

PRIVACY PRESERVING DATA MINING IN HEALTH CARE APPLICATIONS

First A. Dr. D. Aruna Kumari, Ph.d.; Second B. Ch.Mounika, Student, Department Of ECM, K L University, chittiprolumounika@gmail.com; Third C. A. Sai Kavya, Student, Department Of ECM, K L University, akaveetikavya@gmail.com; Fourth D. M. Anvesh Babu, Student, Department Of ECM, K L University, mrbl4.anvesh@gmail.com

Abstract

In recent years data analysis and protection are the two important parameters in any organization to improve the efficiency and privacy. In hospital management, privacy plays a crucial role. In order to hide the patient's related databases effective data mining algorithms are to be implemented. In this project we are going to create a database and perform different algorithms on it, so as to make sure that private data is preserved. Our project mainly deals with choosing the best algorithm for performing the privacy preservation on the databases mined. In our project we use a weak tool to implement randomization technique and analyze how efficiently the data is preserved.

Keywords: Randomization, Weka, Privacy.

1. Introduction

In this digital age, with the advent of hybrid technologies personal information is being used in many aspects for example online banking, subscription to any news letters, registration in any websites for further access, entering patient details to a hospital database. Large number of databases is generated daily. For analyzing this huge stream of data many organizations encourage Data mining in order to make good decisions. While mining the data there is chance that sensitive information is exposed. If the sensitive information is released it gives threat to the individual to avoid this various privacy methods are applied. For example consider a hospital database which contains an attribute "date of birth". This is represented as 11thnov 1993 in general manner. This field is used to analyze which disease is frequently occurring at what age. However during the analyzing of hospital databases for patient's record the patient age should not be revealed. Here some algorithms are implemented and date of birth is manipulated as **93. Here by doing such process the accurate age of the patient is disclosed and the occurrence of the disease will also be predicted so that the health statics will be predicted in well manner. In this paper we are going to analyze different methods of privacy preserving data mining such as randomization, anonymization and also by using

WEKA tool we are going to apply randomization technique on the clinical dataset and analyze the results.

1.1 Privacy Preserving in Data Mining

Privacy preserving data mining is an area of data mining to protect sensitive information. It gives a new ray to the field of datamining without interpreting the underlying data. Classification of privacy preserving data mining is given below [5].

Data Hiding	Data Perturbation <	Value Distortion -	Data Anonymization Data Swapping Other Randomization Techniques						
		Probability Distri	Analytical Method						
	Secure Multi-Party Computation (SMC) / Cryptographic Protocols								
	Distributed Data Mining (DDM)								
	Association Rule H	ubation							
Rule Hiding	Classification Rule	· · ·	Data Blocking { Parsimonious Downgrading						

2. Related Work

Privacy preserving data mining has many applications and to secure the individual privacy we have many methods like anonymization, data swapping, randomization.

2.1 Anonymization

In the model the many attributes in the dataset may be in-conjunction with the other values that are used to uniquely identify the records. For example if the fields like DOB and ZIPCODE are used to uniquely identify the records by removing the sensitive fields from the values. The main idea in this model Is to reduce granularity that is not to uniquely identify the Kth record from the (K-1) set of records. Generalization and suppression are two techniques implemented in anonymization method which are used to secure the individual data record. In generalization, original value is masked and displays the value within the interval. Suppression represents the sensitive values are masked by (*).



These two illustrated below

AGE	WEIGHT	DISEASE
20	45	ENT
40	60	HEART DISEASE

Table 2. Before generalization

AGE	WEIGHT	DISEASE
[15,30]	45	ENT
[35,55]	60	HEART DISEASE

Table 3. After generalization

NAME	ZIP CODE	DATE OF BIRTH
SAI	522004	ENT
DEEPTHI	522009	HEART DISEASE

Table 4. Before suppression

NAME	ZIP CODE	DATE OF BIRTH
SAI	5220**	ENT
DEEPTHI	5220**	HEART DISEASE

Table 5. After suppresion

2.2 Data Swapping

Additive and multiplication of noise are not only the distortion techniques. Data swapping is another for value distortion. Data swaping is dependent on the values of the neighboring the data unlike randomization which is implemented on the independent data. In this technique values of the records are interchanged and original data is not revealed to the researchers and privacy is preserved. Drawback of data swapping is the accurate results are not shown.

2.3 Randomization

This algorithm is used over the centralized data mining and allows the secure to sensitive data of an individual. For example, in hospital databases it has the patients related values like name,age,address,disease,contact. To apply the randomisation on this data we used the WEKA tool. We add some randomised value to the attributes which we want and we get modified dataset.

3. Description Of Work

The purpose of privacy preserving data mining is to analyze the data and also to maintain the privacy of an individual data by modifying the original dataset value to a nearest neighboring values. This can be done through number of techniques like data hiding, blocking, data randomization. Randomization is one of the technique which is used to modify the data from the original dataset. It is a prevailing method in present privacy preserving data mining studies. It disguise the values of the records by adding noise to the original data. The below figure shows, random noise that is random number is added to the original dataset and the original data now changes to the randomized dataset. This randomized dataset modifies the sensitive information of the individual data like age, contact number, pincode, address. So that the information is not revealed and also researchers can do the research on the particular application area.

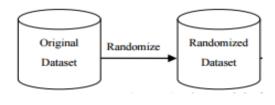


Fig 1. Model of randomization

The randomization method provides an effective yet simple way of preventing the user from learning precise data, which can be easily carried out at data collection phase to keep privacy data mining, because the noise added to a given record is independent of the behavior of other data records. When the randomization method is implemented, the data collection process consists of two steps. The first step is for the data providers to randomize their data and transmit the randomized data to the end user. In the second step, the end user estimates the original distribution of the data by employing a distribution reconstruction algorithm. In their randomization scheme, a random number is added to the value of a sensitive attribute. For instance, if ai is the value of a sensitive attribute, ai \Box ri, rather than ai, will appear in the database, where ri is a random noise drawn from some distribution. It is displayed that given the distribution of irregular noises, regenerate the distribution of the original data is possible. The method of randomization can be illustrated as follows. Consider a set of data records denoted by $A = \{a1 \dots aN\}$. For record ai ! A, we add a noise component which is drawn from the probability distribution fB (b). These noise components are drawn independently, and are denoted b1 . . . bN. Thus, the new set of distorted records are denoted by a1 +b1 . . . aN +bN. We denote this new set of records by c1 . . . cN. In general, it is assumed that the variance of the added noise is vast enough, so that the original data record values cannot be easily guessed from the distorted data. Thus, the original records cannot be recovered, but the distribution of the original data records can be recovered. Thus, if A be the random variable denoting the data distribution for the original record, B be the random variable describing the noise distribution, and C be the random variable denoting the final data record, we have:

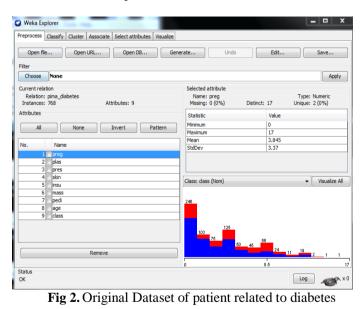
$$C = A + B$$
$$A = C - B$$



Their are two kinds of perturbation are possible with the randomization method. Additive perturbation, randomized noise is added to the data records. The data distributions can be recovered from the randomized records. Multiplicative perturbation, the random projection of random rotation techniques, projection or random rotation techniques are used in order to perturb the records. This method includes random noise based perturbation and randomized response scheme. Hence it results efficiently and high information method.

4. Experimental Results

To implement the randomization technique we used the software called "WEKA" (Waikato Environment for Knowledge Analysis). This has the applications like Explorer, Experimenter, Knowledge flow, simple CLI. In explorer we follow steps like preprocess, classification, clustering, association, attribute selection, visualization. In Preprocessing the data is choosen and modified. In classification train and test learning schemes that classify or perform regression. In clustering the data is divided into clusters. In association we learn association rules for the data. Attributes selection, selects the most relevant attributes in the data. In visualization it views an interactive 2D plot of the data. This tool helps us to get the precise outputs. We take the example of the patient data related to diabetes which consists the attributes like preg. age, class. This sensitive information is modified to nearest neighbouring value which is not seen when the researchers want to research on particular area.



The figure(2) describes the original datset of the patient related to diabetes before the filtering is done. Bar graphs represents the number of people registered for that disease.

Weka Explorer		
Preprocess Classify Cluster Associate Select attributes Visualize		
Open file Open URL Open DB Gene	erate Undo	Edit Save
Filter		
Choose Randomize -5 42		Apply
Current relation	Selected attribute	
Relation: pima_diabetes-weka.filters.unsupervised.instance.Rand Instances: 768 Attributes: 9	Name: preg Missing: 0 (0%) Distinct:	Type: Numeric 17 Unique: 2 (0%)
Attributes	Statistic	Value
All None Invert Pattern	Minimum	0
	Maximum	17
No. Name	Mean	3.845
1 preg 2 plas	StdDev	3.37
3 pres		
4 skin 5 insu	Class: class (Nom)	✓ Visualize Al
6 mass		
7 pedi	246	
8 age		
9 dass		
	103 75 50 45	3
Remove		24 11 19 2 1 1
	- 0 - 6	3.5
Status DK		Log

Fig 3. Preprocessing stage when randomization is applied

The figure(3) represents the preprocessed data after applying the randomization technique.

Preprocess Classify Cluster Associat	e Select attributes Visua	lize						
Classifier								
Choose RandomTree -K 0 -M 1	.0 -5 42 -D							
Test options	Classifier output							
Ose training set	Correctly Class	sified In:	stances	768		100	ł	
Supplied test set Set	Incorrectly Cl	assified :	Instances	0		0	ş	
Suppled test set	Kappa statisti	c		1				
Cross-validation Folds 10	Mean absolute	error		0				
Percentage split % 66	Root mean squa	red error		0				
	Relative absolu	ute error		0	8			
More options	Root relative	squared e	rror	0	8			
	Total Number o	f Instance	23	768				
Start Stop Result list (right-dick for options)				Precision				Area
		1	0	1	1	1	1	
09:25:40 - trees.RandomTree		1	0	1	1	1	1	
	Weighted Avg.	1	0	1	1	1	1	
	=== Confusion I	Matrix ===	-					
	a b <	classifi	ed as					
	500 0 a	= tested	negative					
	0 268 b	= tested	positive					
			III					P.
	•							

Fig 4. Classification of random tree classifier

The figure (4) represents the classification of data after random tree classifier is applied on the randomized data. The classifier output describes the correct instance for the particular class.



ISSN:2319-7900

Weka Classifier Visualize: 09:25:40 - trees.Rar	domTree (pima_diabetes-weka.filters.u 💷 😐	8	🛃 Weka : Instance info
X: class (Nom)	 Y: predicteddass (Nom) 	•	insu: 540.0
Colour: dass (Nom)	Select Instance	,	mass: 38.7 pedi: 0.24
		-1	age: 25.0
Reset Clear Open Save	Jitter 🛛	-1	predictedclass: tested_positive
Plot:pima_diabetes-weka.filters.unsupervised.instan	re.Randomize-S42 predicted		class: tested_positive
t e st e e d s t c c c c c c c c c c c c c c c c c c	x tested positive	т Т	<pre>Plot : 09:25:40 - trees.RandomTree (pima_ Instance: 765</pre>
Class colour			age: 32.0 predictedclass: tested_positive class: tested_positive
tested_nega	tive tested_positive		

Fig 5. Classifier output

The figure(5) describes the classifier error output on the randomized data which we got from the preprocessing stage.

Decome (and) (Date (and and (and (and (and (and (and (and	🔾 Weka Explorer					-	🔮 Weka : Instance info 💷 🗵		_	_
State		: Visualize				_	Dist : 09:21:01 - EM (nime disheter.)			
Lower Description Catere obst Coder mode Catere obst Instance _uniber: 4.2.0 Instance _uniber: 4.2.0 We to the start Start </th <th></th> <th></th> <th></th> <th></th> <th></th> <th>_</th> <th></th> <th></th> <th>A-DLO-L</th> <th></th>						_			A-DLO-L	
Cuter note Outer angut Pre-stage All All All All<	Choose EM -1 100 -N 4 -M 1.0E-6 -5 100								_	
a use ranges: Subject test is State Sub	Cluster mode	Clusterer output					-	Title	Subtitle Su	ubtle Em
Supplet tasts: St. Presentage selt St. 66 Presentage selt St. 66 Classe to dates relation Classes to dates relation (Presentage selt St. 66 Queue to dates relation Classes to dates relation (Presentage selt St. 66 Queue to dates relation Classes to dates relation (Presentage selt St. 66 Queue to dates relation Classes to dates relation (Presentage selt St. 66 Queue to dates relation Classes to dates relation (Presentage selt St. 66 Queue to dates relation Classes to dates relation (Presentage selt St. 66 State State State State </th <th>Use training set</th> <th></th> <th></th> <th></th> <th></th> <th>_</th> <th>plas: 145.0</th> <th></th> <th></th> <th></th>	Use training set					_	plas: 145.0			
Audie miss Image: Since Since Percentage pail 56 Classes 8 1 2 3 (Ven) data 1000 1000 Tope dates for wouldation 1000 1000 Tope dates for wouldation 1000 1000 State: 10.4 (0.122) (0.18) State: 10.4 (0.100) (0.14) State: 10.4 (0.12) (0.14) State: 10.4 (0.12) (0.14) State: 10.4 (0.12) (0.14) State: 10.4 (0.100)	· · ·	Number of cluster	CS: 4						4	
Createst data Artrihute 0 1 2 3 Clease to datas roduaton Pref: 0, 125 (0, 14) (0, 14) (0, 14) Store dates for visuataton Pref: 0, 125 (0, 14) (0, 14) (0, 14) Store dates for visuataton Pref: 0, 125 (0, 14) (0, 14) (0, 14) (0, 14) Store dates for visuataton Pref: 0, 125 Pref: 0, 125 (0, 14)								P	2	
0 Come Student of solutions Artributer 0 1 2 3 (Ven) doms (0.14) (0.22) (0.18) (0.14) (0.22) group doms (0.14) (0.22) (0.18) (0.14) (0.22) group doms mean 1.4509 1.7426 6.5052 4.565 logree athbute piss mean 11.659 1.7426 6.5052 4.565 Start Duo mean 11.659 1.7426 5.5151 5.6111 Claster Moule (0.2101 - DL ping dotter webs flag. Start Duo mean 117.625 101.751 5.631 174.625 101.751 5.631 174.625 101.751 5.631 102.655 101.650 102.655 101.650 102.655	Percentage split % 66		Cluster							
Wond dar: (0.14) (0.22) (0.18) (0.45) W Store dutes for multilation preg Class: toteld_contine Class: toteld_contine Igner athbuts end. 14.559 1.746 6.5564 4.656 Igner athbuts end. 1.559 1.746 6.5564 4.656 Stat Stat Stat Stat Stat State Stat Stat State State State State Stat Stat State State State State State Stat State	Classes to clusters evaluation	Attribute	0	1	2	3		00		
Bits State State <ths< th=""><th>Nom) date w</th><th></th><th>(0.14)</th><th>(0.22)</th><th>(0.18)</th><th>(0.45)</th><th></th><th>1</th><th></th><th></th></ths<>	Nom) date w		(0.14)	(0.22)	(0.18)	(0.45)		1		
Byte subtrive prop Claster:										
Igree attibutes std. dev. 1.555 1.422 Wash Custer Washer 993.01 - 20 jmg. doeter weak file. Start Stop plas Realthe (nyl) Fides (nul) Fides (nul) Start Stop plas Realthe (nyl) Fides (nul) Start Start Stop plas Realthe (nyl) Fides (nul) Start Start Stop plas Realthe (nyl) Fides (nul) Start Start Stop plas Realthe (nyl) Fides (nul) Start Stop plas Realthe (nyl) Fides (nul) Start Stop plas Realthe (nyl) Fides (nul) std. dev. 21.4821 21.4821 21.4821 Realthe (nyl) Fides (nul) std. dev. 11.4824 11.011 12.1821 Realthe (nyl) Fides (nul) std. dev. 12.4824 21.5855 90.1126 12.6835 12.4821 12.4821 std. dev. 12.4824 1.4823 <t< th=""><th>Store clusters for visualization</th><th>preg</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></t<>	Store clusters for visualization	preg								
Start Start Plas Rachte (period for option) maas 137,4259 101,4231 135,453 101,454						4.5615		E2	0.00	T Y
State 137,4359 101,4231 115,633 113,455 Cour Cube (lon) Sectorization State Sold Ger. 11,6284 11,7458 11,714 55 11,7	Ignore attributes	std. dev.	1.3058	1.4023	3.3757	3.4012	Weka Clusterer Visualize: 09:31:01 - EM (pim	a_diabetes-wek	atilte	
Alexant and projection of options Field of err. 11.0884 11.7845 31.174 32.4765 Non-option	Start Stop	plas					X: age (Num) 🔹 🕅	: class (Nom)		•
Solidor Solidor <t< th=""><th>Result list (right-click for options)</th><th>nean</th><th></th><th></th><th></th><th></th><th></th><th>elect Instance</th><th></th><th></th></t<>	Result list (right-click for options)	nean						elect Instance		
pros Action Action <th></th> <th>std. dev.</th> <th>31.0884</th> <th>17.8455</th> <th>30.1754</th> <th>32.6763</th> <th></th> <th></th> <th></th> <th></th>		std. dev.	31.0884	17.8455	30.1754	32.6763				
Beas 66.5586 61.5686 72.4522 72.1552 Respectively. Respeciintermananananananananananananananananananan							Re Clear Open Save	Jt	ter [
std. dev. 22,4591 20,4129 16,9455 17,245 skin men 31,2094 23,1555 10,1129 12,8954 ned. std. dev. 11,4324 11.011 10,2195 15,8954 men 201,2221 25,2352 177,3155 1 1 51 64 men 201,2221 25,2352 177,3155 1 51 64 men 21 51 64 12 51 64 State							Slations, dahates webs films you are institution	an Dandanian Ci	1 during d	
sin sin <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th>×</th> <th></th> <th>642 .</th>								×		642 .
State		sta. dev.	22.4931	20.0129	16.9653	17.200	e			
mem 31,294 23,155 30,112 12,023 3 rd. der. 13,424 13,031 10,215 15,839 3 <t< th=""><th></th><th>skin</th><th></th><th></th><th></th><th></th><th>1. Contract (1. Co</th><th></th><th></th><th></th></t<>		skin					1. Contract (1. Co			
insu mess 201,2221 65,2392 177,915 1 51 61 mass mass 1 51 61 1 50 61 Stata clasteri		nean	31,2094	23,1555	30,1128	12.0813				
insu mess 201,2221 65,2392 177,915 1 51 61 mass mass 1 51 61 1 50 61 Stata clasteri		std. dev.					1 5		1 1 1 1	W 1
1000 mem 201,2021 65,2052 177,5155 21 51 61 ass									100	
and. dev. 145.6506 51.4647 124.76 115.24 11 51 11 asss		insu					14		- 8	11 H
scd. dev. 168.8506 \$1.4617 124.76 11 \$1 1		nean	201.2321	85.2392	177.9105	0	<u>.</u>	****	1 53	6. I.
Sate Clasteri Clasteri Clasteri Clasteri Clasteri Clasteri		std. dev.	148.8806	51.4647	124.76	115.244	21 51	81		Ξ.
State Clasteri Clasteri Clasteri Clasteri Clasteri Clasteri		1811				_	Class colour		_	
380										
α	Status					-	cluster0 cluster1 clu	ister2	cluster3	
	OK .									
						-		_	-	_

Fig 6. dividing the dataset into clusters

The figure(6) describes the clustering analysis of the randomized data.

and the second		and the second second	* it states and states	and the second second			omize-S42		1.1	1000000	on: pima_	Sabetes					1000		-	ļ
No.	preg Numeric	plas Numeric	pres Numeric	skin Numeric	insu Numeric	mass Numeric	pedi Numeric	age Numeric	cli Nor	No.	preg Numeric	plas Numeric	pres Numeric	skin Numeric	insu Numeric	mass Numeric	pedi Numeric	age Numeric	class Nomina	
1	9.0	164.0	78.0	0.0	0.0	32.8	0.148	45.0	test .	1	6.0	148.0	72.0	35.0	0.0	33.6	0.627	50.0	0 tested	
2	4.0	134.0	72.0	0.0	0.0	23.8	0.277	60.0	test 🗐	2	1.0		66.0		a management		and the design of the		0 tested.	
3	5.0	166.0	72.0	19.0	175.0	25.8	0.587	\$1.0	test	3	8.0	183.0	64.0			-	0.672		0 tested.	
4	1.0	79.0	60.0	42.0	48.0	43.5	0.678	23.0) test	4	1.0	89.0	66.0	23.0	94.0	28.1	0,167	21.0	tested.	Ì
5	3.0	78.0	50.0	32.0	88.0	31.0	0.248	26.0) test	5	0.0	137.0	40.0	35.0	168.0	43.1	2.288		0 tested.	
6	7.0	147.0	76.0	0.0	0.0	39.4	0.257	43.0	test	6	5.0	116.0	74.0		and the second sec		and the second second		0 tested.	
7	1.0	71.0	62.0	0.0	0.0	21.8	0.416	26.0	test	7	3.0	78.0	50.0			31.0			0 tested.	
8	4.0	146.0	85.0	27.0	100.0	28.9	0.189	27.0) test	8	10.0	115.0	0.0	0.0	0.0	35.3	0.134		0 tested.	-
9	2.0	91.0	62.0	0.0	0.0	27.3	0.525	22.0	test	9	2.0	197.0	70.0	45.0	543.0				0 tested	
10	12.0	92.0	62.0	7.0	258.0	27.6	0.926	44.() test	10	8.0	125.0	96.0		and a local division of	and the latest of	Annual States		0 tested.	
11	9.0	119.0	80.0	35.0	0.0	29.0	0.263	29.0) test	11	4.0	110.0	92.0	0.0	0.0	37.6	0,191	and the second	tested.	
12	1.0	119.0	88.0	41.0	170.0	45.3	0.507	26.0	test.	12	10.0	168.0	74.0	0.0	0.0	38.0	0.537	34.0	tested.	ĺ
13	6.0	87.0	80.0	0.0	0.0	23.2	0.084	32.0	test	13	10.0	139.0	80.0			and the second second	1.441		0 tested	
14	9.0	164.0	84.0	21.0	0.0	30.8	0.831	32.0) test	14	1.0	189.0	60.0	23.0	846.0	30.1	0.398		0 tested.	
15	4.0	76.0	62.0	0.0	0.0	34.0	0.391	25.0) test	15	5.0	166.0	72.0	19.0	175.0	25.8	0,587	\$1.0	tested.	
16	4.0	136.0	70.0	0.0	0.0	31.2	1.182	22.0) test	16	7.0	100.0	0.0	0.0	0.0	30.0	0.484		0 tested	
17	1.0	130.0	60.0	23.0	170.0	28.6	0.692	21.0) test	17	0.0	118.0	84.0		and and includes	45.8			0 tested.	
18	2.0	101.0	58.0	35.0	90.0	21.8	0.155	22.0) testi	18	7.0	107.0	74.0	0.0	0.0	29.6	0.254	31.0	0 tested.	į
19	8.0	95.0	72.0	0.0	0.0	36.8	0.485	57.0) test	19	1.0	103.0	30.0	38.0	83.0	43.3	0.183		0 tested.	
20	6.0	105.0	70.0	32.0	68.0	30.8	0.122	37.0) test	20	1.0	115.0	70.0			34.6			tested.	
21	2.0	111.0	60.0	0.0	0.0	26.2	0.343	23.() test	21	3.0	126.0	88.0		and a strend state	39.3	0.704		0 tested	
22	9.0	112.0	82.0	24.0	0.0	28.2	1.282	50.0	test	22	8.0		84.0		remaining the	-	0.388		0 tested.	
23	6.0	80.0	66.0	30.0	0.0	26.2	0.313	41.0	test =	23	7.0		90.0				0.451		0 tested.	
1				m)	24	9.0						-		0 tested.	

Fig 7. comparision of datasets

The figure(7) describes the comparision of original dataset and randomized dataset. We can observe that data is modified and the original dataset is secured.

5. Conclusion

From the results above stated it is clearly observed how the randomization algorithm is being implemented on the dataset that is taken and the results are properly analysed how much privacy is being done on the dataset by choosing the algorithm.

References

[1] N. Zhang, "Privacy-Preserving Data Mining", Texas A&M University, pp.19-25, 2006.

[2] Z. Huang, W. Du, B. Chen, "Deriving Private Information from Randomized Data", In Proceedings of the ACM SIGMOD Conference on Management of Data, Baltimore, Maryland, USA, pp.37-48, 2005.

[3] H. Kargupta, S. Datta, Q. Wang, K. Sivakumar, "On the Privacy Preserving Properties of Random Data Perturbation Techniques", In Proceedings of the 3rd International Conference on Data Mining, pp.99-106, 2003.

[4] R. Agrawal, R. Srikant, "Privacy-Preserving Data Mining", ACM SIGMOD Record, New York, vol.29, no.2, pp.439-450,2000.

[5] Methods and Techniques to Protect the Privacy Information in Privacy Preservation Data Mining N.Punitha R.Amsaveni.

[6] VECTOR QUANTIZATION FOR PRIVACY PRESERVING CLUSTERING IN DATA MINING D.Aruna Kumari , Dr.Rajasekhara Rao and M.Suman.

[7] Agrawal R., Srikant R. Privacy-Preserving Data Mining. Proceedings of the ACM SIGMOD Conference, 2000.



[8] D.Aruna Kumari , Dr.K.Rajasekhara Rao,M.Suman "Privacy preserving data mining: LBG Algorithm" in International Journal of Database Management Systems(IJDMS) ISSN : 0975-5705 (Online); 0975-

5985(Print).

[9] D.Aruna Kumari, Dr. K.Rajesekhara Rao, M.Suman published a paper on "Vector quantization for privacy preserving clustering in data mining" in Advanced Computing: An International Journal (ACIJ -Nov 2012)
[10] D.Aruna Kumari, Dr. K.Rajesekhara Rao, M.Suman Tharun Maddu Published a paper on"Compression in privacy preserving data mining" in International Journal of Advanced Computer technology(IJACT ISSN : 2320-0790). April 2013.
[11] D.Aruna Kumari, Dr. K.Rajesekhara Rao and M.Suman Published a paper on "Privacy preserving data mining" in AIRCCJ Computer Science and Information technology (CS &IT)

Biographies

Proceedings.

FIRST A. Dr. D. Aruna Kumari, B. Tech, M. Tech, Ph.d received the Ph.D. degree from the K L University, Vaddeswaram, AndhraPradesh. Currently, She is an associate Professor of K L University. Her teaching and research areas include in data mining and published papers on privacy preserving in data mining.