International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 5, Issue 01, January-2019

PREVENTION OF LEAKAGE SENSITIVE AND PRIVILEGED INFORMATION ANALYSIS

[1]Abhijeet kumar, [2]Mrs.Geeta Tiwari, [3]Mr.kishore Mishra

[1]*M.Tech Scholar CSE, Compucom Inst. Of Tech. & Management,jaipur*
[2]*Asst. Prof CSE Compucom Inst. Of Tech. & Management,jaipur*
[3]*Asst. Prof CSE Compucom Inst. Of Tech. & Management,jaipur*

Abstract— privacy preserving is the most concerning issue. Information related to specific individual needs to be protected, so that it may not harm the privacy. Anonymization method is applied on the dataset, Pseudonymization is a method to substitute identifiable data with a reversible, consistent value. Anonymization is the destruction of the identifiable data. K-anonymity has been used as successful technique it express this methodology achieves enhanced over the distinct l-diversity measure, probabilistic l-diversity measure and k-anonymity through t-closeness measure since only rarer partitioning must be done for a robust secrecy requirement. Then by using of fuzzy c-means clustering data are clustered.

Keywords—Data mining, Anonymization , Fuzzy C-Means Clustering, Privacy preserving.

I. INTRODUCTION

Data mining is, *the removal of hidden predictive information from extensive database*, is a powerful new innovation with incredible potential to enable organizations centered around the most vital info in their data warehouses [9]. This is to protect the identity of the individual this encrypt identifiers like unique number and the name. Whereas the data which is not encrypted provides less or no guarantee. For this privacy purpose k-anonymity proposed, here in collection of data sets it is hard to recognize the identity of the client as of the data sets of information which is sensitive. While using this approach detecting the exact information risk is minimized [10]. Privacy preserving data mining (PPDM) is partitioned towards differs classification. In the survey of PPDM & diverse examination execute in the region of PPDM beneath different classification We will center around estimations that are used to measure the responses happened in light of privacy preserving technique [11]. Fuzzy clustering is an awesome unsupervised system for the examination of information and advancement of models. As a rule, fluffy bunching is more normal than hard grouping.

II. FUZZY C-MEANS CLUSTERING

The FCM operates fuzzy divisioning to such a degree, to the point that an data point can have a place with whole gatherings with various membership range between 0 & 1.This algo works by consigning membership to all data point involving towards each group center based on distance between the cluster center & the data point. Increasingly, the data is close to the group focus more is its membership towards the particular cluster center. Clearly, summation of participation of every data point ought to be equal to one. After, every cycle iteration membership & cluster centers are updated based on formula [12].

III. ANONYMIZATION METHOD

This is to protect the identity of the individual this encrypt identifiers like unique number and the name. Whereas the data which is not encrypted provides less or no guarantee. For this privacy purpose k-anonymity proposed, here in collection of data sets it is difficult to identify the identity of the user from the data sets of information which is sensitive. While using this approach detecting the exact information risk is minimized. [5]
For instance, various common data characteristics such as race, birth, sex, and zip are accessible in public records such as voter list and a particular data set for instance medical data, they can be used to accomplish the identity of the equivalent individual with high probability by linking process, as is shown in the figure 1.

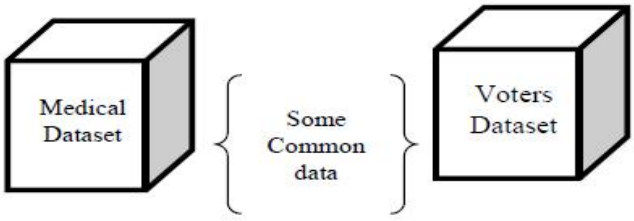

Figure 1: Anonymization Model

The following are common types of data anonymization.

Removal

Completely removing sensitive fields that could be used in any way to identify a person.

Redaction

includes removal and other techniques such as blacking data out on paper.

Encryption

can be very strong and difficult to reverse. It does present a few challenges such as the generation of a sufficiently strong decryption key.

IV. LITERATURE SURVEY

M. Prakash and G. Singaravel [2015],utilized Top-Down Greedy algorithm in which it partitions the high-dimensional space into areas and by the area's representation itencodes data focuses in one area.

Here four privacy measures are studied-

- Distinct l-diversity measure with default value of l = 5.
- Probabilistic l-diversity measure with default estimation of l = 5.
- k-anonymity with t-closeness measure with default estimation of k = 5 and t = 0.15.
- Proposed personalized anonymization approach with default estimation of k = 5, n = 1000 and t = 0.15.[2]

Vijay Sharma[2014]provides a surety that released data is in any case k-anonymous. Numerous methods have been suggested to realize k-anonymity for the specified dataset. It categories these methods into four main fields based on the standard these are based and methods they are applying to accomplish k-anonymous data.Four main approaches to the solution of k-anonymity problem for anonymization of data have been discussed. These methodologies have been diplayed in a basic and easy to understand language, with examples. All these categories include most of the algorithms developed so far for accomplishing k-anonymity [6]

Fei Liu et al [2013]recommend a new k-anonymity algorithm for sensitive characteristics. He divide sensitive attributes values into highly sensitive ones and lowly sensitive ones. Tuples are sorted according to amount of highly sensitive values first and then distributed to best equivalence classes one by one. We destroy association among sensitive attributes values to avoid attack. He introduce information entropy to evaluate diversity of equivalence classes.[7]

Mohammad Reza Zare Mirakabad et .al [2008] recommend procedures to discover several questions about k-anonymity of data. Such questions are, for example, ―is my informationadequatelyunknown?‖, ―which info, if accessible from an outside source, threatens the anonymity of my data?‖.The methodology that they propose affect two properties of k-anonymity that they express as two lemmas. The principal lemma is a monotonicity property that empowers to adapt the A-priori algorithm for k-anonymity. The following lemma is a determinism property that empowers to devise an effective algorithm for δ-suppression.[8]

Xiangwen Liu et .al [2015] propose a personalized extended (alpha, k)-anonymitymodel for the purpose of personalized privacy preservationrequirements in Privacy Preservation Data Publishing technology.Our model combines sensitive attribute value-orientedprivacy preservation method with individual-oriented method,and unifies the privacy protection requirement of above twomethods with Privacy Preservation Level.The experimental results show that our model can providestronger privacy protection with not much time cost. d. Experimental results show that the personalized extended (α, k)-anonymity model can provide stronger privacy protection efficiently.[9]

JordiSoria-Comas et .al[2013] approach combines *k*-anonymity and *ε*-differentialprivacy to reap the best of each approach for anonymizeddata publishing: namely, the reasonably low informationloss incurred by *k*-anonymity and its lack of assumptionson data uses, and the robust confidentialitypromises offeredby *ε*-differential privacy. Theyusage a new defined insensitivemicroaggregation to achieve a *k*-anonymous data setby considering all features as quasi-identifiers; then take the *k*-anonymous microaggregated data set as aninput to which uncertainty is added in order to reach*ε*-differential privacy which show that our combined approach reduces information loss by several orders magnitude[10]

Shyue-Liang Wang et .al[2011]suggested two procedures, *Sensitive Transaction Neighbors*(*STN*) and *Gray Sort Clustering* (*GSC*), by addition/deletion of*QI*items and adding *SI* items to realize*sensitive k-anonymity* ontransactional data. Extensive numerical researches wereassumed to determine the features of the suggestedconcept and approaches.It is

different from *k-anonymity* on transaction in that transactionsmay contain both sensitive items and quasi-identifying items.[11]

Tanashri Karle et .al [2017] main focus of the study is Privacy Preservation using Anonymization Technique and a detailed study two Anonymization Algorithms are explained – Datafly Algorithm and Mondrian Algorithm. Datafly algorithm is more suitable for synthetic dataset while Mondrian algorithm is more suitable for real dataset. Datafly Algorithm performs better when dataset is We have done a detailed study of these two algorithms and achieved a detailed comparison of these two algorithms based on thirteen parameters.[12]

V. MERITS AND DEMERITS OF TECHNIQUES

TECHNIQUES	MERITS	DEMERITS
Anonymization	This technique is utilized to ensure respondents characters while discharging fruitful info.	There are two assaults: the homogeneity assault & the background knowledge assault. Since the limitation of the k-anonymity model show originate from the two suspicions. K anonymity model demonstrate a certain technique of assaults, while in genuine situations there is no motivation behind why the attacker should not attempt other techniques.
t-closeness	Measure the separation between two probabilistic conveyance that were undefinedfrom each other	Information pickup was heavy
ℓ-Diversity	Delicate attribute would have at most same frequency	Homogeneity & background Knowledge assault has needed
Distributed K -Anonymity framework (DKA)	Worldwide Anonymization to guarantee privacy	Utilization & potential were Misused
Perturbation	Independent treatment of the diverse attributes via irritation approach	The technique does not remake the original data values, but only distributions to carry out mining of the data available.
Randomized response	The randomized strategy is a easy method which can easily executed at data accumulation time.	Randomized response method isn't for various attribute databases.
Slicing	Randomization on sensitive attribute	Utility and risk measures not Matched

VI. CONCLUSION

Every privacy preserving strategy has its own particular significance. Anonymizing huge data and dealing with anonymized data sets are nonetheless challenges for classic anonymization processes.common types of data anonymization like removal,redaction and encryption but there are challenges for large amount of data in that case we use clustering techniques ,here we use Fuzzy C -means clustering for cluster data set.

REFRENCES

[1] M. Prakash and G. Singaravel, ―An approach for prevention of privacy breach and information leakage in sensitive data mining‖,Computers and Electrical Engineering,2015.
[2] Wen He,"Research on LBS Privacy Protection Technology in Mobile Social Networks‖", 2017.
[3] B.B.Patil and A.J.Patankar, ―Multidimensional k-anonymity for Protecting Privacy using Nearest Neighborhood Strategy‖, IEEE International Conference on Computational Intelligence and Computing Research, 2013.
[4] Vijay Sharma, ―Methods for Privacy Protection Using K-Anonymity‖,International Conference on Reliability, Optimization and Information Technology,India, Feb 6-8 2014.
[5] Mohammad Reza Zare Mirakabad1 et.al, ―Towards a Privacy Diagnosis Centre: Measuring *k*-anonymity‖, International Symposium on Computer Science and its Applications, 2008.
[6] Jordi Soria-Comas et.al, ―Improving the Utility of Differentially Private Data Releases via k-Anonymity‖, 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications,2013.
[7] Shyue-Liang Wang et.al, ―K-anonymity on Sensitive Transaction Items‖, IEEE International Conference on Granular Computing, 2011.
[8] Tanashri Karle,Prof. DeepaliVora ―Privacy Preservation In BigData using Anonymization Techniques‖, InternationalConference on Data Management, Analytics and Innovation (ICDMAI), 2017.
[9] Mohammadian, M., "Intelligent Agents for Data Mining and Information Retrieval," Hershey, PA Idea Group Publishing, 2004.
[10] TamannaKachwala, SwetaParmar, "An Approach for Preserving Privacy in Data Mining, 2014, IJARCSSE All Rights Reserved .
[11] Dhivakar K , Mohana S "A Survey on Privacy Preservation Recent Approaches and Techniques" International Journal of Innovative
[12] Putri, A walia W., LaksmiwatiHira "Hybrid Transformation in Privacy-Preserving Data Mining" ©2016 IEEE