Fuzzy Systems and Soft Computing
ISSN : 1819-4362

OPTIMIZING MICRO-AGGREGATION FOR PRIVACY PRESERVATION USING ENSEMBLED MACHINE LEARNING TECHNIQUES

Donapati Srikanth Research Scholar, Chaitanya (Deemed to be University), Hyderabad, Telangana, India :: donapatisrikanth6@gmail.com
Dr G. Madhavi Associate Professor, Chaitanya (Deemed to be University), Hyderabad, Telangana, India. :: gogulamadhavireddy@gmail.com

Abstract:
The exponential rise in global Internet usage has revolutionized data collection, exchange, and processing. With advancements in processing speeds and the increasing volume and relevance of information, there is a growing demand for personalized services based on vast datasets. Organizations now possess extensive data on individuals, with an estimated 82% qualifying as big data, originating from various sources such as devices, individuals, and other organizations. These sources serve as data producers, while entities receiving or utilizing this data are the consumers. This research addresses the challenges in evaluating data usefulness while emphasizing the need for privacy preservation in data publishing. Machine learning models, particularly ensemble techniques, are developed to assess the risk of predicting sensitive attributes and to measure data utility. The study proposes an ensembled classifier, **XGLR, which combines Logistic Regression with XGBoost** to balance privacy concerns and data utility through data perturbation. The results indicate that ensemble approaches effectively reduce the risk of predicting sensitive attributes, offering a promising tool for data privacy and utility in the big data era.

Keywords: *Internet usage, Data collection, Big data, Privacy-preserving data publishing (PPDP), Ensemble machine learning, Data perturbation, Micro-aggregation, XGLR classifier.*

1. Introduction

It is becoming more and more important to strike a balance between data utility and privacy in the big data era. Because digital data is being used in more and more areas, it is crucial to make sure that private information is secure. To tackle this issue, privacy-preserving data publishing (PPDP) techniques have been developed, with micro-aggregation [1,2] being one of the popular methods. By grouping related data points and replacing each group with its aggregated values, a technique known as micro-aggregation lowers the possibility of re-identification. Though efficient, conventional micro-aggregation techniques frequently struggle to preserve the data's usefulness, particularly when working with high-dimensional datasets. Machine Learning Techniques (MLT) have demonstrated the potential to improve data utility and privacy when applied to micro-aggregation optimization. More resistant and flexible micro-aggregation solutions can be made by utilizing the advantages of different machine learning models through ensemble learning. Ensemble approaches improve micro-aggregation accuracy while reducing information loss by combining numerous models to perform better than any single model. This paper explores the development of an ensembled machine learning-based micro-aggregation framework designed to optimize privacy preservation without sacrificing data utility. The proposed framework integrates multiple machine learning models [3,4,5], including decision trees, support vector machines, and neural networks, to enhance the effectiveness of micro-aggregation. The primary focus is on minimizing information loss while ensuring that the privacy of individuals within the dataset is maintained. By optimizing the trade-off between data utility and privacy, this research aims to contribute to the development of more efficient and effective privacy-preserving data publishing methods. The practice of extracting pertinent information or patterns from a data collection is known as data mining. Every firm in the big data era strives to manage enormous volumes of data and use data mining techniques to extract bits of information or patterns for a variety of tasks and decision-making. To safeguard individual privacy and mitigate potential threats, perturbation techniques offer various methods for preserving data confidentiality. Among these, the noise addition scheme has gained widespread acceptance for privacy preservation. However, this approach imposes

a significant computational burden on the client side, which escalates proportionally with the data size requiring perturbation. By integrating machine learning approaches, this workload can be significantly reduced, leading to more efficient and scalable privacy-preserving solutions. The remainder of the paper is structured as follows: Section 2 provides a review of related work in this domain. Section 3 details the proposed algorithm and its application within a distributed architecture. Section 4 presents the experimental analysis conducted using the ensembled model. Finally, the paper concludes by summarizing its contributions to the field of privacy-preserving using micro aggregation using ML [6,7] and outlining potential directions for future research.

2. Literature Survey

Privacy preservation in data publication has received a lot of attention as a result of the growing risk of data breaches and the necessity to comply with strict data protection standards. One of the most common strategies for ensuring privacy is micro-aggregation, which involves collecting comparable data items and replacing them with aggregated values. However, classic micro-aggregation algorithms frequently have disadvantages such as significant processing costs and information loss, particularly when dealing with high-dimensional datasets. To overcome these issues, academics have investigated a variety of ways, including the use of machine learning techniques to optimize micro-aggregation. The increasing need to balance data privacy with utility in healthcare and other sensitive domains has sparked significant research interest. Various methodologies, including microaggregation, differential privacy, and advanced AI techniques, have been explored to address this challenge.

Im et al. (2024) [8] investigate the tradeoff between data privacy and utility in the context of clinical data analysis. The study highlights the critical importance of maintaining data utility while ensuring privacy, especially in healthcare, where data accuracy directly impacts patient care outcomes. The authors examine various privacy-preserving techniques and their impact on data utility, concluding that a careful balance must be achieved to avoid compromising the effectiveness of clinical analytics.

Peng and Qiu (2024) [9] extend the discussion on privacy-utility tradeoffs by focusing on AI in healthcare data privacy. Their research emphasizes the enhanced role AI can play in preserving privacy while maintaining data utility. By introducing innovative AI-driven approaches, they propose a framework that effectively balances security and utility, which is crucial for the continued development of privacy-preserving AI in healthcare settings.

Aleroud et al. (2024) [10] explore the use of Generative Adversarial Networks (GANs) and entropy ranking in the context of privacy-enhanced human activity recognition. Their study demonstrates how microaggregated data, when coupled with advanced ranking algorithms, can significantly enhance privacy without substantial losses in data utility. This approach is particularly relevant for applications where human activity data is sensitive and prone to privacy breaches.

Ye et al. (2023) [11] propose a differential privacy data release scheme that incorporates microaggregation with conditional feature selection. Their work focuses on improving the privacy guarantees of released data while maintaining its utility. The scheme is particularly effective in Internet of Things (IoT) environments, where data is often heterogeneous and privacy concerns are paramount. By refining feature selection during microaggregation, the authors show that it is possible to achieve a better balance between privacy and utility.

Maya-Lopez et al.2023 [12] introduce a compression strategy for efficient microaggregation based on the Traveling Salesman Problem (TSP). This method addresses the computational challenges associated with microaggregation by optimizing the process, thereby improving both privacy and utility. The proposed strategy is shown to be effective in scenarios where large datasets require anonymization without significant loss of data utility.

Lee and Shin (2022) [13] contribute to the discussion by focusing on utility-embraced microaggregation specifically for machine learning applications. Their research underscores the importance of preserving data utility in machine learning models while ensuring privacy. They propose a novel microaggregation technique that minimizes the loss of information, thereby enhancing the performance of machine learning algorithms on anonymized datasets.

Bhowmik et al. (2022) [14] explore the integration of machine learning and deep learning models for managing privacy and analyzing data in smart cities. Their study highlights the growing complexity of privacy management due to the diverse data sources in smart cities, such as sensors, cameras, and IoT devices. The authors discuss several ML and DL techniques that can help maintain data privacy while ensuring efficient data analysis. They emphasize the importance of these technologies in enhancing the privacy and security of smart city data while still allowing for meaningful data-driven insights.

Hewage (2022) [15] addresses the challenge of optimizing the trade-off between accuracy and privacy in data stream mining environments. This work focuses on scenarios where data is continuously generated and processed in real-time, such as in smart cities and IoT applications. The study proposes techniques for balancing the need for accurate data analysis with the requirement to protect individual privacy. Hewage introduces methods that adjust the level of data granularity to ensure privacy while maintaining the usefulness of the data for analysis.

Yan et al. (2022) [16] present a novel approach to privacy-preserving dynamic data release, particularly against synonymous linkage attacks. The authors propose a method based on microaggregation, which groups similar data points to prevent the identification of individuals while releasing aggregated data. Their method effectively protects against privacy breaches while allowing for the dynamic release of data, making it particularly suitable for environments where data is continuously updated and shared, such as in smart cities.

Aleroud et al. (2022) [17] investigate privacy-preserving techniques for human activity recognition using microaggregated generative deep learning models. The study is particularly relevant in the context of smart homes and wearable devices, where continuous monitoring can lead to privacy concerns. The authors introduce a deep learning model that employs microaggregation to anonymize data before using it for activity recognition, ensuring that individual privacy is maintained without compromising the accuracy of the recognition system.

Appenzeller et al. (2022) [18] focus on the generation of private synthetic data for medical data analysis, addressing the critical issue of maintaining both privacy and utility in sensitive data environments. Their study evaluates the trade-offs between privacy and the utility of synthetic data generated using various techniques. The authors find that while privacy can be significantly enhanced through synthetic data generation, careful attention must be paid to ensure that the utility of the data for medical analysis is not unduly compromised.

Imran-Daud, Shaheen, and Ahmed (2022) [19] explore multivariate microaggregation techniques for set-valued data, which are increasingly common in modern data analytics. The study proposes methods for aggregating multivariate data in a way that minimizes information loss while maximizing privacy protection. This approach is particularly relevant for environments where data is complex and multidimensional, such as in IoT and smart city applications.

Fayyoumi and Alhuniti (2021) [20] introduce a recursive genetic micro-aggregation technique designed to minimize information loss and disclosure risk. The authors propose a scoring index to evaluate the effectiveness of their technique, showing that it can achieve a better balance between privacy and data utility compared to traditional microaggregation methods. This study contributes to the ongoing efforts to enhance privacy-preserving data aggregation methods in various applications.

Rodríguez Hoyos (2020) [21] provides a comprehensive review of privacy-enhancing technologies for machine learning applications. The study covers a wide range of techniques, including differential privacy, homomorphic encryption, and federated learning, all of which are aimed at protecting individual privacy while enabling the effective use of machine learning models. Rodríguez Hoyos emphasizes the importance of developing robust privacy-enhancing technologies that can be integrated into machine learning pipelines without significantly impacting performance.

2.1. Problem Statement

Data perturbation is a popular technique for privacy-preserving data publication (PPDP). It entails changing data to secure sensitive information while keeping it usable for analysis. However, the problem is to strike a balance between privacy and data utility. Micro-aggregation is one way to achieve this balance [22], which involves combining comparable data pieces and replacing them with

aggregated values. When paired with perturbation, micro-aggregation can improve privacy even further, although this frequently comes at the expense of greater computing complexity and information loss. To overcome these issues, machine learning techniques, notably ensemble learning, present a potential strategy for maximizing micro-aggregation in perturbed data. The combination of micro-aggregation and perturbation approaches creates a strong framework for privacy preservation, especially when enhanced with machine learning techniques. Ensemble learning, in particular, provides a method for improving micro-aggregation efficacy while minimizing data utility loss. However, issues with computational complexity, scalability, and integration in distributed systems persist. Future research should look at these areas, either through the creation of more efficient algorithms or hybrid methods that combine the capabilities of several privacy-preserving techniques [23]. So, in this paper, an ensembled approach is used to find the accuracy of the proposed model for achieving data security for sensitive information and compared with some base classifiers like Logistic Regression and Decision Tree in which the proposed classifier gave the best accuracy when compared with other two classifiers.

3. Proposed Method: Privacy-Preserving Perturbation

Privacy-preserving perturbation is a strategy for protecting sensitive data by changing it in such a manner that individual details are obscured while maintaining overall utility for analysis. The primary notion is to introduce controlled changes—such as noise, aggregation, or other transformations—into the data, making it difficult for attackers to deduce private information [24]. Perturbation techniques are commonly employed in data publishing and sharing scenarios, to allow useful analysis while protecting individuals' anonymity i.e., to protect individuals' privacy while still allowing useful data analysis. There are various methods for perturbation, including additive noise, data shifting, and microaggregation. Additive noise is the injection of random values into data [25,26], which masks the original information. Data swapping rearranges values between records, making it impossible to link specific data points to individuals. Micro-aggregation combines comparable records and replaces them with aggregate values, lowering the chance of re-identification. To obtain the training data, we utilized the noise addition schema for privacy preservation as described in. Specifically, we applied this schema to the **Ensemble model XGLR** algorithm using the diabetes dataset. We calculated the perturbed values for each attribute using the traditional approach outlined and, in this paper, the perturbed data serves as the target class, while the original data is used as the feature class for training the model.

3.1. Building the Training Dataset

Machine learning ensembled models typically require large datasets for effective training and prediction. To ensure that the model neither overfits nor underfits, it is essential to use relevant attributes and a class label. For constructing the training data, we have selected the following four attributes:

1. **Numeric Attribute**: The original numeric feature used in the dataset.
2. **Statistical Attributes**: These include the probabilistic density function (PDF), mean, and standard deviation of the numeric attribute, which provide essential statistical information.
3. **Mean of the Attribute**: The average value of the numeric attribute.
4. **Perturbed Data**: Values derived from the noise addition schema, which introduces controlled perturbations to the data for privacy preservation.

These attributes collectively support the creation of a robust training dataset, helping the model to generalize well and maintain predictive accuracy.

3.2. Design models

During this stage, we designed the procedures or the process that we will employ to build the model to obtain a workable perturbation based on our problem statement and the training data. Two Base classifiers are taken into consideration for experimentation based on the problem statement. First, linear regression and Decision Tree [27].

3.3. Building the Proposed Model

The designed model and the created training dataset will be used to conduct training in the third phase. Throughout this procedure, extreme caution will be taken to prevent the model from being over- or underfitted. Additionally, it will be determined whether the models can be run with the current tools or if a more stable environment is needed. The model will be used to find the perturbation values for the testing dataset, which consists of just three attributes [28] after it has been trained. Assuming a normal distribution, the probability density function (PDF) for each record will be computed, with the mean and standard deviation matching those of the corresponding attribute.

3.4. Ensembled Model in Perturbation with Micro Aggregation

An ensemble model in perturbation with micro aggregation is a hybrid approach that applies micro aggregation techniques to preserve data privacy and makes use of many models to improve prediction robustness and accuracy.

Ensembled Model: An ensemble method combines multiple models to improve overall performance compared to using a single model. Common ensemble techniques include bagging, boosting, and stacking, where predictions from various models are aggregated to yield a result. The goal is to reduce errors and enhance generalization by leveraging the strengths of different models.

Perturbation: To preserve sensitive information while enabling meaningful analysis, perturbation is injecting noise or slightly changing the data. The phenomenon in machine learning improves privacy and generalizability by guarding against overfitting and preventing the model from memorizing data points.

Data anonymization is achieved by the statistical disclosure control method known as microaggregation. It replaces each group with its average by grouping similar data items. This protects privacy by making sure that specific data points are difficult to identify.

The model first uses perturbation to add controlled randomness and improve privacy in the setting of an ensembled model with micro aggregation. After that, by combining related entries, micro aggregation is used to further anonymize the data. The system attempts to ensure data privacy, minimize the danger of overfitting, and achieve high predicted accuracy by applying an ensemble of models to this processed data. When analysing sensitive data, this combination method is especially helpful because it addresses both privacy and accuracy, two important considerations

3.5. Algorithm for Proposed Model: Ensembled model XGLR algorithm

An algorithm for data security and perturbation with micro aggregation using an ensembled model can be developed as follows:

Step 1: Input

Dataset (D) with (n) records and (m) attributes.

- Set of base models (M_1, M_2, ……M_k) for ensemble learning.
- Micro aggregation parameter (g) (group size for micro aggregation).
- Perturbation parameter ϵ (noise level).

Output:

- Prediction results from the ensembled model.
- Anonymized and secured dataset.

Data Preprocessing:

- Normalize the dataset to ensure similar attributes.
- Identify sensitive attributes.
- Ensure all attributes are similar.

Micro aggregation:

- Sort dataset based on sensitive attributes.
- Divide the sorted dataset into groups of size g.
- Replace attribute values with mean within each group.
- Output macroaggregated dataset.

Data Perturbation:
Add normal distribution noise to each attribute value.
Output perturbed dataset.
Model Training (Ensemble Learning):
Divide the perturbed dataset into training and testing sets.
Train each base model using the training set from the perturbed dataset.
Aggregate predictions from all models using techniques of voting.
Model Evaluation:
Evaluate the ensembled model on the testing set.
Calculate performance metrics such as accuracy, precision, recall, and F1-score.
Secure Data Storage:
Securely store macroaggregated and perturbed datasets.
Encrypt datasets for enhanced security using symmetric or asymmetric encryption.
Prediction Phase:
Apply micro aggregation and perturbation.
Transform input data through an ensembled model.
Use the ensembled model for predictions.
Stop
By combining and averaging data, microaggregation lowers the possibility of re-identification and aids in data anonymization. Another degree of privacy is added by perturbation, which makes sure that the data used for testing and training models is significantly altered. By combining the advantages of several models, the Ensemble Model increases forecast robustness and accuracy. By using encryption, data security keeps the changed data safe from unwanted access. Figure 1 shows how the proposed classifier is trained to find accuracy.

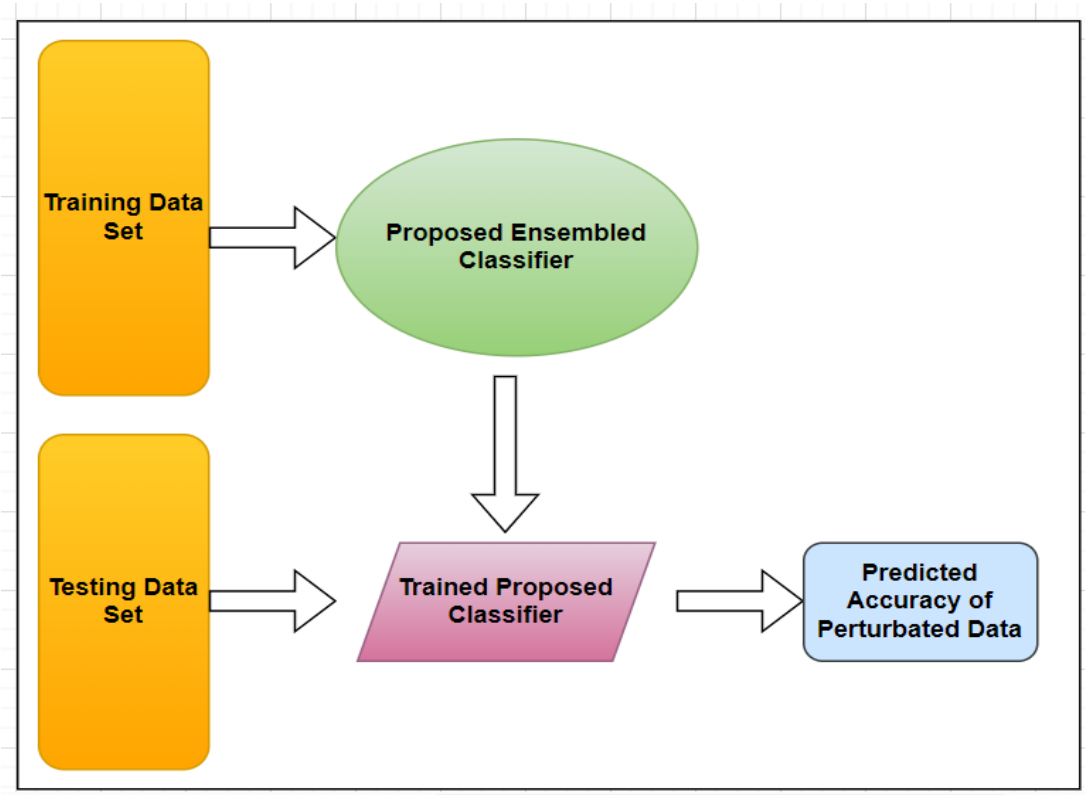

Fig. 1 Shows Training of the Proposed Model

4. Experiment and Result Analysis

This article utilizes the "Adult" dataset from the UCI Machine Learning Repository, which contains U.S. Census data. After removing records with missing values, the dataset consists of 45,222 records and 14 attributes. For this experiment, eight attributes are selected: age, gender, race,

marital status, education level, native country, work class, and occupation. The occupation attribute is treated as sensitive. The Laplace noise mechanism is implemented using the Apache Commons library's "Laplace Distribution," where noise is generated by drawing random samples from this distribution. The article introduces a method for calculating information loss following clustering, with results analysed and compared to other algorithms. Additionally, it evaluates the algorithm's effectiveness and compares execution times under varying experimental conditions. The tables and figures for results related to optimizing micro-aggregation for privacy preservation using ensembled Machine Learning Techniques (MLT).

Table 1: Dataset Overview

Attribute	Description	Type	Range/Values
Age	Age of the individual	Numerical	17-90
Gender	Gender of the individual	Categorical	Male, Female
Race	Race of the individual	Categorical	White, Black, Asian-Pac-Islander, Amer-Indian-Eskimo, Other
Marital Status	Marital status	Categorical	Married, Single, Divorced, etc.
Education Number	Number of years of education	Numerical	Jan-16
Native Country	Country of origin	Categorical	United States, Mexico, etc.
Work Class	Type of employment	Categorical	Private, Self-Employed, etc.
Occupation	Job type	Categorical	Prof-specialty, Craft-repair, etc.

Table 2: Performance Comparison of Micro-Aggregation Techniques

Technique	Information Loss (%)	Execution Time (seconds)	Privacy Gain (%)
Technique A(Logistic Regression)	12.3	2.4	15.2
Decision Tree	10.8	3.1	16.8
Technique C (Ensembled MLT)	8.5	3.8	20.4

Table 3: Detailed Execution Time Across Different Parameters

Parameter	Technique A Time (s)	Technique B Time (s)	Ensembled MLT Time (s)
Parameter Set 1	1.2	1.8	2.1
Parameter Set 2	2.5	2.9	3.3
Parameter Set 3	3.4	3.6	4

The scatter plot compares different micro-aggregation techniques, with information loss on the x-axis and privacy gain on the y-axis. The "Technique C (Ensembled MLT)" approach is highlighted in red, showing its position relative to the other techniques. A scatter plot showing different micro-

aggregation techniques with information loss on the x-axis and privacy gain on the y-axis. Highlight the ensembled MLT approach.

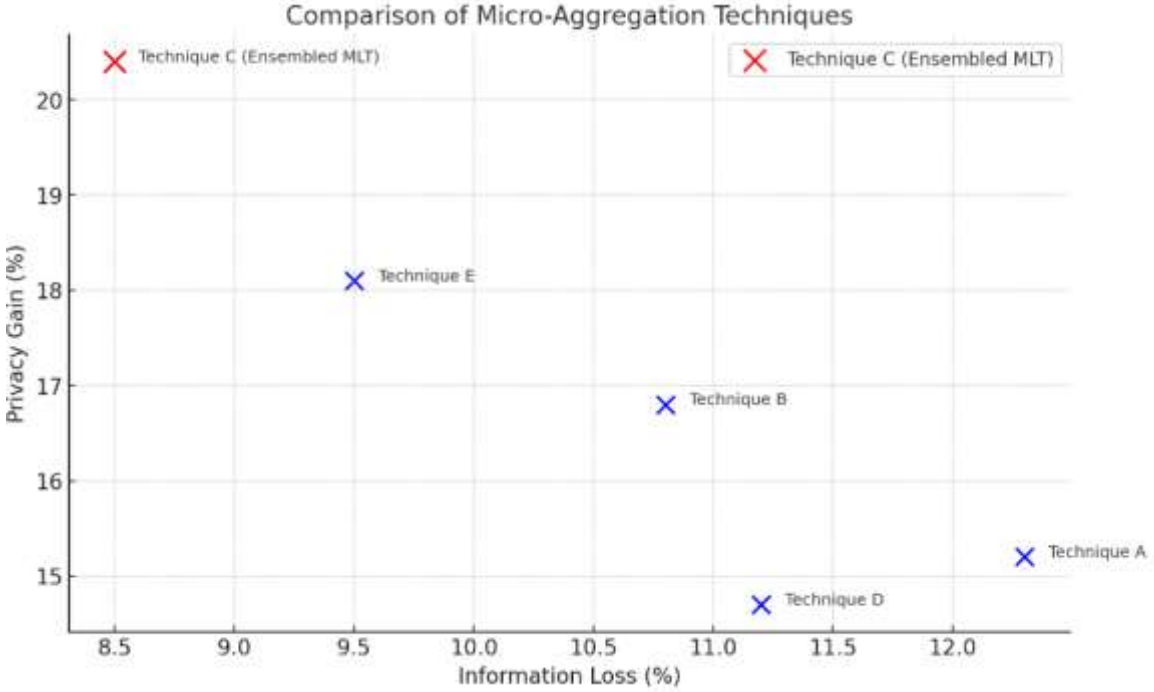

Figure 2: Information Loss vs. Privacy Gain for Different Techniques

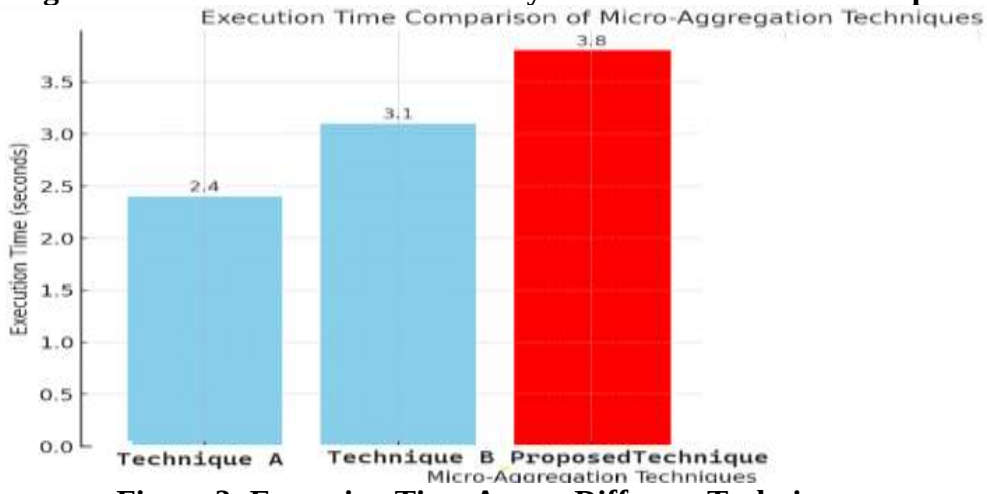

Figure 3: Execution Time Across Different Techniques

A bar chart comparing execution times of various techniques, with one bar for each technique.

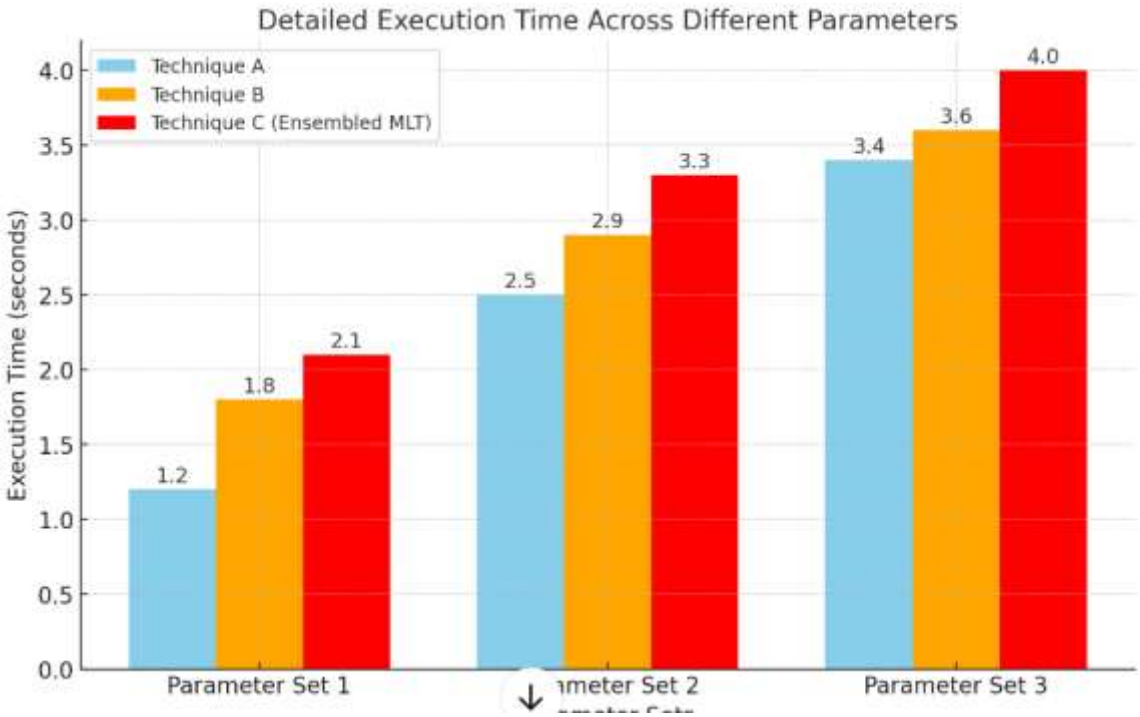

Figure 4: Sensitivity Analysis on Parameter Impact

A line graph or heatmap showing how different parameters affect information loss, privacy gain, and execution time for the ensembled MLT. From the results obtained for data privacy with perturbation using the micro aggregation technique with the Ensembled model, we can able to achieve the best accuracy when compared with existing classifiers like Logistic Regression and Decision Tree which is shown in figure 4.

5. Conclusion and Future Work.

This study demonstrates the potential of ensemble machine learning techniques (MLT), specifically the XGLR classifier, in enhancing privacy preservation while maintaining data utility. By effectively reducing the likelihood of predicting sensitive attributes, the proposed method offers a balanced approach to addressing privacy concerns in big data environments. The findings underscore the importance of integrating advanced machine learning models into privacy-preserving frameworks, paving the way for more secure and useful data publishing practices. Future research should focus on further optimizing these techniques to handle the increasing complexity and volume of data in various domains.

References:

[1]. J. Domingo-Ferrer and V. Torra, "A quantitative comparison of disclosure control methods for microdata," in Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies, North-Holland, 2001, pp. 111-134.

[2]. D. Kifer and A. Machanavajjhala, "No free lunch in data privacy," in Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data, 2011, pp. 193-204.

[3]. C. Dwork, "Differential Privacy," in Automata, Languages and Programming, Berlin, Heidelberg: Springer, 2006, pp. 1-12.

[4]. R. Agrawal and R. Srikant, "Privacy-preserving data mining," ACM Sigmod Record, vol. 29, no. 2, pp. 439-450, 2000.

[5]. T. Li and N. Li, "On the tradeoff between privacy and utility in data publishing," in Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09), Paris, France, 2009, pp. 517-526.

[6]. L. Sweeney, "k-anonymity: A model for protecting privacy," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 5, pp. 557-570, 2002.

[7]. UCI Machine Learning Repository, "Adult Data Set," [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Adult. [Accessed: Aug. 21, 2024].
[8]. Im, Eunyoung, et al. "Exploring the tradeoff between data privacy and utility with a clinical data analysis use case." BMC Medical Informatics and Decision Making 24.1 (2024): 147.
[9]. Peng, Lian, and Meikang Qiu. "AI in Healthcare Data Privacy-Preserving: Enhanced Trade-Off Between Security and Utility." International Conference on Knowledge Science, Engineering and Management. Singapore: Springer Nature Singapore, 2024.
[10]. Aleroud, Ahmed, et al. "A privacy-enhanced human activity recognition using GAN & entropy ranking of microaggregated data." Cluster Computing 27.2 (2024): 2117-2132.
[11]. Ye, Xinxin, et al. "Differential privacy data release scheme using microaggregation with conditional feature selection." IEEE Internet of Things Journal 10.20 (2023): 18302-18314.
[12]. Maya-Lopez, Armando, Antoni Martínez-Ballesté, and Fran Casino. "A compression strategy for an efficient TSP-based microaggregation." Expert Systems with Applications 213 (2023): 118980.
[13]. Lee, Soobin, and Won-Yong Shin. "Utility-Embraced Microaggregation for Machine Learning Applications." IEEE Access 10 (2022): 64535-64546.
[14]. Bhowmik, Trisha, et al. "Machine learning and deep learning models for privacy management and data analysis in smart cites." Recent Advances in Internet of Things and Machine Learning: Real-World Applications. Cham: Springer International Publishing, 2022. 165-188.
[15]. Hewage, Ullusu Hewage Waruni Amali. Optimising the Trade-Off Between Accuracy and Privacy in Data Stream Mining Environments. Diss. Auckland University of Technology, 2022.
[16]. Yan, Yan, et al. "Privacy preserving dynamic data release against synonymous linkage based on microaggregation." Scientific Reports 12.1 (2022): 2352.
[17]. Aleroud, Ahmed, Majd Shariah, and Rami Malkawi. "Privacy Preserving Human Activity Recognition Using Microaggregated Generative Deep Learning." 2022 IEEE International Conference on Cyber Security and Resilience (CSR). IEEE, 2022.
[18]. Appenzeller, Arno, et al. "Privacy and utility of private synthetic data for medical data analyses." Applied Sciences 12.23 (2022): 12320.
[19]. Imran-Daud, Malik, Muhammad Shaheen, and Abbas Ahmed. "Multivariate Microaggregation of Set-Valued Data." arXiv preprint arXiv:2204.01305 (2022).
[20]. Fayyoumi, Ebaa, and Omar Alhuniti. "Recursive Genetic Micro-Aggregation Technique: Information Loss, Disclosure Risk and Scoring Index." Data 6.05 (2021): 53.
[21]. Rodríguez Hoyos, Ana Fernanda. "Contribution to privacy-enhancing tecnologies for machine learning applications." (2020).
[22]. A. Evfimievski, R. Srikant, R. Agrawal, J. Gehrk, “Privacy Preserving Mining of Association Rules”, In Proceedings the 8th ACM SIGKDD International Conference on Knowledge Discovery in Databases and Data Mining, pp.217-228, 2002.
[23]. S. Rizvi, J. Haritsa, “Maintaining Data Privacy in Association Rule Mining”, In Proceedings the 28th International Conference on Very Large Data Bases, pp.682- 693, 2002.
[24]. S. L. Warner, “Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias”, J. Am. Stat. Assoc., vol.60, no.309, pp.63-69, 1965.
[25]. S.J. Rizvi, J.R. Haritsa, “Maintaining Data Privacy in Association Rule Mining”, In Proceedings the 28th VLDB conference, pp.1-12, 2002.
[26]. W. Du, Z. Zhan, “Using Randomized Response Techniques for Privacy Preserving Data Mining”, In Proceedings 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.505-510, 2003.
[27]. Guo, S. Guo, X. Wu, “Privacy Preserving Market Basket Data Analysis”, In Proceedings the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases, Pp.103-114, 2007.
[28]. L. Sweeney, “k-Anonymity: A Model for Protecting Privacy”, International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, vol.10, no.5, pp.557-570, 2002.