IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 4 (April. 2019), ||S (I) || PP 41-49

Privacy Preserving Sensitivity Data Classification Approach for Secured Mining

Boddu Mani kumar[1], KintaliPadmaja[2]
[1,2]Assistant Professor, Department of Computer Science and Engineering, Viswanadha Institute of Technology and Management, Visakhapatnam AP, INDIA.
Corresponding Author: Boddu Mani kumar

Abstract: Privacy preservation in data management and publishing has become an imperative research zone in the period of huge data. Effectively securing individual privacy in data publishing is particularly basic because of variety in close to home inclination and affectability. The outcomes of private data getting distributed are causing mental issues and unsettling influences in person's close to home life. This has set off the necessity to create different methodologies for privacy preservation in data publishing. Diversion hypothesis is one of the methodologies received for privacy preservation in data publishing. An examination of PPDP with PPDM has been done to investigate the utility. K-anonymization strategies have been the focal point of extraordinary research over the most recent couple of years. To frame a premise of advancement, a characterization technique has been defined to gather delicate and non-touchy data independently, where touchy data must be firmly secured. A theoretical methodology has been intended to accomplish customized privacy preservation in data publishing (PPPDP) in view of the order of affectability of person. In view of the affectability characterization, Game hypothesis has been proposed for accomplishing customized privacy preservation connected in the field of data publishing in the money related and banking part. This methodology can be stretched out to different divisions of data publishing like internet based life systems, marriage data, surveys and so on as future research road.

Keywords: Big data, Privacy, k-anonymity, l-diversity, t-closeness, Differential Privacy, Correlation, Privacy-Preserving Data Publishing (PPDP)

Date of Submission: 26-03-2019 Date of acceptance: 09-04-2019

I. INTRODUCTION

Data Anonymization is an innovation that changes over clear content into a non-intelligible structure. Data Anonymization procedure for privacy-saving data publishing has gotten a great deal of consideration as of late. Point by point data (likewise called as miniaturized scale data) contains data about an individual, a family unit or an association. Most prominent Anonymization methods are Generalization and Bucketization. [1]There are number of characteristics in each record which can be ordered as 1) Identifiers, for example, Name or Social Security Number are the traits that can be exceptionally recognize the people. 2) a few characteristics might be Sensitive Attributes(SAs, for example, ailment and pay and 3) some might be Quasi-Identifiers (QI, for example, postal district, age, and sex whose qualities, when taken together, can conceivably recognize a person. Data is considered anonymized notwithstanding when conjoined with pointer or family esteems that immediate the client to the starting framework, record, and esteem (e.g., supporting particular disclosure) and when anonymized records can be related, coordinated, and additionally conjoined with other anonymized records. Data Anonymization empowers the exchange of data over a limit, for example, between two offices inside an organization or between two offices, while lessening the danger of unintended exposure, and in specific situations in a way that empowers assessment and examination post Anonymization [1]. The two strategies contrast in the following stage. Speculation changes the QI-values in each can into "less explicit however semantically reliable" values so that tuples in a similar can can't be recognized by their QI esteems. In Bucketization, one isolates the SAs from the QIs by arbitrarily permuting the SA esteems in each can. The anonymized data comprise of a lot of cans with permuted touchy quality qualities.

A. Speculation

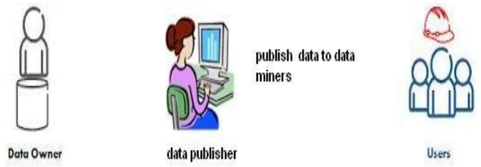
Speculation is one of the usually anonymized approaches, which replaces semi identifier esteems with qualities that are less-explicit however semantically reliable. At that point, all semi identifier esteems in a gathering would be summed up to the whole gathering degree in the QID space. [2] If no less than two exchanges in a gathering have particular qualities in a specific section (for example one contains a thing and alternate does not), at that point all data about that thing in the present gathering is lost. The QID utilized in this procedure incorporates every conceivable thing in the log. Because of the high-dimensionality of the semi identifier, with the quantity of conceivable things in the request of thousands, almost certainly, any speculation strategy would bring about incredibly high data misfortune, rendering the data futile [3]. With the end goal for speculation to be compelling, records in a similar basin must be near one another so that summing up the records would not lose a lot of data. Nonetheless, in high-dimensional data, most data focuses have comparable separations with one another. To perform data investigation or data mining assignments on the summed up table, the data examiner needs to make the uniform dispersion supposition that each incentive in a summed up interim/set is similarly conceivable, as no other dissemination presumption can be legitimized. This essentially diminishes the data utility of the summed up data. And furthermore in light of the fact that each property is summed up independently, relationships between's various properties are lost. So as to contemplate quality relationships on the summed up table, the data expert needs to expect that each conceivable blend of property estimations is similarly conceivable. This is an innate issue of speculation that forestalls powerful examination of quality relationships.

B. Bucketization

The primary, which we term bucketization, is to segment the tuples in T into containers, and after that to isolate the touchy characteristic from the non-delicate ones by haphazardly permuting the touchy quality qualities inside each basin. The sterilized data at that point comprises of the pails with permuted touchy qualities. In this paper [4] we use bucketization as the technique for developing the distributed data from the first table T, albeit every one of our outcomes hold for full-space speculation also. We currently determine our thought of bucketization all the more formally. Parcel the tuples into cans (i.e., evenly segment the table T as indicated by some plan), and inside each container, we apply a free arbitrary stage to the section containing S-values. The subsequent arrangement of cans, meant by B, is then distributed. For instance, in the event that the fundamental table T, at that point the distributer may distribute bucketization B .obviously, for included privacy, the distributer can totally cover the recognizing property (Name) and may in part veil a portion of the other non-touchy qualities (Age, Sex, Zip). For a basin b ϵ B, we utilize the accompanying documentation. While bucketization [1, 4] has preferred data utility over speculation, it has a few impediments. In the first place, bucketization does not counteract participation exposure. Since bucketization distributes the QI esteems in their unique structures, an enemy can see if an individual has a record in the distributed data or not. As appeared, 87 percent of the people in the United States can be extraordinarily recognized utilizing just three characteristics (Birth date, Sex, and Zip code). A small scale data (e.g., evaluation data) as a rule contains numerous different properties other than those three characteristics. This implies the participation data of most people can be derived from the bucketized table. Second, bucketization requires an unmistakable partition among QIs and SAs. Be that as it may, in numerous data sets, it is vague which qualities are QIs and which are SAs. Third, by isolating the touchy quality from the QI characteristics, bucketization breaks the property relationships between's the QIs and the SAs. Bucketization first parcels tuples in the table into basins and after that isolates the semi identifiers with the touchy quality by haphazardly permuting the delicate trait esteems in each can. The anonymized data comprise of a lot of cans with permuted touchy property estimations. Specifically, bucketization has been utilized for anonymzing high dimensional data. Be that as it may, their methodology accept a reasonable division among QIs and SAs. What's more, in light of the fact that the precise estimations of all QIs are discharged, participation data is unveiled. C. Cutting To improve the present best in class in this paper, we present a novel data Anonymization strategy called cutting [1]. Cutting allotments the data set both vertically and on a level plane. Vertical dividing is finished by gathering traits into segments dependent on the relationships among the qualities. Every section contains a subset of properties that are exceptionally corresponded. Level dividing is finished by gathering tuples into pails. At long last, inside each pail, values in every section are arbitrarily permutated (or arranged) to break the connecting between various segments. The fundamental thought of cutting is to break the affiliation cross sections, however to save the relationship inside every segment. This diminishes the dimensionality of the data and jam preferable utility over speculation and bucketization. Cutting jelly utility since it bunches very connected qualities together, and jam the relationships between's such properties. Cutting ensures privacy since it breaks the relationship between uncorrelated qualities, which are rare and along these lines recognizing. Note that when the data set contains QIs and one SA, bucketization needs to break their relationship; cutting, then again, can gather some QI properties with the SA, saving trait connections with the delicate property. The key instinct that cutting gives privacy protection is that the slicing.

II. PRIVACY-PRESERVING DATA PUBLISHING

PPDP can be represented in the form of stages as shown in figure 1 where the first stage contains Data Publisher, who collects information from the individuals and store it in large databases. Second stage where data is published to recipient who can be data miners and also can be adversaries. The overall execution can be divided as data collection phase and data publishing phase. In data collection phase the actual data is collected from record owners by Data Publisher. Data Publisher in turn modifies the data suitable for Data Recipient in a way which ensures privacy this phase is called publishing phase.



(Individual, medical, organization data)   (Data miners, adversaries)

Figure 1 Data collection and data publishing

Privacy preserving techniques

The basic idea of PPDP is to develop Techniques so that the sensitivity of data will not release. PPDP has been classified into following categories:

Randomization method

It is the process of adding noise to the original data in order to mask attributes from disclosure. There are different ways of Randomization [1] the simplest is additive randomization. One of the disadvantages is that results are approximate and has huge information loss.

Data Swapping

It is a method in which values of records are swapped which maintains the statistical inference of the relation in order to preserve privacy. This technique can be used in combination with other frameworks such as *k*-anonymity.

Cryptographic approach

In this approach there are lots of algorithm to implement cryptographic methods and it is a well-defined model for privacy.

Anonymization Approach

The most common approach to preserve sensitivity is to modify the contents of the record owners before publishing the data this approach itself is called Anonymization [2].

In the most basic form of privacy-preserving data publishing (PPDP), the data holder has a table of the form: Explicit Identifier, Quasi Identifier[8], Sensitive Attributes, non-Sensitive Attributes, where Explicit Identifier is a set of attributes, such as name and social security number (SSN), containing information that explicitly identifies record owners, Quasi Identifier is a set of attributes that could potentially identify record owners, Sensitive Attributes consist of sensitive person specific information such as disease, salary, and disability status and Non-Sensitive Attributes contains all attributes that do not fall into the previous three categories. Most works assume that the four sets of attributes are different. Most works assume that each record in the table represents a unique record owner.

III. PRIVACY MODELS AND ATTACK MODELS

There are many privacy models and attack models in the privacy preserving data publishing; following are most relevant to our topic.

Privacy Models

Privacy models which gives different algorithms and different techniques for achieving privacy from different attacks and disclosures.

K-anonymity

K-anonymous database is a database where attributes in row are suppressed or generalized until each row is identical with at least k-1 other rows. K-Anonymity [2] thus prevents definite database linkages. K-Anonymity guarantees that the data released is accurate. Kanonymity proposal focuses on two techniques in particular: generalization and suppression. To protect respondents' identity when releasing microdata, data holders often remove or encrypt explicit identifiers, such as names and social security numbers. De-identifying data, however, provide no guarantee of anonymity. Released information often contains other data, such as birth date, sex, and ZIP code that can be linked to publicly available information to re-identify respondents and to infer information that was not intended for release. One of the emerging concepts in microdata protection [3] is k-anonymity, which has been recently proposed as a property that captures the protection of a microdata table with respect to possible re-identification of the respondents to which the data refer. K-anonymity demands that every tuple in the microdata table released be indistinguishably related to no fewer than *k* respondents. One of the interesting aspect of *k*-anonymity is its association with protection techniques that preserve the truthfulness of the data. The first approach toward privacy protection in data mining was to perturb the input (the data) before it is mined. The drawback of the perturbation approach is that it lacks a formal framework for proving how much privacy is guaranteed. At the same time, a second branch of privacy preserving data mining was developed, using cryptographic techniques [9]. Thus, it falls short of providing a complete answer to the problem of privacy preserving data mining. One definition of privacy which has come a long way in the public arena and is accepted today by both legislators and corporations is that of k-anonymity. The guarantee given by k-anonymity is that no information can be linked to groups of less than k individuals. Generalization for kanonymity losses considerable amount of information, especially for high-dimensional data. Limitations of kanonymity are*:*(1) it does not hide whether a given individual is in the database, (2) it reveals individuals' sensitive attributes , (3) it does not protect against attacks based on background knowledge , (4) mere knowledge of the k-anonymization algorithm can violate privacy, (5) it cannot be applied to high-dimensional data[4] without complete loss of utility , and (6) special methods are required if a dataset is anonymized and published more than once. It preserves record linkage.

l- diversity

The next concept is “l-diversity” [6]. Say if you have a group of k different records that all share a particular quasi-identifier. That’s good, in that an attacker cannot identify the individual based on the quasi-identifier. But if the value they’re interested in, (e.g. the individual’s medical diagnosis) is the same for every value in the group. The distribution of target values within a group is referred to as “*l*-diversity”. Currently, there exist two broad categories of *l*-diversity techniques: *generalization* and *permutation*-based. An existing generalization method would partition the data into disjoint groups of transactions, such that each group contains sufficient records with *l*-distinct, well represented sensitive items. It preserves both record and attributes linkage.

t-closeness

t-closenessformalizes the idea of global background knowledge [10] by making it compulsory the distribution of a sensitive attribute in any equivalence class is close to the distribution of the attribute in the overall table . This effectively limits the amount of individual-specific information an observer can learn. Intuitively, privacy is measured by the information gain of an observer. Before seeing the released table, the observer has some prior belief about the sensitive attribute value of an individual. After seeing the released table, the observer has a posterior belief. Information gain can be represented as the difference between the posterior belief and the prior belief. The novelty of approach is that separate the information gain into two parts: that about the whole population in the released data and that about specific individuals. It preserves both from attribute linkage and probabilistic attack.

Attack Models

When publishing microdata, there are four types of information disclosure threats.

Record Linkage

The first type is membership disclosure [1], when the data to be published is selected from a larger population and the selection criteria are sensitive (e.g., when publishing datasets about diabetes patients for research purposes), it is important to prevent an adversary from learning whether an individuals record is in the data or not.

Table Linkage

The second type is identity disclosure[3], which occurs when an individual is linked to a particular record in the released table. In some situations, one wants to protect against identity disclosure when the adversary is uncertain of membership. In this case, protection against membership disclosure helps protect against identity disclosure. In other situations, some adversary may already know that an individual’s record is in the published dataset, in which case, membership disclosure protection either does not apply or is insufficient.

Attribute Linkage

The third type is attribute disclosure, which occurs when new information about some individuals is revealed, i.e., the released data makes it possible to infer the attributes of an individual more accurately than it would be possible before the release. Similar to the case of identity disclosure, we need to consider adversaries who already know the membership information. Identity disclosure leads to attribute disclosure. Once there is identity disclosure, an individual is re-identified and the corresponding sensitive value is revealed. Attribute disclosure can occur with or without identity disclosure, e.g., when the sensitive values of all matching tuples are the same.

Probabilistic Attacks

The fourth type is Probabilistic attack. There is another family of privacy models that does not focus on exactly what records, attributes, and tables the attacker can link to a target victim, but focuses on how the attacker would change his/her probabilistic belief on the sensitive information of a victim after accessing the published data. In general, this group of privacy models aims at achieving the uninformative principle, whose goal is to ensure that the difference between the prior and posterior beliefs is small.

Table 1 describes the different privacy models and attacks from which they give protection for e.g. K-anonymity preserves data from record linkage that is membership disclosure, l-diversity protects from record and attributes linkage and t- closeness protects from attribute and probabilistic attack.

Table 1: Privacy Models v/s Attack models

Privacy model	Attack model		
	Record Linkage	Attribute Linkage	Probabilistic attack
K-anonymity	√		
l-diversity	√	√	
t-closeness		√	√

IV. Proposed System

Generally in privacy preservation there is a loss of security. The privacy protection is impossible due to the presence of the adversary's background knowledge [10] in real life application. Data in its original form contains sensitive information about individuals. These data when published violate the privacy. The current practice in data publishing relies mainly on policies and guidelines as to what types of data can be published and on agreements on the use of published data. The approach alone may lead to excessive data distortion or insufficient protection. Privacy-preserving data publishing (PPDP) provides methods [8] and tools for publishing useful information while preserving data privacy. Many algorithms like bucketization, generalization have tried to preserve privacy however they exhibit attribute disclosure. So to overcome this problem an algorithm called slicing is used. This algorithm consists of three phases: attribute partitioning, column generalization, and tuple partitioning. Attribute Partitioning this algorithm partitions attributes so that highly correlated attributes are in the same column. This is good for both utility and privacy. In terms of data utility, grouping highly correlated attributes preserves the correlations among those attributes. In terms of privacy, the association of uncorrelated attributes presents higher identification risks than the association of highly correlated attributes because the associations of uncorrelated attribute values is much less frequent and thus more identifiable. Column Generalization: First, column generalization may be required for identity/membership disclosure protection. If a column value is unique in a column, a tuple with this unique column value can only have one matching bucket. This is not good for privacy protection, as in the case of generalization/bucketization where each tuple can belong to only one equivalence-class/bucket. Tuple Partitioning The algorithm maintains two data structures: 1) a queue of buckets Q and 2) a set of sliced buckets SB. Initially, Q contains only one bucket which includes all tuples and SB is empty. For each iteration, the algorithm removes a bucket from Q and splits the bucket into two buckets. If the sliced table after the split satisfies l-diversity, then the algorithm puts the two buckets at the end of the queue Q Otherwise, we cannot split the bucket anymore and the algorithm puts the bucket into SB. When Q becomes empty, we have computed the sliced table. The set of sliced buckets is SB.

V. PPDP Results and Discussions

Consider table as an example of micro data.

Table 1: The Original Tablegeneralization

Age	**Sex**	**Zipcode**	**Disease**
22	M	47906	dyspepsia

22	F	47906	flu
33	F	47905	flu
52	F	47905	bronchitis
54	M	47302	flu
60	M	47302	dyspepsia
60	M	47304	dyspepsia
64	F	47304	gastritis

Generalization [7], [9] is the commonly used anonymized approach, in which quasi-identifier values are replaced with less-specific but semantically consistent values. Then, all quasiidentifier values in a group would be generalized to the entire group. It uses the k-anonymity [4], [8], [9] model of privacy.

Steps

1. Removes identifiers from the data.
2. Partition tuples into buckets.
3. Transform QI values in each bucket with less-specific but semantically consistent values.

Table 2: The generalized Table

Age	**Sex**	**Zipcode**	**Disease**
[20-52]	*	4790*	dyspepsia
[20-52]	*	4790*	flu flu
[20-52]	*	4790*	bronchitis
[20-52]	*	4790*	
[54-64]	*	4730*	flu
[54-64]	*	4730*	dyspepsia
[54-64]	*	4730*	dyspepsia
[54-64]	*	4730*	gastritis

Limitations

- K-Anonymity suffers from the curse of dimensionality [2].
- The data analyst has to make the uniform distribution assumption that every value in each generalized set is equally possible.
- Correlations between different attributes are lost.

Generalization suffers with two types of attacks:-

- Background Knowledge Attack.
- Homogeneity Attack.

Bucketization

Bucketization [6], [3], [10] firstly, Partitions tuples into buckets, and then separate the sensitive attribute from the non-sensitive by randomly permuting the sensitive attribute values within each bucket. It uses l-diversity [5] model of privacy.

Steps

1. Removes identifiers from the data.
2. Partition tuples into buckets.
3. Separate the SAs from the QIs by randomly permuting the SA values in each bucket.

Limitations

- Does not prevent membership disclosure.
- Requires a clear separation between QIs and SAs.
- Breaks the attribute correlations between the QIs and the SAs.

Bucketization suffers with two types of attacks:

- Skewness attack.
- Similarity attack.

Skewness attack is defined as overall distribution is skewed by satisfying the l-diversity and does not prevent membership disclosure. Similarity Attack is defined as sensitive attributes in a column are distinct but semantically similar.

Table 3: The Bucketized Table

Age	**Sex**	**Zipcode**	**Disease**
22 22 33 52	M F F F	47906 47906 47905 47905	dyspepsia flu flu bronchitis
54 60 60 64	M M M F	47302 47302 47304 47304	flu dyspepsia dyspepsia gastritis

Anatomy

Table 4: The Anonymized Tables

Age	**Sex**	**Zipcode**	**Group-ID**
22	M	47906	1
22	F	47906	1
33	F	47905	1
52	F	47905	1
54	M	47302	2
60	M	47302	2
60	M	47304	2
64	F	47304	2

(a) The quasi-identifier table (QIT)

Group-ID	**Disease**	**Count**
1	flu	2
1	dyspepsia	1
1	bronchitis	1
2	gastritis	1
2	flu	1
2	dyspepsia	2

(b) The sensitive table (ST)

Anatomy [10] uses the l-diversity [5] model of privacy. It releases the values of QIs and SAs into two separate tables. The inherent problem in Generalization is that it prevents an analyst from correctly understanding the data distribution inside each QI-group. Anatomy removes this problem by capturing the exact QI-distribution. In this the correlation between the different attribute remain preserve. It uses grouping mechanism.

Steps

1. Partition tuples of microdata into several QI-groups.
2. Create QI Table.
3. Create ST (SA table) which contains SA statistics for each QI group.

Slicing

A novel data anonymization technique called slicing [1] introduced partitions the data set both vertically and horizontally. Vertical partitioning is done by grouping attributes into columns based on the correlations among the attributes. Each column contains a subset of attributes that are highly correlated. Horizontal partitioning is done by grouping tuples into buckets. Finally, in each bucket, values in each column are randomly permutated to break the linking between different columns. Slicing preserves membership disclosure and it is suitable for high dimensional data. The basic idea behind slicing is to break association cross columns, but to preserve the association within each columns. It uses k-anonymity [4], [8], [9] and l-diversity [5] model of privacy.

Steps

1. Attribute Partitioning.
2. Column generalization.
3. Tuple partitioning.

Limitation

• It cannot provide better data utility for an analyst.

Table 5: The Sliced Table

(Age, Sex)	(Zipcode, Disease)
(22,M) (22,F) (33,F) (52,F)	(47905,flu) (47906,dyspepsia) (47905,bronchitis) (47906,flu)
(54,M) (60,M) (60,M) (64,F)	(47304,gastritis) (47302,flu) (47302,dyspepsia) (47304,dyspepsia)

Overlapping Slicing

Overlapping slicing [11] is an enhance version slicing [1]. It partitions attributes both horizontally and vertically like slicing. In horizontal partitioning, tuples are grouped together and in vertical partitioning, highly correlated attributes are grouped together. Sensitive attribute in the relational table should be placed in the each column of a table.

Steps

1. Attribute Partitioning.
2. Column generalization.
3. Tuple partitioning.

Table 6: The Overlapping Slicing Table

(Age, Sex, Disease)	(Zipcode, Disease)
(22,M,flu) (22,F,dyspepsia) (33,F,bronchitis) (52,F,flu)	(47905,flu) (47906,dyspepsia) (47905,bronchitis) (47906,flu)
(54,M,gastritis) (60,M,flu) (60,M,dyspepsia) (64,F,dyspepsia)	(47304,gastritis) (47302,flu) (47302,dyspepsia) (47304,dyspepsia)

VI. CONCLUSION

In this paper, classification of data based on sensitivity is done. Then,wehaveexamined the different privacy-preserving techniques one by one, discussing whether existing techniques are enough to process the big data. All the techniques were briefly argued with an example.K-anonymity protects against identity disclosure, it does not provide sufficient protection against attribute disclosure. The notion of l-diversity attempts to solve homogeneity attack and background knowledge attack but it cannot resolve the problem of attribute disclosure so we have proposed a novel privacy notion called tcloseness. We use Earth Mover’s Distance for tcloseness but it is certainly not perfect. Recent technology which is used to protect privacy is Differential privacy. It provides strong privacy; even adversary has arbitrary external knowledge. Moreover, we discussed pros and cons of each technique. Traditional as well as recent both techniques were reviewed in this paper. In future, we can compare these anonymization techniques with differential privacy using various evaluation criteria. Also, we can use multiple sensitive attributes.

REFERENCES

[1]. Thakkar, Amit, AashiyanaArifbhai Bhatti, and Jalpesh Vasa. "Correlation-based anonymization using generalization and suppression for disclosure problems." Advances in Intelligent Informatics. Springer, Cham, 2015. 581-592.

[2]. Mehta, Brijesh B., and UdaiPratap Rao. "Privacy-preserving unstructured big data analytics: Issues and challenges." Procedia Computer Science 78 (2016): 120-124.

[3]. K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, “Incognito: Efficient Full-domain k-Anony Anonymity,” in Proc.of ACM SIGMOD, 2005.

[4]. Li, Ninghui, Tiancheng Li, and Suresh Venkatasubramanian. "t-closeness: Privacy beyond k-anonymity and ldiversity." Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on. IEEE, 2007.

[5]. Zhu, Dan, Xiao-Bai Li, and Shining Wu. "Identity disclosure protection: A data reconstruction approach for privacypreserving data mining." Decision Support Systems 48.1 (2009): 133-140.

[6]. Soria-Comas, Jordi, and Josep Domingo-Ferrer. "Differential privacy via t-closeness in data publishing." Privacy, Security and Trust (PST), 2013 Eleventh Annual International Conference on. IEEE, 2013002E

[7]. Soria-Comas, Jordi, et al. "t-closeness through microaggregation: Strict privacy with enhanced utility preservation." IEEE Transactions on Knowledge and Data Engineering 27.11 (2015): 3098-3110.

[8]. Shah, Rajesh, and Durgesh Thakur. "Closeness Privacy Measures Using Tree EMD for Data Disclosures."

[9]. Domingo-Ferrer, Josep, David Sánchez, and Jordi SoriaComas. "Database anonymization: privacy models, data utility, and microaggregation-based inter-modelconnections." Synthesis Lectures on Information Security, Privacy, & Trust8.1 (2016): 1-136.

[10]. Lee, Jaewoo, and Chris Clifton. "How much is enough? choosing ε for differential privacy." International Conference on Information Security. Springer, Berlin, Heidelberg, 2011

Authors:

Mr. BodduMani kumar, working as an Assistant Professor in the CSE Department of 'Viswanadha Institute of Technology and Management' who has studied his M.Tech(CSE) from 'Avanthi Institute of Engineering & Technology', Vizianagaram, A.P and B.Tech from 'Visakha Institute of Engineering & Technology', Visakhapatnam. He has three years of teaching and industrial experience and also actively participated in various workshops, seminars and presented papers related to information technology (IT). His area of interests are cloud computing, Networking and Network security. Aim of his life is to receive Doctarate(PhD), Research on advanced topics and serve for his mother country.

Ms. Kintali Padmaja, working as an Assistant Professor in the CSE Department of 'Viswanadha Institute of Technology and Management' who has completed her M.Tech from 'Sri Venkateswara College Of Engineering & Technology', Andhra Pradesh, India and B.Tech from 'GayatriVidhyaParishad College Of Engineering For Women'. She has one year of teaching experience and also participated in various workshops, seminarsand presented papers on data-mining and data-warehousing. Her areas of interests are data-mining and advanced computer technologies.

Boddu Mani kumar. “Privacy Preserving Sensitivity Data Classification Approach for Secured Mining.” IOSR Journal of Engineering (IOSRJEN), vol. 09, no. 04, 2019, pp. 41-49.