Turkish Journal of Computer and Mathematics Education Vol.12 No. 7 (2021), 1659-1668

Research Article

Enhanced Privacy With Disrupted Data By Choosing Features

V. S. Prakash [1], M Gunasekaran [2] Dr. Helen Josephine V.L. P. Murugesan [4]

[1]Assistant Professor, Department of Computer Science, Kristu Jayanti College, Bengaluru – 560077, Karnataka, India, vsprakash@kristujayanti.com
[2]Professor, Department of Computer Science and Engineering, Saveetha School of Engineering Saveetha Institute of Medical and Technical Sciences, Chennai – 602105, Tamilnadu, India ,gunasekaranm.sse@saveetha.com
[3] Associate Professor of MCA, CMR Institute of Technology, Bangalore, E-mail ID: helen.j@cmrit.ac.in
[4] Assistant Professor Level I, Department of Computer Science and Engineering,Bannari Amman Institute of Technology,Sathyamangalam-638 401, Tamilnadu, Indiam, murugesanp@bitsathy.ac.in
(* Corresponding author's e -mail: vsprakash@kristujayanti.com)

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 16 April 2021

Abstract : Privacy shows a key role in data mining applications. This has sparked the expansion of several data mining procedures to protect privacy. In order to facilitate the preservation of privacy in data mining algorithms over generally consist of together horizontal and vertical data Collection, a variety of procedures have been recommended using SMC and different protected structure blocks. Earlier work concentration on maintaining privacy by adjusting a separately adaptable disruption perfect that allows persons to select their individual degree of secrecy. The weakness, however, is that the computational findings for privacy are not adequately discovered. This paper suggested the collection of privacy features in a multi-partitioned dataset. Information can be wrapped for confidentiality by disruptive method as a alias name. In a multi-stakeholder data assessment, the arrangement of data and the collection of features for the data mining conclusion model that builds the fundamental model evidence popular this paper are established. The aim of the improvement ratio technique used in this work is to improve privacy in a multi-part data collection. Not completely features need the security of sensitive data for the top classical. Documents Representation for Privacy Protection Data Mining has occupied measures to improve techniques to create the best classical without breaking the confidentiality of characters. An investigational assessment is performed to evaluation the efficiency of the future Enhanced Privacy with Disrupted Data by Choosing Features [EPDDCS] in multi-partitioned datasets proved by various experimentations on together simulated and actual datasets.

Keywords: Privacy Preserving, Data Mining, Disturbed Data, Choosing Features

Introduction

Clustering is the method of defining collections within multi-dimensional databases, assisted by relationships, by negligible knowledge about their society. Conservative clustering algorithms are clustered around essential databases. However, existing implementations require datasets spread through a wide variety of sites. As a consequence, in discrete database settings, all discrete data is hard on an essential site prior to disturbing conventional algorithms.

There are two different states that insist on the need to perform cluster analysis in a dispersed manner. The initial state arises when the volume of data to be analyzed is comparatively high and commands significant computational effort and is often not feasible to accomplish this mission. The best option is to split the data, collect it in a distributed way and syndicate the results. The second state arises when data is evidently distributed among a variety of geologically isolated units and the costs associated with its centralization are very high. Clustering algorithm based on hybrid global optimization based on dynamic systems approach algorithm Maroosi & Amiri (2010) share information between different solutions. Therefore, the sharing of information is not secured by enabling the adversary's access to further security violation outbreaks.

Some recent requests have a large database that is unlikely to remain entirely in the focal memory, even with robust machines. Maintaining data in the derived memory and data subset grouping separately. Restricted results are retained and, at a later stage, data are obtained for the classification of the entire collection.

The use of an incremental clustering algorithm, in which all elements are automatically moved to the main memory and attached to one of the clusters provided or owed in a novel cluster. Results are reserved and the factor is not required to make room for the other. Using a related mechanism in which multiple algorithms process stored data simultaneously improves clustering efficiency. Here is a boundary series that hinders the application of traditional data mining methods in dispersed databases. The technique usually integrates the grouping of all distributed databases in the middle unit of the algorithm application. The traditional data mining strategy is strongly refuted as it is important to think about certain subjects in the survival of parallel data with different names. In addition, layouts vary from one another in data structures with variances. Instead, the integration of several databases at a different location is not recommended when the database is massive. If a company maintains massive databases and chooses to run data mining algorithms that cause huge data transmission, resulting in a slow and costly one. In addition, modifications to the distributed database, such as

the addition of new information or the alteration of a previous one, must be reorganised along with or updated with the critical database. As a result, the distributed database requires a very complex approach to updating data with excess transfer of information in the system.

The data is separated on the basis of such instances, such as when the data set is combined and separated into subsets dependent on size. The data partition is usually based on two techniques, namely horizontal and vertical separating the data, as shown in Figure 1 below.

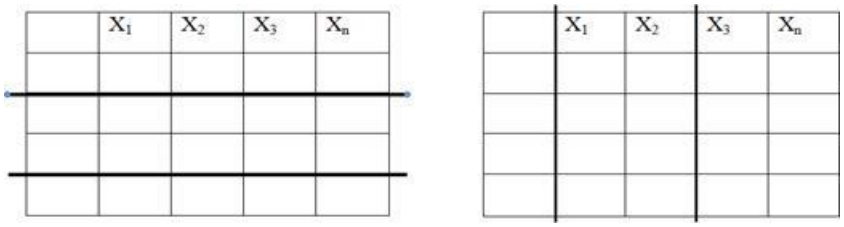

Figure 1 Horizontal and Vertical Partitioning

Above Figure 6.1 describes the horizontal and vertical separating the data. The horizontal separating approach is more utilized and includes horizontally split database producing reliable data subsets. Privacy-Preserving Horizontally Partitioned Linear Programs (Mangasarian 2012) faces interesting problem such as realm occurs and inequality constraints. Therefore, all procedure panels on diverse archives permitting for partition but with parallel set of characteristics.

Alternative technique is to vertically divide the database and create varied data subsets. In this perspective, all algorithm is enabled on identical sales records with a variety of attributes. In addition, vertically separated data through random kernels Mangasarian et al (2008) for Privacy-Preserving privately held classification data without the disclosure of any such data.

Both database separating are popular in a wide variety of research areas, often in dispersed systems and database environments where feasible applications fit into databases. Depending on the sequence of aspects of the cluster analysis, the data separation in the database is calculated. The operational requirement opens up authority in the framework of the delivery of data. In this Data mining algorithms must also be strong sufficient to accommodate these constraints.

The geometric perturbation approach Chen & Liu (2009) to multiparty privacy-preserving collaborative mining protectively combines the disrupted recycled by dissimilar users without much harm of privacy assurance and statistics utility. The current framework focuses only on the set of one service provider facing security breaches with respect to multi-partitioned service providers. Therefore, more concentration on the privacy problems in the condition of many service workers or multi-partitioned dataset is required collaboratively to provide privacy preservation.

Even though the difficulty in data perturbation is overcome, the privacy and security is still a challenge. The privacy conservation is improved by adjusting the adaptable perturbation model that intends in reducing the adversary attacks. The process of selecting the level of privacy like top, high, medium and low enhanced the privacy. In addition the noise further to the unique data preserves the data by increasing the level of privacy. Moreover the security of the system is handled against attacks. Therefore the Enhanced Privacy with disrupted Data by Choosing features process intends in significantly increasing the security level.

Preservation of privacy with disrupted data using feature selection is theoretically designed to improve privacy and protection of disrupted data in multi-divided datasets. The enhanced data-disrupted protection mechanism for multi-partitioned use of feature selection is processed in four different steps. The first step explains how to exchange multi-part data with users. The other third party users are blocked by combinatorial function from participating in the communication. The second step explains the method of clustering multi-partitioned data using a divisive k-mean clustering process.

Third step defines the procedure of improving the applicability of disrupted data in a multi-part dataset. Privacy preservation was provided using an individually customizable disturbance model that allows individuals to select their own privacy stages. The third phase was mainly designed on the purview of enhancing the privacy of individuals. As a result, the computational results were not taken into consideration. The fourth step demonstrates the process of improving the protection and privacy of disrupted data by selecting dataset features. The proposed Enhanced Privacy with Disrupted Data diagram by selecting features is shown in Figure 2.1.

Figure 2.2 describes the protection of privacy that can be achieved by adapting the flexible disturbance model. Multi-partitioned data set consists of data divided from each logical database. Data is primarily divided into two groups, namely the horizontal data partition and the vertical data partition. Horizontal partitioning

requires separate rows to be placed in different tables. The general overview of vertical partitioning is to separate dynamic data from static data in a table where dynamic data is not as much used as static data. Producing a view for both tables restores a unique performance penalty table while improving performance when accessing static data such as statistical analysis.

Previously edition of the privacy preservation arrangement, contentious k implies that clustering is realistic to a multi-part dataset that starts with a person, comprehensive cluster. Divide a cluster at each stage until only one-ton clusters of entity points remain behind. In this cluster assessment, the result is occupied on the basis of cluster selection to split and break.

The divisive k-means cluster is performed for privacy conservation appliance to defeat the issue of ambiguity among the clustered datasets in data perturbation technique that cause unreliability. The concept behind divisive K-means clustering is that an alliance argument symbolizes a cluster. Afterward negotiating the vagueness of the separated dataset, the adaptable perturbation model is performed to improve the privacy conservation system amongst the separated datasets by selecting the privacy levels.

Membership the Multi-Partitioned Dataset Using Combinatorial Function

The principle objective of data disrupted is to alter the statistics so that effective separate data standards are not increased, while maintaining the worth of the data for reviews. Even if a data item is linked to an entity, privacy is not breached because the data does not respond to the actual values of private data. If User A and User B decide to split the data in the database, then initial task is to partition data either in a horizontal and vertical manner suitably.

Gradient descent model is incorporated to partition the dataset horizontally. For vertical divider of data in the dataset, kth element vector method is utilized. Afterward data partition either horizontally or vertically since the database, the data is shared between the users in a harmless and safe way. The security is provided by involving only the two users to participate. The key idea is to block extra third party involving in this data allotment between the users. Then customers partition the data from the database in a vertical way. Users receive their sets of vertically separated data sets.

In order to portion the data sets with dissimilar users lengthways with privacy conservation, the data Perturbation technique is performed. . Data disturbance technique shared the data to dissimilar users and collected all the information in order to obtain one complete set of true data. The combinatorial function is used to reserve data sets that are to be shared between users and third party members for data access to be blocked. The Combinatory function allows users to share the two separate data sets with different users.

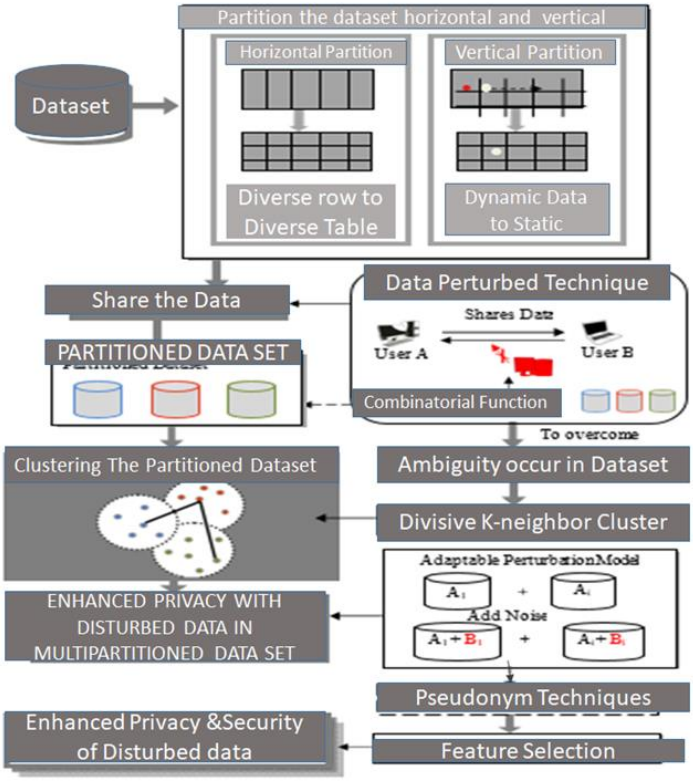

Figure 2.2 Diagram Enhanced Privacy with disrupted Data by Choosing features

Division k-mean clustering is a hierarchical clustering used in the privacy preservation for multi-partitioned dataset. The task performed in categorized clustering is top-down clustering. The main idea behind clustering is to build up altogether datasets in one group. The cluster is partitioned based on horizontal clustering algorithm.

Adaptable Perturbation Model

Adaptable perturbation model intends in increasing the privacy preservation of the data by appropriately choosing the level of privacy. The perturbation technique is processed on setting up noise without modifying the allocation of the unique data. At the same time, diverse geometric methods are performed to the disturbed data to update the unique allocation. Data hammering in conflict to privacy protection is evermore a trade off in adaptable perturbation model. The sum of data generates problem with the unique data distinctly about the data mining consequences.

Adaptable perturbation model is divided into one phase and two phase Disrupted model. The one phase process is to adjoin the noise to the unique information and then apply the renovation algorithm to approximate the unique delivery. The two phase split the province of the dataset into determined periods and customized to integrate diverse individual privacy inclinations. The noise addition to the data preserves the privacy. In addition privacy level selection strategy by the users achieves enhanced privacy preservation.

Function Selection Scheme for Privacy

Enhanced Privacy with Disrupted Data by Choosing Features (EPDDCS) works on disturbed data to attain data mining arrangements as it processes on the unique data. However, the disrupted data does not understand the outcome or class of the data and can only indicate the effects of all the partitioned data in the data sets. Since all partitioned data only recognizes the outcome of the data steps, data protection is maintained. Enhanced privacy with data interruption using feature selection work gathers data using a private security pseudonym technique.

The pseudonym technique is used in the disturbed system to seal the individual disturbed data procedures. Thus, the transformed data groups contain only the pseudonym technique used to analyse the disrupted data in the data collection, i.e. the data miner, without collecting any authentic values. The pseudonym technique produces data noise in disturbed data. Perturbed data declares that the actual data is transferred to some other form of unique data or that the actual database is changed to a custom perturbed database. The disturbed data is created by an analysis of the queries based on the original datasets.

Feature selection scheme for privacy preservation is used mainly for searching feature space. With the help of an optimal type of search, the feature selection results in an optimal solution. The feature selection for privacy preservation uses pseudonym technique and result in an O(N2) worst case search. The feature selection for multi partitioned datasets uses sequential selection, which initiates with the entire feature space and iteratively removes the feature that has the least of criterion function one after the other. Each set of features and the respective values of the sub-features are processed using both the pseudonym set of features and the pseudonym values of the sub-feature. Figure 3 displays the original database of disrupted security info. Both specific and disrupted databases are provided by the framework.

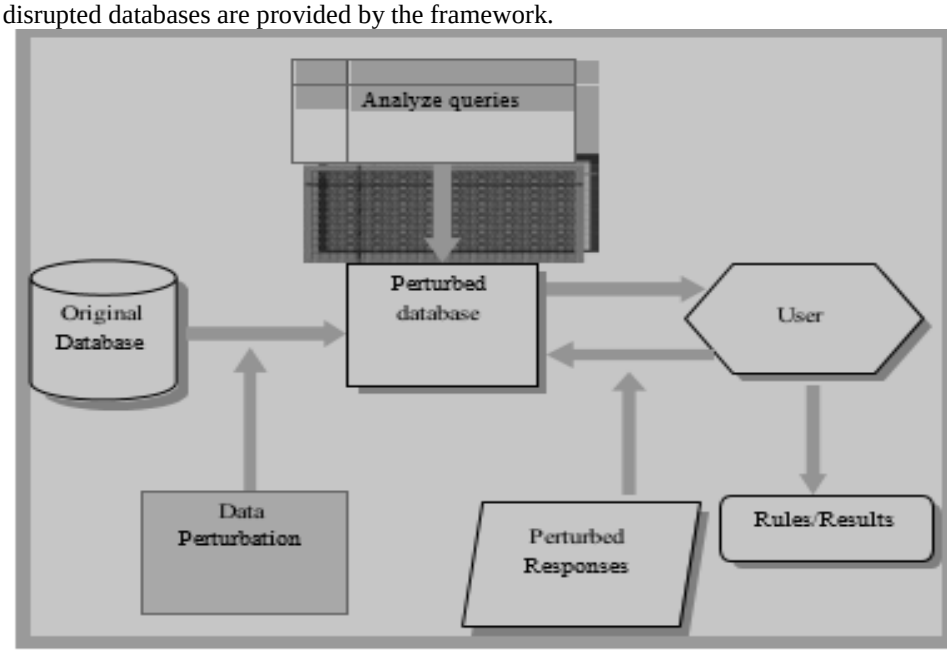

Figure 3 Perturbed or Disrupted Data Model

Let DA be the dataset, and C is the class of the database. When the whole dataset is divided into a partition number to detect a distinctive identifier attribute, the information that is vital to the categories of the DA data set supported by this partitioning will be info (DA) = 0. Since knowledge gain on this set of features is the upper limit on all pure partition, there is no critical categorization of this type of p. The split info for feature X is acknowledged as SIX as processed by

Type equation here.

$$SI_X(DA) = f(x) = \sum_{j=1}^{p} \left(\frac{|DAj|}{|DA|} \log 2 \left(\frac{|DAj|}{|DA|}\right) \quad \right) \qquad \text{……. (1)}$$

Where DAj is the jth separation of the D that fits the X function. The average value of the number of tuples in the different partition for each product is the total number of tuples in D. Tuples distinguishes the information categorization steps from the information advantage provided by similar partitioning. The benefit ratio shall be defined below.

$$Gainration(x) = Gain(X)/SI(X) \qquad \text{…… (2)}$$

Let Di, the D and Cj partition data, indicate different classes. Let Pi be the probability of the D-subjective tuples that fit into the Cj class. The estimated information of D for a number of classes shall be determined by

$$\text{Info}(DA) = \sum_{J=1}^{allclasses} (Pi\ log2\ (Pi)) \qquad \text{…….. (3)}$$

Here is the data (D) of the approximate details of D for the identification of the class name. However, the precise categorization is considered to be an individual function after dividing the data as

$$info_X(DA) = \sum_{J=1}^{No,of\ partitions} \frac{|DAj|}{|DA|} * \text{info} \mid DAi \mid \qquad \text{……..(4)}$$

where $\frac{|DAj|}{|DA|}$ is the weight of the i th parting of data and Infox(DA) is the valued information vital to categorize the tuples. The information gain is labelled as the difference among info (DA) and infoX(DA) ie

$$\text{Gain (X)} = \text{info(DA) - infox (DAi)} \qquad \text{……. (5)}$$

The function containing the highest gain ratio is selected as a terrible feature. However, the information on the disturbed dataset should not be zero, as the ratio tends to be unbalanced. Thus, when the collection of disturbed data is inserted into the respective features, the dishonored split information results when the gain ratio is considered. The drawback for this dimension is that the information obtained from the chosen analysis must be broad. Indicating that, at a minimum, standard experiments of disrupted datasets have been examined. The role of privacy is to ensure that any peer node participating in the network is maintained. Various colluding sets of disrupted data split the dimension in order to select the best function.

The following pseudo code explains the method of Enhanced Privacy Protection with Data Disturbance using Feature Selection.

Phase 1: Compile the occurrences of disturbed pseudonym name info.
Phase 2: Exchange pseudonym data to a real data set.
Phase 3: Based on class distribution, partitioned data is sent to each class.
Phase 4: The # instances are labelled as their individual condition in each class.
Phase 5: The partitioned data set is labelled with both a article set and a class.
Phase 6: Process the gain ratio method for the selection of features.
Phase 7: Set the values for optimal selection of features to enhance the privacy of disrupted data.
Phase 8: Maintain privacy for best data set features

Through the above steps, EPDDCS delivers protection for disrupted data in data mining applications. The next section explains the experimental evaluation of the future EPDDCS scheme and contrasts the findings with current approaches.

EXPERIMENTAL EVALUATION

Enhanced Privacy with Disrupted Data by Choosing Features [EPDDCS] is implemented in Java. The experiments were run on an Intel P-IV machine with 4 GB memory and 4 GHz dual processor CPU. Compare the Enhanced Privacy with Disrupted Data by Choosing Features [EPDDCS] with Privacy-preserving Multiparty Collaborative Mining with Geometric Data Perturbation.

Datasets are effectively partitioned horizontally or vertically using a combination feature. As a result, the scalability of goods or services has decreased. The robust size of the data set to be exchanged preserves the same amount from the beginning of the splitting process using a split k-means clustering algorithm. With the proficient negotiation of ambiguity problem in dataset, adaptable perturbation model is utilized in order to reduce the adversary attacks through addition of noise and privacy level selection. At the end of the day, the privacy and protection of disrupted data in multi-partitioned datasets is protected through a feature selection process. Performance of Enhanced Privacy with Disrupted Data by Choosing Features [EPDDC] is calculated in terms terms of (i) Privacy Assurance, (ii) adversary attack rate and (iii) authenticity.

RESULTS AND DISCUSSION

Enhanced Privacy with Disrupted Data by Choosing Features [EPDDC] is consistently prepared aimed at improving the privacy conservation in multi-partitioned dataset. The EPDDC allow the users a secure sharing of files with other users through data perturbation technique and combinatorial function. Partitioned dataset are clustered based on which the privacy levels are selected through adaptable perturbation model for high privacy level. Additionally the security of the perturbed data is increased through feature selection process. An experimental evaluation test is conducted with benchmark dataset in order to approximation the performance of the EPDDC. The under table and chart describes the ability of EPDDC and compared results with an existing technique termed geometric perturbation approach Chen & Liu (2009) to multiparty privacy-preserving collaborative mining.

PRIVACY ASSURANCE

Certainly, all users resolve to preserve the gain of privacy guarantee that is favoring optimized perturbations in multi-partitioned dataset. The principle objective of each method in providing privacy guarantee is to secure information about the data in dataset, without leaking individual information about participants to the adversaries.

$$\text{Privacy Guarantee} = \frac{\text{Secure (Data)}}{\text{Multipartitioned Dataset}}$$

Both the enhanced perturbed data using feature selection and geometric perturbation approach intends in achieving better security and the values are elaborated in the Table 1.

Table 1 User Density vs. Privacy Assurances

User Density	Privacy Assurances (%)	
	Geometric Perturbation Approach	EPDDC
10	80	94
20	78	88
30	75	87
40	71	85
50	67	80
60	62	77
70	56	72

The Table 1 describes the privacy and shared data of the parties in the environment. The privacy guarantee of enhanced privacy with disrupted data using feature selection is compared with the existing geometric disruption approach to multi-party privacy-preserving collaborative mining.

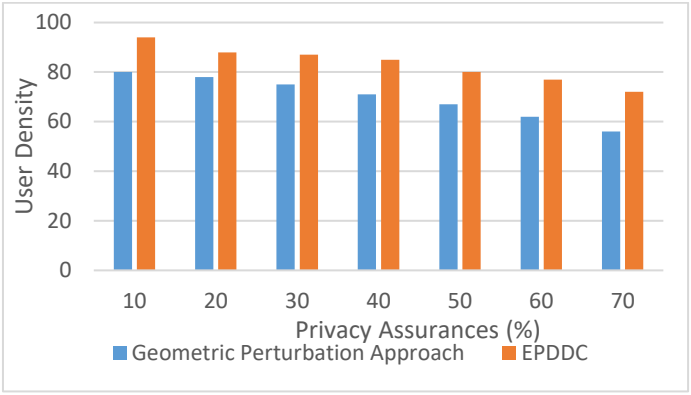

Figure 4 User Densities vs. Assurances

Figure 4 describes the privacy assurances provided to the perturbed data. Enhanced Privacy with Disrupted Data by Choosing Features achieved better privacy assurances of about 15-25% compared with geometric perturbation approach. As perturbed data are arranged based on pseudonym name, the partitioned data are sent to individual classes with exchange of pseudonym data to authentic set of data guaranteeing the privacy in EPDDC. Whereas in geometric perturbation approach Chen & Liu (2009) the data provider conveys encoded disturbed data to the service provider, thus searching data providers is unable to solve any useful information from eavesdropping.

Success Rate

Success rate of the perturbed data is decided based on the level of privacy and security offered. Each data is preserved in order to fall under illegal transformation. Therefore the data is set with minimum satisfaction level to identify the success rate. Generally, success rate defines the privacy and vulnerability of data. If the users are not satisfied with the privacy and on any good perturbation in multi-partitioned dataset, the data is set to unsuccessful rate. Success rate is defined as

$$\text{Success Rate} = \frac{\text{Data (High Privacy + Security)}}{\text{Dataset}_{\text{Minimum Satisfaction Level}}}$$

A Table 2 and Figure 5 show the difference between providers transmits encrypted perturbed data to the service provider, thus sensitive data providers and geometric perturbation approach in terms of success rate with minimum cost and benefit, justifying that Enhanced Privacy with Disrupted Data by Choosing Features process.

Table 2 Minimum Satisfaction Level vs. Success Rate

Minimum Satisfaction Level	Success Rate (%)	
	Geometric Perturbation Approach	EPDDC
0.5	41	55
1	38	51
1.5	35	44
2	31	40
2.5	28	38
3	26	35
3.5	24	33

Table 2 describes the success rate of providing privacy and security for perturbed data from unauthenticated data access. The success rate of the Enhanced Privacy with Disrupted Data by Choosing Features process.is compared with geometric perturbation approach Chen & Liu (2009).

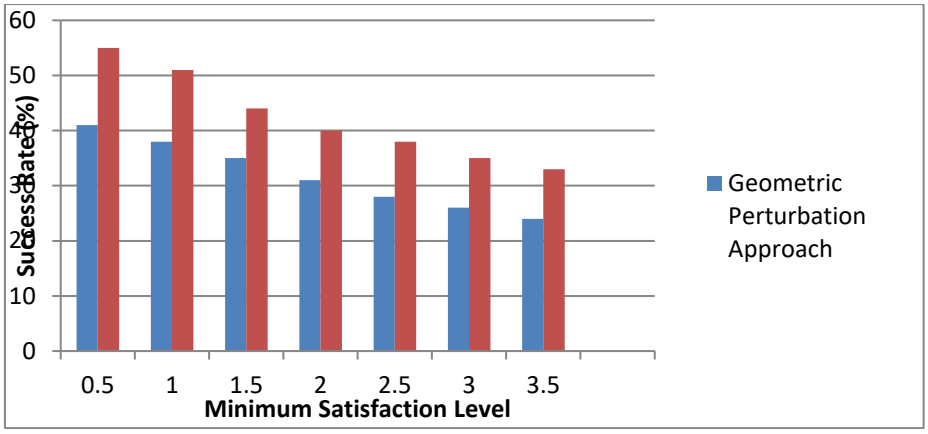

Figure 5 Minimum Satisfaction Level vs. Success Rate

Figure 5 describes the success rate of blocking unauthenticated access of data from third party. Success rate of EPDDC is about 25-32% compared to geometric perturbation approach. As the separated set of data are marked with together feature set and classes facilitates the procedure of feature selection that leads to better success rate in EPDDC. Since each subset holds very different distributions in geometric perturbation approach which results in different perturbation with minimum success rate.

Authenticity

Authenticity is characterized as the quality or state of authenticity, trustworthiness or authenticity of disrupted data in a multi-partitioned dataset. As the size of the partitioning data set increases, the authenticity factor of the data sets should be high to improve the partitioning process. The purpose of improved privacy for disrupted data using feature selection is to preserve the privacy and confidentiality of authentic data.

Table 3 Number of Parties vs. Authenticity

Number of Parties	Authenticity (%)	
	Geometric Perturbation Approach	EPDDC
1	76	89
2	73	84
3	69	77
4	66	74
5	61	70
6	59	68
7	54	64

Table 3 the validity of the partitioned data in the dataset is defined in Table 3. The authenticity level of Enhanced Privacy with Disrupted Data by Choosing Feature is contrasted with the Chen & Liu (2009) geometric disruption approach

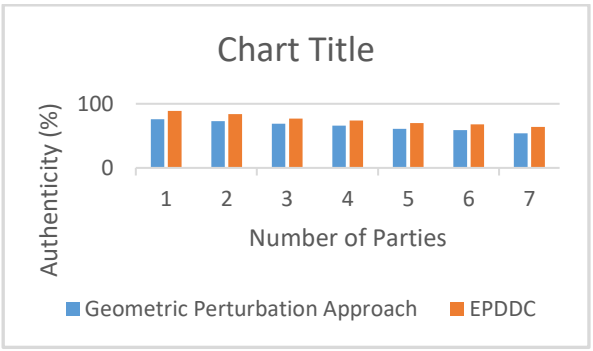

Figure 6 Number of Parties vs. Authenticity

Figure 6 describes the authenticity in restricting unauthenticated access of data by third user that tries to get into the transaction of message among parties. Enhanced Privacy with Disrupted Data by Choosing Feature allows better authenticity of about 9-14% compared with geometric perturbation approach. As the feature collection values from gain ratio are sorted for optimal feature selection, the authenticity of the data is maximized to a better level in EPDDC. While geometric perturbation approach is able to provide authenticity and privacy only to one service provider and is not applicable for multi-partitioned dataset.

Finally, the experimental evaluation justified the efficiency of EPDDC in achieving better privacy and security with the feature selection process of perturbed data. Perturbed data arrangement based on pseudonym name and the feature collection values sorting based on gain ration increased the authenticity. The EPDDC achieved better results in terms of high privacy

Assurance for about 10-15%, higher success rate of about 20-25% and better authenticity of about 7-12% compared with geometric perturbation approach Chen & Liu (2009)to multiparty privacy-preserving collaborative mining.

CONCLUSION

Enhanced Privacy with Disrupted Data by Choosing Feature justified higher privacy Assurance and better security. In the EPDDC, The data miner has discovered an immense amount of data to create a classification model. Typically, the classification of individual instances retains more details. Data mining processing work depicted the collection of features using gain ratio technique for the best feature as a structure. The set of features is ordered from the data mining system. The sensitive and non-sensitive function enables the classification of data model for the protection of data privacy by data miners.

Sharing of information with other users is made more secure with partitioning of datasets efficiently in together horizontal and vertical method. Combinatorial function restricts the third party participating in the data sharing and dataset are partitioned. Partitioned dataset are clustered efficiently which increases the privacy conservation system by adapting the one stage and two stage Disrupted model. In addition the privacy preservation is improved through the selection of privacy levels.

Adaptable perturbation model is utilized that added noise to the innovative information preserving the data. Disrupted data are arranged based on pseudonym, the partitioned data are sent to individual classes with exchange of pseudonym data to authentic set of data guaranteeing the privacy in EPPDFS. The partitioned set of data are marked with together feature set and classes facilitates the procedure of feature selection. The feature collection values from gain ratio are sorted for optimal feature selection that improves the authenticity of the data. Finally, the privacy and security of the perturbed data is still improved in EPPDFS.

Enhanced Privacy with Disrupted Data by Choosing Features process [EPDDCF] performs fine compared to geometric perturbation approach to multiparty privacy-preserving collaborative mining. The experimental evaluation carried out with bank data sets collected from common e-business / e-commerce sites demonstrated the enhanced performance of EPDDCF. Performance of the proposed enhanced privacy with disrupted data using feature selection justified satisfactory results in terms of privacy assurance, success rate and authenticity.

REFERENCES

1. Kun Liu, Hillol Kargupta, et. Al., “Random Projection-Based Multiplicative Data Perturbation for
2. Privacy Preserving Distributed Data Mining” , IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 18, NO. 1, JANUARY 2006.
3. P. Kamakshi , Dr. A. Vinaya Babu, “Preserving Privacy and Sharing the Data in Distributed Environment using Cryptographic Technique on Perturbed data”, JOURNAL OF COMPUTING, VOLUME 2, ISSUE 4, APRIL 2010, ISSN 2151-9617.
4. Ying peng Sang, Hong Shen et. Al., “Effective Reconstruction of Data Perturbed by Random Projections”, IEEE TRANSACTIONS ON COMPUTERS, VOL. 61, NO. 1, JANUARY 2012.
5. Jaideep Vaidya, et. Al., “Privacy-Preserving Kth Element Score over Vertically Partitioned Data”, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 21, NO. 2, FEBRUARY 2009.
6. Dhiraj, S.S.S. et. Al., ‘Privacy preservation in kmeans clustering by cluster rotation’, IEEE Region 10 Conference TENCON 2009 – 2009 .

7. Natwichai, Juggapong , “An approximation algorithm for privacy preservation of associative classification”, International Conference on Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), 2010 .
8. Keshavamurthy, B.N. et. Al., ‘Privacy preserving Naive Bayes classification using trusted third party and different offset computation over distributed databases’, 1st International Conference on Parallel Distributed and Grid Computing (PDGC), 2010.
9. Gopalakrishnan, R., Mohan, A., Sankar, L. P., & Vijayan, D. S. (2020). Characterisation On Toughness Property Of Self-Compacting Fibre Reinforced Concrete. In Journal of Environmental Protection and Ecology (Vol. 21, Issue 6, pp. 2153–2163).
10. Deivanai, P. et. Al., “A hybrid data anonymization integrated with suppression for preserving privacy in mining multi party data”, International Conference on Recent Trends in Information Technology (ICRTIT), 2011 .
11. Fan Zhang et. Al., ‘Data perturbation with statedependent noise for participatory sensing’, Proceedings IEEE INFOCOM, 2012 .
12. Tholkapiyan, A.Mohan, Vijayan.D.S, A survey of recent studies on chlorophyll variation in Indian coastal waters, IOP Conf. Series: Materials Science and Engineering 993 (2020) 012041, 1-6.
13. Rajaram, A., & Palaniswami, S. (2010). Detecting malicious node in MANET using trust based cross-layer security protocol. *Intern J Comput Science Information Technologies*, *2*, 130-137.
14. Li Liu , Murat Kantarcioglu et. Al., ‘The applicability of the perturbation based privacy preserving data mining for real-world data’, Science direct on Data & Knowledge Engineering 65 (2008) 5–21.
15. Keke Chen et. Al., “Privacy-Preserving Multiparty Collaborative Mining with Geometric Data Perturbation”, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 20, NO. 12, DECEMBER 2009.
16. Li Liu et. Al., “The Applicability of the Perturbation Model-based Privacy Preserving Data Mining for Real-world Data”, Sixth IEEE International Conference on Data Mining Workshops, 2006. ICDM Workshops 2006.
17. Prakash V. S, A. Shanmugam, “Enhanced privacy preservation with perturbed data using feature selection‟, Journal of Theoretical and Applied Information Technology(JATIT), E-ISSN 1817-3195/ISSN 1992-8645, Vol. 58, no. 3, 2013.