# An Advance Approach of Privacy Preservation Using the Concept of Machine learning and Data Modification Techniques

[*]Sonia Bajaj, [#]Naushaba Aafreen Khan

[*]Head of Department, [#]M.tech Scholar, Computer Science and Engineering, G H Raisoni University, M.P, India. [*]sonia.bajaj@ghru.edu.in, [#]naushabaaafreen123456@gmail.com

***Abstract*— Huge amount of data stored in databases so it is necessary to develop an effective technique for analysis of such data. Data mining process is used to extracts the useful and knowledgeable information from the large databases. Privacy preserving data mining techniques are used to extract the relevant information from the large amount of data and protect the sensible information. This paper represents review of privacy preserving techniques and machine learning techniques like secure multiparty computation technique, data modification technique and target function technique.**

**Keywords:—** ***Data Mining techniques, Privacy Preserving, Data Modification, Machine learning, Target Function technique.***

## I. INTRODUCTION

Data mining is the process of finding useful and interesting patterns from large amount of databases. Privacy preserving data mining has emerged as a very important research area in data mining. Knowledge discovery through a combination of different databases generates security issue. Although data mining results do not violate privacy means it cannot be assured that an illegal person will not access the data. Data mining techniques try to identify regularities in data, which are hidden and hard to discover by individual persons. Classification of data mining techniques is shown in fig.1.

Data mining is a process of discovering unknown interesting patterns from huge amount of data. Data mining techniques are used to access the data from different perspectives and convert it into the useful form i.e. information. This information can be used to improve performance.

Data mining techniques involves integration of different techniques such as database and warehouse technology, machine learning, statistics high performance computing, artificial intelligence, pattern recognition, neural network and data visualization. The data mining techniques are also used on Bio-Database for finding environmental conditions.

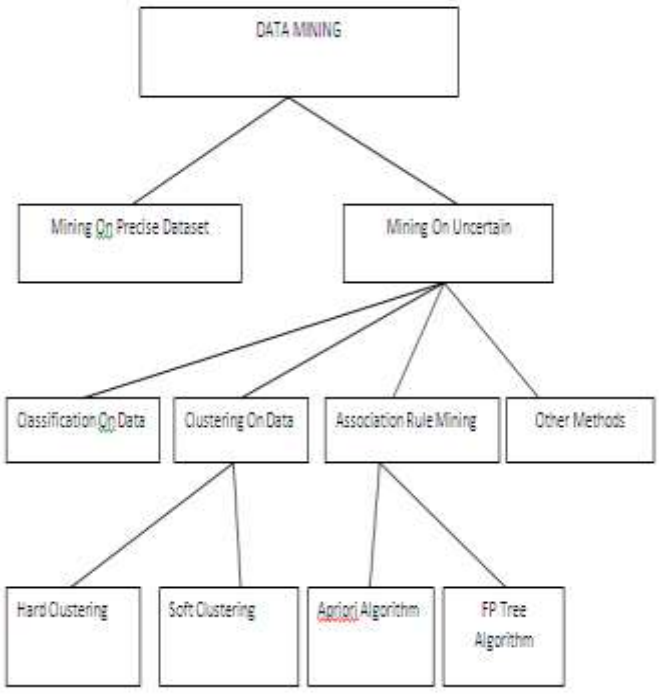


**Fig.1: Classification of Different Data Mining Techniques**

The architecture of a data mining system may have the following important components (Fig.2)

 DOI : 10.35291/2454-9150.2020.0518 © 2020, IJREAM All Rights Reserved.

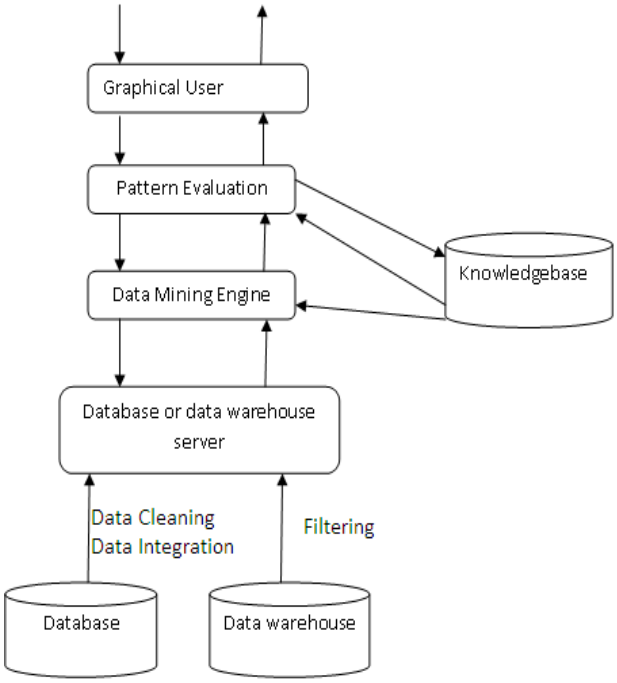


**Figure 2: Architecture of a typical data mining system**.

**1. Database, data warehouse:** Collection of data is called database and collection of one or more databases is called data warehouse. Initially we perform data cleaning and data integration operations before using the data.

**2. Data warehouse server:** It is responsible for accessing the relevant and meaningful data from the large amount of data and it's totally depends on user's mining request.

**3. Knowledge base:** Knowledge base is the main knowledge or domain which is responsible for guiding, searching and evaluating the interesting patterns. This knowledge is useful for organizing the attributes.

**4. Data mining engine:** Data mining system is essential part of the data mining system. It is responsible for performing clustering, classification, regression etc.

**5. Pattern evaluation module:** This module is used for evaluating the interesting patterns. it is compulsory to insert the results of evaluation so pattern interestingness can confine the search only for the interesting patterns.

**6. Graphical user interface:** Graphical user interface module is used to communicate between users and the data mining system. With the help of this interface user communicates with the system.

## II. PRIVACY ISSUES IN DATA MINING

Privacy preservation [7] [8] is an important concept because when we transfer the data from one place to another place it is necessary to provide security to that data. Preservation means maintaining the data in its original position. Privacy is required at the time of communication because unauthorized persons can disclose our private data as publically. Various techniques are used for providing privacy to data is describe below.

1. Data is modified earlier than deliver it to the receiver.

2. Data is scattered between different sites, so we require best protocol to get best global results without any fault or failure.

3. While using a best model we can get best results.

Privacy preserving data mining has emerged as a very important research area in data mining. Knowledge discovery through a combination of different databases generates security issue. Although data mining results do not violate privacy means it cannot be assured that an illegal person will not access the data. Data mining techniques try to identify regularities in data, which are hidden and hard to discover by individual persons.

Regularities or unknown patterns are to be revealed over the entire database, rather than on individuals. So we have to find such disclosure of patters, the data mining process has to access and use individual persons information.

## III. PRIVACY PRESERVING DATA MINING TECHNIQUES

The basic concept behind privacy preserving data mining techniques is given below.

1) Data Swapping Technique: This technique were introduced by Dalenius in year 1982 but only for categorical data set. The basic idea of the technique is to keeps all the original values in data set. This technique actually replaces the original data set by another one [10].

After implementation of this techniques a new data swapping techniques is introduced for modifying the original data set. Main focus of this technique is on pattern preservation. Mainly this technique preserves the classification rules

2) Aggregation Technique: Aggregation technique is also known as generalization technique. Manish Sharma et al. proposed a method in which he preserves the individual privacy by perturbing the original data set. This method can change k no. of data records at a time. The value of an attribute is derived from average of all the values.

3) SMC Technique (Secure Multipart Computation) developed by Yao in year 1982. There are several SMC (secure multiparty techniques). Generally these techniques uses primitive computations like secure sum computation, secure set union computation and secure size of set intersection.

4) Suppression Technique: In this method the data values are suppressed into micro data. Data suppression technique is used to protect an privacy from unauthorized access.

5) Target Function Technique: Target function is an important machine learning technique. It reduces the problem of over fitting and it also reduces the problem of under fitting.

In order to find sensitive data values an unauthorized

 DOI : 10.35291/2454-9150.2020.0518 © 2020, IJREAM All Rights Reserved.

person like hacker or attacker can use several approaches. Privacy preserving techniques are used to protect confidential data from unauthorized persons.

## IV. FLOW CHART OF PROPOSED WORK

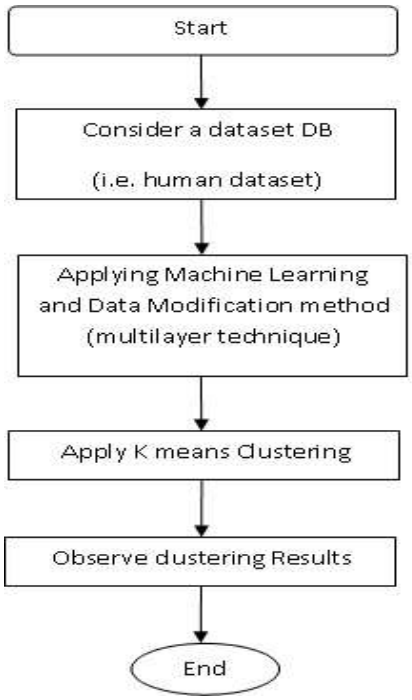


**Fig2 :-Flow Chart of Proposed Work**

## V. CONCLUSION

This paper represents review of privacy preserving techniques and machine learning techniques like secure multiparty computation technique, data modification technique and under fitting technique.

All the techniques discussed here are only an approximation for our goal of privacy preservation. We will further refine those techniques and also develop some efficient techniques. We will try to develop more efficient techniques for reducing the computation cost.

## REFERENCES

[1] V.S. Verykios et al. State-of-the-art in privacy preserving data mining. SIGMOD Record, 33 (1): 50-57, 2017.

[2] US Department of Labor. Executive order 13145. Available from http://www.dol.gov/oasam/regs/statutes/eo13145.htm, Feb 8, 2015.

[3] M. Thuraisingham. national security, privacy, and civil liberties. SIGKDD Explorations, 4 (2): 1-5, 2016.

[4] S. E. Fienberg. Privacy in an e-commerce world: Data mining, matching and disclosure limitation. Statistical Science, 21:143-154, 2016.

[5] R. Agarwal et al. Privacy –preserving data mining. In Proc. Of the ACM SIGMOD

[6]J. C. Wortmann. Security control methods for statistical databases: A comparative study. ACM Computing Surveys21 (4): 515-556, 2010.

[7] N. Adam and J. C. Wortmann. Security control methods for statistical databases: A comparative study. ACM Computing Surveys21 (4): 515-556, 2010

[8] V. S. Iyenger. Transforming data to satisfy privacy constraints. In Proc. Of SIGKDD'02, Edmonton, Alberta, Canada, 2008.

[9] R. Sarathy. a new masking approach for numerical data. Management Science, Forthcoming, 2008.

[10] J. Gehrke. Privacy preserving mining of association rules. In Proc. Of the Eighth ACM SIGKDD International Conference on Knowledge and Data Mining, pages 217-228, 2005.

[11] J.R Hartisa. Maintaining data privacy . In Proc. of the 28th VLDB Conference, pages 682-693, Hong-Kong, China, 2007.

[12] Y. Saygin, V. S. Verykios and A. K. Elmagarmid. Privacy preserving association rule mining. In RIDE, pages 151-158, 2006.

[14] Z. Zhan et al Using randomized response techniques for privacy preserving data mining. In Proc. of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 515-510, Washington DC, USA, August 2003.

[15] Y. Zhu et al randomization for privacy preserving data mining. In Proc. of the Tenth ACM SIGKDD International Conference on Knowledge discovery, pages 761-766, Seattle, Washington, USA, August 2004.

[16] A. C Yao. Protocols for secure computations. In Proc. of the 23rd Annual IEEE Symposium on Foundation of Computer Science, 2000.

 DOI : 10.35291/2454-9150.2020.0518 © 2020, IJREAM All Rights Reserved.