been built on the grounds that selective data alteration or sanitization is an NP-hard problem and therefore heuristics may be used to deal with complexity problems.

Confusion over centralized data perturbation-based association rules

An optimum sanitization is a systematic evidence of an NP Hard challenge in the exploration of association rules for hiding sensitive big articles collections. The following is the basic issue that was dealt with in this work. If D is the root database, R is a collection of important association rules that can be exploited by D, and Rh is an array of rules in R. How do we convert the D_ database into the published database D_, so that all R laws, except the Rh rules, also have to be mined from D_? The heuristic proposal to modify the data was focused on data interference; in fact, the process involved changing a chosen range of 1-value to 0-values, such that the support for sensitive laws is reduced to a maximum value for the usefulness of the published database. The usefulness of this work is calculated as the amount of un-sensitive laws hidden from the side effects of the method of data alteration. The sanitation of big sensitive items to sanitizing sensitive laws is then extended to include later jobs. The methods used in the work were either to avoid the sensitive rules by covering their repeated item sets, or to reduce the trust of the sensitive rules by taking them below the user-specified threshold. Both tactics have contributed to three methods to hide delicate laws. The main feature about this respect was the ability to convert a 1-value to a 0-value and a 0-value to a 1-value in the binary database. This versatility in data alteration had the side effect that a non-frequent law might become popular apart from the secret norms of non-sensitive association. This is what we call the 'ghost law.' Since critical rules are concealed, both hidden, non-sensitive rules and frequent (ghost rules) rules are seen as less useful than the database posted. Therefore, the heuristics used for this subsequent work should be more sensitive to utility problems as safety is not affected.

Design Engineering
ISSN: 0011-9342 | Year 2021
Issue: 8 | Pages: 15852-15862

RESULTS AND DISCUSSION

Hiding Failure

This calculation calculates the relationship between the sensitive rules in the updated dataset. It is known as a fraction of the sensitive association rules derived from the changed data set which are separated from the original data set according to the sensitive associating rules.

Currently,

$$\text{Hiding Failure HF} = |SR\ (D')|/|SR\ (D)| \quad \text{- (7.1)}$$

$|SR\ (D')|$ - The changed dataset D' is detected by responsive rules

$|SR\ (D)|$ - In the initial dataset D the sensitive rules appear.

The hiding loss could, if necessary, be 0 percent.

Misses Cost

This calculation calculates the ratio of non-sensitive rules concealed as a hiding influence. The perishable law is calculation. The following is calculated:

$$\text{Misses Cost}\ \ MC=|NSR(D)|-|NSR(D')|\ /\ |NSR(D)| \quad \text{- (5.2)}$$

NSR(D) - Set in the initial dataset D with all non-sensitive laws

NSR (D) - Sets in the updated data collection with all non-sensitive laws.

Artifactual Pattern

The ratio of the laws discovered which are objects should be calculated. It tests the side effects of the fantasy law. The following is calculated:

Artifactual Pattern $AP =|AR'|-|AR \cap AR'|\ /\ |AR'|$ (7.3)

AR - Association laws used in the initial D dataset

AR' - Set of rules of association included in D.

Dissimilarity

By contrasting their histograms, this calculation quantifies the discrepancy between the initial and the sanitized data sets. The horizontal axis displays their respective frequencies in the dataset and vertical axis. The following is calculated:

$$\text{Dissimilarity Diss}(D,D') = \frac{1}{\sum_{i=1}^{n} fD(i)} X \sum_{i=1}^{n} [fD(i) - fD'(i)] \quad (7.4)$$

$fX(i)$ - the i-th dataset X frequency

n - Number in initial dataset D of different objects

ISSN: 0011-9342 | Year 2021
Issue: 8 | Pages: 15852-15862

Dataset	No. of Transactions	No. of Items	Avg. Length	Type
Retail	88,162	16,470	10.30	Sparse
Mushroom	8,124	119	23	Dense
Pumsb	49,046	2,113	74	Dense
Chess	3196	75	37	Dense
Connect	67,557	129	43	Dense

Table1DataSourceCharacteristics

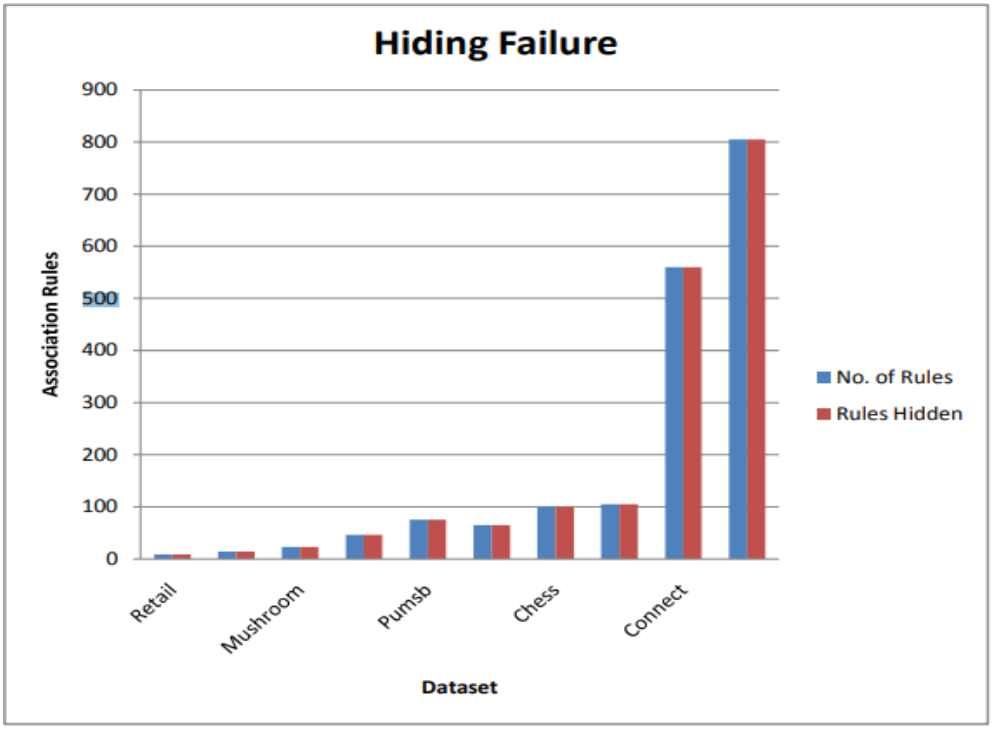

Figure 2 Hiding Failure

ISSN: 0011-9342 | Year 2021
Issue: 8 | Pages: 15852-15862

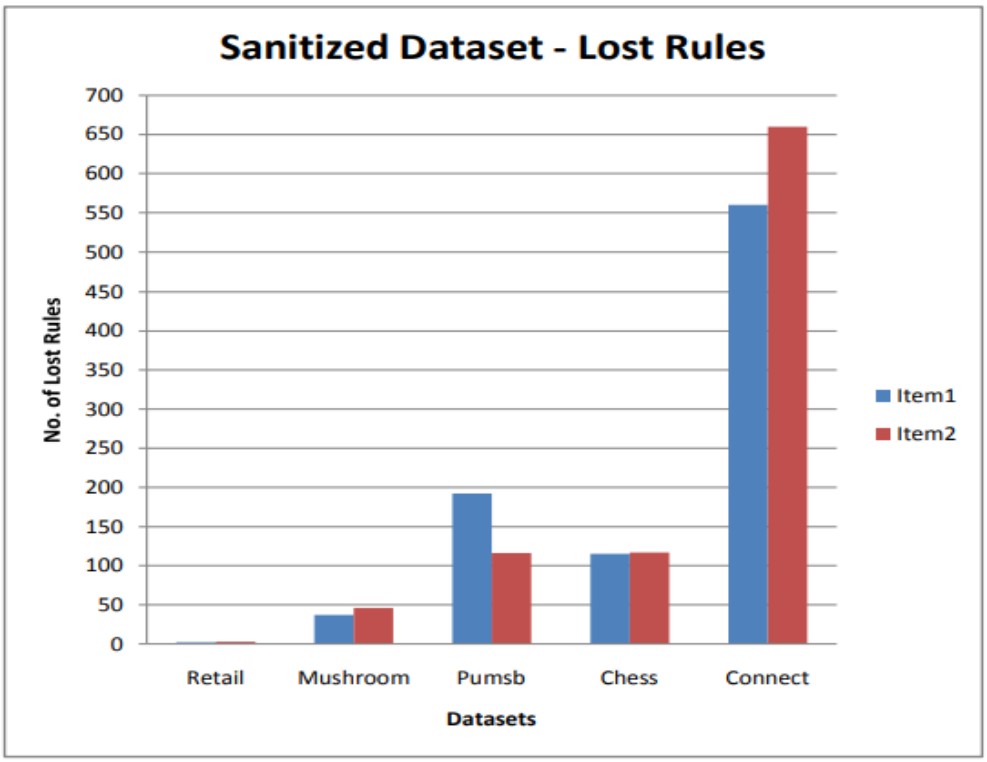

Figure 3 Lost Rules in HSARH

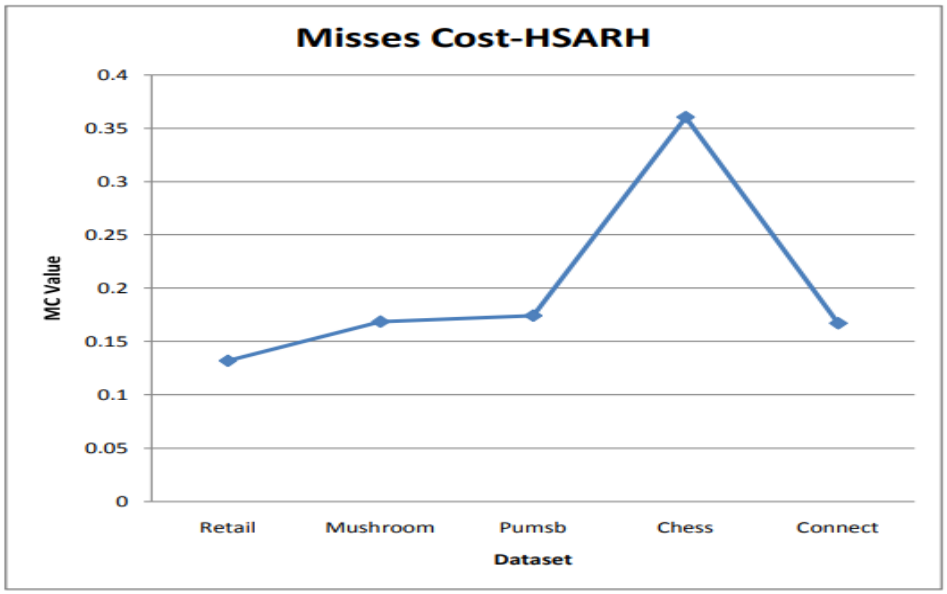

Figure 4 Misses Cost in HSARH

ISSN: 0011-9342 | Year 2021
Issue: 8 | Pages: 15852-15862

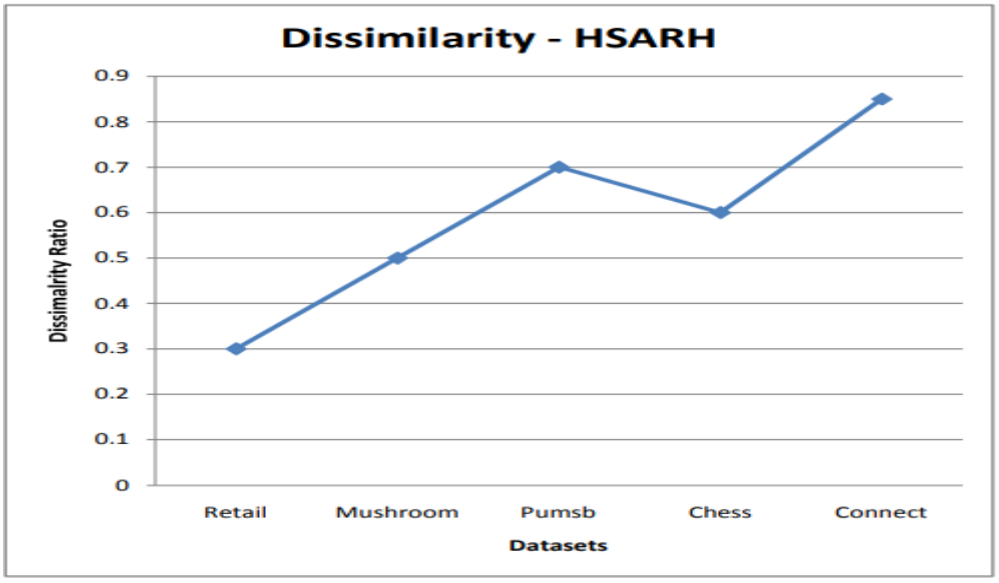

Figure 5 Dissimilarity in HSARH

Comparison of HSARH and existing technologies

The approach is comparable to familiar approaches, such as ISL, DSR and EHSAR, to show the utility of the proposed methodology. These current strategies include the input of objects and modify the dataset by the laws. ISL technology expands support if the object is on the rule's LHS. If the item is on RHS, DSR can lower help. EHSAR switches things with the law of representative association.

Datasets	Sensitive Items	No of Rules	No. of Rules Hidden			
			ISL	DSR	EHSAR	Proposed Method
Retail	41	9	6	8	9	9
	48	14	11	13	14	14
Mushroom	63	23	18	22	23	23
	39	46	38	12	15	46
Pumsb	4940	75	71	72	75	75
	7092	65	56	62	65	65
Chess	36	100	89	95	100	100
	40	105	96	101	105	105
Connect	85	560	545	552	560	560
	109	805	784	796	805	805

Table 2Hiding failure comparison in HSARH to existing methods

ISSN: 0011-9342 | Year 2021
Issue: 8 | Pages: 15852-15862

Experimental Results

The missed side effects in GBSARH are seen in Figure 7.7. Different datasets appear on the X axis. The cumulative number of missing registers is calculated and displayed in Y axis from the sanitized data collection. It is obvious that there are less lost laws in the supermarket data set.

The GBSARH side effects are shown in Figure 7.8. The various databases are shown in X-Axis. In the sanitized dataset, the amount of ghost rules is computed and shown in the Y axis. Mushroom and supermarket data sets have no fantasy law. When opposed to the amount of missing rules in both databases, there are less ghost rules.

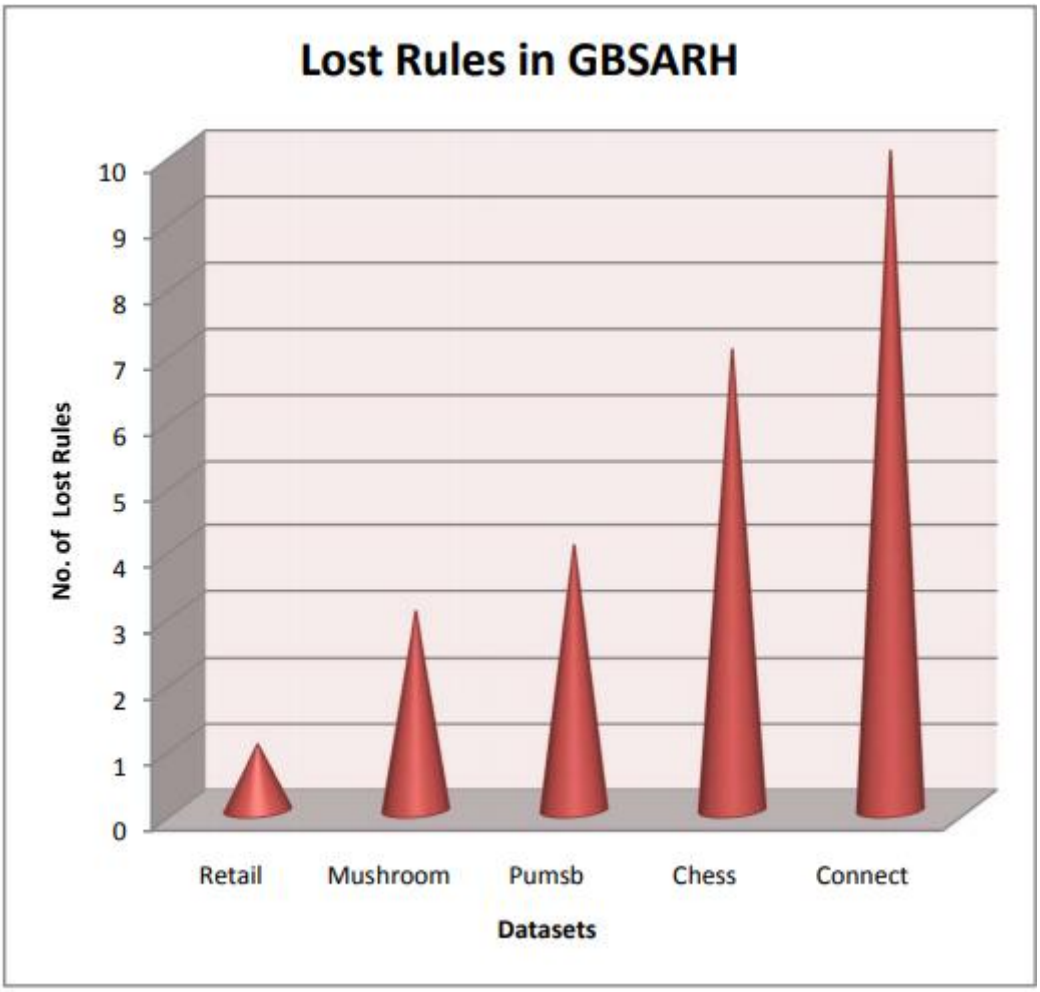

Figure 6 Lost Rules in GBSARH

CONCLUSION

This research work summarizes the contributions and provides recommendations for potential work in this portion.Data mining is a new study guideline for data mining and computational databases. Data preservation The Rule Hiding Association is one of the data mining privacy protection activities. The key goal is to create algorithms for changing the initial dataset such that even during the mining phase the private information is kept private. In the hiding phase, side effects may occur. The rules that are not responsive are lost or new rules are created. To ensure the usability of the data collection the hiding system has to take account of side effects.

ISSN: 0011-9342 | Year 2021
Issue: 8 | Pages: 15852-15862

REFERENCES

1. Abul, O., Atzori, M., Bonchi, F., & Giannotti, F. (2007, October). Hiding sensitive trajectory patterns. In Data Mining Workshops, 2007. ICDM Workshops 2007. Seventh IEEE International Conference on (pp. 693-698). IEEE.
2. Aggarwal, C. C., & Philip, S. Y. (2008). A general survey of privacy-preserving data mining models and algorithms. In Privacy-preserving data mining (pp. 11- 52). Springer US.
3. Aggarwal, C. C., & Philip, S. Y. (2008). Privacy-preserving data mining: A survey. In Handbook of database security (pp. 431-460). Springer US.
4. Aggarwal, C. C., & Yu, P. S. (2007, April). On privacy-preservation of text and sparse binary data with sketches. In Proceedings of the 2007 SIAM International Conference on Data Mining (pp. 57-67). Society for Industrial and Applied Mathematics.
5. Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB (Vol. 1215, pp. 487-499).
6. Agrawal, R., &swami A. (1993, June). Mining association rules between sets of items in large databases. In Acmsigmod record (Vol. 22, No. 2, pp. 207-216). ACM.
7. Amiri, A. (2007). Dare to share: Protecting sensitive knowledge with data sanitization. Decision Support Systems, 43(1), 181-191.
8. Atallah, M., Bertino, E., Elmagarmid, A., Ibrahim, M., &Verykios, V. (1999). Disclosure limitation of sensitive rules. In Knowledge and Data Engineering Exchange, 1999.(KDEX'99) Proceedings. 1999 Workshop on (pp. 45-52). IEEE.
9. Belwal, R. C., Varshney, J., Khan, S. A., Sharma, A., & Bhattacharya, M. (2008, October). Hiding sensitive association rules efficiently by introducing new variable hiding counter. In Service Operations and Logistics, and Informatics, 2008. IEEE/SOLI 2008. IEEE International Conference on (Vol. 1, pp. 130- 134). IEEE.
10. Berberoglu, T., & Kaya, M. (2008, May). Hiding fuzzy association rules in quantitative data. In Grid and Pervasive Computing Workshops, 2008. GPC Workshops' 08. The 3rd International Conference on (pp. 387-392). IEEE.
11. Bertino, E., Fovino, I. N., &Provenza, L. P. (2005). A framework for evaluating privacy preserving data mining algorithms. Data Mining and Knowledge Discovery, 11(2), 121-154.
12. Bononi, L., Bracuto, M., D'Angelo, G., &Donatiello, L. (2005, June). Concurrent replication of parallel and distributed simulations. In Principles of Advanced and Distributed Simulation, 2005. PADS 2005. Workshop on (pp. 234-243). IEEE.
13. Borhade, S. S., & Shinde, B. B. (2014). Privacy preserving data mining using association rule with condensation approach. International Journal of Emerging Technology and Advanced Engineering, 4(3), 292-296.
14. Brijs T., Swinnen G., Vanhoof K., and Wets G. (1999), The use of association rules for product assortment decisions: a case study, in: Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining, San Diego (USA), August 15-18, pp. 254-260. ISBN: 1-58113-143-7.
15. Brin, S., Motwani, R., Ullman, J. D., &Tsur, S. (1997, June). Dynamic itemset counting and implication rules for market basket data. In ACM SIGMOD Record (Vol. 26, No. 2, pp. 255-264). ACM.

ISSN: 0011-9342 | Year 2021
Issue: 8 | Pages: 15852-15862

16. K. Bhargavi. An Effective Study on Data Science Approach to Cybercrime Underground Economy Data. Journal of Engineering, Computing and Architecture.2020;p.148.
17. S. Jessica Saritha. AN EFFICIENT APPROACH TO QUERY REFORMULATION IN WEB SEARCH, International Journal of Research in Engineering and Technology. 2015;p.172.
18. K BALAKRISHNA,M NAGA SESHUDU,A SANDEEP. Providing Privacy for Numeric Range SQL Queries Using Two-Cloud Architecture. International Journal of Scientific Research and Review. 2018;p.39
19. K BALA KRISHNA, M NAGASESHUDU. An Effective Way of Processing Big Data by Using Hierarchically Distributed Data Matrix. International Journal of Research.2019;p.1628
20. P.Padma, Vadapalli Gopi,. Detection of Cyber anomaly Using Fuzzy Neural networks. Journal of Engineering Sciences.2020;p.48.
21. Clifton, C., Kantarcioglu, M., & Vaidya, J. (2002, November). Defining privacy for data mining. In National Science Foundation Workshop on Next Generation Data Mining (Vol. 1, No. 26, p. 1).
22. Sehgal.P, Kumar.B, Sharma.M, Salameh A.A, Kumar.S, Asha.P (2022), Role of IoT In Transformation Of Marketing: A Quantitative Study Of Opportunities and Challenges, Webology, Vol. 18, no.3, pp 1-11
23. Kumar, S. (2020). Relevance of Buddhist Philosophy in Modern Management Theory. Psychology and Education, Vol. 58, no.2, pp. 2104–2111.
24. Roy, V., Shukla, P. K., Gupta, A. K., Goel, V., Shukla, P. K., & Shukla, S. (2021). Taxonomy on EEG Artifacts Removal Methods, Issues, and Healthcare Applications. Journal of Organizational and End User Computing (JOEUC), 33(1), 19-46. http://doi.org/10.4018/JOEUC.2021010102
25. Shukla Prashant Kumar, Sandhu Jasminder Kaur, Ahirwar Anamika, Ghai Deepika, MaheshwaryPriti, Shukla Piyush Kumar (2021). Multiobjective Genetic Algorithm and Convolutional Neural Network Based COVID-19 Identification in Chest X-Ray Images, Mathematical Problems in Engineering, vol. 2021, Article ID 7804540, 9 pages. https://doi.org/10.1155/2021/7804540
26. Dasseni, E., Verykios, V. S., Elmagarmid, A. K., &Bertino, E. (2001, April). Hiding association rules by using confidence and support. In International Workshop on Information Hiding (pp. 369-383). Springer Berlin Heidelberg.
27. Dehkordi, M. N., Badie, K., & Zadeh, A. K. (2009). A Novel Method for Privacy Preserving in Association Rule Mining Based on Genetic Algorithms. JSW, 4(6), 555-562.
28. Dhutraj, N., Sasane, S., &Kshirsagar, V. (2013). Hiding sensitive association rule for privacy preservation. IEEE Transactions on Knowledge and Data Engineering Year.