International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 8 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT183838

A Robust Privacy Preserving of Multiple and Binary Attribute by Using Super Modularity with Perturbation

Priya Ranjan, Raj Kumar Paul
Computer Science and Engineering, Vedica Institute of Technology, Bhopal, Madhya Pradesh, India

ABSTRACT

With the increase of digital data on servers different approach of data mining is applied for the retrieval of interesting information in decision making. A major social concern of data mining is the issue of privacy and data security. So privacy preserving mining come in existence, as it validates those data mining algorithms that do not disclose sensitive information. This work provides privacy for sensitive rules that discriminate data on the basis of community, gender, country, etc. Rules are obtained by aprior algorithm of association rule mining. Those rules which contain sensitive item set with minimum threshold value are considered as sensitive. Perturbation technique is used for the hiding of sensitive rules. The age of large database is now a big issue. So researchers try to develop a high performance *platform* to efficiently secure these kind of data before publishing. Here proposed work has resolve this issue of digital data security by finding the relation between the columns of the dataset which is based on the highly relative association patterns. Here use of super modularity is also done which balance the risk and utilization of the data. Experiment is done on large dataset which have all kind of attribute for implementing proposed work features. The experiments showed that the proposed algorithms perform well on large databases. It work better as the Maximum lost pattern percentage is zero a certain value of support.

Keywords : Association Patterns , Decision Making , Aprior Algorithm , Data Mining , Perturbation Technique

I. INTRODUCTION

Data mining is actually the process or technique of discovering patterns in huge data bases .It mainly involves artificial intelligence, machine learning, statistics, and database systems. The main goal of the data mining process is to extract information from a data base and convert them into an understandable structure for our further usage. Apart from the raw analysis, it involves following aspects that is database and data base management, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, processing of newly found structures, imagination, and online information updating.Data mining is referred as KDD process i.e., knowledge discovery in databases.The Data Mining process can be defined with the following steps:

(1) Selection step
(2) Pre-processing step
(3) Conversion step
(4) Data Mining
(5) Interpretation/Evaluation.

According to the CRISP-DM or Cross Industry Standard Process for Data Mining, there are following six stages:

(1) Business Consideration
(2) Data Consideration
(3) Data Creation
(4) Modelling process
(5) Evaluation
(6) Deployment

PRIVACY MODELS TYPES

1. Non-interactive privacy model: Database is to be sanitized first and then release it.
2. Interactive privacy model: Database is to be asked multiple questions and then answered adaptively.

II. ISSUES IN DATAMINING

Confidentiality: This is the main problem that arises in any huge databases. But sometimes it is needed because of law such as medical databases or for other business interests. However, in some situations the data sharing can lead for a mutual benefit.

Theoretical Challenges in High Dimensionality: Actually real datasets are extremely high dimensions, and this makes the process of privacy preservation tremendously difficult from a computational and effectiveness point of view as well.

Data Perturbation : In this technique data is modified by a random process. It apparently manipulates the sensitive data values by applying some mathematical formula such as addition, subtraction or any other. There are two types of data preprocessing that is attribute wise coding and coded data set wise. Average region method is used that disperse the continuous dataset. Sativa Lohiya and Lata Ragha has given a discrete formula: At (max) - At (min)/nt = size. At is continuous attribute, nt is number of discrete, and size is the length of the discrete interval.

Data Modification technique : This technique actually modifies the data set before its release. It actually modifies data in such a manner that the privacy is preserved in the released data set, whereas the data quality is improved very high that may serve the purpose of the release.

Noise Addition in Statistical Database : This technique was originally used for statistical databases that can maintain data quality in along with the privacy of individuals. It proved fruitful privacy preserving data mining as well.

Data Swapping : This technique was first applied by Dalenius and Reiss in the year 1982, for modification of categorical values in of secure statistical databases context. This method actually keeps all original values in the data set itself, while making the record re-identification difficult at the same time.

Suppression : In this technique sensitive data sets are deleted or suppressed prior to their release. This technique actually protect an individual person's privacy from intruders. An intruder have different tact and approaches in order to predict a sensitive value. Take for instance, a classifier, built a released data set, which is capable to predict a suppressed attribute value. However, this type of suppression of values results in information loss.

Distributed Privacy Preserving : As the name implies distributed ,it is methods that allows computation of aggregate statistics over the entire data set without having loss of the privacy of the individual datas from the various participants. Participants can assemble to get aggregate results, but may not be fully trustful in terms of distribution of personal datasets. So we have partitioned the datasets horizontally or vertically. This problem can be generalized across k number of parties by k argument function f(a1. . . bk). Many data mining algorithms are used for repetitive computations of many such primitive functions as the scalar dot product, secure sum etc. To compute the function f(a, b) or f(a1. . . , bk), a protocol is to be designed for exchanging information in such a manner that the function is computed without compromising the data's privacy.

Association Rule Mining : In our day to day life there exist a large number of databases in various fields like business management, administration, banking and social, health care services, environmental protection, security, politics and so on. Take for instance For example; point-of-sale does not contains any information about the customer's personal interests, age and occupation. Market basket analysis provides new perception of customer behaviour and has lead to higher business profits by maintaining better customer relations, good customer retention, improved product placements, better product development and fraud detection as well. Market basket analysis is not restricted to retail shopping but has also been applied in other business areas that includes

- credit cards transactions,
- telecommunication services
- commercial or banking services
- Insurance claims
- Study of medical patient case histories.

This mining technique generalises the market basket analysis and is beneficial in many other fields including genomics, text data analysis and Internet intrusion detection.

MARKET BASKET ID	MARKET BASKET CONTENT
1	Orange juice, soda water
2	Milk, orange juice, bread
3	Orange juice, butter
4	Orange juice, bread, soda water
5	Bread

Table: 1.1 Market Basket Analysis

Whenever a customer goes after purchasing through a point of sale, his contents in the basket are noted down. It provides large collections of market basket data providing information regarding the items sold and combinations of items sold to a particular customer.

The above table reveals that:

_ orange juice is in the four baskets,

_ soda water is in two baskets

_ half of the orange juice baskets also contain soda with it

_ The baskets having soda also contain juice

Association rules can be defined as follows: Let It = { it1,it2… itm} be a set of items. Let Dt be task data. Each transaction Tn is a set of items such that Tn ⊆ It. Each transaction is associated with an identifier, called TId. Suppose Ar be a set of items. A transaction Tn is said to contain Ar if and only if Ar⊆ Tn. An association rule is an implication of the form Ar =>Br, where Ar⊂It, Br⊂It and Ar⋂ Br = Nil. The rule Ar => Br holds in the transaction set D with support Sp, where Sp is the percentage of transaction in D that contains A ∪B. Support(s) can be defined as the percentage or fraction of records containing (Ar ∪ Br) to the total number of records in the database.

Support (A➔B) (AUB) / D

Confidence (A➔B) = (AUB) / X

Cryptographic Technique

In order to encrypt the sensitive data, cryptographic techniques are applied while preserving the data. This technique is very popular due to its security and safety features in case of sensitive attributes Different algorithms are available for this technique .But it has many disadvantages.

Condensation Approach : This is another approach introduced by Charu C. Aggarwal and Philip creates a constrained clusters in the data set and then produces pseudo-data. In this method we may contract or condense the data into variable groups of predefined size. For each group, there are certain statistics.

III. APPLICATIONS OF PRIVACY-PRESERVING DATA MINING

There are various privacy-preserving data mining applications in the following fields:

– Medical Databases
– Bioterrorism Applications
– Homeland Security Applications
• Credential verification problem
• Identity thievery
• Web camera examination
• Video-surveillance
• The watch-list problem
– Genomic privacy.

PREVENTION OF OUTSOURCE DATABASE

In [11] perturbation of dataset is done for providing security of the data that resides at server. Many of the organizations locate the data at server for general updates in price list, group, etc. This is done in the case of the numeric set of values.

$D(x, y) = 1/N (\sum E(y-x)^2)$

Without knowledge of whole method and parameter one can predict approx dataset which is much closer to the original set. One more methods is Linear Least Squares Error Estimation method.

$X(y) = Kx / Ky((y - \mu) + \mu)$

Where Kx and Ky are covariance of x & y while μ is mean of x.

BASIC MEASURES OF ASSOCIATION RULE

:Mainly the association rules can be categorised in two stages detecting the most occurring frequent items and generating the rules through it. The two important basic measures for the association rules are support (s) and confidence (c). Since there is very large database and the users mainly focus on frequently purchased items, usually limit value of support and confidence are basically defined by the user. Basically there are two parameters of Association Rule Mining (ARM) support and confidence. Support(s) of an association rule can be said as the proportion/fraction of all records comprises of A U B to the total number of records in the database. The count associated with each item is incremented by one on every turn the item is meets a different transaction Tn in database Db in its scanning process. The support count does not consider the amount of the item taken into account. The Support(s) is calculated by the following

Support (A->B) =(AUB) / Db.

Confidence : Confidence of an association rule can be said as proportion/fraction of all records that comprises of A U B to the total number of records that contain A, whereas the percentage if increases above the limit of confidence then an interesting association rule X->Y will be generated. Confidence (A->B) = (AUB) / A is a measure of confidence or strength of this rule, Let us assume that confidence of the association rule A->B is 80%, that is 80% of the transactions containing A also containing B along with it, similarly the rules specified minimum confidence is also pre-defined by users.

A priori Algorithm

Assume Db as a set of database transactions and there each transaction Tn contains set of items, called Tid. Suppose I= {I1, I2,..., Im} be an item set. An item set consist of collection of k items in a item set. If an item k set fulfils the minimum support (Min_sup) then it is considered as frequent k item set, denoted by Lk. Earlier Apriori algorithm generates a set of candidates, that is candidate k-item sets, denoted by Ck. If that candidate item sets fulfils the minimum support value then it is a frequent item sets. The description of the algorithm is as follows:

1. Assume minimum support threshold value as min_sup and the minimum confidence threshold value as min_conf.
2. The dataset is scanned first, the candidate 1-itemsets, c1, in which the number of occurrences of a single item in the dataset is found out. The set of frequent 1-itemsets, l1, is then found out, having those candidate 1-itemsets in c1 containing minimum

support. The algorithm uses l1 ∞ l2 to generate candidate 2-itemsets, c2.

3. The dataset is scanned again, from which 2 frequent itemsets and l2, is then determined, which contains only those candidate 2-itemsets in the c2 contains minimum support. Candidate 3-itemsets, c3 is then generated by l2 ∞ l2.

4. Repetitively we have to scan the dataset and then compare support count of each candidate in Ck-1 having min_sup, and generate the lk-1 , join the lk-1 ∞ lk-1 to generate Ck.

PRIVACY PRESERVING DATA MINING TECHNIQUES

The randomization method: In this method noise is added to the actual data in order to mask actual values of the attributes. The noise that is added is big enough that individual record values can be easily recovered

The k-anonymity model and l-diversity: This model was developed as there is possibility of indirectly identifying of records from the public databases as the combinations of the record attributes is able to properly identify each and every record. In this method, the granularity of data representation is reduced by generalization and suppression method.

Distributed privacy preservation: In various cases, individual entities want to derive aggregate results from the data sets that are partitioned around the entities. The partitioning can be horizontal (the records are distributed around multiple entities) or may be vertical (when the columns are distributed across multiple entities).

The randomization method : This method is traditionally being used in for data distortion by probability distribution and methods such as surveys which gives elusive answer bias due to privacy concerns.Suppose there is a set of data records referred by $Xr = \{xr1 \ldots xrN\}$. In each record x of Xr a noise component is added which is removed from the probability distribution fY (yr). These noise components are drawn independently, and are denoted $yr1 \ldots yrN$. Thus, the new set of distorted records are denoted by $xr1 + yr1 \ldots xrN + yrN$. The new set of records are denoted by $zr1 \ldots zrN$. So, if Xr be the random variable denoting the data distribution for the original record, Yr be the random variable describing the noise distribution, and Zr be the random variable denoting the final record, we have:

$$Zr = Xr + Yr$$
$$Xr = Zr - Yr$$

Thus the N instantiations of the probability distribution Zr are known, however the distribution Yr is known publicly. If there are large number of values of N, the distribution Zr can be estimated by methods such as kernel density estimation. If we subtract Yr from the approximated distribution of Zr, the original probability distribution Xr can be easily determined.

The *K*-Anonymity Framework: In this model, the most extreme feature in a table is that it strongly reveals private information, by joining it with other tables. In addition, the sensitive feature is a feature serves as the class label of each record. There are set of three features {Zip, Gender, and Age} is known as quasi-identifier feature set, while the feature {Diagnosis} is the sensitive feature.

Table 3.1 Patient Diagnosis Records in a Hospital

ZIP	GENDER	AGE	DIAGONSIS
4791*	Person	[35-39]	Cancer
4790*	Person	[30-34]	HIV+
4791*	Person	[35-39]	Flu
4791*	Person	[35-39]	Obesity
4790*	Person	[30-34]	Cancer
4790*	Person	[30-34]	Flu

Table 3.2 The K-anonymity Protected Table when K=3

ZIP	GENDER	AGE	DIAGONSIS
47918	Male	35	Cancer
47906	Male	33	HIV+
47918	Male	36	Flu
47916	Female	39	Obesity
47907	Male	33	Cancer
47906	Female	33	Flu

In Privacy Preserving Data mining technology field, k-anonymity has gained a lot of attention in the past years. The guarantee given by k-anonymity is that no information can be linked to groups of less than k individuals.

IV. ASSOCIATION RULE HIDING ALGORITHMS

A. Heuristic Approach This approaches can be further divided into distortion based schemes and blocking based schemes. For hiding the sensitive item sets, the distortion based scheme alters certain items in particular choose transactions that is from present to absent and absent to present. In the blocking based scheme certain items in selected transactions are replaced with unknowns.

B. Border Revision Approach This approach actually modifies borders in the network of the frequent and infrequent item sets collection and cover sensitive association rules. It takes the limit of the non sensitive frequent item sets and gradually modifies the data that causes minimum effect on the quality to collect the hide sensitive rules. Researchers have proposed various border revision approach algorithms BBA (Border Based Approach) that is Maxm–Minm1 and Maxm-Minm2 that are used to hide sensitive association rules.

PROPOSED METHODOLOGY

The data set is a combination of object of data and their attribute. An item is an attribute that comprises of value, e.g. {Race=black}. Take for example item set that is A, is a group of one or more set of items, e.g. {Foreign worker=Yes, City=NYC}. A classification law is an appearance A→ Ci, where Ci is a class item (decision of yes/no), and A is a group of items containing no class item, e.g. {Foreign worker=Yes, City=NYC} --> {hire=no}. A is called as the basis of the law.

Support: Support of an association law is defined like the percentage/fraction of data that consist of A U B to the total number of records in the database, Let D be the amount of datasets or records in the database.

Support (A→B) = (AUB) / D

Confidence: Confidence of an association law is defined as the percentage/fraction of the number of transactions contain A U B to the total number of records that contain A, in case if the

NAME	ID	AGE	SALARY	PURCHASE ITEMS		
Rahul	10001	24	27000	Jeans	Jacket	Shoes
Arun	10002	34	52000	Trouser	Jeans	Shoes
Rohit	10003	22	21000	jacket	Shoes	
Kuldeep	10004	39	23000	Jacket	Shoes	Jeans

percentage exceeds the threshold of confidence then an interesting association rule X→Y can be generated.

Confidence (A→B) = (AUB) / A

Table 4.1 Unprocessed dataset

Dr = { rahul, 10001, 24, 27000, jeans, jacket, shoes; arun, 10002, 34, 52000, trouser, jeans, shoes; rohit,

10003, 22, 21000, jacket, shoes; kuldeep, 10004, 39, 23000, jacket, shoes, jeans; }

Then in above dataset one pattern is obtained from the sequence of data that is name, Id, age, salary, convince, commodity member. So collection of this dataset in proper format is necessary for proper operation on the whole set.

Pre-Processing

As the dataset obtained from the above steps contain many unnecessary information which one need to be removed for making proper operation on those sets. This can be understood as let the name be the same as it is in the original set so to put this column in the original dataset is not necessary and it can be removed move from the above set of vectors, while if to hide information of the salary of the individual then one has to make from the original, therefore this kind of numeric data which need to be hide is perturbed by our method. Special attention should be paid for the text data as in our D dataset if commodities like items. is perturbed from the original then each row need one new combination from the older one this can be understand as the let the original data is bike, house then after perturbation it will be like car, house. So by doing this one can easily manipulate the whole information and no one get complete and correct information from the perturbed copy of the set.

RULE	SUPPORT	CONDITION	FILTER RULE
A1, B1 →C	9	MS > 10	
A1, B2 →C	11		A1, B2 →C
A1, B3 →C	12		A1, B3 →C
A2, B1 →C	9.5		
A2, B2 →C	7		
A2, B3 →C	13		A2, B3 →C

Table 4.2 Pre-processed dataset.

One important operation of this step is to separate those column which need to be perturbed in the numeric and text category as they need to be perturbed separately by different steps. So selection of the desired column from the whole dataset is done in this module of preprocessing.

Multi-attribute Super modularity:

In this step whole multi attributes are replace by its hierarchy value in the super modularity tree, while replacing it is required to balance the dataset utility and risk by making required changes.

Generate Rules

There are various pattern finding techniques in the dataset for finding the most frequent data. The most commonly used algorithm is aprior algorithm. Assume we have four items in the present dataset that is {jeans, Trouser, Shoes, jacket}In the first scan of our present dataset with their corresponding support. Let minimum support is 15:{jeans} → 20 {trouser} → 25 {shoes} → 30 {jacket} → 25

Then in second pass the set of items is find and there corresponding supports are:

{jeans, trouser } → 12 {jeans, shoes } → 15 {jeans, jacket} → 14

{trouser, jeans} → 12 { trouser, shoes } → 18 { trouser, jacket} → 19

{ shoes, trouser } → 18 { shoes, jeans} → 15 { shoes, jacket} → 22

{ jacket, jeans} → 14{ jacket, shoes } → 22 {jacket, trouser} → 19

So item set have minimum support are:

{jeans, shoes } → 15 { trouser, shoes } → 18 { trouser, jacket} → 19 { shoes, trouser } → 18 { shoes, jeans} → 15 { shoes, jacket} → 22 { jacket, shoes } → 22 {jacket, trouser} → 19

Then move for third scan of dataset for three item in item set:

{trouser, shoes, jeans } → 12 { trouser, jacket, shoes} → 10

{ shoes, trouser, jacket} → 16 {shoes, jeans, jacket} → 17

Separate Sensitive Rule

Now from the generated rule one can get bunch of rules then it is required to separate those rules from the collection into sensitive and non- sensitive rule set. Those rules which contain sensitive items are identified as the sensitive rules while those not containing are indirect rules. This can be understood as the Let A, B →C where A is set of sensitive item then this rule is sensitive rule, where B, C are non sensitive items. If D, B→ C is a rule and D is the non sensitive item set then this rule is not sensitive rule.

For deeper understanding consider an example let two attribute values set first is for direct items which is represent by the A = {female, black , foreign, …….etc }. In the similar fashion other items are considered as the non sensitive items set that is represent by B = { graduate, male, white,……..etc.}. C is the set of values whose value is binary in terms such as literate / illiterate, old / young, 50k> / 50k<, ……..etc.

So those rules which contain sensitive items in their set are consider as the sensitive rule for example [female, graduate] → 50k< this is same like A, B → C. In the similar fashion those rules which not contain the sensitive items set value are considered as non-sensitive rule for example [male, graduate] → 50k< this is same like D, B → C.

Now all sensitive rules which cross minimum support value is need to be perturb. This can be understand by below example.

Sensitive Pattern Hiding:

So in order to hide pattern, {X, Y), we can decrease its support to be lesser than user-provided minimum support transaction (MST). (1) Increase the support of X, the left hand side of the rule, but not support of X → Y.

(2) Decrease the support of the item set X →Y .For the second case, if we only decrease the support of Y, the RHS or right side of the rule will lessen the confidence very fast rather than simply reducing the support of X → Y.

We have to reduce the right side or the RHS item Y of the rule correspondingly. So for the rule Bread→ Milk can generate reduce the support of Y only. Now it needs to find that for how many transaction this need to be done. So calculation of that number is done by

((Rule_support – Minimum_ support) * Total_transaction)/100

Above formula specify the number of transaction where one can modify and overall support of that hiding pattern is lower than the minimum support.

Table 4.3 Number of session to hide sensitive dataset.

RULE	SUPPORT	FORMULA	NO. OF PERTURB SESSION
A1, B2 →C	11	((11-10)*1000)/100	10
A1, B3 →C	12	((12-10)*1000)/100	20
A2, B3 →C	13	((13-10)*1000)/100	30

Proposed Algorithm:

For this algorithm t is a transaction, T is a set of transactions, P is used for pattern, RHS (R) can be defined as Right Hand Side of the rule, LHS (R) can be defined as left hand side of the rule Here P is the pattern, S is the support of the rule, H is the set of hidden items.

Hiding Rules Algorithm:

Input: A source database D, A minimum support in Transaction (MST).

Output: The sanitized database D, where rules containing X on Left Hand Side (LHS) or Right Hand Side (RHS) will be hidden.

Steps of algorithm:

1. P[c] ← Aprior(D) // s = support
2. Loop I = For each P
3. If Intersect(P[I], H) and P[I] > MST
4. New_transaction ← Find_transaction(P[I], MST)
5. While (T is not empty OR count = New_treansaction)
6. If t←T have XUY rule then
7. Remove Y from this transaction
8. End While
9. EndIf
10. End Loop

In proposed algorithm input is original dataset (DS), MST threshold and output contain perturbed dataset (PDS). In whole algorithm frequent rules (FR) are generated then rules are filter by sensitive rule. Then in-order to suppress those discriminating rules (DR) find number of sessions to perturb and perturb those session where those item set is present.

V. EXPERIMENT AND RESULT

This section presents the experimental evaluation of the proposed perturbation and de-perturbation technique for privacy prevention. All algorithms and utility measures were implemented using the MATLAB tool. The tests were performed on an Intel Core i3 2.00GHz machine, with 4 GB RAM, and executing under Windows 10 Professional environment.

Evaluation Parameters

Missed Pattern: Representing the number of Sensitive patterns still present in dataset even after applying adopted privacy preserving technique.

Privacy Percentage: This specifies the percentage of the privacy provide by the adopting technique.

Support	Originality Percentage	
	Previous work	Proposed Work
13	69.0549	93.6895
14	69.0237	93.5859
15	69.0612	93.4823
16	69.0937	93.2751

Results

Table 5.1 Comparison of proposed and previous work on the basis of Originality percentage.

Table 5.2 Comparison of proposed and previous work on the basis of Risk values.

Support	Missed Patterns Percentage	
	Previous work	Proposed Work
14	100	0
15	100	0
16	100	0
17	100	0

From table 5.2 it is obtained that the risk value of the dataset is reduced after applying the proposed work. In other words previous work has reduced the risk value but to less extent.

From the above table it is obtained that proposed work has highly maintain the originality of the dataset after applying the perturbation algorithm.

Here by change in the sigma value originality of the previous work is move around 69% while proposed work has originality around 93%. Here pattern preservation has less affect on dataset while previous approach was having higher affect .

Support	Risk Value	
	Previous work	Proposed Work
13	-1.2938e+04	-6.2071e+04
14	-1.2765e+04	-1.4593e+04
15	-1.2816e+04	-6.6866e+04
16	-1.2781e+04	-7.1661e+04

Table 5.3 Comparison of proposed and previous work on the basis of Missed Patterns.

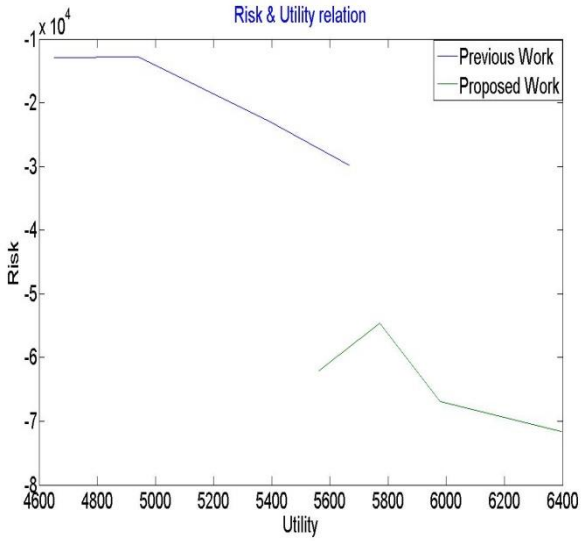
From above table it is obtained that proposed work has not preserve all sensitive patterns in the dataset.

Support	Utility Value	
	Previous work	Proposed Work
13	4.9523e+03	5.5622e+03
14	4.9416e+03	5.7708e+03
15	4.9883e+03	5.9794e+03
16	4.9667e+03	6.3965e+03

While previous work do not apply any approach for pattern preservation so no affect on those patterns are present after previous approach. Here all sensitive information is hidden in proposed work.

Table 5.4 Comparison of proposed and previous work on Utility Value basis.

From above table it is obtained that proposed work has increase the utility value of the dataset after applying the proposed work. As the previous work was having lower utility value.



From above graph it is obtained that proposed work has increased the utility value of the perturbed dataset as compared to the previous work. While one more evaluation is obtained that risk of the proposed outcome is quite low as compare to the previous work. In other words, previous work was having lower utility value.

VI. CONCLUSION

In this work, a set of algorithms and techniques were proposed to solve privacy-preserving data mining problems. The experiments showed that the proposed algorithms perform well on large databases. It work better as the Maximum lost pattern percentage is zero a certain value of support. Then this work shows that false patterns value is zero. Comparison with the other algorithm it is obtained that including the differential privacy and then directly hide the sensitive information. It is shown in the results that accuracy of the perturbed dataset is preserved for low support values as well. Here proposed work has resolve the multi party data distribution problem as well as different level trust party get different level of perturbed dataset copy.

VII. REFERENCES

[1]. Data Mining From Wikipedia, The free Encyclopaedia.

[2]. Jiawei Han, Micheline Kamber, Jian Pei "Data Mining Concepts and Techniques," Intelligent Database Systems Research Lab, School of Computing Science, Simon University, Canada.

[3]. Jaydip Sen "Privacy Preserving Data Mining Applications Challenges and Future Trends," Presentation to Innovation Lab, Tata Consultancy Services, Kolkata, 2nd ICCCT, MNNIT, Allahabad Sept 2011.

[4]. R Agrawal, R. Srikant "Privacy-Preserving Data Mining," Proceedings of the ACM SIGMOD Conference, 2000.

[5]. Toon Calders and Sicco Verwer "Three naive Bayes approaches for discrimination-free classification," July 2010.

[6]. Faisal Kamiran and Toon Calders "Data preprocessing techniques for classification without discrimination," Nov 2011.

[7]. Sara Hajian, Josep Domingo-Ferrer, Antoni Martinez-Balleste " Discrimination Prevention in Data Mining for intrusion and Crime detection," Conference on Computational Intelligence in Cyber Security (CICS), IEEE Symposium 2011.

[8]. F Kamiran, T Calders, M Pechenizkiy Discrimination Aware decision Tree Learning," Data Mining (ICDM), 2010 IEEE 10th International Conference 2010.

[9]. S Hajian, J Domingo-Ferrer "A Methodology for direct and Indirect discrimination Prevention in Data Mining," IEEE Transactions on knowledge and data engineering, 2013.

[10]. Charu C.Agarwal "A General Survey of Privacy-Preserving Data Mining Models and Algorithms ," IBM T. J. Watson Research Center Hawthorne, Philip S. Yu Chicago.

[11]. Cynthia Dwork and Frank McSherry in collaboration with Ilya Mironov and Kunal Talwar. " Privacy Preserving Data Mining".

[12]. Shuvro Mazumder " Privacy Preserving Data Mining," Department of Computer Science, University of Washington, Seattle.

[13]. Rakesh Agrawal, Ramakrishnan Srikant "Fast Algorithms for Mining Association Rules ," IBM Almaden Research Centre .

[14]. Aleksandra Korolova "Protecting Privacy When mining and sharing user data," a dissertation Submitted to stanford university for doctor of philosophy August 2012.

[15]. Songtao Guo "Analysis of and Techniques for Privacy Preserving Data mining ,"A dissertation submitted to The University of North Carolina at Charlotte for Doctor of Philosophy 2007.

[16]. Arie Friedman "Privacy Preserving Data Mining," Research Thesis submitted to the Technion | Israel Institute of Technology for Doctor of Philosophy 2011.

[17]. Bhupendra Kumar Pandya,Umesh Kumar Singh ,Keerti Dixit "A Robust Privacy Preservation by Combination of Additive and Multiplicative Data Perturbation for Privacy Preserving Data Mining,". Institute of Computer Science Vikram University, Ujjain, International Journal of Computer Applications(IJCA) , June 2015.

[18]. Arvind Batham, Mr.Srikant Lade, Mr. Deepak Patel "A Robust Data Preserving Technique by K-Anonymity and Hiding Association Rules,"Research paper , Rajiv Gandhi Technical University, Bhopal ,International Journal of Advanced Research in Computer Science and Software Engineering(IJARCSSE) January 2014.

[19]. Mohamed R. Fouad, Khaled Elbassioni, Member, IEEE, and Elisa Bertino, Fellow, IEEE "A Supermodularity-Based Differential Privacy Preserving Algorithm for Data Anonymization"

IEEE Transactions On Knowledge And Data Engineering July 2014

[20]. Vehuad Lindell "Privacy Preserving Data Mining," Department. of CSE Israel and Benny Pinkas STAR Lab, Santa Clara, CA.

[21]. Cecilina Parng, Presentation on Theme "Data Mining "Jan 2014

[22]. Markus Hegland " The Apriori Algorithm," WSPC/Lecture Notes series a Tutorial By CMA,Australian National University March 2005.

[23]. Ramakrishnan Srikant " Privacy Preserving Data Mining Challenges & Opportunities," .

[24]. Mrs. Bharati M. Ramageri " Data Mining Techniques And Applications," Lecturer Modern Institute of Technology and Research,Computer Application Deptt. , Pune, Indian Journal of Computer Science and Engineering (IJCSE) Vol. 1,2014.