# Privacy-Preserving Association Rule Mining Considering Multi-objective through an Evolutionary Algorithm

Darshana H. Patel, Hiral Kotadiya, and Avani R. Vasant

*Abstract*—**The COVID-19 pandemic has led to an increase in digitization. With the strict social and physical distancing measures in place, new routines require accessing the internet for most online services which have led to the explosive growth of data. As a consequence, data mining technologies are used for the extraction of useful information from a huge compilation of such digital data. Thus, the desire to mine data from varied sources to discover behaviors and patterns among entities such as customers, diseases, and environmental conditions is on the rise which can be accomplished by association rule mining. However, such pattern discovery by association rule mining also discloses the personal information of an individual or organization. Thus, the challenge of association rule mining is privacy preservation wherein confidentiality of sensitive rules should be maintained while releasing the database of third parties. Privacy-preserving association rule mining is the process of modifying the original database to hide the sensitive rules for preserving privacy. Thus, the paper emphasizes multiple objectives like minimizing the side effects of hiding sensitive rules. i.e. reduce the number of ghost rules, lost rules, and hiding failure along with the increase in utility of the data.**

*Index Terms*—**Data mining, association rule mining, privacy-preserving association rule mining, evolutionary algorithm, genetic algorithm.**

## I. Introduction

With the technological revolution, a huge amount of data is being collected, and as a consequence, data mining technologies are used for the extraction of useful information from the huge compilation of digital data. However, this immense quantity of data, if publicly available, can be employed for growth and development as well as in several applications. In the field of data mining, an enormous amount of data is processed to acquire certain fruitful data. Major techniques of data mining are predictive and descriptive which can be further classified as classification, prediction, clustering, summarization, and association [1] as depicted in the below-given Fig. 1.

The Association functionality of data mining is gaining immense popularity as it determines the probability of the co-occurrence of items from raw data [2]. Then, association rules are formed based on relationships between these co-occurring items. These association rules are beneficial for analyzing and then foretelling the behavior of the customer. However, it has the shortcoming is that such raw data may sometimes reveal the private or sensitive information of an individual or organization. This will certainly discourage an individual to share their data. Privacy-preserving association rule mining offers the solution to limit this deficiency by hiding the sensitive rules [3]. As a result, privacy-preserving association rule mining has become an active research area.

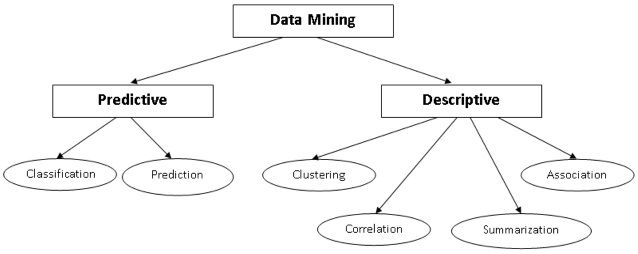

Fig. 1. Techniques of data mining.

### A. Motivation

Let us consider a scenario that exemplifies the necessity of applying association rule hiding algorithms to protect sensitive knowledge.

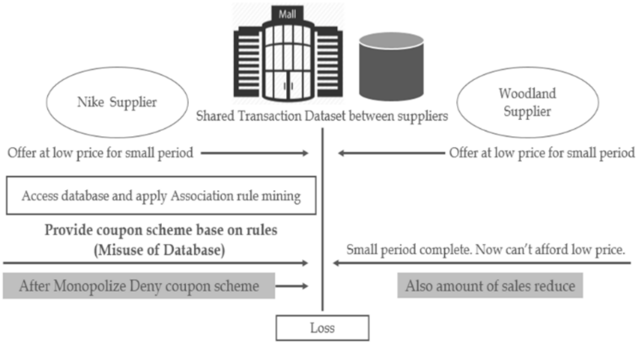
Mall
Nike Supplier
Shared Transaction Dataset between suppliers
Woodland Supplier
Offer at low price for small period
Offer at low price for small period
Access database and apply Association rule mining
Provide coupon scheme base on rules
(Misuse of Database)
Small period complete. Now can't afford low price.
After Monopolize Deny coupon scheme
Also amount of sales reduce
Loss

Fig. 2. Need of hiding sensitive rules.

The distribution of data between more parties is valuable but occasionally it releases private information. There are two denim suppliers namely Nike and Woodland as shown in Fig. 2. Suppose the Nike supplier fetches the database and finds some sensitive rules with high confidence. They offer their products at reduced prices, provided that access to the mall's database of customer purchases. By using an association rule mining tool, they started doing market basket analysis. Nike company now runs a coupon marketing campaign. The campaign cuts heavily into the sales of Woodland, which increases the prices, based on lower sales. Then because of that monopoly, Nike customers may increase and woodland brands cannot afford low prices because of the lack of customers. During the next negotiation with Nike, they are unwilling to offer a low price due to reduced competition. After monopolizing, Nike customers deny coupon schemes. So ultimately all-over swelling decreases and malls have to

Manuscript received July 29, 2021; revised October 16, 2021.

Darshana H. Patel is with the I. T. Department, V. V. P. Engineering College, Rajkot, India (e-mail: darshana.h.patel@gmail.com).

Hiral Kotadiya is with Weboccult Technologies at Ahmedabad, Gujarat, India.

Avani R. Vasant is with Babaria Institute of Technology, Vadodara, Gujarat, India.

DOI: 10.7763/IJCTE.2022.V14.1303

suffer loss. In other words, the aforementioned scenario indicates that a mall should sanitize competitive information (and other important corporate secrets of course) before delivering access to the database.

### B. Applications

Privacy-preserving association rule mining covers a wide range of applications [4], [5] such as marketing, health care, disease diagnoses, video surveillance, recommended system, catalog design, and store layout, retailing, banking, fraud detection, insurance, telecommunication, etc.

## II. Association Rule Mining

The task of association rule mining is to find the frequently occurring items or patterns amongst the database and mine them for fetching associations or relationships amongst such items or patterns [6]. Given a set of transactions, association rules can be obtained which will predict the occurrence of an item based on the occurrences of other items in the transaction. The common approaches used for association rule mining are Apriori and FP-growth [6].

### A. Apriori Algorithm

It is a significant algorithm for finding association rules based on frequent itemsets. Apriori uses a bottom-up approach and it is designed to operate on a database containing transactions as shown in Fig. 3 below.

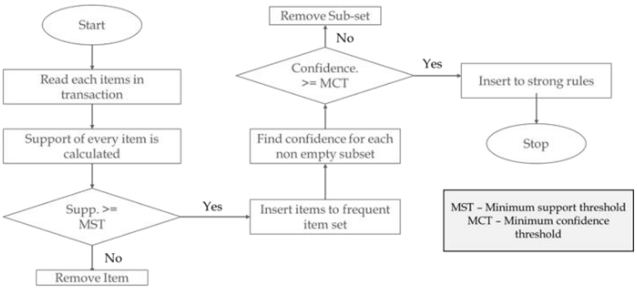
Start
Read each items in transaction
Support of every item is calculated
Supp. >= MST
Yes
No
Remove Item
Insert items to frequent item set
Find confidence for each non empty subset
Confidence. >= MCT
No
Remove Sub-set
Yes
Insert to strong rules
Stop
MST – Minimum support threshold
MCT – Minimum confidence threshold

Fig. 3. Apriori algorithm.

### B. FP-Growth Algorithm

The FP Growth algorithm was introduced in 2000 to discover frequent items without candidate generation. It uses a divide and conquer strategy to generate association rules. This process primarily works in the two-phase construction of frequent pattern trees and extraction of frequent items as shown in Fig. 4 below.

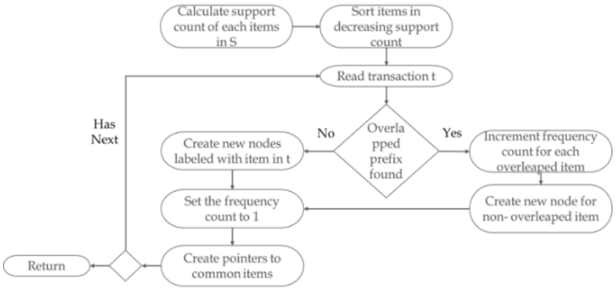
Calculate support count of each items in S
Sort items in decreasing support count
Read transaction t
Overla pped prefix found
No
Yes
Create new nodes labeled with item in t
Increment frequency count for each overleaped item
Set the frequency count to 1
Create new node for non- overleaped item
Create pointers to common items
Has Next
Return

Fig. 4. FP-Growth algorithm.

Thus, the main difference between apriori and FP growth are as given below:

a) Apriori uses a breadth-first search whereas FP growth practices the divide and conquer method.
b) Apriori discovers frequent items with a large number of candidate generations while FP growth generates frequent items without candidate generation.
c) Apriori algorithm requires enough memory space as compared to FP growth.
d) Apriori method executes multiple scans for generating a candidate set however FP growth scans the dataset twice only.

Both apriori and FP growth generate frequent items wherein association rules can be formed. In data mining, strong association rules are those that satisfy minimum support and minimum confidence threshold. Association rules are mostly measured based on two thresholds namely a) support and b) confidence.

a) Support: It denotes how frequently a set of items existence in the dataset [1]. It represents the popularity of an item set. Suppose there is $X$ item/itemset available in dataset D [1], then,

$$\text{Support of } X = \frac{Number\ of\ transaction\ in\ which\ X\ apperes}{Total\ number\ of\ transaction\ in\ D}$$

b) Confidence: It denotes the hood of one item purchase when the second item is purchased [1]. Suppose there are antecedent $X$ and Consequent $Y$ in association rule [1], then,

$$\text{Confidence of } X \quad Y = \frac{Support\ of\ (X\ U\ Y)}{Support\ of\ X}$$

## III. Privacy-Preserving Association Rule Mining

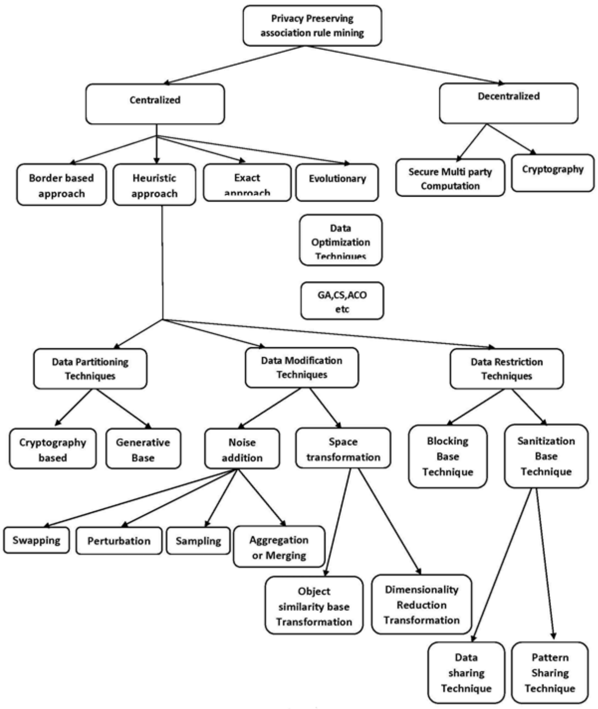
Privacy Preserving association rule mining
Centralized
Decentralized
Border based approach
Heuristic approach
Exact approach
Evolutionary
Data Optimization Techniques
GA,CS,ACO etc
Secure Multi party Computation
Cryptography
Data Partitioning Techniques
Data Modification Techniques
Data Restriction Techniques
Cryptography based
Generative Base
Noise addition
Space transformation
Blocking Base Technique
Sanitization Base Technique
Swapping
Perturbation
Sampling
Aggregation or Merging
Object similarity base Transformation
Dimensionality Reduction Transformation
Data sharing Technique
Pattern Sharing Technique

Fig. 5. Privacy-preserving association rule mining approaches.

Privacy-preserving association rule mining [7] aims to conceal the association rules in such a way that no sensitive or private information can be extracted from the database. In today’s world, privacy is a major concern and people are very much concerned about their sensitive information which they don’t want to share. Hence, protecting private information is becoming extremely crucial, which can be achieved through privacy-preserving association rule mining [8]. The taxonomy of privacy-preserving association rule mining is as depicted in below given Fig. 5.

For sharing data in privacy-preserving, considering in broader aspect, following two different scenarios [9] exist specifically:

a) Centralized Scenario: Firstly, the data owner conceals the micro-data by applying various PPDM techniques before publishing *it to the data miner which then performs different* data mining tasks on such concealed data. In this scenario, data owners/data miners are independent of managing privacy issues.

b) Distributed Scenario: The data owners can also be the miners and get collective outcomes on the amalgamation of their records. This is a situation where privacy is ensured on the results of data mining.

In a Centralized scenario, further classification can be done as Border, Heuristic, Exact and Evolutionary [10]-[12] whereas distributed scenario includes secure multiparty computation and Cryptographic approaches [13] for privacy-preserving of an association rule. The concept and approach along with the advantages and disadvantages of all these different methods are described below-given Table I:

TABLE I: COMPARISON OF DIFFERENT PRIVACY-PRESERVING ASSOCIATION RULE HIDING METHODS

| Method | Concept | Approach | Advantage | Disadvantage |
|---|---|---|---|---|
| Border | Border theory in which the upper border is frequent itemsets and lower border is the minimum of infrequent itemsets and the border is a union between them. | Reduce the support or confidence of rules | Maintains the quality of data | Border |
| Assistant channel | Channel 2 | Channel 3 | … | Channel 1 |
| Exact | A non-heuristic approach is based on constraint and hence one which satisfies the constraint is selected. | Linear or Integer programming | Less modification on the original data | High computational cost |
| Heuristic | Solving the problem by a practical method to produce solutions that may not be optimal but are sufficient considering the deadline. | Data modification or partitioning | Simple, Fast, and efficient | Undesirable side effects and does not provide the global optimal solution |
| Evolutionary | Based on aspects of natural evolution wherein it follows a randomized search procedure with a candidate set of random individual solutions called population. | Optimization algorithms | Lower computational cost | Provides optimal solution |
| Secure multiparty computation | Used for two or more parties for collectively performing computation by sharing the database | Functions such as secure sum and secure set union | Safeguards the susceptible data | More complex |
| Cryptography | Utilized for outsourcing the mined association rule to a cloud server. | Encryption and decryption algorithms | Secures the mining of association rules over | Cryptography |

## A. Related Work

The contextualization of relevant papers [14]-[19] with their approach, performance parameters, advantages, and open issues are mentioned below in this Table II:

TABLE II: RELATED WORK

| Paper Title & Year of Publication | Approach | Parameters | Advantages | Open Issues |
|---|---|---|---|---|
| Privacy-preserving in association rule mining using an improved discrete binary artificial bee colony & year 2019 [14] | Improved Binary Artificial Bee Colony (IBABC) approach | Execution time and the side effects such as hiding failure and the data utility | More effective and provides the balance between exploration and exploitation. | Computational time can be improved. |
| A Grid-Based Swarm Intelligence Algorithm for Privacy-Preserving Data Mining & year 2019 [15] | Grid-based method particle swarm optimization (GMPSO) | Run time, Missing cost, Artificial cost, Hiding Failure, Database dissimilarity. | Better effectiveness and efficiency as well as improves side effects other than this effect. | A minor side effect of database dissimilarity (Dis) can be improved. |
| The technique for Optimization of Association Rule Mining by Utilizing Genetic Algorithm & year 2019 [16] | Genetic Algorithm | Execution time and accuracy | Provides better accuracy. | Distributed scenarios and other evaluation parameters can be considered for implementation. |
| A multi-analysis on privacy preservation of association rules using hybridized approach & year 2019 [17] | Genetic Algorithm with Crow Search Algorithm (GA-CSA) | Hiding Failure, Information preservation rules and degree of modification | The sensitive data is restored efficiently. | The objective function comprises other parameters that can be deliberated. |
| Efficient Association Rules Hiding Using Genetic Algorithms & year 2018 [18] | Efficient association rules are hidden using a genetic algorithm (EARH-GA) | Execution time, utility, and accuracy. | Provides better accuracy and utility. | Consideration of hiding a set of rules in one optimization run instead of hiding one rule in every run can be further thought. |
| Association Rule Hiding using Cuckoo Optimization Algorithm & year 2016 [19] | Cuckoo optimization algorithm for the sensitive association rules hiding (COA4ARH). | Hiding failure, Lost rules, and ghost rules | Few side-effects and better performance as compared to other algorithms | Self-adaptive The mechanism can be added to the proposed approach for better performance. |

## IV. Preliminary Information and Problem Statement about Proposed Work

Let $D$ be a set of transactions in a dataset, denoted as $D=\{T_1, T_2,\ldots\ldots., T_n\}$. Each record $Tr$ is defined as a set of data items, $Tr=\{d_1,d_2,\ldots\ldots\ldots.d_k\}$, $S$ be a set of a sensitive item or sensitive pattern, denoted as $S=\{s_1, s_2, \ldots\ldots.., s_m\}$ and $D$' be the sanitized dataset.

Considering the data, find the frequent itemsets and interesting or strong rules by utilizing association functionality and then preserve the privacy of that rules along with reducing the side effects such as a lost rule, ghost rule, and hiding failure caused during preserving the privacy of association rules.

**Lost rules:** Some non-sensitive rules are falsely hidden during the process, known as lost rules [20].

$$\text{Lost rules} = \frac{\#non{-}sensitive\ rules\ in\ D\ {-}\#Non{-}sensitive\ rules\ in\ D`}{\#non{-}sensitive\ rules\ in\ D}$$

where D is the original dataset. $D$` is sanitized dataset.

**Ghost rules:** Some artificial rules are generated during this process known as ghost rules. [20]

$$\text{Ghost rules} = \frac{\#patterns\ in\ D`\ {-}\#Commomn\ pattern\ in\ D\ \&\ D`}{\#pattern\ in\ D}$$

where $D$ is the original dataset. $D$` is sanitized dataset.

**Hiding failure:** Some sensitive patterns remain as it is after the hiding process, we call this Hiding Failure [21].

$$\text{Hiding failure} = \frac{\#sensitive\ rules\ in\ D}{\#sensitive\ rules\ in\ D'}$$

where $D$ is the original dataset. $D$` is sanitized dataset.

**Data Utility:** Data utility is used to measure the effectiveness of the data sanitization algorithm and it is associated with the lost rules [21] which can be calculated as

Data utility = 1 – (number of lost rules/ number of rules)

**Genetic parameters:** The significant genetic parameters [22] required for implementing the proposed algorithm PPARM-GA are as follows:

a) Tournament selection: In Tournament selection two chromosomes are selected randomly from the population and more fit of these two is selected for the mating pool.

b) Single point crossover: The parent's chromosome is split into two portions such as head and tail. Similarly, the head of one chromosome combines with the tail of another chromosome in the mating pool. Crossover is usually applied in a GA with a high probability – pc. Empirical studies have shown that better results are achieved by a crossover probability of between 0.65 and 0.85.

c) Mutation operator: It attempts to introduce some random alteration of the genes. The mutation is usually applied in a GA with a low probability – pm. We should account for how our solution run is performing. That is if staked in local optima we should increase mutation or if it does not converge we should decrease mutation probability. Mutation probability ranges from 0.35 to 0.1.

d) Replacement operator: In this operator, some of the chromosomes of the initial population will replace with some of the chromosomes of offspring. In weak parent replacement, a weaker parent is replaced by a strong child

### A. Architectural View of Proposed Work

The generic view of privacy-preserving association rule mining is as shown in the below-given Fig. 6:

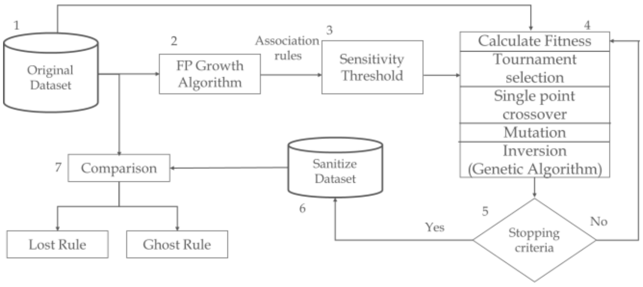

Fig. 6. The generic view of privacy-preserving association rule mining.

Here, as shown in the above Fig. 6 the original dataset is given as input. Then, the FP-Growth algorithm will be applied to generate strong association rules. Amongst those rules, having a certain sensitivity threshold will be provided to the proposed evolutionary algorithm namely PPARM-GA until the terminating condition gets satisfied. Due to which sanitized dataset will come into existence and hence comparing it with the original dataset several lost and ghost rules are generated as a side effect of hiding sensitive association rules. Thus, the main aim is to minimize this side effect and maintain the utility of data at the same time.

### B. An Example

Let's illustrate an example of a sample dataset. Following the steps provided in the workflow of the genetic algorithm leads to the below sequence of steps.

Step-1 Original dataset and sensitive pattern.

D = {T1, T2, T3, T4}

T1 = {1,2,3}

T2 = {1,3}

T3 = {1,4,5}

T4 = {2,5}

Sp = Sensitive pattern = {1,3}

Step – 2 Binary encoding to apply genetic algorithm

| |
|---|
| 11100 |
| 10100 |
| 10010 |
| 01001 |

Step – 3 Let's count fitness value: (Base on proposed fitness function)

| Dataset | Fitness value |
|---|---|
| 11100 | 6.34 |
| 10100 | 4.34 |
| 10010 | 4.5 |
| 01001 | 5.0 |

Transaction having lower fitness value will be selected for modification

Step – 4 Tournament selection process:

| Dataset | Fitness value | T-selection | |
|---|---|---|---|
| 11100 | 6.34 | 10100 | 10010 |
| 10100 | 4.34 | | |
| 10010 | 4.5 | 10010 | |
| 01001 | 5.0 | | |

Step – 5 Single point crossover and mutation operator

| Population | Offspring | Mutation | Fitness |
|---|---|---|---|
| 11100 | 10100 | 00100 | 5.5 |
| 10100 | 11010 | 11110 | 4.34 |
| 10010 | 11100 | 11000 | 3.34 |
| 01001 | 11010 | 11010 | 4.5 |

Step – 6 Replacement to improve the fitness of parents.

| Population | Offspring | Replacement |
|---|---|---|
| 11100 | **00100** | 11100 |
| **10100** | 11110 | **00100** |
| 10010 | 11000 | 10010 |
| 01001 | 11010 | 01001 |

Step – 7 After replacement sanitized dataset is generated as shown below:

| |
|---|
| 11100 |
| 00100 |
| 10010 |
| 01001 |

Step – 8 Decoding of sanitizing dataset

D = {T1, T2, T3, T4}

T1 = {1,2,3}

T2 = {3}

T3 = {1,4,5}

T4 = {2,5}

### C. PPARM-GA: Proposed Algorithm

Here, Fitness Value has been proposed and it is defined as:

$$\text{Fitness} = 1/\sum_{i=1}^{k} Count(Sp)\ in\ Tr + size(Ti) + \sum_{i=1}^{n} di$$

```
1. Input: Original Database D, SARs, MCT, MST, N, Replace
2. Output: Transform D into D′
3. FS → Frequent Item set (D)
4. AR → Generate Association Rules (FS)
5. SAR → Select Sensitive Association Rules (AR)
6. WHILE SAR{} != Ø OR generation != N
7. Fitness: $\mathbf{1/\sum_{i=1}^{k} Count(Sp)\ in\ Tr + size(Ti) + \sum_{i=1}^{n} di}$
8. Selection: Base on fv
9. Crossover: Tr * Tr+1
10. Mutation: Select Tr , Change 1 to 0 or 0 to 1 randomly
11. Fitness: $\mathbf{1/\sum_{i=1}^{k} Count(Sp)\ in\ Os + size(Oi) + \sum_{i=1}^{n} oi}$
12. Replace: Tr Δ Os
13. Stop
14. D vs D′
```

## V. Performance Analysis

The minimum specification required for its implementation is a processor of 1.80 GHz, RAM of 2 gigabytes, hard disk with 10 gigabytes, Windows 7 OS or higher with 64 bit, and MATLAB R2013B tool.

The dataset description [24]-[27] along with its source has been specified in the below-given Table III:

TABLE III: Dataset Description

| Source | Dataset | No. of Instance | No. of attribute |
|---|---|---|---|
| UCI Repository | Voting-records | 101 | 15 |
| GitHub | Breast cancer record | 1000 | 50 |
| MathWorks | Karate | 34 | 34 |
| MathWorks | Transactional Data | 1000 | 50 |

The association rules were generated using Apriori and FP-growth algorithm considering the four different datasets and the time is taken by each of the datasets is depicted in the below-given Fig. 7. It can be concluded that FP-Growth takes less time for execution as compared to the Apriori algorithm.

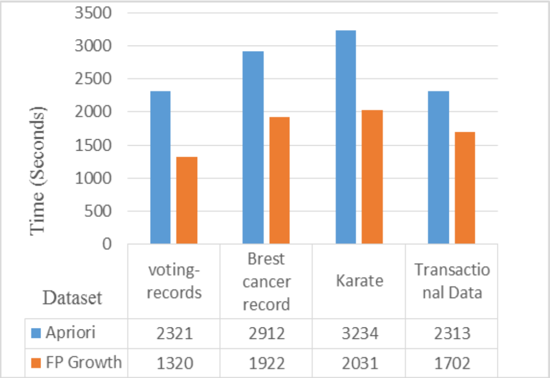

Fig. 7. Comparison of execution time using Apriori and FP growth algorithm.

Then, the number of association rules achieved considering the FP-growth algorithm and setting various support and confidence threshold for all the respective datasets are shown in the below-given Table IV.

Now, the proposed algorithm PPARM-GA has been implemented and the side effects obtained such as several lost rules and ghost rules have been analyzed for all the four datasets in the below Fig. 8.

TABLE IV: ASSOCIATION RULES

| Dataset | \|D\| | No. of Association Rules | Support | Confidence |
|---|---|---|---|---|
| Voting-records | 101 | 1264 | 0.2 | 0.6 |
| | | 428 | 0.3 | 0.7 |
| | | 13 | 0.35 | 0.8 |
| Breast cancer | 10k | 55 | 0.2 | 0.6 |
| | | 24 | 0.3 | 0.7 |
| | | 0 | 0.35 | 0.8 |
| Karate | 3.5 | 29 | 0.2 | 0.6 |
| | | 4 | 0.3 | 0.7 |
| | | 0 | 0.35 | 0.8 |
| Transactional bakery | 10k | 83 | 0.2 | 0.6 |
| | | 17 | 0.3 | 0.7 |
| | | 4 | 0.35 | 0.8 |

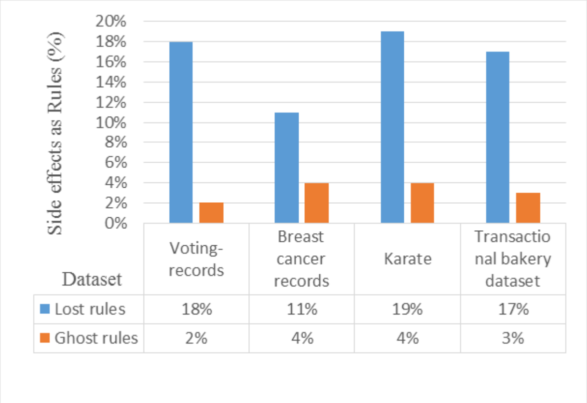

Fig. 8. Side effects as a number of lost and ghost rules using PPARM-GA.

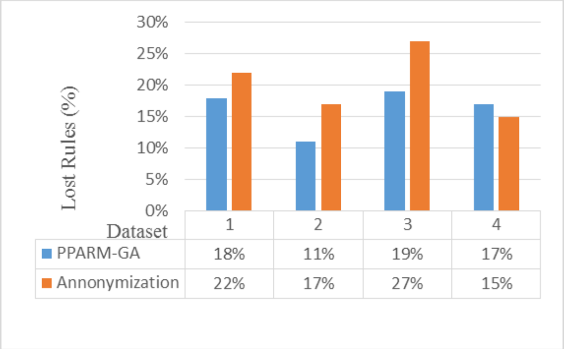

Fig. 9. Generation of Lost rules for various datasets.

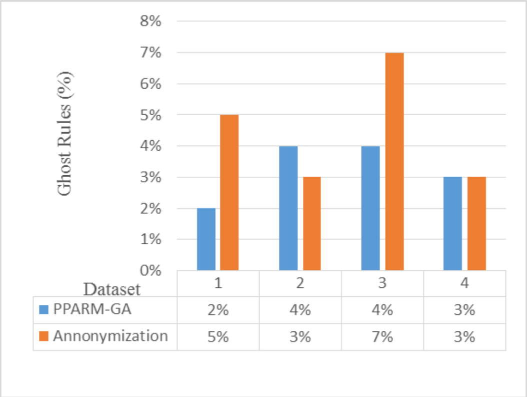

Fig. 10. Generation of Ghost rules for various datasets.

Furthermore, a comparison of the existing method anonymization [23] and the proposed method PPARM-GA considering the side effects such as the generation of lost rule and ghost rule considering four different datasets has been carried out. It can be concluded from the below Fig. 9 and Fig. 10 that fewer lost rules and ghost rules are produced by PPARM-GA as compared to anonymization.

The CPU time i.e execution time taken by four different datasets setting confidence threshold as 30%,40%,50%, and 60% by utilizing the PPARM-GA and anonymization algorithm is calculated. As shown below given Fig. 11-14, it can be seen that PPARM-GA takes a minor more time as compared to the anonymization algorithm.

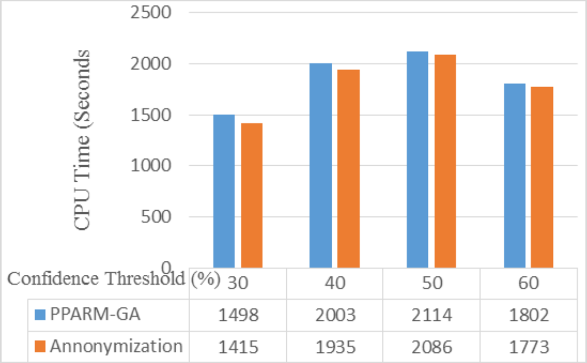

Fig. 11. CPU time considering confidence threshold for the Voting-records dataset.

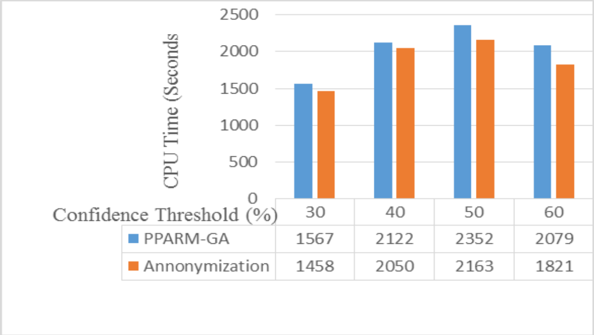

Fig. 12. CPU time considering confidence threshold for Breast cancer dataset.

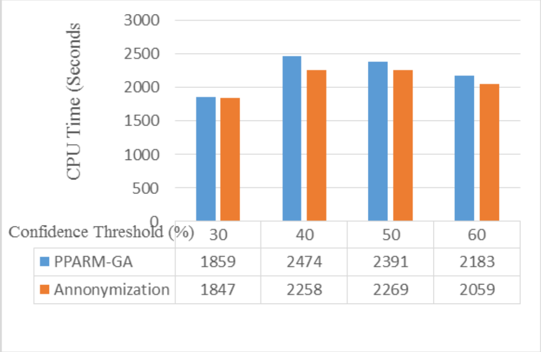

Fig. 13. CPU time considering the confidence threshold for the Karate dataset.

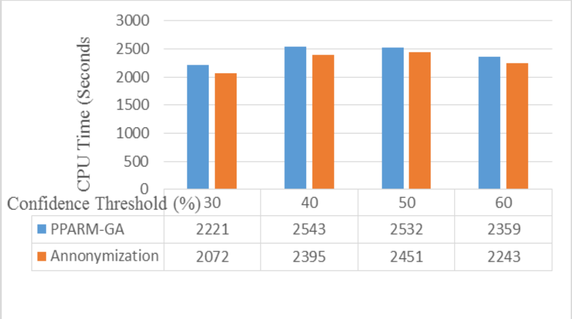

Fig. 14. CPU time considering confidence threshold for the Transactional bakery dataset.

Now, one of the important parameters that is hiding failure was calculated using PPARM-GA and anonymization algorithms considering the sensitive percentage of FI (frequent items) as depicted in the below-given Fig. 15. It can be concluded that PPARM-GA successfully hides all the sensitive items as compared to anonymization.

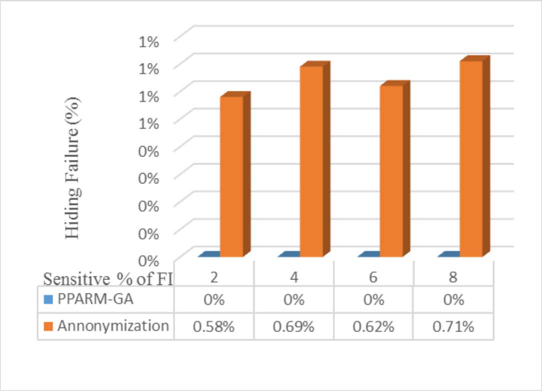

Fig. 15. Hiding failure considering the sensitive percentage of FI.

Finally, the utility of data by which the effectiveness of the sanitization algorithm namely PPARM-GA and anonymization considering all the datasets can be measured is calculated. It can be concluded that as shown in the below given Fig. 16 PPARM-GA is more effective as compared to the anonymization algorithm in terms of data utility.

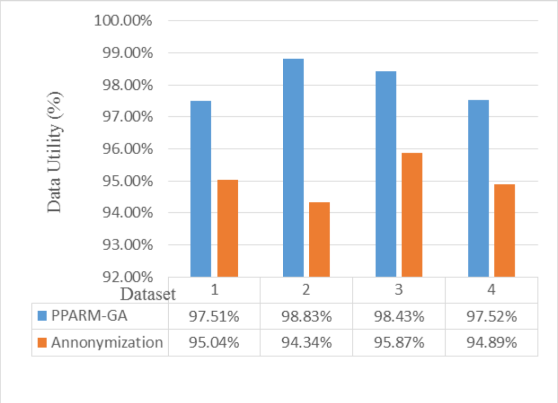

Fig. 16. Data utility considering the different datasets.

## VI. Conclusion and Future Scope

This paper presents a privacy-preserving algorithm based on the evolutionary concept specifically genetic algorithm GA called PPARM-GA (privacy-preserving association rule mining using a genetic algorithm) to protect sensitive association rules with multiple objectives like reducing the side effect and maintaining the data utility.

The association rules were generated considering the FP-growth algorithm since it requires less time for execution as compared to Apriori. Now, privacy-preserving association rule mining was carried out to hide the sensitive rules. While hiding the sensitive rules, sanitized dataset gets created which gives rise to various side effects. The side effects in terms of lost rule, ghost rule, and hiding failure are taken into deliberation. However, the PPARM-GA outperforms the existing anonymization algorithm by reducing all the considered side effects on different datasets.

Furthermore, data utility and CPU time factors were also well-thought-out for performance analysis. It can be concluded that PPARM-GA provides better data utility as compared with anonymization though it takes a minor more CPU time for its execution.

Thus, PPARM-GA is found to be more effective than the existing method of anonymization. The work can further be extended by considering the distributed scenario as future scope. Also, an effort to reduce the CPU time and other parameters such as missing cost and artificial cost, etc. can be addressed.

## Conflict of Interest

The submitted work was carried out with no conflict of interest. Hence, the authors declare no conflict of interest.

## Author Contributions

Prof. Darshana Patel analyzed the data and wrote the paper. Miss Hiral Kotadiya conducted the research and Prof. Avani Vasant provided guidance in the preparation of the paper and has checked for plagiarism and grammar. All the authors had approved the final version.

## References

[1] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, 1st ed. Addison Wesley Longman Publishing, Co. Inc., 2005.

[2] R. Agrawal, T. Imielinski, and A. Swami., "Mining association rules between sets of items in large databases," in *Proc. the ACM SIGMOD International Conference on Management of Data (ACM SIGMOD '93)*, Washington, USA, May 1993.

[3] J. Panackal and A. Pillai, "Privacy-preserving data mining: An extensive survey," in *Proc. International Conference on Multimedia Processing, Communication, and Information Technology*, 2013.

[4] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy-preserving mining of association rules," *Information Systems*, vol. 29, pp. 343-364, 2004.

[5] R. Agrawal and R. Srikant, "Privacy-preserving data mining," in *Proc. ACM Conf. Management of Data*, 2000, pp. 14-19.

[6] T. Kumbhare and S. Chobe, "An overview of association rule mining algorithms," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, pp. 927-930, 2014.

[7] D. Patel, S. Shah, and A. Vasant. "Privacy preservation of class association rules and its optimization by utilizing genetic algorithm," *International Journal of Engineering &Technology*, 2018.

[8] A. Telikani and A. Shahbahrami, "Optimizing association rule hiding using a combination of border and heuristic approaches," *Applied Intelligence*, vol. 47, pp. 544-557, 2017.

[9] P. Darshana and K. Radhika, "Privacy-preserving data mining: A parametric analysis," in *Proc. the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications*, pp. 139-149, 2017, Springer Singapore.

[10] V. S. Verykios, A. K. Elmagarmid, E. Bertino, Y. Saygin, and E. Dasseni, "Association rule hiding," *IEEE Trans. Knowledge and Data Eng.*, vol. 16, no. 4, pp. 434-447, 2004.

[11] M. Aarthna and P. Khushboo, "Analysis of methodologies for hiding sensitive frequent itemsets using a border-based approach," *International Journal of Advanced Research in Computer Science*, vol. 6, no. 8, Nov-Dec 2015.

[12] S. Anju and S. Umesh Kumar, "Approaches for privacy-preserving data mining by various associations rule hiding algorithms – A survey," *International Journal of Computer Applications*, vol. 134, no. 11, 2016.

[13] L. Zhang, W. Wang, and Y. Zhang, "Privacy-preserving association rule mining: taxonomy, techniques, and metrics," *IEEE Access*, vol. 7, pp. 45032-45047, 2019.

[14] A. Telikani, A. Gandomi, A. Shahbahrami, and M. Naderi, "Privacy-preserving in association rule mining using an improved discrete binary artificial bee colony," *Expert Systems with Applications*, vol. 144, 2020.

[15] T.-Y. Wu, J. C.-W. Lin, Y. Zhang, and C.-H. Chen, "A grid-based swarm intelligence algorithm for privacy-preserving data mining," *Applied Sciences*, vol. 9, no. 4, pp. 774, 2019.

[16] D. Patel, S. Shah, and A. Vasant, "Technique for optimization of association rule mining by utilizing genetic algorithm," *Recent Advances in Computer Science and Communications*, vol. 13, no. 1, 2020.

[17] G. S. Navale and S. N. Mali, "A multi-analysis on privacy preservation of association rules using the hybridized approach," *Evolutionary Intel.*, 2019.

[18] N. K. Bux, M. Lu, J. Wang, S. Hussain, and Y. Aljeroudi, "Efficient association rules hiding using genetic algorithms," *Symmetry*, vol. 10, no. 11, p. 576, 2018.

[19] M. Afshari, M. Dehkordi, and M. Akbari, "Association rule hiding using a cuckoo optimization algorithm," *Expert Systems with Applications*, vol. 64, pp. 340-351, 2016.

[20] S. Rahat and A. Sohail, "Privacy-preserving in association rules using a genetic algorithm," *Turkish Journal of Electrical Engineering and Computer Science*s, 2014.

[21] A. Telikani and A. Shahbahrami, "Data Sanitization in association rule mining: An analytical review," *Expert Systems with Applications*, vol. 96, pp. 406-426, 2018.

[22] C. W. Lin, T. P. Hong, J. W. Wong, G. C. Lan, and W. Y. Lin, "GA-based approach to hide sensitive high utility itemsets," *Scientific World Journal*, 2014.

[23] P. Garach and D. Patel, "Privacy protection of class association rules produced by medical datasets," in *Proc. IEEE 5th International Conference for Convergence in Technology*, March 2019.

[24] L. Sweeney, "K-Anonymity: A model for protecting privacy," *International Journal Uncertain Fuzziness Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[25] Zoo Data Set. *Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/zoo

[26] Breast Cancer Wisconsin (Diagnostic) Data Set. *Machine Learning Repository*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)

[27] Congressional Voting Records Data Set. *Machine Learning Repository*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/congressional+voting+records

**Darshana H. Patel** is currently working as the head and assistant professor of the Department in Information Technology at V. V. P. Engineering College, Rajkot, Gujarat, India. She holds a Ph.D. in computer engineering and has more than 13 years of experience in academics. She is having a life membership of ISTE and also coordinated several workshops and short-term training programs at VVP. Dr. Darshana has authored several research papers which are published in prominent International Journals and also guided several research candidates. Her area of research includes machine learning, data mining, data science, etc.

**Hiral Kotadiya** is currently working as a researcher in Weboccult Technologies at Ahmedabad, Gujarat, India. She has completed a master's in information technology engineering and is having more than 5 years of research experience. Her area of research includes machine learning, data mining, pattern recognition, etc.

**Avani R. Vasant** is working as a professor and the head of the Department of Computer Science Engineering at Babaria Institute of Technology, Vadodara, Gujarat, India. She is a member of the Governance body at BITs Edu. Campus. She holds a Ph.D. in computer engineering. As an academic and researcher, she holds more than 20 years of experience, actively participates in research projects, had lectured in different domains. Dr. Avani has published numerous research papers in renowned proceedings and has also guided several research candidates. Her research specialization includes an area of machine learning, pattern recognition, etc.