Genetic Algorithm Based Confidential Data Protection in Privacy Preserving Data Mining

Dr. S. Vijayarani[1], M. Janakiram[2]

Assistant Professor, School of Computer Science and Engineering, Bharathiar University, Coimbatore, India [1]

M.Phil Research Scholar, School of Computer Science and Engineering, Bharathiar University, Coimbatore, India [2]

Abstract: Nowadays data sharing becomes obvious for successful completion of several tasks. In many places, data sharing has created lot of problems i.e. data modification, data mishandling and data misinterpretation. If we wish to share the datasets to others, first we have to verify whether the dataset has any sensitive or confidential data, if yes, then it is necessary and essential to protect that confidential data and then it can be shared to others. Privacy preserving data mining helps to perform the data mining tasks in a secured manner. Many privacy techniques and algorithms have been proposed by many researchers. This research work Genetic Algorithm based Masking Technique (GAMT) mainly concentrated on protecting confidential numerical attributes in a dataset using the genetic algorithm. Genetic algorithm concept is used for modifying the data items of the confidential attribute and to generate new data values for those attributes. The performance of Genetic Algorithm based Masking Technique is compared with the existing technique, named as additive noise. Results showed that GAMT has produced better results.

Keywords: Data mining, Privacy, Genetic algorithms, Additive noise, X-means, Filtered and Simple EM Clustering.

I. INTRODUCTION

Data mining is the process of discovering knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories. The discovered knowledge can be applied for decision making, process control, information management and query processing.

In recent years, data mining has been viewed as a threat to privacy because of the vast proliferation of electronic data maintained by [2] corporations. A number of techniques have been proposed for modifying or transforming the data which preserves sensitive data. The privacy preserving in data mining concerns the protection of confidential data from unauthorized users. The confidential data may be numerical, categorical or both. The protection of confidential data may give personal security to the workers of a company, government employers and sometimes it concerns with national security.

A key problem that arises in any mass collection of data is that of confidentiality. Many Industries enforce privacy while sharing their data to others. Because sharing of data can be beneficial to others and give mutual gain. Hence the privacy of industries' confidential data should be preserved from other corporate or public sectors. The problem of privacy preserving data mining [4] has numerous applications such as: medical databases, bioterrorism applications, homeland security applications including credential validation problem, identity theft, web camera surveillance, video surveillance, watch list problem and genomic privacy.

In order to protect the confidential data items, many techniques, methods and algorithms [5] [6] [7] are used in privacy preserving data mining. Some of the important research problems in the literature of privacy preserving data mining are statistical disclosure control; k-Anonymity, query auditing, cryptographic techniques and association rule hiding. In this research work, we have proposed a new masking technique namely Genetic Algorithm based Masking Technique for protecting the confidential numeric data items. Some of the existing techniques used for protecting confidential numerical data items are additive noise, rounding, perturbation and micro aggregation.

The rest of this paper is organized as follows. In section 2, we present an overview of the related works. Section 3 discussed about the problem definition and the proposed solution. Section 4 gives the experimental results and the performance analysis of the proposed technique and existing technique. Conclusions and future enhancements are given in section 5.

II. RELATED WORKS

S Md Zahidul Islam and Ljiljana Brankovic conducted a study on [8] privacy preserving data mining: A noise addition framework used a novel clustering technique. This study has presented a framework that used a few novel noise addition techniques which protected individual privacy and also maintained a high data quality. Noise is added to all attributes i.e. both numerical and categorical. The authors also presented a novel technique called DETECTIVE for clustering categorical values and this is used for noise addition purpose. A security analysis is also presented for measuring the security level of a data set.

Unil Yun, Jiwon Kim conducted a study [9] on "A fast perturbation algorithm using tree structure for privacy preserving utility mining". This study proposed a fast perturbation algorithm which is based on tree structure and it more rapidly performed database perturbation processes for preventing confidential data. Authors also presented

wide experimental outcome between their proposed method and the state of art algorithms by using both real and synthetic datasets. Authors have concluded that the proposed method has given excellent privacy preservation and this method required less execution time. The proposed algorithm also takes care of the scalability issue.
Sridhar Mandapati, Ratna babu chekka and Dr. Raveendra babu bhogapathi conducted a study on [10] a hybrid algorithm for privacy preserving in data mining. This work proposed a hybrid evolutionary algorithm using Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). Both GA and PSO in the proposed system worked with the same population. In the proposed framework, k-anonymity is accomplished by generalizing the original dataset. The hybrid optimization is used for searching optimal generalized feature set.

Vijayarani. S, Tamilarasi A, have discussed a study on [11] confidential numeric data protection in privacy preserving data mining. First the authors modified the original confidential data and then the modified data is shared by others. In this research work, authors have proposed two new techniques namely bit++ and bit—for protecting the confidential numeric attribute. The performances of the proposed techniques are compared with the existing techniques and they found the proposed techniques accuracy is good.

Domingo ferrer J et al, [12] [13] [14] discussed micro aggregation technique in statistical disclosure control technique. Raw micro data, i.e., individual records or information vectors are grouped into tiny aggregates before publication. Each combination should contain a minimum of k data vectors to stop revealing of individual data, wherever k could be a constant price predetermined by the data protector. Today, no precise polynomial algorithms are offered to perform best micro aggregation i.e. with lowest variability loss. Numerous strategies mentioned within the literature are ranking of data items is partitioned off into groups of fixed size. Within the multivariate case, ranking is performed by projecting data vectors onto one axis. The authors have characterized candidate optimal solutions to the multivariate and univariate micro aggregation issues. Within the univariate case, two heuristics supported, i.e. hierarchical clustering and genetic algorithms are introduced, and these are data oriented, hence they struggled to protect natural data aggregates. Within the multivariate case, fixed size and hierarchical clustering micro aggregation algorithms are given that do not need data to be projected onto a single dimension; such strategies clearly reduced variability loss as compared to conventional multivariate micro aggregation on projected data.

Hillol Dargupta and Krishnamoorthy sivakumar conducted a study on [15] Random data perturbation techniques and Privacy preserving data mining. This work first comments the random objects (particularly random matrices) which have “predictable” structures in the spectral domain and then it developed a random matrix-based spectral filtering technique to recover original data from the dataset distorted by adding random values. The projected method works by comparing the spectrum generated from the observed data with that of random matrices. This paper presented the theoretical foundation and extensive experimental results to show that in many cases random data distortion preserves very little data privacy. The analytical framework presented also points out several possible avenues for the development of new privacy-preserving data mining techniques.

III. PROBLEM DEFINITION

The core work of this research is to secure the sensitive numerical data items in the database. If the data holder wants to share or outsource the data to other third parties to perform data mining tasks. However, the data holder is not ready to give the original database, because it contains the sensitive data items. Hence, data holder modifies the sensitive data items by using protection techniques and provides it to the third parties. Here, we have to ensure that the protection techniques will not affect the data mining results. The modified database also provides the same data mining result as the original database results.

Proposed solution: In this research work, first the confidential numeric attributes are selected from the original database (D). Then the confidential data items are modified by the proposed protection technique genetic algorithm based masking technique and the existing technique, i.e. additive noise. This modified database D′ can be given to the organizations and data mining researchers. The data mining techniques, for example classification, clustering and association rule can be applied to D′. It should produce the same result as we get through D. In this research work, we have analyzed the statistical accuracy, privacy protection accuracy and the clustering accuracy. For clustering, the x-means, filtered and simple EM clustering algorithms are used. The system architecture of the proposed work is represented in figure 1.

i. Identify the confidential data items
ii. Modification
a. Proposed Technique
Genetic Algorithm based Masking Technique (GAMT)
b. Existing technique
Additive noise
iii. Performance Analysis.

The synthetic employee-income dataset is created with 20K instances. This data set has four attributes; emp_no, emp_name, age and income from which three are numerical and one is categorical. From this, income attribute is considered as confidential. Different sizes of data set with 3K, 5K, 10K and 20K instances are used for implementing the proposed and the existing protection techniques.

Additive Noise [11]

Additive noise is an existing protection technique for modifying the numerical data items. The principle of this concept is that a noise is added to trouble the sensitive data items of the original database.

DOI 10.17148/IJARCCE.2016.5442

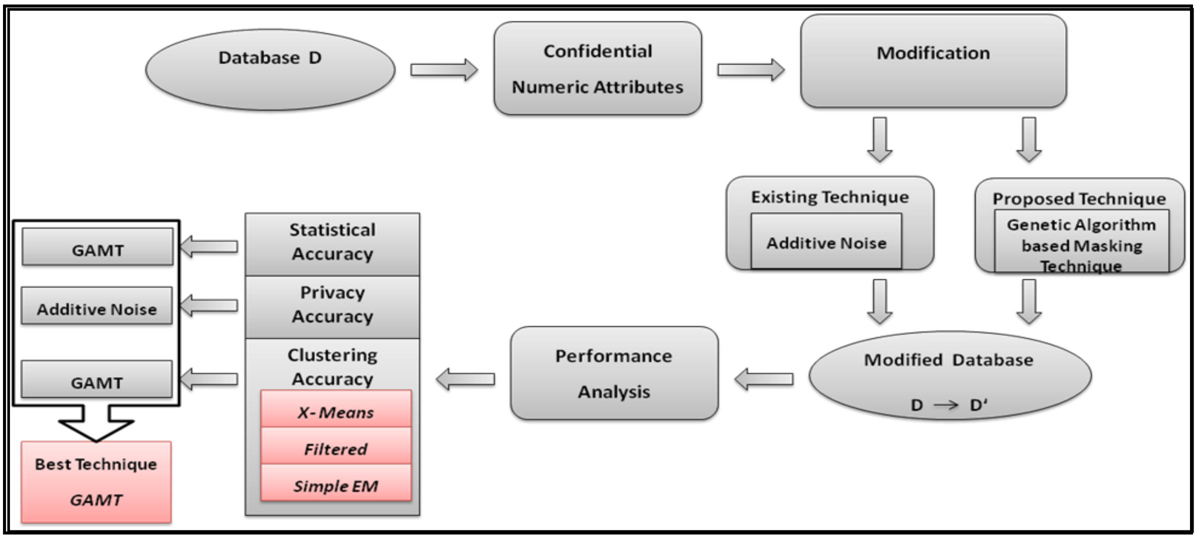

Figure 1: System Architecture

The data items of the sensitive numeric attribute are modified by adding noise value to the original data item or subtracting noise value from the original data item. The algorithm for additive noise technique is presented below.[xvii.]

```
i. Consider a database D which consists of T tuples, where
   D = {$T_1$,$T_2$,...$T_n$).
ii. Each tuple T in D consists of set of attributes
   T={$A_1$,$A_2$,....$A_p$} where $A_j \in T$, $T \in D$ and j=1,2,....p
iii. Identify the confidential numeric attribute $CA_R$
iv. Calculate the mean value ($\overline{X}$) of the data items $CA_R$
```

$$\overline{X} = \frac{\sum_{i=1}^{n} d_i}{n}$$

```
//n represents number of data items
v. Initialize countgre=0 and countmin=0
   //find the data items which are greater than or equal to mean
vi. if ($d_i$>=mean) then {
   // form a group which contains the data items which are greater than mean value
   6.1 Store $d_i$ into group 1
   6.2 Countgre = countgre+1}
   //find the data items which are less than mean value
vii. else{
   //form a group which contains the data items which are less than mean value
   7.1 Store $d_i$ into group2
   7.2 Countmin=countmin+1 }
viii. Calculate the value of noise1=(2*$\overline{X}$)/countgre
ix. Calculate the value of noise2=(2*$\overline{X}$)/countmin
x. //subtract the noise1 value from the data items of the group 1
   10.1 {for (i=1 to n)
   10.2 $d_{i=}$ $d_i$-noise1}
xi. //add the noise2 value from the data items of the group 2
   11.1 {for (i=1 to n)
   11.2 $d_{i=}$ $d_i$-noise2}
xii. Verify mean value of D′ which is same as D.
xiii. Verify the sum of noise1 and noise2 which is 0.
xiv. Repeat the same process for all the sensitive attributes
xv. Get the new modified data set D′
xvi. Stop
```

Algorithm for Genetic Algorithm based Masking Technique GAMT

This Genetic Algorithm based Masking is a new technique proposed for modifying the confidential data items of the data set. The fitness function calculation of this technique is to sort the sensitive data items in descending order and assign rank to each data items.

Input: Given Database *D*, *D={A_1,A_2,...A_N}* Where A_i is the attributes of *D*. *N*=|*D*|

Select the sensitive attributes $SA_i \in D$ {Where *i=1, 2, ..,n. n*=|SA_i|}

Population $P=SA_i$

Output:

Generate the Improved Population *P′* with the modified sensitive attributes SA_i'

Algorithm:

```
1. Initialization
      Consider the Population $P= SA_i$
      Select the sensitive numerical attributes one by one.
2. Fitness Function
      2.1 Arrange the data items $SA_i[data_j]$ in descending order.
      2.2 Assign rank to $SA_i[data_j]$. Rank starts from 1 to n.
3. Selection
      Select the two data items of $SA_i$ such as $SA_i[data_j]$, $SA_i[data_{j+1}]$ which has the highest rank.
4. Crossover
      Perform three point crossover from LSB
         $P'$=Crossover [$SA_i[data_j]$, $SA_i[data_{j+1}]$]
```

5. Repeat the steps 3 and 4 until no pair of data items found.
6. *Mutation*
   Consider the single data item. If $data_j$ is equal to $data_m$ in SA_i,
   $$SA_i' = Mutate[SA_i[data_m]]$$
   $$P' = P' \cup \{ SA_i' \}$$
7. Repeat steps 1 to 6 until no sensitive attribute (SA_i) is found.
8. *Terminate* the process with the required improved population P′.

The next step is selection i.e. to get the first two data items from the confidential attributes which has the highest rank. The next step is crossover; here three point crossovers is performed into the selected two data items. The same process is repeated for the remaining data items. At last if there is any individual data item is left that will be masked by mutation.

Mutation occurs from least significant bit (LSB) of the specified data item to most significant bit. First interchange the position of one's and ten's. Still the original data item alive interchange the ten's and hundred's and so on.

IV. EXPERIMENTAL RESULTS

The performance factors used for measuring the efficiency of the existing and the proposed protection techniques are statistical accuracy, privacy protection and clustering accuracy. The proposed protection technique is implemented using Visual Studio and MATLAB.

The experiment was conducted in Intel core 2 Duo processor with a CPU clock rate of 2.4 GHz. 500 GB Hard Disk and 3 GB RAM running windows operating system. The average percentages of these three confidential attributes are given in the following performance factors.

Statistical Accuracy

To find the statistical accuracy, the mean value of the data items of the confidential attribute(s) is considered. First, the mean value is calculated for original database D and then the mean value is calculated for the modified database D′ i.e. after performing modification using the existing and the proposed protection techniques. The mean values of D and D′ are compared. Table 1 shows the statistical accuracy.

Table 1: Statistical Accuracy of GAMT and Additive Noise

Dataset	GAMT (%)	Additive Noise (%)
3000	100	99.99
5000	100	100
10000	100	99.99
20000	100	99.99

Figure 2 illustrates the statistical mean accuracy of the existing and the proposed protection techniques. By analyzing the results, it is found that, for all the data sets, the GAMT has produced better results than the additive noise technique.

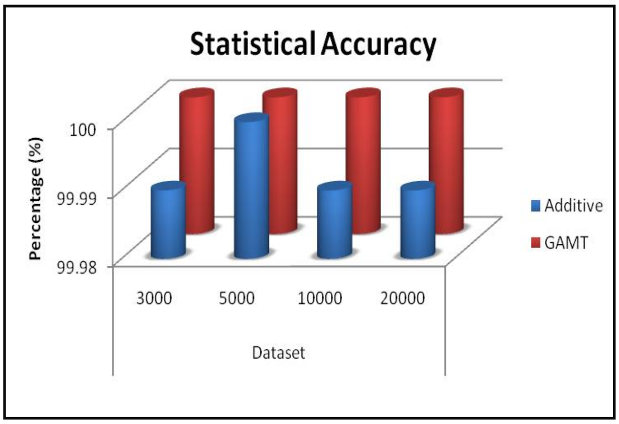

Figure 2: Statistical accuracy of GAMT and Additive Noise

Privacy Protection Accuracy

The privacy protection accuracy ensures that all the data items in the confidential attribute of the original database are modified by the existing and the proposed masking techniques.

Table 2: Protection Accuracy of GAMT and Additive Noise

Dataset	GAMT (%)	Additive Noise (%)
3000	99.93	100
5000	99.98	100
10000	99.99	100
20000	99.99	100

This performance factor is mainly used for finding out whether the protection techniques have properly modified the confidential data items or not. The original confidential data items are compared with the modified data items to verify whether both have the same value. If both the data items have different values, then privacy protection is good. Based on the results represented in Table 2 and illustrated in Figure 3, it is found that additive noise technique is better which has 100% privacy protection accuracy.

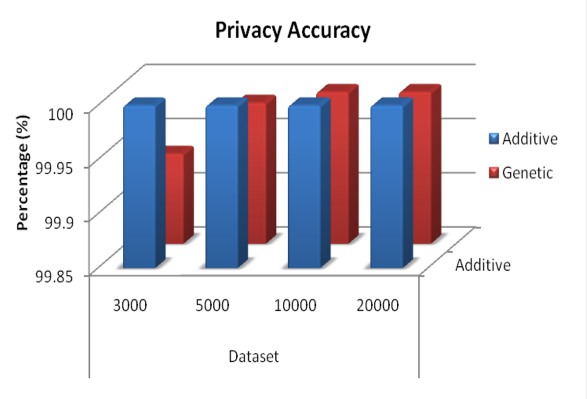

Figure 3: Privacy protection of GAMT and Additive Noise

Clustering Accuracy

In order to find the clustering accuracy, three clustering algorithms are used. They are X-means clustering, Filtered clustering and Expectation-Maximization algorithms.

X-means Clustering Algorithm

X-Mean [16] is extended version of k-means algorithm. In this X-means algorithm the centers are attempted to be split in its region. This algorithm starts with k which is equal to the lower bound of given range and continuously adds centroids where they are needed until the upper bound is reached. During this process, the centroid set that achieves the best score is recorded, and this is considered as final output.

1. Improve-Params

The Improve-Params operation is simple: it consists of running conventional k-means to convergence.

Table 3: X – means Accuracy of GAMT and Additive Noise

	Data set	Cluster 3 (%)	Cluster 5 (%)	Cluster 7 (%)	Cluster 10 (%)
GAMT	**3000**	99.53	99.26	97.53	96.13
	5000	100	99.84	94.99	98.53
	10000	99.79	98.65	97.58	96.30
	20000	99.94	99.98	99.57	96.14
Additive Noise	**3000**	99.53	99.19	97.48	96.81
	5000	99.92	99.92	98.67	98.33
	10000	99.84	98.61	97.68	95.84
	20000	99.92	98.48	99.29	96.13

2. Improve-Structure

The improve structure operation find outs where new centroid should appear.

From the Table 3 it is observed that GAMT has produced better accuracy compared to additive Noise. This is represented in Figure 4.

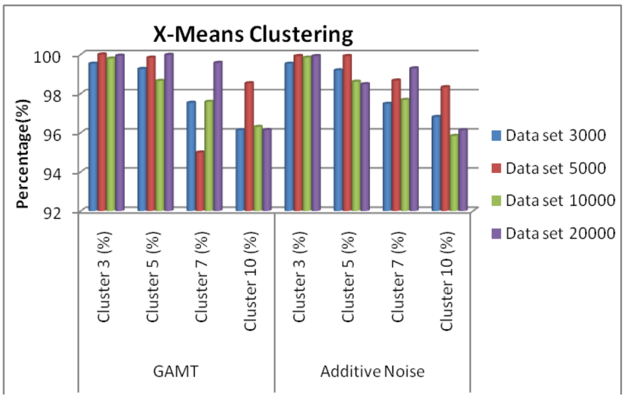

Figure 4: X-means Clustering Accuracy

Filtered Clustering Algorithm

This algorithm is based on storing the multi dimentional data points in Kd binary tree. First step is to compute a Kd tree for the given data points. For each node of the Kd tree, a set of candidate centers are maintained. If u is a leaf node, then the distances are computed from its associated data point to all the candidates in Z and assign the datapoint to its nearest center.

Algorithm: Filtered Cluster (KdNode n, Candidate dataSet D) {

```
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);}}}
```

Table 4: Filtered Cluster Accuracy of GAMT and Additive Noise

	Data set	Cluster 3 (%)	Cluster 5 (%)	Cluster 7 (%)	Cluster 10 (%)
GAMT	**3000**	100	99.80	97.18	97.05
	5000	99.92	97.96	96.35	98.18
	10000	99.69	98.60	97.64	95.91
	20000	99.99	98.49	99.38	99.02
Additive Noise	**3000**	99.93	99.79	96.90	95.16
	5000	100	98.19	96.05	97.83
	10000	99.84	98.54	97.60	95.91
	20000	99.94	98.51	99.30	99.02

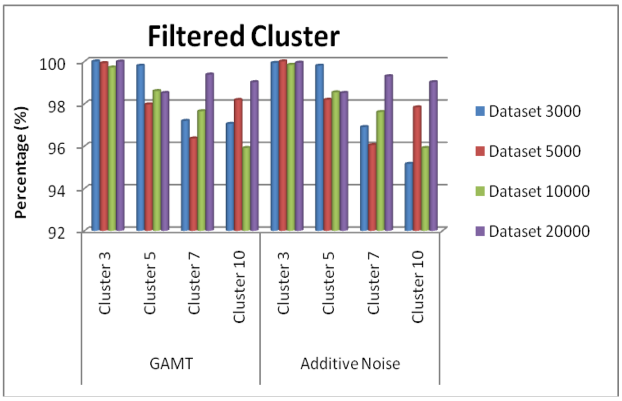

Figure 5: Filtered Clustering Accuracy

Figure 5 represents the filtered clustering accuracy of GAMT and additive noise. Table 4 also shows GAMT has higher accuracy compared to additive noise.

Expectation Maximization Algorithm [18]

The expectation maximization (EM) algorithm is a widely used maximum likelihood estimate of data distribution when data is partially missing or hidden [19]. The two steps are:

1. E (Exception) Step: This step is responsible to estimate the probability of each element belong to each cluster - $P(C_{-j}|X_{-K})$. Each element is composed by an attribute vector (X_K). The relevance degree of the points of

DOI 10.17148/IJARCCE.2016.5442

each cluster is given by the likelihood of each element attribute in comparison with the attributes of the other elements of cluster C_i.

$$P\left(\frac{C_j}{X}\right) = \frac{\left(|\Sigma_j(t)|^{-\frac{1}{2}} \exp^{hk} P_j(t)\right)}{\left((\Sigma_{k=1}^{M} |\Sigma_j(t)|^{-\frac{1}{2}} \exp^{nj} P_k(t))\right)}$$

Where X is input dataset
M is the total number of clusters
t is an instance and initial instance is zero.

2. M (Maximization) step: This step is responsible to estimate the parameters of the probability distribution of each class for the next step. First it computes the mean (M_i) of class j which is obtained through the mean of all points in function of the relevance degree of each point. Each iteration the covariance matrix is calculated using bayes theorem. The probability of occurrence of each class is computed through the mean of probabilities (Cj) in function of the relevance degree of each point from the class. Where x input dataset, $P_j(t+1)=1/N$

$$P_i(t+1) = \frac{1}{N}\sum_{K=1}^{N} P\left(\frac{C_j}{X_K}\right)$$

M is the total number of clusters t is an instance and initial instance is zero [19]
Table 5 shows the EM clustering accuracy.

Table 5: Simple EM Accuracy of GAMT and Additive Noise

	Data set	Cluster 3 (%)	Cluster 5 (%)	Cluster 7 (%)	Cluster 10 (%)
GAMT	**3000**	100	99.31	99.79	98.37
	5000	100	99.36	96.87	95.56
	10000	100	99.83	99.90	99.77
	20000	99.99	99.96	99.98	95.28
Additive Noise	**3000**	99.17	99.90	99.83	97.78
	5000	99.86	98.75	99.56	95.12
	10000	99.89	99.88	99.05	99.87
	20000	99.97	99.93	99.87	99.88

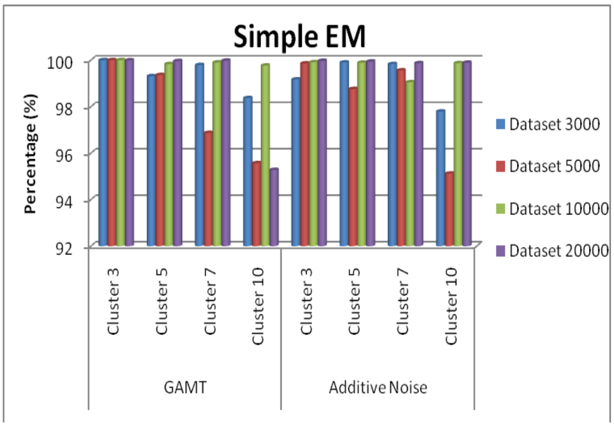

Figure 6: Simple EM Clustering Accuracy

From the Figure 6 it is observed that GAMT provides better accuracy than additive noise.

V. CONCLUSION AND FUTURE ENHANCEMENT

Confidential data protection and knowledge extraction are very complicated tasks in the data mining domain. This research work has discussed about confidential numeric data protection. New protection technique GAMT is proposed. The performance of the proposed technique is compared with the existing technique additive noise. By analyzing the experimental results, we come to know that the proposed GAMT protection technique has produced better results compared with the additive noise technique. In future, we would develop new protection techniques which will provide much better results compared to other techniques. In addition to this, new protection technique is to be developed for protecting the categorical attributes

REFERENCES

[1] Shweta Taneja et al, "A Review on Privacy Preserving Data Mining : Techniques and Research Challenges", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2310-2315
[2] R.Natarajan1, Dr.R.Sugumar, M.Mahendran, K.Anbazhagan, "A survey on Privacy Preserving Data Mining", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 1, Issue 1, MARCH 2012
[3] Li, Xiao-Bai and Sarkar, Sumit, "A Data Perturbation Approach to Privacy Protection in Data Mining" (2004). ICIS 2004 Proceedings, Paper 80.
[4] Charu C. Aggarwal, Philip S. Yu, "An Introduction to Privacy-Preserving Data Mining", Privacy-Preserving Data Mining Models and Algorithms, pp 1-9, 2008
[5] Jiawei Han, Micheline Kamber, Data Mining:Concepts and Techniques, The Morgan Kaufmann publisher, 2006
[6] Charu C. Aggarwal, Philip S. Yu, "A Survey of Randomization Methods for Privacy-Preserving DataMining", Privacy-Preserving Data Mining Models and Algorithms, pp 137-154, 2008
[7] Mohammad Naderi Dehkordi, Kambiz Badie, Ahmad Khadem Zadeh, "A Novel Method for Privacy Preserving in Association Rule Mining Based on Genetic Algorithms", Journal Of Software, Vol. 4, NO. 6, AUGUST 2009
[8] S Md Zahidul Islam , Ljiljana Brankovic, "Privacy preserving data\mining", Knowledge-Based Systems, Vol. 24 (2011) 1214–1223
[9] Unil Yun, Jiwon Kim, "A fast perturbation algorithm using tree structure for privacy preserving utility mining", Expert Systems with Applications, Vol. 42 (2015) 1149–1165.
[10] Sridhar Mandapati, Ratna babu chekka and Dr. Raveendra babu bhogapathi, "A hybrid algorithm for privacy preserving in data mining", I.J. Intelligent Systems and Applications, 2013, 08, 47-53.
[11] Vijayarani. S, Tamilarasi. A, "Confidential numeric data protection in privacy preserving data mining", International Journal of Computational Intelligence and Informatics, Vol. 1: No. 2, July - September 2011.
[12] Domingo-Ferrer J & Torra V, "Aggregation Techniques for Statistical Confidentiality", In: Aggregation operators: new trends and applications, pp. 260-271. Physica-Verlag GmbH, Heidelberg(2002a).
[13] Domingo-Ferrer J & Mateo-Sanz J.M, "Practical Data Oriented Micro Aggregation for Statistical Disclosure Control", IEEE Transactionas on Knowledge and Data Engineering, Vol. 14, no. 1, pp. 189-201,(2002b).
[14] Domingo-Ferrer J & Torra V, "Ordinal, Continuous and heterogeneous k-anonymity through microaggregation", Data Mining and knowledge Discovery, Vol.11, No. 2. pp. 195-212, 2005.
[15] Hillol Dargupta and krishnamoorthy sivakumar, "Random data perturbation techniques and Privacy preserving data mining", IEEE International Conference on Data Mining, 2003.
[16] Dan Pelleg, Andrew Moore, "X-means: Extending K-means with Efficient Estimation of the Number of Clusters".
[17] Angela Y. Wu, Senior Member, IEEE, An Efficient k-Means

DOI 10.17148/IJARCCE.2016.5442

Clustering Algorithm: Analysis and Implementation, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 7, JULY 2002.

[18] Prajwala T R, Sangeeta V I, "Comparative Analysis of EM Clustering Algorithm and Density Based Clustering Algorithm Using WEKA tool. ", International Journal of Engineering Research and Development, Volume 9, Issue 8 (January 2014), PP. 19-24.

[19] A Fast Convergence Clustering Algorithm Merging MCMC and EM Methods ,David Sergio Matusevich, Carlos Ordonez, Veerabhadran Baladandayuthapani, proceedings of the 22nd ACM international conference on Conference on information & knowledge management, October 2013, Pages 1525-1528.

BIOGRAPHY

Dr. S. Vijayarani, MCA., M. Phil., Ph.D is working as Assistant Professor in the Department of Computer Science, Bharathiar University, Coimbatore. Her fields of research interest are data mining, privacy and security issues in data mining and data streams. She has authored a book and published papers in the international journals and presented research papers in international and national conferences.

Mr. M. Janakiram has completed M.Sc in Computer Science. He is currently pursuing his M.Phil (Computer Science) in the Department of Computer Science, Bharathiar University, Coimbatore. His field of research interest is privacy preserving data mining. He has presented research papers in international and national conferences.

DOI 10.17148/IJARCCE.2016.5442