Framework for Privacy Preserving Classification in Data Mining

Dr. Yogesh Kumar Sharma[1], Ghouse Mohiyaddin Sharif G.M[2]
[1]Associate Professor, Department of Computer Science, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan
[2]Research Scholar, Department of Computer Science, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan

Abstract: In the present period of developing innovation the information gathered by associations has the necessity to protect the security of the people. It needs to keep up protection of the people since clients delicate information is put away online over the brought together vault. The procedures like anonymization, randomization are utilized to accomplish the protection. In any case, anonymization prompts certain level of data misfortune while safeguarding protection. To beat this disadvantage, cross breed approach is utilized. The proposed framework includes blend of two strategies i.e anatomization and irritation methods. The semi identifiers like postal division, age, sexual orientation of a man does not appear to be critical to ensure but rather these fields when connected with some different traits can uncover the character or delicate data of a person. The cross breed technique centers around the objective of saving protection by examining and irritating the semi identifiers in the touchy information of clients put away on incorporated information archive without making any misfortune the data.

Keywords— Privacy preserving; Sequential pattern mining; Anonymity; Randomization; Secure multiparty computation; Sequential pattern hiding

1. INTRODUCTION

Inside the Privacy Preserving Data Publishing (PPDP) network, averting delicate data about people from being induced is a best need. This is known as "anonymization". One of the key ideas in PPDP is the exchange off that is characteristically present when "anonymizing" information: adjusting the expansion in security with the abatement in data quality. The dominant part of past work has concentrated on the troublesome issue of characterizing and estimating security [1], [2]. This paper investigates the opposite side of the exchange off: data quality. A considerable measure of the time, shortsighted measures are produced to give a gauge of the data quality, or measurable methods are acquired from the SDC (Statistical Disclosure Control) people group. While hearty, these assessment methods regularly neglect to catch the subtleties that can be available while assessing particular anonymization assignments, for example, speculation 1 . Data estimates that objective particular anonymization undertakings tackle this issue; anyway contrasting the consequences of diverse measures is a progressing issue. "Speculation" alludes to making an esteem vaguer, for example, changing all events of "apple" and "banana" to "organic product". 2 A "dataset" is a two dimensional table where columns speak to free records (tuples) and sections speak to different properties that depict the records and separate them from each other.

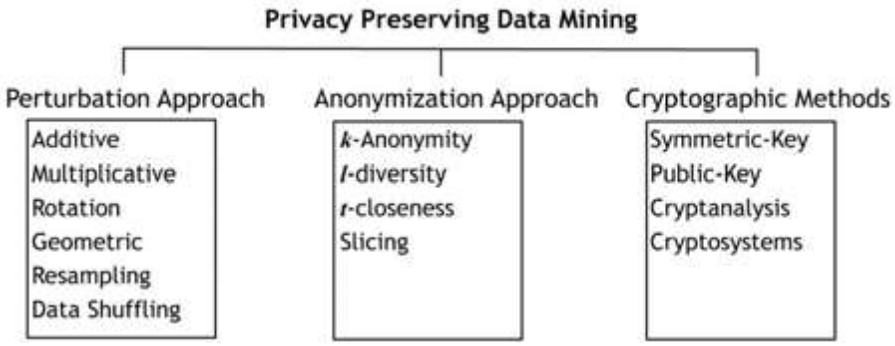

Fig 1. Privacy Preserving Data Mining Technique

In PPDP, the data nature of an anonymized dataset is regularly assessed by estimating the closeness. The dataset is assessed in a way that applies to any situation – we allude to this as estimating the "dataset quality" or "dataset data misfortune". These sorts of strategies are examined in Section II.

On the other hand, if the reason for the dataset is particular and known, the data quality can be estimated in regard to that reason. Protection Preserving Data Mining (PPDM; an organization of PPDP) centers around this sort of information, where the nature of the dataset itself is less imperative than the nature of the yielded information mining3 results delivered from the dataset.

Basic intentions are classification4 and bunching 5 [2]. Numerous examples in the dataset can be lost after anonymization, regardless of whether the dataset itself seems to hold the vast majority of its measurable data [6]-[8]. Thus, data measures have been composed that particularly take a gander at the impact of anonymization on information mining results, and we examine these in Section III. We call this sort of data quality, "information mining quality" or "information mining data misfortune". It ought to be noticed that we make a qualification between "data misfortune" and "data quality" due to the suggested relative nature of "misfortune" – this paper centers around measures that think about a dataset previously, then after the fact change. "Data quality" could allude to this when correlation, yet additionally to the nature of an detached dataset

(with no correlation). "Data misfortune" gives greater specificity. Proportions of data misfortune are additionally usable in situations outside of protection safeguarding, such as information ascription/information cleaning 6 . In these occasions, data gain is the objective.

1.1. Private information investigation

Information is given to an information accumulation by people, the overseer of the information accumulation at that point turns into the information proprietor. Information examination can be performed by the information proprietor or the information proprietor can outsource the information examination to other parties. Regardless, the protection worries of the included people ought to be tended to and considered constantly.

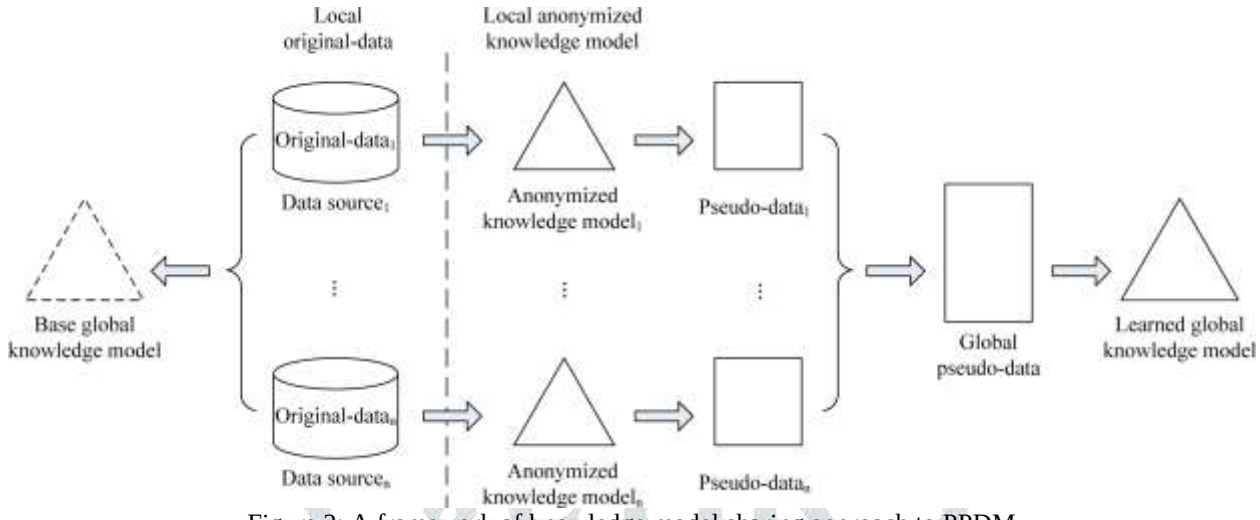

Figure 2: A framework of knowledge model sharing approach to PPDM

The subject of how to acquire legitimate outcomes furthermore, information without taking in the hidden private information is contemplated in private information investigation, additionally alluded to as security safeguarding information mining [2], [4]. Private information investigation is achievable in the accompanying ways:

• Private information investigation over unique information. In this situation, calculations are performed over the first private or even classified information.

– Data investigation is performed by the information proprietor. No other gathering will take in the information, and the consequences of the examination will remain "in house".

– Data mining is performed over the first information and afterward the acquired information is distributed. The distributed learning is secured against protection spills in a way that it doesn't uncover delicate data about the fundamental information. This is achievable by disinfecting the learned information and alluded to as the protection saving learning distributing [7], [3].

– One or a few gatherings claim secret information and another gathering performs a calculation over them. Secure multiparty calculation [15], [8] and secure multiparty calculation over conveyed informational indexes are fields that review cryptographic instruments that permit to register a capacity over classified information without getting the hang of whatever else than what can be gained from the yield of the capacity.

• Data investigation over purified information. In this situation, information is purified and after that mutual or distributed for investigation. This is alluded to as protection - saving information distributing (PPDP) [1], [20]. Purification is generally accomplished as a change of the information that gives pseudonymity, secrecy, or on the other hand security hazard decrease by summing up, veiling, randomizing, or even smothering a few information. Whatever is left of this paper bargains solely with this situation.

2. PPDM models

There three fundamental models of protection safeguarding are:

1. Trust outsider model: It is accepted that there exists an outsider that to whom every one of the information is given. It is comprehended that nobody other than that specific outsider has any entrance to the information. The point is to actualize security safeguarding procedures to keep up the secrecy of this outsider.

2. Semi fair model: Very gathering takes after specific conventions utilizing right configurations yet it is allowed to utilize any convention amid the execution on the off chance that it feels the security is undermined.

3. Malignant model: Since the semi-genuine model does not give insurance to all applications the noxious model is utilized. The malignant model is allowed to utilize any ensure it wishes to secure protection. It is hard to create proficient conventions for this model.

VII. Protection PRESERVING TECHNIQUES\

The protection saving systems are ordered into five classifications:

Anonymization based Privacy Preserving:

The information in the table comprises of 4 distinct qualities:

1. Express Identifiers: It is an arrangement of traits that recognize a proprietor record unequivocally.

2. Semi Identifier: These are an arrangement of traits if joined with a freely accessible tuple would recognize the proprietor.

3. Touchy Identifiers: A trait which contains delicate data about the proprietor e.g. pay.

4. Non-Sensitive Identifiers: If uncovered such traits make no protection issues.

Anonymization is an approach where touchy data around an individual is to be covered up. Semi identifiers when consolidated to a publically accessible database can uncover touchy data. For e.g. in the event that a voters ID database is joined with the representative database of an

organization, delicate data like the pay of a specific worker can be uncovered. So k-anonymization includes stowing away or the adjustment of certain semi identifiers with the end goal that when the information is imprint in for information mining the semi identifiers are not unveiled. Where if there is 1 semi identifier then the information sent for information mining will have k-1 tuples. Along these lines securing protection. This is expert by speculation and concealment. In spite of the fact that the anonymization strategy guarantees that the changed information is right yet it endures overwhelming data misfortune. This technique isn't resistant to comparative assault and foundation learning assault. Constraints of the k-namelessness demonstrate are, first, it might be hard for the proprietor of a database to figure out which of the qualities are accessible or which are not accessible in outside tables. The second confinement is that the k-namelessness display accept a specific strategy for assault, while in genuine situations; there is no motivation behind why the aggressor ought not attempt different strategies.

The word reference significance of irritated is being disrupted or vexed. So also in information mining bother implies supplanting or disquieting the first qualities with some manufactured information esteems to such an extent that the factual data of the information is safeguarded. The information records don't relate to the person's real information. So in a vindictive assault, cross connecting is impossible to recover delicate data. In this way protecting security. Consequently the irritated information are meaning less and contain just factual data. Annoyance is finished by adding commotion to the first information, information swapping or engineered information age.

Randomized Response based Privacy Preserving

Randomized reaction is a measurable method. In this the information is mixed so that the focal place can't tell if the information that is originating from the client contains genuine or false data.

The data got from every individual client is mixed and if the quantity of clients is more, the combined data of all clients can be assessed with an entirely decent precision. This is extremely helpful for choice tree characterization since choice tree grouping depends on total estimations of a dataset, instead of individual information things. The information accumulation process is a twostep procedure. Amid initial step, the information suppliers randomize their information and transmit the randomized information to the information collector. In second step, the information beneficiary recreates the first conveyance of the information by utilizing a dispersion reproduction calculation.

Cryptography base Privacy Preserving

Cryptographic systems depend on the essentials of conveyed processing, where numerous gatherings meet up to register results or offer non delicate mining results and maintaining a strategic distance from divulgence of touchy data. Cryptographic procedures are useful as a result of two reasons. In vertically divided information between various associates, the individual elements may have diverse properties of same arrangement of records and if there should arise an occurrence of on a level plane apportioned information, singular records are spread out over different elements, every one of which has a similar arrangement of characteristics.

3 PREVIOUS WORK

Numerous protection procedures have been proposed in the setting of factual databases while few security strategies have been proposed with regards to information mining (see, for instance, Agrawal what's more, Srikant 2000, Du and Zhan 2002, Lindell and Pinkas 2000, Oliviera and Zaane 2002). A fantastic study, on the current protection techniques for factual databases, has been exhibited in (Adam and Wortmann 1999). These strategies have been sorted in three principle gatherings, in light of the methodologies they take, for example, question limitation, information bother, and yield irritation. Among these strategies information irritation technique is the most direct to execute. We just need to bother the information. At that point we can utilize any current programming (eg. DBMS) to get to the information with no further limitations on handling. That isn't the situation with question limitation and yield annoyance. Hence, we feel that this technique is the most reasonable for information mining applications, what's more, as far as anyone is concerned no different techniques have been researched in this unique circumstance.

The least difficult variant of added substance information annoyance was proposed by Kim (1986). This technique experiences inclination identified with the changes (Type An), inclination identified with the connection between private qualities (Type B) and inclination identified with the connection among private and non-private qualities (Type C). Quite a long while later, Tendick and Matloff (1994) displayed an adjusted rendition of arbitrary information annoyance, which is free of Type An and Sort B predisposition.

In 2002, Wilson and Rosen (2002) looked at GADP strategy with a gullible clamor expansion technique called Simple Added substance Data Perturbation (SADP) strategy in the setting of information mining. They recommended the presence of a new predisposition called Type Data Mining (DM) inclination and consequently endeavored to demonstrate that GADP strategy isn't sans predisposition in the setting of information mining.

Estivil-Castro and Brankovic (1999) proposed an information irritation method by adding clamor to the class trait. The method accentuated the example protection as opposed to acquiring fair factual parameters.

Information mining from a reasonable data hones point of view was first examined in [15]. O'Leary contemplated the effect of the OECD rules in learning disclosure. The key finding of this examination was that the OCDE rules couldn't foresee or address numerous imperative issues with respect to learning revelation, and along these lines a few standards are excessively broad or unenforceable. Our work here is symmetrical to that one in [15]. We examine the impact of the OECD standards with regards to PPDM ordering them in various gatherings of importance. Specifically, we demonstrate that the OECD rules are acknowledged worldwide and in this way they speak to the essential parts for institutionalization in PPDM. We talk about how the network in PPDM could infer a few standards and strategies from the OECD rules.

All the more as of late, Clifton et al. examined the significance of PPDM as an establishment for additionally explore in this field [3]. That work presents a few definitions for PPDM and examines a few measurements for data revelation in information mining. The work in [3] is reciprocal to our work. The essential objective of our work is to advance institutionalization issues in PPDM. Our exertion incorporates the outline of security standards and strategies, and prerequisites for the improvement and sending of specialized answers for PPDM.

4. PROPOSED ALGORITHM

1-In this paper, a semi-genuine model for enemy is utilized, where each gathering takes after accurately the convention of secure processing capacity however inquisitively endeavor to deduce information about different gatherings. A key outcome which is likewise utilized in this work is the sythesis hypothesis. We state it for the semi-legit display.

Hypothesis (1): "Assume that g is secretly reducible to f also, that there exists a convention for secretly processing f.

At that point there exists a convention for secretly registering g". Freely the arrangement hypothesis states if a convention comprises of a few sub-conventions, and can be appeared to be secure other than the summons of the subprotocols, on the off chance that the sub-conventions are themselves secure, at that point the convention itself is likewise secure. A definite exchange of this hypothesis, and in addition the verification, can be found in [26].

2-The proposed calculation exhibits a technique for secretly registering information mining process from disseminated sources without unveiling any data about the sources or their information aside from that uncovered by last arrangement result. The proposed calculation builds up a answer for privacy– saving k-closest neighbor arrangement which is one of the generally utilized information mining errands.

The proposed calculation figures out which of the nearby results are the nearest by recognizing the class of least weight utilizing K closest neighbors. We accept that traits of the occasion required for characterization are not private (the protection of the inquiry example isn't ensured). In this way, it is important to secure the security of the information sources i.e. a site/party Si isn't permitted to pick up anything about any of the information of the different gatherings, however is trusted not to intrigue with other gatherings to uncover data about the information.

3-The possibility of the proposed calculation depends on discovering K-closest neighbors of each site, at that point scramble also, scrambles the nearby $di\ min$ with homomorphic encryption and its class yi with the general population key ei sent from Encryption Decryption Management Server (EDMS). The outcomes from all destinations are consolidated to deliver the stage table at EDMS and occurrence with least weight with its class is resolved as the class of questioning case which is exchanged to questioning site.

Each site adapts nothing about different locales. Since the KNN calculation executed locally for each site Si. The standard information mining calculation is K closest neighbor for each site/party Si will be as per the following:

1-Determine the parameter K=number of closest neighbors previously.

2-Calculate the separation between the inquiry occurrence and all the preparation tests utilizing
Euclidean separation calculation.

3-Sort the separations for all the preparation tests what's more, decide the closest neighbor in light of the Kth least separations.

4-Since this administered learning, get every one of the classes of preparing information for the arranged esteem which falls under K.

5-Use the larger part of closest neighbor as the expectation esteem.

Documentations:

(x) intends to scramble information x utilizing an uncommon encryption calculation E. ; alludes to scrambling information x utilizing an uncommon calculation E with a predefined key k.

The Integrated PPDM Algorithm of K Nearest Classifier is as per the following:

1-Require: m parties, yi class esteems, l quality values, Xp question occasion { 1, $x2$,... xl}
2-Pi and Qi are expansive security prime numbers.
$Ni = Pi \times Qi$
3-(e, di) speak to the encryption and decoding keys of RSA calculation are produced at Encryption Decryption Management Server (EDMS).
4-$di\ min$ speaks to least neighbor separate with dominant part class in respect to inquiry occasion , and class mark yi is the relating class of $di\ min$
5-For I =1 ... m do/producing encryptiondecryption keys
6-EDMS creates (, di) utilizing RSA Algorithm ;
7-Transport ei to Party ;
8-End For/creating encryption-unscrambling keys
9-For I =1 ... m do/examine m parties, registering $di\ min$ and encryption process
10-Party Si locally registers $di\ min$ and its class esteem Cli as per K closest calculation with respect to question example X .
11-Encrypt $di\ min$ as in Eq. (2) to get homormophic encryption $Ei\ di\ mi$.
12-RSA encodes Pi to $Ci = Eei\ Pi$ and class mark yi to $Cli = Eei\ y$;
13-Transport $Ei\ di\ m$, Ci , and Cli to EDMS;
14-End For/figuring $di\ min$ and encryption process
15-For I =1 ... m do/Decryption process at EDMS
16-Decrypt $Ei\ di\ min$ according to Eq. (3) to get $di\ min$ and Cli to get yi
17-End For/decoding process
18-Construct the mapping table that maps the relative distinction between $di\ min$ with all $dj\ min\ i \neq j$ and $i, j \in 1, m\ to + 1, -1$
19-Calculate the weight for each column in the mapping table by including the column components and get the aggregate.
20-Determine the worldwide min remove which compares to min weight in the mapping table.
21-Get the anticipated class that match worldwide min remove (min weight in the mapping table).

1-Require: m parties, yi class esteems, l quality values, Xp question occasion { 1, $x2$,… xl}
2-Pi and Qi are expansive security prime numbers.
$Ni= Pi \times Qi$
3-(e, di) speak to the encryption and decoding keys of RSA calculation are produced at Encryption Decryption Management Server (EDMS).
4-$di\ min$ speaks to least neighbor separate with dominant part class in respect to inquiry occasion
, and class mark yi is the relating class of $di\ min$
5-For I =1 … m do/producing encryptiondecryption keys
6-EDMS creates (, di) utilizing RSA Algorithm ;
7-Transport ei to Party ;
8-End For/creating encryption-unscrambling keys
9-For I =1 … m do/examine m parties, registering $di\ min$ and encryption process
10-Party Si locally registers $di\ min$ and its class esteem Cli as per K closest calculation with respect to question example .
11-Encrypt $di\ min$ as in Eq. (2) to get homormophic encryption $Ei\ di$.
12-RSA encodes Pi to Ci=$Eei\ Pi$ and class mark yi to $Cli = Ee$;
13-Transport $Ei\ d$, Ci , and Cli to EDMS;
14-End For/figuring $di\ min$ and encryption process
15-For I =1 … m do/Decryption process at EDMS
16-Decrypt $Ei\ di\ min$ according to Eq. (3) to get $di\ min$ and Cli to get yi
17-End For/decoding process
18-Construct the mapping table that maps the relative distinction between $di\ min$ with all
$dj\ min\ i \neq j$ and $i, j \in 1, m\ to + 1, -1$
19-Calculate the weight for each column in the mapping table by including the column components and get the aggregate.
20-Determine the worldwide min remove which compares to min weight in the mapping table.
21-Get the anticipated class that match worldwide min remove (min weight in the mapping table).

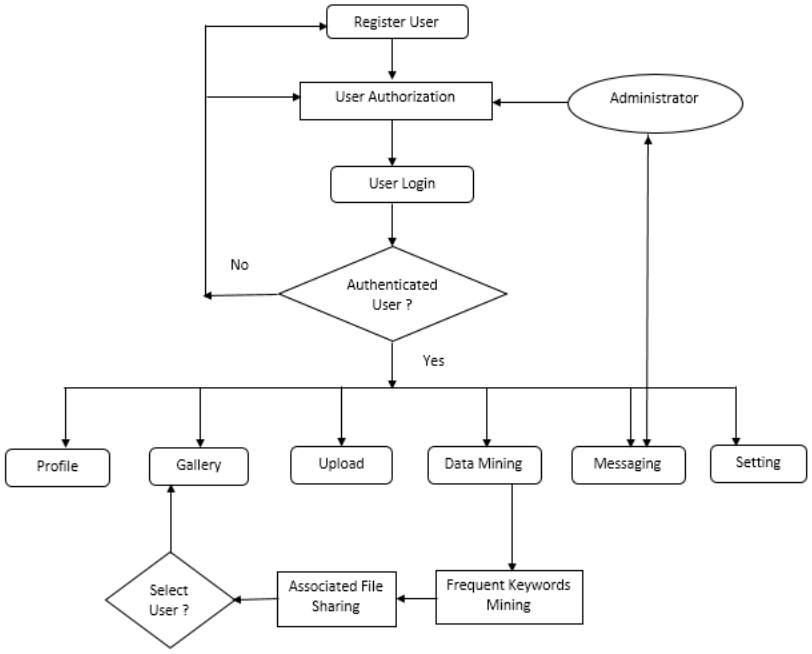

Figure 3: Data Flow Diagram of working System

5 CONCLUSION

In this paper, we attempt to set up the foundation for additionally look into in the territory of Security Preserving Data Mining (PPDM). We set forward institutionalization issues in PPDM. Despite the fact that our work depicted in this paper is primer and reasonable in nature, we contend that it is a crucial essential for institutionalization in PPDM.

Our essential objective in this work is to consider a typical system for PPDM, prominently regarding definitions, standards, arrangements, and necessities. The benefits of a structure of that nature are:

(an) a typical structure will abstain from confounding designers, professionals, and numerous others keen on PPDM;

(b) selection of a typical structure will hinder conflicting endeavors in various ways, and will empower merchants and engineers to make strong advances later on in the PPDM territory.

Our commitments in this paper can be condensed as takes after: 1) we portray the issues we look in characterizing what data is private in information mining, and talk about how protection can be disregarded in information mining; 2) we characterize protection conservation in information mining in view of clients' close to home data what's more, data concerning their aggregate movement; 3) we depict the general parameters for portraying situations in PPDM; 4) we examine the ramifications of the Organization for Economic Collaboration and Development (OECD) information security standards in learning revelation; 5) we recommend a few approaches for PPDM in view of

instruments acknowledged around the world; and 6) we propose a few prerequisites for the advancement of specialized arrangements and to manage the organization of new specialized arrangements..

6 REFERENCES

[1]. T. S. Chen, W. Bin Lee, J. Chen, Y. H. Kao, and P. W. Hou, "Reversible privacy preserving data mining: A combination of difference expansion and privacy preserving," J. Supercomput., vol. 66, no. 2, pp. 907–917, 2013.

[2]. E. Bertino, I. N. Fovino, and L. P. Provenza, "A framework for evaluating privacy preserving data mining algorithms," Data Min. Knowl. Discov., vol. 11, no. 2, pp. 121–154, 2005.

[3]. S. Agrawal, J. R. Haritsa, and B. A. Prakash, "FRAPP: A framework for high-accuracy privacy-preserving mining," Data Min. Knowl. Discov., vol. 18, no. 1, pp. 101–139, 2009.

[4]. Z. Wang, X. Pang, Y. Chen, H. Shao, Q. Wang, L. Wu, H. Chen, and H. Qi, "Privacy-preserving Crowd-sourced Statistical Data Publishing with An Untrusted Server," IEEE Trans. Mob. Comput., vol. PP, no. c, p. 1, 2018.

[5]. X. Liu, R. Deng, K. K. R. Choo, Y. Yang, and H. H. Pang, "Privacy-Preserving Outsourced Calculation Toolkit in the Cloud," IEEE Trans. Dependable Secur. Comput., vol. 5971, no. c, pp. 1–14, 2018.

[6]. S. Qiu, B. Wang, M. Li, J. Liu, and Y. Shi, "Toward Practical Privacy-Preserving Frequent Itemset Mining on Encrypted Cloud Data," IEEE Trans. Cloud Comput., vol. X, no. X, pp. 1–12, 2017.

[7]. A. Kaur, "A Hybrid Approach of Privacy Preserving Data Mining using Suppression and Perturbation Techniques," no. Icimia, pp. 306–311, 2017.

[8]. P. Chahar and S. Dalal, "Deadlock Resolution Techniques : An Overview," International Journal of Scientific and Research Publications, vol. 3, no. 7, pp. 1–5, 2013.

[9]. Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "PPHOPCM: Privacy-preserving High-order Possibilistic c-Means Algorithm for Big Data Clustering with Cloud Computing," IEEE Trans. Big Data, vol. 7790, no. c, pp. 1–1, 2017.

[10]. X. Ding, P. Liu, and H. Jin, "Privacy-Preserving Multi-keyword Top-k Similarity Search Over Encrypted Data," IEEE Trans. Dependable Secur. Comput., vol. 5971, no. c, pp. 1–1, 2017.

[11]. S. Sharma and D. Shukla, "Efficient multi-party privacy preserving data mining for vertically partitioned data," Proc. Int. Conf. Inven. Comput. Technol. ICICT 2016, vol. 2, 2017.

[12]. V. Baby, "Distributed threshold k-means clustering for privacy preserving data mining," pp. 2286–2289, 2016.

[13]. L. Li, R. Lu, K. K. R. Choo, A. Datta, and J. Shao, "Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases," IEEE Trans. Inf. Forensics Secur., vol. 11, no. 8, pp. 1547–1861, 2016.

[14]. P. S. Wang, F. Lai, H. C. Hsiao, and J. L. Wu, "Insider Collusion Attack on Privacy-Preserving Kernel-Based Data Mining Systems," IEEE Access, vol. 4, pp. 2244–2255, 2016.

[15]. S. Khan, T. Gorhe, and R. Vig, "Enabling Multi-level Trust in Privacy Preserving Data Mining," pp. 1369–1372, 2015.

[16]. S. Liu, Q. Qu, L. Chen, and L. M. Ni, "SMC: A Practical Schema for Privacy-Preserved Data Sharing over Distributed Data Streams," IEEE Trans. Big Data, vol. 1, no. 2, pp. 68–81, 2015.

[17]. M. Kantarcioglu, "Incentive-compatible privacy-preserving distributed data mining," Proc. - IEEE 13th Int. Conf. Data Min. Work. ICDMW 2013, p. 859, 2013.

[18]. G. Li and Y. Wang, "Privacy-Preserving Data Mining Based on Sample Selection and Singular Value Decomposition," 2011 Int. Conf. Internet Comput. Inf. Serv., pp. 298–301, 2011.

[19]. Y. Miao, X. Zhang, K. Wu, and J. Su, "An efficient algorithm for privacy preserving maximal frequent itemsets mining," Proc. - 2011 4th Int. Symp. Parallel Archit. Algorithms Program. PAAP 2011, vol. 1, pp. 115–118, 2011.

[20]. U. Rani, S. Dalal, and J. Kumar, "Optimizing performance of fuzzy decision support system with multiple parameter dependency for cloud provider evaluation," International Journal of Engineering & Technology, vol. 7, pp. 166–170, 2018

[21]. L. Xin, "Practical anonymous subscription system with privacy preserving data mining," ICSESS 2011 - Proc. 2011 IEEE 2nd Int. Conf. Softw. Eng. Serv. Sci., pp. 138–141, 2011.

[22]. H. Li, "Study of Privacy Preserving Data Mining," Proc. IEEE, pp. 700–703, 2010.

[23]. M. D. Singh, P. R. Krishna, and A. Saxena, "A privacy preserving Jaccard similarity function for mining encrypted data," IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON, pp. 1–4, 2009.

[24]. Y. Shen, H. Shao, and Y. Li, "Research on the personalized privacy preserving distributed data mining," 2009 2nd Int. Conf. Futur. Inf. Technol. Manag. Eng. FITME 2009, pp. 436–439, 2009.

[25]. Y. Shen, J. Han, and H. Shao, "Research on privacy-preserving technology of data mining," 2009 2nd Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2009, vol. 2, pp. 612–614, 2009.

[26]. J. Zhu, "A new scheme to privacy-preserving collaborative data mining," 5th Int. Conf. Inf. Assur. Secur. IAS 2009, vol. 1, pp. 468–471, 2009.

[27]. K. Murugesan, M. R. Ghalib, J. Gitanjali, J. Indumathi, and D. Manjula, "A pioneering Cryptic Random Projection based approach for privacy preserving data mining," 2009 IEEE Int. Conf. Inf. Reuse Integr., pp. 437–439, 2009.

[28]. J. Wang, Y. Luo, Y. Zhao, and J. Le, "A Survey on Privacy Preserving Data Mining," 2009 First Int. Work. Database Technol. Appl., pp. 111–114, 2009.

[29]. Z. Zhou, L. Huang, and Y. Yun, "Privacy Preserving Attribute Reduction Based on Rough Set," 2009 Second Int. Work. Knowl. Discov. Data Min., pp. 202–206, 2009.

[30]. S. Wu and H. Wang, "Research on the privacy preserving algorithm of association rule mining in centralized database," Proc. - Int. Symp. Inf. Process. ISIP 2008 Int. Pacific Work. Web Min. Web-Based Appl. WMWA 2008, pp. 131–134, 2008.