Data Mining: Random Swapping based Data Perturbation Technique for Privacy Preserving in Data Mining

Ajmeera Kiran **Dr. D. Vasumathi**

ABSTRACT---Data mining is a process of collecting unknown data from different data sources and such data are very much useful for various decision-making Processes. Data mining process utilizes such sensitive information for analyzing purpose but, privacy preservation of such sensitive data is very much important in every data mining applications. For Example, in-patient Health records some of the sensitive attributes like PID, Age, and Disease Name should not be disclosed to the third party which will lead to privacy violation of the individuals. Hence, a new model should be designed to preserve the privacy of such Sensitive data before it make publicly available. In this paper, an accurate and efficient PPDM (Privacy Preserving Data Mining) technique is implemented in order to preserve the private information about individuals. In the Existing System, traditional Geometric data perturbation (Gaussian Noise Based) technique preserved the individual privacy with some information loss. In the Proposed Paper, an efficient and effective Random Swapping based data perturbation technique is proposed which is mainly focuses on preserving the sensitive attributes and also attaining accurate classification results with minimum Information loss. In Proposed Framework, the accuracy, error rates is compared with a Naïve Bayes classification algorithm and J48 decision tree Algorithm and results are analysed using Weka 3.8 tool. Proposed Random Swapping Based perturbation technique improved the Accuracy and reduced the error rates with minimum information loss with compared to the existing system.

Keywords—Data Mining, Privacy Preserving, Data Perturbation, Random Swapping, Naive Bayes Classification, J48 decision tree, Geometric Data Perturbation.

1. INTRODUCTION

In Early years, the data size may be very low and such data will be stored in small storage devices like magnetic tapes, hard discs, etc, but, in recent years there is a huge usage of data rapidly increased and utilization of such data also increased in many organizations like banking, business organizations, financial departments, medical insurance companies, social sciences, statistics etc [1] [2], for their professional growth. To analyze such voluminous data of different tools and techniques are necessary. Data mining plays an important role in processing such large volume of data [3] [4]. Data mining process extracts the useful information from heterogeneous sources. Data which is sharing between the users may contain private information about persons or some sensitive information [6] which cannot be disclosed. If such sensitive data is shared it may lead to the Privacy violation or Privacy breaching of individuals [7] [8]. To overcome such challenge a new technique is introduced for preserving privacy called Privacy

Preserving Data Mining (PPDM) [9]. Presently different privacy preserving data mining algorithms playing a crucial role in preserving sensitive individual data from the last few decades in the field of data mining [10]. Every result of privacy preserving data mining is directly influencing the Success of data mining Process. Different Existing methods have some limitations in preserving individual sensitive information [11] [12]. Different PPDM (Privacy Preserving Data Mining) techniques are classified as follows namely: Data modification based approach, Cryptographic based methods, data transformation based techniques, anonymization based methods, query Auditing based methods, randomization based approach etc [13] [14]. The main drawback of existing methods is that the perturbed data (Modified data) is mainly influenced by the Noise that is being added to the Original data, by this way sensitive data are transformed into a new form [15].

In the Existing Method, novel traditional Geometric data perturbation methods are preserved the individual sensitive data privacy with some limited information loss [16]. In this proposed Paper, an effective and efficient Random Swapping based data perturbation technique is proposed that is mainly focused on privacy preservation of preserving individual sensitive attributes and also obtaining data good Classification results [17]. The proposed framework can perturb the sensitive attributes of more than one column at a time. In the Proposed system, accuracy, error rates are compared with the help two popular algorithms like Naïve Bayes classification algorithm and J48 decision tree Algorithm and results are analyzed using popular data mining tool i.e., Weka 3.8 tool. Although, the Proposed Random Swapping Based perturbation method improved the Accuracy as well as reduced the error rates with minimum information loss with compared to existing methods and also it is possible to generate original data values from the perturbed data. In this Paper, comparison results are represented with the help of tables and charts.

2. RELATED WORKS

Data mining is a continuous process of Extracting, Storing, analyzing voluminous data which are generated

Revised Manuscript Received on June 10, 2019.

Ajmeera Kiran, JNTUH College of Engineering Hyderabad, T.S,INDIA. (E-mail: kiranphd.jntuh@gmail.com)
Dr. D Vasumathi, JNTUH College of Engineering Hyderabad, T.S,INDIA. (E-mail: rochan44@gmail.com)

from heterogeneous data sources. Nowadays, most of the data is publicly available, data miner can get access to whole information such as individual private data also. Such sensitive data is successfully achieved by data miner which may lead to privacy attacks. This issue is motivated to preserve the private data before making it public. To accomplish this task in data mining privacy preserving data mining came into existence [18].

The initial investigation about data swapping is founded by Resis (1980).In this paper, authors are primarily focussed on interchanging of Sensitive information in a record. The Swapping Method protects the univariate Distribution of Sensitive variables. Following are the different author‟s perspectives regarding Swapping technique.

Verykios, K. Bertino, I.N.Fovino, L.P.Provenza, Y.Saygin and Theodoridis (2004), Presented a Novel technique on privacy-preserving data mining methods, which is based on five Measurements like Data modification, Data Distortion, Data or Rule Hiding, Data mining algorithms, and also privacy preservation for some centralized and distributed data.

W. Du and Zhan (2002) illustrated the Methods to building Decision trees by the utilizing secure scalar, Secure Sum, secure Union for protecting original data. The main drawbacks of this method are it performs multiple database scans which degrade the performance of the system.

R. Aggarwal and R.Srikanth (2000) defined a new perturbation method using random data perturbation technique for protecting user‟s private data and constructed decision trees. This method generated the less accuracy with the perturbed dataset.

HillolKargupta, SouptikDatta, Qi Wang, and K.Siva Kumar,(2003), has Exhibited privacy preserving data mining of user data with random noise generated. They have defined different random based methods for generating random matrices.

TanveerJahan, G.Narasimha, C.V.GururRao (2012), has elaborated clustering on distorted Data for Privacy-preserving Data mining. They presented Sparsified singular value decomposition (SSVD).it is better than the existing SVD system.

Lindell and B.Pinkas (2002), has Focused on Different cryptographic-based privacy-preserving data mining techniques. It is mainly used for the smaller databases. This paper achieved high security by using the cryptographic methods.

Shweta T, ShanshankKhanna, T.Sugandha, and Ankitha, (2014), have illustrated an efficient Hybrid C-Tree technique for providing security to the sensitive data in privacy preserving data mining. They used Special Characters and ASCII codes for the encryption process. It is very much useful for protecting the medical type datasets.

A.Srivastava and G.Srivastav (2015), presented a technique by protecting sensitive data medical type dataset in privacy preserving data mining. Here K-Anonymity techniques are utilized for protecting individual private data for E-Health records and achieved best accuracy results.

Yifeng XU and Jie Liu (2010) have exhibited a random response method and geometric data perturbation method. This method can protect only numerical type data sets.it has Better privacy protection than existing techniques.

Jie Liu and Yifeng XU (2010) have projected geometric data perturbation with random response technique. This paper only concentrates on the Numerical typed dataset.it is not applicable for categorical typed datasets. It is maintained good accuracy results with compared to the existing data perturbation techniques.

Chhinkaniwala H and Garg S (2011) have defined different techniques and Challenging issues with, related to privacy-preserving data mining. In this paper, the authors focused on the taxonomy of Different privacy-preserving data mining techniques and also discussed various limitations of Existing Privacy-preserving data mining methods.

M.Reza and SomayyehSeifi (2011) has illustrated a new classification technique for evaluating privacy preserving data mining by using data perturbation methods.in this paper, two methods are used namely K-Anonymization and Data Perturbation techniques. The Authors maintained a fixed stability between data utility and data privacy.

H. Chhinkaniwala and S. Garg, (2013), presented an effective Multiplicative perturbation technique based on the tuple value. They utilized K-Means Clustering algorithm for obtaining better accuracy and calculated recall, precision, and CMM.

Mr.Kiran Patel (2013) has defined a novel approach for classification of the data streams for Privacy-preserving data mining.in included two major steps of data mining like Preprocessing of data and also data Stream data mining. They introduced a Hoeffding technique with minimum information loss.

G. Manikandan, et al., (2013), have proposed a data transformation technique using Normalization for data achieved good accuracy and also enhanced the performance of data mining algorithms.

Tarique Ahmad et al., (2014), have defined a min max normalization based approach for preserving the privacy of the sensitive attributes in a dataset. Original dataset values are modified using min max normalization before data mining begins. The Experimental are proven that proposed k-means algorithm is preserved the both accuracy and privacy.

Patel Brijal et al.,(2015),has illustrated the concept of clustering algorithm for preserving sensitive attributes in a dataset. Proposed method succeeded in protecting the sensitive and critical information in a dataset and achieved good data mining results with minimum information loss.

Anjana Patel, et al., (2016), have demonstrated the concept of geometric data perturbation on modified and randomized data using k-means clustering algorithm. Proposed method is used to check the correctness and achieved good accuracy results. The experimental results proven that proposed method will reconstruct the original data values without ant information loss.

3. PROBLEM DEFINITION

Existing traditional Privacy preserving data mining algorithms has following limitations

- High Computational Complexity-Traditional privacy preserving techniques are faced computational Complexities while handling huge volume of data.
- Poor Scalable - Poor scalability and inefficiency problems occur due to the existing traditional privacy preserving techniques for large datasets.
- Less reliability and Poor Security- By using existing privacy preserving techniques will lead to the less reliability and give less security to the sensitive data.
- Lack of Data Reconstruction- By using Traditional privacy preserving data mining methods data reconstruction is not possible.

All of these Drawbacks are motivated us to carry out the proposed technique which will solve all those issues.

4. PROPOSED RANDOM SWAPPING METHOD

In recent years ,there are many works has been defined to preserve the individual privacy and also to protect the sensitive information about the individuals like data perturbation, data transformation and anonymization etc in the field of data mining by modifying the original dataset into perturbed dataset.

The primary goal of the proposed research work is to transform the Original dataset values D into a Modified form of Dataset values D‟ that majorly fulfils the privacy requirements with minimum information loss with compared to existing techniques. Proposed system mainly focuses on the Random data Swapping based method to modify the original dataset values into modified form of dataset. The modified form of data values are very difficult to predict by the attacker while publishing or sharing the data [19].In proposed system, Data modification is performed on the selective attributes in a dataset by using by any of the following technique.

- Data Swapping or Data Sampling
- Data Blocking
- Data Masking

Data perturbation using Masking Method

Data Masking is another form of data perturbation technique. In this method original sensitive attributes are replaced with symbols like „*‟ or „+‟.It is very similar to data blocking method but in this method symbols are used for masking of attributes.

Data Perturbation with blocking

Blocking is the one of the basic method of perturbing original data by replacing the sensitive information by substituting the attributes with „?‟ or „$‟ or with any meaningless characters.

Data perturbation with Data Swapping or Sampling method

Data swapping is a well-known and popular data perturbation technique. Data swapping can be defined as, the process of exchange sensitive information between two individuals by preserving the sensitive information about the individuals [20]. In this method Original Individual records are replaced with new values so that original dataset is completely replaced to preserving the sensitive attributes in a dataset. By using data swapping method, data mining process achieved good accuracy with compared to existing noise addition methods without breaching the privacy of the individuals. The main advantage by using data swapping technique is that it can be applied along with other privacy preserving data mining techniques like k-anonymity and randomization [21].

In this Paper, an efficient and effective random Swapping based data perturbation technique is proposed that mainly focuses on preserving sensitive attributes and also attaining data Classification with minimum Information loss. Proposed Random Swapping Based perturbation technique improved the Accuracy as well as reduced the error rates with minimum information loss with compared to the existing system and also it is possible to generate original data values from the perturbed data.

Various data mining techniques in data mining for privacy preserving data mining will relay and evaluated on two factors namely Data Utility and Data Privacy. The Primary goal of each data modification technique is to ensure maximum level of privacy preservation of sensitive data along with high data utility factor.

4.1 Flow Chart Of Applied Framework

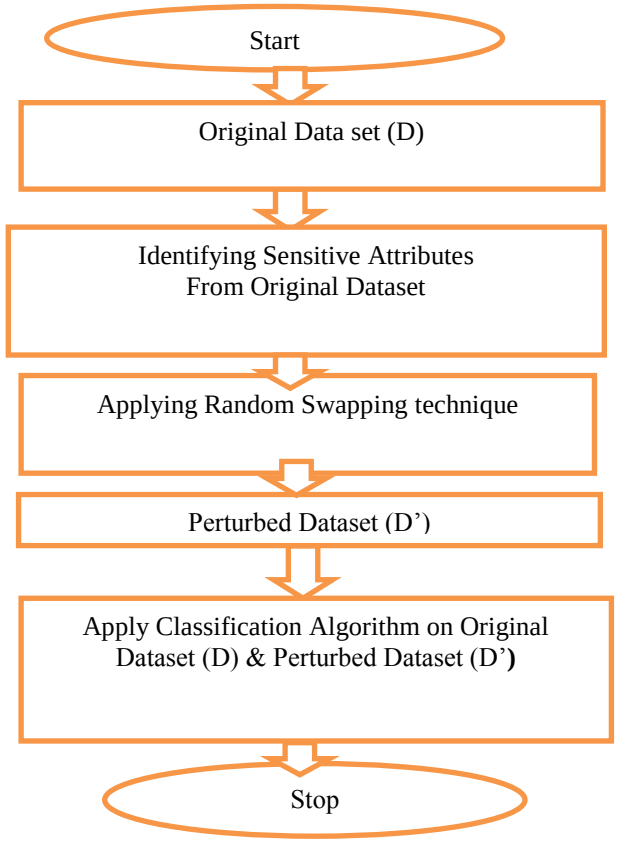

Figure 1: Work flow Diagram of Proposed System

Various data mining techniques in data mining for privacy preserving data mining will relay and evaluated on two factors namely Data Utility and Data Privacy. The Primary goal of each data modification technique is to ensure maximum level of privacy preservation of sensitive data along with high data utility factor.

Data Privacy: Privacy of the data can be measured in terms of complexity that an attacker can have to obtain original data values from the modified dataset. The Proposed framework utilizes data transformation technique to modify the original sensitive information for preserving the attributes. Proposed technique gives higher level of data privacy to the sensitive information.

Data Utility: Data Utility can be defined as the level of sensitive data is preserved after performing data mining process.

4.2 Proposed Random Swapping Algorithm

Procedure: Perturbation of Data Using Random Swapping Method.

Input: Original Data Set D, Sensitive Attributes

Intermediate Output: Modified (Perturbed) data Set D'.

Output: Classification Results R and R' of Datasets D and D'.

Step 1: Given input dataset D with tuple size n and extract Sensitive Attributes [S]

Step 2: Apply Random Swapping method on Sensitive Attributes [S].

(In the Random Swapping Method Attributes are Swapped randomly using Java Mappers. Using Proposed technique the Data reconstruction possible)

Step 3: Create perturbed dataset D' by replacing Sensitive attributes in original dataset D.

Step 4: Apply Naïve Bayes classification algorithm on original dataset D having sensitive attributes S.

Step 5: Apply Naïve Bayes classification algorithm on Perturbed dataset D' having Perturbed sensitive attributes S.

Step 6: Compare and analyze the results of Steps 4 and step 5 for analyzing accuracy of proposed Method with Naïve Bayes Classification algorithm

Step 7: Apply J48 Decision tree algorithm on original dataset D having sensitive attributes S

Step 8: Apply J48 Decision tree algorithm on Perturbed dataset D' having perturbed sensitive attributes S.

Step 9: Compare and analyze the results of Steps 7 and step 8 for analyzing accuracy of proposed Method with J48 Decision tree algorithm.

4.3 Proposed Random Swapping Architecture

The Primary objective of the proposed technique is preserve the sensitive attributes in a dataset. Following figure 1.Illustrates the main functionality of the proposed method. Initially Proposed method collects the information from the UCI machine Repository. In Next step user identifies the Sensitive attributes present in the whole Dataset and each every Sensitive attributes are distorted using the Random Swapping Method. Data which is distorted is very difficult for the attacker to identify the original dataset values form the distorted one. After Generating Perturbed dataset, both original dataset and perturbed dataset is analyzed by Data mining process.

Finally, proposed method maintains both accuracy and Privacy with minimum information loss. Reconstruction of original data is possible from the perturbed dataset using the proposed method.

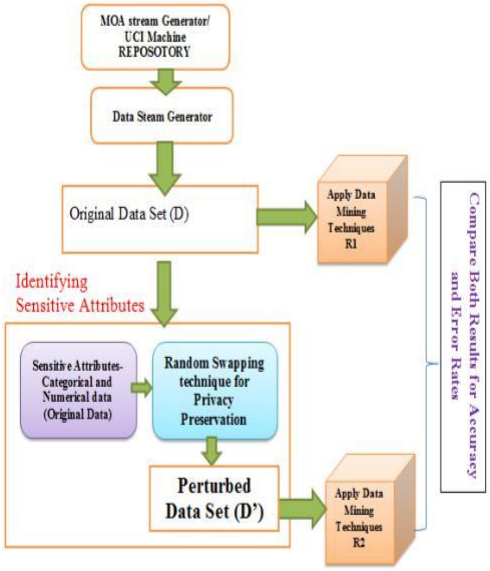

Figure 2: System Architecture of Proposed System

5. EXPERIMENTAL RESULTS

A. Data Sets

Table 1: Description of Adult Dataset

Datasets	Description
Adult Dataset	Total Instances: 48,842 Attributes: 6

In general many dataset are utilized for analyzing the data mining process like Cancer dataset, Bank dataset, Cover type dataset, Heart dataset, Adult dataset. In proposed framework Adult dataset is used for analyzing.

Table 2: Description of Attributes in Adult Dataset

Attribute	Data type
Age	Numeric
Fnlwgt	Numeric
Work Class	Text
Education	Text
Marital Status	Text
Education Num	Numeric
Capital gain	Numeric
Occupation	Text
Race	Text
Relationship	Text
Sex	Text
Native Country	Text
Hours Per Week	Numeric
Less Greater	Numeric

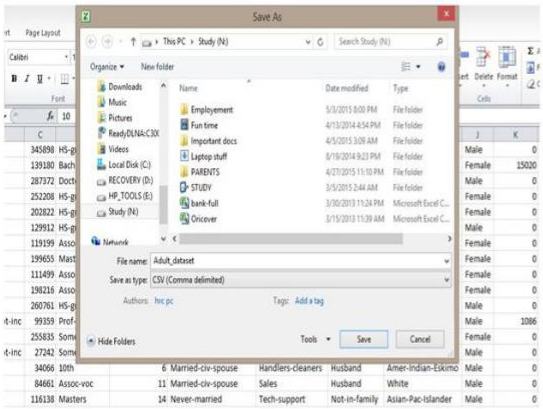

Figure 3: Collecting Data set From UCI library

The Above figure 3 illustrates the process of collecting data from the UCI machine repository [22]. UCI machine library contains the all kinds of datasets.

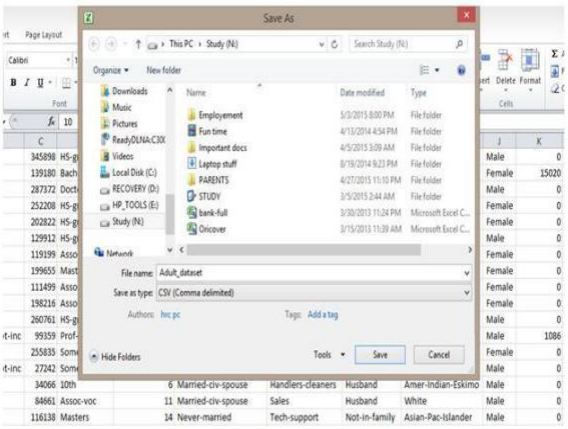

Figure 4: conversion of .XLSX file into .CSV

The Figure 4: Illustrates the Conversion process of the file from .XSLX file format to .CSV file format.

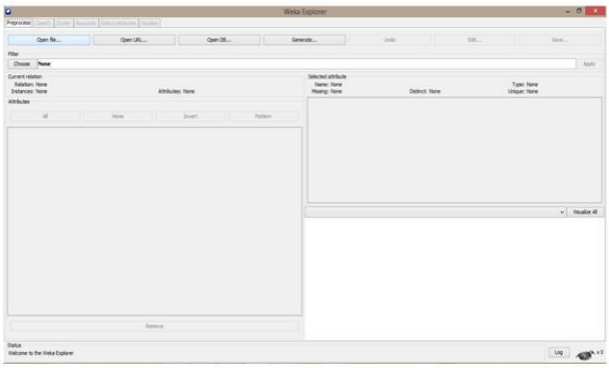

Figure 5:Main window of Weka 3.8

The Figure 5 demonstrates the usage of Data mining Tool Weka 3.8 user interface to perform various data mining applications [23].Weka is a well-known and very popular data mining tool which consists of various predefined data mining algorithms.

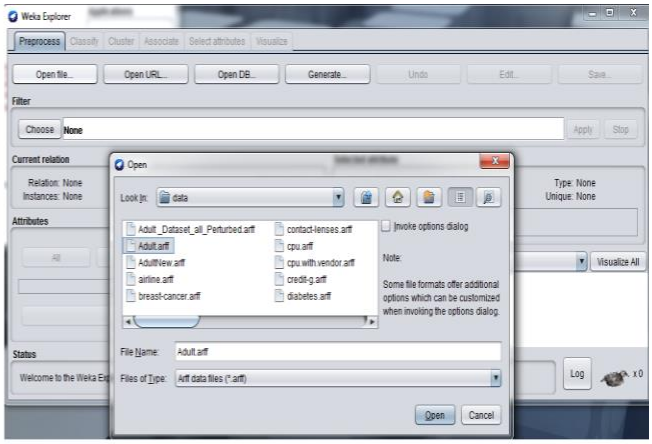

Figure 6: Loading of Data set into Weka 3.8

The above figure 6 defines the process of loading a simple Dataset into the Weka 3.8 tool for analyzing purpose.

B. Pre-Processing of Data

Pre-processing is the primary step in data mining. The Primary Objective to perform Pre-processing technique is to Discover and eliminate non-relevant data from the dataset which includes null values, delimiters, unspecified attribute values etc. By performing pre-processing the performance and efficiency of the data mining algorithm will improve. After performing data pre-processing, perturbed dataset is generated from original dataset and both datasets are compared using data mining tool.

C. Experimental Results

In the implementation Process, Original dataset is modified using data transformation technique such as Min-Max Normalization technique to attain the perturbed dataset. Both the datasets (original and Perturbed) are compared with the two well-known algorithms Naïve Bayes and J48 in Weka 3.8 data mining Tool.

Weka 3.8 is a well-known and very popular data mining tool which consists of various predefined data mining algorithms. Proposed algorithm utilized Naive Bayes classification algorithm and J48 decision tree algorithm from the set predefined data mining algorithms present in Weka tool.

The Following tables 3 & 4 demonstrates the classification results of both algorithms Naive Bayes classifier algorithm and J48 Decision tree algorithm for Single column as well as More than two columns at time. Both the tables illustrate the comparison of modified and Original values and their efficiencies. In the Existing framework, traditional Geometric data perturbation techniques preserved the individual privacy with some information loss. And also Existing System, Data reconstruction is not possible but using proposed scheme data reconstruction is made easy. Proposed framework is evaluated using random swapping based technique on adult dataset and has more advantages with compared to the existing methods. Proposed method works for Categorical type also and gives better accuracy results and also Error

rates are reduced by using the proposed framework with compared to existing noise based System. Experimental results are presented in the form of graphs and table for better understanding of proposed algorithm.

5.1. Performance Analysis Of The Proposed Approach

Dataset 1: Adult dataset

	Single Column (Numerical)				Single Column (String)			
	AGE				GENDER			
	NB		J48		NB		J48	
	Original	Perturbed	Original	Perturbed	Original	Perturbed	Original	Perturbed
Correctly Classified Instances	83.25%	83.48%	86.07%	86.06%	83.25%	83.35%	86.07%	86.00%
Correctly Classified Instances	16.74%	16.51%	13.92%	13.93%	16.74%	16.64%	13.92%	13.99%
Kappa Statistics	0.49	0.50	0.58	0.58	0.49	0.49	0.58	0.58
Mean Absolute Error	0.1746	0.1727	0.1956	0.1977	0.1746	0.174	0.1956	0.1937
Root Mean Squared Error	0.3741	0.3703	0.3201	0.3277	0.3741	0.3731	0.3201	0.3211
Relative Absolute Error	0.4795	0.4744	0.5371	0.5431	0.4795	0.4778	0.5371	0.5321
Root Relative Squared Error	0.8768	0.8679	0.7502	0.7564	0.8768	0.8745	0.7502	0.7525
Time Consumed(sec)	0.22	0.16	6.46	6.21	0.22	0.18	6.46	6.54

Table 3: .Comparison of Naive Bayes and J48 Algorithms for Single Column

Table 3 presents the result of naïve Bayes classification algorithm along with J48 Decision tree algorithm on both original and perturbed datasets for Single Column at a time such as age and Gender attributes along with different privacy measurement values. Such as Kappa Statistic's, Root means Squared Error, Mean Absolute Error ,relative absolute error, root relative squared error and Time taken to complete the Classification result.

NB	Age		Gender	
	Original	Perturbed	Original	Perturbed
Correctly Classified Instances	83.25%	83.48%	83.25%	83.35%
Correctly Classified Instances	16.74%	16.51%	16.74%	16.64%

Table 3.1.Accuracy Comparison of Naive Bayes Values for Age and Gender Attribute

From the Table 3, The above Table 3.1, the Classification accuracy of proposed algorithm using Naive Bayes technique on Original Dataset for age attribute was 83.25 % for Correctly Classified Instances and incorrectly Classified was 16.74 %. At the same time, the accuracy of Naive Bayes algorithm on Perturbed Dataset was 83.48% for Correctly Classified Instances and incorrectly Classified was 16.51%. And for Gender Attribute on Original Dataset was 83.25% for Correctly Classified Instances and incorrectly classified was 16.74 %.At the same time, the accuracy of Naive Bayes Algorithm on Perturbed Dataset was 83.35% for Correctly Classified Instances and incorrectly Classified was 16.64%. So Privacy preservation of Original dataset is well-preserved with Negligible Information loss.

J48	Age		Gender	
	Original	*Perturbed*	*Original*	*Perturbed*
Correctly Classified Instances	86.07%	86.06%	86.07%	86.00%
Correctly Classified Instances	13.92%	13.93%	13.92%	13.99%

Table 3.2. Accuracy Comparison of J48 Values for Age and Gender Attribute

From the Table 3, The above Table 3.1, the Classification accuracy of proposed algorithm using J48 technique on Original Dataset for age attribute was 86.07 % for Correctly Classified Instances and incorrectly Classified was 13.92 %. At the same time, the accuracy of J48 algorithm on Perturbed Dataset was 86.06% for Correctly Classified Instances and incorrectly Classified was 13.93%. And for Gender Attribute on Original Dataset was 86.07% for Correctly Classified Instances and incorrectly classified was 13.92 %. At the same time, the accuracy of Naive Bayes algorithm on Perturbed Dataset was 86.00% for Correctly Classified Instances

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

And Incorrectly classified instances was 13.99%.so privacy preservation of Original dataset is well preserved with negligible information loss.

	TWO COLUMNS							
	AGE GENDER				EDUCATION GENDER			
	NB		J48		NB		J48	
	Original	Perturbed	Original	Perturbed	Original	Perturbed	Original	Perturbed
Correctly Classified Instances	83.25%	83.13%	86.07%	86.01%	83.25%	83.33%	86.07%	85.91%
Correctly Classified Instances	16.74%	16.86%	13.92%	13.98%	16.74%	16.66%	13.92%	14.08%
Kappa Statistics	0.49	0.48	0.58	0.58	0.49	0.49	0.58	0.58
Mean Absolute Error	0.1746	0.176	0.1956	0.1972	0.1746	0.174	0.1956	0.1934
Root Mean Squared Error	0.3741	0.3758	0.3201	0.3218	0.3741	0.3731	0.3201	0.3211
Relative Absolute Error	0.4795	0.4833	0.5371	0.5416	0.4795	0.4777	0.5371	0.5311
Root Relative Squared Error	0.8768	0.8809	0.7502	0.7541	0.8768	0.8743	0.7502	0.7525
Time Consumed(sec)	0.22	0.23	6.46	6.32	0.22	0.22	6.46	6.46

Table 4: Accuracy Comparison of Naive Bayes and J48 Algorithms for Two Columns at a Time

Table 4 presents the result of naïve Bayes classification algorithm along with J48 Decision tree algorithm on both original and perturbed datasets for Single Column at a time such as age and Gender attributes along with different privacy measurement values. Such as Kappa Statistic's, Root means Squared Error, Mean Absolute Error ,relative absolute error, root relative squared error and Time taken to complete the Classification result.

NB	AGE, GENDER		EDUCATION ,GENDER	
	Original	*Perturbed*	*Original*	*Perturbed*
Correctly Classified Instances	83.25%	83.13%	83.25%	83.33%
Correctly Classified Instances	16.74%	16.86%	16.74%	16.66%

Table 4.1.Comparisoion of Accuracy with Naive Bayes Algorithm for two Columns

From the Table 4, The above Table 4.1, the Classification accuracy of proposed algorithm using Naive Bayes technique on Original Dataset for age attribute was 83.25 % for Correctly Classified Instances and incorrectly Classified was 16.74 %. At the same time, the accuracy of Naive Bayes algorithm on Perturbed Dataset was 83.13% for Correctly Classified Instances and incorrectly Classified was 16.86%. And for Gender Attribute on Original Dataset was 83.25% for Correctly Classified Instances and incorrectly classified was 16.74 %.At the same time, the accuracy of Naive Bayes algorithm on Perturbed Dataset was 83.35% for correctly Classified Instances and incorrectly classified was 16.66%. So Privacy preservation of Original dataset is well-preserved with Negligible Information loss.

J48	AGE, GENDER		EDUCATION ,GENDER	
	Original	*Perturbed*	*Original*	*Perturbed*
Correctly Classified Instances	86.07%	86.01%	86.07%	85.91%
Correctly Classified Instances	13.92%	13.98%	13.92%	14.08%

Table 4.2. Accuracy Comparison of J48 Values for Age and Gender Attribute at a Time

From the Table 4, The above Table 4.2, the Classification accuracy of proposed algorithm using J48 technique on Original Dataset for age attribute was 86.07 % for Correctly Classified Instances and incorrectly Classified was 13.92 %. At the same time, the accuracy of J48 algorithm on Perturbed Dataset was 86.01% for Correctly Classified Instances and incorrectly Classified was 13.98%. And for Gender Attribute on Original Dataset was 86.07% for Correctly Classified Instances and incorrectly classified was 13.92 %. At the same time, the accuracy of Naive Bayes algorithm on Perturbed Dataset was 85.91% for Correctly Classified Instances and incorrectly Classified was 14.08%. So Privacy preservation of Original dataset is well-preserved with Negligible Information loss.

	Three Columns				*Three columns*			
	Education Gender Hours Per Week				*Gender, Marital Status, Hours_per_Week*			
	NB				*J48*			
	Original	Perturbed	Original	Perturbed	Original	Perturbed	Original	Perturbed
Correctly Classified Instances	83.25%	83.14%	86.07%	85.96%	83.25%	83.15%	86.07%	85.96%
Incorrectly Classified Instances	16.74%	16.58%	13.92%	14.03%	16.74%	16.85%	13.92%	14.03%
Kappa Statistics	0.49	0.48	0.58	0.58	0.49	0.48	0.58	0.58
Mean Absolute Error	0.1746	0.175	0.1956	0.2003	0.1746	0.175	0.1956	0.2003
Root Mean Squared Error	0.3741	0.3753	0.3201	0.325	0.3741	0.3753	0.3201	0.325
Relative Absolute Error	0.4795	0.4805	0.5371	0.5499	0.4795	0.4805	0.5371	0.5499
Root Relative Squared Error	0.8768	0.8794	0.7502	0.7615	0.8768	0.8794	0.7502	0.7615
Time Consumed(sec)	0.22	0.16	6.46	6.29	0.22	0.22	6.46	6.26

Table 5: .Comparison of Naive Bayes and J48 Algorithms for more than Single Column (Three Columns)

Table 5 presents the result of naïve Bayes classification algorithm along with J48 Decision tree algorithm on both original and perturbed datasets for Single Column at a time such as age and Gender attributes along with different privacy measurement values. Such as Kappa Statistic's, Root means Squared Error, Mean Absolute Error ,relative absolute error, root relative squared error and Time taken to complete the Classification result.

NB	**Education , Gender, Hours Per Week**		**Gender, Marital Status, Hours_per_Week**	
	Original	***Perturbed***	***Original***	***Perturbed***
Correctly Classified Instances	83.25%	83.14%	83.25%	83.15%
Correctly Classified Instances	16.74%	16.58%	16.74%	16.85%

Table 5.1. Accuracy Comparison of Naive Bayes Values with Three Columns at time

From the Table 5, The above Table 5.1, the Classification accuracy of proposed algorithm using Naive Bayes technique on Original Dataset for age attribute was 83.25 % for Correctly Classified Instances and incorrectly Classified was 16.74 %. At the same time, the accuracy of Naive Bayes algorithm on Perturbed Dataset was 83.14% for Correctly Classified Instances and incorrectly Classified was 16.58%. And for Gender Attribute on Original Dataset was 83.25% for Correctly Classified Instances and incorrectly classified was 16.74%.

At the same time, the accuracy of Naive Bayes algorithm on Perturbed Dataset was 83.15% for Correctly Classified Instances and incorrectly Classified was 16.85%. So Privacy preservation of Original dataset is well-preserved with Negligible Information loss.

J48	**Education , Gender, Hours Per Week**		**Gender , Marital Status, Hours_per_Week**	
	Original	***Perturbed***	***Original***	***Perturbed***
Correctly Classified Instances	86.07%	85.96%	86.07%	85.96%
Correctly Classified Instances	13.92%	14.03%	13.92%	14.03%

Table 5.2. Accuracy Comparison of J48 Values with Three Columns at time

From the Table 5, The above Table 5.2, the Classification accuracy of proposed algorithm using J48 technique on Original Dataset for age attribute was 86.07 % for Correctly Classified Instances and incorrectly Classified was 13.92 %. At the same time, the accuracy of J48 algorithm on Perturbed Dataset was85.96% for Correctly Classified Instances and incorrectly Classified was 14.03%. So Privacy preservation of Original dataset is well-preserved with Negligible Information loss.

From table 6, the table illustrates the comparison of Naïve Bayes Algorithm for different Columns like Single Column, Two Column, and Three Columns. Proposed method Provides privacy for categorical data sets with minimum information loss with compared to existing method Gaussian noise techniques.

Adult Data Set	Existing Gaussian Noise Based Method For NB	Proposed Random Swapping Based Method for NB
Single Column	**83.25%**	**83.48%**
Two Columns	**82.86%**	**83.13%**
Three Columns	**82.07%**	**83.14%**

Table 6.Accuracy Comparison of Existing and Proposed with NB Algorithm

From the Table 7, Comparison of the accuracy with NB Algorithm on Adult Dataset for Single Column in Existing System is 83.25% and for the Proposed method is 8348%.as well as for the Two Columns the accuracy is 82.86% in existing Method and for the Proposed method is 83.13%.and also for Three Columns the Accuracy is 82.07% in existing Method and for the Proposed method is 83.14%.So Privacy of the Original data set is preserved with minimum information loss.

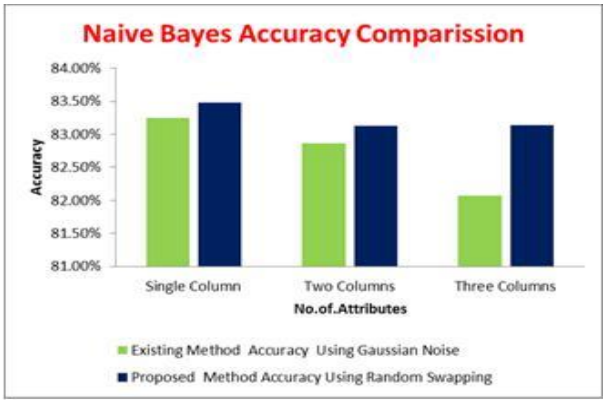

Figure 7: Accuracy Comparison of Existing and Proposed with NB Algorithm

From Figure 7, the graph shows that comparison NB Algorithm for different Columns like Single Column, Two Column, and Three Columns. Proposed method Provides privacy for categorical data sets with minimum information loss with compared to existing method Gaussian noise techniques.

Adult Data Set	Existing Gaussian Noise Based Method For J48	Proposed Random Swapping Based Method for J48
Single Column	**85.90%**	**86.06%**
Two Columns	**85.75%**	**86.01%**
Three Columns	**85.64%**	**85.96%**

Table 7. Accuracy Comparison of Existing and Proposed with J48 Algorithm

From the Table 7, Comparison of the accuracy with J48 on Adult Dataset for Single Column is 85.90% in existing and for the Proposed method is 86.06%.as well as for the Two Columns the accuracy is85.75% in existing Method and for the Proposed method is 86.01%.and also for Three Columns the Accuracy is 85.64% in existing Method and for the Proposed method is 85.96%.So Privacy of the Original data set is preserved with minimum information loss.

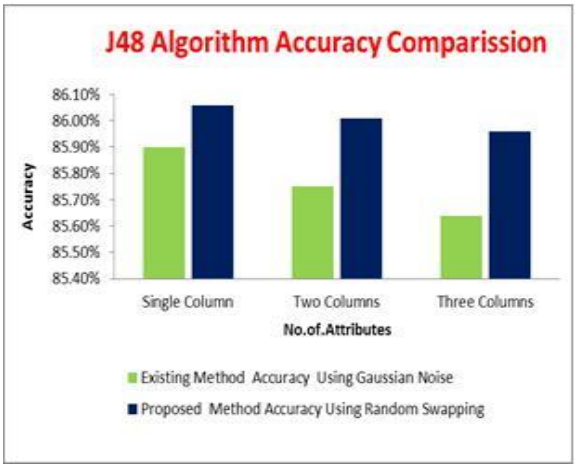

Figure 8: Accuracy Comparison of Existing and Proposed with J48 Algorithm

From Figure 8, the graph shows that comparison J48 Algorithm for different Columns like Single Column, Two Column, and Three Columns. Proposed method Provides privacy for categorical data sets with minimum information loss with compared to existing method Gaussian noise techniques.

5.2. Comparison Analysis Of Proposed Technique

The proposed framework uses two popular algorithms Naive Bayes and J48 for estimation of correct classification values among perturbed data set and original data set. Accuracy result of the perturbed data set are less than the accuracy result of the original data set, but it is possible to achieve privacy preservation for sensitive attributes with minimum information loss and loss can be minimized further.

5.2.1. Accuracy Comparison With Existing Method

Adult Data Set	Techniques Applied	Original Data Accuracy	Accuracy	
			Existing Method	Proposed Method
Age	Naive Bayes	**83.25%**	**83.08%**	**83.48%**
	J48	**86.07%**	**85.90%**	**86.07%**
Gender	Naive Bayes	**83.25%**	**83.26%**	**83.35%**
	J48	**86.07%**	**85.80%**	**86.00%**

Age ,Gender	Naive Bayes	**83.25%**	**82.98%**	**83.13%**
	J48	**86.07%**	**85.75%**	**86.01%**
Education Gender	Naive Bayes	**83.25%**	**82.33%**	**83.33%**
	J48	**86.07%**	**85.77%**	**85.91%**
Education, Gender, Hours Per Week	Naive Bayes	**83.25%**	**81.93%**	**83.14%**
	J48	**86.07%**	**85.64%**	**85.96%**
Marital Status , Gender , Hours Per Week	Naïve Bayes	**83.25%**	**80.60%**	**83.15%**
	J48	**86.07%**	**85.91%**	**85.96%**

Table 8: Comparison of Accuracy of Existing and Proposed with NB Algorithm and J48 Algorithm

Table 8 illustrates the Comparison values of Accuracy with Two well-known algorithms such as NB Classification algorithm and J48 Decision tree Algorithm by perturbing Different Columns at a Time. The Accuracy of the Original dataset without modifying the value for single Column (Age) was 83.25%. After perturbation the Accuracy for single Column (Age) with NB Algorithm was 83.08% in Existing Method for 83.48% for the proposed random Swapping Method. The Accuracy of the Original dataset without modifying the value for single Column (Gender) was 83.25%. After perturbation the Accuracy for single Column (Gender) was 83.26% in Existing Method and for Proposed random Swapping Method is 83.35%. The Accuracy of the Original dataset without modifying the value for Two Columns Column (Age & gender) was 83.25%. After perturbation the Accuracy for single Column (Age & gender) with NB Algorithm was 82.98% in Existing Method for 83.13% for proposed random Swapping Method. The Accuracy of the Original dataset without modifying the value for Two Column at a time (Education &Gender) was 83.25%. After perturbation the Accuracy for Two Columns at a time (Education &Gender) was 82.33% in Existing Method and for Proposed random Swapping Method is 83.33%. The Accuracy of the Original dataset without modifying the value for Three Columns at a time (Education, Gender, Hours per Week) was 83.25%. After perturbation, the Accuracy for Three Columns At a time (Education, Gender, Hours per Week) with NB Algorithm was 81.93% in Existing Method for 83.14% for proposed random Swapping Method. The Accuracy of the Original dataset without modifying the value for Three Columns at a time (Gender Marital Status Hours_per_Week) was 83.25%. After perturbation, the Accuracy for Three Columns At a time (Gender Marital Status Hours_per_Week per Week) with NB Algorithm was 80.60% in Existing Method for 83.15% for proposed random Swapping Method. So Privacy of the Original dataset by using Naïve Bayes Algorithms is preserved with minimum information loss.

At the Same time, Comparison values of Accuracy with J48 Decision tree Algorithm by perturbing Different Columns at a Time. The Accuracy of the Original dataset without modifying the value for single Column (Age) was 86.07%. After perturbation the Accuracy for single Column (Age) with J48 Algorithm was 80.90% in Existing Method for 86.07% for the proposed random Swapping Method.

The Accuracy of the Original dataset without modifying the value for single Column (Gender) was 86.07%. After perturbation the Accuracy for single Column (Gender) was 85.80% in Existing Method and for Proposed random Swapping Method is 86.00%.

The Accuracy of the Original dataset without modifying the value for Two Columns Column (Age & gender) was 86.07%. After perturbation the Accuracy for single Column (Age & gender) with J48 Algorithm was 85.75% in Existing Method for 86.01% for proposed random Swapping Method.

The Accuracy of the Original dataset without modifying the value for Two Column at a time (Education &Gender) was 86.07%. After perturbation the Accuracy for Two Columns at a time (Education &Gender) was 85.77% in Existing Method and for Proposed random Swapping Method is 85.91%.

The Accuracy of the Original dataset without modifying the value for Three Columns At a time (Education, Gender, Hours Per Week) was 83.25%. After perturbation, the Accuracy for Three Columns At a time (Education, Gender, Hours per Week) with J48 Algorithm was 85.64% in Existing Method for 85.96% for proposed random Swapping Method.

The Accuracy of the Original dataset without modifying the value for Three Columns at a time (Gender Marital Status Hours_Per_Week) was 86.07%. After perturbation, the Accuracy for Three Columns At a time (Gender Marital Status Hours_Per_Week per Week) with NB Algorithm was 85.91% in Existing Method for 85.96% for proposed random Swapping Method.

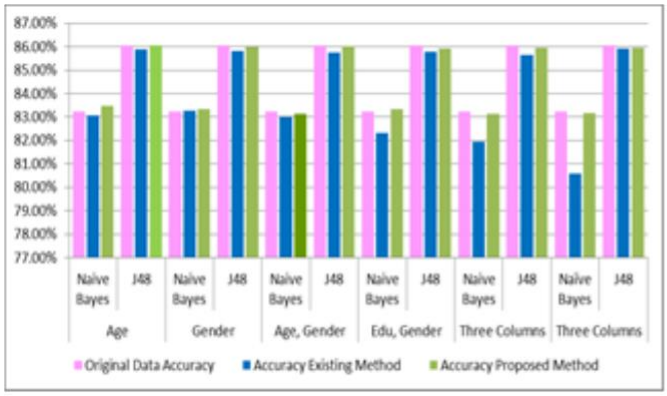

Figure 9: Accuracy Comparison of Existing and Proposed with NB Algorithm NB Algorithm and J48 Algorithm

Figure 9 represents the Pictorial representation of the Table 8 and proven that Proposed System has high data accuracy with compared to the existing privacy preserving method.

5.2.2 Performance Analysis Of Execution Time By Varying Data Size

Figure 10 illustrates the projected implementation time taken by changing data dimension, For the Data Dimension of 10000; implementation time obtained by the proposed framework is 160 ms by the Naïve Bayes Classification algorithm and 1320 ms for J48 Decision tree algorithm. And For the Data Dimension of 20000, implementation time obtained by the proposed framework is 160 ms by the Naïve Bayes Classification algorithm and 3330 ms for J48 Decision tree algorithm. As well as For the Data Dimension of 30000, implementation time obtained by the proposed framework is 250 ms by the Naïve Bayes Classification algorithm and 6070 ms for J48 Decision tree algorithm. And also For the Data Dimension of 40000, implementation time obtained by the proposed framework is 320 ms by the Naïve Bayes Classification algorithm and 10680 ms for J48 Decision tree algorithm. Finally, For the Data Dimension of 50000, implementation time obtained by the proposed framework is 520 ms by the Naïve Bayes Classification algorithm and 12520 ms for J48 Decision tree algorithm.

Data Size	Time(ms) With NB Algorithm	Time(ms) With J48 Algorithm
10000	160	1320
20000	160	3330
30000	250	6070
40000	320	10680
50000	520	12520

Table 9: Time Comparison of Existing and Proposed with NB Algorithm and J48 Algorithm.

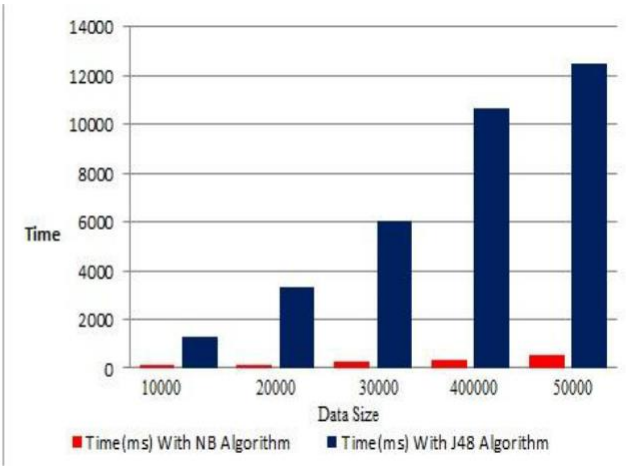

Figure 10: Time Comparison of Existing and Proposed with NB Algorithm and J48 Algorithm.

5.2. 3. Error Rate Comparision

Adult Data Set	Techniques Applied	Original Mean Absolute Error Value	Mean Absolute Error Value of	
			Existing Method	Proposed Method
Age	Naive Bayes	**0.1746**	**0.177**	**0.1727**
	J48	**0.1956**	**0.1989**	**0.1977**
Gender	Naive Bayes	**0.1746**	**0.1745**	**0.174**
	J48	**0.1956**	**0.1933**	**0.1937**
Age Gender	Naive Bayes	**0.1746**	**0.1775**	**0.176**
	J48	**0.1956**	**0.2002**	**0.1972**
Education Gender	Naive Bayes	**0.1746**	**0.1813**	**0.174**
	J48	**0.1956**	**0.1918**	**0.1934**
Education, Gender, Hours Per Week	Naive Bayes	**0.1746**	**0.1839**	**0.175**
	J48	**0.1956**	**0.1961**	**0.2003**
Marital Status , Gender , Hours Per Week	Naïve Bayes	**0.1746**	**0.1921**	**0.175**
	J48	**0.1956**	**0.1939**	**0.2003**

Table 10: Comparison of Mean Absolute Error Value of Existing and Proposed with NB Algorithm and J48 Algorithm

Table 10 illustrates the Comparison values of mean absolute value with two well-known algorithms such as NB Classification algorithm and J48 Decision tree Algorithm by perturbing Different Columns at

a time. The mean absolute error value of the Original dataset without modifying the value for single Column (Age) was 0.1746.After perturbation the mean absolute error value for single Column (Age) with NB Algorithm was 0.177 in Existing Method for 0.1727 for proposed random Swapping Method. The mean absolute error value of the Original dataset without modifying the value for single Column (Gender) was 0.1746After perturbation the mean absolute error value for single Column (Gender) was 0.1745 in Existing Method and for Proposed random Swapping Method is 0.174. The mean absolute error value of the Original dataset without modifying the value for Two Columns Column (Age & gender) was 0.1746.After perturbation the mean absolute error value for single Column (Age & gender) with NB Algorithm was 0.1775 in Existing Method for 0.176 for proposed random Swapping Method. The mean absolute error value of the Original dataset without modifying the value for Two Column at a time (Education &Gender) was 0.1746After perturbation the mean absolute error value for Two Columns at a time (Education &Gender) was 0.1813 in Existing Method and for Proposed random Swapping Method is 0.174. The mean absolute error value of the Original dataset without modifying the value for Three Columns At a time (Education, Gender, Hours Per Week) was 0.1746.After perturbation, the mean absolute error value for Three Columns At a time (Education, Gender, Hours Per Week) with NB Algorithm was 0.1839 in Existing Method for 0.175 for proposed random Swapping Method. The mean absolute error value of the Original dataset without modifying the value for Three Columns at a time (Gender, Marital Status Hours_Per_Week) was 0.1746.After perturbation, the mean absolute error value for Three Columns At a time (Gender, Marital Status Hours_per_Week) with NB Algorithm was 0.1921 in Existing Method for 0.175 for proposed random Swapping Method. So mean absolute error value of the Original dataset by using Naïve Bayes Algorithms is preserved with minimum information loss.

At the Same time, Comparison values of Accuracy with Two well-known algorithms such as NB Classification algorithm and J48 Decision tree Algorithm by perturbing Different Columns at a Time. The mean absolute error value of the Original dataset without modifying the value for single Column (Age) was 0.1956. After perturbation the mean absolute error value for single Column (Age) with J48 Algorithm was 0.1989 in Existing Method for 0.1977 for proposed random Swapping Method. The mean absolute error value of the Original dataset without modifying the value for single Column (Gender) was 0.1956After perturbation the mean absolute error value for single Column (Gender) was 0.1933 in Existing Method and for Proposed random Swapping Method is 0.1937. The mean absolute error value of the Original dataset without modifying the value for Two Columns Column (Age & gender) was 0.1956After perturbation the mean absolute error value for single Column (Age & gender) with J48 Algorithm was 0.2002 in Existing Method for 0.1972 for proposed random Swapping Method. The mean absolute error value of the Original dataset without modifying the value for Two Column at a time (Education &Gender) was 0.1956.After perturbation the mean absolute error value for Two Columns at a time (Education &Gender) was 0.1918 in Existing Method and for Proposed random Swapping Method is 0.1934. The mean absolute error value Three Columns At a time (Education, Gender, Hours per Week) was 0.1956. After perturbation, the mean absolute error value for Three Columns At a time (Education, Gender, Hours Per Week) with J48 Algorithm was 0.1961 in Existing Method for 0.2003 for proposed random Swapping Method. The mean absolute error value of the Original dataset without modifying the value for Three Columns at a time (Gender, Marital. Status Hours_Per_Week) was 0.1956.After perturbation, the mean absolute error value for Three Columns At a time (Gender, Marital_Status, Hours_Per_Week) with NB Algorithm was 0.1939 in Existing Method for 0.2003 for proposed random Swapping Method.

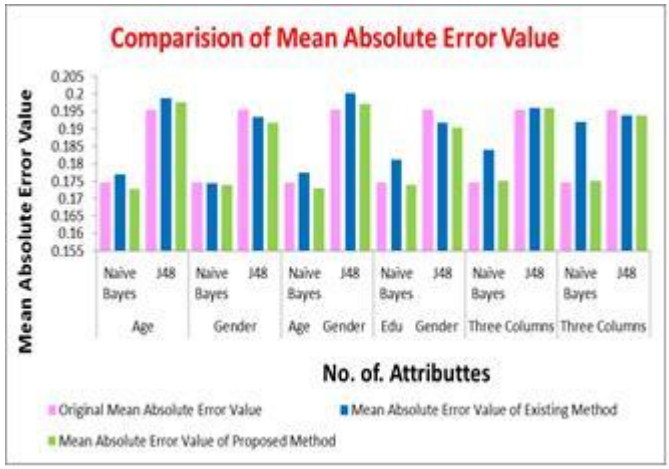

Figure 11: Comparison of Existing and Proposed of Mean Absolute Error Value with NB Algorithm and J48 Algorithm

Figure 11 represents the Pictorial representation of the Table 10 and proven that Proposed System has very low mean absolute value compared to the existing privacy preserving methods.

Adult Data Set	Techniques Applied	Original Root Mean Squared Error Values	Root Mean Squared Error Values of	
			Existing Method	Proposed Method
Age	Naive Bayes	0.3741	0.3775	0.3703
	J48	0.3201	0.3221	0.3217
Gender	Naive Bayes	0.3741	0.3745	0.3731
	J48	0.3201	0.3242	0.3211
Age Gender	Naive Bayes	0.3741	0.379	0.3758
	J48	0.3201	0.3259	0.3218

Education Gender	Naive Bayes	**0.3741**	**0.3781**	**0.3731**
	J48	**0.3201**	**0.3264**	**0.3211**
Education, Gender, Hours Per Week	Naive Bayes	**0.3741**	**0.3822**	**0.3753**
	J48	**0.3201**	**0.3292**	**0.325**
Marital Status , Gender , Hours Per Week	Naïve Bayes	**0.3741**	**0.4037**	**0.3753**
	J48	**0.3201**	**0.3269**	**0.325**

Table 11: Comparison Root Mean Squared Error Value of Existing and Proposed NB Algorithm and J48 Algorithm

Table 11 illustrates the Comparison values of Root Mean Squared Error Value with Two well-known algorithms such as NB Classification algorithm and J48 Decision tree Algorithm by perturbing Different Columns at a Time. The Root Mean Squared Error Value of the Original dataset without modifying the value for single Column (Age) was 0.3741.After perturbation the Accuracy for single Column (Age) with NB Algorithm was 0.3775 in Existing Method for 0.3703 for proposed random Swapping Method. The Root Mean Squared Error Value of the Original dataset without modifying the value for single Column (Gender) was 0.3741. After perturbation the Accuracy for single Column (Gender) was 0.3745 in Existing Method and for Proposed random Swapping Method is 0.3731

The Root Mean Squared Error Value of the Original dataset without modifying the value for Two Columns Column (Age & gender) was 0.3741.After perturbation the Root Mean Squared Error Value for single Column (Age & gender) with NB Algorithm was 0.379 in Existing Method for 0.3758 for proposed random Swapping Method. The Root Mean Squared Error Value of the Original dataset without modifying the value for Two Column at a time (Education &Gender) was 0.3741.After perturbation the Root Mean Squared Error Value for Two Columns at a time (Education &Gender) was 0.3781 in Existing Method and for Proposed random Swapping Method is 0.3731.

The Root Mean Squared Error Value of the Original dataset without modifying the value for Three Columns At a time (Education, Gender, Hours Per Week) was 0.3741.After perturbation, the Root Mean Squared Error Value for Three Columns At a time (Education, Gender, Hours Per Week) with NB Algorithm was 0.3822in Existing Method for 0.3753 for proposed random Swapping Method. The Root Mean Squared Error Value of the Original dataset without modifying the value for Three Columns at a time (Gender, Marital Status Hours_Per_Week) was 0.3741.After perturbation, the Root Mean Squared Error Value for Three Columns At a time (Gender, Marital Status Hours_per_Week per Week) with NB Algorithm was 0.4037 in Existing Method for 0.3753 for proposed random Swapping Method. So Root Mean Squared Error Value of the Original dataset by using Naïve Bayes Algorithms is minimized with minimum information loss.

At the Same time, Comparison values of Accuracy with Two well-known algorithms such as NB Classification algorithm and J48 Decision tree Algorithm by perturbing Different Columns at a Time. The Root Mean Squared Error Value of the Original dataset without modifying the value for single Column (Age) was 0.3201. After perturbation the Accuracy for single Column (Age) with J48 Algorithm was 0.3221 in Existing Method for 0.3217 for proposed random Swapping Method. The Root Mean Squared Error Value of the Original dataset without modifying the value for single Column (Gender) was 0.3201. After perturbation the Root Mean Squared Error Value for single Column (Gender) was 0.3242 in Existing Method and for Proposed random Swapping Method is 0.3211.

The Root Mean Squared Error Value of the Original dataset without modifying the value for Two Columns Column (Age & gender) was 0.3201 .After perturbation the Root Mean Squared Error Value for single Column (Age & gender) with J48 Algorithm was 0.3259 in Existing Method for 0.3218 for proposed random Swapping Method. The Root Mean Squared Error Value of the Original dataset without modifying the value for Two Column at a time (Education &Gender) was 0.3201. After perturbation the Root Mean Squared Error Value for Two Columns at a time (Education &Gender) was 0.3264 in Existing Method and for Proposed random Swapping Method is 0.3211.

The Root Mean Squared Error Value of the Original dataset without modifying the value for Three Columns at a time (Education, Gender, Hours per Week) was 0.3201.After perturbation, the Root Mean Squared Error Value for Three Columns At a time (Education, Gender, Hours per Week) with J48 Algorithm was 0.3292 in Existing Method for 0.325 for proposed random Swapping Method. The Root Mean Squared Error Value of the Original dataset without modifying the value for Three Columns at a time (Gender, Marital Status Hours_Per_Week) were 0.3201. After perturbation, the Root Mean Squared Error Value for Three Columns At a time (Gender, Marital Status, Hours_Per_Week) with NB Algorithm was 0.3269 in Existing Method for 0.325 for proposed random Swapping Method.

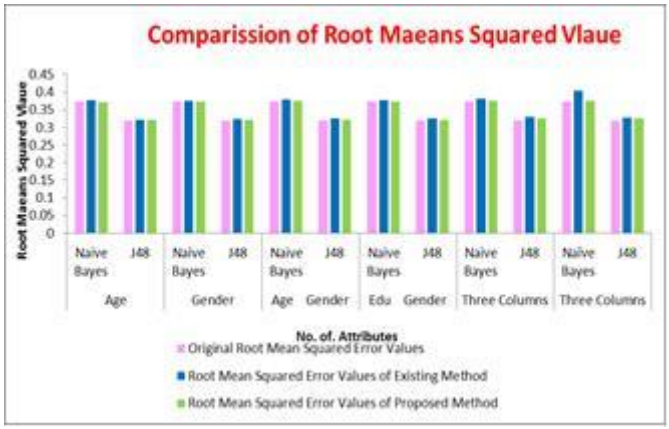

Figure 12: Comparison Root Mean Squared Error Value of Existing and Proposed NB Algorithm and J48 Algorithm

Figure 12 represents the Pictorial representation of the Table 11 and proven that Proposed System has minimum root means squared value with compared to the existing privacy preserving methods

6. CONCLUSION

Due to the huge Data explosion in our day to day life in recent years, achieving a higher level of privacy of an individual becomes a Challenging task in data mining. Preserving privacy for sensitive data is a critical issue in many Applications like Banking, Health, and Education in data mining.

There is a need for obtaining effective data mining results while maintaining the highest Confidentiality at the same time protecting disclosure of individual sensitive information had led to the existence of privacy preserving data mining. Still Balancing Privacy and accuracy are the challenging issues. From last two decades, there are many efforts have been attempted to preserve privacy and also obtaining a high level of accuracy.

In this proposed framework, an efficient random swapping based data perturbation technique is defined to preserve the sensitive individual data. The accuracy results of Existing and Proposed methods are compared with two well-known algorithms such as Naïve Bayes(Classifier algorithm) and J48(Decision tree Algorithm).In this proposed random swapping data perturbation method perturbs sensitive data by multiple columns in one iteration, which is a challenging issue in the field of data mining.

In existing Geometric data perturbation method having moderate data accuracy, with more data loss. To overcome such problem with the minimum data loss random swapping based method is introduced. The proposed method achieves more accurate results and minimum data loss compared to an Existing method. This work can be improved further with 100% accuracy by using an efficient data mining technique in preserving the privacy and also it may be extended by applying different normalization techniques for efficient privacy preservation for sensitive data. This work can also be enhanced further by adding more efficient lossless data Transformation techniques and also for the data streams by minimizing the information Loss.

REFERENCES

1. Data Mining, Melbourne, Florida, U.S.A., pp. 211-218 (2003).
2. M. Chen, J. Han, and P. Yu, "Data mining: An Overview from a database Prospective", IEEE Trans. on Knowledge and Data Engineering, vol. 8, no. 6, pp. 866-883, Dec. 1996.
3. C.L.P. Chen, C.Y. Zhang, "Data Intensive applications, challenges,Techniques and technologies: A survey on Big Data", Information Sciences, vol. 275, pp.314-347, 2014.
4. Alcalá, J, Fernández, A, Luengo, J, Derrac, J, García, S, Sánchez, L & Herrera, F 2010, „KEEL data-mining software tool: Data set repository, integration of algorithms and ti on Privacy and Security no‟s in Data Mining and Machine Learning‟, vol. 4, no. 3,pp. 127 – 128.
5. Xu, Lei, Chunxiao Jiang, Jian Wang, Jian Yuan, and Yong Ren, „Information security in big data: privacy and data mining", pp.1149-1176, 2014.
6. Atallah, M., Elmagarmid, A., Ibrahim, M., Bertino, E., Verykios.V:Disclosure limitation of sensitive rules, Workshop on Knowledge and Data Engineering Exchange, 1999.
7. Dunning, Larry A., and Ray Kresman, "Privacy preserving data sharing with anonymous ID assignment", IEEE Transactions on Information Forensics and Security, Vol.8, no.2, pp.402-413, 2013.
8. Li, Yaping, Minghua Chen, Qiwei Li, and Wei Zhang, "Enabling multilevel trust in privacy preserving data mining", IEEE Transactions on Knowledge and Data Engineering, Vol.24, no.9, pp.1598-1612, 2012.
9. Ms. Ompriya Kale, Ms. Prachi Patel, "A Survey on Privacy Preserving Data Mining Techniques", Global journal of Advanced Engineering Technologies, ISSN: 2277-6370, vol2, Issue3-2013.
10. Ricardo mendes,Joao p. vilela, „Privacy-Preserving Data Mining: Methods, Metrics, and Application‟, IEEE-2017 .
11. V.S.Mahalle,PankajJogi,ShubhamPurankar, SamikshaPinge, UrvashiIngale ," Data Privacy Preserving using Perturbation Technique", Asian Journal of Convergence in Technology Volume III, Issue III,ISSN No.:2350-1146, I.F-2.71, Dec, 2017.
12. Nikunj Kumar Patel, „Data Mining: Privacy Preservation Using Perturbation Technique, June, 2015.
13. AobakweSenosi, George Sibiya, „Classification and Evaluation of Privacy Preserving Data Mining: A Review", 2017 IEEE.
14. M.Reza,SomayyehSeifi," Classification and Evaluation the PPDM Techniues by using a data Modification -based framework", IJCSE,Vol3.No2 Feb,2011.
15. AjmeeraKiran, D.Vasumathi "Data Mining: A Study of Multiplicative Data Perturbation Technique for Privacy Preservation",intheICNTET2018:International Conference on New Trends in Engineering & Technology Tirupathi Highway, Tiruvallur Dist., Chennai, India, September 7-8, 2018. (Communicated AND accepted).
16. Twinkle Ankleshwaria, Prof. J. S. Dhobi," Geometric Data Perturbation Approach for Privacy Preserving in data Stream Mining" Engineering Universe for Scientific research and Management, Impact factor 3.7, Volume 6, Issue 4, April 2014.

Knowledge and Information Systems (2005) 7: 387–414,DOI 10.1007/s10115-004-0173-6.
18. HinaVaghashia, AmitGanatra, 2015. " A Survey: Privacy Preservation Techniques in Data Mining " , In. Proceedings of International Journal of Computer Applications (0975 – 8887) Volume 119 – No.4, June 2015.
19. Mehta, Brijesh B., and UdaiPratapRao, "Privacy preserving big data publishing: a scalable k-anonymization approach using MapReduce", IET Software, Vol. 11, No. 5, pp. 271-276, 2017.
20. Ienberg S., McIntyre J.: Data Swapping: Variations on a Theme by Dalenius and Reiss. Technical Report, National Institute of Statistical Sciences, 2003.
21. Vaidya, Jaideep, BasitShafiq, Wei Fan, Danish Mehmood, and David Lorenzi, "A random decision tree framework for privacy-preserving data mining", IEEE transactions on dependable and secure computing, no. 5, pp. 399-411, 2014.
22. UCI Machine Learning Repository http://archive.ics.uci.edu/ml/datasets.html
23. Weka 3.6 Data Mining Tool , https://www.cs.waikato.ac.nz/ml/weka/downloading.html

Ajmeera Kiran is currently working as Full Time Research Scholar in the Department of Computer Science and Engineering at Jawaharlal Nehru Technological University, Kukatpally, Hyderabad, and Telangana, India. He Received B. E Degree in Computer Science and Engineering from Vasavi College of Engineering Affiliated to Osmania University, Hyderabad. M.Tech Degree in Information Technology from Jawaharlal Nehru Technological University, Kukatpally, Hyderabad. He has Published 04 Research Papers in International Conferences and 1 in International Journals. His area of research interest includes data mining, Information Security, and Network Security.

Dr.D.Vasumathi presently Professor of CSE, Professor in charge student's welfare, NSS Coordinator at JNTUH College of Engineering, Jawaharlal Nehru Technological University Hyderabad, Telangana-India. She served and held several academics and administrative positions including hostel warden (JNTUHCEH), Office in charge of examinations (JNTUHCEH), Additional Controller of Examinations (JNTUH), Member of Board of Studies at various JNTUH affiliated colleges and student advisor. She received B.Tech and M.Tech degrees in Computer Science and Engineering from JNT University, Hyderabad and Ph.D in CSE from JNTU, Hyderabad. She is recipient of national award of Savitribai Phule award in 2017. She has guided 12 Ph.D thesis, 60 M.Tech and B.Tech projects and she has published more than 50 research papers in National and International Journals and conferences including IEEE, ACM, Springer and Inderscience publishers. She has 18 years of experience in teaching. She has organized 10 workshops and 3 refreshment courses. She served as confidential team member for EAMCET and Police Recruitment. Her areas of interest are Data Mining, Big data Analytics, Cloud Computing and Computer Networks. She is a life member of ISTE and IEEE.