Procuring Privacy Preservation in Data Mining by applying Normalization and Randomization

Subhashree P¹, Dr. G.Gunasekaran²

¹Research Scholar CSE, Sathyabama Institute of Science and Technology ²Principal, JNN Institute of Engineering, Chennai,

Abstract:

The principle target of Data mining is to separate verifiable, beforehand obscure and conceivably valuable patterns from information. The removed examples are applied in a few areas like advertising, climate determining, picture investigation and in clinical finding. At the point when the association begin circulating the information for mining measure, the protection of information might be penetrated. There is a need to protect the security of people which can be accomplished by utilizing Privacy safeguarding information mining. Security Preserving procedures are the current one to conquer this issue. It gives fitting information mining results without improving the first information esteems and subsequently gives protection and exactness. In this paper, we have examined the utilization of standardization methods in accomplishing security. We have looked at the consequences of these procedures and from the exploratory result, it tends to be inferred that our methodology safeguard both security and privacy.

Keywords: Privacy, Clustering, Accuracy, K-Means, Normalization.

I.INTRODUCTION

Information Mining is the way toward discovering irregularities, examples and connections inside huge informational indexes to foresee results of private and sensitive data. The traditional privacy preserving methods are not up to a mark. Protection in terms of legitimate and business issues may keep the clients from straightforwardly sharing delicate information. Delicate information typically incorporates data regarding individual medical data, economic privacy etc. Thus, Information Mining without penetrating information security is a significant test .Hence the strategies that permit the information extraction from information, while protecting privacy are known as PPDM procedures.

In this paper, we have analysed the use of normalization techniques like Min-Max normalization, Z-score normalization, decimal scaling methods and randomization methods regarding privacy and accuracy. K-means clustering algorithm and a statistical approach of randomization methods are discussed to ensure privacy and accuracy.

II LITERATURE SURVEY

The investigation of Privacy-Preserving Data Mining strategies began since 2000 [1] covering, development approximately in two categories: Bother Base strategy [1, 2] and Secure Multi-Party Computation Base procedure [3, 4].

In [1] a lot of cross breed information change is utilized to address security issue. Misclassification blunder is utilized to gauge the proficiency of the bunching technique in the work.

In [2] a methodology is proposed to save the data utilizing fluffy rationale. Fluffy rationale is an essential control framework that depends on the degrees of condition of the information and the yield relies upon the condition of the information and pace of progress of this state. The blemishes related with this methodology is, it maps all qualities in the scale 0-1.

In writing [7, 8], an AI calculation to be specific, Very Fast machine Leaning (VFML) has been proposed, though this technique relies upon deciding an upper limit to be applied as information things test in each progression of the calculation.

In [9] min-max standardization procedure is utilized for protecting security. Here analysis was done only with numerical data.

In [10] maintaining data privacy using association rule mining was done.

III PROPOSED WORK

The objective of information standardization is to lessen and even dispose of information repetition. Various normalization techniques are available in literature. Here, we set forward a methodology, for security protecting utilizing Min-max Normalization.

S.No	Name	age	Gender
1	Priya	4	Female
2	Mithun	12	Male
3	Lakshmi	20	Female
4	Chithra	26	Female
5	Karthik	30	Male

Table 1(a) Original Data

S.No	Name	age	Gender
1	Priya	10	Female
2	Mithun	34	Male
3	Lakshmi	59	Female
4	Chithra	77	Female
5	Karthik	90	Male

Table 1(b) MinMax Normalized data

Min-max standardization plays out a direct change on the first information. For mapping a value, v of an attribute A from range [minA, maxA] to a new range [newminA, newmaxA], and the computation is given by.

$$\frac{v - \min_{A}}{\max_{A} - \min_{A}} (new_{max_{A}} - new_{min_{A}}) + new_{min_{A}}$$

Where V is the new value in the required range.

The principle preferred position of Min-Max standardization is that, it monitors the connections between the first information esteems.

Table 1(a) is the example informational collection utilized for our investigation. Table 1(b) is the comparing standardized qualities for the 'Age' property subsequent to applying min-max standardization. The steps involved in our approach is depicted as a flow diagram as shown below. Fig. 1 shows the proposed framework



Entire Procedure of our proposed work is shown below



K-Means clustering:

Bunching strategies are utilized to distinguish gatherings of comparative articles in a multivariate informational index. The calculation is portrayed underneath



IV Randomization

The Randomization strategy is one of the moderate and effective methods, used to guarantee the security of every client. The value of randomization method is that, it needn't bother with information on dissemination of different records in the information. The randomization technique can be executed at information assortment time, and doesn't need the use of a believed worker containing all the first records to play out the anonymization cycle. In the current methods of Privacy safeguarding Data Mining, the idea of randomization is a wonderful methodology. Information disclosure measure gives the soundness among utility and protection. This is accomplished by coordinating clamour inside information. At last, this is sent to the beneficiary gadget. The beneficiary gadget gets the information by dissemination reproduction calculation. Here, clamour is mixed into the information at information assortment time. The Randomization strategy is handily actualized at information assortment time, in light of the fact that the additional commotion is unconstrained of the deed of other information records.

Primary objective for the exploration of protection safeguarding Data Mining is to build up these sorts of results that may furnish the security of information alongside information consistency and secrecy with low computational intricacy.

By this randomization procedure, we secure the information of client by giving them some irregular changes inside their information. Prior to sending those information, we look at not many important subtleties and furthermore create a few commotions in network. Two methodologies are the kind of mathematical randomization and furthermore the thing set randomization. Here, the commotion might be built through including or by increasing the arbitrary qualities with the mathematical information or through eliminating the genuine things and furthermore consolidating not many "copy" values inside the traits set.



The above figure shows plainly, that subsequent to perusing the dataset, we segment the information into semi and delicate trait. We are worrisome, semi characteristic with the Root Mean Square estimation of past segment. After, we join information of semi and touchy trait, at that point picking, the irregular section. Presently, isolate section into three arrangements, for example, 1, 2, and 3. After the division of the section, sort entire information as indicated by

segment. Presently select irregular line. At last, separate line in three classifications, for example, 1, 2, 3 and Sort entire information as per line. In this way, Randomization has been applied on dataset.

IV RESULTS AND SIMULATIONS

In this paper, we have utilized min-max standardization to accomplish security and exactness during information mining and precision is tried utilizing K-implies grouping. We have likewise determined the adequacy and of the method utilizing K implies Clustering calculation. The accompanying depictions depend on the result of K-mean bunching calculation for 2 groups with the first and Normalized information.

Clustered value	es are:
cluster1	
5 [10, 15, 20, 2 [,] cluster2	4, 30]
5 [37, 40, 45, 50	0, 60]

Cluster f	for origi	nal data

Cluster for Min-Max data

[53, 58, 66, 74,

Clustered values are:

18, 26, 32, 42]

901

cluster1

luster2

Table 2 depicts the grouping aftereffect of unique and standardized information. Table 3 and Fig.2 sums up the examinations among Original and the Normalized Data.

K is 2	1- Cluster	2- Cluster
Original Data	[10,15,20,24,30]	[37,40,45,50,60]
MinMax Data	[10,18,26,32,42]	[53,58,66,74,90]

Original Data	Min-Max
10	10
15	18
20	26
24	32
a30	42
37	53
40	58
45	66
50	74
60	90

Table IIINormalization Process

The below figure shows the comparison among original and normalized data. We have taken Z-score data and decimal scaling data also for comparison.



We have thought about two boundaries for examination of the exhibition of the proposed work over existing work

- a. Privacy
- b. Execution Time

The table illustrates the privacy level of existing and proposed work

Existing Work	Proposed Work
6.233	12.61



The above diagram shows the visual portrayal of security level of existing and proposed work.

The execution season of existing and proposed work is given in the table below

Existing Work	Proposed Work
0.5	0.44

The Graphical representation of execution time comparison is shown below



V CONCLUSION

In this work, we have examined the utilization of standardization procedure to safeguard information security. We did the three standardization procedure on an example informational index containing 10 components with values 10, 15, 20, 24, 30, 37, 40, 45, 50, 60 are utilized and from the investigations and results, it is obviously clear that Min-Max standardization has less number of misclassification blunders. This paper likewise proposed a sorted randomization PPDM strategy to help comprehend the issue. It unmistakably shows that the presentation of the proposed function as contrast with existing work is more proficient.

VI REFERENCES

 [1] Agrawal, R. and Srikant, R., "Privacy-Preserving Data Mining,", Proceeding of the ACM SIGMOD Conference on Management of Da-ta, Dallas, Texas, U.S.A., pp. 439-450 (2000).

[2] Kumari, P. K., Raju, K. and Rao, S. S., "Privacy Preserving in Cluster-ing Using Fuzzy Sets," Proceedings of the 2006 International Conference on Data Mining, Las Vegas, Nevada, U.S.A., pp. 26 29 (2006).

[3] Clifton, C., Kantarcioglu, M., Vaidya, J., Lin, X. and Zhu, M. Y., "Tools for Privacy Preserving Distributed Data Mining," ACM SIGKDD Explorations Newsletter, Vol. 4, pp. 28 34 (2002).

[4] Meregu, S. and Ghosh, J., "Privacy-Preserving Distributed Clustering Using Generative Models," Proceedings of the 3th IEEE International Conference on Data Mining, Melbourne, Florida, U.S.A., pp. 211-218 (2003).

[5] G.Manikandan,N.Sairam, R.Sudan and B.Vaishnavi "Shearing based Data Transformation Approach for Privacy Preserving Clustering " Third International Conference on Computing,Communication and Networking Technologies (ICCCNT-2012), SNS College of Engineering,Coimbatore,July 26-28,2012.

[6] B.Karthikeyan,G.Manikandan,Dr.V.Vaithiyanathan,A Fuzzy Based Approach for Privacy Preserving Clustering in Journal of Theoretical and applied information Technology, Vol 32(2), October 2011,118-122.

[7] Hulten, G., Spencer, L. and Domingos, P., "Mining Time-Changing Data Streams," Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francis-co, California, U.S.A., pp. 97 106 (2001).

[8] G.Manikandan,N.Sairam,S.Sharmili,S.Venkatakrishnan, Data Masking – A few new Techniques – international conference on research and development prospects on engineering and technology(ICRDPET -2013), E.G.S Pillay Engineering college, Nagapattinam, march 29-30,2103.

[9] Domingos, P. and Hulten, G., "Mining High-Speed Data Streams," Proceedings of the Association for Computing Machinery 6th Inter-national Conference on Knowledge Discovery and Data Mining, Bos-ton, U.S.A., pp. 71 80 (2000).

[10] S. J. Rizvi and J. R. Haritsa. Maintaining data privacyin association rule mining. InProceedings of the 28thInternational Conference on Very Large Data Bases,Hong Kong, China, August 2002.

[11] A.S.Shanthi, , Dr. M. Karthikeyan" A Review on Privacy Preserving DM"IEEE International Conference on Computational Intelligence and Computing, 2012.

[12] Mohnish Patel, PrashantRichariya, AnuragShrivastava, (2013),,,,,A review paper on Privacy-Preserving Data Mining", Reviewarticle onScholars Journal of Engineering and Technology (SJET), pp.359-361.

About Author



Mrs.P.Subhashree has done M.E at Anna University of Technology, Trichy and B.Tech at JJCET Trichy. She had 6+ years of teaching experience. Pursuing Ph.D. at Sathyabama Institute of Science and Technology. Chennai.



Dr.G.Gunasekaran, Principal J.N.N Institute of Engineering Chennai, has obtained his B.E from National Engineering College and M.E from Jadavpur University and Ph.D. also from the same University and he has published 13 International Journals and has attended 7 International Conferences. He has got a profound experience of 27 years.