

# Procuring Privacy Preservation in Data Mining by applying Normalization and Randomization

Subhashree P<sup>1</sup>, Dr. G.Gunasekaran<sup>2</sup>

<sup>1</sup>Research Scholar CSE, Sathyabama Institute of Science and Technology

<sup>2</sup>Principal, 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 Learning (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  $[\min_A, \max_A]$  to a new range  $[\text{newmin}_A, \text{newmax}_A]$ , and the computation is given by.

$$\frac{v - \min_A}{\max_A - \min_A} (\text{new\_max}_A - \text{new\_min}_A) + \text{new\_min}_A$$

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.









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 values are:
-----
cluster1
-----
5
[10, 15, 20, 24, 30]
cluster2
-----
5
[37, 40, 45, 50, 60]

```

Cluster for original data

```

Clustered values are:
-----
cluster1
-----
5
[10, 18, 26, 32, 42]
cluster2
-----
5
[53, 58, 66, 74, 90]

```

Cluster for Min-Max data

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

Table II -cluster Result

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]











