International Journal of Research
Available at https://edupediapublications.org/journals

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 08
April 2016

Study of Various Techniques for Privacy Preserving in Data Mining

Prateek Kumar Singh[1]
M. tech Student CSE Department, RGPV University, Madhya Pradesh Lakshmi Narain College of Technology, Jabalpur, MP, India Prateek8030@gmail.com

Naazish Rahim[2]
Assistant Professor HOD, CSE Department, RGPV University, Madhya Pradesh, Lakshmi Narain College of Technology, Jabalpur, MP, India naazish.rahim786@gmail.com

Sujeet Tiwari[3]
Assistant Professor CSE Department, RGPV University, Madhya Pradesh Lakshmi Narain College of Technology, Jabalpur, MP, India

Neelu Sahu[4]
Assistant Professor IT Department, CSVTU University, Chhattisgarh Government Engineering College, Bilaspur, Chhattisgarh, India neelu.sahu.12@gmail.com

ABSTRACT

Extracting previously unknown patterns from massive volume of data is the main objective of any data mining algorithm. In current days there is a tremendous expansion in data collection due to the development in the field of information technology. The patterns revealed by data mining algorithm can be used in various domains like Image Analysis, Marketing and weather forecasting. As a side effect of the mining algorithm some sensitive information is also revealed. There is a need to preserve the privacy of individuals which can be achieved by using privacy preserving data mining. In this paper we discuss different techniques used for privacy preservation in data mining. There are many Techniques available for privacy preserving we will study some of the important technique from them.

Keywords: Anonymization; Condensation; Cryptography; Distributed Data Mining; Perturbation; Privacy Preserving Data Mining (PPDM); Randomized Response.

II. INTRODUCTION

Since we are in an era of information explosion, it is very important to be able to find out useful information from massive amounts of data. Consequently, various data mining techniques have been developed. Data mining is often applied to fields such as marketing, sales, finance, and medical treatment. Besides, the rapid advance in Internet and communications technology has led to the emergence of data streams. Due to the consecutive, rapid, temporal and unpredictable properties [1,2] of data streams, the study of data mining techniques has transformed from traditional static data mining to dynamic data stream mining.

In recent years, enabled by the rapid development of various telecommunication technologies, many companies have improved their competitive edge by forming strategic alliances or information outsourcing, one after another. Consequently, many companies frequently expose private data while engaging in data analysis activities, which has led to grave threats to data privacy [4, 6]. For example, online marketing companies usually employ information technology outsourcing with a data mining company for cluster mining, in order to earn greater profits and to find the best target groups of customers. Therefore, how to preserve private data without disclosure while obtaining an accurate mining result in the process of mining will become increasingly difficult, which in turn has led to the development of Privacy- Preserving Data Mining techniques. Nonetheless, traditional Privacy-Preserving Data Mining is not applicable in a data stream environment which requires dynamic updating [7]?
This paper gives a survey of various privacy preserving data mining methods and analyses the representative methods for privacy preserving data mining, as well as points out their advantages and disadvantages.

There have been two types of privacy in data mining [9]. The first type of privacy is that the data is altered so that the mining result will preserve certain privacy. The second type of privacy is that the data is manipulated so that the mining result is not affected or minimally affected. The aim of privacy preserving data mining researchers is to develop data mining techniques that could be applied on data bases without violating the privacy of individuals. Many techniques for privacy preserving data mining have come up over the last decade. Some of them are statistical, cryptographic, randomization methods, k-anonymity model, l-diversity and etc.

The rest of this paper is organized as follows. In section 2 we will study the introductions of various methods of privacy preservation. In section 2.3 Classification framework of PPDM is discussed. In section 3 different PPDM technique for privacy preserving is analyzed. In section 4 contains the conclusions and future work.

II Related Work

2.1 METHODHS OF PRIVACY PRESERVATION:

2.1.1 The randomization method: The randomization method is a technique for privacy-preserving data mining in which noise is added to the data in order to mask the attribute values of records [3, 4]. The noise added is sufficiently large so that individual record values cannot be recovered. Therefore, techniques are designed to derive aggregate distributions from the perturbed records. Subsequently, data mining techniques can be developed in order to work with these aggregate distributions.

2.1.2 The k-anonymity model and l-diversity: The *k*-anonymity model was developed because of the possibility of indirect identification of records from public databases. This is because combinations of record attributes can be used to exactly identify individual records. In the *k*-anonymity method, we reduce the granularity of data representation with the use of techniques such as generalization and suppression. This granularity is reduced sufficiently that any given record maps onto at least *k* other records in the data. The *l*-diversity model was designed to handle some weaknesses in the *k*-anonymity model since protecting identities to the level of *k*-individuals is not the same as protecting the corresponding sensitive values, especially when there is homogeneity of sensitive values within a group. To do so, the concept of intra-group diversity of sensitive values is promoted within the anonymization scheme [11].

2.1.3 Distributed privacy preservation: In many cases, individual entities may wish to derive *aggregate results* from data sets which are partitioned across these entities. Such partitioning may be horizontal (when the records are distributed across multiple entities) or vertical (when the attributes are distributed across multiple entities). While the individual entities may not desire to share their entire data sets, they may consent to limited information sharing with the use of a variety of protocols. The overall effect of such methods is to maintain privacy for each individual entity, while deriving aggregate results over the entire data.

2.1.4 Downgrading Application Effectiveness: In many cases, even though the data may not be available, the output of applications such as association rule mining, classification or query processing may result in violations of privacy. This has led to research in downgrading the effectiveness of applications by either data or application modifications. Some examples of such techniques include association rule hiding [8], classifier downgrading [12], and query auditing [10].

2.5 *Encryption:* Encryption technique solves the problem of data privacy easily. Use of encryption techniques makes easy to conduct data mining among mutual un-trusted parties, or even between competitors. In distributed data mining encryption technique is used due to its privacy concern. Neglecting the efficiency of Encryption, it is used in both approaches of distributed data mining i.e. horizontally partitioned data and that on vertically partitioned data.

2.2. ASSESSMENT OF PRIVACY PRESERVING ALGORITHM

2.2.1 Heuristic-Based Techniques

A number of techniques have been developed for a number of data mining techniques like classification, association rule discovery and clustering, based on the premise that selective data modification or sanitization is an NP-Hard problem, and for this reason, heuristics can be used to address the complexity issues.

2.2.2 Cryptography-Based Techniques

A number of cryptography-based approaches have been developed in the context of privacy preserving data mining algorithms, to solve problems of the following nature. Two or more parties want to conduct a computation based on their private inputs, but neither party is willing to disclose its own output to anybody else. The issue here is how to conduct such a computation while preserving the privacy of the inputs. This problem is referred to as the Secure Multiparty Computation (SMC) problem. In particular, an SMS problem deals with computing a probabilistic function on any input, in a distributed network where each participant holds one of the inputs, ensuring independence of the inputs, correctness of the computation, and that no more

2.2.3 Reconstruction-Based Techniques

A number of recently proposed techniques address the issue of privacy preservation by perturbing the data and reconstructing the distributions at an aggregate level in order to perform the mining.

2.3 CLASSIFICATION FRAMEWORK FOR PPDM

To reduce the PPDM taxonomy into four levels: data distribution, purposes of hiding, data mining algorithms, and privacy preserving techniques (see figure 1).

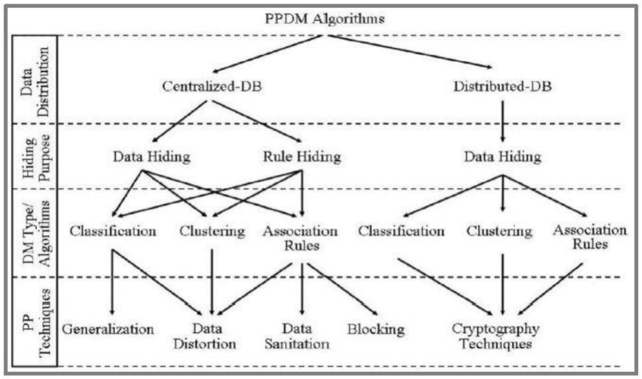

Figure 1: The Taxonomy of PPDM Algorithms

2.3.1 PPDM TECHNIQUES

For Recent years have witnessed extensive research in the field of PPDM. As a research direction in data mining and statistical databases, privacy preserving data mining received substantial attention and many researchers performed a good number of studies in the area. Since its inception in 2000 with the pioneering work of Agrawal & Srikant [7] and Lindell & Pinkas [8], privacy preserving data mining has gained increasing popularity in data mining research community. PPDM has become an important issue in data mining research [10-11]. As a outcome, a whole new set of approaches were presented to allow mining of data, while at the same time leaving out the releasing any secretive and sensitive information. The majority of the existing approaches can be classified into two broad categories [9]:

(i) Methodologies that protect the sensitive data itself in the mining process, and

(ii) Methodologies that protect the sensitive data mining results (i.e. extracted knowledge) that were produced by the application of the data mining. The first category refers to the methodologies that apply perturbation, sampling, generalization or suppression, transformation, etc. techniques to the original datasets in order to generate their sanitized counterparts that can be safely disclosed to untrustworthy parties. The goal of this category of approaches is to enable the data miner to get accurate data mining results when it is not provided with the real data. Secure Multiparty Computation methodologies that have been proposed to enable a number of data holders to collectively mine their data without having to reveal their datasets to each other.

The second category deals with techniques that prohibits the disclosure sensitive knowledge patterns derived through the application of data mining algorithms as well as techniques for downgrading the effectiveness of classifiers in classification tasks, such that they do not reveal sensitive information. In difference to the centralized model, the Distributed Data Mining (DDM) model accepts that the individual's information is distributed across multiple places. Algorithms are developed within this area for the problem of efficiently receiving the mining results from all the data through these distributed sources. A simple method to data mining over multiple sources that will not share data is to run existing data mining tools at each place independently and combine the results [12]. However, this will often fail to give globally valid output. Issues that cause a difference between local and global results include:

(i) Values for a single entity may be divided across sources. Data mining at individual sites will be unable to detect cross-site correlations.

(ii) The same item may be duplicated at different sites, and will be over-biased in the results.

(iii)At a single site, it is likely to be from a similar population.

PPDM tends to transform the original data so that the result of data mining task should not defy privacy constraints. Following is the list of five dimensions on the basis of which different PPDM Techniques can be classified [13]:

i. Data distribution

ii. Data modification

iii. Data mining algorithms

iv. Data or rule hiding

v. Privacy preservation

Data or Rule Hiding: This dimension refers to whether raw data or grouped data should be hidden. Data hiding means protecting sensitive data values, e.g. names, social security numbers etc. of some people. And Rule hiding means Protecting Confidential Knowledge in data, e.g. association rule. The difficulty for hiding aggregated data in the form of rules is very difficult, and for this purpose, typically heuristics have been developed.

Data Distribution: This dimension refers to the distribution of data. There are some of the approaches are developed for centralized data, while others refer to a distributed data scenario. Distributed data scenarios can be divided as horizontal data partition and vertical data partition. Horizontal distribution refers to these cases where different sets of records exist in different places, while vertical data distribution refers where all the values for different attributes reside in different places.

Data Modification: Data modification is used with the aim of change the unique values of a database that wants to be allowed to the public and in this way to guarantee high privacy protection. Methods of data modification include:

i. **Perturbation:** This is able to replacing attribute value by a new value (changing a 1-value to a 0-value, or adding noise)

ii. **Blocking:** which is the replacement of an existing attribute value with a "?"

iii. **Swapping:** This refers to interchanging values of individual record.

iv. **Sampling:** This refers to losing data for only sample of a population.

v. **Encryption:** many Cryptographic techniques are used for encryption.

Data Mining Algorithm: The data mining algorithm for which the privacy preservation technique is designed:

1. Classification data mining algorithm
2. Association Rule mining algorithms
3. Clustering algorithm

Privacy Preserving Techniques:

1. Heuristic-based techniques: It is an adaptive modification that modifies only selected values that minimize the effectiveness loss rather than all available values.

2. Cryptography-based techniques: This technique includes secure multiparty computation where a computation is secure if at the completion of the computation, no one can know anything except its own input and the results. Cryptography-based algorithms are considered for protective privacy in a distributed situation by using encryption techniques.

3. Reconstruction-based techniques: where the original distribution of the data is reassembled from the randomized data.

Based on these dimensions, different PPDM techniques may be classified into following five categories [13-15, 21, 22].

1. Anonymization based PPDM
2. Perturbation based PPDM
3. Randomized Response based PPDM
4. Condensation approach based PPDM
5. Cryptography based PPDM

We discuss these in detail in the following subsections.

III. PPDM TECHNIQUES FOR PRIVACY PRESERVING

3.1 Anonymization based PPDM

The basic form of the data in a table consists of following four types of attributes:

(i) Explicit Identifiers is a set of attributes containing information that identifies a record owner explicitly such as name, SS number etc.

(ii) Quasi Identifiers is a set of attributes that could potentially identify a record owner when combined with publicly available data.

(iii) Sensitive Attributes is a set of attributes that contains sensitive person specific information such as disease, salary etc.

(iv) Non-Sensitive Attributes is a set of attributes that creates no problem if revealed even to untrustworthy parties.

Anonymization refers to an approach where identity or/and sensitive data about record owners are to be hidden. It even assumes that sensitive data should be retained for analysis. It's obvious that explicit identifiers should be removed but still there is a danger of privacy intrusion when quasi identifiers are linked to publicly available data. Such attacks are called as linking attacks. For example attributes such as DOB, Sex, Race, and Zip are available in public records such as voter list.

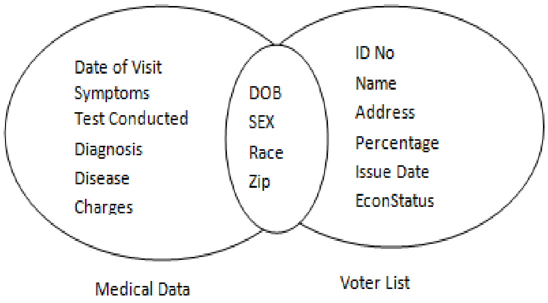

Fig.3 Linking Attack

Such records are available in medical records also, when linked, can be used to infer the identity of the corresponding individual with high probability as shown in fig.3.

Sensitive data in medical record is disease or even medication prescribed. The quasi-identifiers like DOB, Sex, Race, Zip etc. are available in medical records and also in voter list that is publicly available. The explicit identifiers like Name, SS number etc. have been removed from the medical records. Still, identity of individual can be predicted with higher probability. Sweeney [16] proposed k-anonymity model using generalization and suppression to achieve k-anonymity i.e. any individual is distinguishable from at least k-1 other ones with respect to quasi-identifier attribute in the anonymized dataset. In other words, we can outline a table as k-anonymous if the Q1 values of each raw are equivalent to those of at least k- 1 other rows. Replacing a value with less specific but semantically consistent value is called as generalization and suppression involves blocking the values. Releasing such data for mining reduces the risk of identification when combined with publically available data. But, at the same time, accuracy of the applications on the transformed data is reduced. A number of algorithms have been proposed to implement k-anonymity using generalization and suppression in recent years.

Although the anonymization method ensures that the transformed data is true but suffers heavy information loss. Moreover it is not immune to homogeneity attack and background knowledge attack practically [14]. Limitations of the k-anonymity model stem from the two conventions. First, it may be very tough for the owner of a database to decide which of the attributes are available or which are not available in external tables. The second limitation is that the k-anonymity model adopts a certain method of attack, while in real situations; there is no reason why the attacker should not try other methods. However, as a research direction, k-anonymity in combination with other privacy preserving methods needs to be investigated for detecting and even blocking k-anonymity violations.

3.2 Perturbation Based PPDM

Perturbation being used in statistical disclosure control as it has an intrinsic property of simplicity, efficiency and ability to reserve statistical information. In perturbation the original values are changed with some synthetic data values so that the statistical information computed from the perturbed data does not differ from the statistical information computed from the original data to a larger extent. The perturbed data records do not agree to real-world record holders, so the attacker cannot perform the thoughtful linkages or recover sensitive knowledge from the available data. Perturbation can be done by using additive noise or data swapping or synthetic data generation.

In the perturbation approach any distribution based data mining algorithm works under an implicit assumption to treat each dimension independently. Relevant information for data mining algorithms such as classification remains hidden in inter-attribute correlations. This is because the perturbation approach treats different attributes independently. Hence the distribution based data mining algorithms have an intrinsic disadvantage of loss of hidden information available in multidimensional records. Another branch of privacy preserving data mining that manages the disadvantages of perturbation approach is cryptographic techniques.

3.3 Randomized Response Based PPDM

Basically, randomized response is statistical technique introduced by Warner to solve a survey problem. In Randomized response, the data is twisted in such a way that the central place cannot say with chances better than a pre-defined threshold, whether the data from a customer contains correct information or incorrect information. The information received by each single user is twisted and if the number of users is large, the aggregate information of these users can be estimated with good quantity of accuracy. This is very valuable for decision-tree classification. It is based on combined values of a dataset, somewhat individual data items. The data collection process in randomization method is carried out using two steps [14]. During first step, the data providers randomize their data and transfer the randomized data to the data receiver. In second step, the data receiver rebuilds the original distribution of the data by using a distribution reconstruction algorithm. The randomization response model is shown in fig.4.

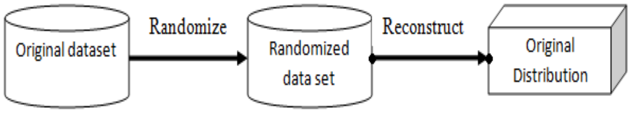

Fig.4 Randomization Response Model

Randomization method is relatively very simple and does not require knowledge of the distribution of other records in the data. Hence, the randomization method can be implemented at data collection time. It does not require a trusted server to contain the entire original records in order to perform the anonymization process [2]. The weakness of a randomization response based PPDM technique is that it treats all the records equal irrespective of their local density. These indicate to a problem where the outlier records become more subject to oppositional attacks as compared to records in more compressed regions in the data [8]. One key to this is to be uselessly adding noise to all the records in the data. But, it reduces the utility of the data for mining purposes as the reconstructed distribution may not yield results in conformity of the purpose of data mining.

3.4 Condensation approach based PPDM

Condensation approach constructs constrained clusters in dataset and then generates pseudo data from the statistics of these clusters [19]. It is called as condensation because of its approach of using condensed statistics of the clusters to generate pseudo data. It creates sets of dissimilar size from the data, such that it is sure that each record lies in a set whose size is at least alike to its anonymity level. Advanced, pseudo data are generated from each set so as to create a synthetic data set with the same aggregate distribution as the unique data. This approach can be effectively used for the classification problem. The use of pseudo-data provides an additional layer of protection, as it becomes difficult to perform adversarial attacks on synthetic data. Moreover, the aggregate behavior of the data is preserved, making it useful for a variety of data mining problems [2]. This method helps in better privacy preservation as compared to other techniques as it uses pseudo data rather than modified data. Moreover, it works even without redesigning data mining algorithms since the pseudo data has the same format as that of the original data. It is very effective in case of data stream problems where the data is highly dynamic. At the same time, data mining results get affected as huge amount of information is released because of the compression of a larger number of records into a single statistical group entity [14].

3.5 Cryptography Based PPDM

Consider a scenario where multiple medical institutions wish to conduct a joint research for some mutual benefits without revealing unnecessary information. In this scenario, research regarding symptoms, diagnosis and medication based on various parameters is to be conducted and at the same time privacy of the individuals is to be protected. Such scenarios are referred to as distributed computing scenarios [17].The parties involved in mining of such tasks can be mutual un-trusted parties, competitors; therefore protecting privacy becomes a major concern. Cryptographic techniques are ideally meant for such scenarios where multiple parties collaborate to compute results or share non sensitive mining results and thereby avoiding disclosure of sensitive information. Cryptographic techniques find its utility in such scenarios because of two reasons: First, it offers a well-defined model for privacy that includes methods for proving and quantifying it. Second a vast set of cryptographic algorithms and constructs to implement privacy preserving data mining algorithms are available in this domain. The data may be distributed among different collaborators vertically or horizontally.

All these methods are almost based on a special encryption protocol known as Secure Multiparty Computation (SMC) technology. SMC used in distributed privacy preserving data mining consists of a set of secure sub protocols that are used in horizontally and vertically partitioned data: secure sum, secure set union, secure size of intersection and scalar product. Although cryptographic techniques ensure that the transformed data is exact and secure but this approach fails to deliver when more than a few parties are involved. Moreover, the data mining results may breach the privacy of individual records

IV. CONCLUSION

Here in this paper a wide survey has been done on the various approaches for privacy preserving data mining and analyzed the major algorithms for data mining with their drawbacks... A complete study is done on for privacy preserving mining, and the methods for handling horizontally and vertically partitioned data Privacy preserving in mining has recently emerged as a vital field of study. As a new comer, Privacy-preserving in mining may offer a wide application prospect but at the same time it also brings us many issues / problems to be answered. In this study, we conduct a comprehensive survey on 29 prior studies to find out the current status of Privacy-preserving in mining. The limitation of privacy preserving as the increase in the dimension also analyzed and application which can employ the privacy algorithm is also studied.

In future work application of various optimizations should be deeply researched because Privacy should be achieved with accuracy.

REFERENCES

[1] Syed Md. Tarique Ahmad, et al "Privacy Preserving in Data Mining by Normalization" . IN: Proc. Of *International Journal of Computer Applications (0975 – 8887),Volume 96– No.6, June 2014.*

[2] S. Vijayarani, et al "Data Transformation Technique for Protecting Private Information in Privacy Preserving Data Mining". In: Proc. of Advanced Computing: An International Journal (ACIJ), Vol.1, No.1, November 2010.

[3]. Agarwal, R., Imielinski, T., Swamy, A. "Survey on privacy preservation in data mining", Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-210, 1993.

[4]. Srikant, R., Agarwal, R "Mining generalized association rules", In: VLDB'95, pp.479-488, 1994.

[5]Agrawal, R., Srikant, R, "Privacy-Preserving Data Mining", In: proceedings of the 2000 ACM SIGMOD on management of data, pp. 439-450, 2000.

[6] Ahmed HajYasien. Thesis on "PRESERVING PRIVACY IN ASSOCIATION RULE MINING" in the Faculty of Engineering and Information Technology Griffith University June 2007.

[7] R. Agrawal and R. Srikant. "Privacy Preserving Data Mining",ACM SIGMOD Conference on Management of Data, pp: 439-450,2000.
[8] Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining", Journal of Cryptology, 15(3), pp.36-54, 2000.

[9] Aris Gkoulalas-Divanis and Vassilios S. Verikios, "An Overview of Privacy Preserving Data Mining", Published by The ACM Student Magazine, 2010.

[10] 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.

[11] 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.

[12] 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.

[13] 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.
510, 2003.

[14] Gayatri Nayak, Swagatika Devi, "A survey on Privacy Preserving Data Mining: Approaches and Techniques", ternational Journal of Engineering Science and Technology, Vol. 3 No. 3, 2127-2133, 2011.

[15] Wang P, "Survey on Privacy preserving data mining", International Journal of Digital Content Technology and its Applications, Vol. 4, No. 9, 2010.

[16] Sweeney L, "Achieving k-Anonymity privacy protection uses generalization and suppression" International journal of Uncertainty, Fuzziness and Knowledge based systems, 10(5), 571-588, 2002.

[17] Benny Pinkas, "Cryptographic Techniques for Privacy preserving data mining", SIGKDD Explorations, Vol. 4, Issue 2, 12-19, 2002.

[18] D. Agrawal and C. Aggarwal, "On the Design and Quantification of Privacy Preserving Data Mining Algorithms", PODS 2001. pp: 247-255.

[19] L. Sweeney, "Achieving k-Anonymity Privacy Protection Using Generalization and Suppression", International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, vol.10, no.5, pp.571-588,2002.

[20] A. Machanavajjhala, J. Gehrke, D. Kifer, "l-Diversity: Privacy Beyond k-Anonymity", ACM Transactions on Knowledge Discovery from Data, pp.24-35,2007.

[21] T. Truta, B. Vinay, "Privacy Protection: p-Sensitive k-Anonymity Property", In Proceedings of the 22nd International Conference on Data Engineering Workshops, pp. 94-103, 2006.

[22] R.C.Wong, J.Y.Li, A.W. Fu, "(a, k)-Anonymity: An Enhanced k- Anonymity Model for Privacy-Preserving Data Publishing", In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.754-759, 2006.

[23] N.H. Li, T.C. Li, "t-Closeness: Privacy beyond k-Anonymity and l- Diversity", In Proceedings of the 23rd International Conference on Data Engineering, pp.106-115, 2007.

[24] X.K. xiao, Y.F. Tao, "M-Invariance: Towards Privacy Preserving Re-Publication of Dynamic Datasets", In Proceedings of the ACM Conference on Management of Data (SIGMOD), pp.689-700, 2007.

[25] X.K. Xiao, Y.F. Tao, "Personalized Privacy Preservation", In Proceedings of the ACM Conference on Management of Data (SIGMOD), pp.229-240, 2006.

[26] G. Loukides, J.H. Shao, "An Efficient Clustering Algorithm for k- Anonymisation", International Journal of Computer Science And Technology,vol.23, no.2, pp.188-202, 2008.

[27] J.L. Lin, M.C. Wei, "Genetic Algorithm-Based Clustering Approach for k-Anonymization", International Journal of Expert Systems with Applications,vol.36, no.6, pp.9784-9792, 2009.

[28] L.J. Lu, X.J. Ye, "An Improved Weighted-Feature Clustering Algorithm for k-Anonymity", In Proceedings of the 5th International Conference on Information Assurance and Security, pp.415-419, 2009.

[29] Z.H. Wang, J. Xu, W. Wang, B.L. Shi, "Clustering-Based Approach for Data Anonymization", Journal of Software,vol.21, no.4, pp.680-693, 2010.

[30] M. Kantarcioglu, C. Clifton, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data", IEEE Transactions on Knowledge and Data Engineering, vol.16, no.9, pp.1026-1037, 2004.

[31] Lindell, Yehuda, Pinkas, "Privacy preserving data mining", In Proceedings of the Advances in Cryptology–CRYPTO, pp.36– 54,2000.

[32] J. Vaidya, C. Clifton, "Privacy Preserving Association Rule Mining in Vertically Partitioned Data", In Proceedings of the 8th ACM SIGKDD

International Conference on Knowledge Discovery and Data Mining, pp.639-644, 2002.
[33] I. Ioannidis, A. Grama, M.J. Atallah, "A Secure Protocol for Computing Dot-Products in Clustered and Distributed Environments", In Proceedings of the 31st International Conference on Parallel Processing, pp.379-384, 2002.

[34] J. Vaidya, C. Clifton, "Privacy Preserving Association Rule Mining in Vertically Partitioned Data", In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.639–644, 2002.

[35] W.L. Du, Z.J. Zhan, "Building Decision Tree Classifier on Private Data", In Proceedings of the IEEE International Conference on Data Mining Workshop on Privacy, Security, and Data Mining, pp.1-8, 2002.

[36] J. Vaidya, C. Clifton, "Privacy Preserving Naive Bayes Classifier for Vertically Partitioned Data", In Proceedings of the 2004 SIAM International Conference on Data Mining, pp.522–526, 2004.

[37] J. Vaidya, C. Clifton, "Privacy-Preserving k-Means Clustering over Vertically Partitioned Data", In Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.206–215, 2003.

[38] Yao, C. Andrew, "How to Generate and Exchange Secrets", In Proceedings of the 27th IEEE Symposium on Foundations of Computer Science, pp.162-167, 1986. Sachin Janbandhu et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5279-5283.

Prateek kumar Singh Received the B.E degree in information Technology from R.I.T.E.E College, Raipur, India, in 2012. And pursuing M.Tech from the LNCT, Jabalpur, India. Email - Prateek8030@gmail.com

Naazish Rahim Working as an HOD & Assistant Professor, Department of Computer Science & Engineering at Lakshmi Narain College of Technology, Jabalpur, MP, India. Email-naazish.rahim786@gmail.com

Neelu Sahu Working as an Assistant Professor, Department of Information Technology Government Engineering College , Bilaspur, Chhattisgarh, India
Email- neelu.sahu.12@gmail.com