Hindawi
Scientific Programming
Volume 2021, Article ID 3611182, 9 pages
https://doi.org/10.1155/2021/3611182

Research Article

A Novel Method to Solve Real Time Security Issues in Software Industry Using Advanced Cryptographic Techniques

B. Gobinathan,[1] **M. A. Mukunthan**,[2] **S. Surendran**,[3] **K. Somasundaram**,[4] **Syed Abdul Moeed**,[5] **P. Niranjan**,[5] **V. Gouthami**,[5] **G. Ashmitha**,[5] **Gouse Baig Mohammad**,[6] **V. K. Shanmuganathan**,[7] **Yuvaraj Natarajan**,[8] **K. Srihari**,[9] **and Venkatesa Prabhu Sundramurthy**[10]

[1]*Department of Computer Science and Engineering, Jaya Sakthi Engineering College, Chennai 602 024, Tamil Nadu, India*
[2]*Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai 600 062, Tamilnadu, India*
[3]*Department of Computer Science and Engineering, Tagore Engineering College, Chennai 600127, TamilNadu, India*
[4]*Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamilnadu, India*
[5]*Department of CSE, Kakatiya Institute of Technology and Science, Warangal, Telangana, India*
[6]*Department of Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad, India*
[7]*Dept of Mechanical Engineering, J.N.N. Institute of Engineering, Chennai, India*
[8]*ICT Academy, Chennai, India*
[9]*Department of Cse, Snsct, Coimbatore, India*
[10]*Department of Chemical Engineering, Addis Ababa Science and Technology University, Ethiopia*

Correspondence should be addressed to Venkatesa Prabhu Sundramurthy; venkatesa.prabhu@aastu.edu.et

Received 22 September 2021; Revised 19 October 2021; Accepted 21 October 2021; Published 28 December 2021

Academic Editor: M. Pallikonda Rajasekaran

In recent times, the utility and privacy are trade-off factors with the performance of one factor tends to sacrifice the other. Therefore, the dataset cannot be published without privacy. It is henceforth crucial to maintain an equilibrium between the utility and privacy of data. In this paper, a novel technique on trade-off between the utility and privacy is developed, where the former is developed with a metaheuristic algorithm and the latter is developed using a cryptographic model. The utility is carried out with the process of clustering, and the privacy model encrypts and decrypts the model. At first, the input datasets are clustered, and after clustering, the privacy of data is maintained. The simulation is conducted on the manufacturing datasets over various existing models. The results show that the proposed model shows improved clustering accuracy and data privacy than the existing models. The evaluation with the proposed model shows a trade-off privacy preservation and utility clustering in smart manufacturing datasets.

1. Introduction

Data anonymization [1] is considered as a major factor for data preservation in data mining. The result maintains the data privacy and sensitivity of data for data publishing. Since a large portion of the information distributing or information mining schemes go through linkage assaults [2] or probabilistic inference assault [3], the anonymization process can either be achieved through attributes suppression or generalization. This technique either causes data loss or can lead to NP-hard problem. Therefore, it is essential to use privacy preservation models with lightweight mechanism to avoid various constraints like NP-hard and data loss problems.

While maintaining the privacy, it is essential to address the data utility as it maintains the accuracy of the context

and can provide data usefulness while displaying the dataset to the users. Elimination of appropriate data leads to diminishing the sensitive information [4]. Therefore, it is essential to maintain the trade-off between the utility and privacy metrics in data pulsing schemes [5].

Issues associated with privacy in statistical database have been investigated thoroughly. The privacy-preserving techniques in data mining have been the active research study. There are various privacy-preserving clustering methods [4, 6–22], which are used as privacy preserved data mining models to secure the data and to balance it with data utility for providing equilibrium. In all these techniques, the frequency of sensitive information formed from the cluster is not uniform. It is found that certain information unlike original datasets exists more rapidly in each cluster. This leads to probabilistic inference attack. Hence, it is very necessary to use uniformly distributed clustered sensitive information. Therefore, the distribution of data records using a specific value of sensitive attribute among all classes is needed based on a neighborhood method. This could lead to the attainment of anonymized clusters with sensitive information that are distributed uniformly. In this regard, we develop a model that maintains the sensitivity of data while considering privacy and utility as its atmost concerns. The main contributions of the paper are given as follows:

(1) The authors develop a novel trade-off scheme that maintains the utility-privacy of smart manufacturing data

(2) The utility is handled by a metaheuristic algorithm, and the privacy is maintained using a lightweight cryptographic algorithm

(3) Finally, the model is tested over various smart manufacturing datasets in order of validating the model over privacy-utility metrics.

The outline of the paper is given as follows. Section 2 provides the related works. Section 3 provides the data anonymization and data clustering technique that include preliminaries and the algorithm required to resolve the problem. Section 4 provides the evaluation of the proposed work with existing methods. Section 5 concludes the work with possible directions of the future.

2. Related Works

Liu et al. [23] proposed the minimum monotonicity of privacy-utility which constitutes differential privacy limits. The study first defined the restricted concession of varying privacy and suggested limited data protection monotony based on technological irrelevance. The study also has shown in theory various distinct data protection structures.

Ruddell et al. [24] offered a statistical overview of the effectiveness and privacy relationship in Los Angeles. The trade depends on the economic field and type of services. This study provides guidelines for the better and more ethically balanced collection and distribution of customer knowledge.

Cho et al. [25] offered a guideline for automated biomedical databases, optimising the concept of differentiating privacy from meaning.

Asikis and Pournaras [26] studied the requirements for secrecy which balance maximum utility, minimal privacy and low consumption, and maximal privacy where the utility of aggregation calculations relates to accuracy. Data security settings should be universally applied as system-wide rules and checks. Each customer may, however, often be individually enforced or decided by offers.

Valdez, and Ziefle [10] explored how personal health information sharing is required and beneficial to protect security and confidentiality. The data exchange value, mode of data, and privacy are maintained by the data recipient, and users were asked to select the ideal communication scenario. The use of data for clinical studies and physical disorders was objecting to the sharing of mental disease and heavy de-anonymisation data for commercial purposes but was not interested.

3. ECC-ACO Algorithm

The ECC-ACO algorithm is illustrated in Figure 1. The study applies the privacy of datasets using the lightweight ECC algorithm prior the formation of cluster and its relevant classes. The clusters are formed in such a manner that the formed cluster has reduced clustering error. To validate the clustered samples, we use classification to classify the relevant class samples. The data are transmitted or published if there exist a reduced classification error than the threshold level. The privacy of data is not compromised or adjusted based on the clustered instances for maintaining the utility.

3.1. Preliminaries. The datasets published by the owners usually have private and public attributes. The public attributes are known among all users, and it is publically available.

Formally, let I represent M items with proper subsets $\rho(I)$. Consider an itemset X, which is regarded as the subset of the itemset I, where $X \subseteq I$. The transaction T is represented in the form of tuple (t_{id}, X), where X is the itemset and t_{id} differentiates the transactions with same items.

Item set X is supported for the purpose of transaction T if $X \subseteq T$. The transactional database D over the itemset I with N transactions and the T_{nm} value is assigned for 1 of n^{th} transaction and has an m item or else T_{nm} is 0.

The probability of intruders to reveal the user identity through the publically available attributes is called as QID attributes. At the time of data publishing, certain attributes are kept confidential, and these data are called as sensitive attributes (SA). The datasets usually possess the information like gender, age, and zipcode, and they are called as QID attributes. Further, the disease of an individual is called as sensitive attribute, and the identity of a person is a public attribute. It is easier for an attacker to reveal the private or sensitive user information without their knowledge. Knowing the QID information of a neighbor, the attack can link with their disease, and such attack is called as linking

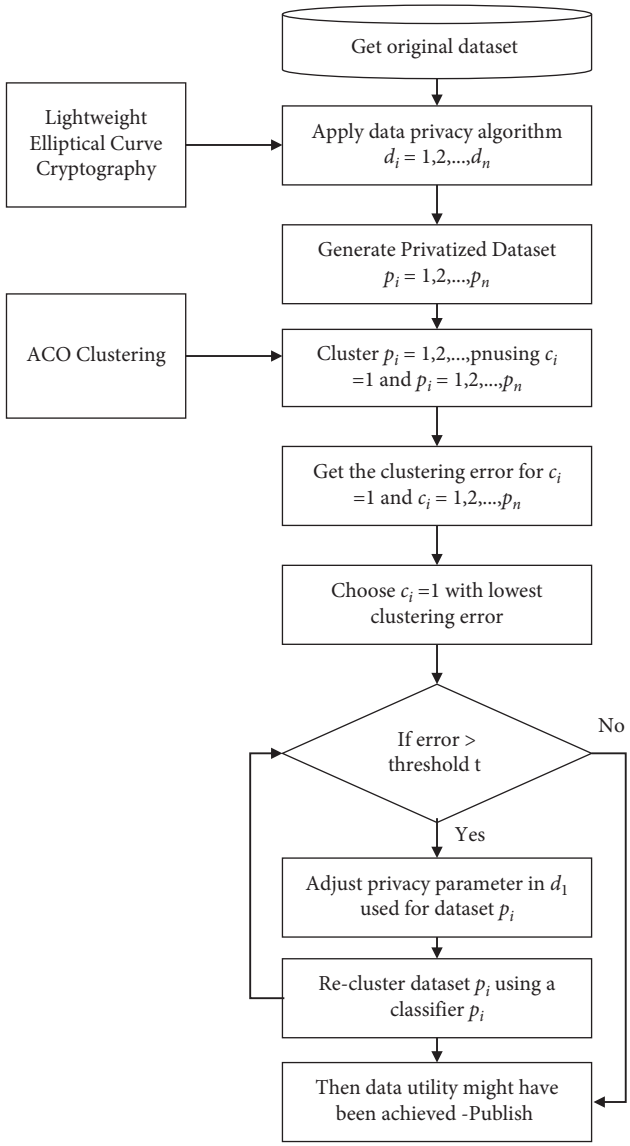

Figure 1: Overview of ECC-ACO.

attack or identity disclosure attack. However, the conclusion of disease may change, and this is linked with another attack named as probabilistic inference attack. This attack is common in place of a more frequent occurrence of one sensitive attribute information than others. This can lead to an assumption of predicting the most common disease based on QID values. The proposed clustering method developed in this paper provides better resilience against the above two attacks.

3.2. ACO Clustering Process. The data clustering process or the partitioning process is given in [20] in the form of suitable steps, which is given in Algorithm 1. After the partitioning of data into clusters, the size of each cluster may vary. Hence, it is very necessary to adjust the size of clusters to obtain k-anonymity. The cluster value lesser than k value is merged with other clusters. This reduces the deterministic penalty since the combinations of all the clusters are same, which is given in Algorithm 2.

To improve the distributed clusters uniformity, the study uses a K-nearest neighbor classifier after clustering the datasets using ACO. The classifier generates the equivalent classes of clustered samples with uniformity, and hence the information provided for data privacy is considered sensitive. In each cluster, KNN finds the equivalent class of best neighbor by maintaining the sensitive information of each data sample. Algorithm 3 provides the pseudocode of ECC-ACO with classifier.

4. Results and Discussion

The implementation of ECC-ACO is carried out in Java on Eclipse IDE. The study considers two different experiments over smart manufacturing datasets. The ECC-ACO is tested with Smart Energy Data: Aachen/Cologne Smart Factory [21] and Power System Datasets [22]. It can effectively be applied on power system analysis or any application that generates rapid data. The model is tested using three different metrics that include average equivalence class size after classification, discernibility cost, and KullbackLeibler (KL) divergence. Further, the performance of utility is estimated in terms of F-measure, accuracy, and execution time.

4.1. Cluster Output. The conventional methods work on anonymization, which forms equivalence class or groups; however, it fails to provide data related to cluster size. The model uses both these parameters as the cluster significantly provides the difference between ACO and ECC-ACO clustering algorithm based on cluster size. The experiments are conducted on synthetic dataset and adult dataset case one. Further, the results provide the cluster size formed by ACO.

The study considers six different parameters that include the total number of ants (m), a parameter to control the pheromone τ for the sample selection choice probability α, a parameter to control the local heuristic factor η for the sample selection choice probability β, the pheromone persistence coefficient ρ, the pheromone reward factor R, and the pheromone penalty factor P_p.

Figure 2 provides the total number of clusters and cluster frequency of case one. Efficient clusters are created by using the proposed ant colony optimization (ACO) clustering scheme over the given datasets with different sensitive values say 5, 10, and 15 shown in (Figures 2(a)–2(c)) with a total of 10000 records and fixed k value. In this experiment, the value of S is varied in order to provide variations between various clusters as per different frequencies. Similarly, the results are obtained by varying the value of k, i.e., 50 (Figure 2(d)) and 100 (Figure 2(e)) with fixed S value. Finally, Figure 2(f) shows the clusters formed by elliptic curve cryptography-ant colony optimization (ECC-ACO) algorithm with fixed k (=300) value and fixed S (=2) value, i.e., case one. The results show that ECC-ACO produces equivalence clusters of

Input: privacy datasets
Output: unequally clustered datasets
Step 1. Pheromone matrix initialization, where the matrix elements are set with small values.
Step 2. Ants initialization

(a) Start initial process of iteration
(b) Reset the memory, cluster center matrix, and weight matrix

Step 3. Random selection of data object (other than memory list)
Step 4. Cluster selection using the process of exploitation and exploration

(a) Exploitation using the greedy algorithm
(b) Exploration allots probabilities of nodes and chooses the higher probability ant in stochastic manner
(c) This is regarded as the cluster center, which is used for forming the cluster

Step 5. Update the ants

(a) Update the memory, cluster center matrix, and weight matrix
(b) If the memory list is not full, go back to Step 3
(c) Else continue with the process
(d) End

Step 6. Find the fitness
Step 7. Update the value of pheromone
Step 8. Check end condition

(a) If total iterations > maximum iterations

(i) End the process

(b) Else go to Step 6
(c) End

Algorithm 1: Partitioning step.

Input: unequal clusters from the ACO algorithm
Output: k-anonymity table
Step 1. For each cluster

(a) Find the size of cluster

Step 2. If the size of cluster is lesser than maximum value of size of a cluster

(a) Merge the clusters

Step 3. Else
Step 4. Divide the groups into subgroups, s.t. each groups will have at least k-tuples
Step 5. End

Algorithm 2: Adjusting step.

Input: privacy datasets
Output: classified sensitive samples
Step 1. Sort private dataset using sensitive attributes
Step 2. Split the sorted datasets into subgroups
Step 3. If there exist identical attributes

(a) Cluster using ACO

Step 4. Else

(a) Go to Step 1

Step 5. End
Step 6. Repeat over other clusters

Algorithm 3: Continued.

(a) If the size of cluster size = Dmin
(b) Distribute the data sample among the clusters
(c) Create single cluster element
(d) Add nearest neighbor in the cluster
(e) Eliminate unwanted instances
(f) End

Step 7. In case two, the value of S is greater than K

(a) Distribute the instances to all clusters, equally
(b) Add the record from Dmin to single element cluster
(c) Follow Step 6(c)
(d) Split Dmin values from each cluster
(e) Follow Step 8
(f) Compete the KN values and add the neighboring values to each cluster similar to case one
(g) Replace QID values with center values in each cluster to anonymize the data clusters.
(h) Repeat the process

Step 8. Calculate the centroid
Step 9. Use Euclidean distance to calculate nearest instances along the centroid in clusters
Step 10. Classify the similar instances using classifiers
Step 11. Merge the equivalent class samples

Algorithm 3: The pseudocode of ECC-ACO with classifier.

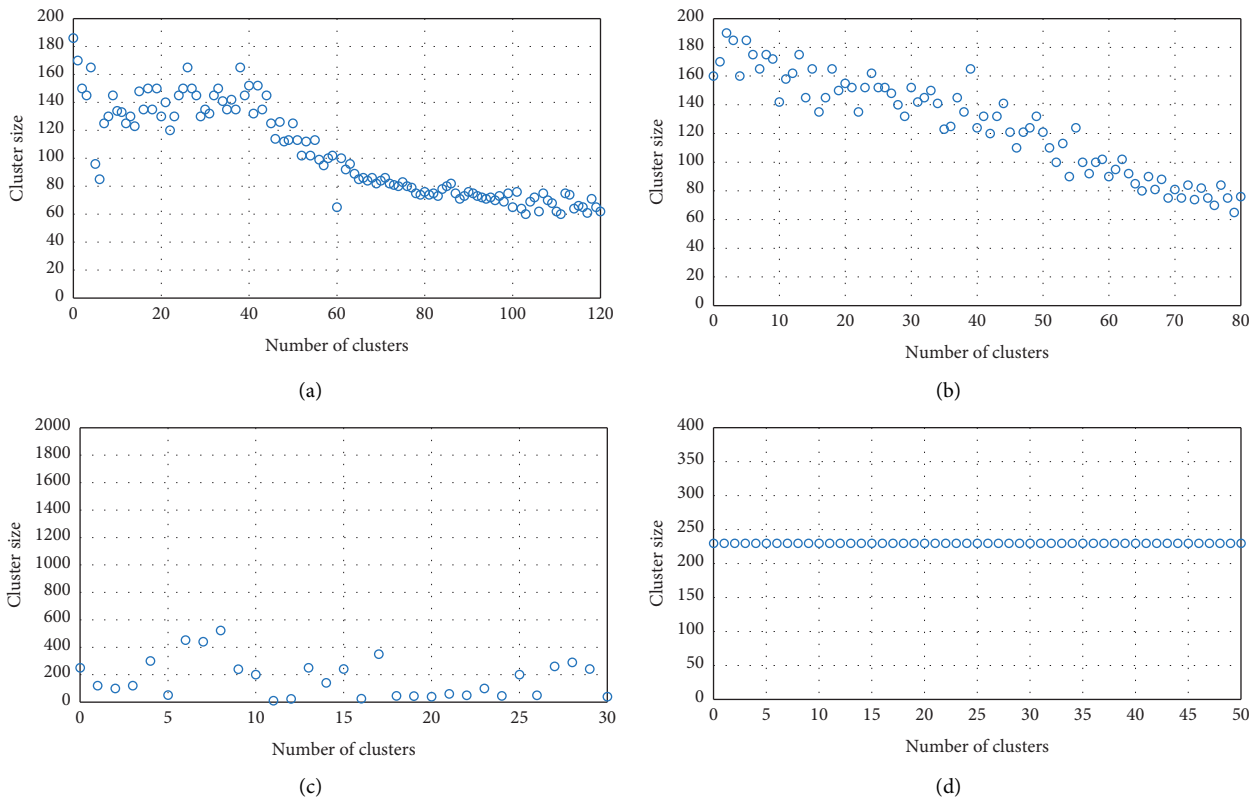

Figure 2: Continued.

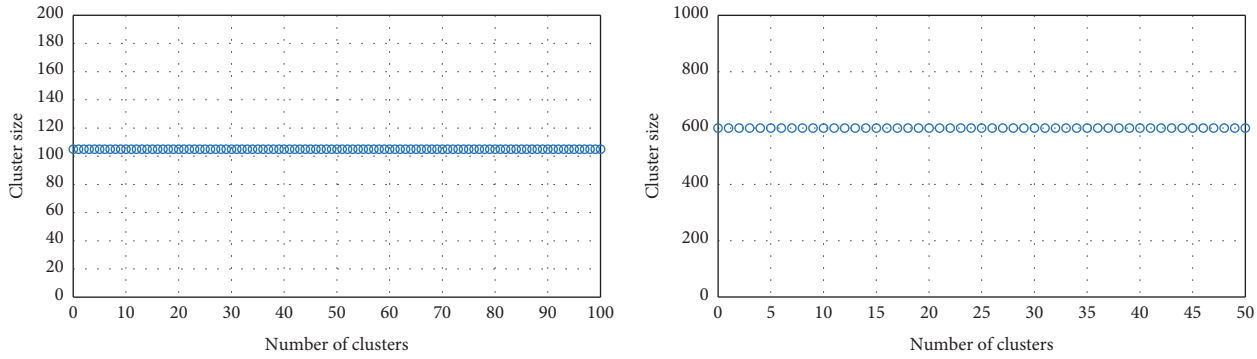

Figure 2: Clustered output for proposed algorithms on all three datasets: (a) $s = 15$ and $k = 50$; (b) $s = 10$ and $k = 50$; (c) $s = 5$ and $k = 50$; (d) $s = 100$ and $k = 10$; (e) $s = 10$ and $k = 50$; (f) $k = 300$.

Table 1: Accuracy on dataset 1.

K	4		10		14		40	
Algorithm	J48 classifier	NB classifier	J48 classifier	NB classifier HDFS	J48 classifier	NB classifier	J48 classifier	NB classifier
Incognito	84.2	81.64	84.2	81.33	84.2	81.6	84.2	81.6
Mondrian	84	81.82	83.96	81.82	83.99	81.83	84.04	81.84
Datafly	84.24	81.84	84.24	81.84	84.28	81.84	84.26	82
ECC-ACO	84.6	81.83	84.6	81.83	84.6	81.83	84.6	81.83

differing sizes, whereas the ECC-ACO algorithm forms uniform size.

Finally, the ECC-ACO is compared with models in terms of detection accuracy and F-measure. Here, the study considers the classification of clustered samples in order to find the accurate classes. The above two measures define the total number of correctly classified instances classes with classifiers like Naïve Bayes and J48 and experimented over smart manufacturing dataset. The experiments on case one and case two are carried out on HDFS.

4.2. Comparison with Existing Methods. Tables 1 and 2 show the results of percentage of optimal clustering of smart manufacturing data instances with varying values of k, say 5, 10, 25, and 30. The results of optimal clustering exhibit optimal results with smart manufacturing data while the instances are generated using the two different classifiers after the process of clustering that includes Naive Bayes and J48 classifier.

Tables 3 and 4 show the results of classified instances using the classifiers with varying values of k. The Hadoop system is improved by varying the parameters in order to increase the classification accuracy. The result shows that significant improvement is noted with increased random shuffling prior data chunk creation. It is seen that the performance of the ECC-ACO algorithm exhibits closer and optimal performance with J48 classifier. In case of Naive Bayes, the ACO clustering algorithm obtains improved performance than other algorithms. The overall performance shows that ECC-ACO exhibits better classification accuracy and F-measure than existing algorithms.

4.3. Data Utility and Privacy Metrics. The model is tested with three different metrics that include equivalence class size, discernibility cost, and KullbackLeibler divergence to find the utility losses associated with the smart manufacturing datasets on the privacy preserved dataset. It is seen that the proposed model acquires reduced loss in data than other methods. Tables 5 and 6 show the details of discernibility cost with various experiments conducted on different datasets. The results on the model show optimal discernibility cost than other models. Also, it shows a reduced cost than other methods, which shows that the model is better than other with optimal performance on data utility.

Table 7 shows average values of class size on dataset 1, and Table 8 shows average values of class size on dataset 1.

Table 9 shows the results of degree of privacy for the manufacturing datasets with increasing clustering values. The results show that the ECC-ACO achieves higher degree of privacy with increasing k samples than the ACO model.

4.4. Analysis. The result of evaluation shows the level of confidence of P^{-1} that helps to maintain the privacy of data even where there exists an attack on the data. The study considers an individual QID attribute that offers a sensitive attribute to the given unique clustered sample. The results also show that the probability of finding the clustered data samples provides improved data utility using ACO

Table 2: F-measure on dataset 1.

K	4		10		14		40	
Algorithm	J48 classifier	NB classifier	J48 classifier	NB classifier HDFS	J48 classifier	NB classifier	J48 classifier	NB classifier
Incognito	0.836	0.808	0.838	0.801	0.836	0.806	0.836	0.806
Mondrian	0.832	0.808	0.83	0.808	0.83	0.808	0.831	0.808
Datafly	0.84	0.809	0.839	0.809	0.839	0.809	0.838	0.814
ECC-ACO	0.842	0.809	0.842	0.809	0.842	0.809	0.842	0.809

Table 3: Accuracy on dataset 2.

K	4		10		14		40	
Algorithm	J48 classifier	NB classifier	J48 classifier	NB classifier HDFS	J48 classifier	NB classifier	J48 classifier	NB classifier
Incognito	82.99	81.8	83.03	81.3	83.04	81.8	83.04	81.8
Mondrian	83	81.94	83.02	81.95	83.06	81.94	83.02	81.93
Datafly	83.2	81.96	83.21	81.96	83.2	81.98	83.2	81.98
ECC-ACO	85.32	81.84	85.32	81.84	85.32	81.84	85.32	81.84

Table 4: F-measure on dataset 2.

K	4		10		14		40	
Algorithm	J48 classifier	NB classifier	J48 classifier	NB classifier HDFS	J48 classifier	NB classifier	J48 classifier	NB classifier
Incognito	0.818	0.809	0.819	0.81	0.819	0.809	0.819	0.809
Mondrian	0.818	0.809	0.818	0.809	0.819	0.809	0.818	0.809
Datafly	0.821	0.81	0.821	0.81	0.821	0.81	0.821	0.81
ECC-ACO	0.849	0.808	0.849	0.808	0.849	0.808	0.849	0.808

Table 5: Discernibility cost ($\times 10^8$) on dataset 1.

k	GA	ACO	PSO	BSO	ECC-ACO
50	0.6	0.83	2.35	0.21	0.15
100	6.9	0.93	2.43	0.23	0.15
150	6.8	1.23	1.33	0.23	0.15
200	6.83	0.6	1.32	0.24	0.15
250	3.25	0.32	1.2	0.22	0.15
300	3.33	0.21	0.9	0.25	0.15

Table 6: Discernibility cost ($\times 10^8$) on dataset 2.

k	GA	ACO	PSO	BSO	ECC-ACO
2	2.46	1.32	2.43	1.21	1
5	2.46	1.32	2.43	1.21	1
10	2.46	1.32	2.43	1.21	1
25	2.46	1.32	2.43	2.23	2.1
50	2.46	1.32	2.43	4.65	4.65

clustering, and it proves an improved clustering result than other models.

The study also observes that the trade-off often provides the smart manufacturing datasets with improved data utility and privacy. With increasing cluster size, the trade-off between the utility and privacy is hence formed without comprising the trade-off between them. However, with increased dataset size, the data privacy tends to get reduced that tends to get compromised while the data utility is maintained. The reduction of data privacy also occurs with increased number of data samples in the smart manufacturing datasets.

Table 7: Average values of class size on dataset 1.

k	GA	ACO	PSO	BSO	ECC-ACO
50	60	7	75	2	2
100	60	6	38	2	2
150	40	5	50	2	2
200	30	5	38	2	2
250	30	5	18	2	2
300	25	5	15	2	2

Table 8: Average values of class size on dataset 1.

k	GA	ACO	PSO	BSO	ECC-ACO
2	845	12	1932	729	729
5	852	12	874	623	623
10	231	12	323	187	187
25	121	12	241	95	95
50	64	12	173	14	14

Table 9: Privacy degree.

Case	k	ACO	ECC-ACO
Dataset 1	2	1423	3227
	5	1423	3227
	10	1423	3227
	15	1365	3234
	25	2476	3875
Dataset 2	50	3	5
	100	3	5
	150	3	5
	200	3	5
	250	3	5
	300	3	5

5. Conclusion

In this paper, the trade-off between the utility and privacy of smart manufacturing datasets is maintained, and it is provided in an optimal manner. The utility through ACO clustering and privacy through ECC clustering offers an optimal trade-off between them such that the model does not sacrifice one factor for the other. The extensive evaluation conducted on the smart manufacturing datasets shows reduced cost, optimal divergence, and accurate class size. It further opts for an improved privacy of data while considering the utility metrics like F-measure and clustering accuracy. The privacy result proves that the data are distributed in a secured way, and the sensitive information is protected against various attacks. The ECC-ACO is not algorithm specific; however, on any data type with the proposed model, it can be applied. It is highly scalable to larger datasets. Also, the major limitation of the study is that the method can be applied on a domain-specific and not on multiple applications. In future, the ECC-ACO can be used to tune the running time as it is directly affected by the formation of increased number of clusters. The privacy of datasets can be improved for other attacks.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] L. Sweeney, "k-anonymity: a model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[2] W. Y. Lin, D. C. Yang, and J. T. Wang, "Privacy preserving data anonymization of spontaneous ADE reporting system dataset," *BMC Medical Informatics and Decision Making*, vol. 16, no. 1, p. 58, 2016.

[3] J. Soria-Comas, J. Domingo-Ferrer, D. Sanchez, and S. Martinez, "t-closeness through microaggregation: strict privacy with enhanced utility preservation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3098–3110, 2015.

[4] K. Mivule, C. Turner, and S.-Y. Ji, "Towards A differential privacy and utility preserving machine learning classifier," *Procedia Computer Science*, vol. 12, pp. 176–181, 2012.

[5] R. C. W. Wong, A. W. C. Fu, K. Wang, and J. Pei, "Minimality attack in privacy preserving data publishing," in *Proceedings of the 33rd international conference on Very large data bases*, pp. 543–554, VLDB Endowment, Vienna Austria, September 2007.

[6] K. Keke Chen and L. Ling Liu, "Privacy-preserving multiparty collaborative mining with geometric data perturbation," *IEEE Transactions on Parallel and Distributed Systems*, vol. 20, no. 12, pp. 1764–1776, 2009.

[7] I. Polato, R. Ré, A. Goldman, and F. Kon, "A comprehensive view of Hadoop research-A systematic literature review," *Journal of Network and Computer Applications*, vol. 46, pp. 1–25, 2014.

[8] T. Asikis and E. Pournaras, "Optimization of privacy-utility trade-offs under informational self-determination," 2017, https://arxiv.org/abs/1710.03186.

[9] J. Joy, D. Gray, C. McGoldrick, and M. Gerla, "K privacy: towards improving privacy strength while preserving utility," *Ad Hoc Networks*, vol. 80, pp. 16–30, 2018.

[10] A. Calero Valdez and M. Ziefle, "The users' perspective on the privacy-utility trade-offs in health recommender systems," *International Journal of Human-Computer Studies*, vol. 121, pp. 108–121, 2019.

[11] A. Majeed, "Attribute-centric anonymization scheme for improving user privacy and utility of publishing e-health data," *Journal of King Saud University - Computer and Information Sciences*, vol. 31, no. 4, pp. 426–435, 2019.

[12] J. C.-W. Lin, W. Gan, P. Fournier-Viger et al., "High utility-itemset mining and privacy-preserving utility mining," *Perspectives in Science*, vol. 7, pp. 74–80, 2016.

[13] 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.

[14] U. Yun and 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, 2015.

[15] A. W.-C. Fu, K. Wang, R. C.-W. Wong, J. Wang, and M. Jiang, "Small sum privacy and large sum utility in data publishing," *Journal of Biomedical Informatics*, vol. 50, pp. 20–31, 2014.

[16] K. Mivule and C. Turner, "A comparative analysis of data privacy and utility parameter adjustment, using machine learning classification as a gauge," *Procedia Computer Science*, vol. 20, pp. 414–419, 2013.

[17] J.-S. Yeh and P.-C. Hsu, "HHUIF and MSICF: novel algorithms for privacy preserving utility mining," *Expert Systems with Applications*, vol. 37, no. 7, pp. 4779–4786, 2010.

[18] M. Rodriguez-Garcia, M. Batet, and D. Sánchez, "Utility-preserving privacy protection of nominal data sets via semantic rank swapping," *Information Fusion*, vol. 45, pp. 282–295, 2019.

[19] S. Gao, J. Ma, C. Sun, and X. Li, "Balancing trajectory privacy and data utility using a personalized anonymization model," *Journal of Network and Computer Applications*, vol. 38, pp. 125–134, 2014.

[20] Y. Kao and K. Cheng, "An ACO-based clustering algorithm," in *Proceedings of the International Workshop on Ant Colony Optimization and Swarm Intelligence*, pp. 340–347, Springer, Berlin, Heidelberg, September 2006.

[21] "Smart Energy data: aachen/Cologne smart factory," 2015, https://data.lab.fiware.org/dataset/smart_energy_data-_aachen__cologne_smart_factory.

[22] "Power system datasets and gas pipeline datasets," 2015, https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets.

[23] H. Liu, Z. Wu, C. Peng, F. Tian, and L. Lu, "Bounded privacy-utility monotonicity indicating bounded tradeoff of differential privacy mechanisms," *Theoretical Computer Science*, vol. 816, pp. 195–220, 2020.

[24] B. L. Ruddell, D. Cheng, E. D. Fournier, S. Pincetl, C. Potter, and R. Rushforth, "Guidance on the usability-privacy tradeoff for utility customer data aggregation," *Utilities Policy*, vol. 67, Article ID 101106, 2020.

[25] H. Cho, S. Simmons, R. Kim, and B. Berger, "Privacy-preserving biomedical database queries with optimal privacy-utility trade-offs," *Cell systems*, vol. 10, no. 5, pp. 408–416, 2020.

[26] T. Asikis and E. Pournaras, "Optimization of privacy-utility trade-offs under informational self-determination," *Future Generation Computer Systems*, vol. 109, pp. 488–499, 2018.