

Achieving Highest Privacy Preservation in Data Mining Using Data Modification Technique

Naushaba Aafreen Khan¹, Sonia Bajaj²

¹M.tech Scholar, Computer Science and Engineering Department, G H Raisoni University, M.P. ²Head of Department, Computer Science and Engineering, G H Raisoni University, M.P.

Submitted: 20-03-2021	Revised: 01-04-2021	Accepted: 05-04-2021

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.

DOI: 10.35629/5252-030312301234 Impact Factor value 7.429 | ISO 9001: 2008 Certified Journal Page 1230



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	







das Relation (1)	Vokaros hatanthines -Option"	58-131			
(hirse)	herest				
d in family					
(and the second	and the second				
Instead	3.10				
- Contraction	0.000	00220			
	ADC 0.10	ranani di selamini kerem	e stelener	i.	
Printer Distant	form the	fails score of	2 BELLE		
- accession and a second					
inter strikelte	Takes we	1.00		Carrier	
	Antideur.	. 8	L fey	1.1	.1
241			供	40	(5
Instantige data grand	6		2.6	81 3	i an
A	wian		61.30	sizh d	tiasi
	ter ph		1.19	1411	Let
	1249	00180.40	THE DOMEST	actor inclusion	a de la
	Tak Mexi 1	table and the in-	and mar 1.1	-	
	1.000				
	- her of	estation in table	(m)		
	Connect Tel	1000			
	F 36.14	81.			
	15 12				
	14				
Fig. 2.	Clusteri	0 ng on (origina	al datas	et
Fig. 2.	Clusteri	0 ng on (origina	al datas	et
* Fig. 2.	Clusteri	on on o	origina	al datas	et.
Fig. 2.	Clusteri	ong on d	origina	al datas	et
Fig. 2.	Clusteri	ong on d	origina	al datas	et.
Fig. 2.	Clusteri	ong on d	origina	al datas	.et
Fig. 2.	Clusteri ni sale activo 2014 (1012)	ng on d	origina	al datas	.et.
Fig. 2.	Clusteri Clusteri alaan	ong on d	origina	al datas	.et.
Fig. 2.	Clusteri Clusteri a Met Menne Prof (1012)	o ng on o	origina	al datas	.et.
Fig. 2.	Clusteri clusteri adat	ong on d	origina	al datas	et.
Fig. 2.	Clusteri Clusteri al Mar. derman	o ng on o	origina	al datas	.et.
Fig. 2.	Clusteri Clusteri a Mat Anter-Strint (2012) Anter di unanteri	o ng on o	origina	al datas	.et.
Fig. 2.	Clusteri Clusteri Material (10-12)	ng on o	prigina	al datas	.et.
Fig. 2.	Anter of Least are of A	ng on o		al datas	et.
Fig. 2.	Clusteri Clusteri al Main Anter et descare et de State et descare et de State se de set	ng on o	origina ense	al datas	et.
Fig. 2.	Anter of constants of the formula of	ng on o	origina succes	al datas	et.
Fig. 2.	Anter of instance of formation of the sector	ng on o ng on o pet turo side		al datas	et.
Fig. 2.	Clusteri Clusteri al Mate Anter Hisid (10:1) Ante i instance i Nate i instance i Curie ostani Europeine pinip Curie ostani Europeine	pert may 1465 relation to 1465 relation and may hill bes	enser and and and and and and and and and and	al datas	eet.
Fig. 2.	Clusteri Clusteri an Ade Anter Clusteri Ade downed Ade	ng on o		al datas	et.
Fig. 2.	Clusteri Clusteri attes Head (10:1) Atte i instance i Sate of instance of Sate of instance of Sate Sate of Instance of Sate Sate Sate Sate Sate Sate Sate Sate	ng on o ng on o peri tran 166 ngeot et an r		al datas	et.
A Fig. 2. Fig. 2. A ³⁴⁰ asse months asse months asses months asses asses bit when bit when bit when bits when	Clusteri Clusteri a Mate Mate di matana di Stata di Stata di Stata di Stata di Stata di Stata di Stata di Stata di Stata di Stata Stata di Stata di Stata di Stata di Stata Stata di Stata di Stata di Stata di Stata Stata di Stata di Stata di Stata di Stata di Stata Stata di Stata di Stata di Stata di Stata di Stata Stata di Stata di Stata di Stata di Stata di Stata di Stata Stata di Stata di Stata Stata di Stata	ang on o pet the 146 plot of sen hi he plot of sen hi he hi he	enser enser an an an an	al datas	et.
Fig. 2. Fig. 2.	Clusteri clusteri an det anter times Adde di instant (10:12) Adde di instant (10:12) Adde di instant (10:12) Adde di instant (10:12) Adde di instant (10:12) Clust obde no di anter Adde di instant (10:12) Adde di instant (10:12) Clust obde no di anter Adde di instant (10:12) Adde dinstant (10:1	ng on o pertenses steller spinet ernes steller fill has the second	origina encore lates is encore encore is is is is is is is is is is is is is	al datas	et.
Fig. 2. Fig. 2. Fig. 2.	Anter di castante di Anter di Castante di Castante di Castante di Anter di Castante di Castante di Anter di Castante di Castante di Castante di Castante di Anter di Castante	ang on o ng on o pool of any fill has life its its its its	Drigina 8825394 00 823594 19 8.35 8.45 8.45 8.45 8.45 8.45 8.45 8.45 8.4	al datas	et.
Fig. 2. Fig. 2. Fig. 2.	Clusteri Clusteri al Mate anter Strid (10:1) Atter i instanto (1) Cure saturation Cure saturation Atter i instanto (1) Cure saturation Atter i instanto (1) Atter i instanto	pert may 1465 replace an according for a second for a sec	origina enset anset anset anset anset anset anset anset	al datas	et.
Fig. 2.	Clusteri Clusteri ana dat ana data ana	Rent states Add part states Add part states Add Rent Rent Sci Sci Sci Sci Sci Sci Sci Sci	2000 2000 2000 2000 2000 2000 2000 200	al datas	et.
Fig. 2.	Clusteri Clusteri Anter Clusteri Anter Clusteri Ant	ng on o ng on o pet two 550 piot et act fil bas fil bas fil bas	Drigina 800094 80000 80000 80000000000	al datas	et.
A Fig. 2. Fig. 2.	Clusteri Clusteri al Mate Anter Hand (10-13) Ante Hanswick State License H State License H Sta	ng on o ng on o pat mo 146 plot at set 15 sa 16	201331 201311 2013111 201311 201311 201311 201311 201311 201311 201311 201311 2	al datas	et.

Fig. 3. Clustering results of updated dataset

Current Datione (18 (88) 1 1 5 28)

N 15 8 0

С

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

0 4 0



under fitting. Comparison between earlier and proposed work 3.

Serial Number	Original Data Values	Base Paper	Propo sed Syste m
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

International Journal of Advances in Engineering and Management ISSN: 2395-5252

IJAEM

Volume: 03

Issue: 03

DOI: 10.35629/5252

www.ijaem.net

Email id: ijaem.paper@gmail.com