International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Studies on the Use of Various Noise Strategies for Perturbing Data in Privacy-Preserving Data Mining

Dr. Prarthana A. Deshkar[1], Dr. Jaikumar M. Patil[2], Dr. Pornima B. Niranjane[3], Prof. Vaishali Niranjane[4], Dr. Nisha Thakur[5], Vaibhav D. Dabhade[6]

Submitted: 08/10/2023 **Revised**: 29/11/2023 **Accepted**: 09/12/2023

Abstract: Privacy Preserving Data Mining techniques broadly fall under the categories of randomization and value distortion. Randomization replaces the existing value with a non existing value whereas value distortion modifies each value in the database. Data perturbation is widely used randomization technique that promises both valid data mining result and secured privacy. Existing research works were conducted with additive and multiplicative forms of data perturbation. Privacy is measured with the increasing level of error rate based on different types of attacks. Attacks in data perturbation are normally noise filtering methods which are based on linear and nonlinear filtering schemes. Linear filtering process produces the output data which will be in linear combination of the input values. The perturbed copies in additive and multiplicative data perturbation is generated with Laplace noiseWhen the input data is in linear distribution and is subjected to perturbation, linear filtering schemes are used to reconstruct the data and analyze the privacy measure. In nonlinear filtering schemes, the output of the filtering process will not be in linear distribution. This means that for a non linear data distribution, linear noise filtering schemes cannot be used for accurate analysis. Attacks considered so far are based on only linear filtering schemes. The perturbed models are evaluated with both privacy and utility of data mining. The work describes both linear and non linear type of attacks over the generated perturbed data. Attacks for additive data perturbation include Maximum A Posteriori (MAP) and Principal Component Analysis (PCA) based filtering method.

Keywords: *Privacy Preserving, Data Mining, Data perturbation, Maximum A Posteriori, Principal Component Analysis (PCA)*

1. Introduction

Sophisticated and powerful technologies today are likely to collect data from different applications and create a massive data repository. Confronted with the massive collection of data, the need for better managerial choices arose in the form of data summarization, extracting the fundamental nature of the information stored and the innovation of new patterns from the raw data. Data mining emerged as a powerful new technology for extracting hidden information from large amount of data[1][2]. The process is also referred as knowledge discovery and it helps to analyze data from different perspectives. Various data mining tools are used to predict the future behavior and trends of the data that helps business people to enhance their decisions accordingly. Technically, Data Mining is defined as the process of discovering information from large amount of data. In other words, it can be said that data mining is the method of mining knowledge from the archived collection of data[3].

The process in knowledge discovery includes the following steps.

- Data Preprocessing includes Data cleaning, Data integration and data transformation.
- Data cleaning involves the removal of noise and inconsistent data from the data warehouse. • Data integration combines data from different data sources and finally data transformation selects the attributes and transforms into different forms appropriate for data mining.
- Data Mining applies intelligent methods in order to extract patterns from the data.
- Pattern Discovery is used to discover new patterns.
- Knowledge Presentation displays the knowledge inferred from the pattern

[1]*Assistant Professor, Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur*
prarthana.deshkar@gmail.com
[2]*Associate Professor, Department of CSE, Shri Sant Gajanan Maharaj College of Engineering, Shegaon*
jaimpatil1011@gmail.com
[3]*Assistant Professor. Department of CSE Babasaheb Naik College of Engineering,*
pornimaniranjane@gmail.com
[4]*Assistant Professor, Department of Electronics and Telecommunication Yeshwantrao Chavan College of Engineering, Nagpur*
niranjane_vaishali@yahoo.com
[5]*Assistant Professor, Faculty of Engineering and Technology, DMIHER(DU), Sawangi, Wardha*
nishat.feat@dmiher.edu.in
[6]*Assistant Professor, MET Institute of Engineering, Nashik*
vaibhavdabhade@rocketmail.com

Sensitive Data In Data Mining

Identifying new patterns from the historical collection of data is a trendy probing technique in data mining. Nevertheless, if the data values are sensitive then discharging information about noteworthy patterns or trends will lead to significant risk to privacy. For example, if data mining is performed on patient health records or customers purchase records in a super market; sensitive data may include patients‟ personal details about diseases and the customers‟ personal details respectively. Sensitive data disclosure susceptibility is one of the major decisive security threats that can result in negotiating the security of data mining applications[4-7]. It happens when the sensitive data is disclosed to unauthorized data miners and such unauthorized access may lead to severe information security and data privacy violations. It thus raises the need to identify methods that focuses on how to precisely learn and discharge the noteworthy patterns in a data set that contains sensitive information, while providing meticulous assurance to privacy of the persons whose information is collected and stored.[8-10]

2. Need for Privacy Preservation

Individual‟s privacy is always a common thought and a concept which is always dealt with when handling data for enterprise decision making. With the enhancement of computer technologies and wide usage of web applications, privacy threats pose different technical challenges in today‟s information era. Privacy is defined as the users right to preserve and obscure their personal information and have controlled private information disclosure. Various data mining techniques are used to extract constructive patterns from vast collection of data. It supports domains like medical diagnosis, national security, marketing, weather forecasting etc.,[11] Though advantageous in various domains, it is still a challenging issue when users privacy is breached. For example, medical diagnosis involves patient‟s health records to be mined for successful patterns. This may include mining of patient‟s private information also. When the need for successful patterns from the data owners increases, same happens with the privacy threat level of the users. Marketable issues are also allied with the privacy concern. Many organizations collect customer‟s data for specific purposes. When different departments of the organization share the data, it should be ensured that private sensitive data is not despoiled. In such scenarios, need to enhance data mining algorithms to include privacy preservation emerged. Hence, many researches are committed to tackle the privacy preservation in data mining. Various techniques are proposed to hide private information either by removing the reserved data or by adding noise. The need for privacy preservation can be summarized as follows.[12]

- Discovering useful patterns from the sensitive data.
- Use of single and multiple databases to extract information about individuals.
- Posing open ended queries to find relationships about individuals and groups of individuals.
- Inferring associations between records to retrieve sensitive information
- Non-Predictive aspect of information about individuals during data mining.
- Private and public nature of information gained from data mining
- The potential to create new clusters based on generated patterns. The risks to privacy posed by data mining include, but are not limited to, the following: illegal access to personal data, the revelation of uncomfortable information, the use of personal data for purposes other than those for which it was originally obtained, etc[13]. Companies nowadays often choose to keep any data found from the actions of their consumers, thanks to the prevalence of technology and the ease with which gear can be acquired. As a result, there's rising anxiety that data collectors would use the information they unearth to spy on their customers.[14]

Privacy Preserving Data Mining

A branch of data mining termed "Privacy Preserving Data Mining (PPDM)" has arisen to solve the aforesaid difficulties. In recent years, it has grown in many different ways. The primary goal of PPDM is to prevent confidential data from falling into the wrong hands. In recent years, PPDM has emerged as a promising new area of study for data scientists. The effectiveness and security of data mining methods are examined. PPDM is thus being considered for a dual purpose.[15][16] The first is to prevent users' private information from falling into the wrong hands, and the second is to mine useful information from secure databases. In a word, the goal is to get valid data mining findings without gaining access to sensitive information. Figure 1 depicts the organizational framework for data mining that protects individual privacy.[17]

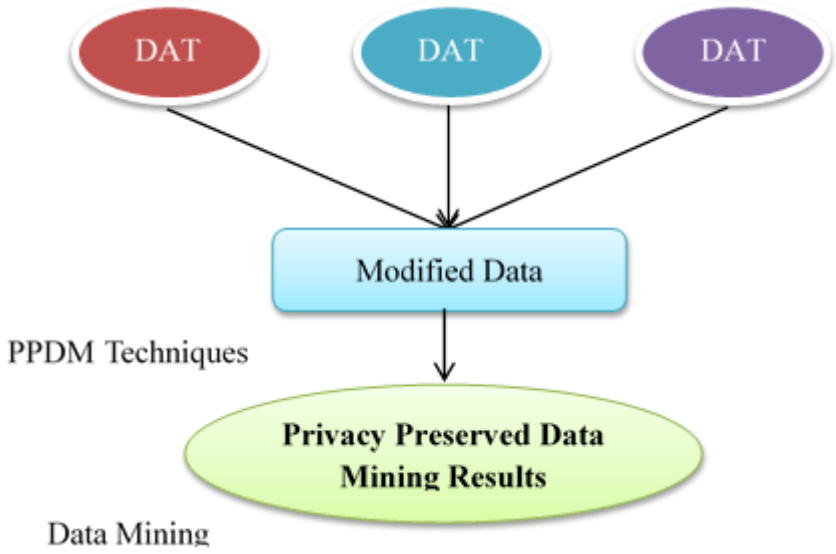

Fig 1 Block diagram of PPDM

PPDM techniques mainly involve two major methods

- Data Perturbation methods that mask the data so that sensitive data is not revealed

- Cryptographic methods that use some key values to encrypt the sensitive data. Generally there are two fundamental problems in PPDM: collection of privacy preserved data and mining the sensitive data across several private platforms.[3][18] The main goal of PPDM algorithms is to retrieve valid information from data warehouse at the same time preserving the private sensitive data. There are many PPDM methods developed in recent years, nevertheless there is no standardized approach to achieve privacy and utility. For getting valid results efficiently, the following dimensions of subjects need to be considered.[4]

- How the data is distributed against the enterprise?

- How do PPDM methods work?

- Mining methodologies and Privacy preservation methods The largest part of the enterprises is in need of secure data transmission and security preservation when the data is stored in the data warehouse. Normally PPDM techniques add values to traditional data mining methodologies by working with hidden sensitive data. The issue faced here is how to do achieve better data mining result from the distorted data. The target of a PPDM algorithm is

- To prevent the identification of sensitive data

- To apply the algorithm to huge collection of data

- To have less exponential computational complexity

To secure the sensitive information during data mining, various techniques have been proposed. Value distortion and randomization are the two basic approaches towards privacy preservation. Randomization method replaces the existing item with non existing item whereas distortion changes every item in a transaction. Data perturbation is also a common approach to preserve the sensitive data. This method perturbs the sensitive data before it is released for data mining.[5][6] The main advantage of perturbation based method is that it preserves the statistical properties of the data and thus the data mining results will be accurate. The drawback is that there is no proper framework for measuring the privacy. However, recent perturbation techniques try to measure the amount of that is guaranteed. Another limb of PPDM uses cryptographic methods for preventing leakage of sensitive information during data mining.[7][16] It proves perfect for two main reasons: one is, it offers a well defined privacy preserving model and the other is, there are various tools of cryptographic algorithms for implementing privacy preservation. Nevertheless, it has also been proved that cryptographic methods do not maintain the utility for data mining models. Though it is helpful in preventing privacy leaks, it falls short of providing a better data mining model. Anonymization techniques also have received a lot of attention to privacy preservation process. The data here is segregated into public and private columns. Public data is made available to all users of the data table whereas private columns are not released in external tables. The privacy measure is that an attacker may not be able to relate private data not more than k individuals. It is made possible by providing every combination of public columns with the private columns in at least k rows. K-anonymity model enforces privacy in context of releasing public data[8]. Data owner ensures that attackers will not be able to link the data which is gained from the database to individual's sensitive data from

whom the data is collected. Differential privacy has also gained popularity in privacy preserving data models. It necessitates that any process to the database should be insensitive to the changes applied in any individual's record. Privacy is guaranteed with aggregate composition of all the individual private records. This method has many advantages over previous methods. It relies on mathematical definition to privacy and hence it becomes easy to prove whether a technique is conventional to differential privacy and to figure out calculations that can or cannot be made in this skeleton. Differential privacy does not depend on attacker's background knowledge. This independence makes data owners to share the data without any concern about past and future releases. It also guarantees the privacy when there is joint reconstruction of two independent differentially private data[10]

Distributed Privacy Preservation

In a decentralized setting, people may seek to pool their individual copies of altered data in order to get meaningful aggregate findings. The primary goal of distributed privacy preservation is to reach consensus on the working out of cumulative information across all data without disclosing any sensitive details. Entities in a dispersed setting don't have faith in one another when it comes to sharing data. Both vertical and horizontal data partitioning are possible in a distributed setting.[11] In a vertically partitioned database, different entities may see different fields from the same set of entries. In horizontal partitioning, individual records are extended among various entities with the same set of characteristics. The difficulty of distributed privacy preservation is complicated by the fact that vertical and horizontal partitioning schemes provide distinct difficulties. Secure multiparty computing, which is used in cryptographic techniques, is also used to preserve privacy in distributed data. Cryptographic approaches, on the other hand, calculate functions over inputs received from a number of different receivers without disclosing each of the inputs individually.[12]

3. Data Perturbation

The most widely used method for PPDM is the Data Perturbation approach. Perturbation approach is used in applications where the data owners wish to allow the data for collaborative data mining without compromising individual's privacy. Data publishing for research, releasing the data to the third party data miners are examples where data perturbation can be applied. Statistical Disclosure Control (SDC) techniques help in data perturbation by releasing the statistics of the data without the leakage of private sensitive information. Statistical methods necessitate some kind of data modification to ensure the privacy protection.[13] This lies between data encryption where the data set is fully encryption with minimum or no data mining utility and no modification which allows maximum utility of data mining. Data Perturbation modifies the sensitive data in such a way that both privacy and data mining utility are balanced. The main challenge in statistical disclosure control techniques lies in modifying the data with minimum information loss while mining. Two broad categories of masking methods can be dealt with: Perturbative and Non-Perturbative masking methods. Both the techniques are used for continuous and categorical attributes.[14] Continuous attributes are numerical and arithmetic operations are allowed on it. For example, age, income etc are continuous attributes. A categorical attribute limits the value set to it. An ordered relationship can be defined over the categorical attribute. For example, marital status, gender etc., are categorical attribute[5][15]

Perturbative Masking method

In Perturbative masking method the dataset is modified before publishing. The original statistics of the data will be overwritten with the new unique statistics in the perturbed dataset. This enables to preserve statistical confidentiality. The techniques for perturbation should be in such a way that the new statistics should not result in different results that would be obtained from the original statistics. Random noise addition, resampling, swapping the data and aggregation are some of the perturbative masking methods[3][6].

Where X is the original data, A is the transformation method, and Z is the noise introduced, Y is the result of using the perturbative masking approach. When it comes to perturbative masking methods, the noise additive approach is by far the most used. The approach introduces some chance variation into the baseline readings[4][11]. The impact of adding noise to the original sensitive data is determined by the statistical characteristic of the noise. Various noise addition strategies are presented to retain the statistical features of the original data. When perturbing sensitive data with uncorrelated noise, a vector of a particular attribute is swapped out with another vector that is the original vector plus a vectors of normal distribution noise.[12]

Perturbation using correlated noise – Addition of correlated noise preserves the correlation coefficients. The covariance matrix of the perturbed data and the noise is proportional to the covariance matrix of the original data.

The insertion of correlated noise works well with continuous data. Algorithms above address the usage of additive noise technique for perturbation of sensitive

properties. The disturbance caused by correlated noise may also be multiplicative. Multiplicative data perturbation techniques keep sensitive information private while maintaining data usefulness. When numerous transformations are applied, a transformation-invariant model for data mining is the outcome.[7] Rotational perturbation, projectional perturbation, and geometric perturbation are the three varieties of multiplicative perturbation. Because it maintains the distance and inner product, which is necessary for classification and clustering methods, multiplicative data perturbation is helpful in the classification and clustering process. This ensures that the accuracy of the modified data mining model is equivalent to that of the original data model.[5]

Additive Data Perturbation

Additive Data Perturbation approach preserves the data privacy by adding a randomly generated noise to the original data. Assuming the original data denoted as X and the random noise as Z, the perturbed copy (Y) is obtained by the Equation (1)

$$\mathbf{Y = X + Z} \quad \textbf{(1)}$$

This method preserves the privacy of the sensitive data by adding some random noise at the same time making sure that the random noise still preserves the original data from the data so that the data mining patterns can still be accurately estimated. The data owner returns a value of X+Z of the original dataset, where Z is drawn from a certain distribution. A large amount frequently used distributions are the uniform distribution over a specified interval [-α,+ α] and Gaussian distribution with mean μ = 0 and standard deviation σ. In this interpretation, the raw data have the same distribution as the random variable but are considered to be independent random variables. Randomized perturbation experiments use such independent samples. The data provider also supplies the function of cumulative distribution and the perturbed values. There are two common methods for introducing additive noise into data. Membership in the Value Class Distribution of Values The perturbation in the value distribution technique is based on a random value drawn from a predetermined distribution. Estimating the density of a function is a typical challenge in data analysis as well as in security applications. The density data may be used to issues of grouping, classification, and the like. Using additive noise to perturb data allows for a fair estimation of the underlying density function.

Single Level Gaussian Additive Data Perturbation

The framework for additive Gaussian noise at single level trust is depicted in Figure 2. As illustrated in the figure, the sensitive attributes are extracted from the original data set and is taken up for perturbation. Random Gaussian noise is generated using Gaussian function and is added to the sensitive attribute to generate the perturbed copy

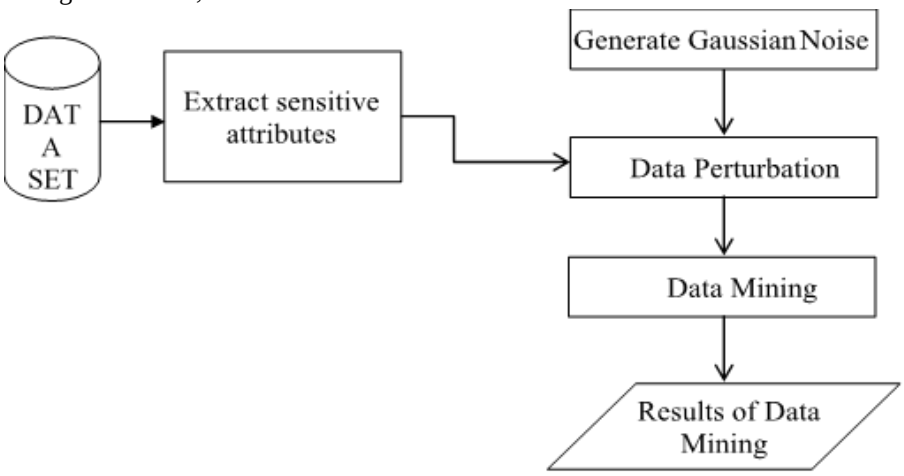

Fig 2 Additive Gaussian Data Perturbation at Single Level Trust

The algorithmic steps of the process involved are shown in Algorithm 1.

Input : Original Data set X, Covariance C_x, Noise N
Output : Perturbed data P
Generate random Gaussian noise GN ~ N in (0, σ2Cx)
for sensitive attributes
for all values in each sensitive attributes
Construct P = X + GN
End for
End for

Algorithm 1 Single Level Gaussian Additive Perturbation

Single Level Geometric Data Perturbation using Gaussian Noise

From the study of literature, it is found that out of the three representative methods of multiplicative data perturbation, geometric data perturbation is found to be effective to knowledgeable attackers. Hence, geometric data perturbation is taken into consideration for the research findings. The framework for multiplicative data perturbation using Gaussian noise at single level trust can be described in the diagram shown in Figure 3

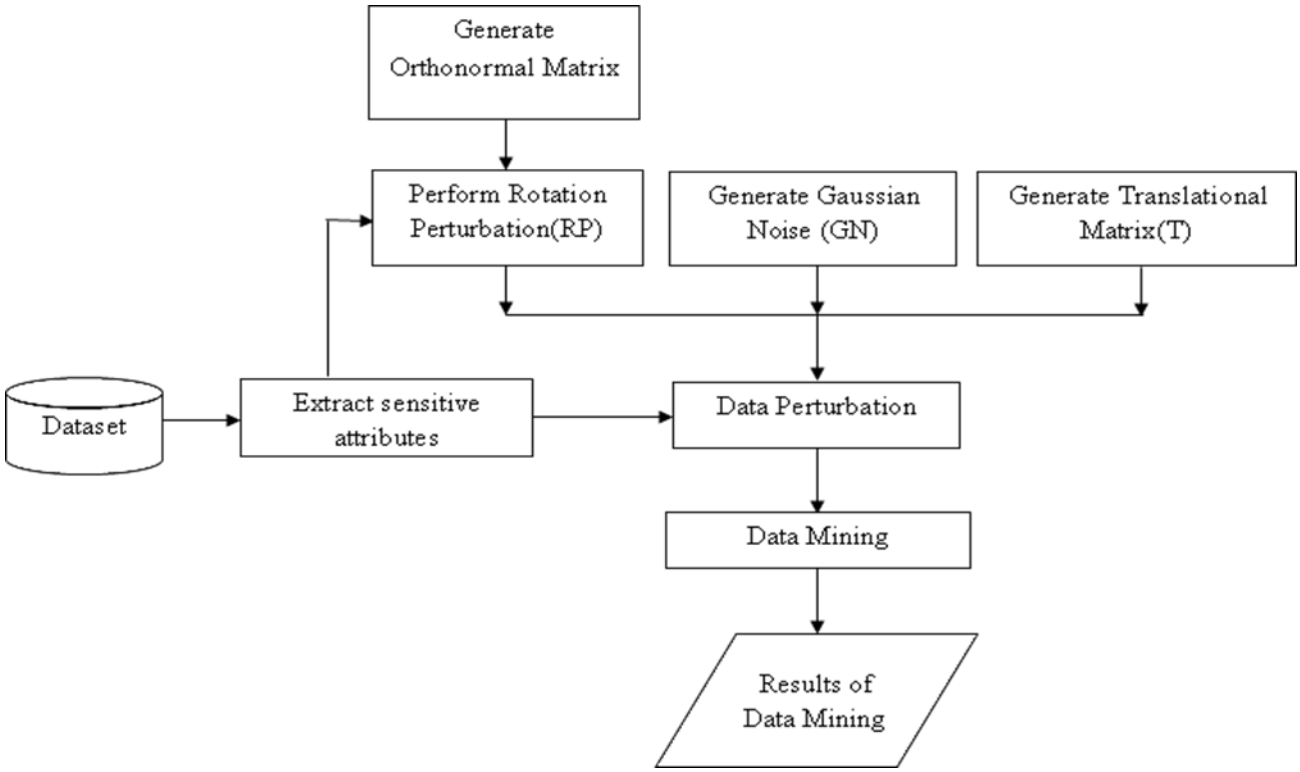

Fig 3. Multiplicative Gaussian Data Perturbation at Single Level Trust

4. Experimental Evaluation

In this section, experiments are carried out to evaluate the effectiveness of Multiplicative Gaussian noise technique over additive Gaussian noise. Since data perturbation involves addition of noise to the sensitive attributes, privacy is measured by analyzing the error that is obtained in reconstructing the original data by removing noise. Various noise filtering schemes are used for this purpose. In the experimental evaluation, privacy precision is computed by testing the perturbed data with PCA and ICA based noise filtering schemes. Data mining utility is also evaluated by utilizing classification model over the perturbed data. The proposed research work is experimented with bank and credit card data set from UCI repository. The bank dataset contains 25,000 instances with 16 attributes. The test data has 500 instances. Out of the 16 attributes, age and balance are taken as the sensitive attribute. In the credit card data set 20000 instances for train data and 6000 instances for test data is available. There are 24 attributes out of which age and credit-limit are taken as the sensitive attributes. The experiments are carried out in MATLAB for measuring the privacy level and in WEKA for computing the classifier accuracy. It is assumed that the attackers have

the knowledge on the distribution of noise, mean and covariance of the original data and the perturbed data.

Additive Gaussian noise data perturbation

Let the original data values be X = {25, 27, 31, 32}. Gaussian noise is generated randomly using mean as 0 and variance as σ2Cx. Cx is the covariance matrix of X and the value of σ2 is assumed to be 0.05. The randomly generated Gaussian noise is obtained by adding mean (which is assumed as 0) with σ2Cx * rand (N,1).

This gives the values of Gaussian noise as

GN = {-0.9212, 1.8126, -0.19322, 1.22547}

With the Gaussian noise, perturbed data is obtained by adding the noise to the sensitive data values X.

Perturbed data = X + GN

Peturbed data = {25-0.9212, 27+1.8126, 31-0.19322, 32+1.22547}

= {24.0788, 28.8126, 30.80678, and 33.22547}

Estimation of Classifier model accuracy

Gaussian additive and multiplicative perturbed copies of the data are utilized for evaluating classifier algorithms such as Decision Tree classifier, Naïve Bayes classifier and KNN classifier. It is verified that the perturbed copies with additive and multiplicative techniques has the same data mining utility as that of the original data. The utilities of decision tree and naïve Bayes classifier models are measured. Table 1 below shows the classifier accuracy of different models under single level trust.

Table 1: Single-Level Trust Accuracy of Classifiers for Gaussian Data Disturbances

	Bank Dataset			**Credit Card Dataset**		
Classifier Accuracy	**Decision Tree**	**Naïve Bayes**	**KNN**	**Decision Tree**	**Naïve Bayes**	**KNN**
Original Data	92.76	90.24	86.42	79.82	42.3	75.12
Gaussian Additive	91.16	91.81	78.42	77.42	52.3	42.22
Gaussian Multiplicative	87.25	87.76	87.42	75.62	55.22	69.78

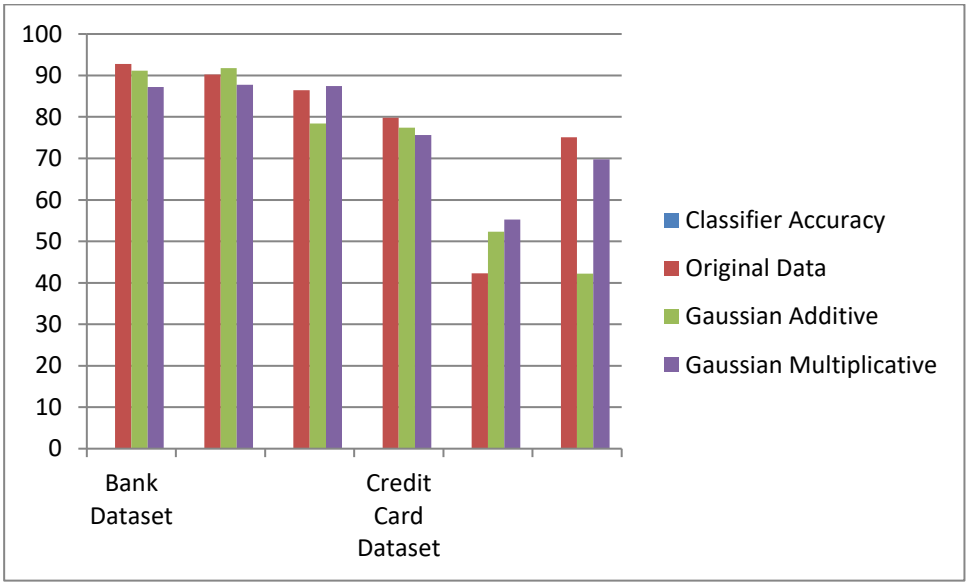

Fig 4. shows the results obtained for all the three classifier models under multi-level trust.

The results clearly show that Gaussian Additive scheme provides nearly equal classifier accuracy as that of the original data for Decision Tree and Naïve Bayes classifier algorithm. Gaussian multiplicative technique is better for KNN classifier model for both bank and credit card dataset. This is because multiplicative scheme performs an orthonormal transformation which preserves the Euclidean distance of the data points. Hence, the distance based classifier model suits to multiplicative perturbation providing better classifier accuracy.

5. Conclusion

The main focus of this research is to enable efficient data mining process with the distorted data. Data perturbation approach utilizing Gaussian noise is presented. The framework is experimented with both Gaussian additive and Gaussian multiplicative schemes. In Gaussian additive method, randomly generated Gaussian noise is added to the sensitive data. The perturbed data is evaluated for privacy precision and classifier accuracy. Under Gaussian multiplicative method, rotation matrix, translation matrix and random Gaussian noise components are added to the sensitive data. The perturbed data is then experimented for classifier accuracy. Algorithms are tested under single level and multilevel trust of the data miners. Under single level trust there is no significant change in the results. In multilevel trust scenario, it is proved that even if the number of perturbed copies available increases, there is no diversity gain in the reconstruction of the original data. Compared to Gaussian additive scheme, Gaussian multiplicative scheme gives almost identical results for all levels of trust. Experimental results proved that Gaussian multiplicative scheme finds better solution to privacy preservation and utility preservation when compared to Gaussian additive method when the classifier model is developed using KNN algorithm.

References

[1] S. M. Jesus and O. C. Rodríguez, "Estimating excess noise from deep sea mining: a simulated test case," *OCEANS 2023 - Limerick*, Limerick, Ireland, 2023, pp. 1-6, doi: 10.1109/OCEANSLimerick52467.2023.10244695.

[2] T. Zhang, X. Liu, J. Wang and W. K. Victor Chan, "Improve Data Mining Performance by Noise Redistribution: A Mixed Integer Programming Formulation," *2023 IEEE International Conference on Smart Internet of Things (SmartIoT)*, Xining, China, 2023, pp. 190-195, doi: 10.1109/SmartIoT58732.2023.00034.

[3] P. Dubey and A. Rajavat, "Effective K-means clustering algorithm for efficient data mining," *2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)*, Vellore, India, 2023, pp. 1-6, doi: 10.1109/ViTECoN58111.2023.10157179.

[4] T. Zhang, X. Liu, J. Wang and W. K. Victor Chan, "Improve Data Mining Performance by Noise Redistribution: A Mixed Integer Programming Formulation," *2023 IEEE International Conference on Smart Internet of Things (SmartIoT)*, Xining, China, 2023, pp. 190-195, doi: 10.1109/SmartIoT58732.2023.00034.

[5] P. Lin, S. Peng, Y. Xiang, C. Li, X. Cui and W. Zhang, "Structure-Oriented CUR Low-Rank Approximation for Random Noise Attenuation of Seismic Data," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-13, 2023, Art no. 4504713, doi: 10.1109/TGRS.2023.3297999.

[6] M. F. Uddin, "An Enhanced Machine Learning Approach to Identify Noise and Detect Relevant Structures for Predictive Modeling," *2023 9th International Conference on Information Technology Trends (ITT)*, Dubai, United Arab Emirates, 2023, pp. 55-60, doi: 10.1109/ITT59889.2023.10184237.

[7] L. Wang and T. Zhang, "The Application of Data Mining Algorithm in the Legal Protection of Personal Data," *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)*, Dalian, China, 2022, pp. 1339-1342, doi: 10.1109/ICDSCA56264.2022.9988630.

[8] Y. Lu, X. -D. Liu, J. -J. Song, T. -H. Xu and E. C. C. Tsang, "Data Noise Suppression-based Attribute Reduction," *2022 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, Toyama, Japan, 2022, pp. 89-94, doi: 10.1109/ICWAPR56446.2022.9947188.

[9] Y. Lu, X. -D. Liu, J. -J. Song, T. -H. Xu and E. C. C. Tsang, "Data Noise Suppression-based Attribute Reduction," *2022 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, Toyama, Japan, 2022, pp. 89-94, doi: 10.1109/ICWAPR56446.2022.9947188.

[10] S. Wang, P. Song, J. Tan, B. He, Q. Wang and G. Du, "Attention-Based Neural Network for Erratic Noise Attenuation From Seismic Data With a Shuffled Noise Training Data Generation Strategy," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-16, 2022, Art no. 5918916, doi: 10.1109/TGRS.2022.3197929.

[11] R. Churchill and L. Singh, "Topic-Noise Models: Modeling Topic and Noise Distributions in Social Media Post Collections," *2021 IEEE International Conference on Data Mining (ICDM)*, Auckland, New Zealand, 2021, pp. 71-80, doi: 10.1109/ICDM51629.2021.00017.

[12] J. Shan, Y. Lin and X. Zhu, "A New Range Noise Perturbation Method based on Privacy Preserving Data Mining," *2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS)*, Dalian, China, 2020, pp. 131-136, doi: 10.1109/ICAIIS49377.2020.9194850.

[13] C. Boonseng, R. Boonseng and K. Kularbphettong, "Noise and Vibration Analysis of Dry-Type Power

Transformer for Monitoring and Data Mining Applications," *2019 22nd International Conference on Electrical Machines and Systems (ICEMS)*, Harbin, China, 2019, pp. 1-5, doi: 10.1109/ICEMS.2019.8922194.

[14] Y. Shi, "Multi-Dimensional Processing for Big Data with Noise," *2019 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, Shenyang, China, 2019, pp. 686-690, doi: 10.1109/ICPICS47731.2019.8942582.

[15] S. Yaji and B. Neelima, "Optimizing Privacy-Preserving Data Mining Model in Multivariate Datasets," *2019 PhD Colloquium on Ethically Driven Innovation and Technology for Society (PhD EDITS)*, Bangalore, India, 2019, pp. 1-3, doi: 10.1109/PhDEDITS47523.2019.8986965.

[16] W. Li, H. Zhu, W. Liu, D. Chen, J. Jiang and Q. Jin, "An Anti-Noise Process Mining Algorithm Based on Minimum Spanning Tree Clustering," in *IEEE Access*, vol. 6, pp. 48756-48764, 2018, doi: 10.1109/ACCESS.2018.2865540.

[17] H. S. Aggarwal, A. Kansal and A. Jamshed, "Noisy information and progressive data-mining giving rise to privacy preservation," *2017 3rd International Conference on Advances in Computing,Communication & Automation (ICACCA) (Fall)*, Dehradun, India, 2017, pp. 1-5, doi: 10.1109/ICACCAF.2017.8344673.

[18] Y. Sei and A. Ohsuga, "Private True Data Mining: Differential Privacy Featuring Errors to Manage Internet-of-Things Data," in *IEEE Access*, vol. 10, pp. 8738-8757, 2022, doi: 10.1109/ACCESS.2022.3143813.