Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 4, August 2025, pp. 3202∼3217
ISSN: 2302-9285, DOI: 10.11591/eei.v14i4.9113

COMATS: a cuckoo-mimicking data anonymization scheme for preserving sensitive preferences in transaction data

Dedi Gunawan[1], Yusuf Sulistyo Nugroho[1], Fatah Yasin Al Irsyadi[1], Diah Priyawati[1], Arini Nur Rohmah[1], Bambang Sukoco[2], Syful Islam[3]

[1]Department of Informatics Engineering, Faculty of Communication and Informatics, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia

[2]Department of Law, Faculty of Law, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia

[3]Department of Computer Science and Engineering, Faculty of Engineering, Gopalganj Science and Technology University, Gopalganj, Bangladesh

Article Info

Article history:

Received Aug 8, 2024
Revised May 27, 2025
Accepted Jul 5, 2025

Keywords:

COMATS
Brood parasitism behavior
Data anonymization
Sensitive personal preference
Transaction database

ABSTRACT

Sharing customer transaction data is becoming more perceived in e-commerce and retail industries. Even though the act derives benefits for companies, it may end up in certain privacy threats, such as sensitive personal preferences disclosure. Therefore, the data owner should take measures to minimize the threats. Data anonymization is one of the solutions that has been suggested to address the issue. However, there are still underlying problems, specifically in diminishing the amount of information loss and item loss, as well as maintaining data properties of the anonymized dataset. This paper proposes a unique data anonymization scheme called COMATS. It adopts the brood parasitism behavior of cuckoo birds in laying their eggs into host nests. The scheme incorporates item insertion technique and item suppression technique. The robustness of the proposed scheme lies in its strategy for selecting suppressed items and determining the inserted items. To ensure its efficacy, the proposed method is evaluated in several experiments. The experimental results suggest that the COMATS can guarantee privacy protection by reducing the probability of a successful attack. Additionally, it can also reduce the number of item losses and preserve better data utility in comparison to existing data anonymization schemes.

Corresponding Author:

Dedi Gunawan
Department of Informatics Engineering, Faculty of Communication and Informatics
Universitas Muhammadiyah Surakarta
Surakarta, Indonesia
Email: dedi.gunawan@ums.ac.id

1. INTRODUCTION

E-commerce and retail companies are actively accumulating the data of their customers from keyword searches, visited products, and transaction records. In some specific events the generated data from the e-commerce system can reach more than 490,000 sales transactions in a second [1]. Therefore, the gathered data is then saved to a database in Terabyte size [2]. Nowadays, Society 5.0 encourages companies to work collaboratively with other institutions to analyze their data [3], including transaction databases. The transaction databases offer numerous advantages for business institutions when the companies can conduct data analysis through data mining and big data technology as well as machine learning for marketing purposes [4].

Various data analysis tasks can be performed from the transaction data such as personalized item recommendation [5], observing consumption habits [6], and influencing the customer in shopping behavior [7]. Even though such activity is becoming more recognized, the lack of awareness of data owners to privacy issues may end up in severe privacy violations which may also impact a certain financial condition of the customers. Moreover, the high number use of payment tools such as credit cards is prone to be targeted by attackers to hijack sensitive information and commit fraudulent actions. Therefore, due to its importance, privacy issues are enforced by laws and formal regulations such as the General Data Protection Regulation (GDPR), The Electronic Communications Privacy Act (ECPA), The Children's Online Privacy Protection Rule (COPPA), and other laws for specific types of data.

Discerning the fact that sharing transaction databases may bring some concealed threats i.e., exposing sensitive personal preferences, data owners should take measures to safeguard the customer's privacy. Sensitive personal preference in transaction data context refers to a set of purchased items that allows the data analyst to infer the sensitive personal preference of a specific customer. Preserving sensitive preferences is essential to avoid users getting product promotion email spam and phishing. To illustrate the usefulness of hiding personal preferences we provide the following example. Referring to Table 1, Bane has several items in his shopping cart, among these items there is a set of items that can be used to infer his personal preference. Let us consider Bane has transaction t_3, he bought $\{beer, wine, whiskey, chips, popcorn\}$ with the **IID** $\{25, 46, 57, 110, 112\}$, the transaction data clearly show that most of the products in Bane's transaction can be categorized as alcoholic drinks and snacks.

Table 1. Illustration of transaction database D

Tid	Cname	IID
t_1	John	1,2,3,8,10
t_2	Jane	12,17,18,100
t_3	Bane	25,46,57,110,112
t_4	Martini	22,23,28,49
t_5	Aston	11,31,52,93,110
t_6	Felix	4,6,7,9,10,12
t_7	Nitani	11,31,52,8,101
t_8	Marlin	1,16,46,72,99
t_9	Ben	55,102
t_{10}	Doet	13,31

Thus, if the database is shared transparently, malignant data recipients may perform some data analysis to deduce sensitive information such as the private preference of any individual in D. Data anonymization can be an effective action to unfold the privacy issues in data publishing or sharing. Since the last decades, several data anonymization schemes that follow either generalization or suppression techniques have been suggested to ensure privacy preservation in transaction data publishing [8]-[11]. The generalization technique refers to reconstructing specific values in a dataset to another less specific value while semantically consistent. Meanwhile, suppression aims to remove specific values such that the values cannot be observed from the dataset. The existing data anonymization methods have been tested and the results show that the methods successfully achieve the goals. Employing generalization methods to satisfy the k-anonymity in transaction dataset results in excessive information loss [8], [12]. In addition, it may also fail to hide customer personal preference due to the specific items only changing to their general value, allowing irresponsible data recipients to picture the customer's personal preference [13]. Therefore, the item suppression-based method is a more realistic solution to achieve the goal than the generalization-based method.

Considering the trade-off between privacy and data utility, each method has its benefits and drawbacks. Therefore, balancing the trade-off between those two remains a challenge in designing the data anonymization method. Depending on the application, the trade-off can be set by the database owner. For example, if the application is for strict security measures the the security and privacy protection level is set to maximum, resulting in significantly low data utility. To realize this, a data anonymization algorithm that combines item suppression and item insertion strategy is proposed. The proposed method follows the brood parasitism behavior of cuckoo birds *(Cucunus canorus)* in nesting their eggs. The cuckoo bird is one of the invasive birds that lay eggs in other birds' nests for reproduction [14]. To successfully breed, the cuckoo birds sometimes remove several eggs of the other host birds that already exist in the nest and deposit their eggs on it [15]. The cuckoo bird has interesting behavior in selecting a targeted nest and determining which eggs should be removed from

the nest of the host bird. Figure 1 illustrates a cuckoo bird selecting a host nest. Therefore, we adopt its behavior in nesting for the item selection and removal process of our proposed method.

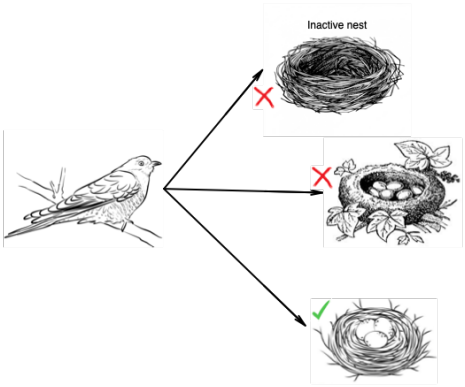

Figure 1. The illustration of cuckoo bird for nest selection

2. RELATED WORK

An early work in data anonymization has been suggested in [16] namely k-anonymity. The work aims to prevent personal identity leakage in a released micro-dataset. Most of the existing studies on privacy protection data are implemented in micro-dataset. The microdata table is composed of some attributes such as identity attribute (IA), quasi-identifier attribute (QA), and sensitive attribute (SA). A table is said to be safe to release if it satisfies the k-anonymity principle, i.e., when at least k-1 existing records have the same values in the data table.

The existing data anonymization schemes are mainly based on generalization or suppression techniques. These techniques have successfully boosted privacy protection in microdata release [17], [18]. The essential concept of the data generalization strategy is semantically altering a genuine value of the QA to its general value for instance, a specific profession is altered with its profession category (programmer→IT Employee→Tech professional). On the other hand, the suppression approach attempts to discard specific SA values in the data table to make it undiscovered.

Adopting a generalization-based data anonymization method can successfully guarantee privacy protection in the data release and it provides flexibility in data anonymization [19]. Unfortunately, the generalization-based methods could result in overmuch item loss when applied to transaction datasets [20]. The drawback exists since items in a data record are transformed into their general category even though these items are distinct from each other. As an illustration, for example, there are several items like $Milk, Cheese, Yogurt$ that exist in a record, when the generalization method is employed the items are changed to $Dairy$. The term item loss refers to the number of items that disappear and cannot be observed from the database $\widetilde{D}$. Furthermore, as the number of item losses increases, the data utility of the $\widetilde{D}$ decreases.

A different approach to solving both issues has also been proposed in [21]. The method is called ρ-uncertainty which employs a partial suppression technique. The method adopts a heuristic solution that can maintain the distribution of data as well as preserve important item correlations in the database. Unfortunately, the scheme is specifically intended to preserve sensitive information in association rule mining and therefore may not apply to various data analyses. A recent interesting idea has been proposed in [22] as a solution to anonymize transaction data, the method is called $\ell\rho$-suppression. The scheme also follows the suppression method by determining two parameters ℓ and ρ to limit the number of suppressed items, as a result, it can successfully minimize the number of item loss and data utility loss.

Instead of applying generalization or suppression technique, a swapping strategy is proposed in [23]. Applying the method can guarantee to cover the personal tendency or preference of the data subject. Moreover, the method does not evict or insert items into the transaction databases, thus, the item loss can be maintained. Regardless of its advantages, the method changes the structure of items that compose the swapped transaction records. Consequently, item correlation among them cannot be maintained and it leads to some degree of

information accuracy. A notion of considering items that have a certain sensitivity level is proposed in [23]. The proposed method differentiates items based on their sensitivity level and performs a clustering strategy to generate an anonymized database. A noise addition strategy that adopts the differential privacy concept is introduced in [24]. The technique is specifically constructed to preserve privacy for frequent item mining tasks using local differential privacy. While privacy protection in machine learning mainly focuses on the model privacy protection from unauthorized modification [25]. Designing a data anonymization algorithm that fits all various types of databases is nearly impossible since each database has different characteristics and properties. Therefore, there is no one-fits-all algorithm to anonymize database [26].

Despite the various methods that have been proposed for anonymizing transaction datasets, the issues of minimizing item loss and preserving data utility are still challenging. Thus, in this paper, a cuckoo-mimicking behavior method for anonymizing transaction dataset COMATS is suggested. The proposed method adopts the cuckoo bird behavior in the reproduction process where they lay their egg and expel the host bird's egg from the host nest. The investigation in [27] found that cuckoo birds are more likely to lay their eggs in an active nest where the host egg has almost similar pattern and size to its egg. Further investigation in [28] found that cuckoo birds prefer to settle their eggs in a nest that contains the minimum number of eggs. The bird selects the eggs that have similarities with their own such as color, size, and pattern. In this paper, we consider the active nest as the transaction record containing sensitive items that should be anonymized. The COMATS not only can hide sensitive personal preferences in transaction datasets but also does not induce significant item loss and data utility loss.

3. PRELIMINARIES AND PROBLEM STATEMENT

3.1. Transaction dataset

A collection of customer transaction data records is referred to as transaction dataset D. The dataset is constructed from a set of attributes such as transaction record identity Tid, customer's name or identity number Cid, and a set of the purchased items, Pid. As portrayed in Table 1, the D, incorporates a set of transaction record T where T is a collection of distinct customer transaction records that are stored in D, thus $T = \{t_1, t_2, t_3, \ldots, t_n\}$. Each transaction record t_q is a non-empty set of collection of items $i = \{1_1, 1_2, 1_3, \ldots, i_n\}$ from the global item I.

3.2. Taxonomy item generalization

In real life, items are usually categorized under certain criteria or based on their similarity. Item categorization brings some benefits such as for creating product catalogs and marketing purposes [29]. The item category can be depicted in a taxonomy generalization graph. The graph shows the hierarchy of items from the most general perspective to the most specific one. The illustration of the taxonomy generalization graph is depicted in Figure 2.

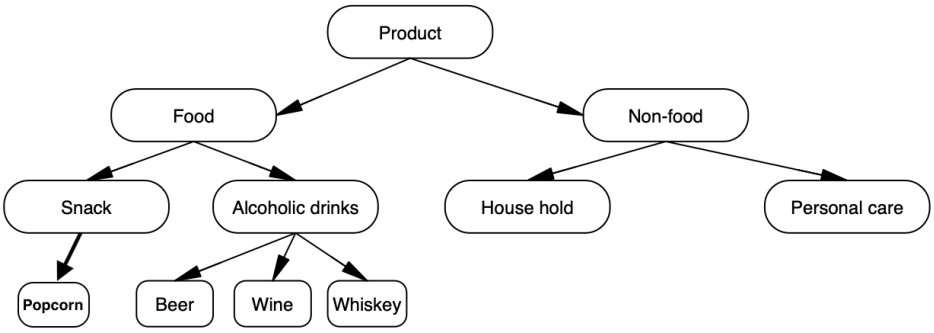

Figure 2. Taxonomy item categorization

Each item category G may have at least one item i_q as its respective member. In addition, each item i_q only belongs to one category G_j without any overlapping membership to another G'_j. For example, the item $beer$ is an item of alcoholic drinks only and it does not fit into another category. The number of G, where $G = \{G_1, G_2, ..., G_n\}$ in a transaction dataset depends on the dataset owner in managing their items.

3.3. Transaction record

A transaction record t_q is composed of at least an item i_q that belongs to specific G, where $i_q \subset G$. In other words, the t_q contains a set of G_j. Therefore, by observing the item $i_q \in t_s$ one can understand the items' category that represents the preferences of the customer.

3.4. Sensitive item

The work in [30] suggested that the user or database owner can pre-define whether an item is sensitive or non-sensitive. Therefore, the items I can be classified into two groups namely sensitive items I_s and non-sensitive items I_{ns}, thus $I = I_s \cup I_{ns}$. Each item i_q is only belong to either I_s or I_{ns} and $I_s \neq I_{ns}$. By holding the transaction data one can easily investigate the customer's preference due to the purchased items ID $PIID$ being tied to a specific customer ID Cid. In practice, the Cid is sometimes omitted when the database goes to other parties to avoid re-identification.

- Definition 1 (sensitive transaction): a transaction record t_q is composed of items i_q. Each transaction record t_q which contains $i_s \subset I_s$ is referred to as sensitive transaction t_s and the collection of t_s is denoted as T_S. In contrast, t_q that does not contain any i_s is referred to as a non-sensitive transaction, T_{NS}.
- Definition 2 (sensitive preference): each customer has a particular personal preference that can be reflected in their purchased product. Personal preference refers to a set of item categories from taxonomy generalization G that exist in the sensitive transaction record t_s. The personal preference is called a sensitive preference if it is composed of a set of sensitive items i_s from the same G. In real-world applications, given a customer transaction database, processing items straightly is more functional than manipulating item categories in various data analyses. Therefore, to hide the sensitive preference we define a weight of the category in which its items exist in t_q and denote it as wG. The wG represents the number of items belonging to G_j that exist in t_q. In addition, one is said to have a sensitive preference if the number of $i_s \in G_j$ in his/her transaction record is greater than the number of non-sensitive items i_{ns} of wG_j. To illustrate this concept, recall an example in the introduction section and inspect Figure 3. We can inspect that the wG of the alcoholic drink category is higher since it contains three sensitive items, i.e., $wG_{ad} = 3$ than that of the snack category which has only two items $wG_{snack} = 2$.

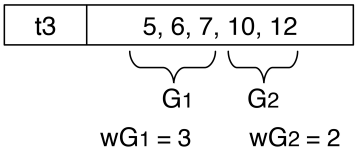

Figure 3. Weight of categories in a transaction record

- Definition 3 (privacy definition): a data anonymization scheme is considered to successfully generate an anonymized database $\widetilde{D}$ from D if it successfully and precisely removes sensitive items i_s from t_s such that the wG_j of i_s in t_s becomes lower than that of non-sensitive item. Once it is achieved, adversarial access for observing the sensitive personal preference in the t_s cannot be performed. In addition, the method should also be able to insert as low as possible the number of non-sensitive items to the t_s to reduce the number of item losses and minimize data utility loss.
- Definition 4 (item loss, $ItLoss$): the data anonymization schemes always change the content in D. To successfully hide the personal sensitive preference some items $i_s \in D$ might be expelled from t_s. The removal process causes the reduction of the item frequency which is called item loss, $ItLoss$. The $ItLoss$ quantifies the inequality between the number of items in the original dataset D and that of the anonymized $\widetilde{D}$. To avoid excessive item loss the data anonymization method should be able to control the removal process. It is expected that the method results in a low value for $ItLoss$, which suggests that the content of $\widetilde{D}$ is similar to that of the D.
- Definition 5 (data utility loss, $UtLoss$): as the items of $\widetilde{D}$ differ from its original D, the amount of information contained in $\widetilde{D}$ is also lower than that of the D. Therefore, measuring the $UtLoss$ is necessary to evaluate the performance of a data anonymization method in maintaining data utility. The lower the $UtLoss$ resulted the better the performance of the algorithm.

4. PROPOSED METHOD

The proposed method COMATS follows the brood parasitism behavior of the cuckoo bird and it becomes the unique feature of COMATS in adopting two strategies at once i.e., item removal and item insertion. The COMATS has five main steps, which are; i) reading and examining $t_q \in D$, ii) collecting the sensitive transaction record t_s to T_s, iii) computing the weight of wG to determine the weight of each category that exists in t_s, iv) selecting and removing $i_s \in t_s$, and v) substituting the removed item with non-sensitive item i_{ns}. The item removal aims to eliminate sensitive information while the item insertion acts to maintain data utility and data properties. Calculating the weight of sensitive itemset is also another main feature of COMATS. The weight represents the degree of sensitivity of itemset and it plays an important role in reducing item loss. The framework of COMATS is presented in Figure 4.

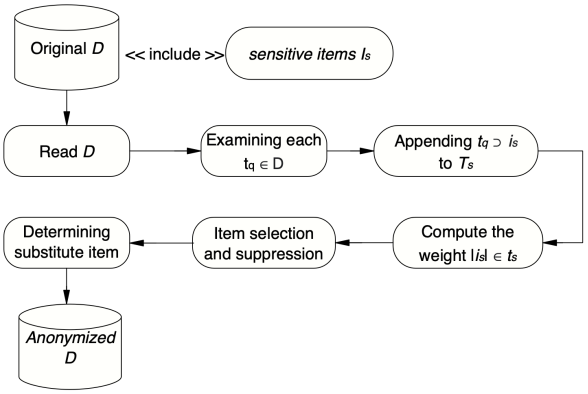

Figure 4. The COMATS framework

4.1. Reading and examining $t_q \in D$

The process of reading and examining $t_q \in D$ can be performed in one scan. The method checks each t_q and determines whether the t_q contains sensitive items i_s. In this step, the database owner should provide a set of sensitive items i_s. If the scheme finds $i_s \in t_q$, the t_q is considered a sensitive transaction t_s and will be appended to T_S otherwise to T_{NS}. Later on, only T_S is subjected to the data anonymization process. The pseudo-code of the first and second steps can be inspected in Algorithm 1.

Algorithm 1: Reading and examining t_q

Input: D, i_s
Result: T_S and T_{NS}
Scan D
$\forall t_q$ in D
if t_q *contains* i_s **then**
    $t_s \leftarrow t_q$
    Append t_s to T_S
    return T_S
else
    Append t_q to T_{NS}
    return T_{NS}
end

4.2. Computing sensitive itemset weight $wG \in t_s$

This step plays a critical role in our proposed method since it investigates whether a transaction has a sensitive preference. The procedure works by inspecting categories G_j in the t_s and computing the wG of each $G_j \in t_s$. The wG_j represents the degree of sensitivity of itemset and it is determined by calculating the number of items i_q that exist in t_s. Prior to processing it, the database owner needs to provide a taxonomy generalization

tree that contains a set of item categories G and their respective items i_q. The pseudo-code in Algorithm 2 represents the weight computation.

Algorithm 2: Computing wG

Input: G, T_S
Result: wG_j
Read T_S
$\forall t_q$ in T_S
if $i_q \in G_j$ **then**
  $wG_j += 1$
  return wG_j
end

4.3. Selecting and removing sensitive item

The cuckoo birds are likely to select a host nest with the least number of eggs. The proposed method adopts their characteristic where it computes the item length of each sensitive transaction record t_s, $|t_s|$. The t_s which has the least item length $Min|t_s|$ becomes the priority for the data modification process. Finally, the t_s is sorted according to its $|t_s|$ in ascending order, and select the one with $Min|t_s|$. As it has been described in section 2, the cuckoo bird selects the nest with the minimum number of eggs and randomly selects an egg in the nest to be removed. However, adopting its behavior arbitrarily into the scheme may severely induce side effects i.e., $ItLoss$ and $UtLoss$.

Therefore, determining $i_s \in t_s$ as the candidate for the suppression process is crucial to thwart excessive $ItLoss$ and $UtLoss$. To minimize the $ItLoss$ and $UtLoss$ the proposed scheme selects item $i_s \in t_s$ which has the highest frequency f among other i_s in the t_s as the candidate of suppressed item Ci_s. Intuitively, when the selected i_s which has the highest frequency is removed from t_s, the reduction of occurrence frequency f of $i_s \in D$ does not result in the disappearance of the $i_s \in D$. Therefore, it allows the i_s to remain observable in D for analysis. In a case there is more than one sensitive item i_s having the same value f co-exists in t_s, the algorithm will randomly select the one that has the highest item frequency. The detailed procedure is depicted in Algorithm 3 .

Algorithm 3: Item selection for suppression and insertion

Input: $T_S, i_s \in I_S, i_{ns} \in I_{NS}, G$
Result: $\widetilde{D}$
$\forall t_s \in T_S$
compute and sort $|t_q|$
select $Min|t_q|$ count the f of $i_s \in \mathcal{D}$
select i_s with the highest f
$Ci_S \leftarrow i_s$
if *# of* $Ci_s > 1$ **then**
  select $i_s \in Ci_s$ randomly
  remove i_s from t_s
else
  remove i_s from t_s
end
$\forall i_{ns} \in I_{NS}$
if $G \supset i_{ns} \equiv G \supset i_s$ **then**
  count the f of $i_{ns} \in \mathcal{D}$
  if f *of* $i_{ns} \simeq f$ *of* i_s **then**
    insert i_{ns} to t_s
  end
end
generate anonymized transaction records, $\widetilde{T_S}$

4.4. Selecting item insertion

Once the cuckoo bird removes an egg from the host's nest, the following step is starting to lay its egg. In this scheme, we consider the egg's pattern as the category G of the items and its size as the item frequency f. Therefore, to select an inserted item we consider finding non-sensitive items i_{ns} which have the same category as the selected item $i_s \in Ci_s$ since items from the same category have close similarity. Another consideration to select the item for the insertion process is the item's occurrence frequency f. The objective of selecting i_{ns} which has similar f to item $i_s \in Ci_s$ is mainly to keep the item discoverable. In addition, it can also avoid excessive data utility loss specifically when data recipients conduct data analysis using data mining processes such as frequent itemset mining and association rule mining. In detail, the procedure of item insertion is presented in Algorithm 3.

4.5. Combining T_{NS} and $\widetilde{T_S}$

The final step of the COMATS is combining the set of modified sensitive transaction records with the set of non-sensitive records. As has been depicted in Algorithm 4, the merging process straightforwardly reads both T_{NS} and $\widetilde{T_S}$ and then saves it to an anonymized database $\widetilde{\mathcal{D}}$.

Algorithm 4: Combining T_{NS} and $\widetilde{T_S}$

Input: $T_{NS}, \widetilde{T_S}$
Result: $\widetilde{D}$
1 load T_{NS}
2 load $\widetilde{T_S}$
3 $\widetilde{\mathcal{D}} \leftarrow T_{NS} \cup \widetilde{T_S}$

5. EXPERIMENT RESULT

The experiment is carried out by utilizing an online retail dataset from the UCI machine learning repository and a liquor dataset that has different properties. Both datasets are real-life data that are available online and the detailed properties of the datasets are presented in Table 2. We involve several metrics to analyze our proposed method's performance. The computation is performed using the Google Cloud computing platform with 2 vCPU and 13GB of RAM for evaluating the performance of the proposed method. The first metric is related to privacy protection issues namely probability success attack Pr that is generated from query answering [31]. The second evaluation measures the data utility of an anonymized transaction database such as frequent itemset mining FI and association rule mining AR [32]. Another evaluation metric that can be adopted is dissimilarity, which reflects the amount of item loss. The metric measures the similarity of item frequency between the items in the original data and that of the anonymized data. The last is the performance metric which evaluates the computation cost of the algorithms [33]. In comparison with other existing algorithms such as split personality (SPLIT) [34], and direct anonymization (DIRECT) [35] is also conducted to evaluate the quality and performance level of our proposed methods.

Table 2. Properties of $\mathcal{D}$

Properties	Datasets	
	Online retail	Liquor
# of distinct item	2,603	4,026
# of record	540,455	52,131
# of total items in $\mathcal{D}$	2,363,344	410,619
Average record length	4	8

5.1. Probability success attack

The probability of successful attack metric adopts the query answering technique in [34]. Once the data recipient obtains the anonymized dataset, they can conduct various analyses including linking sensitive items to the individual. The probability of the data recipient linking an individual to his/her sensitive preference is called the probability of a successful attack, $Pr_s(X)$ where X indicates the dataset. In this experiment, we assume an attacker has a set of itemset as his/her prior knowledge to conduct a linking attack to the databases.

Since predicting exact attacker knowledge is nearly impossible [36], we generate a set of itemset with a length from 1 item to 4 items in the Online retail dataset and from 1 item to 5 items for the Liquor dataset. The probability of a successful attack refers to the number of sensitive transactions t_s that occur when data recipients conduct a query using the sensitive itemset to the anonymized dataset $\widetilde{D}$. Assume, the probability of an attacker successfully compromising the original D as Pr_sD, where $0 \leq Pr_s(X) \geq 1$, while the probability of a successful attack in $\widetilde{D}$ is denoted as to $Pr_s\widetilde{D}$. In general, when an original database had been anonymized the value of $Pr_s(\widetilde{D}) < Pr_s(D)$. Experiment results indicate that the proposed solution can significantly reduce the probability of a successful attack in the $\widetilde{D}$. Figure 5 shows that the Pr_sD is 0.143 after the COMATS is executed the $Pr_s(\widetilde{D})$ becomes 0.005, while that of the SPLIT is 0.125, and that of the $Pr_s(\widetilde{D})$ from DIRECT is 0. In addition, all three methods can successfully reduce the $Pr_s(\widetilde{D})$ to 0, indicating that all of them work well in the liquor dataset. These results imply that the proposed method achieves the modest value of the probability of a successful attack in all datasets.

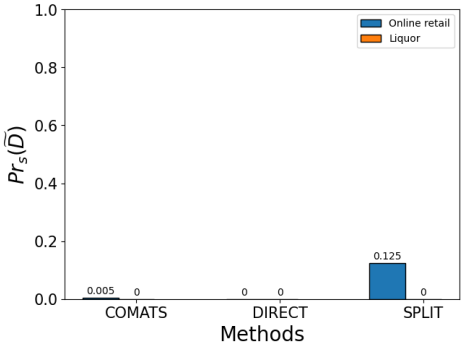

Figure 5. Successful attack probability in $\widetilde{D}$

5.2. Data utility metric

5.2.1. Item loss

Item loss examines the count of items that disappear from a database after the database has been anonymized. To measure it, we adopt the concept of data dissimilarity $Diss$ from [37]. The metric counts the distinction between the data histogram in an original database and that of the anonymized database. The notation $fx(i)$ expresses the number of appearances of item i in dataset x. In (1) declares the calculation of $Diss$.

$$Diss(D, \widetilde{D}) = \frac{1}{\sum_{i=1}^{d} f\mathcal{D}(i)} \times \left| \sum_{i=1}^{d} f\mathcal{D}(i) - \sum_{i=1}^{\widetilde{d}} f\widetilde{\mathcal{D}}(i) \right| \tag{1}$$

The value of $Diss$ spans from 0-1, where the lower value indicates the closeness of frequency of item i between the original database D and that in the anonymized database $\widetilde{D}$.

Table 3 shows that the dissimilarity values resulting from COMATS are 0.025 and 0.085 for the Online retail dataset and the Liquor dataset respectively. These values are lower compared to that of DIRECT for both datasets with the values are 0.029 and 0.110. Meanwhile, SPLIT can achieve the lowest value of dissimilarity since it does not remove or alter any item in the $\widetilde{\mathcal{D}}$.

Table 3. Data dissimilarity of $\widetilde{\mathcal{D}}$

Methods	Datasets	
	Online retail	Liquor
COMATS	0.025	0.085
DIRECT	0.029	0.110
SPLIT	0.000	0.000

The experimental result as depicted in Figure 6 implies that the proposed method induces lower item loss compared to that of the DIRECT. However, the SPLIT method has the best performance in minimizing item

loss since it does not add or remove items from the dataset. These conditions are mainly due to the DIRECT performing item eviction in the database to achieve an anonymized dataset. The strategy causes some items to disappear from the dataset, affecting its similarity. Meanwhile, since the SPLIT does not remove or modify any items, it allows all the items to remain observable in the dataset. Even so, our proposed method can successfully reduce excessive item loss in the anonymized database and can be an alternative solution to tackle that issue. The dissimilarity between an original dataset and the anonymized one for Online retail data can be observed in Figures 7 to 9.

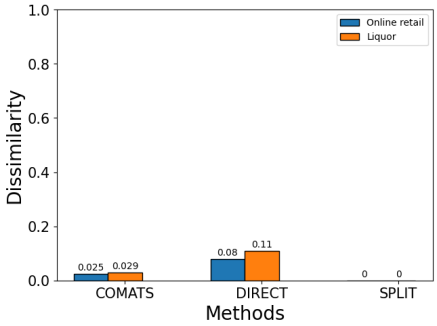

Figure 6. Dissimilarity value comparison

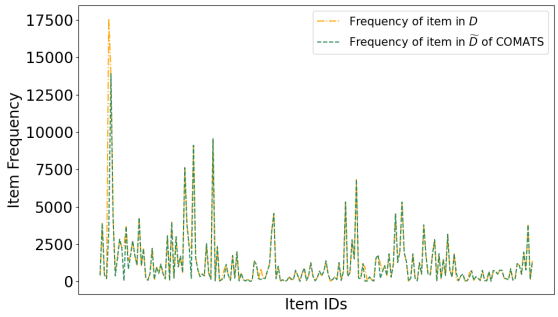

Figure 7. Item frequency in original Liquor dataset D and that of $\widetilde{D}$ generated by COMAT

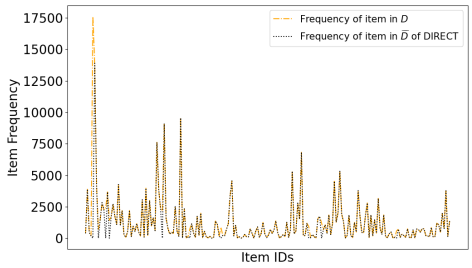

Figure 8. Item frequency in original liquor dataset D and that of $\widetilde{D}$ generated by DIRECT

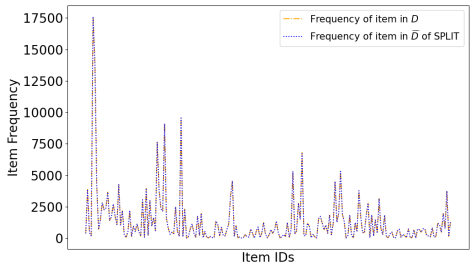

Figure 9. Item frequency in original liquor dataset D and that of $\widetilde{D}$ generated by SPLIT

In addition, as can be seen in Figures 10 to 12, the dissimilarity value of the Liquor dataset and its anonymized one indicates the same pattern as that in the Online retail dataset. There are some differences between item frequency in the original Liquor dataset and that of the anonymized version. COMATS and DIRECT cause some changes in item frequency while SPLIT maintains item frequency in its anonymized dataset. In this case, our proposed method has better performance compared to that of DIRECT. Unsurprisingly, since the SPLIT does not remove or add any items to the database, the item frequency of the anonymized database $\widetilde{D}$ generated by SPLIT has the same pattern as that of the original one.

5.2.2. Frequent itemset mining and association rule mining

Data utility points out the usability of a database for specific analysis. Logically, as a database is anonymized, its utility experiences a decrease compared to that of the original one. The idea in [38] is adopted in this paper, where the data utility indicates the ratio between an original database D and that of $\widetilde{D}$. To evaluate the data utility, one can compute the distant similarity value $U(D, \widetilde{D})$ of frequent itemset mining results and association rules mining results between an anonymized and the original database. The equation to calculate the data utility is described in (2) and (3).

$$U(D, \widetilde{D}) = \frac{|F_D \cap F_{\widetilde{D}}|}{|F_D \cup F_{\widetilde{D}}|} \tag{2}$$

$$U(\widetilde{\mathcal{D}}, \mathcal{D}) = \frac{|AR_{\mathcal{D}} \cap AR_{\widetilde{\mathcal{D}}}|}{|AR_{\mathcal{D}} \cup AR_{\widetilde{\mathcal{D}}}|} \tag{3}$$

The F_D notation indicates the frequent itemset attained from the original database D while $F_{\widetilde{D}}$ represents the frequent itemset obtained from an anonymized database $\widetilde{D}$. The higher the value of $U(D, \widetilde{D})$, the more data utility is preserved.

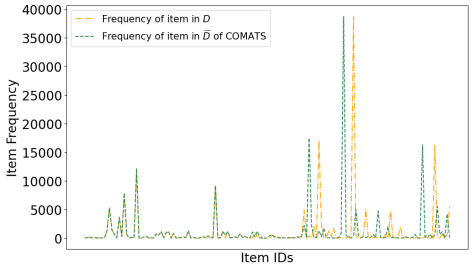

Figure 10. Item frequency in original Online retail D and that of $\widetilde{D}$ generated by COMAT

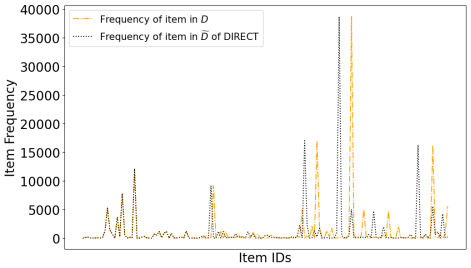

Figure 11. Item frequency in original Online retail$D$ and that of $\widetilde{D}$ generated by DIRECT

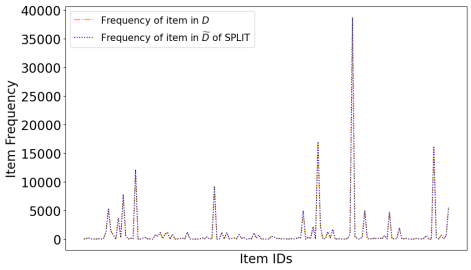

Figure 12. Item frequency in original Online retail D and that of $\widetilde{D}$ generated by SPLIT

The frequent itemset mining test is conducted in both the original database and the anonymized database. To perform this test we need to determine a minimum support threshold value $minSup$ to obtain a set of frequent itemset from the databases. The term $minSup$ refers to the percentage of minimum occurrence of itemset out of all the records in the database. Since the items in the dataset are very sparse, we vary the value of $minSup$ that is relatively low i.e., 0.001%, 0.004%, and 0.007%.

Referring to Figure 13, the proposed method consistently works in all conditions. The generated anonymized database $\widetilde{D}$ from the COMATS results in relatively high similarity with around 0.9. Even though in the Online retail dataset the DIRECT and SPLIT outperform the proposed method, and the similarity value is lower than that of other methods, its discrepancy is not too significant. However, in another experiment with the Liquor dataset, we obtain a slightly different result. As can be observed in Figure 14, the proposed method only outperforms DIRECT and SPLIT when the support value is 0.007%. These results might be affected by the assumption of the adversary prior knowledge since all the modified items are based on that knowledge. In addition, these results occur due to the proposed method performing item substitution that changes the structure of items that compose the transaction records in $\widetilde{D}$. For example, originally the transaction record t_3 contains {5,6,7,10,12} however, after its being anonymized the items in t_3 changes to {5,6,7,8,12} which generates a difference itemsets.

Experiment to observe the association rule mining for further data utility investigation is also performed using both datasets. The association rules mining evaluates the data utility that pictures item correlation in a database. To perform the task we determine two parameters namely minimum support threshold $minSup$ and minimum confidence threshold $minConf$. The value of $minSupp$ is the same as that in frequent itemset mining tasks, while the $minConf$ values are 0.02%, 0.05%, and 0.08%. Subsequently, the notations $AR_{\mathcal{D}}$ and $AR_{\widetilde{\mathcal{D}}}$ represent the set of rules in the original database and that in the anonymized database.

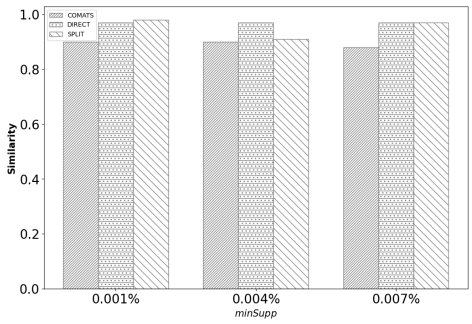

Figure 13. Data utility of frequent itemset mining of the anonymized Online retail dataset

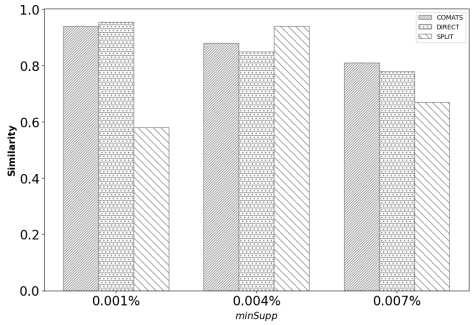

Figure 14. Data utility of frequent itemset mining of the anonymized Liquor dataset

The evaluation result on the Online retail dataset can be investigated in Figure 15. It shows that the data utility of the anonymized database generated from the proposed method COMATS achieves better results in comparison to that of the SPLIT in any scenario. The similarity of the mining result in the Online retail dataset from that of the COMATS is about 0.9, while the similarity result of that of $SPLIT$ is only around 0.5. The proposed method has a strategy to minimize the number of items collision which does not affect the number of records in the database. In contrast, the $SPLIT$ method generates new records in the database, resulting in significant degradation of the mining results. Compared to the similarity generated by the DIRECT method, our proposal has a slightly lower similarity with a difference of around 0.05. Meanwhile, the experiment using the Liquor dataset indicates that the DIRECT outperforms our proposed method and SPLIT. Referring to Figure 16, the COMATS only achieves 0.25 similarity while DIRECT and SPLIT obtain around 0.5 and 0.05 similarity values respectively. These results are mainly caused by the item substitution process that changes the item composition and item correlation of the transaction since both aspects greatly influence the association mining results.

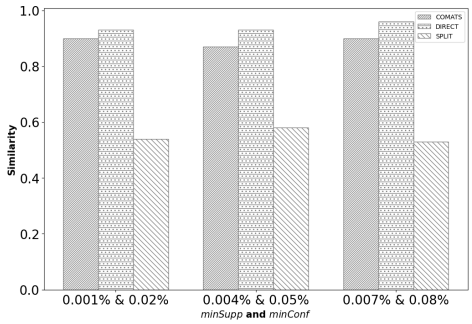

Figure 15. Data utility of association rule mining of the anonymized Online retail dataset

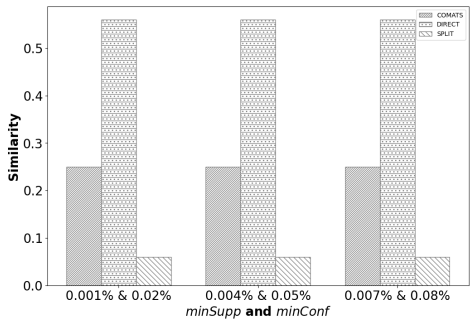

Figure 16. Data utility of association rule mining of the anonymized Liquor dataset

Overall, the proposed method works considerably acceptable to generate an anonymized database that can protect sensitive information from linking attacks while maintaining data utility.

5.3. Computation cost

The performance evaluation related to the computational cost measures the computation resources required in generating an anonymized dataset. The evaluation compares the cost of the proposed method with that of the several existing algorithms. Table 4 shows that the CPU system time taken by the COMATS is similar to that of DIRECT. In addition, it is more than two times lower compared to that of the SPLIT. This result is achieved since the COMATS performs weight computation only to the transaction records that contain sensitive items. Moreover, by discriminating transactions that do not contain sensitive items the scanning process can be minimized. In terms of memory usage, our proposed method takes the smallest amount of memory usage compared to that of the others. Interestingly, COMATS also obtains the smallest memory increment in completing the program execution. These results are achieved due to the COMATS performing item suppression

only to the item i_s that exists in T_s, not in the whole transaction record. Additionally, the method determines substitute items that come from the item with the same category, as a result, the search space in finding the substitution items becomes smaller, and finally, it can reduce the computation cost. Overall, the computational cost efficiency of COMATS is at the top compared to that of DIRECT and SPLIT. We should also note that since COMATS investigates items and its category, while at the same time, it also computes the weight of a transaction record, it will require high computation resources as the number of items and its category as well as the number of transactions grows.

Table 4. Computation cost

Method	CPU time sys (ms)	Peak memory (MiB)	Memory increment (MiB)
COMATS	115	220.56	0.02
DIRECT	114	222.46	1.03
SPLIT	142	243.02	13.97

6. CONCLUSION

In this paper, a method for anonymizing transaction databases to prevent sensitive personal preferences disclosure is proposed. The method namely COMATS mimics the behavior of a cuckoo bird in breeding its offspring. It consists of five main steps to generate an anonymized database. The method performs item substitution where it selects a non-sensitive item to substitute the sensitive item. Therefore, it allows the protection of personal privacy, results in minimum item loss, and maintains data utility.

Experiment results suggest that the proposed method can significantly minimize the probability of a successful attack in the anonymized database. The proposed method can also minimally reduce the number of item losses compared to that of the DIRECT. Furthermore, it can also preserve more data utility for association rule mining in comparison to the existing method, i.e., SPLIT. In terms of preserving data utility for frequent itemset mining, our proposal has a slightly lower performance since it performs item substitution that leads to a change in the item composition of the transaction records.

Evaluation of the computational cost also indicates that the proposed method can effectively utilize computational resources compared to the other algorithms. It is suggested that considering an efficient procedure in the algorithm to process a huge amount of transaction data is necessary. Overall, our proposed method successfully increases the privacy protection level while maintaining data utility in an anonymized database.

Since the proposed method uses item substitution, the database owner should be very careful when adopting the method for health-related databases or other relevant databases due to some data correlations may change and result in inaccurate treatment.

FUNDING INFORMATION

This research is fully funded by the Directorate General of Higher Education, Research and Technology (DRTPM), Ministry of Education, Culture, Research and Technology of the Republic of Indonesia, under the Fundamental Research scheme with grant number 108/E5/PG.02.00.PL/2024, 007/LL6/PB/AL.04/2024, 196.10/A.3-III/LRI/VI/2024.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dedi Gunawan	✓	✓	✓	✓	✓	✓			✓	✓		✓		✓
Yusuf Sulistyo Nugroho	✓	✓		✓		✓	✓	✓		✓				✓
Fatah Yasin Al Irsyadi					✓		✓	✓		✓				
Diah Priyawati			✓			✓	✓	✓		✓				✓
Arini Nur Rohmah				✓	✓					✓	✓		✓	
Bambang Sukoco						✓				✓			✓	
Syful Islam	✓			✓						✓	✓			

C	: **C**onceptualization	I	: **I**nvestigation	Vi	: **Vi**sualization
M	: **M**ethodology	R	: **R**esources	Su	: **Su**pervision
So	: **So**ftware	D	: **D**ata Curation	P	: **P**roject Administration
Va	: **Va**lidation	O	: Writing - **O**riginal Draft	Fu	: **Fu**nding Acquisition
Fo	: **Fo**rmal Analysis	E	: Writing - Review & **E**diting		

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author [DG] on request.

REFERENCES

[1] G. Huang *et al.*, "X-engine: An optimized storage engine for large-scale e-commerce transaction processing," in *SIGMOD '19: Proceedings of the 2019 International Conference on Management of Data*, 2019, pp. 651–665, doi: 10.1145/3299869.3314041.

[2] S. Suguna, M. Vithya, and J. I. C. Eunaicy, "Big data analysis in e-commerce system using hadoopmapreduce," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 2016, pp. 1-6, doi: 10.1109/INVENTIVE.2016.7824798.

[3] X. Zheng and Z. Cai, "Privacy-preserved data sharing towards multiple parties in industrial iots," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 968–979, 2020, doi: 10.1109/JSAC.2020.2980802.

[4] D. Anggoro and P. Rahmatullah, "The implementation of subspace outlier detection in k-nearest neighbors to improve accuracy in bank marketing data," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 2, pp. 545–550, 2020, doi: 10.30534/ijeter/2020/44822020.

[5] H. Chen, "Personalized recommendation system of e-commerce based on big data analysis," *Journal of Interdisciplinary Mathematics*, vol. 21, no. 5, pp. 1243–1247, 2018, doi: 10.1080/09720502.2018.1495599.

[6] S. Kaabi and R. Jallouli, "Overview of E-commerce Technologies, Data Analysis Capabilities and Marketing Knowledge," in *Digital Economy. Emerging Technologies and Business Innovation 4th International Conference, ICDEc 2019*, Beirut, Lebanon, 2019, pp. 183–193, doi: 10.1007/978-3-030-30874-2_14.

[7] L. T. Khrais, "Role of Artificial Intelligence in Shaping Consumer Demand in E-Commerce," *Future Internet*, vol. 12, no. 12, pp. 1-14, 2020, doi: 10.3390/fi12120226.

[8] G. Ghinita, P. Kalnis, and Y. Tao, "Anonymous publication of sensitive transactional data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 2, pp. 161–174, 2011, doi: 10.1109/TKDE.2010.101.

[9] G. Loukides, A. Gkoulalas-Divanis, and B. Malin, "Coat: Constraint-based anonymization of transactions," *Knowledge and Information Systems*, vol. 28, pp. 251–282, 2011, doi: 10.1007/s10115-010-0354-4.

[10] S. L. Wang, Y. C. Tsai, H. Y. Kao, and T. P. Hong, "On anonymizing transactions with sensitive items," *Applied Intelligence*, vol. 41, no. 4, pp. 1043–1058, 2014, doi: 10.1007/s10489-014-0554-9.

[11] X. Liu, X. Feng, and Y. Zhu, "Transactional Data Anonymization for Privacy and Information Preservation via Disassociation and Local Suppression," *Symmetry*, vol. 14, no. 3, pp. 1–22, 2022, doi: 10.3390/sym14030472.

[12] C. C. Aggarwal, "On K-anonymity and the Curse of Dimensionality," in *VLDB '05: Proceedings of the 31st international conference on Very large data base*, 2005, pp. 901–909.

[13] D. Gunawan and M. Mambo, "Data anonymization for hiding personal tendency in set-valued database publication," *Future Internet*, vol. 11, no. 6, pp. 1-16, 2019, doi: 10.3390/fi11060138.

[14] "Nest Parasitism," Encyclopedia of Biodiversity: Second Edition, vol. 5, pp. 501–509, 2013.

[15] S. K. Robinson and S. I. Rothstein, "Nest Parasitism," in *Encyclopedia of Biodiversity*, pp. 365–376, 2001, doi: 10.1016/B0-12-226865-2/00209-1.

[16] P. Samarati, "Protecting respondents identities in microdata release," *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, no. 6, pp. 1010–1027, 2001, doi: 10.1109/69.971193.

[17] L. Sweeney, "Achieving k -anonymity privacy protection using generalization and suppression," *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 10, no. 5, pp. 571-588, 2002, doi: 10.1142/S021848850200165X.

[18] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "ℓ-Diversity: Privacy beyond k-anonymity," *Proceedings - International Conference on Data Engineering (ICDE'06)*, Atlanta, GA, USA, 2006, pp. 24-24, doi: 10.1109/ICDE.2006.1.

[19] R. Aufschläger *et al.*, "Anonymization Procedures for Tabular Data: An Explanatory Technical and Legal Synthesis," *Information*, vol. 14, no. 9, pp. 1–34, 2023, doi: 10.3390/info14090487.

[20] O. Vovk, G. Piho, and P. Ross, "Methods and tools for healthcare data anonymization: a literature review," *International Journal of General Systems*, vol. 52, no. 3, pp. 326–342, 2023, doi: 10.1080/03081079.2023.2173749.

[21] X. Jia, C. Pan, X. Xu, K. Q. Zhu, and E. Lo, "ρ-Uncertainty Anonymization By Partial Suppression," in *19th International Conference, Database Systems for Advanced Applications (DASFAA)*, 2014, vol. 8422, no. PART 2, pp. 188–202, doi: 10.1007/978-3-319-05813-9_13.

[22] D. Gunawan, Y. S. Nugroho, F. Y. Al Irsyadi, I. C. Utomo, I. Andreansyah, and S. Islam, "$\ell\rho$-suppression: A privacy preserving

data anonymization method for customer transaction data publishing," in *2022 International Conference on Information Technology Systems and Innovation (ICITSI)*, Bandung, Indonesia, 2022, pp. 171-176, doi: 10.1109/ICITSI56531.2022.9970910.
[23] B. Zhang, J. C.-W. Lin, Q. Liu, P. Fournier-Viger, and Y. Djenouri, "A (k, p)-anonymity Framework to Sanitize Transactional Database with Personalized Sensitivity," *Journal of Internet Technology*, vol. 20, no. 3, pp. 801–808, 2019.
[24] H. Wu, R. Ran, S. Peng, M. Yang, and T. Guo, "Mining frequent items from high-dimensional set-valued data under local differential privacy protection," *Expert Systems with Applications*, vol. 234, 2023, doi: 10.1016/j.eswa.2023.121105.
[25] M. Selvarathnam, R. Ragel, C. C. Reyes-Aldasoro, and M. Rajarajan, "Privacy vs Utility analysis when applying Differential Privacy on Machine Learning Classifiers," in *2023 19th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Montreal, QC, Canada, 2023, pp. 306-311, doi: 10.1109/WiMob58348.2023.10187829.
[26] D. Gunawan, Y. S. Nugroho, and Maryam, "Swapping-based Data Sanitization Method for Hiding Sensitive Frequent Itemset in Transaction Database," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, pp. 693–701, 2021, doi: 10.14569/IJACSA.2021.0121179.
[27] C. Yang, L. Wang, W. Liang, and A. P. Moller, "How cuckoos find and choose host nests for parasitism," *Behavioral Ecology*, vol. 28, no. 3, pp. 859–865, 2017, doi: 10.1093/beheco/arx049.
[28] L. Wang, C. Yang, G. He, W. Liang, and A. P. Moller, "Cuckoos use host egg number to choose host nests for parasitism," *Proceedings of the Royal Society B: Biological Sciences*, 2020, vol. 287, no. 1928, doi: 10.1098/rspb.2020.0343.
[29] D. Shen, J. D. Ruvini, M. Somaiya, and N. Sundaresan, "Item Categorization in the e-Commerce Domain," in *CIKM '11: Proceedings of the 20th ACM international conference on Information and knowledge management*, 2011, pp. 1921–1924, doi: 10.1145/2063576.2063855.
[30] T.-P. Hong, C.-W. Lin, K.-T. Yang, and S.-L. Wang, "Using TF-IDF to hide sensitive itemsets," *Applied Intelligence*, vol. 38, no. 4, pp. 502–510, 2013, doi: 10.1007/s10489-012-0377-5.
[31] Q. Zhang, N. Koudas, D. Srivastava, and T. Yu, "Aggregate query answering on anonymized tables," in *2007 IEEE 23rd International Conference on Data Engineering*, Istanbul, Turkey, 2007, pp. 116-125, doi: 10.1109/ICDE.2007.367857.
[32] T. Y. Prawira, S. Sunardi, and A. Fadlil, "Market Basket Analysis To Identify Stock Handling Patterns & Item Arrangement Patterns Using Apriori Algorithms," *Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika*, vol. 6, no. 1, pp. 33–41, 2020, doi: 10.23917/khif.v6i1.8628.
[33] O. Gkountouna, K. Lepenioti, and M. Terrovitis, "Privacy against aggregate knowledge attacks," in *2013 IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, Brisbane, QLD, Australia, 2013, pp. 99-103, doi: 10.1109/ICDEW.2013.6547435.
[34] E. Adar, "User 4xxxxx9: Anonymizing query logs," Query Logs Workshop WWW, 2007.
[35] M. Terrovitis, N. Mamoulis, and P. Kalnis, "Local and global recoding methods for anonymizing setvalued data," *The VLDB Journal*, vol. 20, no. 1, pp. 83–106, 2011, doi: 10.1007/s00778-010-0192-8.
[36] L. Zhang,W.Wang, and Y. Zhang, "Privacy Preserving Association Rule Mining: Taxonomy, Techniques, and Metrics," *IEEE Access*, vol. 7, pp. 45032-45047, 2019, doi: 10.1109/ACCESS.2019.2908452.
[37] S. Oliveira and O. Zaiane, "Privacy preserving frequent itemset mining," in *CRPIT '14: Proceedings of the IEEE International Conference on Privacy, Security and Data Mining*, 2002, vol. 14, pp. 43–54.
[38] J. C.-W. Lin, T.-Y. Wu, P. Fournier-Viger, G. Lin, J. Zhan, and M. Voznak, "Fast algorithms for hiding sensitive high-utility itemsets in privacy-preserving utility mining," *Engineering Applications of Artificial Intelligence*, vol. 55, pp. 269–284, 2016, doi: 10.1016/j.engappai.2016.07.003.

BIOGRAPHIES OF AUTHORS

Dedi Gunawan is an Associate Professor in Department of Informatics Engineering, Universitas Muhammadiyah Surakarta. He received B.Eng. from electrical engineering at Universitas Muhammadiyah Surakarta. In 2014 he obtained Master degree in Computer Science and Information Engineering from National Dong Hwa University, Taiwan. He also received a Ph.D. degree from Kanazawa University, Japan in 2019. Currently, his research interests include data mining, machine learning, privacy enhancing technology, data anonymity, privacy preserving data mining, and privacy preserving data publishing. He can be contacted at email: dedi.gunawan@ums.ac.id.

Yusuf Sulistyo Nugroho is an Assistant Professor at the Department of Informatics, Universitas Muhammadiyah Surakarta, Indonesia. He received his Ph.D. degree from the Nara Institute of Science and Technology, Japan in 2020. His research interests include empirical software engineering, software documentation, and mining software repositories. He can be contacted at email: yusuf.nugroho@ums.ac.id.

Fatah Yasin Al Irsyadi is an Assistant Professor at the Department of Informatics, Universitas Muhammadiyah Surakarta, Indonesia. He received his Master degree from Universitas Gadjah Mada. His research interests include VR/AR-based game education, game education for inclusive education, and Islamic game education. He can be contacted at email: fatah.yasin@ums.ac.id.

Diah Priyawati is an Assistant Professor in Department of Informatics Engineering, Universitas Muhammadiyah Surakarta. She received B.Eng. from the department of electrical engineering at Universitas Muhammadiyah Surakarta. In 2015 she obtained Master degree in electrical engineering from Universitas Gadjah Mada, Indonesia. Currently her research interests including machine learning, image processing, software engineering, and software testing. She can be contacted at email: diah.priyawati@ums.ac.id.

Arini Nur Rohmah received Bachelor of Computer (Cumlaude) from Universitas Muhammadiyah Surakarta, Indonesia. In 2023, she was awarded three national degree certificates in accordance with a trilateral degree agreement from Università degli Studi dell'Aquila UnivAQ, Italy, for Master of Computer Science, Lappeenranta-Lahti University of Technology LUT, Finland, for Master of Science (Technology) (Distinction), and from Vrije Universiteit Amsterdam VUA, the Netherland, for Master of Science in Computer Science, as a graduate student of Software Engineers for Green Deal (SE4GD), Erasmus Mundus Joint Master Degree (EMJMD) program. Her thesis was in collaboration with Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Spain. Currently, she works as a lecturer at the Department of Informatics, Universitas Muhammadiyah Surakarta, Indonesia. Her research interests are related to software engineering, particularly in verification tools, human computer interaction, interaction design, sustainable and inclusive software, as well as health informatics and digital medicine. She can be contacted at email: arininrohmah@ums.ac.id.

Bambang Sukoco completed a Bachelor of Laws at the Faculty of Law, Muhammadiyah University of Surakarta and a Masters in Law at the Master of Laws at the Islamic University of Indonesia. Currently active as a Lecturer at the Faculty of Law at the Muhammadiyah University of Surakarta, General Secretary of the Central Leadership of the Alumni Association of the Muhammadiyah University of Surakarta and Head of the Legal Department of Association Services & General of the Rhetoric Bureau of the Muhammadiyah University of Surakarta. Apart from that, he is active in various organizations/communities, research activities, and community services. He can be contacted at email: bambang.sukoco@ums.ac.id.

Syful Islam is working as an assistant professor in the Department of Computer Science and Engineering, Faculty of Engineering at Gopalganj Science and Technology University, Gopalganj, Bangladesh. He received his M.E. and Ph.D. degrees from the Nara Institute of Science and Technology, Japan. His research interests include software ecosystems, mining stack overflow, and mining software repositories. He can be contacted at email: syfulcse@gstu.edu.bd.