An Proposed Model for Sensitive Rule Hiding

Kaushal Bhatt[1] Ankit Dongre[2]
[1]P.G. Student [2]Assistant Professor
[1,2] JIT,Borawan, Khargone(M.P)

Abstract—The corporations and users own and store their large quantities of data in digital form. Data mining tools are developed and used to extract useful information from that database. But problem is arise when the data base having private information called as a sensitive information about users and/or their organizations, there are worries that these data mining tools may inadvertently reveal sensitive information along with the originally intended output. That's why, we are interested in developing practical approach for privacy preserving data mining. And this research paper provides an efficient model for privacy preserving data mining.

Key words: Privacy preserving, Data mining, Association rule, Association rule publishing.

I. Introduction

Data mining is one of the task of Knowledge Discovery in Database (KDD) Process of extracting hidden valuable knowledge from transactional database.
The Common Steps involved in KDD is as follows [3]-

A. Data Cleaning:
It is a phase in which noise data and irrelevant data are removed from the collection.

B. Data Integration:
At this stage, multiple data sources, often heterogeneous, may be combined in a common source.

C. Data Selection:
At this step, the data relevant to the analysis is decided on and retrieved from the data collection.

D. Data Transformation:
Also known as data consolidation, it is a phase in which the selected data is transformed into forms appropriate for the mining procedure.

E. Data Mining:
It is the crucial step in which clever techniques are applied to extract patterns potentially useful.

F. Pattern Evaluation:
In this step, strictly interesting patterns representing knowledge are identified based on given measures.

G. Knowledge Representation:
It is the final phase in which the discovered knowledge is visually represented to the user. This essential step uses visualization techniques to help users understand and interpret the data mining results.

As more data is gathered digitally, data mining is becoming an increasingly important tool to transform this data into information [4]. Different Privacy preserving data mining techniques [1] exist and researched by researcher to hide sensitive information. This paper proposed a effective model to hide sensitive information to be published in data mining task. Our proposed model work on the basis of reduction of Support and Confidence. The orientation of our paper are as follows that it provide basic terminology , pictorial presentation of proposed model , advantages of proposed model so that one can get clear idea of technique for privacy preservation data mining.

II. Basic terminology

Some of basic terminology used in our proposed model is as follows-

A. Support:
The rule $X \Rightarrow Y$ holds with support s if s% of transactions in D contain $X \cup Y$ .Rules that have a s greater than a user-specified support is said to have minimum support[2].
Support (S): Occurrence of $(X \cup Y) / N$
Where N: Total Number of Occurrence

B. Confidence:
The rule $X \Rightarrow Y$ holds with confidence c if c% of the transactions in D that contain X also contain Y. Rules that have a c greater than a user-specified confidence is said to have minimum confidence [2].
Confidence(C) : Occurrence of $(X \cup Y)$/ Occurrence of X

III. Proposed model

Each Block of fig.3.1 is described below-

A. Data Base:
Data base contain the transactional data which is mined through different mining tools.

ID	Transactions
T100	Burger, Coke, Juice
T101	Juice, Potato Chips
T103	Juice, Ground Nuts
T104	Coke, Ground Nuts

Table. 1: Example of Transaction Database

B. Indexing:-
Index is data structure that allows the transaction data presented in database to arrange effectively and take less time in retrieval of data.

C. Association Rule (R):
Association rules are if/then statements that help uncover relationships between seemingly unrelated transaction data in a transactional database [5].

D. Rule Containing Sensitive Information(S):
Because we work on transactional data base. A company involved in data mining task does not want to reveal the personal information of any customer or customer purchasing information. That information is considered as sensitive information and rule containing this sensitive information is called as a sensitive rule.

E. Reducing support and confidence:

Association Rule having Support and Confidence above the user specific value is considered as a useful after data mining task. To hide any rule that contain sensitive information our proposed model reduced the Support and Confidence of that rule to below user specific value to hide that rule.

F. Publish Association Rule(R-S) :

After hiding association rule containing sensitive information remaining association rules are published for further business decisions.

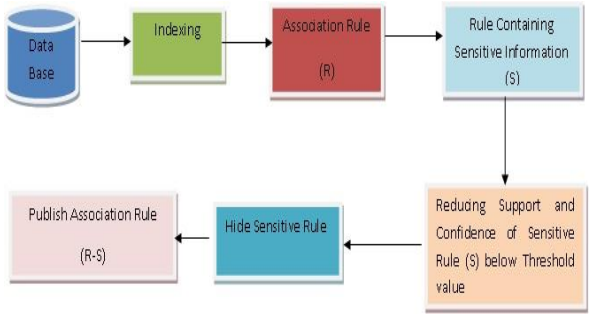

Fig..1: Proposed Model of sensitive Rule Hiding

IV. ADVANTAGE

(1) Simple to implement.
(2) Take less time to execute as compared to previous privacy preserving work.
(3) Hide sensitive data found either on left side or right side of association rule.
(4) No false association rules are generated by applying this model to hide sensitive data.
(5) No need to transform original database for hiding sensitive information.

V. CONCLUSION

In this proposed work, the database privacy problems due to data mining technology are discussed and the model for hiding sensitive rules is presented. The proposed model here presented is useful to hide sensitive association rule efficiently.

REFERENCES

[1] Kaushal bhatt, Ankit dongre ―A survey of sensitive information hiding techniques, International Journal of Emerging Technology and Advanced Engineering(IJETAE) ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 1, January 2014.
[2] Jain Pei, Jiawei Han, Micheline Kamber-Data Mining : Concepts and Techniques 3rd Edition.
[3] Arun K. Pujari- Data Mining Techniques 2nd Edition.
[4] Abedelaziz Mohaisen and Dowon Hong (2008),Privacy Preserving Association Rule Mining Revisited‟‟, Journal of the Computing Research Repository, pp. 1-16.
[5] Agrawal, R., and Srikant (2007), Privacy Preserving Data Mining‟‟, Proceedings of the 19th ACM International Conference on Knowledge Discovery and Data Mining, Canada, pp. 439-450.

