Quasi Attribute Utility Enhancement (QAUE)- A Hybrid Method for PPDP

A. N. Ramya Shree, P. Kiran

Abstract: ***The data analytics has become prominent for today's world because it is defined as the methodology of investigating data sets in order to draw conclusion about the information it contain. The Data Mining is a key part of Data Analytics because it has techniques and tools which help to explore knowledge which is hidden in data. The outcome of data analytics is very crucial to Business organizations because it helps in decision making process. In Data Analytics there are two roles which are very prominent and they are Data publisher and Data Analyzer. Data Publisher is the one who provides data for analytics which is collected from heterogeneous sources. Data Analyzer receives data from Data publisher and uses for data analytics. The main issue involves here is data privacy, which is concerned with the proper treatment of data i.e. approval, discern and regulations. A separate field called PPDP- Privacy Preserving Data Publishing mainly concentrates on how data is shared, used by data analysts and it may be implicit or explicit to organizations (third party) such that it should be safer from untrusted people and attacks. The PPDP offers several approaches to publish data in safe manner and supports data utility, but there is a need of domain specific privacy concern because privacy needs are different based on the domain and in mean time how data is utilized. In the paper a hybrid approach is proposed to preserve data privacy in concern with data publisher which supports domain specific data privacy and utility.***

Keywords: PPDP, PPDM, DW, CH.

I. INTRODUCTION

The Data Privacy became one of the key challenges in Data Analytics. The Privacy Preserving Data Mining approaches concern about preserving privacy of sensitive data belongs to individual /organization and in turn extracting knowledge or patterns from data which is stored in Data Warehouse. The DW isinformationstorage and retrieval system which contains legacyand transactional data which are received fromseveral heterogeneous sources[1].The most important task of PPDM is to convert data into appropriateform for DW in such a way that the sensitivity of data is preserved and mining can be performed. PPDP is a research direction in PPDM which supports publication of useful information while protecting data privacy and focuses on preserving privacy of sensitive data prior to Data Mining [2].The Data Privacy issues and Data Privacy violation attacks which results in loss of privacy are described in the following sections.

Revised Manuscript Received on December 15, 2019.
***** Correspondence Author
A. N. Ramya Shree *, Asst. Professor, CSE Department, RNS Institute of Technology, Bangalore. Email: ramya.rammu@gmail.com
P. Kiran, Assoc. Professor, CSE Department, RNS Institute of Technology, Bangalore. Email: kiranpmys@gmail.com

A. Data Privacy Issues

- The data is stored using either centralized and/or distributed architecture. Related to today's technology data collected from heterogeneous sources and stored in distributed environment. So the data needs to be processed in safe manner.
- The data on social sites are subjected to continuous updates. It is necessary to confine the changing data and preserve data privacy related to data publishing, mining and storage.
- Data retrieval involves complex queries execution and parallel processing. Due to the large size data, other than of transferring the data between the multiple nodes in the network, it is feasible to move the necessary control codes to data storage nodes. So it give raise the need of preserving data privacy which is highly essential.
- Data collected from the heterogeneous sources like various logs, social media, etc. Due to the data heterogeneity nature, it is necessary to identify how, what, when, where and \who has the right access to data and how data is analyzed in data mining process. It becomes very difficult to maintain the privacy of sensitive data i.e. the data which disclosed leads to hazardous effects [3].
- Usage of domain specific prior knowledge to estimate the actual data and able to mine sensitive information.

B. Data Privacy violation attacks

- Identification attack: When we search for specific data in the database, in same time other associated private data gets exposed.
- Malicious attack: Attackers can adopt any techniques to obtain sensitive information of the other parties like do not honor privacy agreements, send wrong information, involve in partnerships with other attackers.
- Semi honest attack: Attackers use computation protocols using that they want to gain others private information.
- Publication attacks: Unstructured data i.e. data which does not have customized structure like emails, log reports, images, audio and video without specific privacy policies and approaches may reveal precious data.
- Domain specific attack: If the adversary attacks with prior knowledge about domain called domain specific

attack. So that attacker can fetch specific information it causes hazardous effects to organization [4].

II. EXISTING PPDP APPROACHES

The various approaches are existing to publish sensitive data in privacy preserving data publishing. The major approaches are described in the following section.

Randomization Approach

It is the process of adding noise to the original data in order to hide attributes from exposure. There are two major types of randomization. 1) Additive randomization - A noise (value) is added to data record values which is taken from probability distribution function .The accuracy depends on how large the distribution function and the amount of randomization. 2) Multiplicative randomization-A noise (value) is multiplied to data record values which is taken from probability distribution function .The accuracy depends on how large the distribution function and the amount of randomization. The limitation of randomization approach is massive information loss and the result obtained is approximate [5].

A. Data Swapping Approach

In Data Swapping approach the values of records are swapped with other records within the database in such a way that statistical inference from the swapped data is equal to the statistical inference of the actual data. The advantage of this approach is the records which are swapped to the same domains leads to enhancement of data accuracy and has less information loss. The limitation of this approach is that the privacy of the data not achieved complete manner because actual availability of sensitive data and it can be subjected to domain specific attack [6].

B. Cryptography

The current Internet is mainly responsible for generation of large amounts of heterogeneous data. The different approaches are deployed to preserve privacy of data at the time of data transfer, storage and retrieval. Cryptography approach is used to preserve data privacy where data converted into secret form, transferred and received in multiple sites. The main problem is data distributed in multiple sites and available in different format.Data mining has to be performed on heterogeneous data which resides on multiple sites. The main drawback of this approach is that it can be used for safe communication but does not support data utility which is most important aspect of Data Analytics [7].

D. Anonymization

To preserve privacy modify the sensitive contents of the record owner data so it ensures privacy prior to publication. This approach is called Anonymization. Each tuple consists of four types of attributes they are: 1) Identifier – These attributes can directly and uniquely identify an individual. 2) Quasi attributes (QA) - These attributes can be linked with generally available data like census data to re-identify individual like gender, age and zip code.3) Sensitive Attribute
- Attributes that an individual wants to hide like disease, salary.4) Non-sensitive attributes –Which are other than ID, QA and Sensitive Attributes. Quasi attributes are mainly used by attackers to link value to any publically available datato identify and extract the individual details. The sensitive attributes which has to be protected from attackers .It can be published only if it is significant for data mining. To preserve the privacy in published data, data publisher will modify tuple in such a waythat user identification is removed and quasi attributes are anonymized such that it ensures data privacy which is specific to particular privacy preserving model.[8][9].The major Anonymization techniques are as follows:

- k-anonymity - In this technique user specific field data, produce a release of the data with guarantees that the individuals cannot be re-identified while the data remain practically useful. If release of data is said to cover the k_anonymity property if the information for each individual contained in the release cannot be distinguished from at least k-1 individuals whose information also appear in the release. Weaknesses in the k-anonymity model are when the sensitive values within a group exhibit homogeneity [10].
- l-diversity- It based on group anonymization that is used to preserve privacy in data sets by decreasing the granularity of a data representation.. The l-diversity model is an extension of the k-anonymity model uses techniques like generalization and suppression. The l-diversity model handles weaknesses in the k-anonymity model, l-diversity requirement ensures “diversity” of sensitive values in each group, but not identifies values which are semantically close[11]
- t-closeness- The t-closeness model extends the l-diversity model by treating the values of an attribute particularly by taking into account the distribution of data values for that attribute. t-closeness specifies if the distance between the distribution of a sensitive attribute in class and the distribution of the attribute in the whole table is not more than a threshold t. A table is said to have t-closeness if all equivalence classes have t-closeness. It is more effective than other privacy preserving mining methods [12].

E. Limitations

The major limitation of existing approaches is data utility. All approaches concentrate mainly on data privacy and there is a need to support utility because data subjected to analytics. The data utility means usefulness of data. When data subjected to mining operations it should yield useful patterns or knowledge to support organization decision making. Data utility differs from one domain to another domain like health care, education, defense etc. So which are sensitive attributes or not related to domainhas to be identified.Another limitation is the types of attributes which are very important to support data utility are not considered when applying PPDP approaches.

III. RESULTS AND DISCUSSIONS

To overcome the limitations of existing PPDP methods a hybrid method is proposed which concentrates on datautility and mean time to preserve data privacy. Thedomain knowledge and utility enhancement methods are described in next sections.

A. Domain knowledge

Itmeans knowledge of an exact specialized field, in contrast to universal knowledge. Person who has domain knowledge, are treated as expert in the field. The domain concept includes the data concerned to the business and its rules.Every domain consists of objects and its interactions with other objects. The objects are real world entities each of which has its own attributes. The container or class contains general characteristics of entities and object is an instance of class. Each object has distinct or common value for each attribute. In health care domain consider Patient class for example,

it contains general attributes like Id, Ssn, Name, Age, Height, Sex, Phone number, Disease, Zip code etc.Each of the attributes mentioned are associated with values some of which are sensitive and others are non sensitive. The term sensitivity is very specific to each individual and its context varies from one individual to other. The sensitive attribute is disease and quasi attributes are age, sex, zip code which are used indirectly by attackers to identify entity who possesses sensitive attributes. When data published to data analyzer without suitable privacy preserving mechanism leads to individual privacy breach.[13][14]

B. QAUE – A Hybrid Approach

To overcome the problem of privacy breach and to maintain data utility a hybrid method proposed to preserve privacy. It mainly dependent on types of data and domain knowledge. The data types mainly identified in data analysis are described in Table-I.Consider the sample dataset as shown in the table Table-II. The attributes related to patient are name, sex, age,nationality, zip, disease .The age is continuous valued and ordinal data. Sex, Nationality and disease are categorical and discrete data. Zip code is ordinal data [15].

Table- I: Data Types

Level	Properties	Example	Types of data
Nominal	Classification	Disease ,Sex	Qualitative
Ordinal	Order classification	Disease severity	Ranked
Ratio or Interval	Equal intervals / Order classification	Patient age, height	Quantitative

Table-II: Sample patient Data

Name	*Sex*	*Age*	*Nationality*	*Zip*	*Disease*
Arnold	M	25	Indian	641054	Gastritis
Saron	M	21	Indian	641064	Viral Inf.
Bob	M	29	South American	641162	Heart Disease
John	M	39	German	641064	Cancer
Jerry	F	32	Switzerland	641162	Cancer
Alice	M	37	Japan	641162	Cancer
Mary	F	52	Indian	671890	Viral Inf.
Joice	F	50	Indian	671892	Cancer
Lee	F	59	Chinese	671882	Aids

The QAUE hybrid approach is mainly considers additional attribute as target labels which specifies which row contains sensitive or non-sensitive information. Every row in the table labeled depends on domain specific sensitive attributes. The target label is associated with quasi identifiers used for utility enhancement. The Table-III shows target labels set for sensitive attributes. The next sections show for each type of quasi attributes how target labels are used to enhance the utility [16].

Table-III: Sample patient Data with Target Label

Name	*Sex*	*Age*	*Nationality*	*Zip*	*Disease*	*Target label*
Arnold	M	25	Indian	641054	Gastritis	Not sensitive
Saron	M	21	Indian	641064	Viral Infection	Not sensitive
Bob	M	29	South American	641162	Heart Disease	Not sensitive
John	M	39	German	641064	Cancer	sensitive
Jerry	F	32	Switzerland	641162	Cancer	sensitive
Alice	M	37	Japan	641162	Cancer	sensitive
Mary	F	52	Indian	671890	Viral Infection	Not sensitive
Joice	F	50	Indian	671892	Cancer	sensitive
Lee	F	59	Chinese	671882	Aids	sensitive

C. Categorical Attribute Encoding

The attribute sex contains discrete values. Normally these values are subjected to generalization and represented as person instead of male or female which leads to loss of data utility. The data transformation technique called one hot encoding is used to represent values as 1 for male and 0 for female. Based on likelihood occurrence frequency the portion of male as 5/9 as 0.55 and female as 1-0.55=0.45[17].

Table-IV: One Hot Encoded sex attribute with target label

M	**1**	Not sensitive
M	**1**	Not sensitive
M	**1**	Not sensitive
M	**1**	Sensitive
F	**0**	Sensitive
M	**1**	Sensitive
F	**0**	Not sensitive
F	**0**	Sensitive
F	**0**	Sensitive

Table-V: Target label based partitioning of sex attribute

M	**1**	Not sensitive
M	**1**	Not sensitive
M	**1**	Not sensitive
M	**1**	sensitive
M	**1**	sensitive
F	**0**	Not sensitive
F	**0**	Sensitive
F	**0**	Sensitive
F	**0**	Sensitive

The target label based partitioning described in Table-IV.The data modified using Categorical Attribute Encoding approach described in Table-V. In the table there are two rows of type sensitive related to male but it is and it is difficult to identify exactly which is sensitive row for value male or female. Because it considers likelihood occurrence which specify either 0.55 as male or female or 0.45 as male or female. So it avoids specific gender identification and helps to maintain privacy and enhance utility.

D. Continuous Attribute Modification using Binning

Group the attribute age w.r.t to target label as shown in Table-VI. Create bins for age values in the table. To create bins first sort the values. Second determine number of bins. The bin range is calculated using formula *max-min/number of bins* i.e. sorted values 32, 36 ,39, 50, 58 .The number of bins equal to 3 then range is 59-32=27is divisible by 3 the range value is 9. i.e. bin1: 30-39, bin2: 50-59. The bin1 average is 36 and bin2 average is 55.replace values as per bin average. The modified data described in Table-VII. It allows us to retain approximate sensitive value and original value, so it supports privacypreservation and utility enhancement [18].

Table-VI: Age categorical attribute grouped w.r.t target

39	sensitive
32	sensitive
37	sensitive
50	sensitive
59	sensitive
25	Not sensitive
21	Not sensitive
29	Not sensitive
52	Not sensitive

Table-VII: Age categorical attribute modified using Binning

36	sensitive
36	Sensitive
36	Sensitive
55	Sensitive
55	Sensitive

E. Categorical attribute generalization using CH

A concept hierarchy that is a total or partialorder of attributes in a relational schema. It can be used to vary the attribute utility when publishing the values. Consider the attribute Nationality, according to concept hierarchy specification, it is a tree structure where root nodes contain generalized values and leaf node contains specific values. The Fig: 1 shows Concept Hierarchy representation for attribute nationality, the root node represents continent and leaf node represents specific country. The attribute nationality is a quasi-identifier, to overcome the disclosure problem form two partitions based on target label as specified in Table-VIII. Perform partial generalization only on sensitive rows as per concept hierarchy specification described in Table-IX.It results in retaining nationality attribute value as per original table and only modification of sensitive rows. It results in data utility enhancement mean time it preserves privacy by partial generalization [19].

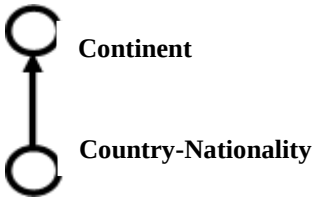

Fig: 1 Concept Hierarchy for attribute nationality

Table- VIII: Attributes partitioned according to target label

Nationality	Target label
Indian	Not sensitive
Indian	Not sensitive
South American	Not sensitive
Indian	Not sensitive
German	sensitive
Switzerland	sensitive
Japan	sensitive
Indian	sensitive
Chinese	sensitive

Table- IX: Modified attributes based on CH

Nationality	Target label
Asia	Not sensitive
Asia	Not sensitive
America	Not sensitive
Asia	Not sensitive
Europe	sensitive
Europe	sensitive
Asia	sensitive
Asia	sensitive
Asia	sensitive

F. Ordinal attribute Zip code generalization using Lattice.

A Zip code used by the US Post Service. The original format consists of five digits. It can be appended with four extra digits that designate a more specific location. The zip code leads to specific location identification so it can be represented complete or partial representation by special symbol * to overcome location disclosureby attackers [20]. The Categorization of Zip code attributes based on target label shown in Table-X.

Table- X: Categorization of Zip code attributes

641054	Not sensitive
641064	Not sensitive
641162	Not sensitive
641064	sensitive
641162	sensitive
641162	sensitive

The attributes are partially generalized based on the region code. The lattice structure used to represent generalization and specialization of values.

The Fig: 2 shows lattice as per target label. The different types of generalizations are complete, partial, semi partial and no generalization. It specified for the zip attribute based on target label. It helps in preserving the privacy and support utility. The Table- XI describes modified Zip code values as per lattice based Ordinal attribute Zip code generalization [21].

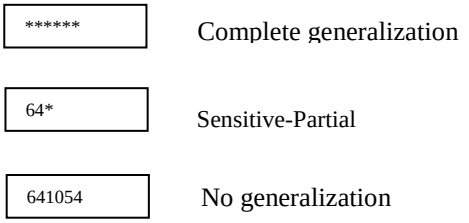

Fig: 2 the lattice based generalization

Table- XI: Modified Zip code values

641054	Not sensitive
641064	Not sensitive
641162	Not sensitive
64*	Sensitive
64*	Sensitive
64*	Sensitive

G. QAUE Hybrid Algorithm

The QUAE Hybrid algorithm used to generate micro data Which contains entity attributes in each row such that data utility enhanced and mean time privacy is ensured.

Algorithm: *QA Utility Enhancement hybrid Algorithm*

//Input: Original Relational Table
//Output: Micro data using QAUE
1. Explore identity, sensitive, quasi, non sensitive attributes in the original relation table based on domain knowledge.
2. For each row assign Target label as sensitive or not sensitive based on quasi identifiers.
3. for each QA
Begin:

- *If QA is binary valued attribute then apply categorical attribute encoding approach.*
- *if QA is continuous then applyContinuous Attribute Modification using Binning*
- *if QA is categorical then apply Categorical Attribute Generalization using CH*
- *if QA is ordinal then apply lattice based generalization*

End:
4. Generate Micro data.

G. Micro data publishing using QAUE Hybrid Method.

The micro data contains field's sex, age, continent, zip and disease. The sex attribute got frequent 1 as 5 and 0 as 4 but difficult to determine exactly male or female .The attacker able to identify 1 as male ten there are 5 male each with same or different age, similarly for female also. Continent details are known not exactly specific location. For data miner disease origin, approximate person ages are going be informative. The proposed approach assumes every tuple categorized into sensitive or non sensitive and clear specification of sensitive, quasi and non sensitive attributes. The micro data published using QUAE approach is shown in Table-XII.

Table- XII: QAUE based Micro data publication

Sex	Age	Continent	Zip	Disease
1	**25**	Asia	641054	Gastritis
1	**21**	Asia	641064	Viral Infection
1	**29**	America	641162	Heart Disease
1	**36**	Europe	64*	Cancer
0	**36**	Europe	64*	Cancer
1	**36**	Asia	64*	Cancer
0	52	Asia	671890	Viral Infection
0	**55**	Asia	67*	Cancer
0	**55**	Asia	67*	Aids

H. Evaluation of QAUE Hybrid Approach

The different categories of information metrics are available for evaluating the utility of data which is outcome of data mining operations. It mainly related to the Data quality of the published table with respect to the data quality of the original table.To measure the utility of attributes of individuals in the published table data, a measure called Weighted Normalized Certainty Penalty is used.Based on data usage domain, weight is assigned to each attribute to reflect its usefulness in the analysis on the modified data. The weighted normalized certainty penalty should be minimal because it reflects low data utility. The proposed approach uses binning based Continuous Attribute Modification approach to minimize the Weighted Normalized Certainty Penalty [22].

Conceptual Hierarchy (CH) is used for the generalization of categorical attributes. Attribute values of different the scale are specified by the CH which is a tree structure where theleaf nodes which are children of predecessor used to measure the generalization in categorical attribute. Discernibility metric measures the number of tuples that are not different from each other.Each row in an equivalence class E incurs a cost |E|. It is equal to the sum of the squares of the sizes of the equivalence classes. The proposed approach helps to reduce the cost by retaining original rows which are not sensitive. The micro data contains field's sex, age, continent, zip and disease. The sex attribute got frequent 1 as 5 and 0 as 4 but difficult to determine exactly male or female. The attacker able to identify 1 as male then there are 5 male each with same or different age, similarly for female also. Continent details are known not exactly specific location.

For data miner disease origin, approximate person ages are going be informative. The UCI Machine Learning Repository Disease Dataset is used for evaluation. The proposed approach assumes every tuple categorized into sensitive or non sensitive and clear specification of sensitive, quasi and non sensitive attributes.

IV. CONCLUSION

The data is very crucial and has to be protected from attackers. Once data published it is difficult to analyze good or bad utilization. Prior to data publishing data publisher has to analyze the attack scenario and carefully deploy privacy preserving techniques. The proposed approach requires every tuple has to be categorized into sensitive or non sensitive and clear specification of sensitive, quasi and non sensitive attributes. It requires domain knowledge to distinguish between attributes. The limitation is proper binning of categorical attributes, usage the of concept hierarchies for nominal attributes and lattice structure usage for special categorical attribute like zip code.

REFERENCES

1. R. J. Han, *Data Mining Techniques*", Morgan Kaufmann publication, 2001.
2. Agrawal and R. Srikanth,"*Privacy-preserving data mining*",in Proceedings of the 2000 ACM SIGMOD conference on Management of Data, Dallas, TX, May 14-19 2000.
3. L. Brankovic and V. Estivill-Castro, "*Privacy issues in knowledge discovery and data mining*", in Proc. Austral. Inst. Comput. Ethics Conf., 1999, pp. 89_99.
4. Kiran P,"*A survey on methods , attacks and metric for privacy preserving data publishing*". Int. J. Comput. Appl. 53(18), 20–28 (2013).
5. B.C. Chen, D. Kifer, K. LeFevre, and A. Machanavajjhala,"*Privacy preserving data publishing. Foundations and Trends in Databases*", 2(1-2):1–167, 2009.
6. ToreDalenius ,Steven P.Reiss "*Data-swapping: A technique for disclosure control*"Journal of Statistical Planning and Inference Volume 6, Issue 1, 1982, Pages 73-85
7. Y. Lindell and B. Pinkas, "*Privacy preserving data mining*" *Advancesin Cryptology*. Berlin, Germany: Springer-Verlag, 2000, pp. 36_54.
8. C. C. Aggarwal and S. Y. Philip, "*A General Survey of Privacy Preserving Data Mining Models and Algorithms"*. New York, NY, USA:Springer-Verlag, 2008
9. B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "*Privacy-preserving data publishing: A survey of recent developments*",ACMComput. Surv.,vol. 42, no. 4, Jun. 2010.
10. Sweeney, ``*k-anonymity: A model for protecting privacy*," Int. J. on Uncertainty, Fuzziness Knowledge.-Based Syst., vol. 10, no. 5, pp. 557_570,2002.
11. A Machanavajjhala, L-diversity: "*privacy beyond k-anonymity",* IEEE,0-7695-2570-9, April 2006
12. Ninghui Li , Tiancheng Li , Suresh Subramanian "*t-Closeness: Privacy Beyond k-Anonymity and l-Diversity*".IEEE,.1109/ICDE.2007.367856, 2007
13. C. N. Modi, U. P. Rao, and D. R. Patel, ``*Maintaining privacy and data quality in privacy preserving association rule mining*," in Proc. Int. Conf.Comput. Commun. Networks. Technol. (ICCCNT), Jul. 2010, pp. 1_6
14. Wu Xiao-dan , Yue Dian-min ,Liu Feng-li ,Wang Yun-feng ,Chu Chao-Hsien ,"*Privacy Preserving Data Mining Algorithms by Data Distortion*" ,International Conference on Management Science and Engineering, 04 September 2007,IEEE Access,Volume5
15. Abou-el-elaAbdouHussien, Nermin Hamza, HeshamHefny "*Attacks on Anonymization-Based Privacy-Preserving:A Survey for Data Mining and Data Publishing"* Journal of Information Security , 2013, 4, 101-112
16. K. LeFevre, D. J. DeWitt and R. Ramakrishna, "*Work-load-Aware Anonymization,"* Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, August2006,pp.277-286.
17. Michael Steinbach, Pang-Ning Tan,Vipin Kumar, "*Introduction to Data mining"*, Pearson, 2006.
18. Xin-She Yang ,"*Introduction to Algorithms for Data Mining*", Academic press,June2019
19. M. Abirami, K. Thirukumar,"*Improving Privacy and Utility in Micro-aggregation Generalization Method using l-diversity"*, International Journal of Advances in Computer and Electronics Engineering ,Issue: 1, May 2016, pp. 28 -34
20. S.Matwin," *Privacy preserving data mining techniques: Survey and challenges in Discrimination and Privacy in the Information Society*". Berlin, Germany: Springer-Verlag, 2013, pp. 209_221.
21. Aashiyana Bhatti, Amit Thakkar, Jalpesh Vasa, "*Correlation Based Anonymization Using Generalization and Suppression for Disclosure Problems*", Advances in Intelligent Systems and Computing, 2015
22. S. Martinez, "*Improving the Utility of Differentially Private Data Releases via k-Anonymity*", Trust, Security and Privacy in Computing and Communications, IEEE, pp. 372-379, 2013

AUTHORS PROFILE

Ms. Ramya Shree A N, B.E, M.Tech, working as Asst.Professor in CSE Department, RNSIT, Bangalore from past 12 years. The research work mainly focuses on Data Mining and Big Data Privacy.

Dr. Kiran P, BE, M.Tech, Ph.D, working as Assoc. Professor in CSE Department, RNSIT, Bengaluru from past 18 years.His research interests include Cryptography, Big Data Analytics, Data Mining, Privacy Preserving Data Mining and Privacy Preserving Data Publishing.