ISSN 2320-9984
Volume 4 Issue 4 | December 2016 ISSN: 2320-9984 (Online)
International Journal of Modern Engineering & Management Research
Website: www.ijmemr.org

Hybrid Transformation Technique Based Privacy Preservation in Data Mining

Virendra Singroure
Research Scholar M.Tech.
Lakshmi Narain College of Technology
Jabalpur (M.P.), [INDIA]
Email: virendrasngrr@gmail.com

Sujeet Tiwari
Assistant Professor
Department of Computer Science & Engineering
Lakshmi Narain College of Technology
Jabalpur (M.P.), [INDIA]
Email: sujeet.tiwari08@gmail.com

***Abstract*—*Data mining is the process of extracting the useful patterns and knowledge from the large amount of databases. Data mining has attracted a big deal of attention in the IT industry and in society in recent years, due to the availability of large amount of data and the imminent need for converting such data into useful information and knowledge. In our work we provide two level securities by using hybrid transformation technique. For performing the clustering operation we use k means clustering technique, in k means clustering technique we divide the given data values into the k number of clusters. For experimental purpose we use a dataset (promice dataset) and perform all operations in weka tool. Weka tool is a data mining tool, by using this tool we can perform the data mining operations like clustering, association and many more. Our work gives the better privacy as compared to the previous work.*

1. INTRODUCTION

Data mining is the 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. Privacy is a state in which one is not disturbed or observed by other persons. Many methods techniques and algorithms are already defined and presented for privacy preserving data mining. Data mining has attracted a big deal of attention in the IT industry and in society in recent years, due to the availability of large amount of data and the imminent need for converting such data into useful information and knowledge. This information and knowledge can be used for the applications like fraud detection, ranging from market analysis, customer retention to production controls and science exploration.

Privacy Preserving Technique

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 the output knowledge of data mining by using several techniques when this output data is valuable and private.

2. PROPOSED WORK

In this work we are going to take a database that is promise database. Now we emphasize on security issues as while communicating data from one place to other we need to provide security to our database for that first we use promise dataset. We apply hybrid transformation technique (i.e. translation and scaling) to the data due to which intruder will have to work a lot in order

to crack this valuable information and our data will be secure for communication. After that we will apply k means clustering techniques in order to check our data is preserved or not.

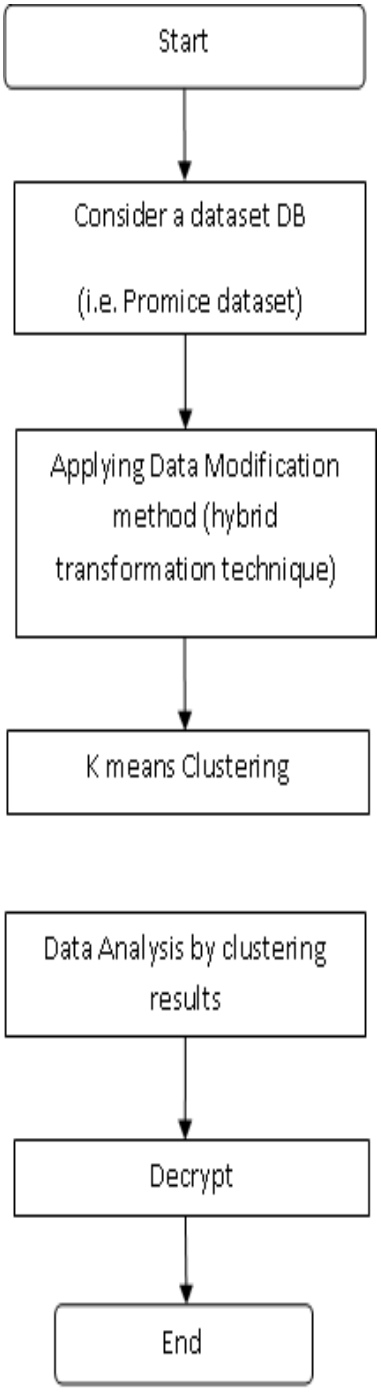

Figure 1: Flow chart

3. Implementation & Result

In the implementation work, we are taking the promise database (i.e. promice dataset) that contains seven attributes Project, TeamExp, ManagerExp, YearEnd, Length, Effort and language then we apply the hybrid 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 and then we implemented this work with the help of Weka tool. For implementation purpose considers a promise database, which is shown in table 1.

Table 1: Promice Dataset

S. No.	Project	Team Exp	Manager Exp	Year End	Length	Effort	language
1	1	1	4	85	12	5152	1
2	2	0	0	86	4	5635	1
3	3	4	4	85	1	805	3
4	4	0	0	86	5	3829	2
5	5	0	0	86	4	2149	1
6	6	0	0	86	4	2821	1
7	7	2	1	85	9	2569	2
8	8	1	2	83	13	3913	1
9	9	3	1	85	12	7854	1
10	10	3	4	83	4	2422	1
11	11	4	1	84	21	4067	3
12	12	2	1	84	17	9051	2
13	13	1	1	84	3	2282	1
14	14	3	4	85	8	4172	1
15	15	4	4	85	9	4977	2
16	16	3	2	85	8	1617	1
17	17	4	3	85	8	3192	1
18	18	4	4	86	14	3437	2
19	19	3	4	87	14	4494	2
20	20	4	2	86	5	840	1
21	21	4	4	86	12	14973	1
22	22	2	4	85	8	5180	1
23	23	2	4	86	5	5775	1
24	24	4	1	87	20	10577	2
25	25	3	4	86	19	3983	2

After applying the hybrid transformation technique result is shown in table 2.

Table 2: Promice dataset after applying the hybrid transformation technique

S.No.	Project No.	Effort before applying the hybrid transformation technique	Effort after applying the hybrid transformation technique
1	1	5152	5777.2
2	2	5635	6308.5
3	3	805	995.5
4	4	3829	4321.9
5	5	2149	2473.9
6	6	2821	3213.1
7	7	2569	2935.9
8	8	3913	4414.3
9	9	7854	8749.4
10	10	2422	2774.2
11	11	4067	4583.7
12	12	9051	10066.1
13	13	2282	2620.2
14	14	4172	4699.2
15	15	4977	5584.7
16	16	1617	1888.7
17	17	3192	3621.2
18	18	3437	3890.7
19	19	4494	5053.4
20	20	840	1034
21	21	14973	16580.3
22	22	5180	5808
23	23	5775	6462.5
24	24	10577	11744.7
25	25	3983	4491.3

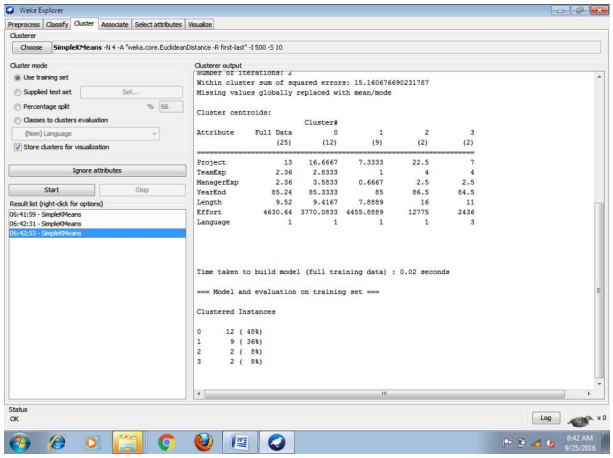

Figure 2: K means clustering before applying the hybrid transformation technique (for k=4)

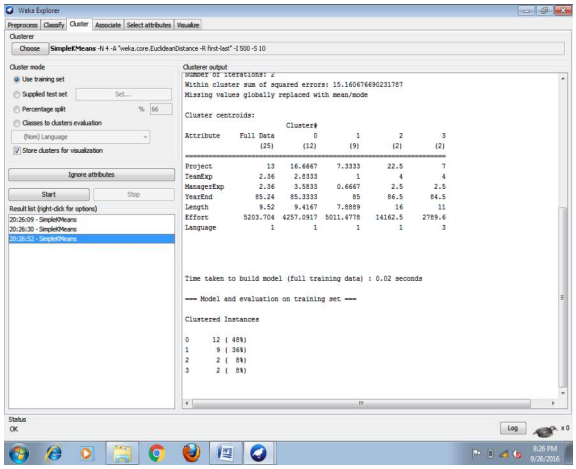

Figure 3: K means clustering after applying the hybrid transformation technique (for k=4)

4. COMPARISON

Obtained results have been compared with the previous work in which author has proposed privacy preservation technique which is based on min _max normalization.

Table 3: Comparison Table

S. No.	Project No.	Effort (Original Data)	Effort after applying the min_max normalization technique	Effort after applying the hybrid transformation technique
1	1	5152	6857	5777.2
2	2	5635	7547.5	6308.5
3	3	805	1547	995.5
4	4	3829	5452.7	4321.9
5	5	2149	3374	2473.9
6	6	2821	4242.2	3213.1
7	7	2569	3921.5	2935.9
8	8	3913	5475.8	4414.3
9	9	7854	9821.5	8749.4
10	10	2422	3744.5	2774.2

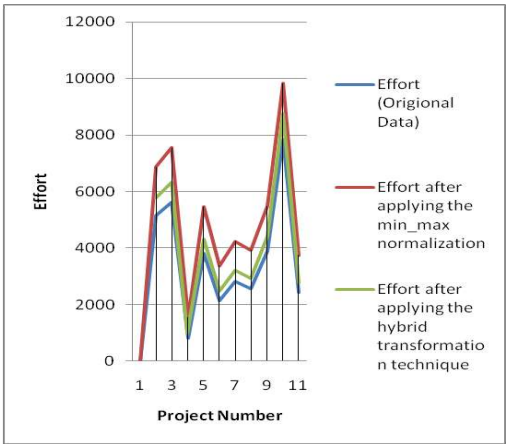

Figure 3: Comparison graph

5. CONCLUSION & FUTURE WORK

This work is based on hybrid transformation technique (i.e. translation and scaling) to provide privacy to our dataset. This technique transforms the original data into privacy- preserved data which maintains the inter relative distance among the data. Our experiments have proven that performing k-means clustering on the modified data produces same clustering results as original data. So we can say we have succeeded for achieving both privacy and accuracy. We have tested this technique for numerical data set.

The future scope of this proposed technique is to extend the same over categorical data and apply other techniques in order to preserve the privacy.

REFERENCES:

[1] **G.Manikandan et. al** "Achieving Privacy in Data Mining Using Normalization". In: Proc. Of Indian Journal of Science & Technology (IJST), Vol.6,No.4,April 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. Muthu Lakshmi, 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 majority36 abstract)," 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.