Achieving Highest Privacy Preservation in Data Mining Using Data Modification Technique

Naushaba Aafreen Khan[1], Sonia Bajaj[2]
[1]M.tech Scholar, Computer Science and Engineering Department, G H Raisoni University, M.P.
[2]Head of Department, Computer Science and Engineering, G H Raisoni University, M.P.

ABSTRACT: Data mining procedure extracts the useful patterns and meaningful knowledge from huge amount of databases. DM has a large deal of concentration in the IT trade in modern years, outstanding to the accessibility of very huge amount of data in addition to need for transforming such data into helpful information. This helpful information can be use in a variety of application areas like FD (fraud detection), MA (market analysis), PC (production control) and SE (science exploration). While we transfer this data from one position to another we have need of privacy preserving techniques since different types of unauthorized persons such as hackers or attackers can release our private data as publically. This work provides very high privacy via hybrid technique. Transformation technique used to change the position, size, shape and direction of the specified data objects. Apply k means clustering technique for data analysis. For investigational purpose we make use of a dataset (customer dataset) and perform all operations in data mining tool. This effort gives the maximum privacy as compared to the earlier work.

KEYWORDS: Data mining, Weka data mining tool, Privacy, k means Clustering, clustering.

I. INTRODUCTION

Data mining procedure extracts the useful patterns and meaningful knowledge from huge amount of databases. DM has a large deal of concentration in the IT trade in modern years, outstanding to the accessibility of very huge amount of data in addition to need for transforming such data into helpful information. This helpful information can be use in a variety of application areas like FD (fraud detection), MA (market analysis), PC (production control) and SE (science exploration). The data mining techniques are also suitable for Bio-Database for analysis and acquiring the similar relations in the condition of market or ecological conditions and many other situations to find the relationships which can tell the root of any ailment at very untimely stage so that appropriate protection can be taken. Medical science is a field where huge amount of data is collected from different sources now the main challenge is to find the suitable information and useful pattern it can be apply for future research to find some useful results for customers but privacy is the major challenge we should be very watchful while transferring data from one side to other side otherwise it may cause some harmful effects.

II. PROPOSED WORK

In this effort we take a database i.e. customer dataset. When we transfer data with the admin initially we need privacy as there is a possibility that someone in between the transmission of data that may change this data which will source many hazards so in regulate to secure our transportation from intruders, we will amend our data. In this work we apply two level securities by using transformation technique. In this work we keep the copy of actual data as it is but before transferring the data to admin node, we put some changes in one copy of the data and apply that copy for transportation in this copy of data we perform data transformation technique. Now we will transfer this valuable copy of data to the admin node where he will apply reverse process which we have applied on client side of data for our database.

III. IMPLIMENTATION WORK

For implementation work, we taking the customer database that contains attributes like C_Age, C_Weight and C_Height then we apply transformation technique over dataset in order to provide privacy. For analysis purpose we use the K means clustering technique. For implemented we use weka tool. Customer dataset is shown in table 1.

SL.NO	C_AGE (IN YEAR)	C_WEIGHT(IN KG)	C_HEIGHT (IN FEET)
1	2	12	2.7
2	10	22	4.8
3	20	47	4.6
4	25	63	5.1
5	12	40	5.8
6	30	57	5.4
7	20	48	5.5
8	18	68	5.7
9	12	70	5.8
10	28	50	5.8
11	26	53	5.5
12	31	43	5.7
13	17	59	5.7
14	30	55	5.8
15	15	63	5.5
16	23	52	5.8
17	37	72	5.5
18	30	67	6.0
19	24	54	6.1
20	16	63	5.6
21	13	61	4.7
22	19	73	5.7
23	34	82	5.5
24	20	75	6.0
25	42	44	5.7

After that we are applying the min_max normalization and transformation technique the dataset is shown in table 2.

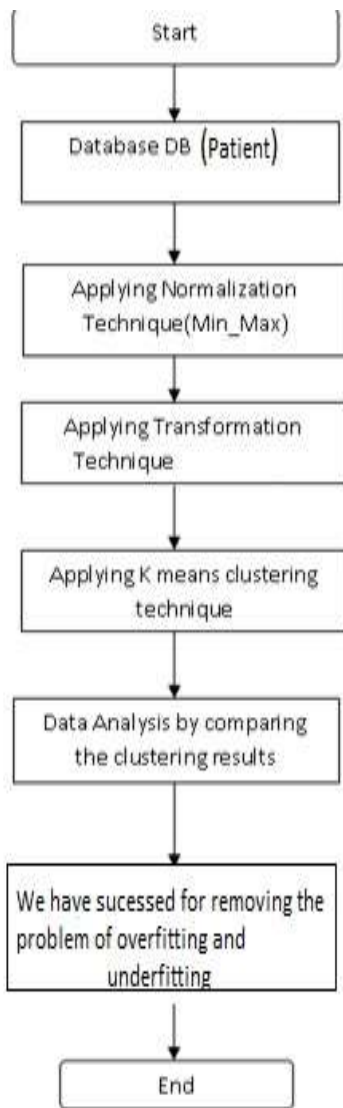| SL.NO	C_AGE (IN YEAR)	C_AGE (IN YEAR) AFTER DATA MODIFICATION TECHNIQUES
1	2	5.0
2	10	25.0
3	20	50.0
4	25	62.49
5	12	30.0
6	30	75.0
7	20	50.0
8	18	70.0
9	12	80.0
10	28	45
11	26	65.0
12	31	47.5
13	17	42.5
14	30	75.0
15	15	37.5
16	23	57.5
17	37	67.5
18	30	75.0
19	24	60.0
20	16	40.0
21	13	31.829
22	19	70.853
23	34	83.048
24	20	102.560
25	42	73.292

Fig. 1. Flow Chart of Work
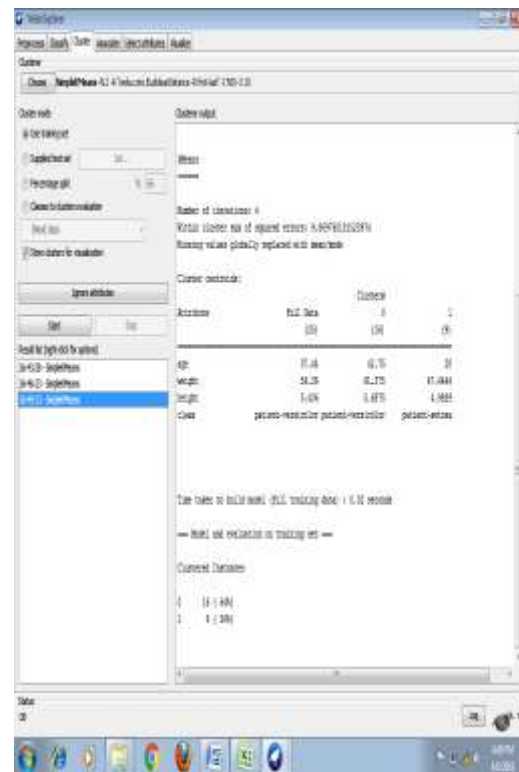

Fig. 2. Clustering on original dataset.

Fig. 3. Clustering results of updated dataset

IV. RESULT & COMPARISON

Results have been compared with the earlier work [1] in which instigator has min_max normalization process for providing privacy. Our approach reduces the difficulty of over fitting and

under fitting. Comparison between earlier and proposed work 3.

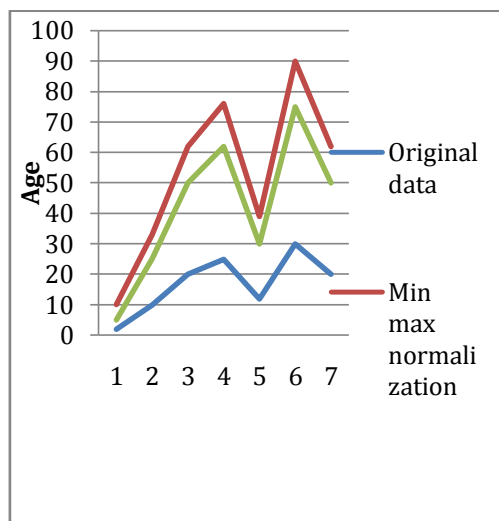| Serial Number	Original Data Values	Base Paper	Proposed System
1	2	10	5
2	10	33	25
3	20	62	50
4	25	76	62.49
5	12	39	30
6	30	90	75
7	20	62	50

Fig. 4.   Comparision between the earlier paper and proposed work.

V. CONCLUSION AND FUTURE WORK

This effort dealt with data adjustment technique for providing privacy preservation. We can enlarge the projected work by applying the concept of spread database for achieving privacy. We can apply other machine learning techniques and privacy preserving technique to provide the maximum security.

REFERENCES

[1]   **Adekitan, A.I., Abolade, J. & Shobayo, O.** Data mining approach for predicting the daily Internet data traffic of a smart university. International Journal of Big Data **6,** 11 (2019).

[2]   **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 2017.

[3]   **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 2016

[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, 2015.

[6]   **Lindell, Y., Pinkas, B,** "Privacy preserving data mining", In: Proceedings of 20th Annual International Cryptology Conference (CRYPTO), 2014

[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, 2013.

[8]   **Han, J. Kamber, M**, "Data mining Concepts and Techniques". Morgan Kaufmann, San Francisco, 2013

[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, 2012

[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, 2011

[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.101-1

IJAEM