International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-5, Issue-6 E-ISSN: 2347-2693

A Novel Approach for Pivacy preserving using Animal Migration Optimization and RSA algorithm

Nivedita Bairagi[1*], Punit K. Johari[2]

[1*]Department of CSE and IT, Madhav Institute of Technology and Science, Gwalior
[2]Department of CSE and IT, Madhav Institute of Technology and Science, Gwalior

**Corresponding Author: bairaginivedita@gmail.com*

Available online at: www.ijcseonline.org

Received: 19/May/2017, Revised: 14/May/2017, Accepted: 10/Jun/2017, Published: 30/Jun/2017

***Abstract*—** In recent years, the quantity of data between organizations, companies and governments has been produced and transmitted with extremely increased in number. Privateness preserving is without doubt one of the primary challenges in a computer world, when you consider that of the large amount of sensitive information on the internet. Additionally, with the quick increase of data mining technologies hidden relationships between items in databases can now be exposed with ease, for the reason of decision making or to determine user's preferences. In the existing work, k-anonymity method used for the safety of sensitive data from the leakage or distribution to unauthorized users. However it is not sufficient for the protection of attribute disclosure. This method is also difficult to reverse the data to get the content. To overcome this problem, we performed the Animal Migration Optimization on the basis of age and then encryption is performed using RSA algorithm for achieving the security of the data and preserve from heavy data loss.

***Keywords*—**Privacy Preservation, Data Modification, Privacy preserving techniques , Animal Migration Optimization and RSA Algorithm.

I. INTRODUCTION

Privacy preserving has originated as an significant anxiety on the subject of the success of the data mining. Privacy preserving data mining (PPDM) deals with keeping the privateness of person's data or sensitive capabilities without losing the utility of the information. People have emerge good sense of the privateness intrusions on their personal data and are very unwilling to share their sensitive data. In current years, the area of privateness has realized rapid advances considering of the increases within the capability to store knowledge. In specified, up to date advances within the data mining area have lead about privacy .The purpose of privacy preserving data mining(PPDM) algorithms is to mined proper know-how from huge quantities of information while defending from thoughtful expertise. The main objectives of a PPDM algorithm is:

- A PPDM algorithm should thwart the discovery of wise knowledge.
- It should be proof against the quite a lot of data mining tactics.
- It will have to no longer compromise the access and the use of non-sensitive information.
- It must not have an exponential computational complexity [1].

Privacy preserving is become increasingly a significant concern for future growth of data mining tactics with best potential access to datasets having sensitive, or personal information. The essential concern for current information mining algorithms is extracting correct information mining results at the same time nonetheless preserving privateness of datasets. As a result of the growing trouble on privateness, a new category of data mining called Privacy Preserving data Mining (PPDM) has been offered. However, the privacy preserving data mining has become an essential concern in contemporary years considering of the huge quantity of exclusive knowledge which is tracked by way of a couple of industry applications. In many instances, the users are unwilling to provide personal data except the privateness of sensitive information is guaranteed. PPDM was once first introduced by Agrawal and Srikant in 2000 [2]. PPDM algorithms are developed by using integrating privateness protection mechanism to hide sensitive information earlier before executing data mining algorithms. Then a couple of exceptional branches with extraordinary goals were developed. Privacy retaining classification methods clog a miner from making a classifier which is capable of forecasting the private information. The major consideration in privacy preserving data mining is the sensitive nature of raw data. The data miner, concurrently mining for complete statistical data about the information, should not be capable

to access information in its normal kind with all the sensitive knowledge. This necessitates more mighty systems in privacy preserving data mining that deliberately alter the data to conceal sensitive expertise also preserve the inherent records of the data which is vital for mining reason. Most of the privacy preserving data mining strategies practice a change which reduces the effectivity of the customary information when it's utilized to information mining approaches or algorithms. Also, there's an average trade-off between privacy and accuracy, however this trade-off is undergo by some specified algorithm which is employed for privacy-protection. Hence, the important concern is to maintain maximum effectiveness of the information by means of satisfying the basic privacy constraints.

The rest of the paper is organized as Taxonomy of Privacy Preserving techniques in section II, related work in section III, proposed methodology in section IV, dataset description in section V, result analysis in section VI, and conclusion in section VII.

II. PRIVACY PRESERVING METHODOLOGY

There are many methodologies which have been accepted for privacy preserving data mining. We can categorize them based on the following measurements [3]:

A. Data Distribution

The first phase elaborates about the distribution of data. Some approaches have been developed for centralized data, while others use distributed data. Distributed data can also be categorized as horizontal distribution and vertical distribution.

B. Data Modification

The second phase elaborates about the data modification scheme. On the whole, data modification is used with a purpose to modify the original data of a database that wants to be released to the public and in this solution to be certain high privacy. It's primary that an information change process must be in concert with the privacy policy adopted by way of an organization.

C. Data Mining Algorithm

The third phase elaborates about the data mining algorithm, for which the data change is taking place. This is really whatever that isn't known earlier, however it facilitates the analysis and design of the data hiding algorithm. For the time being, quite a lot of data mining algorithms had been viewed in isolation of each other. Amongst them, the important suggestions had been developed for classification data mining algorithms, like decision tree, association rule mining, clustering algorithms etc.

D. Data or Rule hiding

The fourth phase elaborates to whether raw information or aggregated information should be hidden. The complexity for hidden aggregated data in the type of rules is more, and thus, heuristics had been developed.

E. Privacy Preservation

The last phase which is the major, discusses about the privacy preserving technique used for the selective changes in the data. Selective changes is required as a way to acquire better utility for the modified data considering the privacy will not be imperiled. The methods which were applied hence are:

- Heuristic-based techniques like adaptive modification that changes only chosen values that minimize the utility loss.
- Cryptography-based techniques like secure multiparty computation where a computation is safe, if on the finish of the computation, no party know anything except its own input and the results [4].
- Reconstruction-based techniques where the randomized data is used to reconstruct the original distributed data.

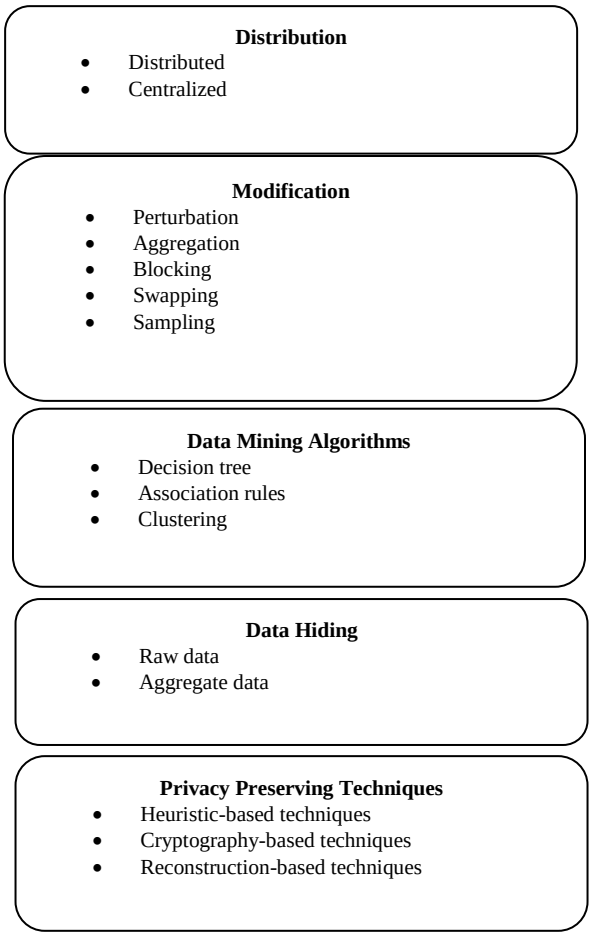

Figure 2 Privacy Preserving Data Mining Classification

III. RELATED WORK

In a general way privacy security mechanism makes use of the generalization and suppression of the sensitive information. It clogs the privacy disclosure of the sensitive information. The privacy security mechanism defers the identity and attributes disclosure. The privateness is obtained through the higher accuracy and consistency of the user information, i.e., the accuracy of the personal information. R. Rajeswari et al. Proposes a privacy persevered access control mechanism for data streams. For the privateness security mechanism it makes use of the combination of both the k-anonymity method and fragmentation procedure [5].

The conventional publishing approaches will cast off the sensitive attributes and generate the plentiful records to attain the intention of privacy protection. Within the enormous data atmosphere, the requirement of utilizing information become more and more diverse, which is past the scope of the conventional approach. Tong Li et al. present a cryptographic data publishing system that preserves the data integrity and attains anonymity without deletion of any attribute or utilization of redundancy [6].

Data anonymization is a promising system within the discipline of privacy preserving data mining used to defend the information in opposition to identity disclosure. Data loss and normal attacks viable on the anonymized data are serious challenges of anonymization. Hence J. Jesu Vedha Nayahi et al. proposed, an anonymization algorithm established on clustering and flexible to similarity attack and probabilistic inference attack [7].

In digital domain, information is kept in digital format. This format of information, consumes less effort and memory. Thus a number of organization and institutes are maintaining their information in this format. Surbhi Sharma et al. present how the one of a kind departments of same group mix their knowledge without harming the confidentiality of the client for making effective selections in efficient and correct method. Thus the process vertically information combination, cryptography and decision mining is established [8].

The sensitive information or private data is an important supply of information for the groups like government and non-governmental group for study and allocation of public money, medical research and pattern evaluation. The main crisis right here is publishing knowledge without revealing the sensitive understanding of members. This sensitive or personal information of any individual is most important to a number of knowledge repositories like scientific knowledge, census data, voter registration data, social network information and client information. M. Prakash et al. Presents a personalized anonymization method which preserves the privateness at the same time the sensitive data is released [9].

Transformation systems had been proposed for transforming numerical and express sensitive attributes. Many algorithms exist within the literature to transform sensitive attributes of numeric information type. But there is no technique for dealing with sensitive attributes of Boolean data, that Boolean attributes do not reveal any private information without compromising data mining outcome. Rupinder Kaur et al. Goal is to present a system for transforming sensitive Boolean attributes [10].

IV. PROPOSED METHEDOLOGY

The k-anonymity [11] is one of the privacy preserving data mining methods which assists us in releasing enormous quantity of data so that it can be used for business or research associated work by a variety of organizations to make sure that privacy of no entity is being put in danger against inference and linking attacks. It mainly used for the protection of data from the disclosure of identity.

In the existing work, personalized anonymization method is used for the safety of sensitive information from the leakage or distribution to unauthorized users. But it is not enough for the protection of attribute disclosure. This method is also difficult to reverse the data to get the content.

In our proposed work, we apply Animal Migration Optimization (AMO) [12] algorithm for partitioning the data in very efficient way. In animal migration optimization, there are two procedures: migration procedure and population updating procedure. In the first procedure, the algorithm simulates how the groups of animal transfer from present position to the new position. In the last procedure, the algorithm simulates how some animals go away from the group and some become member for the period of the migration. Then we applied RSA algorithm for encrypting and decrypting the data.

Proposed Algorithm:

Step:1 Initially fetch the dataset.
Step:2 Partition the dataset by random number.
Step:3 Apply Animal Migration Optimization on each partition.
3.1 Initialize position of each record.
3.2 Evaluate overall fitness and fitness value of each record by fitness function.
Fitness function =sum (x*x,2)
3.3 While stopping criteria not met
3.4 Compare fitness value of record with overall fitness
If (fitness value>=overall fitness)

Update the position

$X_{i,G+1} = X_{i,G} + \delta . (X_{neighborhood,G} - X_{i,G})$

Otherwise

Continue

3.5 If (position $_{new}$ >= position $_{old}$)

Then set updated position as global position

3.6 End

Step:4 Now apply RSA algorithm on the above output

Step:5 Encrypt the dataset

Step:6 End

V. DATA SET DESCRIPTION

The data set used here in the evaluations is the 'Adult' data set which is freely available on UCI Machine Learning Repository, composed of data gathered from the census of USA [13]. The original data of an individual is shown in Table1. From table 1, which is an analogy of original data you may clearly and conveniently determine who's having a different disease. For illustration who are all having Cancer or Flu or Heart disease sickness can be recognized without difficulty that are sensitive attribute. This Microdata is a valuable source of expertise for the scientific study and allocation of public money and trend evaluation.

Table 1 Adult Data

S. No.	UID	Age	Disease
1	55612	29	Cancer
2	55675	21	Flu
3	55627	25	Heart Disease
4	55646	43	Heart Disease
5	55672	48	Flu
6	55655	47	Cancer
7	55647	34	Heart Disease
8	55622	30	Flu
9	55634	36	Cancer
10	55685	55	Flu
11	55681	58	Flu
12	55694	72	Cancer
13	55698	65	Heart Disease
14	55688	59	Heart Disease
15	55690	65	Heart Disease

VI. RESULT ANALYSIS

In this part, the experimental results are done on MATLAB tool. Here, table 1 shows the original adult data which comprises three attributes UID, age and sentient information. In the proposed methodology, the partitioning of data is done over 'age' attribute using animal migration optimization. The partition made by AMO algorithm is shown in table 2, which provide better mined results as compared to earlier used methodology.

Table 2 Data Partitioning with AMO

Command Window

SNO.	Cluster Number	UID	AGE
1	1	55675	21
2	1	55627	25
3	1	55612	29
4	2	55622	30
5	2	55647	34
6	2	55634	36
7	3	55646	43
8	3	55655	47
9	3	55672	48
10	4	55685	55
11	4	55681	58
12	4	55688	59
13	4	55698	65
14	4	55690	65

The table 3 shows the final outcome of the original dataset, Here, RSA algorithm is used over 'UID' attribute on AMO algorithm based partition, which is more secure and accurate as compared to previous methodology.

Table 3 Encrypted adult data with proposed methodology

Command Window

Sl. No.	Non Sensitive UID	Age	Sensitive Information
1	9292	<=30	Flu
2	9292	<=30	Heart Disease
3	9292	<=30	Cancer
7	9292	>=40	Heart Disease
8	9292	>=40	Cancer
9	9292	>=40	Flu
4	9292	<=40	Flu
5	9292	<=40	Heart Disease
6	9292	<=40	Cancer
10	9292	<=60	Flu
11	9292	<=60	Flu
12	9292	<=60	Heart Disease
13	9292	<=60	Heart Disease
14	9292	<=60	Heart Disease

Table 4 Decrypted adult data with proposed methodology

Command Window

Sl. No.	Non Sensitive UID	Age	Sensitive Information
1	55675	21	Flu
2	55627	25	Heart Disease
3	55612	29	Cancer
7	55622	43	Heart Disease
8	55647	47	Cancer
9	55634	48	Flu
4	55646	30	Flu
5	55655	34	Heart Disease
6	55672	36	Cancer
10	55685	55	Flu
11	55681	58	Flu
12	55688	59	Heart Disease
13	55698	65	Heart Disease
14	55690	65	Heart Disease

Table 4 shows the decrypted value means original value of the data set.

The proposed approach contains three main advantages:

- It protects sensitive data with very low information loss.
- Data can also be reconstructed.
- Data is more accurate than traditional methodology.

VII. CONCLUSION

In this paper, we present a novel approach of privacy preserving data mining using animal migration optimization and RSA algorithm .Using AMO algorithm we get the better partitions of the data. One of the major advantage of using AMO algorithm is we are getting less information loss than traditional approach. The second advantage is, here, RSA algorithm is used, which is most secured cryptographic algorithm. The use of RSA algorithm made data reversal problem identical easy, to resolve by encryption and decryption key. The accurateness of data is also achieved so far higher than traditional approach. Thus the proposed system achieves high amount of data utility, minimum information loss and high amount of accuracy.

REFERENCES

[1] M. Patel, A. Hasan , S.Kumar, "*A Survey: Preventing Discovering Association Rules for Large Data Base*", International Journal of Scientific Research in Computer Science and Engineering, Vol.1, Issue.2, pp.30-32, 2013.

[2] Bullarao Domathoti, Rajia Begum and Nageswara Rao.P, "*Privacy Preserving Collaborative Auditing Data Storage Scheme in Cloud Computing*", International Journal of Computer Sciences and Engineering, Vol.3, Issue.5, pp.212-218, 2015.

[3] Md. Riyazuddin, Dr.V.V.S.S.S.Balaram, Md.Afroze, Md.JaffarSadiq, M.D.Zuber, "An Empirical Study on Privacy Preserving Data Mining", ISSN: 2231-5381, Page 687, Volume3Issue6- 2012.

[4] Y. Lindell and B. Pinkas, "Privacy preserving data mining", J. Cryptology, 15(3):177–206, 2002.

[5] R. Rajeswari and Mrs R. Kavitha ,"Privacy Preserving Mechanism for anonymizing data streams in data mining", International conference on current research in Engineering Science and Technology(ICCREST-2016).

[6] Tong Li, Zheli Liu, Zin Li, Chunfu Jia and Kuan-Ching Li, "A Cryptographic Data Publishing System", J. Comput. Syst. Sci. (2016), http://dx.doi.org/10.1016/j.jcss.2016.12.004.

[7] J. Jesu Vedha Nayahi and V. Kavitha," *Privacy and utility preserving data clustering for data anonymization and distribution on Hadoop*", Future Generation Computer Systems, 0167-739X/© 2016 Elsevier.

[8] Surbhi Sharma and Deepak Shukla, "*Efficient multi-party privacy preserving data mining for vertically partitioned data*", Inventive Computation Technologies (ICICT), 10.1109/INVENTIVE.2016.7824852, © 2017 IEEE.

[9] Prakash, M., and G. Singaravel. "*An approach for prevention of privacy breach and information leakage in sensitive data mining*." *Computers & Electrical Engineering* 45 (2015): 134-140.

[10] V.K. Saxena, S. Pushkar, "*Privacy Preserving using Encryption Proxy in Data Security*", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.2, pp.36-41, 2017.

[11] L. Sweeney,"*An Achieving k-Anonymity Privacy Protection using Generalization and Suppression*", International Journal of Uncertainty, Fuzziness and Knowledge-Based System, 2002, pp571-588.

[12] Xiangtao Li, Jie Zhang and Minghao Yin, "*Animal migration optimization: an optimization algorithm inspired by animal migration behavior*," Neural Comput & Applic (2014) 24:1867–1877.

[13] https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data

