# Normalization and Transformation Technique Based Efficient Privacy Preservation In Data Mining

**Ashok Kumar Mishra**
*M. Tech Research Scholar*
*Shri Ram Group of Institution,*
*Jabalpur (M.P.), [INDIA]*
*Email: ashokmishra118@gmail.com*

**Anupam Choudhary**
*Lecturer*
*Kalaniketan Polytechnic College*
*Jabalpur (M.P.), [INDIA]*
Email: *choudharyanupamj@gmail.com*

**Sapna Choudhary**
*Assistant Professor,*
*Head of the Department*
*Department of Computer Science & Engineering*
*Shri Ram Group of Institution*
*Jabalpur (M.P.), [INDIA]*
*Email:choudharysapnajain@gmail.com*

## ABSTRACT

*Data mining is the process of extracting the important information and knowledge from the large amount of databases. Data mining has attracted a big attention in IT industry and in societies in modern year due to the availability of huge amount of data and need for converting such data into the useful information. This important information can be used in various application areas like fraud detection, ranging from market analysis, customer retention to production controls and science exploration. When this stored data is transferred from one place to another we require privacy preserving techniques because different types or hackers or attackers can disclose our private data. In our work we provide two level securities by using normalization and transformation technique. With the help of normalization technique we can convert given data values into the specified range and with the help of scaling transformation technique we can change the position of the given data objects. For performing the clustering operation we use k means clustering technique. For experimental purpose we use a dataset (customer dataset) and perform all operations in weka tool. Our work gives the highest privacy as compared to the previous work.*

***Keywords:***—*Data mining, privacy preservation, normalization and transformation technique, k means clustering technique.*

## I. INTRODUCTION

Data mining is the process in which we extract the useful patterns and knowledge from the large amount of databases. Now a days the databases are very large which consists of so much information but what we want to find is the relevant data from large amount of databases or want to find some interesting patterns which becomes very difficult with normal database management systems but with the use of data mining techniques we can find the hidden patterns and knowledge from large database system. So conclusion is that data mining as the knowledge mining, pattern extraction etc. But before applying the data mining techniques we need to apply some

other 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. The data mining techniques are also used on Bio-Database for analyzing and acquiring the different relations in the food condition of market or environmental conditions and many other conditions to find the relations which can tell the cause of any disease at very early stage so that proper precautions can be taken. Medical science and market analysis is a field where large amount of data is gathered and collected from many sources now the challenge is to find the appropriate information and pattern from that data so that it can be used for further research to find some valuable results for the patients and customers but security is the major issue we should be very careful while sending data from one place to other otherwise it may create some harmful effects.

## II. PROPOSED WORK

In this work we are going to take a database that is customer dataset. We Now discuss about security issues as while communicating the data from one place to other we need to provide security to our database. When we need to communicate this important data with the admin first we need privacy as there is possibilities that someone in between the communication of data may change this important data which will cause many hazards so in order to secure our communications from intruders, we will modify our data. In our work we provide two level security by using normalization and transformation technique. In this technique we keep the original data as it is but before sending the valuable data to admin, we put changes in one copy and use that copy for communication in this copy we perform normalization and transformation technique due to which intruder will have to work a lot in order to crack this valuable data and our data

will be secure for communication. Now we will send this valuable copy to the admin where he will apply the same methods which we have applied on client side for our database.
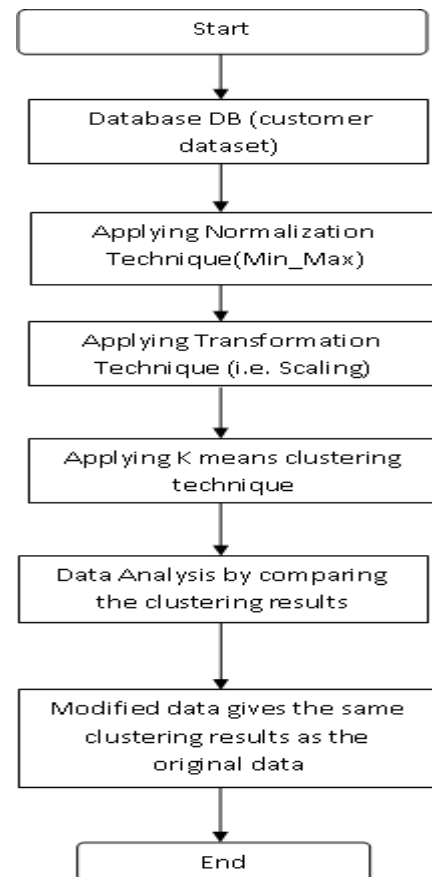


*Figure. 1 Flow chart*

## III. IMPLEMENTATION

In implementation work, we are taking the customer database (i.e. customer dataset) that contains three attributes like customer age, weight and height then we apply the min_max normalization technique and scaling transformation technique on our dataset for providing the highest privacy. For analyzing the dataset we use the K means clustering technique and implemented this work with the help of WEKA tool. Customer dataset is shown in table 1.

## Table 1 Customer dataset

| Serial Number | Customer Age (in year) | Customer Weight ( in kg) | Customer Height ( in feet) |
|---|---|---|---|
| 1 | 2 | 14 | 2.7 |
| 2 | 10 | 17 | 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 |

## Table 2 After applying the normalization and transformation technique

| Serial Number | Customer Age(in year) | Customer Age(in year) After Normalization and Transformation |
|---|---|---|
| 1 | 2 | 6 |
| 2 | 10 | 18 |
| 3 | 20 | 33 |
| 4 | 25 | 40.5 |
| 5 | 12 | 21 |
| 6 | 30 | 48 |
| 7 | 20 | 33 |
| 8 | 18 | 30 |
| 9 | 32 | 51 |
| 10 | 28 | 45 |
| 11 | 26 | 42 |
| 12 | 31 | 49.5 |
| 13 | 17 | 28.5 |
| 14 | 30 | 48 |
| 15 | 15 | 25.5 |
| 16 | 23 | 37.5 |
| 17 | 37 | 58.5 |
| 18 | 30 | 48 |
| 19 | 24 | 39 |
| 20 | 16 | 27 |
| 21 | 13 | 22.5 |
| 22 | 19 | 31.5 |
| 23 | 34 | 54 |
| 24 | 20 | 33 |
| 25 | 42 | 66 |

After applying the min_mix normalization and scaling transformation technique the dataset is shown in table 2.
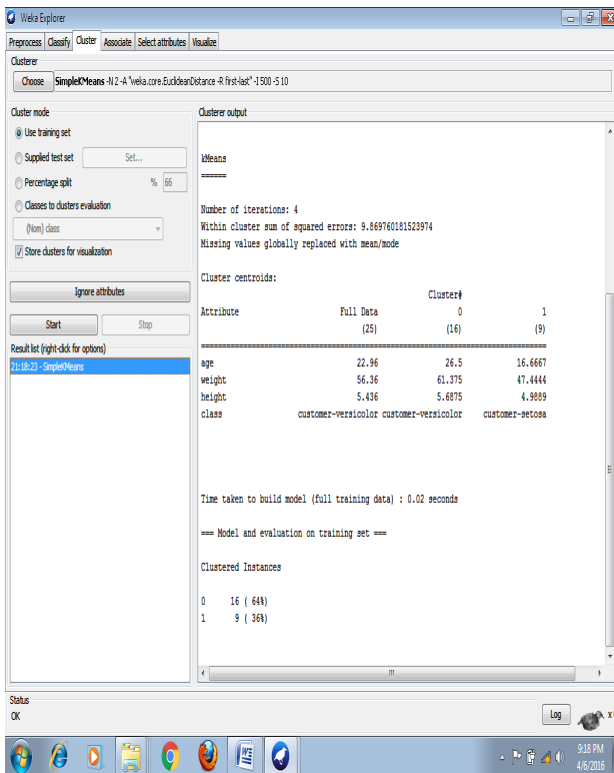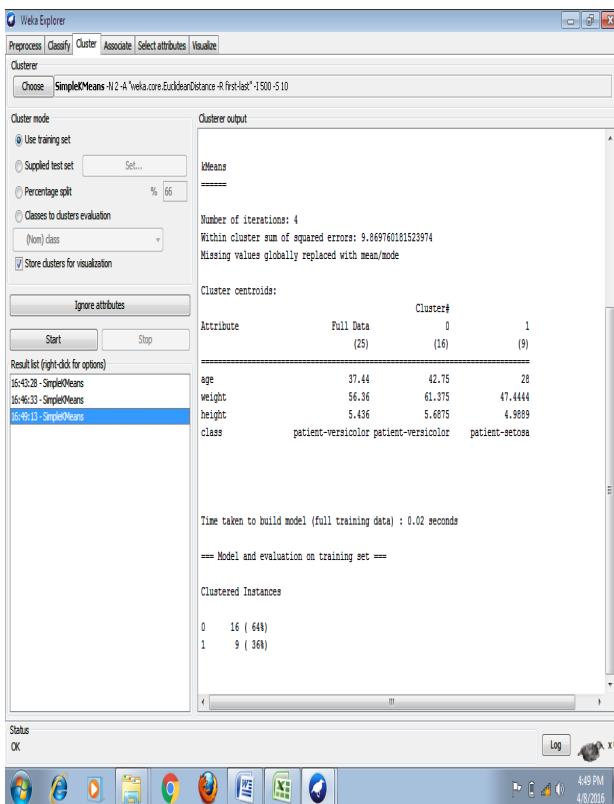
*Figure. 2: Clustering on original dataset*



*Figure. 3: Clustering on modified dataset*

## IV. RESULT & COMPARISON

Obtained results have been compared with the base paper [1] in which author has proposed privacy preservation in data mining based on min _max normalization technique. Proposed approach provides two level security and transforms the original data values into privacy- preserved data maintaining the inter relative distance among the data. The comparison between the base paper and proposed method is shown in table 3.

**Table 3: Comparison between the base paper and proposed work.**

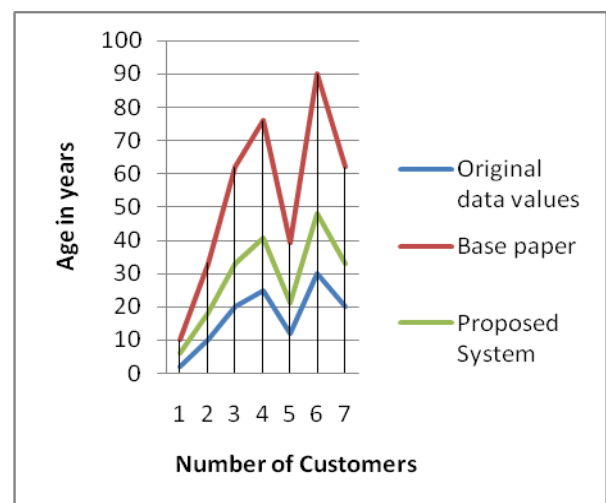| S.No. | Original data values | Base paper | Proposed System |
|-------|---------------------|------------|-----------------|
| 1 | 2 | 10 | 6 |
| 2 | 10 | 33 | 18 |
| 3 | 20 | 62 | 33 |
| 4 | 25 | 76 | 40.5 |
| 5 | 12 | 39 | 21 |
| 6 | 30 | 90 | 48 |
| 7 | 20 | 62 | 33 |



*Figure 3: Comparison between the base paper and proposed work.*

## V. CONCLUSION & FUTURE WORK

In this work we have dealt with min_max normalization and scaling transformation technique 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.

The future scope of this proposed technique is to extend the same over categorical data and apply other techniques for preserving the privacy. We can also extend the proposed work by using the concept of distributed database in order to preserve the privacy and providing fault tolerance. We can also apply other privacy preserving techniques in order to provide the highest security.

## 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 partitioneddata," Ph.D. dissertation, Purdue University, 2004.

\* \* \* \* \*