International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

Implementation of Client Controlled Privacy Preserving Data Model for Mining Decision Rules Using Decision Tree and Association Rules

Md Ilyas, Dhanraj Verma, Manoj kumar Deshpande, Rajeev Raghuvanshi

Abstract*: The privacy-preserving data mining (PPDM) is one of the techniques which are used for mining data dynamically with preserving privacy of the end data owner. In this paper, a PPDM technique for generating the privacy-preserving decision rules is proposed and implemented. The key motive of presenting this privacy-preserving decision rule mining technique is to demonstrate how securely data is aggregated in the PPDM environment, how securely extract them and consumed them with the help of applications. In addition to comparing the state of art methods for mining privacy preserving decision rules for preparing the future directions of research. Therefore two different data models have used namely decision tree and association rule mining. The conducted experiments demonstrate that decision tree-based techniques are superior to the association rule mining based techniques for mining higher dimensional data with higher accuracy and low resource consumption. Therefore in the near future for extending this data model the two concepts are also introduced in this paper.

Keywords*: privacy preserving data mining, decision rule mining, implementation of data mining technique, performance evaluation.

I. INTRODUCTION

Data mining is one of the essential techniques for analyzing data automatically. That is used in various domains of applications for classification, prediction, categorization, pattern understanding and recognition [1]. Therefore it becomes increasingly important for various applications in banking and finance, engineering and medical research, security and investigations and many more [2]. In this context, the different kinds of computational algorithms are applied over the data for recovering the essential patterns which can be used with real-world problems [3]. In this presented work the PPDM [4] is the main area of study and system design.

The PPDM is a technique where the data is mined with the help of a central trusted or semi-trusted authority. Additionally, the specific concentration over security, privacy and the sensitivity of data is maintained for the end data user [5]. In this environment, the different data suppliers are involved who are agreed for disclosing their own part of data attributes for research and common decision-making purposes in various areas of business intelligence, research and development and many more. But they are worried about the sensitivity and privacy of end data owners [6]. Therefore the PPDM techniques are adopted for mining data securely and without harming the sensitivity and privacy of any end data owner.

In this section, the overview of the proposed work is described. In the next section the recent contributions in the domain of PPDM are explored as the literature survey. In next the key objectives of the presented work are discussed with the proposed privacy-preserving data mining technique. Further the implemented technique's performance is measured and finally, the conclusion and future research directions are established.

II. LITERATURE SURVEY

This section provides the various essential privacy-preserving data mining techniques and their core contributions.

Storage and discloser or spillage of sensitive information presents privacy issue. The procedures utilized for information extraction with preserving privacy are called PPDM. ***Ricardo Mendes et al [7]*** studies the most significant PPDM procedures from writing and assess them with the utilization of strategies in applicable fields. Moreover, current difficulties and issues are talked about. ***Yousra Abdul Alsahib S. Aldeen et al [8]*** gives a review of another viewpoint and orderly translation of a rundown distributed writing by means of their association in subcategories. The ideas of accessible PPDM techniques, points of interest, and constraints are talked about. The current methodologies are sorted based on affiliation rule, mutilation, shroud affiliation rule, bunching, scientific categorization, acquainted characterization, dispersed and k-namelessness, redistributed information mining, with aces and corns.

Revised Manuscript Received on November 30, 2019.
***** Correspondence Author
Md. Ilyas*, Dr. APJ Abdul Kalam University, Indore. Email: Prof.md.ilyas@gmail.com
Dr. Dhanraj Verma, Dr. APJ Abdul Kalam University, Indore. Email: dhanrajmtech@gmail.com
Dr. Manojkumar Deshpande, Prestige Institute of Engineering Management & Research, Indore. Email: manojvilasrao@gmail.com.
Rajeev Raghuvanshi, Prestige Institute of Engineering Management & Research, Indore. Email: raj.iet123 @gmail.com

This uncovers the past improvement, present difficulties, future patterns, holes, and shortcomings for upgrades.

Kaihe Xu et al [9] propose a novel structure to accomplish privacy-preserving AI where preparing information are dispersed and in enormous volumes. Creators use information region property of Apache Hadoop and a set number of cryptographic tasks at Reduce() utilized for privacy-conservation. Paillier cryptosystem is widely used as a homomorphic encryption plan to guarantee privacy necessities. ***Ismail San et al [10]*** explore how to handle execution corruption of Paillier cryptosystem. To begin with, use parallelism between tasks in cryptosystem and interleaving among free activities. At that point, create equipment acknowledgment of the plan utilizing field-programmable door exhibits. Group investigation plans to find designs through various calculations, for example, k-implies. Running k-implies over enormous information stores has offered ascend to privacy issues. ***Zakaria Gheid et al [11]*** propose a privacy-preserving k-implies calculation dependent on the multiparty added substance conspire. The structured arrangement is for on a level plane divided information. ***Rongxing Lu et al [12]*** present a Lightweight Privacy-preserving Data Aggregation, for haze processing. LPDA is portrayed by utilizing Paillier encryption, Chinese Remainder Theorem, and single direction hash affix strategies to not just total half and half IoT gadgets' information into one, yet in addition early channel infused false information. LPDA is secure and privacy-improved with differential privacy systems.

Yiannis Kokkinos et al [13] consider disseminated privacy-preserving information mining in decentralized information areas that manufacture a few neural systems. The best neural system is chosen by means of certainty proportion fondness spread and privacy-preserving figuring. In this arrangement, classifiers approve one another. The preparation set of one classifier turns into the approval set for other people. The development is parallelizable and the expense is O(L N) for L classifiers and ***N models. S. Sharma et al [14]*** present research difficulties for planning true privacy-preserving frameworks for medicinal services. Consequently a customized social insurance framework is created and approved for illness supervision. The investigative necessities for concurred parties, they recoup privacy assets, break down existing systems, and talk about tradeoff among productivity, privacy, and quality. The EHRs can conceivably resolve many existing issues related with ailment analysis, with persistent privacy and affectability of therapeutic data. The sharing of patient records between various human services offices has a significant concern. Be that as it may, to settle on viable choices from clinical information, have a lot of information to prepare choice models. ***Yan Li et al [15]*** create two versatile dispersed privacy-preserving calculations dependent on outfit system. The possibility of the methodology is to fabricate a model for each taking part office to precisely learn information appropriation and can move valuable information procured structure their choice models without sharing patient-level sensitive information. The strategy was assessed on diagnosing diabetes of patient records from specific areas.

Hamza Hammami et al [16] recommend a methodology that joins the extraction of regular shut examples in an appropriated domain with keeping up the privacy of destinations during information mining in cloud-based homomorphic encryption. The presentation shows that the system requires less correspondence and calculation overheads. Usage of the brilliant lattice, information gathering could risk the privacy of purchasers. ***Ken Birman et al [17]*** propose a structure for a keen metering that will enable utilities to utilize gathered information adequately while preserving the privacy of purchasers. The sharing of patient records damages the privacy of patients. Subsequently, restorative research is centered around mining affiliation rules without sharing EHR information. ***Nikunj Domadiya et al [18]*** a methodology for PPDARM is proposed for join digging of affiliation rules for all EHR frameworks. This methodology is additionally broke down with the coronary illness dataset. Information examination of private and sensitive data causes privacy issues. To discover viable relations among utility and privacy, Privacy-Preserving Utility Mining (PPUM) has turned into a basic issue. ***Wensheng Gan et al [19]*** give an outline of PPUM. To start with, present utility mining, and afterward present related fundamentals and issues, just as key assessment criteria. Just as favorable circumstances and insufficiencies are featured with specialized difficulties and future headings. ***Jerry Chun-Wei Lin et al [20]*** center around the issues of HUIM and privacy-preserving utility mining (PPUM), and present two calculations to mine and shroud sensitive high-utility itemsets.

G. Kalyani et al [21] an issue has been arranged in context of securing affiliation rules which are sensitive. The proposed strategy chooses exchanges for modifications utilizing the parallel TLBO streamlining method. K-Anonymity may have a few disadvantages. On the exposure restriction side, there is an absence of insurance against quality divulgence. The information utility side, managing genuine datasets is a moving errand to accomplish. ***Balkis Abidi et al [22]*** present another small scale collection HM-PFSOM, in light of fluffy possibilistic grouping. The anonymization procedure is applied per square of comparative information. Therefore, we can diminish data misfortune. This takes into account diminishing the revelation danger of private data. ***Chen-Yi Lin et al [23]*** centers around the information stream and sliding window plan with the reversible privacy-preserving idea is utilized to process ongoing information, which is named as ceaseless reversible privacy-preserving (CRP). Information with CRP calculation can be precisely recuperated Additionally, by utilizing a watermark, the uprightness of information confirmed. The outcomes show that CRP is viable for preserving learning, data misfortune, and privacy exposure chance.

III. PROPOSED WORK

This section discusses the aim and objectives of the proposed work additionally for achieving the required objective a model is also demonstrated with their algorithm steps.

A. System overview

The proposed work is intended to find an efficient and effective technique for mining privacy preserving decision rules [24].

In this environment, three major actors are available first the data owner who have their own data and for some essential task provide their own sensitive and private data to some organization or institution (i.e.

banking, hotel, hospital or other) for consuming some kinds of services offered by the institutions. This data may involve some personal data, such as banking details, credit card information, family member's details, and others. In this environment, the second actor is the institution that gathers various clients' data and can supply and use this data in favor of the end client. But in some conditions for the purpose of business needs, it is required to club their data with other organizations for making business-oriented decisions because the individual part of data is not sufficient for making effective decisions. But, the actual data discloser of end clients can affect the privacy and security of the end client's private and financial life. Therefore, before discloser of the data to any third party the sanitization of private and sensitive data required [25]. In this environment, the third stack holder is the authority where the data is going to be disclosed. This authority is responsible for mining the data securely and privately without disturbing the utility (application oriented) of data, security, and privacy [26]. Therefore, in order to deal with this situation the privacy-preserving data modeling is required. Thus to achieve the privacy-preserving decision mining the following key concepts are needed to be included.

1. Implementation of the cryptographic security at the data supplier end or agreed party who are collaborating data
2. Handling data at the data aggregator end for mining and distributing the decisions and their formations
3. Recovery of data and decisions only and only at the authentic party who are contributed that part of data

Among the three concluded objectives the decision mining is also a considerable part of the system. In this context, two popular techniques namely association rule mining and decision tree-based rule mining techniques are very popular. Thus in this work, both the techniques are modeled with the proposed privacy-preserving systems. The aim of this comparative data modeling is to find the superiority and extension of the modeled technique. This section provides an overview of the proposed concepts and objectives of the work. The next section provides the methodology of work.

B. Proposed methodology

The proposed methodology of privacy-preserving data mining technique includes three main layers for achieving the privacy-preserving decision rule mining. These are demonstrated in "Fig.1".

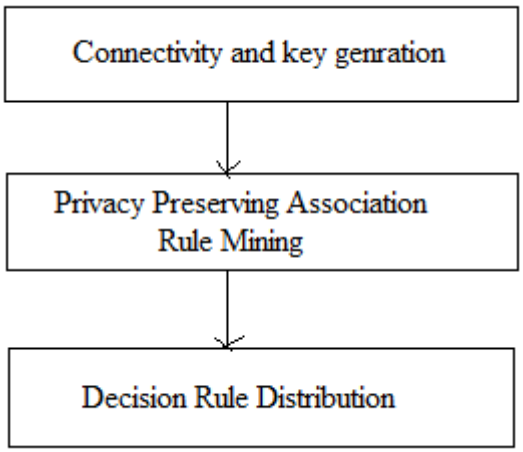

Fig. 1. Proposed System

The proposed layered privacy-preserving technique layered architecture is demonstrated in the above-given "Fig. 1". Additionally, the details are given as:

1. Connectivity and key generation

As we discussed previously the privacy-preserving data models includes the number of parties who are contributing data for securely and privately mining. The contributed data is mined at the server end. Therefore, when a party agreed for supplying the data and established connection then the following process is initiated as demonstrated in "Fig. 2".

The process is initiated when the data supplier (client) wants to combine the data with others and start communication with the rule mining server. The server finds the connection request then accepts the connection. After accepting the connection, the server generates an 8 digit random number as the encryption key and communicates it to the client using the obtained key from server-side, the client process the data using the encryption algorithm. The encryption algorithm processes the input client's attributes and produced ciphered attributes. The cipher is communicated to the server for rule mining purposes.

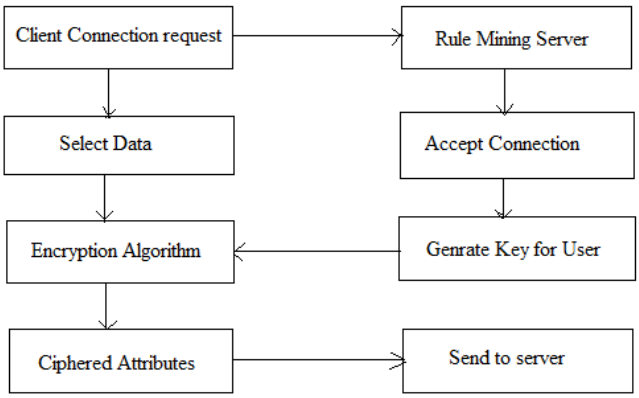

Fig. 2 Connection Process

In order to generate the cipher-text, the following process is used as demonstrated in "Fig. 3".

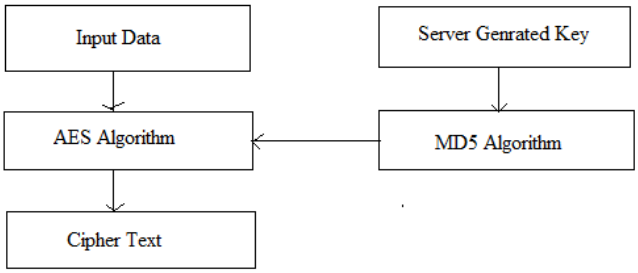

Fig. 3. Encryption Process

According to the described process in "Fig. 3", the encryption technique accepts the input data which is needed to be sent to the server. Additionally, the server-generated key is accepted by the system. The obtained key is passed over the MD5 hash generation algorithm which produces the 128-bit hash code. This code and user input data is passed to the AES algorithm for generating the cipher-text. This cipher-text is sent to the server for further process.

2. Privacy preserving Rule mining

The privacy-preserving rules are mined on the basis of two popular approaches namely Apriori algorithm and C4.5 decision tree algorithm. These algorithms are discussed as:

Apriori Algorithm

Apriori is a classical data mining algorithm for producing association rules. It is designed for transactional databases. In order to understand the transactional database, suppose a shop has a collection of items that are purchased by consumers. Apriori algorithm mainly works on the concept of frequency measurement. [27]:

- ✓ All subsets of a frequent item-set are considered
- ✓ Additionally, all the possible combinations of item-set, are also considered

Prior to starting the procedure, let us set the help limit to half, i.e. just those things are huge for which bolster is over half.

Step 1: created a frequency table which contains all the elements (items) with their occurrence based on the given transactions as given in "Table-I".

Step 2: According to the given "Table-I", those items are having the importance that has the frequency higher than the specified threshold (support).

Table-I: Frequency Table

Items	Frequency
Apple(A)	5
Orange(O)	6
Banana(B)	5
Papaya(P)	5
Dates(D)	2

The support value is provided by experimenter, therefore, those items which are less than support count are considered and remaining items are removed. Thus the above-given table is reformed as "Table-II":

Table-II: Item Purchased by Customer

Items	Frequency
Apple(A)	5
Orange(O)	6
Banana(B)	5
Papaya(P)	5

Thus it means Dates (D) with frequency 2 is rarely consumed.

Step 3: in further to process the data we prepare the subsets of the available items. In this context it is required to keep in mind the combination, AB is the same as BA. Thus we need to create all the possible combinations with two items.

Step 4: now the possible combinations are created with the two items are by using the concept of step 3 we find the frequency in the given transactions. An example of this event is given in "Table-III".

Table-III: Frequency of Each Pair

Items	Frequency
OA	4
OB	3
OP	2
AB	4
AP	3
BP	2

Step 5: after that, we again eliminate the item's pair which is lower than the support count.

Step 6: Now need to create the pairs with the three items. Additionally, start the search over the transaction set for finding the combinations with their frequencies higher than the support count. For example, we found OPB and PBA two items with frequency higher than support.

Table-IV: Frequency of Itemsets

Items	Frequency
OPB	4
PBA	3

This process continuously working until all the combinations are not evaluated. The algorithm of the above-described process is reported using "Table-V", as the algorithm steps.

Table-V: Apriori Algorithm

Process **Apriori** (T, minSupport) {

$L_1 = \{Itemsets\}$;

for $(k = 2;\ L_{k-1}! = \emptyset; k++)\{$

$C_k = genrate\ Candidateset\ using\ L_{k-1}$

// It is Cartesian product $L_{k-1} * L_{k-1}$ and eliminating any $k-1$ size itemset that is below support count

for each *transaction t* ***do*** {

#increment count of all candidates in C_k that contain in t

L_k = candidates in C_k with minSupport

}

}

return $U_k L_k$

}

Decision tree C4.5 or J48

The second method which is used for decision rule mining is the C4.5 or J48 decision tree. That is an extension over a decision tree ID3. Entropy and information gain is used to create data partitions and tree building. The attribute with higher information gain is used to mount on a tree at a higher level. The C4.5 algorithm continuously uses this technique to create sub-lists to complete the tree. Thus information gain measurement requires calculating entropy first. To calculate the entropy supposes the dataset contains two classes, T (True) and F (False). The entropy is basically measured for entire dataset D. Therefore for the given example of binary classification the entropy E for dataset D is defined as:

$$E(D) = -T log_2 T - F\ log_2 F$$

Where T is the ratio of True data instances and F stands for false samples

To reduce the depth of tree while traversing it, selection of the best possible attribute is required for creating branches. The attributes with minimum entropy will be picked. Thus information gain is required to drop in entropy with respect to an attribute during splitting. The information gain, Gain (E, A) for attribute A is given by,

$$Gain(E, A) = Entropy(s) - \sum_{n=1}^{v} \frac{E_v}{E} XEntropy(E_v)$$

The gain can be utilized to decide positions of attributes in the tree. The position of attribute as a node is decided on the basis of gain for the two majors first to create a small size tree and to offer the required level of unfussiness. C4.5 algorithm returns the decision tree as a learning outcome [28]. The following steps can be used for generating decision tree:

INPUT: A set of data (D) with the means of discrete variables.

OUTPUT: A decision tree T which is constructed by passing data set.

A. A node (X) is created
B. If instance in same class.
   a. Make node (X) as leaf node and assign class label C;
C. If the attribute list is empty,
   a. Make node(X) a leaf node and assign a class label of most frequent class;
   b. Choose an attribute with highest information gain, and then marked as test-attribute;
D. If X in role of test-attribute;
   a. Generate a new branch of tree that is suitable for test-attribute from node X;
E. $If\ (B_i == Null)$
   a. Add a new leaf node, of class label of most frequent class**;**
F. Else
   a. Add a leaf node and returned by Generate-decision-tree.

This section first introduces the algorithms which are used for generating the decision rules; now the process involved in this technique is discussed using "Fig. 4".

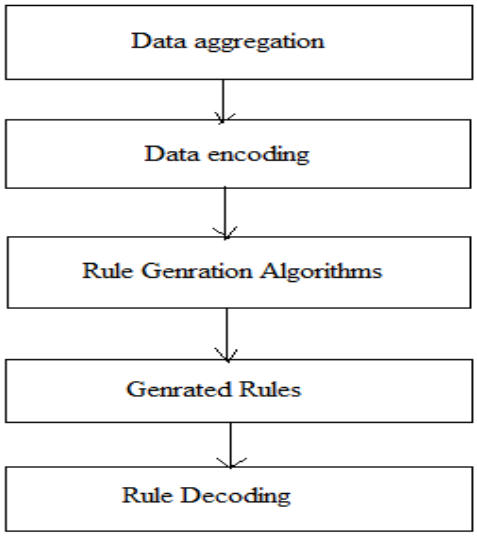

Fig. 4 Rule Generation Process

The received data from all the connected client sources are aggregated over a common dataset. The core cryptographic process transforms entire data attributes, therefore, the data attributes are encoded into symbols first. The encoded symbols are used with the apriori algorithm and C4.5 decision tree algorithm for generating the IF-THEN-ELSE rules. The generated rules are again reversed into their actual encrypted format that process is termed here as the decoding of the symbols used for data processing. In the next section, the data recovery at the client end is described.

3. Decision Rule Distribution

The data distribution for all the clients not required any complex process for recovering the contributed part of data. Therefore the prepared rules are transmitted to all the connected clients.

The clients are usages a similar key and algorithm for decrypting the rules. Using this blind decryption process only those parts of attributes are recovered which are contributed by the parties.

IV. RESULTS ANALYSIS

The proposed privacy persevering rule mining techniques are evaluated in this section. In order to compute and compare the performance of the system is demonstrated using the following performance parameters.

A. Accuracy

In machine learning and data mining, the accuracy of an algorithm is known as the number of decisions correctly made using the trained algorithm. Therefore, it is a ratio of total correctly made decisions for the total samples provided for decision making. The following Eq. can be used for measuring the accuracy of algorithms.

$$accuracy(\%) = \frac{total\ correct\ decisions * 100}{total\ samples\ for\ decision\ making}$$

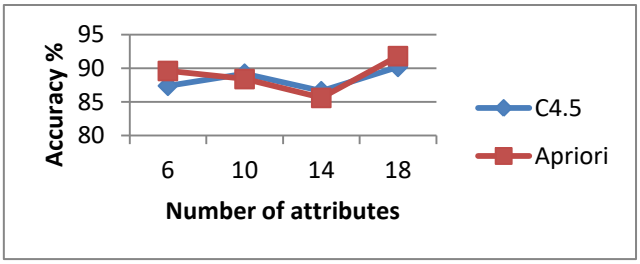

Fig. 5 Accuracy

The accuracy of the privacy-preserving decision-making rule mining techniques for both the kinds of algorithms is reported using "Fig. 5". The accuracy of developed techniques is provided in the Y-axis which is measured in terms of percentage. In addition to that, the X-axis of the algorithm shows the number of attributes involved in experiments. According to the measured performance, the accuracy of both the algorithms is similar but sometimes the performance of decision tree algorithms performed better and sometimes the association rule mining based technique. However, the mean accuracy of the decision tree algorithm is higher as compared to the association rule mining technique because of low ambiguity in decision tree-based rule mining techniques.

B. Number of Rules

The performance of both the privacy-preserving rule mining data model namely association rule mining and the decision tree-based rule mining technique is described in this section by using the number of rule generation. "Fig. 6" shows the number of rules generated using both the privacy reserving rule mining techniques in the Y-axis and the X-axis shows the number of rules involved during experiments.

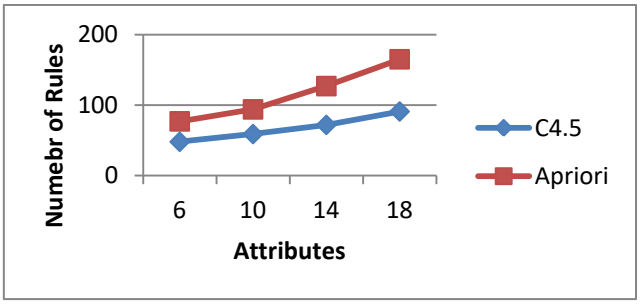

Fig. 6 Number of rules

According to the demonstrated results, the association rules mining based privacy-preserving technique generates the number of rules as compared to the decision tree-based privacy-preserving technique. Therefore for making decisions using the association rules the numbers of comparison cycles are higher than the decision tree-based rule mining technique.

C. Time complexity

The time complexity of an algorithm or process is termed as the amount of time required to process the data according to the implemented algorithm. That can be computed using the following Eq.

$$time\ consumed = Algorithm\ end\ time - start\ time$$

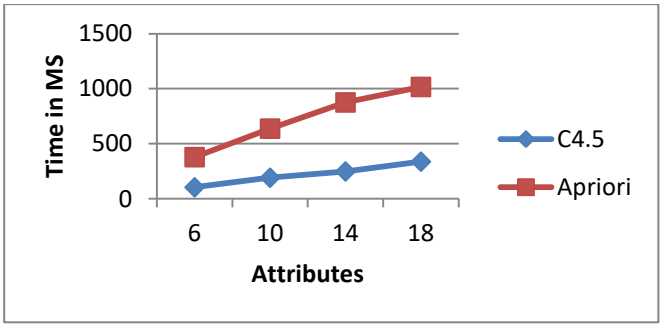

Fig. 7 Time Expenses

The time complexity of association rule-based privacy-preserving rule mining technique and the decision tree-based rule mining technique is described in "Fig.7". The time expenses of the algorithms are measured in terms of milliseconds. In this diagram, the Y-axis shows the time consumed during the experiments and the X-axis shows the number of attributes utilized for experimentation. According to the demonstrated results, the time consumption for the association rule mining technique is higher as compared to the decision tree-based privacy-preserving rule mining technique.

D. Space complexity

The space complexity of a data model is also known as the memory usages of the process or algorithm. Basically when a process initiated for execution the system assigns a fixed amount of main memory for that process.

Among the amount of memory free from expenses is known as the utilized memory. In JAVA that is computed using the following Eq.

$$memory\ usages = total\ assigned - free\ memory$$

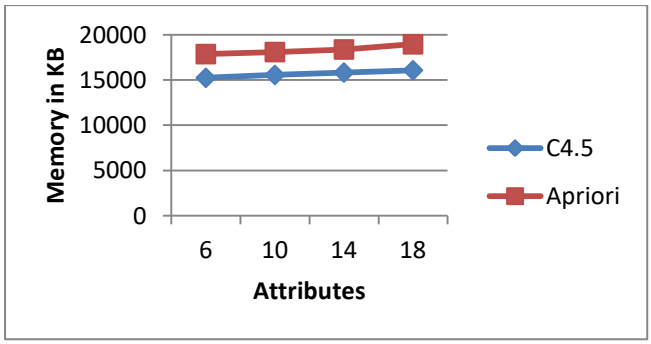

Fig. 8 memory usages

According to the obtained memory usages of association based rule mining and the decision tree-based rule mining technique is demonstrated in "Fig. 8".

The measured memory of the implemented systems is reported in the Y-axis and the X-axis shows the number of attributes involved for privacy-preserving rule mining. The experimental results demonstrate the memory usages of the privacy-preserving rule mining techniques are increases with the number of attributes are increased. Additionally, the memory usages of the association rule mining technique are significantly higher as compared to decision tree-based rule mining for all the conducted experiments.

V. CONCLUSION & FUTURE WORK

This section discusses the findings of the efforts placed in this paper. In addition to future directions of the work is also reported here.

A. Conclusion

The area of data mining applications is increasing rapidly not only in science and technology it is being popular in various other domains of services such as finance, medicinal science, business intelligence and many more. A number of kinds of applications in these industries are utilizing the techniques of data mining such as prediction, classification, understanding of patterns, recognition, quality assurance, and much more. In addition to that, some of the applications need to sanitize and prevent the privacy of the end data owner. Therefore, the proposed work is focused on designing and developing an enhanced technique of privacy-preserving rule mining techniques.

In this presented work the privacy-preserving IF-THEN-ELSE rule mining technique is implemented using apriori algorithm and C4.5 decision tree algorithm. In order to provide security at the end of the data supplier, the AES and MD5 based encryption algorithms are used. After sanitization of sensitive data attributes the data is transmitted to the server end where the data is processed using the implemented algorithms in cryptographic format. After processing data the recovered rules are distributed to all the concerned parties. The rules attributes are recovered at the client's side using the blind decryption process and the decisions are made using the traversing of rules and the available attributes.

The proposed data models are prepared using the JAVA technology additionally for preserving the performance of the implemented system the MySQL Database is used. The implemented system is evaluated using the following performance parameters based on future directions of research is decided. The obtained experimental observations are described in "**Table-VI".**

S. No.	Parameters	Association rules	Decision tree based rules
1	Accuracy	Similar	Similar
2	No. of rules	Higher	Low
3	Memory usages	Higher	Low
4	Time expenses	Higher	Low

Table-VI: performance summary

According to the experimental observations the following conclusion of the work found:

1. Association rules mining technique are computationally costly as compared to decision tree algorithm
2. The amount of rules for association rule mining is higher as compared to the decision tree additionally as the amount of data attributes increase the significant amount of rules also increases.

B. Future work

The aim of the proposed work is to compute the efficient privacy-preserving rules for decision making is accomplished in this work. In addition to that two different kinds of rule mining techniques are implemented and compared. Based on the observed performance the following future work is proposed.

1. During the data aggregation from the different parties, the dimensions of data significantly increase, therefore a dimensionality reduction techniques is also required for handling the data
2. According to the obtained performance, the association rule mining techniques generate a significant amount of decision rules as compared to the decision tree which is not much suitable for enterprise rule mining and distribution. Therefore, in the near future, the decision tree is used for experimentation and system design for PPDM

REFERENCES

1. E. Toch, B. Lerner, E. B. Zion, I. B. Gal, *"Analyzing large-scale human mobility data: a survey of machine learning methods and applications"*, Knowl Inf Syst, https://doi.org/10.1007/s10115-018-1186-x
2. I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, I. Chouvarda, *"Machine Learning and Data Mining Methods in Diabetes Research"*, Computational and Structural Biotechnology Journal 15 (2017) 104–116
3. J. Qiu, Q. Wu, G. Ding, Y. Xu and S. Feng, *"A survey of machine learning for big data processing"*, EURASIP Journal on Advances in Signal Processing (2016) 2016:67, DOI 10.1186/s13634-016-0355-x
4. C. W. Lin, T. P. Hong, H. C. Hsu, *"Reducing Side Effects of Hiding Sensitive Itemsets in Privacy Preserving Data Mining"*, Hindawi Publishing Corporation Scientific World Journal Volume 2014, Article ID 235837, 12 pages
5. P. S. Rao, S. Satyanarayana, *"Privacy preserving data publishing based on sensitivity in context of Big Data using Hive"*, J Big Data (2018) 5:20, https://doi.org/10.1186/s40537-018-0130-y
6. O. Tene, J. Polonetsky, *"Big Data for All: Privacy and User Control in the Age of Analytics"*, 11 Nw. J. Tech. & Intell. Prop., 239 (2013).
7. R. Mendes, J. P. Vilela, *"Privacy-Preserving Data Mining: Methods, Metrics, and Applications"*, Vol. 5, 2017, IEEE
8. Y. A. A. S. Aldeen, M. Salleh, M. A. Razzaque, *"A comprehensive review on privacy preserving data mining"*, SpringerPlus (2015) 4:694, DOI 10.1186/s40064-015-1481-x
9. K. Xu, H. Yue, L. Guo, Y. Guo, Y. Fang, *"Privacy-preserving Machine Learning Algorithms for Big Data Systems"*, 2015 IEEE 35th International Conference on Distributed Computing Systems, 1063-6927/15© 2015 European Union DOI 10.1109/ICDCS.2015.40
10. I. San, N. At, I. Yakut, H. Polat, *"Efficient paillier cryptoprocessor for privacy-preserving data mining"*, Security and Communication Networks 2016; 9:1535–1546, Wiley Online Library, DOI: 10.1002/sec.1442
11. Z. Gheid, Y. Challal, *"Efficient and Privacy-Preserving k-Means Clustering for Big Data Mining"*, IEEE TristCom, Aug 2016, Tianjin, China pp.791 - 798, ff10.1109/TrustCom.2016.0140ff ffhal-01466904f

R. Lu, K. Heung, A. H. Lashkari, A. A. Ghorbani, *"A Lightweight Privacy-Preserving Data Aggregation Scheme for Fog Computing-Enhanced IoT"*, Special Section on Security and Privacy in Applications and Services for Future Internet of Things, Vol. 5, 2017, IEEE

12. Y. Kokkinos, K. G. Margaritis, *"Confidence ratio Affinity Propagation in ensemble selection of neural network classifiers for distributed privacy-preserving data mining"*, Neurocomputing, (2015), vol. 150, pp. 513−528

13. S. Sharma, K. Chen, A. Sheth, *"Towards Practical Privacy-Preserving Analytics for IoT and Cloud Based Healthcare Systems"*, IEEE Internet Computing, March-April 2018
14. Y. Li, C. Bai, C. K. Reddy, *"A Distributed Ensemble Approach for Mining Healthcare Data under Privacy Constraints"*, Inf Sci (Ny). 2016 February 10; 330: 245–259. doi:10.1016/j.ins.2015.10.011
15. H. Hammami, H. Brahmi, I. Brahmi, S. B. Yahia, *"Using Homomorphic Encryption to Compute Privacy Preserving Data Mining in a Cloud Computing Environment"*, EMCIS 2017, LNBIP 299, pp. 397–413, DOI: 10.1007/978-3-319-65930-5_32, Springer International
16. K. Birman, M. Jelasity, R. Kleinberg, E. Tremel, *"Building a Secure and Privacy-Preserving Smart Grid"*, ACM SIGOPS OSR 49(1) pp131–136, http://dx.doi.org/10.1145/ 2723872.2723891.
17. N. Domadiya, U. P. Rao, *"Privacy-preserving association rule mining for horizontally partitioned healthcare data: a case study on the heart diseases"*, Sådhanå (2018) 43:127 Indian Academy of Sciences, https://doi.org/10.1007/s12046-018-0916-9
18. W. Gan, J. C. W. Lin, H. C. Chao, S. L. Wang, P. S. Yu, *"Privacy Preserving Utility Mining: A Survey"*, arXiv:1811.07389v1 [cs.DB] 18 Nov 2018
19. J. C. W. Lin, W. Gan, P. F. Viger, L. Yang, Q. Liu, J. Frnda, L. Sevcik, M. Voznak, *"High utility-itemset mining and privacy-preserving utility mining"*, Perspectives in Science (2016) 7, 74—80
20. G. Kalyani, M. V. P. Chandra Sekhara Rao and B. Janakiramaiah, *"Privacy-Preserving Association Rule Mining Using Binary TLBO for Data Sharing in Retail Business Collaboration"*, Advances in Intelligent Systems and Computing 515, © Springer Nature Singapore Pte Ltd. 2017
21. B. Abidi, S. B. Yahia, C. Perera, *"Hybrid Microaggregation for Privacy-Preserving Data Mining"*, arXiv:1812.01790v1 [cs.CR] 4 Dec 2018
22. C. Y. Lin, Y. H. Kao, W. B. Lee and R. C. Chen, *"An efficient reversible privacy-preserving data mining technology over data streams"*, SpringerPlus (2016) 5:1407, DOI 10.1186/s40064-016-3095-3
23. T. Pawar, Prof. S. Kamalapur, *"A Survey on Privacy Preserving Decision Tree Classifier"*, International Journal of Engineering Research and Applications, Vol. 2, Issue 6, November- December 2012, pp.843-847
24. I. Ray, T. C. Ong, I. Ray, M. G. Kahn, *"Applying Attribute Based Access Control for Privacy Preserving Health Data Disclosure"*, 978-1-5090-2455-1/16/$31.00 ©2016 IEEE
25. L. Urquhart, N. Sailaja, D. McAuley, *"Realising the right to data portability for the domestic Internet of things"*, Pers Ubiquit Comput (2018) 22:317–332, DOI 10.1007/s00779-017-1069-2
26. *"Laboratory Module 8: Mining Frequent Itemsets – Apriori Algorithm"*, available online at: http://software.ucv.ro/~cmihaescu/ro/teaching/AIR/docs/Lab8-Apriori.pdf
27. K. K. Mishra, R. Kaul, *"Audit Trail Based on Process Mining and Log"*, International Journal of Recent Development in Engineering and Technology, Volume 1, Issue 1, Oct 2013

AUTHORS PROFILE

Md. Ilyas, Research Scholar (Pursuing PhD)
Prof.md.ilyas@gmail.com
9827122226, 8109788442

Md. Ilyas Bachelor of Engineering from IET Devi Ahilya University Indore, 2009 and Master of Engineering from Sushila Devi Bansal College of Engineering Indore in Year 2015. Pursuing Ph.D. (CSE) from Dr. APJ Abdul Kalam University, Indore from year 2017. Currently presently doing PhD in Computer Science & Engineering, College of Engineering, Dr. APJ Abdul Kalam University, Indore. He has published more than 8 Research paper National / international journals including. His main research work focuses on Artificial intelligence, Cloud Computing, IoT, Big Data Analytics, Data Mining, Privacy Preserving Data and Computational Intelligence based education. He has 9.8 years of teaching experience.

Dr. Dhanraj Verma, Professor (CSE)
dhanrajmtech@gmail.com
9755301831

Dr. Dhanraj Verma Bachelor of Science from Vikarm University of Ujjain, 1997 and Master of Technology from Devi Ahiliya University, Indore in Year 2007. Ph.D.(CSE) from BU University in year 2013. Currently Presently working as a Professor in Department of Computer Science & Engineering , College of Engineering, Dr. APJ Abdul Kalam University, Indore . He is a member of IEEE & IEEE Computer Society Since 2012, Life member of the CSI since 2012, He has published more than 24 Research paper National / international journals including. His main research work focuses on Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining, IoT and Computational Intelligence based education. He has 18 years of teaching experience and 6 years of Research Experience..

Dr. Manojkumar Deshpande, Director
manojvilasrao@gmail.com
9892061141

Prof. Manojkumar Deshpande has joined PIEMR Indore, as Director, in Jan 2018. In short time, his reforms at various levels, enhanced academic standard of engineering and management education. Before he was having additional charge of Vice Chancellor at Symbiosis University of Applied Sciences, Indore. In SUAS, due to his tireless effort and networking, Technology and Management education boosted. He has also worked as Professor & Associate Dean at MPSTME, SVKM's NMIMS-Mumbai, Shirpur Campus. He spent 12 years in Mumbai and worked with Best Engineering Colleges such as D J Sanghvi College of Engineering. He was also Member, Board of Studies of SVKM's NMIMS University, Mumbai. He has received Doctoral Degree in Computer Engineering from SVKM's NMIMS University, Mumbai in 2011. He is alumnus of formerly CEDTI, NIELET, Aurangabad (M.Tech.) and, SSGMCE, Shegaon (B.E.) Maharashtra, India. He is having 27 years of academic experience at various positions. He is guide for Ph.D. & M.Tech in Computer Engineering. There are 28 publications One Patent (Applied) at his credit. His research focus is Big Data Analytics, Artificial Intelligence, Software Engineering and Multimedia Systems and STEM Education. He is recipient of recipient of Education Leadership Award: Dewang Mehta Education Award, June 2019 & Top Retail Minds Awards, Mar 2019. He is Member of ISTE and involved in conducting Conferences,

Workshops, Industrial Consultancies and Social Activities. He was also member of ASEE, CSI, ACM etc. He visited Kingston University, London, UK, University of Oslo, Norway, Microsoft & Washington State University, Seatle, USA.

Rajeev Raghuvanshi, Assistant Professor (CSE)
raj.iet123 @gmail.com
8103129701, 8878335383

Rajeev Raghuvanshi Bachelor of Engineering from IET Devi Ahilya University Indore, 2009 and Master of Engineering from Medicaps Institute of Technology and Management, Indore in Year 2013. Currently presently working as an Assistant Professor in Department of Computer Science & Engineering, Prestige Institute of Engineering Management & Research Indore. He has published more than 9 Research paper National / international journals including. His main research work focuses on Network Security, Wireless Network, Big Data Analytics, Data Mining, IoT, Solar Energy and Computational Intelligence based education. He has 9.7 years of teaching experience.