IJDI-ERET

INTERNATIONAL JOURNAL OF DARSHAN INSTITUTE ON ENGINEERING RESEARCH AND EMERGING TECHNOLOGIES
Vol. 11, No. 2, 2022
www.ijdieret.in

(Research Article)

Masking Techniques for Confidential Data Protection in Privacy-Preserving Data Mining

Dr. S. Vijayarani[1*], S. Sharmila[2], M. Lavanya[3]

[1*,2,3]*Department of Computer Science, Bharathiar University, Coimbatore, Tamil Nadu, INDIA*

Abstract

Privacy-Preserving Data Mining (PPDM) develops algorithms for altering sensitive data. The private knowledge of a person, industry, or business organization remains private after the usage of data from the database. Data modification is one of the prominent privacy-preserving techniques used to alter confidential information available in the database and guarantees high privacy protection. In this paper, a new masking technique is proposed for hiding sensitive numerical attributes that are later analyzed using clustering algorithms, namely k-means, filtered clusters, and density-based clusters. The proposed technique is used to hide confidential numerical attributes. After modification, the proposed algorithm compares the original and the modified data and ensures that all the data items are altered or not. Experimental evaluation is illustrated using the employee dataset. The accuracy is calculated based on the comparison of the original and the modified data set in terms of data items found in the number of clusters. For every clustering technique, both the original and modified at a set are divided into two, three, four, and five, clusters. Based on the performance metrics, the k-means algorithm gives the best result compared to other algorithms. The results obtained from the proposed technique are compared with the existing approach. The experimental result indicates that the newly developed method is more efficient than the existing approaches.

Keywords: Data masking, Data swapping, Sensitive data, Privacy-Preserving Data Mining, Clustering, Data hiding.

1. Introduction

In the current era, the generation of several kinds of data has increased. Generated data from multiple resources are gathered, efficiently processed, and stored. The fundamental goal of data management is to obtain useful and needed information. Getting information from the data is further useful for improving the business and strategies used in business with the help of data mining tools and statistical tools. The vast benefits of data mining approaches have repeatedly archived their best in consumer profiling, medicine, market-basket analysis, anti-terrorism efforts, fraud detection, and various other domains. One of the vital objectives of data mining is to identify the useful, new, and most needed patterns from the huge quantity of data. These identification processes allow the user to access, alter, and duplicate the stored data.

Confidentiality and privacy are the major issues in data mining. These issues were the heavy blockades that restricted the user from attaining the fullest benefit [1]. Stored data is subjected to an analytical process that provides access permissions only to certain relevant people. Authorization facilities provide secure data retrieval and storage. In this regard, numerical data are particularly emphasized.

Numerical data offers several benefits and security threats. Business Intelligence attains benefits and attention with the help of numerical data. Any third party or illegitimate user can easily breach the confidentiality and privacy of sensitive data. This creates a serious problem in various situations in organizations [2].

Privacy-preserving data mining (PPDM) safeguards highly sensitive data from illegitimate access and disclosure. PPDM consists of two techniques namely, horizontal and vertical data partitioning. Masking techniques were introduced to defend the data from multiple security threats. In data masking, the sensitive information within the data store is replaced by randomly generated data or by any other information [3].

Data masking masks confidential and sensitive information so non-production systems can be duplicated. Several IT organizations use customized or prebuilt refined data masking techniques to maintain referential integrity and preserve the data [4].

[*]*Corresponding Author: e-mail: vijayarani@buc.edu.in*
Tel-+91 9442130862, + 91 8344588980
ISSN 2320-7590 (Print) 2583-3863 (Online)

2. Data Masking

Data masking is one of the rising technologies due to the ever-growing data across industries and numerous data attacks on the enterprise are also continuing as well. Industrialists and business persons are looking for the most eminent practices for shielding their data. Data Masking indicates great promise for numerous futuristic approaches in software testing and development. It is often essential to anonymize data and databases in order to guard them against inappropriate visibility. Managers and security professionals in industries are concerned that the leading information security risk to an organization comes from within and outside of the organization. Data security also minimizes the probability of both internal and external attacks. Data protection is generally concerned with the issue of exposure of data [5].

Data masking hides sensitive data to avoid the exposure of data to users who don't have the right to inspect the data. It is performed as per the access privileges of the user. The vital function of data masking is to mask personal information as non-production data with a pragmatic looking but not with actual information. Data masking assures that data security is preserved by hiding specific data inside a database table which will lessen the risk of data revealing and data breaches [6]. Effectual data masking methods require data to be changed in some way so that actual values are re-constructed, while the table holds the structural and functional meaning of the data without compromising any security credentials.

2.1 Five laws of Data Masking: The general reason for masking data instead of creating random data is that masking permits sensitive information protection. The five laws of data masking are explained here. [7]

2.1.1 Masking must not be reversible: The hidden data in the database cannot be retrieved, which leads to a security credential issue. The original responsive data cannot be recovered from the databases.

2.1.2 The results must be representative of the source data: Highly sensitive data in the database is masked with randomly generated information that is non-sensitive data.

2.1.3 Only mask non-sensitive data if it can be used to recreate sensitive data: Personal information alone can also be masked, not all the stored information needs to be masked. But some impersonal data is also used to rebuild or bind back to the needed sensitive data.

2.1.4 Referential integrity must be maintained: Masking elucidations doesn't interrupt referential integrity. The number of a debit card number is a primary key and replaced as part of masking, then all occurrences of that number correlated through key pairs must also be carefully replaced.

2.1.5 Masking must be a repeatable process: Unproductive and impossible to maintain one-off masking techniques which are ineffective in securing the data. Test data and developed data need to be represented instantaneously once the production data is changed.

2.2 Orders of Data Masking: Data masking has numerous operational methods which have similar efficiency, and all data masking processes look for the same weakness in the data. For this reason, the order of data masking operation is implemented [2] [8].

2.2.1 First order: The content of the data which is the entire main information, comes under the first order of the data. From the huge database masking field's likely names, telephone contact details, date of birth, and other information of personal information fall under this category [5] [21].

2.2.2 Second order: A combination of structure and content of the data involves a second order. The fields in this order data are confidential in their own right. This kind of information doesn't come under the first-order type of masking. Fields in this data are used several times across the data sets of different tables [5].

2.2.3 Third order: The pure form of the structural mask is third order, and it shows the relationship between the information in the tables and inside the tables. The masking data of the third order is currency values (likely purchase amount, salary), the quantity of the product, and fields that are subject to the query "aggregate" [9].

2.2.4 Fourth order: Structural data are masked in the fourth order which deals with the information retrieval through the schema of the database itself rather than the creation of interrelationships among the individual rows of information. The Fourth order emphasizes the field namely foreign key, data type, the existence of the field, expectations, and queries [8] [10].

2.2.5 Fifth order: Metadata information is masked in the fifth order which is RDBMS data. It is used to store within the company for the testing methodology or data within the company and being improved. RDBMS is used to hold the company data, production data, or testing methodology [11] [22].

3. Related Works

S. Vijayarani et.al have proposed the data transformation technique, used for numerical attributes. Protecting sensitive data and also extracting knowledge is a very complicated problem. Based on the above experimental results it has been analysed that the proposed data transformation is a technique used for protecting and modifying sensitive data [6].

Chuang-Cheng Chiu et al have proposed a novel clustering method for conducting the k-anonymity model effectively. The similarity between this method and the proposed method is in the reduction of information distortion. The difference is in the clustering algorithm that is used and the privacy technique [7].

Wang Jian et al have discussed a condensation approach for data mining. This approach used a methodology that condenses the data into multiple groups of predefined size. For each group, certain statistics are maintained. A greater amount of information is lost because of the condensation of a larger number of records into a single statistical group entity. They used the statistics from each group in order to generate the corresponding pseudo-data. The results shown that the proposed method is the best at reducing the amount of information loss [12].

R. Natarajan et.al described privacy-preserving data sharing based on a public-key cryptosystem and its protocol. The importance of protocol like how it allows users to conduct private mining analyses without loss of accuracy was also discussed. The author discussed the execution of protocols exactly how they were specified, the research paper provided a detailed study of inferring hidden links, useful information about other parties and privacy, and concerns of actual data [13].

Alexandre et.al analysed traditional data mining techniques and models, and this research work briefly discussed about privacy preservation techniques. It is primarily concerned with protecting against disclosure of individual data records. This paper discussed an overview of the popular approaches in PPDM, such as: suppression, randomization, Cryptography, and summarization [15].

Selva Rathna et.al reviewed the method of PPDM. The author used the FP-Growth algorithm along with elliptic curve cryptography and elliptic curve-based digital signature algorithms for authentication of sites and verification of data. From the analysis, it was concluded that the method used in this research work had secured the data better. The proposed method used a central authority called an association third party which had collected all the information related to the association rule mining in a secure manner [16].

E. Michael et.al described about the recent works made in privacy preserving data mining techniques with Fuzzy logic, neural network learning, and secured sum and encryption algorithms. The author has discussed the challenges of PPDM. From the analysis, the author identified the best techniques that were suitable for various data environments [16].

4. Data Masking Algorithm

4.1 Existing algorithm: Data swapping was originally described as a statistical disclosure method for databases that hold only categorical attributes. The fundamental idea behind the technique is to convert a database by interchanging the values of sensitive attributes among individual records in the database. Records are exchanged in such a way that low-order frequency counts or marginals are maintained constantly [13]. Even though the original procedure is not used in practice, its basic idea had a clear influence on the subsequent method.

Rank swapping is another variant of data swapping. Although swapping is initially explained for ordinal attributes, rank swapping is also used for any numerical attribute. Values of Xi are ranked in ascending order, and then each value of Xi is swapped with a differently ranked value randomly within the database that is elected within a restricted range [11] [18].

4.2 Proposed algorithm: In aggregation, privacy preservation is predominantly employed to protect the data of any single data record from disclosure and is accomplished in diverse ways by the use of cryptographic algorithms, anonymization methods, randomization techniques, etc. Cryptography is used to encipher the data to prevent stealing.

Data masking is used to increase the security credentials of the data [20]. Data privacy solutions are personalized for each organization. Developed techniques can improve the quality of the data. Confidential data can be intruded by an intruder and is protected by privacy-preserving techniques. Data swapping is introduced to protect continuous and categorical microdata, respectively [23].

In earlier empirical work by these authors on continuous microdata protection, rank swapping has been identified as a particularly well-performing method in terms of the trade-off between loss of data and disclosure threat. The proposed technique uses ascending order to hide the original numerical attribute. In the modified sensitive data, each row is arranged in ascending order. The original confidential data items are compared with the modified data items to verify whether both data items have the same value or not. If both data items have different values, then privacy protection is highly secured.

The proposed methodology hides the most sensitive information. The masking technique is the major work that is important in hiding the information. The main aim of this proposed work is to find a new masking technique for hiding sensitive numerical attributes and to find accuracy, time, and performance using cluster algorithms. The proposed technique uses the ascending order method to hide the original attribute. Every column is arranged in ascending order to modify the sensitive data. The dataset contains employee information, which includes 15000 instances and 3 attributes, namely employee ID, name, and salary. Here, sensitive numerical attribute i.e., salary, is modified using the proposed technique. The existing technique for data swapping is compared with the proposed technique by using performance measures like accuracy, several iterations, and execution time. The original confidential data items are compared with the modified data items to verify whether both data items have the same value. If

both data items have different values, then privacy protection is highly secured. Clustering accuracy is calculated based on the comparison of data items representing the number of clusters in the original dataset and data items of the same number of clusters in the modified dataset. Three different types of clustering techniques are applied. They are k-means, filtered clusters, and make density-based clusters. For every clustering technique, both the original and modified datasets are split into two, three, four, and five clusters. Based on the performance metrics, the k-means algorithm gives the best result compared to other algorithms.

Pesudocode for Proposed algorithm
Input: Given database D with attributes A_1, A_2 …A_n. where n = no of attributes Select sensitive numerical attributes $S_{Ai} \in D$.
Output: Modified Sensitive attributes $S_{Ai}{}'$
Procedure
Step 1: Select the important sensitive numerical attribute from database D
Step 2: Based on the number of digits of the data, make m groups.
Step 3: Consider each group one by one
Step 4: Arrange the MSB, MSB-1, MSB-2… LSB in the ascending order
Step 5: Repeat the same process for all the groups
Step 6: Newly arrived sorted individual number as the modified data
Step 7: Stop the process

5. Clustering Algorithm

5.1 Simple K-means Algorithms: K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) that are fixed Apriori. The main idea is to define k centers, one for each cluster. These centers should be placed differently because different location causes the different result. So, the better choice is to place them as much as possible far away from each other [13] [17].

Pesudocode for k-Simple Means Clustering
Input: Let X = {x_1, x_2, x_3,…..,x_n} be the set of data points and V = {v_1,v_2,…….,v_c} be the set of centres.
Step 1: Randomly select *'c'* cluster centers (cc)
Step 2: Calculate the distance between each data point and cluster centers.
Step 3: Assign the data point to the cc whose distance from the cc is minimum of all the cc
Step 4: Recalculate the new cc using: $v_i=(1/c_i)\sum_{j=1}^{ci} xi$ //Where, *'c_i'* represents the no. of data points in i^{th} cluster.
Step 5: Recalculate the distance between each data point and new obtained cluster centers.
Step 6: If no data point was reassigned then stop, otherwise repeat from step 3

5.2 Filtered Algorithm: The algorithm is based on storing the multidimensional data points in a K-d binary tree. First, compute in a K-d tree for the given data points. The technique maintains a set of candidates centers for each of the node of the K-d tree, if n is a leaf node, then computes the distances from its associated data point to all the candidates in D and assigns the data point to its nearest centers [14] [19].

Pesudocode for Filtered Cluster (KdNode n, CandidatedataSet D)
Procedure
C ← n.cell;
If n is a leaf {
D * ← the closest point in D to n.point;
D * .wgtCent ← D *. WgtCent + n.Point;
D * .count ← D *. Count+1;}
Else D * ← the closest point in D to C's midpoint;
For each (D∈D\ {D *}
If (D is Farther (D *, C))
D←D\ {D};
If (|D|=1) {
D *. wgtCent ← D *. wgtCent+n.wgtCent;
D * .count ← D *. Count+n.count
}
Else {
Filter (n.left, D);
Filter (n.right, D);
}}}

5.3 Density based clustering algorithm: A density-based clustering algorithm has played a vital role in finding nonlinear shape structures based on density. Density-Based Spatial Clustering of applications with Noise (DBSCAN) is the most widely used density-based algorithm [20] [24].

Pesudocode for DBSCAN clustering

Input: Let X = {x_1, x_2, x_3... x_n} be the set of data points. DBSCAN requires two parameters: ε (eps) and the minimum number of points required to form a cluster (minPts).

Step 1: Start with an arbitrary starting point that has not been visited.

Step 2: Extract the neighbourhood of this point using ε (All points which are within the distance are neighbourhood).

Step 3: **If** there is sufficient neighbourhood around this point **then** clustering process starts and point is marked as visited **else** this point is labelled as noise. (Later this point can become the part of the cluster).

Step 4: **If** a point is found to be a part of the cluster, **then** its neighbourhood is also the part of the cluster, step 2 is repeated for all ε neighbourhood points. This is repeated until all points in the cluster is determined

Step 5**:** A new unvisited point is retrieved and processed, leading to the discovery of a further cluster or noise.

Step 6: This process continues until all points are marked as visited.

6. Results and Discussion

This experiment is implemented using POSTGRESQL on a system with an HP processor, 2 GB RAM, and 32-bit Windows 8.1 A performance test is evaluated based on execution time, accuracy, and iteration. The performance of these clustering algorithms is analysed to identify the best accuracy. In this proposed work, the best accuracy, number of iterations, and execution time were calculated. From the experimental results proposed algorithm performs better than the existing Data Swapping technique.

6.1 Dataset Description: To perform the experiments, the dataset was generated synthetically for employee details. The main objective of this research is to hide sensitive information from employee information using clustering algorithms. Employee details contain 15000 instances and 3 attributes, namely employee ID, employee name, and salary. Table 1 shows the original dataset. Here, the salary attribute is considered a numerical sensitive attribute from the employee dataset. The proposed technique uses the ascending order method to hide the original attribute. Every column is arranged in ascending order to modify the sensitive data and the modified data is stored in the modified database. Table 1 shows the modified database.

Table 1. Original/modified database

Original database			Modified database		
Id	Name	Salary	Id	Salary	Salary edited
1	Arome Karlin	14563	1	14563	12000
2	Arome Libil	17864	2	17864	14500
3	Arome Mhana	89653	3	89653	25600
4	Abile Aghot	9832	4	9832	1330
5	Abile Thyro	78900	5	78900	35750
6	Arohit Athin	45000	6	45000	47763
7	Jack Mantro	22700	7	22700	78863
8	Jack Sandy	35780	8	35780	89984
9	Jack Thin	1340	9	1340	9842

Table 2. Clustering accuracy

Clustering techniques	Dataset	2 clusters		3 clusters		4 clusters		5 clusters	
		Existing	Proposed	Existing	Proposed	Existing	Proposed	Existing	Proposed
Simple K-means	3000	95.7	97.6	92.5	93.8	87.9	94.7	93.1	98.9
	8000	95.6	98.4	94.7	97	95.6	98.4	89.01	93
	15000	98.3	100	94.2	96.1	98.3	99.1	92	97.4
Make density-based	3000	91.7	96.8	90	95.4	100	95.4	94.6	95.3
	8000	94.4	95.7	94.8	97.9	93.9	97.6	95.6	96.1
	15000	93.5	97.6	97.8	100	94.5	95.7	92.7	93.4
Filtered	3000	96.5	97.9	93.7	94.8	96.5	96.9	89.4	98.7
	8000	90.6	95.4	98.6	99.6	97.1	99.2	92.7	94.7
	15000	89.8	92.6	97.8	97.9	91.6	94.3	96.3	96.9

Table 2 shows the clustering accuracy using three clustering techniques, namely Simple K-means, make density- based and Filtered clustering. Figure 1 describes the analysis of clustering accuracy in seconds. From the above analysis, it has proved that Simple K-means give the best result

Figure 1. Clustering accuracy

Table 3. Execution time (seconds)

Clustering techniques	Dataset	2 clusters		3 clusters		4 clusters		5 clusters	
		Existing	Proposed	Existing	Proposed	Existing	Proposed	Existing	Proposed
Simple K-means	3000	2	1	3	2	6	2	8	4
	8000	6	3	4	1	9	4	9	8
	15000	7	4	8	3	14	10	14	10
Make density- based	3000	4	1	5	1	5	3	10	7
	8000	3	2	9	7	9	5	13	12
	15000	5	3	6	4	16	10	9	5
Filtered	3000	6	3	3	1	5	2	8	5
	8000	4	1	5	2	12	8	5	4
	15000	5	2	14	8	8	6	7	4

Table 3 shows the performance of execution time using three clustering techniques, namely Simple K-means, Make density-based and Filtered clustering. Figure 2 illustrates the execution time in seconds. From the above analysis, it has proved that Simple K-means gives the best results.

Figure 2. Execution time

Table 4. Number of iterations

Clustering techniques	Dataset	2 clusters		3 clusters		4 clusters		5 clusters	
		Existing	Proposed	Existing	Proposed	Existing	Proposed	Existing	Proposed
Simple K-means	3000	5	4	10	7	8	5	8	6
	8000	8	3	9	4	12	9	9	5
	15000	8	3	14	11	16	11	7	3
Make density-based	3000	12	9	12	6	3	1	12	9
	8000	16	11	10	8	7	5	3	1
	15000	9	8	24	18	16	12	5	4
Filtered	3000	8	6	4	3	8	5	6	2
	8000	4	2	9	3	10	9	4	2
	15000	12	7	18	12	15	11	3	1

Table 4 shows the number of iterations to complete the process using three clustering techniques namely Simple K-means, Make density-based and Filtered clustering. Figure 3 illustrates the number of iterations required to complete the process. From the above analysis, it has proved that Simple K-means gives the best results.

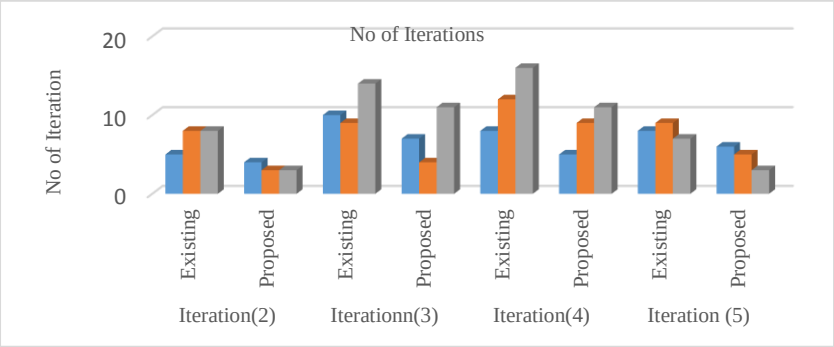

Figure 3. Number of Iterations

7. Conclusions

Data masking is mainly employed to hide confidential information with the help of randomly generated data. The key objective of this research work is to find accuracy in hiding by using clustering algorithms. The proposed technique used the ascending order arrangements of individual digits to hide the original data. Every row is arranged in ascending order to modify the sensitive data. The original confidential data items are compared with the modified data items to verify whether both data items have the same value. If both the data items have different values, then privacy protection is highly secured. Clustering accuracy is calculated based on the comparison of data items with several clusters in the original dataset and data items with the same number of clusters in the modified dataset. Three different types of clustering techniques are applied, they are K-means, filtered clusters, and Make density-based clusters. For every clustering technique, both the original and modified dataset are split into two, three, four, and five, clusters. Based on the performance metrics the K-means algorithm gives the best result compared to other algorithms.

References

1. Ravikumar, G. K., Rabi, B. J., Manjunath, T. N., Hegadi, R. S., & Archana, R. A. (2011). Design of data masking architecture and analysis of data masking techniques for testing. International Journal of Engineering Science and Technology, 3(6).
2. Data Masking: What You Need to Know What You Really Need to Know Before You Begin A Net 2000 Ltd. White Paper.
3. G Sarada, G Manikandan, Dr.N. Sairam, “A Few New Approaches to Data Masking”, International Conference on Circuit, Power and Computing Technology 2015.
4. S. Selvakumar and M. Mohan Priya, "Securing Cloud Data in Transit using Data Masking Technique in Cloud Enabled Multi-Tenant Software Service", Indian Journal of Science and Technology, vol. 9, no. 20, 2016.
5. S. Vijayarani and S. Nithya.” Sensitive Outlier Protection in Privacy Preserving Data Mining”, International Journal of Computer Applications (0975 8887), Volume 33-No. 3, November (2011).

6. Min Li, Zheli Liu, ChunfuJia, Zongqing Dong, "Data Masking Generic Model", Fourth International Conference on Emerging Intelligent Data and Web Technologies 2013.
7. Goyal, C. (2015). "Data Masking: Need, Techniques & Solutions", International Research Journal of Management sociology and Humanities.
8. Data Scrambling Issues *A Net 2000 Ltd. White Paper.*
9. Richard Fine and Llyr Jones, Grid-Tools Ltd BY the Mathematics of Data Masking.
10. Sachin Lodha, Ellora Praharaj and Ravishankar Rajamony. Selecting Quasi-Identifiers. Submitted to International Conference on Database Technology (ICDT), 2007.
11. Domingo-Ferrer, J & Torra, V (2002), "Aggregation Techniques for Statistical confidentiality". In: Aggregation operators: new trends and applications, pp. 260-271. Physica-Verlag GmbH, Heidelberg (2002).
12. Samarati, P (2001), "Protecting respondents' identities in micro data release", IEEE Transactions on Knowledge and Data Engineering, 13(6):1010-1027. 2001.
13. Vassilios S. Veryhios, Elisa Bertino, Igor Nai Fovino Loredana Parasiliti Provenza, Yucel Saygin, Yannis eodoridis, "State-of-the-art in Privacy Preserving Data Mining", SIGMOD Record, Vol. 33, No. 1, March 2004.
14. J. Han, M. Kamber. "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers".
15. R. Agrawal and R. Srikant. "Privacy Preserving Data Mining", ACM SIGMOD Conference on Management of Data, pp: 439-450, 2000.
16. Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining", Journal of Cryptology, 15(3), pp.36-54, 2000.
17. Aris Gkoulalas-Divanis and Vassilios S. Verikios, "An Overview of Privacy Preserving Data Mining", Published by The ACM Student Magazine, 2010.
18. Stanley, R. M. O. and R. Z Osmar, "Towards Standardization in Privacy Preserving Data Mining", Published in Proceedings of 3rd Workshop on Data Mining Standards, WDMS' 2004, USA, p.7-17
19. Elisa, B., N.F. Igor and P.P. Loredana. "A Framework for Evaluating Privacy Preserving Data Mining Algorithms", Published by Data Mining Knowledge Discovery, 2005, pp.121-154.
20. Andreas Prodromidis, Philip Chan, and Salvatore Stolfo, "Metalearning in distributed data mining systems: Issues and approaches". In "Advances in Distributed and Parallel Knowledge Discovery", AAAI/MIT Press, September 2000.
21. S.V. Vassilios , B. Elisa, N.F. Igor, P.P. Loredana, S. Yucel and T. Yannis, 2004, "State of the Art in Privacy Preserving Data Mining" Published in SIGMOD Record, 33, 2004, pp: 50-57.
22. N. Punitha, R. Amsaveni," Methods and Techniques to Protect the Privacy Information in Privacy Preservation Data Mining ", Int. J. Comp. Tech. Appl., Published in IJCTA, NOV-DEC 2011, Vol 2 (6),7. ISSN:2229-6093
23. V.Ciriani, S. DeCapitani diVimercati, S. Foresti, and P. Samarati University degli Studi di Milano, "Micro data protection" 26013 Crema, ltalia, Springer US, Advances in Information Security (2007)
24. Krishnamurty Muralidhar, Rahul Parsa, Rathindra Sarathy, "A general Additive Data Perturbation Method for database Security", management science, Vol. 45, No. 10, October 1999, pp. 1399-1415 DOI: 10.1287/mnsc.45.10.1399

Biographical notes

Dr. S. Vijayarani M.C.A., M.Phil., Ph.D., DCSE. working as an Assistant Professor in the Department of Computer Science, Bharathiar University, Coimbatore. Her research interests include Privacy-Preserving Data Mining, Utility Mining, Text Mining, Web Mining, Image Mining, and Health Care Analytics. She has published more than 200 research articles in International / National journals and conferences. She has also written 20 book chapters. She has produced 33 M.Phil./Ph.D. research scholars. She is a member of various professional bodies like CSI, IAENG, UACEE, INSA, etc

Dr. S. Sharmila, completed her Ph.D. in the Department of Computer Science, Bharathiar University, Coimbatore, Tamilnadu, India. She has completed her MCA., Ph.D., at Bharathiar University, Tamilnadu, India. Her research area includes Association Rule Mining, Association Rule Hiding, Privacy Preserving, and Optimization techniques. She has published papers in International Journals and Conferences.

M. Lavanya completed her M.Sc. Computer Science., in Department of Computer Science, Bharathiar University, Coimbatore, Tamilnadu, India. Her research area includes Privacy-Preserving and Optimization techniques. She has published papers in International Journals and Conferences.