An Improved Rotation-Based Privacy Preserving Classification in Web Mining Using Naïve Bayes Classifier

Subramanian SANGEETHA MARIAMMAL[1]*, Ashok KAVITHAMANI[2]

[1] Department of Computer Science and Engineering, Coimbatore Institute of Technology, Coimbatore-641014, Tamil Nadu, India
ssmcse84@gmail.com (**Corresponding author*)

[2] Department of Electrical and Electronics Engineering, Coimbatore Institute of Technology, Coimbatore-641014, Tamil Nadu, India
kavithamaniashok@gmail.com

Abstract: Recently, the privacy and security of big data has become an important challenge, which requires the privacy preserving data mining techniques to maintain the trade-off between the data utility and privacy. Web mining is the application of data mining techniques for mining the web data. Privacy issues on the web are based on the fact that most users want to maintain a strict anonymity on web applications and activities. Various conventional techniques are used for privacy preservation like condensation, randomization and tree structure etc. The limitations of the existing approaches are maintaining a balance between the data utility and privacy and the scalability problem. The data stream can be secured and classified by privacy preserving techniques such as perturbation, cryptography and machine learning techniques, etc. Here the machine learning technique Naïve Bayes is employed to classify the perturbed data streams and provides an efficient accuracy. In this proposed research work, the UCI web mining datasets are collected, clustering is done with the help of Fuzzy C-Means (FCM), Flip and Rotation Perturbation (FRP) technique is applied to perform data renovation and classification on perturbed data is done using Naïve Bayes classifier. The classifier is improved by using a holdout approach on data separation in training and testing phases. The classification accuracy, computation time, and error rate of the classifier are measured and they are compared with the ones of the existing method. The comparison shows the achieved result of the proposed method. This proposed system is done with the help of MATLAB 2018a.

Keywords: Web mining, Data perturbation, Flip and Rotation Perturbation, Naïve Bayes classification, Fuzzy C-Means clustering.

1. Introduction

Big data analysis with necessary security and privacy preserving aspects is needed for maintaining a proper balance between utility and privacy and for improving the performance of data sharing, managing and controlling (Chamikara et al., 2018). The need for the privacy preservation techniques for web data increases day by day, in order to overcome the privacy conflicts to minimize privacy violations and to increase the usage of web applications without compromising their user's privacy (Wilson & Rosen, 2003). The performance of perturbed data is analyzed using classification accuracy, attack resilience and time consumption (Al-Rubaie & Chang, 2019). The conventional data perturbation techniques endeavoring to increase the accuracy of the perturbed data. Here the machine learning based approach is used for handling big data during data perturbation process (Chao et al., 2009). The large data process scheme uses cloud and IoT for enhancing the performance (Thanga Revathi & Ramaraj, 2017). The data perturbation technique is the process of masking individual data while maintaining confidentiality and underlying database relationships (Jahan et al., 2012). The data perturbation of privacy preserving approach uses data condensation, flip and rotation, micro aggregation, geometric perturbation, random and additive rotations to generate the renovated data (Jahan et al., 2018). Recently, the data mining has played a vital role in large scale data analysis; in various applications like image processing, data mining and weather forecasting the fuzzy based approach is applied for data pre-processing (Kargupta et al., 2005). Various privacy preserving data mining approaches are used for improving the privacy of the sensitive data and the data perturbation is one of the major techniques (Balasubramaniam & Kavitha, 2015). In this approach the sensitive attributes are perturbed with a geometric transformation algorithm of multiplicative data perturbation approach (Gokulnath et al., 2015). The k- anonymization algorithm can be applied to perform data perturbation and the cryptographic approaches also be used for the same purpose but they require a key as an extra parameter in order to sanitize the data and information (Zhou & Yu, 2015). In the field of information processing, data mining is an important approach for facilitating the decision-making process in various real-time applications and it also poses threats to privacy and security

https://doi.org/10.24846/v29i4y202004

aspects while accessing the data streams (Yan et al., 2019). The issues in data handling are reduced through privacy preserving approaches and the privacy preserving data streams are obtained by various techniques like data modification, rule hiding, perturbation and data sanitization (Lee et al., 2019).

By surveying the privacy preserving techniques, the data mining-based models are graphically illustrated in Figure 1.

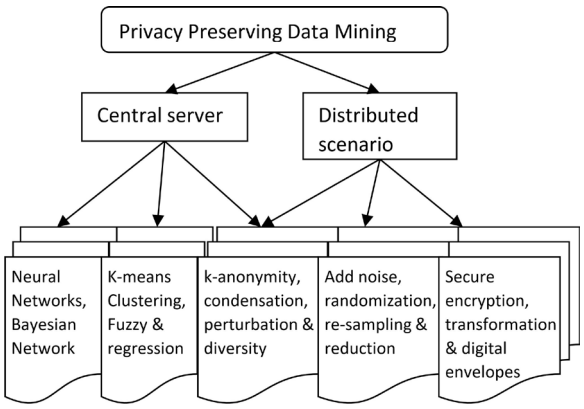

Figure 1. Various privacy preserving data mining techniques

The random rotation perturbation can also be used and it results in a poor model if the samples are not selected properly (Lin et al., 2019). The data within the window size is perturbed and stored and this process continues in order to perturb the entire data stream (Mopuri et al., 2019). Perturbed data is classified by various machine learning approaches and it is compared with the original data to evaluate the performance of the proposed system (Landgrebe & Duin, 2008; Du et al., 2018). In Internet of Things, the large numbers of data streams are traversed across the web and there is a huge demand to ensure the user's privacy (Adeel et al., 2019; Kiran & Vasumathi, 2019). The data privacy preservation and sharing is achieved in IoT with the help of data perturbation and cryptography processes (Chen et al., 2007; Tanuwidjaja et al., 2019). The cloud and IoT based applications produce large scale data streams and the data perturbation methods include unidimensional and multidimensional data renovation processes (Elidan et al., 2002; Chhinkaniwala et al., 2012). This renovation process uses random rotation, geometric perturbation, randomization, and swapping of the data streams for generating the perturbed data streams (Upadhyay et al., 2016). Random perturbation is presented in (Chen & Liu, 2005). Fuzzy based data privacy and perturbation technique is discussed in (Ahmed et al., 2014; Lyu et al., 2018). The aim of this approach is to transform original data stream into a renovated version of the perturbed data, which satisfies the privacy requirements; motive is to overcomes the privacy conflicts, minimizes the privacy violations and increases usage of web applications without compromising the user's privacy. Thus, the main objectives of this approach are achieving better accuracy, guaranteeing a certain level of privacy and maximizing the utility of data streams.

This paper is organized as follows. Section 2 provides a literature survey of related works, Section 3 presents the existing methodology and it also describes the various possible perturbation techniques. Section 4 describes the proposed methodology. Section 5 shows the experimental results and proposes some discussions about the proposed techniques. Finally, the section 6 concludes the present work and provides the perspective of a future work for enhancing the performance of the proposed approach.

2. Related Works

Chamikara et al. (2018) have performed the privacy preserving approach of data stream mining for improving the scalability. The main challenge of this work consists in collecting original data for data sanitization and also improving the security, privacy and utility level of sanitized data. The data perturbation process increases the speed of data streams and maintains the confidentiality of data. Here the rotation perturbation and data condensation techniques are applied. The performance analysis of time complexity, accuracy, scalability and privacy was measured and compared with existing approaches. This method applies several stages to perform data perturbation, which is reduced by the proposed flip and rotate methodology.

Wilson & Rosen (2003) have presented the perturbation technique with the impact of knowledge discovery scheme. Here it maintains

the confidentiality of data using security key. The objective of this work is to protect the data confidentially using various data perturbation techniques. The data mining tool for implementation and measurement uses QUEST(Quick, Unbiased, Efficient Statistical Tree) method. This generates the decision tree for classifying particular test cases. Here the standard deviation and correct classification accuracy of the perturbed data were measured. This method prevents the unauthorized access of sensitive data.

Chao et al. (2009) have applied the classification approaches in privacy preserving data perturbation model. Here the privacy preserving data stream mining is performed for getting perturbed information and it improves the confidentiality of the perturbed data using a cryptographical approach. This approach utilizes incremental mining for average weight evaluation in data perturbation. Online data mining system is employed for updating the real-time applications based on the data streams. The Data Splitting and Perturbation (DSP) algorithm reduces the data leakage by using the numeric attribute handling approach. The standard deviation is calculated to analyze the performance of the perturbed data.

Jahan et al. (2012) have proposed the data perturbation privacy preserving approach using the feature selection method to preserve the sensitive data in the database. Here the data perturbation technique is performed with Singular Value Decomposition (SVD) technique. The data analysis performs classification and clustering on the perturbed data. Here the feature selection algorithm contains feature search and feature subset evaluation phases and its efficiency is determined through the results of the classifier. Here Support Vector Machine Learning uses nonlinear mapping to transform the original data that are linearly separable into high-dimensional data. It finds hyperplane using support vectors and maximum margins.

Balasubramaniam & Kavitha (2015) have presented the geometric data perturbation technique used to perturb the Personal Health Records (PHR). Here the PHR data are stored on the cloud storage in a sanitized way and retrieved through access control devices. In this approach, the access control is performed between user and owner through the service. The perturbation time and encryption times are measured to evaluating the performance results. Here the security and privacy of PHR cloud data are improved and this approach can be applied for medical images and multimedia applications in cloud.

3. The Existing Method

With the reviews of above analysis, the data perturbation technique and machine learning based perturbed data classification is analyzed and the privacy preservation model is studied for improving the data mining process on the web.

3.1 Data Perturbation of Data Stream Mining

This approach was analyzed with various datasets such as Wine Quality, Page Blocks, Epileptic Seizure, Fried, Statlog, HEPMASS and Higgs. Based on the information of a dataset, the attribute level is measured and classified using the machine learning approach. This method is aimed to improve the results of the scalability and efficiency measures. It is used for perturbing high data streams, produced by IoT devices and other real time application's data sources. The clustering of multiple homogeneous sets is used for providing the different groups of features in resulting clusters. Here the fixed size of data chunk is applied for the performance analysis of the perturbation approach. The existing P^2RoCAI is the rotation-based condensation algorithm, which uses the umbrella concept. The K'-P^2RoCAI is a static data perturbation where K' represents the group size and grouping is done using k-means algorithm.

The stream data perturbation algorithm performs grouping based on the buffer size. A fixed k-value of the group representation increases the speed of performance in K-P^2RoCAI k-means stream method. This existing approach has two main compared results such as rotation perturbation and data condensation. The classifiers are decision tree, Naïve Bayes, random tree and J48. The existing method analyzes the data perturbation with the performance measures such as classification accuracy, attack resilience and time consumption. This approach uses k-Nearest Neighbour (kNN)

for testing the classification accuracy of the perturbed data and Friedman's rank test is used for measuring the security level against the privacy attacks. The 10-fold Crossvalidation is applied in the feature selection process in the classification approach. Here the experimental setup is done with MATLAB and Weka 3.6 for implementing and analyzing the performance of the system.

4. The Proposed Methodology

In this proposed methodology, the flip and rotation algorithm is used for data perturbation and the classification of perturbed data is done with the help of Naïve Bayes classifier. The k-means clustering algorithm is a subset of unsupervised machine learning approach and it handles the data based on groupings.

Initially, the raw database is normalized for performing clustering using fuzzy C-Means algorithm. Then flip and rotation technique is applied in order to perturb the data. This perturbed data is then classified using Naïve Bayes classifier. Here the training and testing data are framed by using holdout method. The classification accuracy and time consumption measures are obtained and compared with the ones of other existing approaches. The overall process flow of the proposed methodology is shown in Figure 2. The data perturbation results in the privacy preservation and improves the security of the data.

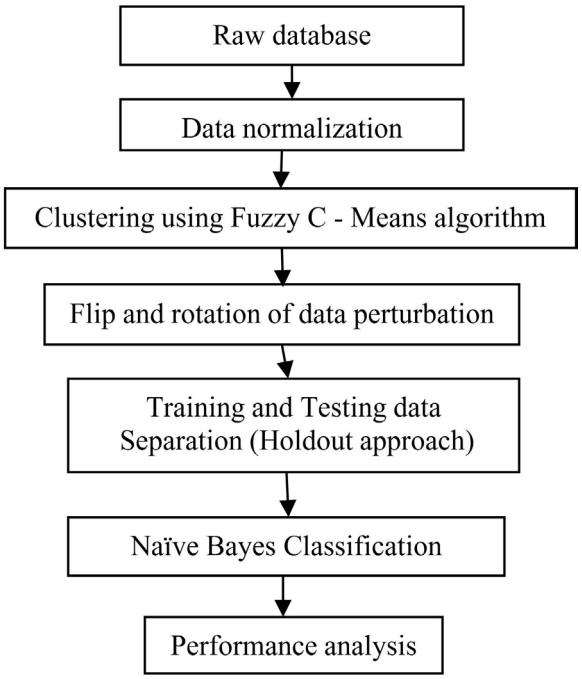

Figure 2. Overall process flow of the proposed methodology

Steps to perform data perturbation and classification:

i. Raw data selection

ii. Clustering using Fuzzy C-Means (FCM) algorithm

iii. Flip and Rotation Perturbation (FRP)

iv. Holdout approach for the separation of data into training and testing phases at 0.1%

v. Naïve Bayes (NB) Classification

vi. Measurement of performance analysis

4.1 Dataset Description

The following datasets are obtained from the existing system (Chamikara et al., 2018). The datasets are Wholesale Customers, Page Blocks and Epileptic Seizure. Each dataset has a different a set of attributes and different classes that are based on the value of attributes, termed as features. Here the data is protected by using the perturbation privacy preserving technique which doesn't require any key in order to achieve data perturbation. But cryptography techniques produce key overhead issues. Table 1 shows the dataset descriptions.

Table 1. Dataset

Dataset	Number of records	Number of attributes	Number of classes
Epileptic Seizure	11,500	178	5
Page Blocks	5473	10	5
Wholesale Customers	440	7	2

i. WCDS (Wholesale Customers Data Set)

Wholesale Customers database has 440 records, 7 attributes and 2 classes (UC Irvine Machine Learning Repository, 2014).

ii. PBDS (Page Blocks Data Set)

It contains 5473 records, 10 attributes and 5 classes (UC Irvine Machine Learning Repository, 1995).

iii. ESDS (Epileptic Seizure Data Set)

The Epileptic Seizure recognition data base contains 11500 records, 178 attributes and

5 classes (UC Irvine Machine Learning Repository, 2017).

4.2 Fuzzy C-Means (FCM) Clustering

The data perturbation approach is the value distorted approach and it transforms the original data into another form. Here the first phase of data perturbation is done by grouping the data using a fuzzy approach and this fuzzy logic is used to perform clustering on large volume data. It allows the grade membership for every data stream. Initially, it assigns the grade of membership values M. Then, the cluster centers are determined based on the following equation:

$$c_j = \frac{\sum_{i=1}^{N} M_{ij}^{\ m} \cdot x_i}{\sum_{i=1}^{N} M_{ij}^{\ m}} \quad (1)$$

where N is the number of clusters and x is the data in the cluster. Then the membership function is updated based on the cluster centers. The membership function is defined in equation 2.

$$M_{ij} = \frac{1}{\sum_{K=1}^{C} \left(\frac{\|x_i - c_{ij}\|}{\|x_i - c_{ij}\|} \right)^{\frac{2}{m-1}}} \quad (2)$$

Based on these two equations, the database is grouped into a certain number of classes. While clustering, the data is shuffled into different groups. By combining all the groups of data, the first phase perturbation results are obtained.This experiment is done for WCDS, PBDS and Epileptic Seizure recognition datasets. The clustering approach helps to perform data perturbation. The experimental result of clustering using FCM is illustrated in Figures 3-5.

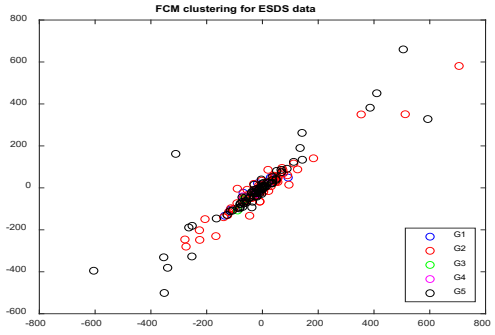

Figure 3. Result of clustering using Fuzzy C-Means (FCM) approach of ESDS

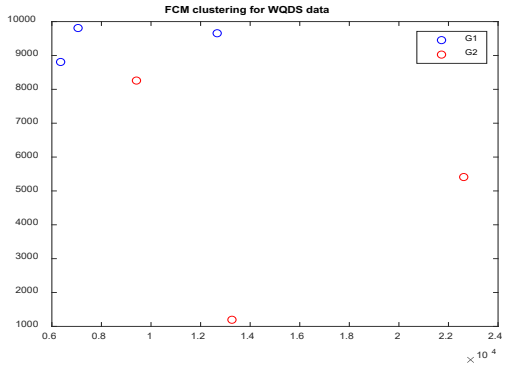

Figure 4. Result of clustering using Fuzzy C-Means (FCM) approach of WCDS

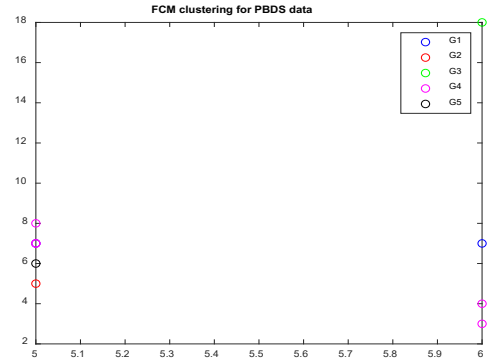

Figure 5. Result of clustering using Fuzzy C Means (FCM) approach of PBDS

4.3 Flip and Rotation Perturbation (FRP)

Generally, the data perturbation approach can be able to perform data partitioning, data modification, data restriction and data ownership rights preservation; in data modification, the data is modified in such a way that the privacy is preserved in the released dataset. This data perturbation performs data shuffling and data rotation processes is aimed to improve the results of privacy preserving approach.

The flip process in data perturbation is given by

$$flip\ perturb = \left(perturbed\ data\ row \right)^T \quad (3)$$

The rotation perturbation of clustered data is given as

$$rotate\ perturb = rot90\left(flip\ perturbed \right) \quad (4)$$

The FRP technique generates the perturbed data and maintains the trade-off between data utility and data privacy. ESDS contains EGG – Electroencephalogram signal values for

recognition and it identifies the seizure and no-seizure category by determining the class value of the classifier. Here the FCM clustering and FRP techniques are applied for perturbing the data and information. This approach preserves the user's information without disturbing the original content and the preserved data classification increases the utility of data without compromising the user's privacy. The input data is then shuffled and it is transformed into the perturbed data. The clustered data perturbation forms the matrix for grouping the data in the database. Then the perturbed data is classified by using Naïve Bayes classifier.

4.4 Naïve Bayes Classifier Using Holdout Approach

The data perturbation is the technique of shuffling the data using clustering and rotation-based transformation. The classification is performed on the perturbed data using Naïve Bayes classification by splitting the data in the training set and testing set by using the hold out method with 0.1 % for testing purpose. In Naïve Bayes classification, the classifier is designed based on the probability. The probability of attributes x comes under a specific category or a class C_k is determined by the following equation 5.

$$P(c_k / x) = \frac{P(c_k)P(x / c_k)}{P(x)} \quad (5)$$

The final classifier output from Naïve Bayes is computed based on the probability values obtained through the probability model.

$$y = argmax\, P(c_k)\prod_{i=1}^{n} P(x_i / c_k) \quad (6)$$

Classifying various groups of data creates the class value and the utility of the perturbed data is measured by using the group information. The proposed design of privacy preserving model uses three different datasets and they are classified by using Naïve Bayes classifier. Here the data perturbation technique is implemented by clustering and rotation- based perturbation approaches. In this proposed system, the input data is shuffled through the flipping process and produces the first phase perturbation results; then the rotation process is applied on it to generate the final perturbed data.

5. Results and Discussion

As shown above, the proposed design of data perturbation model is achieved with FRP technique and the classification of the perturbed data is carried out by using the Naïve Bayes classifier.

ESDS: This proposed approach comprises a new privacy preserving data mining algorithm. Here the Wholesale Customers, Page Blocks and Epileptic Seizure Data Sets are applied on the proposed perturbation and classification model. The classification accuracy of Epileptic Seizure Data Set (ESDS) is evaluated as illustrated in Figure 6. For large cluster k-values, higher than k = 5, the accuracy percentage is increased.

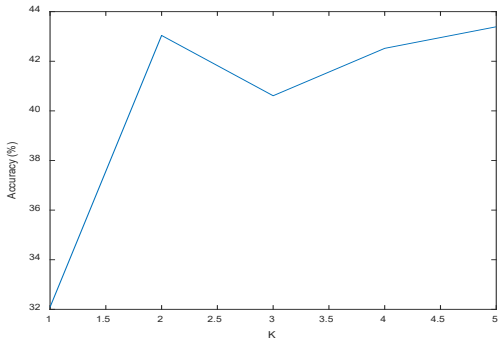

Figure 6. Classification accuracy of the perturbed data for large scale 'k' values – ESDS dataset

Based on the variation of attributes, the performance of ESDS and the time consumption are determined as shown in Figure 7.

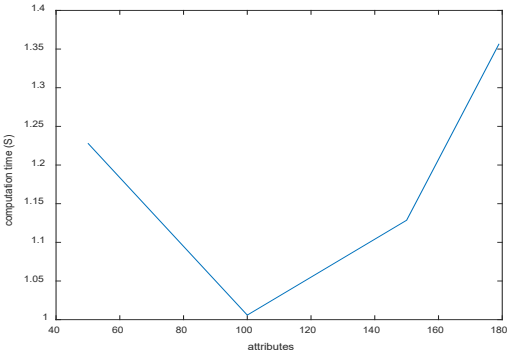

Figure 7. Time consumption versus number of attributes – ESDS dataset

The ESDS uses ElectroEncephaloGram (EEG) signal's attributes. Here the labeled classes are framed with five groups of instances. The EEG signal has 178 dimension vectors for five groups and the attributes represent the properties of EEG

brain signal – eyes open and closed, identifying the region of tissues, location of tumor and recording the seizure activity. The computation time is determined with tuples of ESDS and varies depending on the tuples of the dataset, as presented in Figure 8.

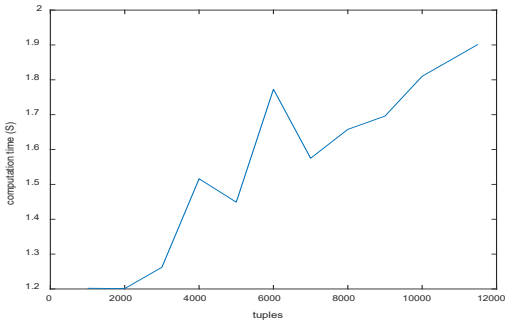

Figure 8. Time consumption versus number of tuples – ESDS dataset

WCDS: The WCDS has 8 attributes with 440 instances. These eight attributes are classified in two groups using Naïve Bayes classification on the perturbed data. The classification accuracy is verified for different number of clusters in order to determine the performance of the proposed method. The classification accuracy of WCDS is shown in Figure 9. Thus, it can be noticed that the classification outputs of Naïve Bayes approach give better results than the one of the existing approaches for different cluster sizes. Here 'k' value is taken up to five and it is used to determine the characteristics of attributes along with the class values. This experiment facilitates the process of web analytics and it is used for an effective decision-making process.

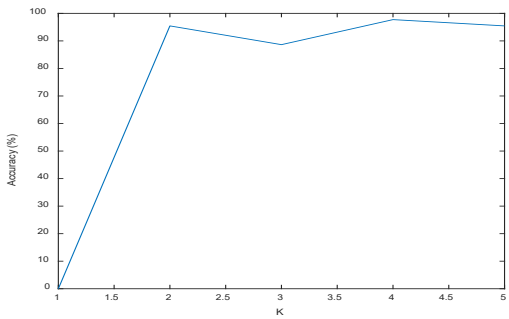

Figure 9. Classification accuracy of perturbed data for large scale 'k' values – WCDS dataset

The computation time of data perturbation and the classification of perturbed WCDS data are determined for analyzing the performance of the proposed system. The time consumption of WCDS based on attribute changes is shown in Figure 10 and the time consumption based on tuples variation of WCDS is given in Figure 11.

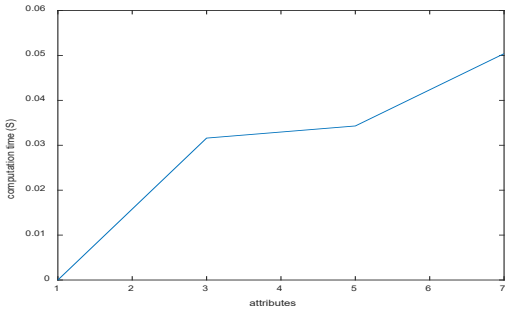

Figure 10. Time consumption versus number of attributes – WCDS dataset

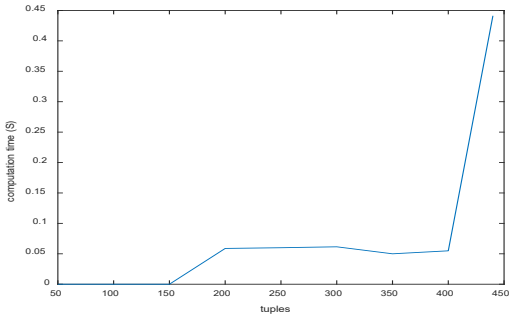

Figure 11. Time consumption versus number of tuples – WCDS dataset

PBDS: The PBDS has 10 attributes with 5473 instances. The classification accuracy of PBDS is shown in Figure 12.

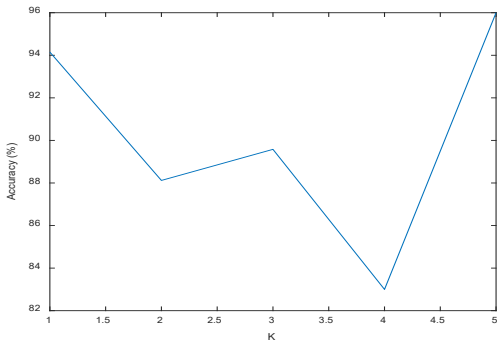

Figure 12. Classification accuracy of perturbed data for large scale 'k' values – PBDS dataset

Here 'k' value is taken up to five and for higher k values the accuracy is increased. The time consumption of PBDS based on attribute changes is shown in Figure 13 and the time consumption based on tuples variation of PBDS is given in Figure 14.

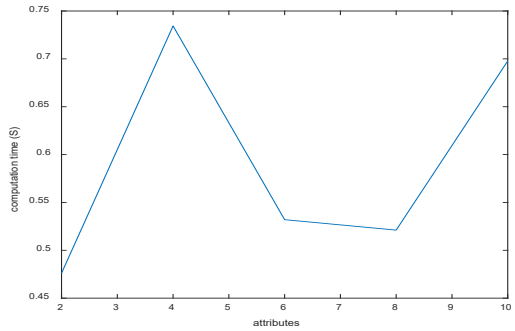

Figure 13. Time consumption while changing number of attributes – PBDS dataset

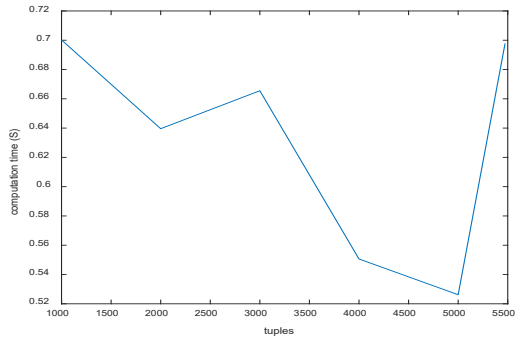

Figure 14. Time consumption while changing number of tuples – PBDS dataset

i. Performance metrics

The ESDS, PBDS and WCDS are perturbed through privacy preserving model using data perturbation and classification approaches. Here the proposed FRP improves the results when compared with the ones of the other existing works. In this approach, the classification accuracy and time consumption are determined for large 'k' values, attributes and tuples. The proposed Naïve Bayes classification results of ESDS,WCDS and PBDS are given in Table 2 and the data perturbation results of ESDS, WCDS and PBDS are shown in Tables 3, 4 and 5 along with the metrics accuracy, error rate and computation time.

Thus the rotation-based privacy preserving data classification using Naïve Bayes classifier is tested for ESDS, WCDS and PBDS datasets. Here the performance analysis of data perturbation technique and the classification using Naïve Bayes classifier have been achieved in order to measure the trade-off between the data utility and privacy. The classification accuracy of the perturbed data is determined for the proposed approach on different datasets and it is compared with other existing approaches (Chamikara et al., 2018). The error rate and computation time were also evaluated and compared, and they produced better results than the ones of the existing method.

Table 2. Performance results of Naïve Bayes Classifier

Number of attributes - ESDS	**Computation time**	**Accuracy**
50	1.2282	42.61
100	1.0059	42.96
150	1.1286	42.52
179	1.3568	43.39
Number of attributes - WCDS	**Computation time**	**Accuracy**
2	-	-
3	0.0316	68.18
5	0.0343	88.64
7	0.0504	95.45
Number of attributes - PBDS	**Computation time**	**Accuracy**
2	0.4761	87.93
4	0.7344	87.93
6	0.5320	91.92
8	0.5211	93.42
10	0.6978	96

Table 3. Wholesale Customers Data Analysis

WCDS	**Accuracy**	**Error rate**	**Computation time**
Original	64.54	35.46	-
Existing K_means	60.35	39.65	0.9744
Proposed	95.45	4.55	0.441

Table 4. Epileptic Seizure Data Analysis

ESDS	**Accuracy**	**Error rate**	**Computation time**
Original	43.57	56.43	-
Proposed	44.78	55.22	1.7341

Table 5. Page Blocks Data Analysis

PBDS	**Accuracy**	**Error rate**	**Computation time**
Original	90.85	9.15	-
Existing K_means	39.11	60.89	0.7462
Proposed	96.00	4.00	1.0661

ii. Attack resilience

The activity of hacking the original data from the perturbed data is called attack. The proposed method's resilience is evaluated by performing the three types of attacks namely naïve estimation, ICA based and known I/O attacks.

These attacks are applied on the perturbed data for 10 iterations at a noise factor of 0.3 using FastICA package. The attack resilience of the proposed method is evaluated using the minimum and average values obtained for each data set and it is used for identifying the overall resilience performance.

Using the information from Table 6, the rank is generated for the datasets through Fried Man Rank (FMR) test and it is compared with the one of the existing approach.

Table 6. Attack Analysis

Parameter	Nmin	Navg	Icamin	Icavg	Iomin	ioavg
PBDS	1.4915	1.7222	0.5696	0.7073	0.00002	0.1566
ESDS	1.1254	1.4386	0.5801	0.6875	0.00009	0.0017
WCDS	1.6764	1.7889	0.6853	0.7506	0.0177	0.1228

The below Table 7 shows the FMR values computed through Analysis of Variance (ANOVA) procedure.

Table 7. FMR comparison

System	FMR
Existing	0.03
Proposed	0.3366

From Table 7, it is observed that the FMR value for the proposed method is higher when compared to the one of the existing method. Generally, the higher FMR value indicates the best resilience performance. From the above experimental results it is proven that the proposed method has the best resilience against all the three types of attacks. Hence the proposed data perturbation approach's efficiency is proved by both the accuracy and the attack resilience metrics.

6. Conclusion and Future Scope

The proposed data perturbation and classification model is applied on the datasets ESDS, PBDS and WCDS and produces better results than the ones obtained by the existing approaches. This data perturbation technique can be applied on different datasets and used for web analytics. Here the Naïve Bayes classifier gives a better performance than other machine learning approaches. The FRP is easily achieved after performing FCM clustering algorithm and the perturbed data maintains the trade-off between privacy and accuracy. The proposed data perturbation approach uses data shuffling in order to preserve the privacy of the data. The implementation of the proposed system is done using MATLAB 2018a. The proposed system results have a better accuracy and less computation time on perturbed data, while comparing with the one of the existing approaches and improve the overall performance of the system. In future, the proposed method can be integrated with various machine learning approaches, data perturbation models and dimensionality reduction approaches using optimization algorithms for improving the classification and perturbation performance of the system.

REFERENCES

Adeel, M., Tejedor, J., Macias-Guarasa, J. & Lu, C. (2019). Improved perturbation in direct detected OTDR Systems using matched filtering, *IEEE Photonics Technology Letters, 31*(21), 1689-1692. DOI: 10.1109/LPT.2019.2940297

Ahmed, S. Md. T., Haque, S. & Tauhid, S. M. F. (2014). A Fuzzy based approach for privacy preserving clustering, *International Journal of Scientific & Engineering Research*, *5*(2), 1067-1071.

Al-Rubaie, M. & Chang, J. M. (2019). Privacy preserving Machine learning: Threats and Solutions, *IEEE Security & Privacy*, *17*(2), 49-58. DOI: 10.1109/MSEC.2018.2888775

Balasubramaniam, S. & Kavitha, V. (2015). Geometric Data Perturbation-Based Personal Health Record Transactions in Cloud Computing, *The Scientific World Journal*. Article ID 927867. DOI: 10.1155/2015/927867

Chamikara, M. A. P., Bertok, P., Liu, D., Camtepe, S. & Khalil, I. (2018). Efficient data perturbation for privacy preserving and accurate data stream mining, *Pervasive and Mobile Computing*, *48*, 1-19. DOI: 10.1016/j.pmcj.2018.05.003

Chao, C.-M., Chen, P.-Z. & Sun, C.-H. (2009). Privacy-preserving classification of data streams, *Tamkang Journal of Science and Engineering*, *12*(3), 321-330.

Chen, L., Goldgof, D. B., Hall, L. O. & Eschrich, S. A (2007). Noise-based feature perturbation as a selection method for micro array data. In *International Symposium on Bioinformatics Research and Applications* (pp. 237-247).

Chen, K. & Liu, L. (2005). A Random rotation perturbation approach to privacy preserving data classification, *Geometric Data Perturbation*. Georgia Institute of Technology.

Chhinkaniwala, H., Patel, K & Garg, S. (2012). Privacy preserving data stream classification using data perturbation technique. In *International Conference on Emerging Trends in Electrical Electronics and Communication Technologies*. DOI: 10.13140/2.1.1339.2964

Du, M., Wang, K., Xia, Z. & Zhang, Y. (2018). Differential privacy preserving of training model in wireless Big Data with edge computing, *IEEE Transactions on Big Data*, *6*(2), 283-295. DOI: 10.1109/TBDATA.2018.2829886

Elidan, G., Ninio, M., Friedman, N. & Schuurmans, D. (2002). Data perturbation for escaping local maxima in learning. In *Eighteenth National Conference on Artificial Intelligence* (132-139).

Gokulnath, C., Priyan, M. K., Balan, E. V., Rama Prabha, K. P. & Jeyanthi, R. (2015). Preservation of privacy in data mining by using PCA based perturbation technique. In *2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Chennai* (pp. 202-206). DOI: 10.1109/ICSTM.2015.7225414

Jahan, T., Narsimha, G. & Guru Rao, C. V. (2012). Data perturbation and feature selection in preserving privacy. In *Proceedings of the International Conference on Wireless and Optical Communications Networks*, 20-22 Sept. 2012, Indore, India (pp. 1-6). DOI: 10.1109/WOCN.2012.6335531

Jahan, T., Pavani, K., Narsimha, G. & Rao, C. (2018). A Data perturbation method to preserve privacy using fuzzy rules. In Bhateja, V., Tavares, J., Rani, B., Prasad, V. & Raju, K. (eds.), *Proceedings of the Second International Conference on Computational Intelligence and Informatics (ICCII 2017)*. *Advances in Intelligent Systems and Computing, vol 712* (pp. 9-16), Springer Singapore. DOI: 10.1007/978-981-10-8228-3_2

Kargupta, H., Datta, S., Wang, Q. & Sivakumar, K. (2005). Random data perturbation technique and privacy-preserving data mining, *Knowledge and Information systems, 7*(4), 387-414. DOI: 10.1007/s10115-004-0173-6

Kiran, A. & Vasumathi, D. (2019). Data mining: Random swapping based data perturbation technique for privacy preserving in data mining, *International Journal of Recent Technology and Engineering*, *8*(IS4), 764-777.

Landgrebe, T. C. W. & Duin, R. P. W. (2008). Efficient multiclass ROC Approximation by decomposition via confusion matrix perturbation analysis, *IEEE Transactions on Pattern analysis and Machine Intelligence*, *30*(5), 810-822.

Lee, T., Edwards, B., Molloy, I. & Su, D. (2019). Defending against neural network model stealing attacks using deceptive perturbations. In *IEEE Security and Privacy Workshop*, 19-23 May 2019, San Francisco, USA (pp. 43-49). DOI: 10.1109/SPW.2019.00020

Lin, Y., Qu, Y., Zhang, Z. & Su, H. (2019). Towards A Guided Perturbation for privacy protection through detecting adversarial examples with provable accuracy and precision. In *Proceedings of the International Conference on Computational Sciecne and Computational Intelligence*, 5-7 Dec. 2019, NV, USA (pp. 107-112). DOI: 10.1109/CSCI49370.2019.00025

Lyu, L., Bezdek, J. C., Law, Y. W., He, X. & Palaniswami, M. (2018). Privacy preserving collaborative fuzzy, *Data and Knowledge Engineering, 116,* 21-41. DOI: 10.1016/j.datak.2018.05.002

Mopuri, K. R., Ganeshan, A. & Babu, R. V. (2019). Generalizable Data-Free Objective for Crafting Universal Adversarial Perturbations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(10), 2452-2465. DOI: 10.1109/TPAMI.2018.2861800

Tanuwidjaja, H. C., Choi, R. & Kim, K. (2019). A Survey on Deep Learning techniques for privacy-preserving. In *Proceedings of the International Conference on Machine Learning for Cyber Security* (pp. 29-46).

Thanga Revathi, S. & Ramaraj, N. (2017). Data privacy preservation using data perturbation technique, *International Journal of Soft Computing and Artificial Intelligence*, *5*(2), 10-12.

UC Irvine Machine Learning Repository (1995). *Page Blocks Classification Data Set*. Available at: <https://archive.ics.uci.edu/ml/datasets/Page+Blocks+Classification>.

UC Irvine Machine Learning Repository (2014). *Wholesale customers Data Set*. Available at: <https://archive.ics.uci.edu/ml/datasets/Wholesale+customers>.

UC Irvine Machine Learning Repository (2017). *Epileptic Seizure Recognition Data Set*. Available at: <https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition>.

Upadhyay, S., Sharma, C., Sharma, P., Bharadwaj, P. & Seeja, K. R. (2016). Privacy preserving data mining with 3D rotation transformation, *Journal of King Saud University-Computer and Information Sciences*, *30*(4), 524-530.

Wilson, R. L & Rosen, P. A. (2003). Protecting data through 'Perturbation' techniques: The impact on knowledge discovery in databases, *Journal of Database Management*, *14*(2), 14-26. DOI: 10.4018/978-1-59140-471-2.ch003

Yan, K., Du, Y. & Ren, Z. (2019). MPPT Perturbation optimization of photovoltaic power systems based on solar irradiance data classification, *IEEE Transactions on Sustainable Energy, 10*(2), 514-521.

Zhou, Z.-H. & Yu, Y. (2015). Ensembling local learners through Multimodal perturbation, *IEEE Transactions on Systems, Man and Cybernetics-Part B*, *35*(4), 725-735.