Privacy Preservation in Data Mining: Algorithmic Analysis on Opposition Intensity-based Cuckoo Search Algorithm by varying Scaling Factor

[1]G.K.Shailaja and [2]Dr.C.V.Guru Rao
[1]Associate Professor
Dept. of IT, Kakatiya Institute of Technology and Science-Warangal,
Telangana, India-506015
gkshailaja68@gmail.com
[2]Professor and Director of Evaluation,
Dept. of CSE, S.R.Engineering College, Warangal, Telangana, India-506371

Abstract— Privacy Preserving Data Mining (PPDM) is a significant aspect of data preservation without losing the privacy of the individuals. Thus, PPDM has become a most important area of research and a number of methods and techniques are being developed concerning the hiding of sensitive information. The most of the conventional research on this topic deals with the transformation of the original data into a different form and here the data guarantee is not addressed. Thus, this paper intents to develop novel privacy preservation model for data mining by following three major phases (a) Data sanitization (v) key generation (c) Restoration. In the data sanitization process, the sensitive fields of data are chosen and they are hidden using the optimal key generated in the key generation phase. The hidden message is transferred from the source to the destination and in the receiver side, the restoration process take place with the same key. The major novelty of this model lies in the optimal key selection and here the optimal key is selected using the Opposition Intensity-based Cuckoo Search Algorithm (OI-CSA), which is the extended version of Cuckoo Search algorithm (CS). Finally, the proposed model is evaluated in terms of analyzing scaling factor β using four datasets namely T10, Chess, Retail, and T40 for four research issues such as Hiding Failure (HF) rate, Information Preservation (IP) Rate, and False Rule generation (FR) and Degree of Modification (DM).

Keywords—PPDM; Data Sanitization;Data Preservation;Key Extraction;Restoration; OI-CSA;Scaling Factor.

I. INTRODUCTION

The digital revolution of data has made the digitized information easier to capture, process, store and distribute. In the recent days, the electronic data is increasing drastically in the digitization of modern world [1] [2]. Data mining plays a crucial role in extracting knowledge from a huge volume of data and it is the core process in the knowledge discovery of databases. In different fields like the weather forecasting, medical diagnosis, marketing, customer relationship management etc. data mining is being greatly employed over the decades. In the whole process of data mining, there is sensitive individual information (medical and financial information) that needs to be hidden from the attackers[3] [4] [5]. The mined information can in the form of patterns, rules, clusters or classification models. The analysis of individual's socio-economic trends is essential for society. If this considers the data disclosure, then the privacy concern is a necessary term. In data mining, it is essential to maintain the ratio between privacy protection and knowledge discovery. The main intension of this PPDM is to preserve personal information under the data mining techniques [6] [7] [8].

Privacy Preserving Data Mining (PPDM) is the major research area concerned with the protection of the privacy of individual data without sacrificing the utility of the data. In PPDM, before performing the data distribution tasks, there is a necessity to employ various methods to preserve the privacy of the data [9] [10] [11]. There are various PPDM techniques that are categorized based on the centralized and distributed forms of data mining as Anonymization based, randomized response, cryptography based, perturbation based and condensation based approaches. The horizontally partitioned data and the vertically partitioned data are the types of the distributed database scenario [12] [13]. In the Horizontally partitioned data, the data base is split into numerous non-overlapping horizontal partitions, whereas in the vertically partitioned data, the data sets are split based on the attributes and the count of the transaction. The data sanitization is the process of hiding the sensitive information of the document to wider audience. In the sensitization process, the frequent item set in the document is hidden with the rule hiding algorithm by eliminating its confidence or support.

In the Anonymization based PPDM, the sensitive person specific information are identified and hidden, hence the retaining of the sensitive data becomes easier. This model also ensures that the transformed data is true and it is not immune to homogeneity attack [14] [15]. Apart from this, it suffers from the drawbacks like heavy information loss. In the statistical disclosure control, the Perturbation based PPDM approach is being used as it is simple, most efficient and has the ability of reserving the statistical information. In this process, the original sensitive values are transformed into synthetic data values and it becomes complex for the attackers to extract the sensitive knowledge of the data from the available data [16]. This technique suffers from the drawbacks like loss of information and lack of trusted regeneration of data. The Randomized Response based PPDM twists the sensitive data of the users in such a way that it becomes un-accessed by attackers [17]. This model is only suitable for smaller database and not for multiple attribute databases. Various conventional models have existed for the privacy preservation data mining. Still, the data security is in the infant stage that needs further improvement in the future. Therefore, during the design of a PPDM technique, it is essential to .override the major challenges like information loss, accuracy, recover original data after hiding, cost, consistency and Security, data quality [18] [19] [20]. In the hiding strategy, the accuracy is directly related to the information loss. Moreover, the Consistency tells about the degree of missed data. The tendency to enhance the data sizes are represented by scalability. The term security tells about the degree of protection of data against damage, loss.

The major contribution of this research work is highlighted below:

- The proposed novel privacy preservation in data mining model is constructed by following three major phases (a) data sanitization (b) Key generation and (c) data restoration.
- The data sanitization is the initial process and here the association rules are extracted from original database. Once, the sensitive fields are chosen, they are preserved using the optimal key.
- The optimal key generation is the major challenge and it is achieved here with the help of OI-CSA. Then, the preserved data with key is transferred from the sender to the receiver.
- In the receiver side, the data restoration takes place with the same key. Finally, the performance of the proposed model by varying the scaling factor β for different data sets like T10, Chess, Retail, and T40 are evaluated with respect to Hiding Failure (HF) rate, Information Preservation (IP) Rate, and False Rule generation (FR) and Degree of Modification (DM).

The rest of the paper is organized as: Section II portrays the literature works concerned with the privacy preservation in data mining and Section III depicts the architecture of the proposed privacy-preserving data mining. Section IV explains the steps involved in data preservation and Section V specifies the objectives and key encoding of data sanitization and restoration. Section VI discusses the results acquired and Section VII provides a strong conclusion to this research.

II. Literature Review

A. *Related Works*

In 2018, Menaga and Revathi [21] proposed an innovative Privacy Preserving Data Mining (PPDM) using the least lion optimization algorithm (LLOA). In the proposed model, the sanitization process is accomplished in two stages viz. rule mining and secret key generation. For the input database, the association rules were mined using the whale optimization algorithm and on the basis of the fitness function, the rules were validated. The secrete key was generated using the lion optimization algorithm (LOA) with least mean square (LMS). Further, LLOA converted the original database into the sanitized database with the aid of the secrete key.

In 2015, Xu *et al.* [22] proffered Bit Vector-based Efficient MASK (BV-EMASK) for privacy preserving frequent pattern mining. The bit vector is formed by distorting the original database before mining. These data sets were horizontally partitioned into subsets for solving the scalable problem.

In 2015, Zhu and Li [23] formulated hybrid partial hiding algorithm (HPH) with the intention of enhancing the privacy preservation associated with the association rule mining. The raw data is hidden and transformed using the data perturbation algorithm by following two major strategies like data perturbation and query restriction.

In 2017, Rehman and Sharma [24] developed Improved Apriori algorithm with the objective of reducing the time consumed for generating the frequent item sets. The minimum support count was generated subsequent to the identification of the frequent items in the available database. On the basis of the support count, the association rules were generated. Finally, the proposed model was compared with the existing model in terms of the count of iterations and the time taken by iterations to create the sanitized data.

In 2017, Kalyani *et al.* [25] formulated a novel data distortion approach in order to address the issues related to the classification rule hiding. The privacy requirements were employed only to centralized data before executing the classification algorithm. Further, the knowledge hiding approach is used in order to protect the sensitive knowledge from leaking out before sharing the data with others.

III. Architecture of the Proposed Privacy-Preserving Data Mining

A. *Formulated Architecture*

The architecture of the proposed PPDM scheme using Opposition Intensity-based Cuckoo Search Algorithm (OI-CSA) is illustrated in Fig. 1. The proposed architecture encloses two major phase viz. (a) data sanitization and (b) data restoration. Initially, the sanitation process takes place in which a key is selected to preserve the sensitive data in a much protective way. The major challenge of key selection is to choose the most optimal key for hiding the sensitive data effectively that can override the challenges like loss of information, unauthorized access to confidential data etc. The sanitized data is send to the receiver via the communication channel and on the receiver side, the data restoration process take place. In the data restoration phase, the authorized person retrieves the data using the same key.

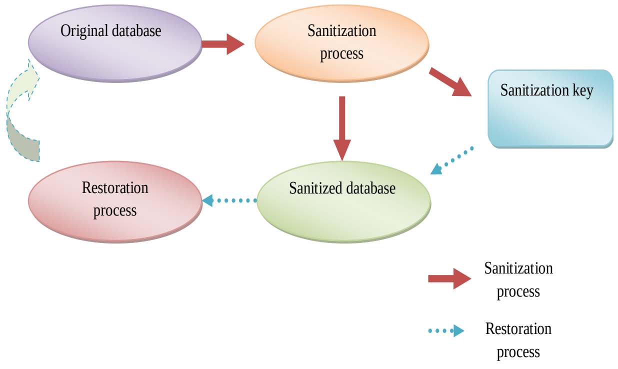


Fig. 1. Framework of the proposed PPDM model

IV. Steps Involved in Data Preservation

A. Sanitization Process

In the sanitization process, the snipped key matrix and the original database O as well as A_2 are binarized and these binarized resultant key matrix is repeatedly fed as input to the rule hiding process. In XOR function of the rule hiding process, the binarized form of O as well as the identical matrix dimensions is XOR-ed and added up with one (1) to generate the sanitized database O'. The mathematical formula for the sanitized database is shown in Eq. (1). The sanitized database O' acquired from sanitization process obtains the sensitive rules SRs as well as association rules subsequent to the sanitization of association rules obtained from sanitized database $(B^{'})$. In order to achieve the objective function, the original database O extracts the relative association rules preceding to the sanitization of association rule B. Fig.2 represents the schematic diagram of the proposed sanitization process.

$$O^{'} = (A_2 \oplus O) + 1 \qquad (1)$$

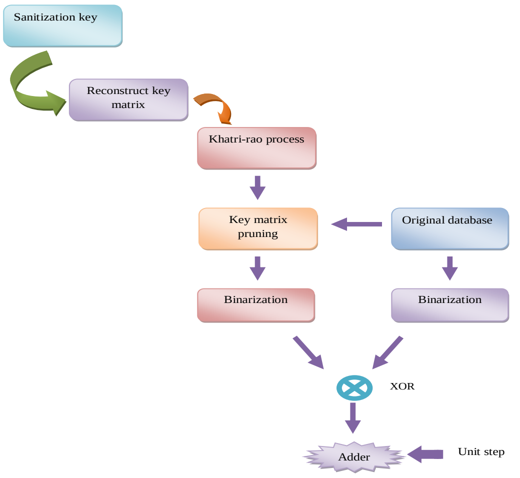


Fig. 2. Sanitization process of the implemented data preservation algorithm

B. Key Generation

The solution transformation process takes place in the key generation process. The key is represented as A and it is converted by the khatri-rao product. In the initial process of solution transformation, the key A is restructured into A_1 having the matrix dimensions as $\left[\sqrt{M_O^{''}} \times O_{\max}\right]$. The highest transaction length is depicted as $O_{\max}$ and the count of the transactions is denoted as M_O. In addition, the term $M_O^{''}$ denotes the nearest highest perfect square of M_O.

For instance, the key value of $A = \{1,2,1\}$ and in the restructured process A is reconstructed to form key matrix A_1 with the dimension $\left[\sqrt{M_O^{''}} \times O_{\max}\right]$. The reconstructed key matrix A_1 is depicted in Eq. (2).

$$A_1 = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix}_{\left[\sqrt{M_O^{''}} \times O_{\max}\right]} \qquad (2)$$

The key matrix A_2 is constructed by two identically restructured A_1 matrixes using Khatri-rao product. It can be mathematically represented $A_2 = A_1 \otimes A_1$ and here the kronecker product is indicated by $\otimes$. The key matrix A_2 has the dimension as $\left[\sqrt{M_O} \times O_{\max}\right]$ and it is further reduced reduced in terms of dimension size of original database. The key generation process is accomplished on the basis of Khatri-rao product and here the generated matrix takes the dimensions similar to original database O and produces $A_2\left[\sqrt{M_O} \times O_{\max}\right]$. Then, by means of hiding the sensitive rules, the rule hiding process is accomplished with the intention of achieving the sanitized database O'. Further, the sanitized database O' is send via the suitable communication channel to reach the receiver.

C. Restoration process

The restoration process takes place in the receiver side, in which the binarization of sanitized database O' as well as generated key matrix A_2 takes place. Then, from the binarization block, the binarized S_d is minimized from unit step unit by the subtraction. The restored database is obtained by performing the XOR function on the subtracted binarized database and key matrix. Further, Eq. (1), Eq. (2), Eq. (4) and proposed OI-CSA update are exploited to form the sanitizing key A_2. Then, the sanitized database O' by generated and the the lossless restoring take place as per Eq. (3). The restored data is denoted using the term $\hat{O}$. Fig. 3 depicts the architecture of the restoration process.

$$\hat{O} = \left(O^{'} - 1\right) \oplus A_2 \qquad (3)$$

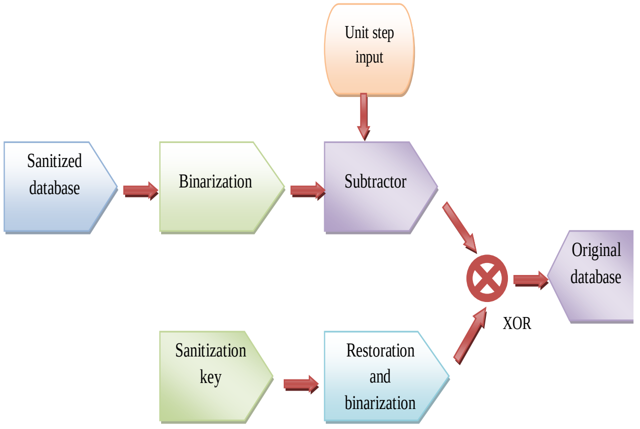


Fig. 3. Restoration process of the proposed data preservation algorithm

V. Objectives and Key Encoding of Data Sanitization and Restoration

A. Objective Function

The foremost objective of this research is to attain the objective function for preserving data as specified in Eq. (4).

$$\min F = \max(F_1, F_2, F_3, F_4) \tag{4}$$

The term F_1, F_2 F_3 and F_4 denotes the normalized HF rate, normalized MD rate, normalized IP rate and the normalized FR rate, respectively. The mathematical formula for F_1, F_2 F_3 and F_4 are shown in Eq. (5), Eq. (7), Eq. (9) and Eq. (11), respectively.

HF rate: In Eq. (5) the HF rate is denoted as f_1 and the worst f_1 of all iterations is denoted as $\max(f_1)$. The fraction of the sensitive rules denoted in sanitized database O' is shown in Eq. (6). In O', the count of the sensitive rules is denoted as $f_1 = \left|B' \cap SRs\right|$.

$$F_1 = \frac{f_1}{\max(f_1)\forall iterations} \tag{5}$$

$$f_1 = \frac{\left|B' \cap SRs\right|}{\left|SRs\right|} \tag{6}$$

IP rate: The notation f_2 in Eq. (7) denotes IP rate and it is "the rate of non-sensitive rules which are concealed in O'". Ip is rate is nothing but the reciprocal of information loss and its mathematical formula is depicted in Eq. (8).

$$F_2 = \frac{f_2}{\max(f_2)\forall iterations} \tag{7}$$

$$f_2 = 1 - \frac{\left|B - B'\right|}{\left|B\right|} \tag{8}$$

FR: f_3 in Eq. (9) denotes FR and it is the "the rate of artificial rules produced in O'". The mathematical formula for f_3 is shown in Eq. (10).

$$F_3 = \frac{f_3}{\max(f_3)\forall iterations} \tag{9}$$

$$f_3 = \frac{\left|B - B'\right|}{\left|B'\right|} \tag{10}$$

DM: f_4 in Eq. (11) depicts DM rate and it is the "count of modifications carried out in O' from O". The mathematical formula for f_3 is shown in Eq. (12), in which the Euclidean distance between the original database O and sanitized database O' is represented as $dist$.

$$F_4 = \frac{f_4}{\max(f_4)\forall iterations} \tag{11}$$

$$f_4 = dist(O, O') \tag{12}$$

B. Key Encoding

The encoding of the keys (chromosome) A in the sanitization process takes place with the help of the proposed OI-CSA algorithm. Then, from the keys generated from A_1 to A_M, the optimal key is selected by the proposed OI-CSA algorithm. Fig. 4 depicts the key encoding process and the length of the key (chromosome) is allocated as $\sqrt{M_O''}$.

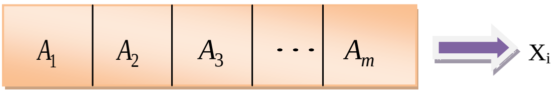


Fig. 4. Key Encoding

C. Proposed OI-CS Algorithm

The conventional CS algorithm has the advantages like high robustness and enhanced global search ability, but, suffers from the drawbacks like slow convergence and easily falling into local optimum. Thus to override these challenges the existing CS algorithm is enhanced by means of modifying the opposition intensity γ as per Eq. (13). The notation $X_i^{(w)}$ and X_i^t depicts the worst solution and the current solution. The values of opposition intensity denoted as γ ranges from 0 to 1. The subsequent position and the positive step size scaling factor are indicated as X_i^{t+1} and β, respectively. The parameter s denotes the step size. $N(s,\tau)$ Indicates levy distribution exploited to describe the step size of arbitrary walk.

The pseudo code of the proposed OI-CSA algorithm is shown in Algorithm 1.

$$X_i^{t+1} = X_i^t + \beta N(s,\tau) - \gamma\left[X_i^{(w)} - X_i^t\right] \tag{13}$$

Algorithm 1: Pseudo code of OI-CSA

```
Objective function, $f(X) = X = (X_1, X_2 ... X_D)^T$
Generate initial population of $n$ host nests $Xi(i = 1,2....n)$
While $(t < MaxGeneration)$ or terminate the process
Determine a cuckoo (assume $i$ ) arbitrarily by Lévy distribution;
quality/fitness is computed ; $F_i$
Choose a nest among $n$ (assume $j$ ) arbitrarily;
```

```
quality/fitness are computed; $F_j$
If $(F_i > F_j)$
Solution update by Eq. (13);
else
$\gamma$ value is determined
Update the solution using Eq. (15)
End if
Building new nest at new location
best solutions are maintained;
The best solution is determined by ordering the solutions;
End while
Terminate
```

VI. Results and discussions

A. Simulation Procedure

The proposed OI-CSA method based privacy preservation of sensitive data is implemented in JAVA and the resultant of each of the analysis is recorded. The enhancement in the performance of the proposed model is validated by varying the scaling factor (β value). The experimentation was performed by means of four datasets namely, T10, Chess, Retail, and T40. The value of β is varied from 0.2, 0.4, 0.8 and 1.0 for four datasets with respect to normalized HF rate F_1, normalized MD rate F_2 , normalized IP rate F_3 and normalized FR rate F_4 is observed.

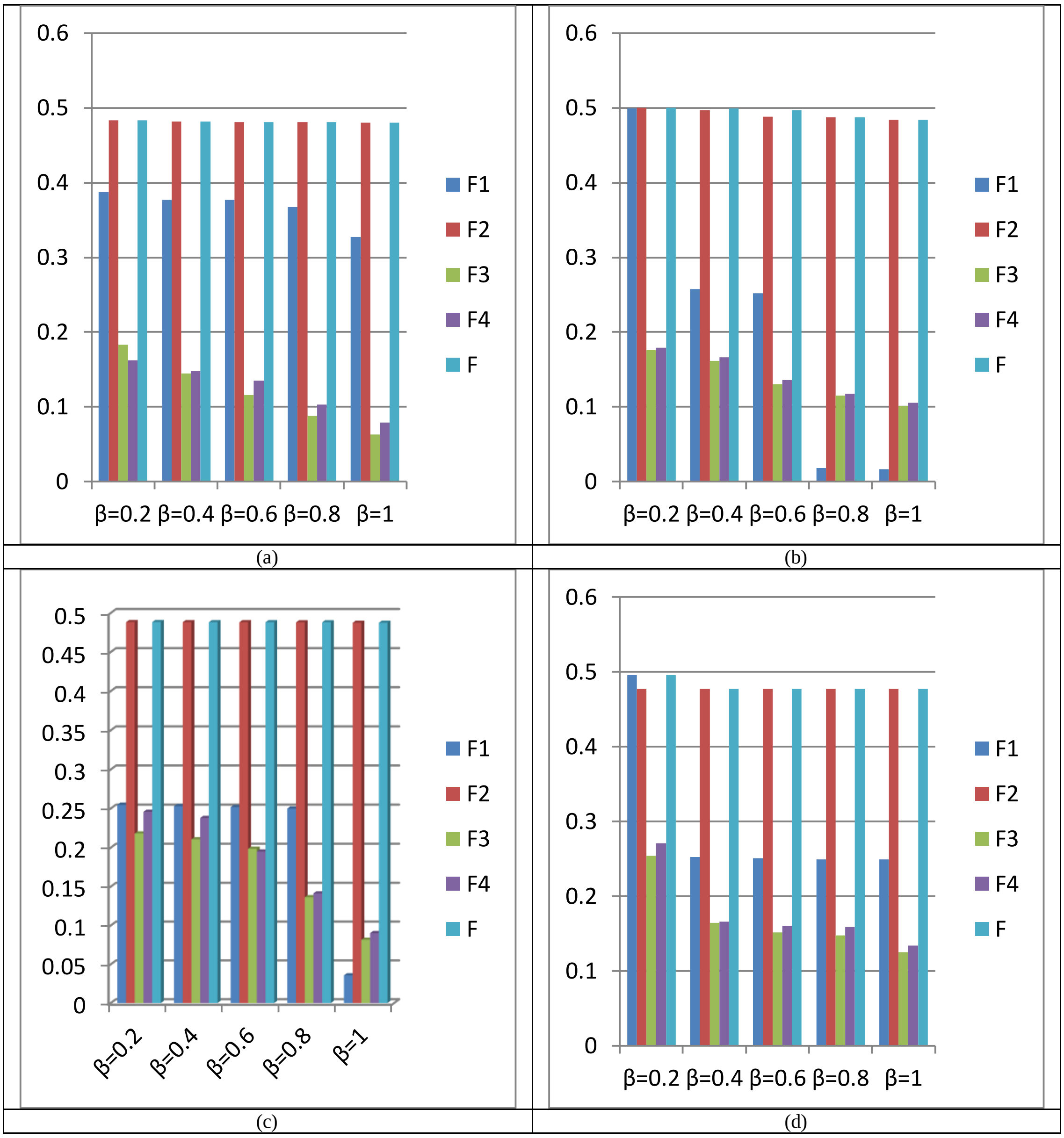
B. Effect on varying β for different databases

In Fig. 4(a), the effect of varying β from 0.2, 0.4, 0.6, 0.8 and 1.0 for retail database for F_1, F_2 F_3 and F_4 is given. The lowest value of F_1 is recorded in $\beta = 1.0$ and the value of F_1 at $\beta = 0.2$, $\beta = 0.4$ $\beta = 0.6$ and $\beta = 0.8$ is 0.387, 0.37725, 0.3765, 0.3675, respectively. The lowest value of F_2 is recorded in $\beta = 1.0$ and the value of F_2 at $\beta = 0.2$ is 0.483438, $\beta = 0.4$ is 0.481496, $\beta = 0.6$ is 0.480873 and $\beta = 0.8$ is 0.480821. The lowest value of F_3 is found in $\beta = 1.0$ and it is 65.5%, 56.4%, 45.5% and 28.17% better than $\beta = 0.2$, $\beta = 0.4$ $\beta = 0.6$ and $\beta = 0.8$, respectively. In terms of F_4 , the lowest value is recorded in $\beta = 1.0$ and it is 51.3%, 46.7%, 46.4% and 23.4% better than $\beta = 0.2$, $\beta = 0.4$ $\beta = 0.6$ and $\beta = 0.8$, respectively.

The value of F_1, F_2 F_3 and F_4 obtained by varying β is shown in Fig. 4(b) concerning T40 database. In which, F_1 attains the minimal value in $\beta = 1.0$ and the values recorded by F_1 at $\beta = 0.2$ is 0.4995, $\beta = 0.4$ is 0.25725, $\beta = 0.6$ is 0.252 and $\beta = 0.8$ is 0.017882. The value of F_2 is lowest in $\beta = 1.0$ and the corresponding value is 0.487322 and the other values at various β values are 0.48818, 0.488084, 0.488012 and 0.487916 at $\beta = 0.2$, $\beta = 0.4$ $\beta = 0.6$ and $\beta = 0.8$, respectively. The least value of F_3 is attained when $\beta = 1.0$ and the corresponding value is 0.081602, the other values are 0.218041, 0.210655, 0.198636 and 0.136411 for $\beta = 0.2$, $\beta = 0.4$ $\beta = 0.6$ and $\beta = 0.8$, respectively. The value of F_4 attains the minimal value of 0.090111 when $\beta = 1.0$ and it is 41.2%, 36.6%, 22.4% and 10.13% better than when $\beta = 0.2$, $\beta = 0.4$ $\beta = 0.6$ and $\beta = 0.8$, respectively.

Fig. 4(c) represents the effect of varying β from 0.2, 0.4, 0.6, 0.8 and 1.0 for T40 database for F_1 , F_2 F_3 and F_4 . The value of F_1 is lowest for $\beta = 1.0$ and it is 85.92% better than $\beta = 0.2$, 85.8% better than $\beta = 0.4$, 85.8% better than $\beta = 0.6$ and 85.7% better than $\beta = 0.8$. The lowest value of F_2 is recorded in $\beta = 1.0$ and the value of F_2 at $\beta = 0.2$ is 0.48818, $\beta = 0.4$ is 0.488084, $\beta = 0.6$ is0.488012 and $\beta = 0.8$ is 0.487916. The value of F_3 records the lowest value in $\beta = 1.0$ and it is 62.5%, 61.2%, 58.9% and 40.17% better than $\beta = 0.2$, $\beta = 0.4$ $\beta = 0.6$ and $\beta = 0.8$, respectively. In terms of F_4 , the lowest value is attained when $\beta = 1.0$ and it is 63.3%, 62.1%, 53.7%, 36.2% better than $\beta = 0.2$, $\beta = 0.4$ $\beta = 0.6$ and $\beta = 0.8$, respectively.

Fig. 4(d) illustrates the different values obtained for F_1 , F_2 F_3 and F_4 by varying β for T10 database. The lowest value of F_1 is recorded in $\beta = 1.0$ and the corresponding value is 0.24925, the other values are 0.4955, 0.25225, 0.251 and 0.24925 for $\beta = 0.2$, $\beta = 0.4$ $\beta = 0.6$ and $\beta = 0.8$, respectively. The lowest value of F_2 is recorded as 0.477291 in $\beta = 1.0$ and the other values are 0.477682, 0.477552, 0.477552 and 0.477376 at $\beta = 0.2$, $\beta = 0.4$ $\beta = 0.6$ and $\beta = 0.8$, respectively. $\beta = 0.2$, $\beta = 0.4$ $\beta = 0.6$ and $\beta = 0.8$, respectively. In terms of F_3 , the lowest value is 0.124899 and it is found in $\beta = 1.0$ and it is 50.7% better than $\beta = 0.2$, 24% better than $\beta = 0.4$, 17.5% better than $\beta = 0.6$ and 15.18% better than $\beta = 0.8$. The F_4 records the lowest value in $\beta = 1.0$ and the corresponding value is 0.133773 and it is 50.5% better than $\beta = 0.2$, 19.3% better than $\beta = 0.4$, 16.5% better than $\beta = 0.6$ and 15.8% better than $\beta = 0.8$.



(a) (b) (c) (d)

Fig. 5. Performance analysis of the proposed model over the conventional model for data preservation algorithm using database (a) Chess, (b) Retail, (c) T40 and (d) T10

VII. Conclusion

This paper developed a novel PPDM method by following three major phases (a) data sanitization (b) key extraction (v) data restoration. The data sanitization was accomplished initially and here the sensitive data were mined. Then, the mined sensitive data were preserved using the optimal key. The preserved message was send from the transmitter to the receiver via the suitable communication channel. In the receiver side, with the same optimal key, the preserved message was reconstructed in the data restoration phase. The optimal key selection was the major challenge and it was satisfied using OI-CSA, which was the extended version of CS. Finally, the performance of the proposed model in terms of varying scaling factor β is evaluated for four datasets chess, retail, T40, and T10 for four research issues such as HF rate, IP Rate, and FR and DM. In terms of F_3 for t101, the lowest value is 0.124899 and it is found in $\beta = 1.0$ and it is 50.7% better than $\beta = 0.2$, 24% better than $\beta = 0.4$, 17.5% better than $\beta = 0.6$ and 15.18% better than $\beta = 0.8$. In terms of F_4 concerning retail dataset, the lowest value is recorded in $\beta = 1.0$ and it is 51.3%, 46.7%, 46.4% and 23.4% better than $\beta = 0.2$, $\beta = 0.4$, $\beta = 0.6$ and $\beta = 0.8$, respectively.

References

[1]. J. Mandala, M. V. P. C. S. Rao,"Privacy preservation of data using crow search with adaptive awareness probability",Journal of Information Security and Applications,Vol.44,pp. 157-169,February 2019.

[2]. Z. Guan, Y. Zhang, L. Wu, J. Wu, J. Hu,"APPA: An anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT"Journal of Network and Computer Applications,vol.125,pp.82-92,January 2019.

[3]. T. Grinshpoun, T. Tassa, V. Levit, R.Zivan,"Privacy preserving region optimal algorithms for symmetric and asymmetric DCOPs",Artificial Intelligence, Vol.266,pp.27-50, January 2019.

[4]. S. A. Abdelhameed, S. M. Moussa, M. E. Khalifa,"Restricted Sensitive Attributes-based Sequential Anonymization (RSA-SA) approach for privacy-preserving data stream publishing Knowledge-Based Systems, Vol.164,pp. 1-20,January 2019.

[5]. M. Rodriguez-Garcia, M. Batet, D. Sánchez,"Utility-preserving privacy protection of nominal data sets via semantic rank swapping",Information Fusion, Vol.45,pp.282-295,January 2019.

[6]. N. Kaaniche, M. Laurent, "Data security and privacy preservation in cloud storage environments based on cryptographic mechanisms",Computer Communications, Vol.111,pp.120-141,October 2017.

[7]. D. Sánchez, M. Batet, "Privacy-preserving data outsourcing in the cloud via semantic data splitting Computer Communications, Vol.110,pp.187-201,September 2017.

[8]. H. Wang, Z. Xu, "CTS-DP: Publishing correlated time-series data via differential privacy Knowledge-Based Systems, Vol.122,pp. 167-179,April 2017.

[9]. K. K. Tripathi, "Discrimination Prevention with Classification and Privacy Preservation in Data mining", Computer Science, vol.79,pp. 244-253,2016.

[10]. D. Li, C. Chen, Q. Lv, L. Shang, N. Gu, "An algorithm for efficient privacy-preserving item-based collaborative filtering Future", Generation Computer Systems, Vol.55,pp. 311-320,February 2016.

[11]. E. G. Komishani, M. Abadi, Fatemeh Deldar, "PPTD: Preserving personalized privacy in trajectory data publishing by sensitive attribute generalization and trajectory local suppression", Knowledge-Based Systems, Vol.94,pp.43-59,February 2016.

[12]. Y. Li, C. Bai, C. K. Reddy, "A distributed ensemble approach for mining healthcare data under privacy constraints ",Information Sciences, Vol.330,pp.245-259,February 2016.

[13]. X. Yang, R. Lu, H. Liang, X. Tang, "SFPM: A Secure and Fine-Grained Privacy-Preserving Matching Protocol for Mobile Social Networking", Big Data Research, Vol.3,pp.2-9,April 2016.

[14]. U.Yun, J. Kim, "A fast perturbation algorithm using tree structure for privacy preserving utility mining", Expert Systems with Applications, Vol.42,no.3, pp.1149-1165,February 2015.

[15]. K. Zhang, X. Liang, M. Baura, R. Lu, X. (Sherman) Shen, "PHDA: A priority based health data aggregation with privacy preservation for cloud assisted WBANs", Information Sciences, Vol.284,pp.130-141,November 2014.

[16]. A. Fahad, Z. Tari, A. Almalawi, A. Goscinski, A.Mahmood, "PPFSCADA: Privacy preserving framework for SCADA data publishing", Future Generation Computer Systems, Vol.37,pp.496-511,July 2014.

[17]. A. Bilge, H. Polat, "An improved privacy-preserving DWT-based collaborative filtering scheme", Expert Systems with Applications, Vol.39,no.3,3841-3854,February 2012.

[18]. X. Zhang, C. Liu, S. Nepal, J. Chen, "An efficient quasi-identifier index based approach for privacy preservation over incremental data sets on cloud", Journal of Computer and System Sciences, Vol.79,no.5,pp.542-555, August 2013.

[19]. A. Waqar, A. Raza, H. Abbas, M. K. Khan, "A framework for preservation of cloud users' data privacy using dynamic reconstruction of metadata", Journal of Network and Computer Applications, Vol.36,no.1,pp.235-248,January 2013.

[20]. G. U. Yong-hao, "An automatically privacy setting algorithm based on Rasch model", The Journal of China Universities of Posts and Telecommunications, Vol.20,pp.17-20,December 2013.

[21]. D. Menaga and S. Revathi, "Least lion optimisation algorithm (LLOA) based secret key generation for privacy preserving association rule hiding," in IET Information Security, vol. 12, no. 4, pp. 332-340, 7 2018.

[22]. C. Xu, J. W. Hong, w. Dan, Y. Pan, "An Improved EMASK Algorithm for Privacy-Preserving Frequent Pattern Mining",Computational Intelligence and Security, pp 752-757, 2015.

[23]. J. Zhu, Z. Li, "Privacy Preserving Association Rule Mining Algorithm Based on Hybrid Partial Hiding Strategy", LISS, pp 1065-1070, 2015.

[24]. S. Rehman, A. Sharma, "Privacy Preserving Data Mining Using Association Rule Based on Apriori Algorithm", Advanced Informatics for Computing Research, pp 218-226,July 2017.

[25]. G. Kalyani, M. V. P. C. S. Rao, B. Janakiramaiah, "Privacy-Preserving Classification Rule Mining for Balancing Data Utility and Knowledge Privacy Using Adapted Binary Firefly Algorithm",Arabian Journal for Science and Engineering,vol.43,no.8,pp 3903–3925,August 2018.