Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 8, Issue, 03, pp. 28325-28329, March, 2016

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

FUZZY MEMBERSHIP FUNCTIONS IN PRIVACY PRESERVING DATA MINING

[1,*]Selva Rathna and [2]Dr. Karthikeyan, T.

[1]Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India
[2]P.S.G Arts and Science College, Bharathiyar University, Coimbatore, Tamil Nadu, India

ARTICLE INFO

Article History:
Received 17th December, 2015
Received in revised form
24th January, 2016
Accepted 26th February, 2016
Published online 31st March, 2016

Key words:

Fuzzy Logic,
Membership Function,
Privacy Preserving Data Mining,
WEKA, Clustering,
Classification.

ABSTRACT

The aim of this paper is to identify the impact of the fuzzy based privacy preserving method for protecting sensitive data in Privacy preserving data mining. Fuzzy based member ship functions like Bell shape, S- Shape and PI shape member ship functions are applied on standard database to generate sanitised database. Further, Popular K-mean Clustering algorithm of data mining is applied on the sanitised database and the results are compared. WEKA tool is used for testing data mining algorithm on privacy preserved database generated using various fuzzy member ship function. This analysis will help to develop new Fuzzy Based privacy preserving techniques which can be applied for Privacy preserved data mining.

Citation: Selva Rathna and Dr. Karthikeyan, T. 2016. "Fuzzy membership functions in privacy preserving data mining", *International Journal of Current Research*, 8, (03), 28325-28329.

INTRODUCTION

In recent years, data mining has been viewed as a threat to privacy because of the widespread proliferation of electronic data maintained by corporations. Privacy preserving Data mining means hiding the sensitive values of a database and performing data mining operations on the sanitised database. Effective preservation techniques will enable the data base owner to prepare privacy preserved data base which can be further used by a third party for any data mining operations like clustering, classification, association rule mining etc. Privacy-preserving data mining has various applications such as for bio-surveillance, facial de-identification, and identity theft. A number of techniques have been proposed for modifying or transforming data to preserve privacy which are effective without compromising security. In Section 2, earlier researches related to Privacy preserving data mining and use of Fuzzy logic techniques in maintaining privacy of the sensitive data is discussed. In Section 3, Fuzzy based privacy preservation with various member ship function is discussed. In Section 4, analysis of data mining algorithm on privacy preserved data using various fuzzy approaches is discussed. In Section 5, results of the analysis are presented. In Section 6, conclusion of the analysis is given.

Privacy Preserving Data Mining Survey

Privacy preserving Data mining

A recent survey on the techniques used for privacy-preserving data mining may be found in Xueyun (2014) which reviews main PPDM techniques based on a PPDM framework and compare the advantages and disadvantages of different PPDM technique. It also discusses the open issues and future research trends in PPDM. Malik (2012), describes the current scenario of Privacy preserving data mining and propose some future research directions. In Shweta (2014), all Cryptography and Random Data Perturbation methods techniques of PPDM is studied. Chris (2002) illustrates the application of certain techniques for preserving privacy on experimental dataset, and reveals their effects on the results. Jian (2009) intends to reiterate several privacy preserving data mining technologies clearly and then proceeds to analyze the merits and shortcomings of these technologies. Methods such as *k*-anonymity, *l*-diversity, *t*-closeness, classification, association rule mining are all designed to prevent identification, though

****Corresponding author: Selva Rathna***
Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India.

the final goal is to preserve the underlying sensitive information. Various techniques and frameworks have been developed for Privacy preserving data mining. The fundamental framework of Privacy preserving data mining is as follows.

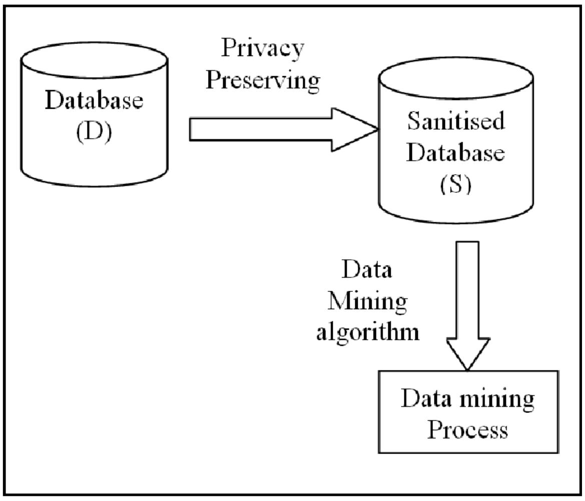


Figure 1. Fundamental Framework of Privacy preserving data mining

Privacy Preserving Data mining with Fuzzy Logic

Fuzzy logic can be applied in Privacy preserving data mining for anonymization of data to preserve sensitive data in a database. Certain membership functions can be used to fuzzify the data so that the private date is not disclosed during data mining process. The sanitised data based created after fuzzification process should be compatible for applying data mining algorithms like classification, clustering etc along with the privacy preservation. The challenge in the process is privacy should be achieved without or less data loss. Various researches have been made with Fuzzy logic in Privacy preserving data mining. Mukkamala (2011) compared a set of fuzzy-based mapping techniques in terms of their privacy-preserving property and their ability to retain the same relationship with other fields. Jian (2014) proposed a method to extract global fuzzy rules from distributed data with the same attributes in a privacy-preserving manner. Cano (2009), proposed a fuzzy c-regression method to generate synthetic data which allows third parties to do statistical computations with a limited risk of disclosure. Torra (2009) made a study on intuitionistic fuzzy clustering and Kasugai (2013) studies the applicability of fuzzy k-member clustering to privacy preserving pattern recognitionl. k-member clustering is a basic technique for achieving k-anonymization, in which data samples are summarized so that any sample is indistinguishable from at least k - 1 other samples. Honda (2012) proposes a fuzzy variant of k-member clustering with the goal of improving the quality of data summarization with k-anonymity. The proposed anonymization method is also applied to collaborative filtering. Tanak (2014) proposed a secure framework for privacy preserving fuzzy co-clustering for handling both vertically and horizontally distributed cooccurrence matrices. A method to hide fuzzy association rule is proposed by Sathiyapriya (2011) using modified apriori algorithm in order to identify sensitive rules to be hidden.

Fuzzy Membership Functions for PPDM

Based on the study made in Section 2, Fuzzy logic membership functions are used to anonymizing the selective data of the database for maintaining privacy of the sensitive data. The privacy and information loss due to application of privacy preservation process has to be maintained zero or at least minimum. Also, the application of privacy preservation process should not affect the data mining process. Membership function (MF) is curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The fuzziness of a fuzzy set is determined by its membership function. The shapes of the membership functions are important for a particular problem since they effect on fuzzy inference system. They may have different shapes like Triangular, Trapezoidal, Bell Shape, PI shape etc. Shi, Yigang (200) has proposed the formula for Triangular MF and Zhoa (2002) has proposed Trapezoidal MF. In this paper, the following Fuzzy membership functions are selected for analysis.

- S-Membership Function
- Z - Membership Function
- PI –Membership Function
- Bell Membership Function
- Elliptical Type 2 Member ship function

S-Fuzzy Membership Function

The S-Shaped function takes three parameters as input and produces the modified membership plane or property plane. By using S-Shaped membership function the shape of the input is modified i.e. fuzzy domain values are modified. The formula for S-Shaped membership function is given in Equation (1).

$$f(x,a,b,c,d) =$$

$$= \begin{cases} 0, & x \le a \\ 2\left(\frac{x-a}{b-a}\right)^2, & a \le x \le \frac{a+b}{2} \\ 1-2\left(\frac{x-a}{b-a}\right)^2, & \frac{a+b}{2} \le x \le b \\ 1, & b \le x \le c \end{cases} \qquad \text{.........................} (1)$$

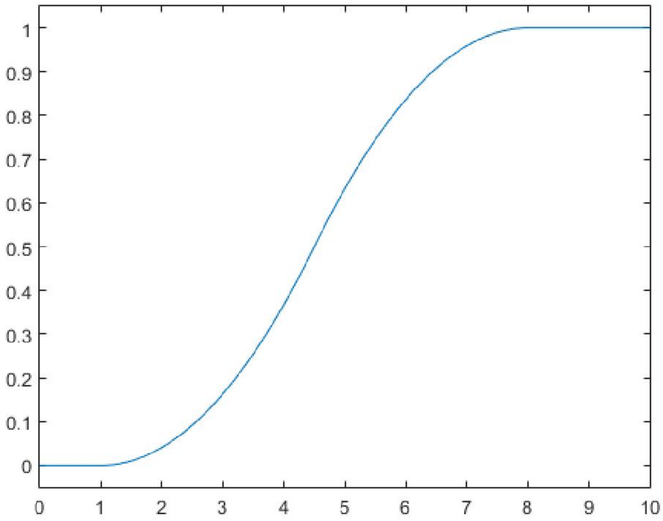


Figure 2. Shape of S- Fuzzy membership function

S-Shaped function has *a, b* and *c* as input parameters where *a* = minimum value, *c* = maximum value and *b= (a + c) / 2* for given range. For any point *i, j,* membership value $\mu_{x(i,j)}$ is computed using a, b & c. The shape of S-Fuzzy membersip member is shown in Figure 2.

Z-Fuzzy Membership Function

The shape of Z-Fuzzy membersip member is shown in Figure 3.

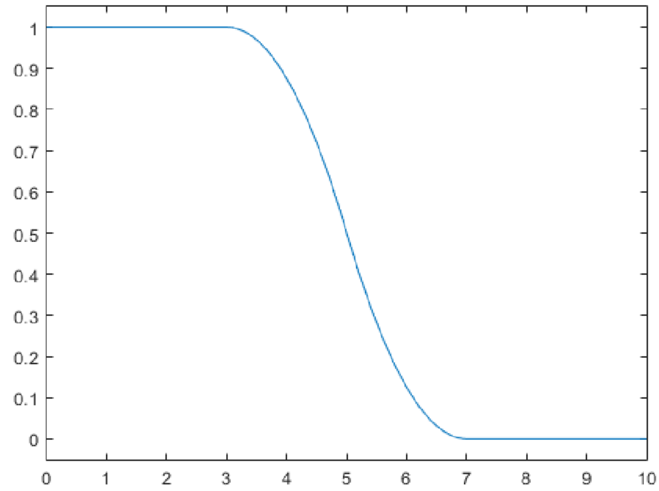


Figure 3. Shape of Z- Fuzzy membership function

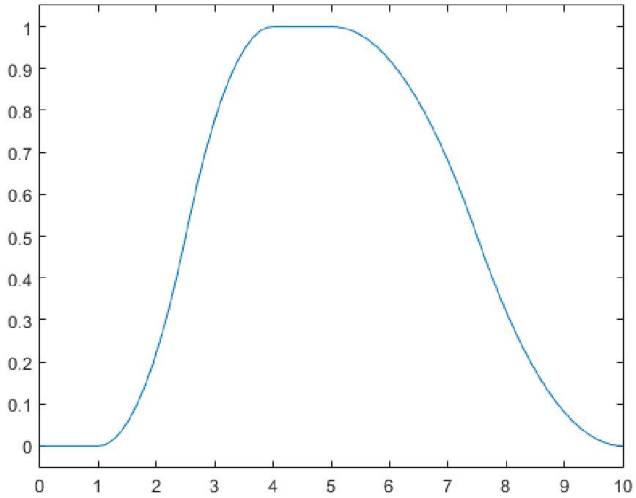


Figure 4. Shape of PI- Fuzzy membership function

The Z-Shaped function takes three parameters as input and produces the modified membership plane or property plane. By using Z-Shaped membership function the shape of the input is modified i.e. fuzzy domain values are modified. The formula for Z-Shaped membership function is given in Equation (2)

$$f(x,a,b,c,d) = \begin{cases} 1\,, & x \le a \\ 1 - 2\left(\frac{x-a}{b-a}\right)^2, & a \le x \le \frac{a+b}{2} \\ 2\left(\frac{x-a}{b-a}\right)^2, & \frac{a+b}{2} \le x \le b \\ 0\,, & b \le x \le c \end{cases} \quad \ldots\ldots\ldots.\ (2)$$

S-Shaped function has *a, b* and *c* as input parameters where *a* = minimum value, *c* = maximum value and *b= (a + c) / 2* for given range. For any point *i, j,* membership value $\mu_{x(i,j)}$ is computed using a, b & c.

PI-Fuzzy Membership Function

The PI-Shaped function takes four parameters *a, b, c* and *d* as input and produces the modified membership plane or property plane. Figure 3 shows PI shaped membership function. Parameter *a* and *d* located in the feet and *b* and *c* is located in the top of the curve. PI membership function is the product of *S* membership function and *Z* membership function. The formula of PI fuzzy membership function is as given in Equation (2).

$$f(x,a,b,c,d) = \begin{cases} 0\,, & x \le a \\ 2\left(\frac{x-a}{b-a}\right)^2, & a \le x \le \frac{a+b}{2} \\ 1 - 2\left(\frac{x-a}{b-a}\right)^2, & \frac{a+b}{2} \le x \le b \\ 1\,, & b \le x \le c \\ 1 - 2\left(\frac{x-c}{d-c}\right)^2, & c \le x \le \frac{c+d}{2} \\ 2\left(\frac{x-d}{d-c}\right)^2, & \frac{c+d}{2} \le x \le d \\ 0\,, & x \ge d \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots.\ (3)$$

Bell-Fuzzy Membership Function

The Bell-Shaped function takes three parameters *a*, *b* and *c* as input and produces the modified membership plane or property plane. *b* is always positive and *c* is located in the center of the curve. The formula of PI fuzzy membership function is as given in Equation (3)

$$f(x,a,b,c,d) = \frac{1}{1 + \left|\frac{x-c}{a}\right|^{2b}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

Figure 5 shows Bell shaped membership function.

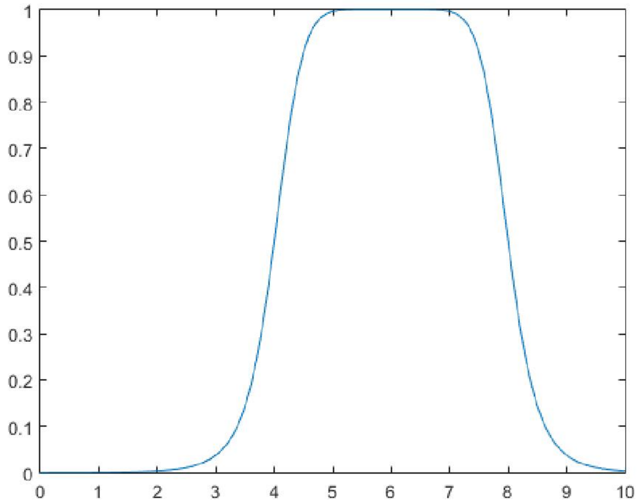


Figure 5. Shape of Bell Fuzzy membership function

Elliptical Type 2 - Fuzzy Membership Function

Fuzzy system with two inputs can be written as

$$y = q\frac{\sum_{j=1}^{J}\sum_{i=1}^{I}\underline{W_{ij}}f_{ij}}{\sum_{j=1}^{J}\sum_{i=1}^{I}\underline{W_{ij}}} + (1-q)\frac{\sum_{j=1}^{J}\sum_{i=1}^{I}\overline{W_{ij}}f_{ij}}{\sum_{j=1}^{J}\sum_{i=1}^{I}\overline{W_{ij}}} \quad \ldots\ldots\ldots (5)$$

where $\underline{W_{ij}} = \underline{\mu_{1i}}\ \underline{\mu_{2j}}$ and $\overline{W_{ij}} = \underline{\mu_{1i}}\ \underline{\mu_{2j}}$ and the parameter q is the weighting parameter which reflects the sharing of the contribution of the upper and the lower MFs.

Table 1. Data used for Analysis

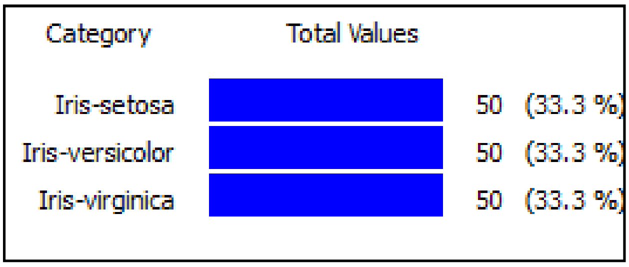
Name of the dataset	Number of Rows	No. of Columns	No. of classes
IRIS	150	4	3
GLASS	214	9	7

Table 2. Mean Absolute error

Fuzzy MF	Mean Absolute Error							
	5 Clusters		10 Clusters		15 Clusters		20 Clusters	
	Iris	Glass	Iris	Glass	Iris	Glass	Iris	Glass
Bell	0.10	0.14	0.06	0.15	0.03	0.13	0.04	0.13
PI	0.16	0.15	0.10	0.14	0.09	0.12	0.11	0.12
Z	0.24	0.17	0.13	0.16	0.13	0.14	0.16	0.13
S	0.10	0.15	0.08	0.14	0.05	0.12	0.05	0.12
ET 2	0.14	0.15	0.06	0.16	0.07	0.19	0.05	0.11
Original	0.08	0.16	0.06	0.16	0.03	0.12	0	0.12

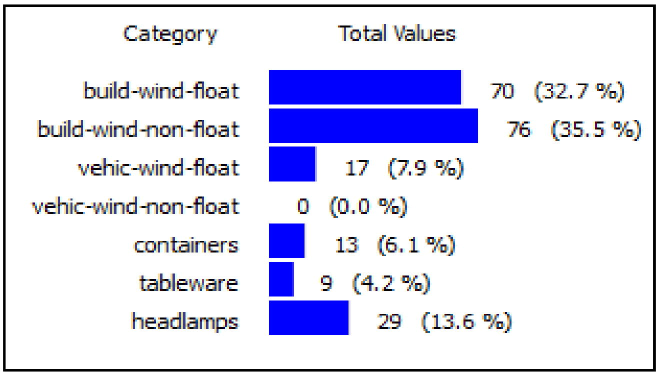
Table 3. Within Cluster Sum of Square Error

Fuzzy MF	Within Cluster Sum of Square Error							
	5 Clusters		10 Clusters		15 Clusters		20 Clusters	
	Iris	Glass	Iris	Glass	Iris	Glass	Iris	Glass
Bell	163	108.6	131	86.6	98	45.9	75	27.5
PI	160	74.5	116	62.1	92	52.2	71	26.4
Z	220	127.3	156	78.9	125	63.8	101	57.4
S	140	74.5	100	62.1	72	52.2	62	26.4
ET 2	152	121.6	83	61.0	60	55.3	47	48.7
Original	6.29	66.1	4.62	38.5	2.1	32.6	1.56	14.2



Category	Total Values
Iris-setosa	50 (33.3 %)
Iris-versicolor	50 (33.3 %)
Iris-virginica	50 (33.3 %)

IRIS



Category	Total Values
build-wind-float	70 (32.7 %)
build-wind-non-float	76 (35.5 %)
vehic-wind-float	17 (7.9 %)
vehic-wind-non-float	0 (0.0 %)
containers	13 (6.1 %)
tableware	9 (4.2 %)
headlamps	29 (13.6 %)

GLASS

Use of Membership Function in Privacy Preserved Data mining

Using Fuzzy based membership function, data values in a database are anonymize the sensitive values. In this paper, using S-Shape, PI shape & Bell shape membership functions, data standard database are converted to fuzzified data. The sanitized database generated using the fuzzy membership function are also used for analysis. IRIS & GLASS database from UCI standard Library are used in the analysis. The details of the dataset used are given in Table 1. K-mean clustering algorithm which is a popular data mining algorithm has been selected for analysis on the data sets before and after privacy preservation using the fuzzy approach.

Privacy preserved K-Mean Clustering

K-Mean Clustering Algorithm

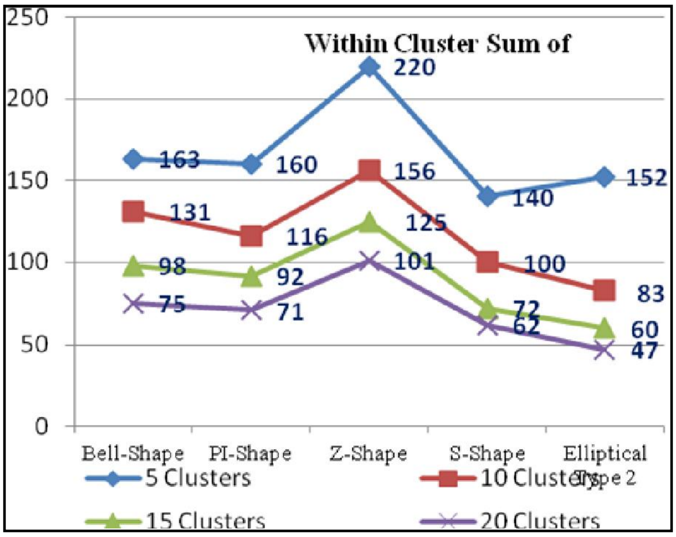
K-Mean clustering is a popular data mining algorithm used for grouping the data into various clusters. *k*-means clustering aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Given a set of observations (x_1, x_2, …, x_n), where each observation is a *d*-dimensional real vector, *k*-means clustering aims to partition the *n* observations into k ($\leq n$) sets $\mathbf{S} = \{S_1, S_2, \ldots, S_k\}$ so as to minimize the within-cluster sum of squares (WCSS). In other words, its objective is to find:

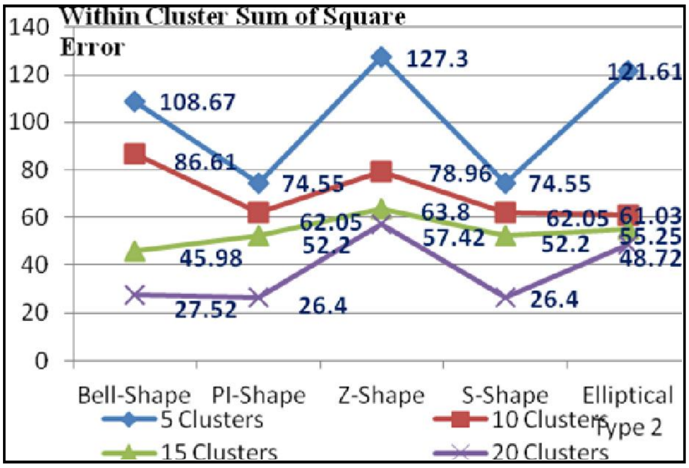
$$\underset{\mathrm{S}}{\mathrm{argmin}} \sum_{i=1}^{k} \sum_{x \in S1} ||x - \mu_i^{\,2}||$$

where μ_i is the mean of points in S_i.

The two key features of *k*-means algorithm which make it efficient are

- Euclidean distance is used as a metric and variance is used as a measure of cluster scatter.
- The number of clusters *k* is determined by running diagnostic checks to yield best results.





Privacy preserved K-Mean Clustering

In this analysis, K-Mean clustering is applied on the data base before and after applying fuzzy based privacy preservation to analyse the capability of the data to use it for data mining process along with maintaining privacy. Clustering error is compared with the original data. Weka 3.6.12 and Orange tools are used in the analysis of data mining process before and after applying fuzzy based privacy preservation on data sets. The findings and analysis results are presented in Section 5.

Results of Privacy preserved K-Mean Clustering

The performance of the fuzzy member ship functions on privacy preserved K mean clustering can be evaluated using the mean absolute error and within cluster Sum of Square Error. The comparison of various fuzzy membership functions used for Privacy preserved K mean clustering is shown in Table 2 and Table 3. Figure 6 and Figure 7 plots the comparison of within cluster sum of square errors for various membership functions.

Conclusion

Fuzzy logic system based algorithms will be used for Privacy Preserving Data mining considering to its simplicity and ease of implementation when compared to other cryptographic methods. This research helps to evaluate and analyze the impact of various membership functions to achieve greater secured data retrieval. This will lead to further researches to analyze the fuzzy based methods for various privacy preserved data mining algorithm to improve the current trends and algorithms available for Privacy preserved data mining.

REFERENCES

Charu. C., Agarwall, Philip, S. Yu, “Privacy Preserving Data Mining, Models and Algorithms”, ISBN 978-0-387-70991-8, 2008 (book style)

Jaideep, V., Chris, C., Michael, Z., “Privacy preserving Data Mining”, 2006, ISBN-13: 978-0-387-25886-8 (book style)

Jaiwan, H., Michaline, J., Jain, P. “Data Mining Concepts and Techniques”, Third Edition, 2012, ISBN 978-0-12-381479-1, (book style)

Jyotirmayee, R., Raghvendra, K.., “FP Tree Algorithm using Hybrid Secure Sum Protocol in Distributed Database”, *International Journal of Scientific & Engineering Research* Volume 4, Issue3, 2013, pp: 1 – 5, 2013 (journal style)

Pathak, F.A.N., Pandey, S.B.S., “Distributed changing neighbors k-secure sum protocol for secure multiparty computation”, Nirma University *International Conference on Engineering* (NUiCONE), pp: 1-3, 2013. (journal style)

Sheikh, R, Kumar, B., Mishra, D.K., Changing Neighbors k-Secure Sum Protocol for Secure Multi-Party Computation, (IJCSIS) *International Journal of Computer Science and Information Security*, Vol. 7, No. 1, 2010. (journal style)

Teo, S.G., Lee, V., Shuguo Han, “A Study of Efficiency and Accuracy of Secure Multiparty Protocol in Privacy-Preserving Data Mining”, 26th International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp: 85-90, 2012 (journal style)

Du, W. and M.J. Atallah, “Secure Multiparty Computation Problems and Their Applications: A Review and Open Problems,” In *proceedings of new security paradigm workshop*, Cloudcroft, New Maxico, USA, pages 11-20, Sep. 11-13 2001. (journal style)

Yanguang Shen, Hui Shao, Jianzhong Huang, “Research on Privacy Preserving Distributed C4. 5 Algorithm”, Third International Symposium on Intelligent Information Technology Application Workshops, IITAW '09. pp:216-218, 2009. (journal style)

Yehuda Lindell, and Benny Pinkas, “Secure Multiparty Computation for Privacy preserving data mining”, The *Journal of Privacy and Confidentiality*, pp: 59-98, 2009. (journal style)

****