



Privacy Preservation in Data Mining Using Mixed Transformation Technique

Pinkal Jain

Research Scholar

*Vindhya Institute of Technology and Science
Jabalpur (M.P.), [INDIA]*

Email: pinku029jain@gmail.com

Sandeep Nandanwar

Assistant Professor,

*Department of Computer Science and Engineering
Vindhya Institute of Technology and Science
Jabalpur (M.P.), [INDIA]*

Email: sandeepnandanwar@gmail.com

ABSTRACT

The new era of information communication and technology (ICT), everyone wants to store/share their Data or information in online media, like in cloud database, mobile database, grid database, drives etc. What is needed is to find the relevant information or data from these large databases is a challenge for which we use data mining techniques. Privacy preservation is the major issue we should be very careful while sending data from one place to other. For securing that information from those kinds of unauthorized people we proposed and implement one of the technique in which we use mixed transformation technique (i.e. Translation & Scaling) for providing the highest privacy with taking the human database Division of data into the groups of similar objects is called Clustering. In proposed work, k means clustering technique has been used. Proposed mechanism has been found satisfactory and provides the highest privacy in clustered database environments which performs better than existing.

Keywords:— *Data mining, mixed transformation technique, K means clustering technique.*

I. INTRODUCTION

In recent years, the data mining became a very interesting topic for the researcher due to its vast use in modern technology of computer science but due to its vast use it faces some serious challenges regarding data privacy and data privacy became an interesting topic. Many methods techniques and algorithms are already defined and presented for privacy preserving data mining. These privacy preserving techniques can be classified mainly in two approaches

1. Data modification approach
2. Secure Multi-party Computation approach

Data Mining in last few decades has become very useful as the database are increasing day by day many people now connected with the computers so it becomes necessary for computer researchers to make the data so fast to access, also need to find right data. The term Data Mining emphasize on the fact of extracting the knowledge from large amount of data, so data mining is the process through which we collect knowledgeable data from very large data.

Now a days the databases are very large which consists so much information but what we want to find is the relevant data from large database or want to find some patterns which becomes very difficult with normal DBMS but with the use of data mining techniques we can find the hidden patterns and information from large database system. So we can also term data mining as the knowledge mining, pattern extraction etc. But before applying data mining techniques we need to apply some processes which we known as preprocessing of data. Although data mining is one of the step involved in process of knowledge discovery but still it becomes more popular by name then that.

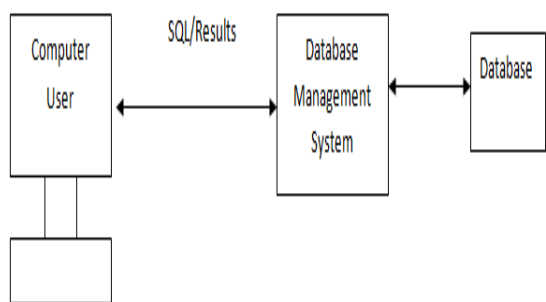


Figure 1: Database Access

2. PROBLEM STATEMENT

Privacy preservation in data mining is an important concept because when the data is transferred or communicated between different parties then its compulsory to provide security to that data so that other parties do not know what data is communicated between original parties .Preserving in data mining means hiding output knowledge of data mining by using several methods when this output data is valuable and private. Mainly two techniques are used for this one is Input privacy in which data is manipulated by using different techniques and other one is the output privacy in which data is altered in order to hide the rules.

3. OBJECTIVE OF THE STUDY

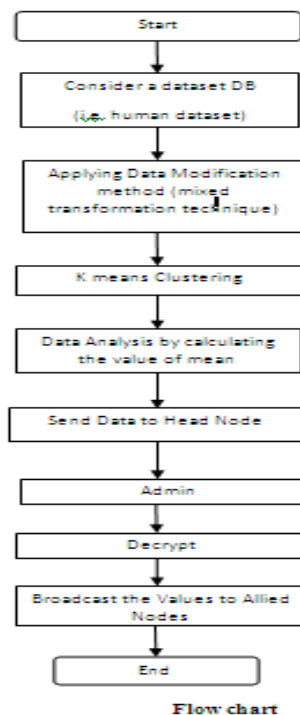
The main objective of this thesis:

When this stored data is transferred the problem of privacy is arises. The main objective of this thesis is to provide the highest privacy in order to secure our data from unauthorized persons. For achieving this goal we use mixed transformation technique and k means clustering technique. This approach transforms the original data values to privacy-preserved data maintaining the inter relative distance among the data.

4. IMPLEMENTATION

In the implementation work, we are taking the human database (i.e. human dataset) that contains five attributes age, weight, height, temperature and class then we apply the mixed transformation technique i.e. translation and scaling with the help of Weka tool for providing the highest privacy. For data analysis we apply the K means clustering technique.

Flow Chart



Flow chart

Figure 2: Flow Chart

Table 1. Original Dataset (Human Dataset)

S.No.	Age(year)	Weight(kg)	h e i g h t (feet)	Tempera- ture (°C)	Class
1	2	15	2.1	37	1
2	10	18	4.6	37.5	1
3	20	49	4.9	38	1
4	25	65	6.1	37.3	1
5	12	42	4.8	37.2	1
6	30	59	5.9	37.2	1
7	20	49	5.7	37	1
8	28	68	6	37.3	1
9	32	72	5.9	37.8	1
10	18	52	5.3	37.5	1

Table 2 After Applying the mixed Transformation

S.No.	Age (year)	Weight (kg)	h e i g h t (feet)	T e m - peratu re (°C)	Class
1	5.0	5.0	5.0	5.0	1
2	25.0	9.477	67.5	31.315	1
3	50.0	55.746	75.0	57.631	1
4	62.49	79.626	105.0	20.789	1
5	30.0	45.298	72.5	15.526	1
6	75.0	70.671	100.0	15.526	1
7	50.0	55.746	95.0	5.0	1
8	70.0	84.104	102.50	20.789	1
9	80.0	90.074	100.0	47.105	1
10	45	60.223	85.0	31.315	1

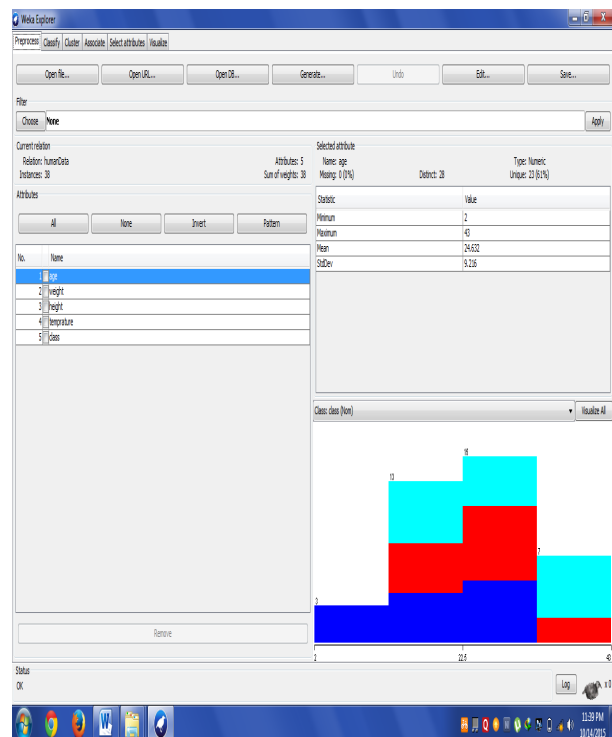


Figure 3: Shows snapshot after inserting the data set in weka tool

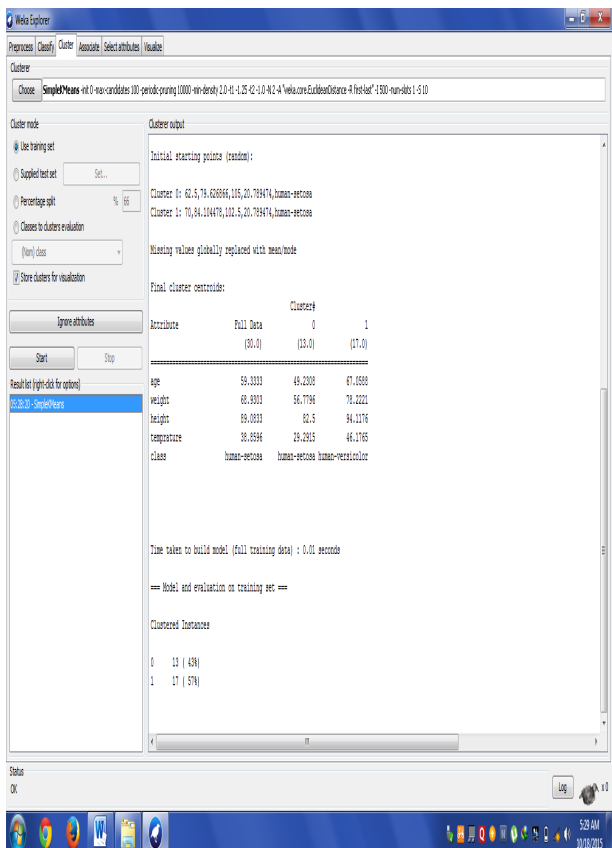


Figure 4: After Applying the K means clustering

5.COMPARISON

Comparison between the Min max normalization technique and proposed mixed transformation technique.

Table 3. Comparison Between Database

S. No.	Person	Original data	Min_max Normalization	Proposed System (Mixed transformation)
1	1	2	10	5
2	2	10	33	25
3	3	20	62	50
4	4	25	76	62
5	5	12	39	30
6	6	30	90	75
7	7	20	62	50

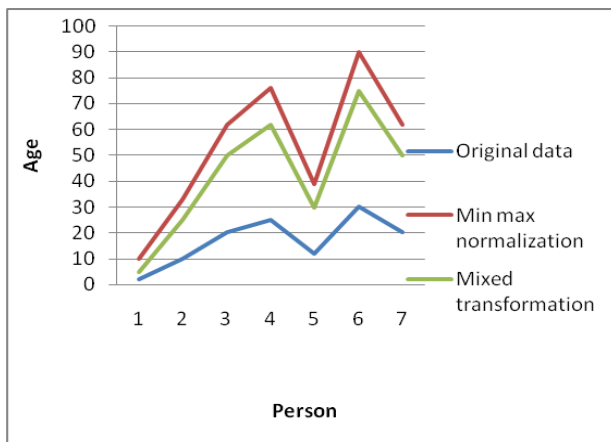


Figure 5: Comparison Graph

6. CONCLUSION:

In proposed work we have dealt with mixed transformation technique (i.e. translation and scaling) to preserve data privacy. This approach transforms the original data values to privacy-preserved data maintaining the inter relative distance among the data. Our experiments have proven that performing k-means clustering on the distorted data produces same clustering results as

original data. Thus we can say we have succeeded in achieving both accuracy and privacy. We have tested this technique for numerical data set.

REFERENCES:

- [1] Syed Md. Tarique Ahmad, et al “Privacy Preserving in Data Mining by Normalization” . IN: Proc. Of International Journal of Computer Applications (0975 – 8887),Volume 96– No.6, June 2014.
- [2] S. Vijayarani, et al “Data Transformation Technique for Protecting Private Information in Privacy Preserving Data Mining”. In: Proc. of Advanced Computing: An International Journal (ACIJ), Vol.1, No.1, November 2010.
- [3] Agarwal, R., Imielinski, T., Swamy, A. “Mining Association Rules between Sets of Items in Large Databases”, Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-210, 1993.
- [4] Srikant, R., Agarwal, R “Mining generalized association rules”, In: VLDB’95, pp.479-488, 1994.
- [5] Agrawal, R., Srikant, R, “Privacy-Preserving Data Mining”, In: proceedings of the 2000 ACM SIGMOD on management of data, pp. 439-450, 2000
- [6] Lindell, Y., Pinkas, B, “Privacy preserving data mining”, In: Proceedings of 20th Annual International Cryptology Conference (CRYPTO), 2000.
- [7] Kantarcioglu, M., Clifto, C, “Privacy-

- Preserving distributed mining of association rules on horizontally partitioned data”, In IEEE Transactions on Knowledge and Data Engineering Journal, IEEE Press, Vol 16(9), pp.1026-1037, 2004.
- [8] Han, J. Kamber, M, “Data mining Concepts and Techniques”. Morgan Kaufmann, San Francisco, 2006.
- [9] Sheikh, R., Kumar, B., Mishra, D, K, “A Distributed k- Secure sum Protocol for Secure Multi Site Computations”. Journal of Computing, Vol 2, pp.239-243, 2010.
- [10] Sheikh, R., Kumar, B., Mishra, D, K, “A modified Ck Secure sum protocol for multi party computation”. Journal of Computing, Vol 2, pp.62-66, 2010.
- [11] Jangde, P., Chandel, G, S., Mishra, D, K.,.: ‘Hybrid Technique for Secure Sum Protocol’ World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 vol 1, No. 5,198-201, (2011).
- [12] Sugumar, Jayakumar, R., Rengarajan, C (2012) “Design a Secure Multi Site Computation System for Privacy Preserving Data Mining”. International Journal of Computer Science and Telecommunications, Vol 3, pp.101-105.
- [13] N V Muthu Lakshmi, Dr. K Sandhya Rani, “Privacy Preserving Association Rule Mining without Trusted Site for Horizontal Partitioned database”, International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, pp.17-29, 2012.
- [14] N V Muthulakshmi, Dr. K Sandhya Rani, “Privacy Preserving Association Rule Mining in Horizontally Partitioned Databases Using Cryptography Techniques”, International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 3 (1), PP. 3176 – 3182, 2012.
- [15] J.Vaidya, “Privacy preserving data mining over vertically partitioned data,” Ph.D. dissertation, Purdue University, 2004.
- [16] J. Vaidya, C. Clifton, M. Kantarcioglu, and A. S. Patterson, “Privacy preserving decision trees over vertically partitioned data,” in ACM Transactions on Knowledge Discovery from Data, vol. 2, no. 3, 2008,pp. 14–41.
- [17] Y. Shen, H. Shao, and L. Yang, “Privacy preserving c4.5 algorithm over vertically distributed datasets,” in International Conference on Networks Security, Wireless Communications and Trusted Computing, vol. 2. Wuhan, Hubei: IEEE computer society, April 2009, pp. 446–448.
- [18] O.Goldreich, S. Micali, and A. Wigderson, “How to play any mental game or a completeness theorem for protocols (extended majority36abstract),” in STOC ’87 Proceedings of the nineteenth annual ACM symposium on Theory of computing, New York, 1987, pp. 218–229.
- [19] N. Adam and J. C. Wortmann. Security control methods for statistical databases: A comparative study. ACM Computing Surveys, 21 (4): 515-556, 1999.

- [20] T. Dalenius and S. P. Reiss. Data Swapping: A technique for disclosure control. *Journal of Statistical Planning and Inference*, 6(1):73-85, 1982.
- [21] S. E. Fienberg and J. McIntyre. Data swapping: Variations on a theme by Dalenius and Reiss. *Journal of Official Statistics*, 21:309-323, 2005.
- [22] K. Murlidhar and R. Sarathy. Data Shuffling –a new masking approach for numerical data. *Management Science*, Forthcoming, 2006.
- [23] V.S.Iyenger. Transforming data to satisfy privacy constraints. In *Proc. of SIGKDD'02*, Edmonton, Alberta, Canada, 2002.
- [24] S. Rizvi and J.R Hartisa. Maintaining data privacy in association rule mining. In *Proc. of the 28th VLDB Conference*, pages 682-693, Hong-Kong, China, 2002.
- [25] Y. Saygin, V. S. Verykios and A. K. Elmagarmid. Privacy preserving association rule mining. In *RIDE*, pages 151-158, 2002.
- [26] A.V. Evfimievski, R. Srikant, R. Agarwal and 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, 2002.

* * * * *