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

Evaluating the Effectiveness of Clustering-Based K-Anonymity and KNN Cluster for Privacy Preservation

Dhananjay M. Kanade[1], Prof. Dr. Shirish S. Sane[2]

Submitted: 27/05/2023 **Revised**: 15/07/2023 **Accepted**: 28/07/2023

Abstract: Due to the increasing number of data-driven innovations, privacy preservation has emerged as a paramount concern in the domain of data anonymity. Several techniques have been proposed to address this issue, and this paper aims to evaluate the three most popular ones. The study will look into clustering-based k-anonymity with KNN Cluster and K-Member with K=5 and 10. The importance of preserving personal information has become more apparent in the age of big data. Due to the increasing processing capabilities and the amount of data collected, the risk of unauthorized access and breaches has increased. This has prompted the need for effective data anonymization strategies. One of the most common methods of protecting personal information is by grouping similar people into clusters. This method ensures that each member of the group is indistinguishable from the others within the cluster. Another method is the KNN Cluster algorithm, which takes into account the proximity of the individuals to the feature. Finally, the K-Member algorithm is designed to identify the most representative of a given dataset. The paper aims to analyze and compare the three most popular methods for protecting personal information. We performed experiments with varying values of k, such as k=5 and k=10, to evaluate their privacy preservation effectiveness. The study is conducted on a scale of data utility, computational efficiency, and information loss. The results of the study will be analyzed and compared to provide a comprehensive understanding of the various limitations and strengths of each approach. This research will also help policymakers, data scientists, and data custodians make informed decisions when it comes to implementing anonymization strategies. The paper provides an in-depth evaluation of the clustering-based KNN Cluster, K-Member, and K-Anonymity techniques, focusing on their privacy effectiveness when protecting varying values of k. Its findings will help advance the field of privacy-enhancing mechanisms in the context of data-driven applications, and it will facilitate the creation of more robust and efficient methods

Keywords: *Data anonymization, Privacy preservation, Clustering-based k-Anonymity, KNN Cluster K-Member, Privacy-enhancing techniques.*

1. Introduction

Privacy has become an increasing concern as the amount of data collected and processed in the digital age continues to rise. The rise of digital systems and the interconnectedness of various technologies have resulted in the accumulation of vast amounts of sensitive and personal information[1], [2]. This data includes medical records, personal information, and behavioral and financial details. The availability of such data can pose a threat to people's privacy and could allow unauthorized access and misuse. A crucial technique that can address this issue is data anonymization. This process involves transforming and manipulating the data in such a manner that it becomes incredibly hard to identify the individuals in the dataset. The main objective of this process is to strike the right balance between the utility of the data and the protection of individuals' privacy[3].

Due to the increasing number of organizations and individuals demanding the protection of their personal information, data anonymization gained popularity. These organizations are bound by regulations and have to ensure that the security of their data is maintained. Also, since researchers and scientists often use sensitive data for modeling and analysis, it is important that these mechanisms are properly implemented. Various data anonymization techniques have been developed by practitioners and researchers to achieve the objectives of preserving the privacy of individuals while maximizing the data's utility[4], [5]. Different algorithms and methods can be used to achieve data anonymization, which provides a unique level of protection[6]. One of the most widely used methods is the k-anonymity algorithm, which ensures that the various individuals in a dataset can't be distinguished by the information that has been released.

The k-anonymity algorithm was first presented by Sweeney in 2002[7]. It is a method that involves grouping similar individuals in a way that makes them appear to be indistinguishable. This method can prevent the unauthorized access and use of sensitive information about individuals. The development of clustering-based methods for k-anonymity has gained widespread

[1]*ResearchScholar, Department of Computer Engineering, K. K. Wagh Institute of Engineering Education and Research, Nashik, Savitribai Phule Pune University, Pune, Maharashtra, India,dmkanade@kkwagh.edu.in*
[2] *Professor and Research Guide, Department of Computer Engineering, K. K. Wagh Institute of Engineering Education and Research, Nashik, Savitribai Phule Pune University, Pune,Maharashtra, India*

attention. These techniques are used to identify groups of individuals within a dataset using a clustering algorithm. The method can be used for anonymization because it allows users to group similar people together. It is a flexible and scalable method that can be used for protecting personal information.

Other methods that are related to clustering are also being proposed to enhance the anonymization process' effectiveness. One of these is KNN Cluster, which takes into account the proximity of the individuals in the feature space to form clusters. This method offers a variety of advantages, such as its ability to adapt to different data distributions. One of the most innovative approaches is the K-Member method, which takes into account the characteristics of the individuals in the dataset to select the most representative sample. This ensures that the released data maintains its privacy while also providing high utility.

The paper aims to analyze and compare the various anonymization techniques, namely the K-Member, K-Anonymity, and K-Cluster. It will also compare these techniques' performance in terms of preserving privacy and utility in diverse scenarios. The paper will look into the various aspects of anonymization techniques and their limitations, strengths, and trade-offs. We will evaluate their performance through various factors, such as computational efficiency, utility, and loss of information. Experiments will also be conducted on different datasets to evaluate the outputs of the different methods.

The findings of this study will contribute to the growing body of knowledge regarding the privacy preservation and anonymization of information. The results will act as a guide for policymakers, data custodians, and researchers in picking suitable methods based on their specific needs. In the future, this research will help develop efficient and effective methods for protecting privacy in big data. The study provides an extensive analysis of the clustering-based anonymization techniques known as K-Member, k-Anonymity, and K-NN Cluster. It will examine their respective performance in the context of anonymization. Doing so will help advance the field of privacy preservation and enable informed decision-making in choosing suitable methods.

2. Literature Review

Data anonymization is a critical technique used to protect privacy and mitigate the risks associated with the collection and analysis of personal and sensitive information. Numerous studies have been conducted to explore and evaluate different approaches and algorithms for achieving effective data anonymization. This literature review aims to provide a comprehensive analysis of the existing research in the field, focusing on the comparison and evaluation of anonymization techniques such as Clustering-based k-Anonymity, KNN Cluster, and K-Member.

One of the earliest works in the field of systematic clustering for k-anonymization was conducted by Kabir et al.[8]. They proposed an efficient systematic clustering method for achieving k-anonymity. The study highlighted the importance of preserving data utility while protecting privacy and demonstrated the effectiveness of their proposed approach through experiments on various datasets.

In a similar vein Sun et al.[9] introduced extended k-anonymity models to address the issue of sensitive attribute disclosure. Their work focused on enhancing the privacy preservation capabilities of k-anonymity by considering additional attributes and constraints. The study provided insights into the limitations of traditional k-anonymity and proposed extensions to improve privacy protection.

Loukides et al.[10] explored efficient and flexible anonymization techniques for transaction data. Their study aimed to achieve high levels of privacy while maintaining the usability and value of transaction data for analysis purposes. The research presented a framework for anonymization and emphasized the need for balancing privacy and data utility.

In the context of selecting suitable anonymization algorithms, Ben Fredj et al.[11] highlighted the importance of abstracting anonymization techniques. They argued that understanding the fundamental principles and characteristics of anonymization algorithms is essential for selecting appropriate generalization techniques. The study provided a comprehensive comparison of different anonymization techniques, enabling researchers and practitioners to make informed decisions.

Chen et al.[12] proposed a novel clustering-based anonymization technique for privacy protection in mobility social network services. Their research focused on improving privacy preservation while maintaining the quality of service in location-based social networks. The study demonstrated the effectiveness of the proposed method through extensive experiments and comparisons with existing techniques.

Privacy preservation in the healthcare domain is of utmost importance due to the sensitive nature of medical data. Lee et al.[13] presented a utility-preserving anonymization approach for health data publishing. Their work aimed to balance privacy preservation and data utility in healthcare datasets, enabling secure

sharing and analysis of medical information while protecting patient privacy.

Geographic partitioning for data anonymization was explored by Croft et al.[14]. They compared different approaches of geographic partitioning and evaluated their effectiveness in achieving privacy protection. The study focused on anonymization techniques applicable to geographic datasets, highlighting the challenges and trade-offs involved in preserving privacy.

Salas et al.[15] provided an overview of privacy techniques, anonymization methods, and their challenges in the era of big data. The research emphasized the need for effective privacy protection in the face of growing data volumes and advanced analytics techniques. The study discussed the strengths and weaknesses of existing privacy-preserving solutions, laying the foundation for further advancements.

Improving privacy preservation in collaborative filtering systems was addressed by Wei et al.[16]. Their research proposed enhancements to k-anonymity-based privacy preservation techniques in collaborative filtering, aiming to protect users' preferences while maintaining recommendation accuracy. The study provided insights into the challenges of privacy preservation in collaborative filtering and proposed solutions to enhance user privacy.

El Ouazzani et al.[17] introduced a new technique for privacy protection in big data, focusing on k-anonymity without a predefined threshold value of k. Their work aimed to address the challenge of determining an appropriate value of k for different datasets. The study provided a flexible and adaptive approach to achieving k-anonymity, allowing privacy preservation in diverse scenarios.

A comprehensive analysis of privacy-preserving solutions developed for online social networks was conducted by Majeed et al.[18]. The study reviewed various privacy techniques, such as k-anonymity and differential privacy, and evaluated their effectiveness in preserving privacy in the context of online social networks. The research highlighted the importance of privacy protection in the era of social media and identified key challenges and potential solutions.

Takaki et al.[19] focused on reasonable setting values for anonymization algorithms in the context of online educational data analysis support systems. Their research aimed to determine suitable parameter values for anonymization techniques to achieve an appropriate balance between privacy preservation and data utility. The study provided insights into the selection of parameters for effective privacy protection in educational data analysis.

Kiran et al.[20] proposed a k-anonymization approach for privacy preservation using data perturbation techniques in data mining. Their study focused on achieving privacy protection while maintaining data utility in the context of data mining applications. The research presented a novel technique for data perturbation, ensuring privacy preservation without significantly compromising data quality.

Caruccio et al.[21] developed a decision-support framework for data anonymization with applications to machine learning processes. Their work aimed to assist data custodians in selecting appropriate anonymization techniques based on the specific requirements of machine learning tasks. The research provided insights into the impact of anonymization on machine learning performance and introduced a framework for informed decision-making.

Neto et al.[22] focused on privacy preservation in multi-domain Internet of Things (IoT) environments. Their research explored anonymization techniques to enable privacy by anonymization in the collection of similar data from different IoT domains. The study emphasized the importance of privacy protection in IoT applications and introduced techniques to achieve anonymization in multi-domain scenarios.

In conclusion, the literature review demonstrates the extensive research conducted on data anonymization techniques such as Clustering-based k-Anonymity, KNN Cluster, and K-Member. The studies highlighted the challenges and trade-offs involved in achieving privacy preservation while maintaining data utility. They proposed innovative approaches, evaluated their effectiveness through experiments and comparisons, and addressed domain-specific requirements, such as healthcare data, social networks, educational data, and IoT environments. The findings and insights from these studies contribute to the advancement of privacy-preserving techniques and provide guidance for selecting suitable anonymization methods in various application domains.

3. Algorithms used for comparison

The k-anonymity method is a privacy-preserving algorithm that clusters individuals into a cluster to ensure that each member is indistinguishable from the others. This method combines the advantages of k-anonymity and clustering algorithms to maintain data utility while protecting the privacy of sensitive information[23], [24].

The initial steps in the clustering-based method of k-anonymity are shown below.

- **Cluster Formation**: The initial step in the clustering process is to divide the data into clusters using similarity measures. There are various types of clustering algorithms that can be used for this process, such as the k-means, hierarchical, and density-based. The choice of the algorithm depends on the type of data and its desired clustering quality.
- **Cluster Generalization**: Once a cluster has been formed, certain attributes of the group are generalized to ensure that no one within the cluster can be uniquely identified. Doing so involves replacing values of categories and ranges.
- **Anonymity Verification**: After the generalization process, the next step is to verify the k-anonymity achievement. This involves checking if the cluster's members share the same attributes. If not, then further clustering or generalization procedures are necessary.
- **Privacy and Utility Trade-off**: The tradeoff between protecting the privacy of a cluster and maintaining the data utility of a cluster is known as the utility trade-off. For instance, if a high level of generalization is implemented to achieve k-anonymization, it can increase the protection of the privacy but decrease the data utility. On the other hand, a low level of generalization is implemented to maintain the data utility.

3.1. KNN Cluster:

The KNN cluster is a variant of the k-Anonymity method that uses the K-nearest neighbor algorithm for establishing clusters. This is a widely used machine learning algorithm for performing regression and classification tasks. In the KNN Cluster framework, the algorithm determines the cluster's members based on their attributes. Follow the steps in the KNN cluster's algorithm to establish a cluster. It performs various tasks such as classification and regression[25], [26].

- **Calculate Similarity**: The first step in determining the similarity of a pair of individuals is to analyze the distance between them. This can be done by using a distance metric, such as the Manhattan distance or Euclidean distance. The value of the distance determines the dissimilarity between the two attributes.
- **Determine K**: The KNN cluster's next step is to identify the number of nearby neighbors. This can be done by calculating the value of K. The choice of the K value is dependent on the clustering quality and the dataset's characteristics. For instance, a larger K value can result in more clusters, while a lower one can result in smaller ones.
- **Find K Nearest Neighbors**: The KNN cluster algorithm finds the closest neighbors of each individual in the given set of data. It takes into account the individuals' attributes and distances.
- **Cluster Assignment**: When the cluster's K nearest neighbors have been identified, the group's members are randomly chosen based on their attributes. In the case of a tie, the group's members can be randomly assigned.

KNN Cluster iterates through the collected data continuously until all of its members are assigned to the cluster. Its objective is to create clusters composed of individuals with similar attributes in order to provide privacy protection. Eq.1 depict the KNN cluster

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \ldots. 1$$

where, x and y= "attribute value of two individuals", $x_i \; and \; y_i$ = "attribute values of the i^{th} attribute" and n = "total no. of attributes".

3.2. K-Member:

Another variation of the clustering method known as k-Member is focused on the members of clusters. This method assigns each individual to a group based on its similarities to the members of the cluster. Doing so ensures that the k-anonymities of the clusters are maintained. Follow the steps in the K-Member algorithm. It assigns each member of a cluster a group based on their similarity to the others in the cluster[27], [28].

- Calculate Similarity: The first step in implementing this method is to determine the similarity between the individuals in the cluster. This can be done by using various metrics such as the cosine, Jaccard, and Euclidean distance. The choice of the similarity metric should also be based on the dataset's characteristics and attributes.
- Determine K: The value of K, which is similar to that of KNN Cluster, needs to be determined to determine the similarity of a new member to the cluster's existing members.
- Find K Similar Members: The first set of similar members in the dataset is identified by the similarity metric that's chosen. These members act as a reference for the cluster assignment of new individuals.
- Cluster Assignment: Once the most similar members of a cluster have been identified, the new member is assigned to the group with the highest similarity. Doing so ensures that the new member will have a high level of similarity with its existing clustermates.

The K-member algorithm will repeat the steps until all clusters have been assigned. It aims to form clusters

where the members are similar to one another, which would result in k-anonymization.

Individuals are assigned clusters through the K-Member algorithm, which considers their similarity to the members of the group. Similarity can be measured by using various metrics, like the Jaccard coefficient.

The Jaccard similarity between *x* and *y*, calculates as in Eq.2

$$J(x,y) = \frac{|x \cap y|}{|x \cup y|} \dots 2$$

where, $x \cap y$ = "intersection of attribute values between individuals *x* and *y*", $x \cup y$ = "union of attribute values"

The formula divides the intersection of an attribute's value into its constituent elements and divides the resulting union between two people. This ratio specifies the degree of similarity between the two individuals, with values varying from zero to one being completely similar. The K-Member algorithm takes into account the similarity measure to determine the cluster's assignment of a new member. The algorithm assigns the new individual to the cluster with the highest similarity with its members.

4. Methodology

4.1. Load Dataset:

To begin the data anonymization process, the first step is to load the dataset. In this case, we will use the "Adult" dataset[29]. The Adult dataset contains information about individuals, including demographic attributes, employment details, and income levels.

After accessing the dataset, we can proceed to the next step.

4.2. Preprocessing:

In this step, we perform preprocessing tasks to ensure the data is in a suitable format for attribute selection and anonymization. Two effective methods for dataset cleaning are:

a. Missing Data Handling:

- One common issue in real-world datasets is the presence of missing values. Missing data can negatively impact the quality of the anonymization process. Two effective methods for handling missing data are:
- Removal of Instances: If the number of instances with missing values is relatively small compared to the overall dataset, removing those instances can be a reasonable approach.
- Imputation: Another method is to fill in the missing values with appropriate replacements, such as the mean, median, or mode of the attribute. Imputation helps retain more data for the anonymization process.

b. Outlier Detection and Treatment:

Outliers are data points that deviate significantly from the majority of the dataset. These data points can skew the anonymization process and affect the overall data quality. Outlier detection methods, such as the z-score or interquartile range (IQR), can be applied to identify and handle outliers.

4.3. Attribute Selection:

In the attribute selection phase, we identify and select the quasi-identifiers (QI) and categorical attributes from the dataset. Quasi-identifiers are attributes that can potentially lead to re-identification of individuals when combined with external knowledge. Examples of QI attributes in the Adult dataset may include age, education, occupation, and marital status.

Categorical attributes are variables that take on a limited number of distinct values. These attributes need special consideration during the anonymization process. By selecting the appropriate QI and categorical attributes, we can ensure that the anonymization techniques are applied to the relevant parts of the dataset.

Tree Generation:

Once the QI and categorical attributes are selected, a tree structure is generated based on the selected attributes. The tree structure helps in partitioning the dataset into clusters and identifying similar individuals within each cluster. Various algorithms, such as k-anonymity or l-diversity, can be applied on these clusters to achieve the desired level of privacy and data utility.

The tree generation process involves constructing a decision tree or a hierarchical clustering tree based on the selected attributes. The tree is built by recursively splitting the dataset into smaller subsets based on attribute values. This process helps identify groups of individuals who share similar attribute values within each branch or cluster of the tree.

The generated tree structure serves as a foundation for applying clustering-based anonymization techniques, such as K-Member or KNN Cluster, which can further enhance the privacy protection of the dataset. The data anonymization process begins with loading the dataset, followed by preprocessing steps to handle missing data and outliers. Attribute selection involves identifying the quasi-identifiers and categorical attributes. Finally, a tree structure is generated based on the selected attributes, forming the basis for applying clustering-based anonymization techniques.

5. Evaluation parameters

In the context of data anonymization, performance comparison plays a crucial role in evaluating the effectiveness and efficiency of different anonymization techniques. Here are some aspects of performance comparison that can be explored:

- **NCP Graph Comparison**:

NCP (Normalized Certainty Penalty) is a metric used to measure the level of information loss in anonymized data. It quantifies the degree to which the anonymized

data deviates from the original dataset. A comparison of NCP graphs allows for assessing the information loss incurred by different anonymization techniques. The graph can plot NCP values on the y-axis against the level of anonymity or generalization on the x-axis, showcasing how different techniques perform in terms of preserving data utility.

- **NCP Graph with Different Dataset Sizes:**

Another important aspect of performance comparison is evaluating the scalability of anonymization techniques with varying dataset sizes. By generating NCP graphs for different dataset sizes, it becomes possible to analyze the performance of different techniques in handling large datasets. This can provide insights into the scalability and efficiency of the techniques, allowing users to determine their suitability for different data volumes.

- **Time Comparison Graph:**

Time efficiency is a crucial factor in selecting an appropriate anonymization technique. A time comparison graph can display the execution time or processing time of different techniques on the y-axis against various parameters, such as the size of the dataset or the level of anonymity on the x-axis. This graph helps in identifying techniques that offer faster processing times, aiding in the selection of efficient anonymization methods for specific use cases.

- **Memory Utilization Graph:**

Memory utilization is another critical aspect to consider when comparing anonymization techniques. The graph can illustrate the memory consumption of different techniques against parameters such as dataset size or level of anonymity. By comparing memory utilization, organizations can make informed decisions about the scalability and resource requirements of different techniques, especially when dealing with large datasets.

These performance comparison graphs provide visual representations of the effectiveness, efficiency, scalability, and resource requirements of different anonymization techniques. They aid in selecting the most suitable technique based on the specific requirements, such as the desired level of anonymity, dataset size, processing time, and memory constraints. However, it is important to note that the actual results may vary depending on the specific implementation, hardware resources, and dataset characteristics.

6. Results and output

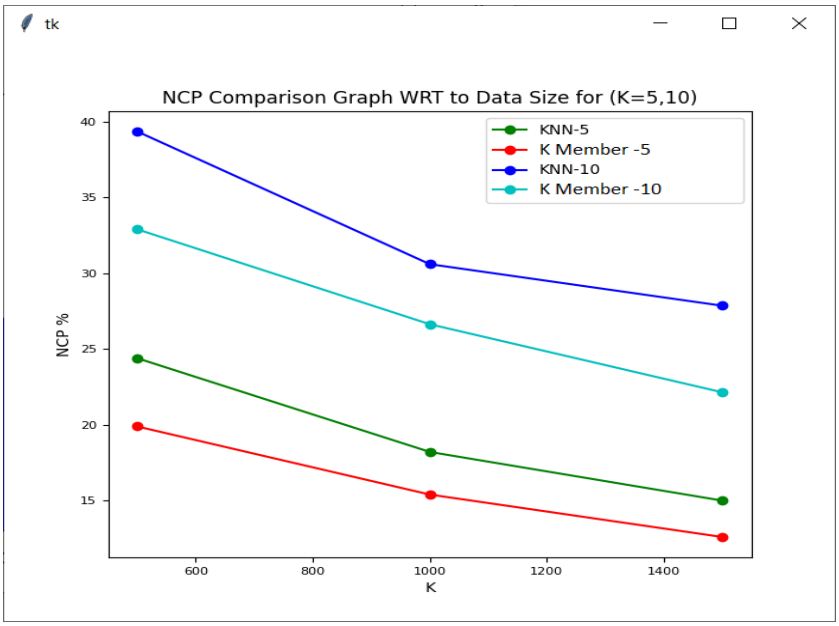

Fig 1 NCP comparison wrt to data size for (K = 5,10))

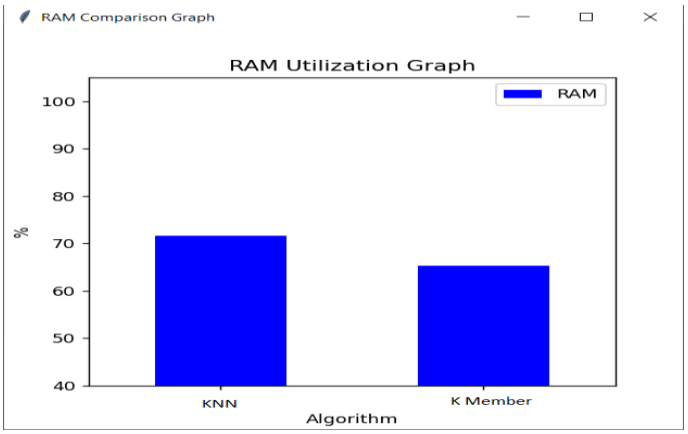

Fig 2 RAM utilization comparison

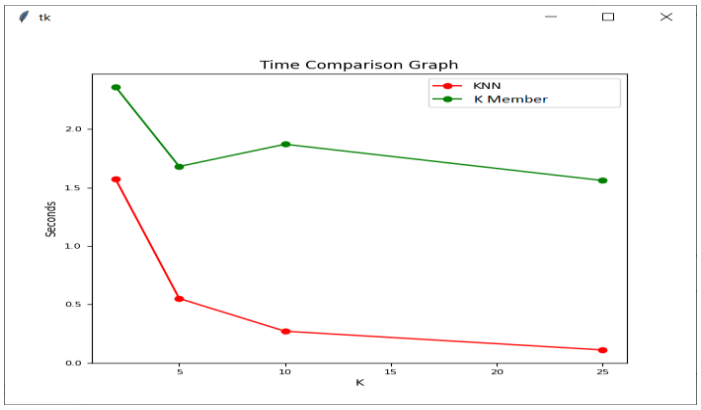

Fig 3 Time comparison graph

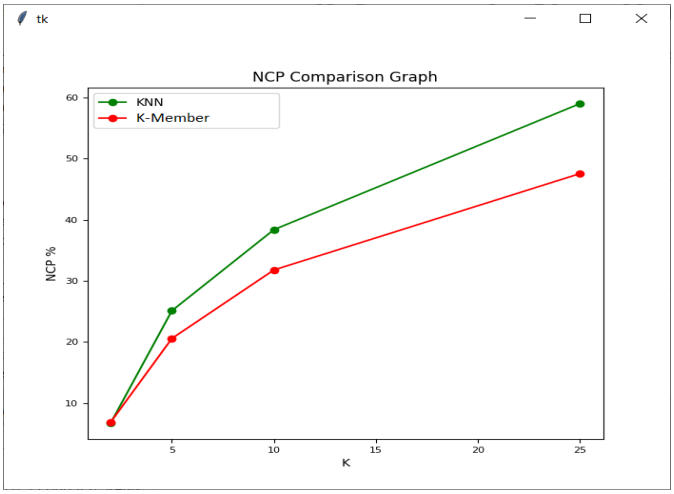

Fig 4 NCP comparison of KNN and K-member

The NCP graph compares the anonymization performance of different methods with different data sizes where k=5 and k=10. This study discusses the results as in figure-1,2,3,4.

- NCP% Decreases with Increasing Data Size: The graph shows how the NCP% decreases as the data grows. This indicates that, as organizations store more information, the level of loss in their anonymized records also decreases. This suggests that having larger datasets can result in better data utility and enable the grouping and generalization of similar records.
- Prominent Output of K-Member with k=5:The graph compares the performance of various methods with different data sizes. One of the most common techniques used is K-Member with a k=5. This indicates that it achieves a lower level of loss and provides better data utility than other methods. It suggests that having a small k value allows for effective anonymization while still maintaining privacy.

- RAM Utilization: The K-Member's efficiency when it comes to memory utilization is shown by comparing its performance with that of other methods. This means that it requires less resources when it comes to anonymization, which can be beneficial for large datasets.
- Time Comparison: The results of the time comparison reveal that KNN performs well when it comes to processing time. Compared to other methods, it has demonstrated faster execution, which indicates that it is suitable for applications that prioritize time efficiency.
- NCP% Comparison: The K-Member algorithm performs well in the NCP% comparison, indicating that it minimizes the amount of information loss while maintaining the privacy of the data. This suggests that it has a better balance between utility and privacy preservation.

The NCP comparison graph provides useful information on the anonymization performance of different methods with varying data sizes. Among the prominent techniques that stand out is K-Member with k=5, which balances the privacy and utility of the data. In addition, K-Member achieves an impressive time efficiency rate. These results can help the user choose the appropriate algorithm for their needs.

7. Conclusion and future scope

The discussions and comparisons that were conducted in the previous sections revealed various aspects of data anonymity. The evaluation metrics used, such as RAM utilization, time comparison, and NCP%, allowed us to gain a deeper understanding of the performance of different methods. The results indicated that the K-Member cluster with k=5 performed well in terms of data utility and privacy. It also exhibited efficient processing time and was suitable for time-sensitive applications. In addition, it was able to use efficient RAM utilization. In the future, data anonymization research will focus on developing new methods that can improve the efficiency of the process while preserving the privacy of individuals. Some of these include the development of advanced measures for protecting l-diversity, differential privacy, and t-closeness. Through in-depth evaluations and comparisons of different domains and datasets, we will be able to gain a deeper understanding of the limitations and effectiveness of these techniques. Integrating machine learning methods with data anonymization enables more efficient modeling and analysis processes. This discipline is focused on developing frameworks and algorithms that leverage the privacy-preserving capabilities of such data. Future research directions in the field of data anonymity will allow it to advance and provide enhanced privacy protection while facilitating secure data exchange and analysis in diverse domains. This will benefit both parties as it enables privacy-conscious decision-making and ensures the integrity of sensitive information.

References

[1] D. Slijepčević, M. Henzl, L. Daniel Klausner, T. Dam, P. Kieseberg, and M. Zeppelzauer, "k-Anonymity in practice: How generalisation and suppression affect machine learning classifiers," Comput. Secur., vol. 111, p. 102488, 2021, doi: 10.1016/j.cose.2021.102488.

[2] K. Wang, W. Zhao, J. Cui, Y. Cui, and J. Hu, "A K-anonymous clustering algorithm based on the analytic hierarchy process," J. Vis. Commun. Image Represent., vol. 59, pp. 76–83, 2019, doi: 10.1016/j.jvcir.2018.12.052.

[3] S. Chester, B. M. Kapron, G. Srivastava, and S. Venkatesh, "Complexity of social network anonymization," Soc. Netw. Anal. Min., vol. 3, no. 2, pp. 151–166, 2013, doi: 10.1007/s13278-012-0059-7.

[4] K. Guo and Q. Zhang, "Fast clustering-based anonymization approaches with time constraints for data streams," Knowledge-Based Syst., vol. 46, pp. 95–108, 2013, doi: 10.1016/j.knosys.2013.03.007.

[5] S. Fletcher and M. Z. Islam, "An anonymization technique using intersected decision trees," J. King Saud Univ. - Comput. Inf. Sci., vol. 27, no. 3, pp. 297–304, 2015, doi: 10.1016/j.jksuci.2014.06.015.

[6] V. Khetani, Y. Gandhi, S. Bhattacharya, S. N. Ajani, and S. Limkar, "INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Cross-Domain Analysis of ML and DL : Evaluating their Impact in Diverse Domains," vol. 11, pp. 253–262, 2023.

[7] L. Sweeney, "Achieving k-anonymity privacy protection using generalization and suppression," Int. J. Uncertainty, Fuzziness Knowlege-Based Syst., vol. 10, no. 5, pp. 571–588, 2002, doi: 10.1142/S021848850200165X.

[8] M. E. Kabir, H. Wang, and E. Bertino, "Efficient systematic clustering method for k-anonymization," Acta Inform., vol. 48, no. 1, pp. 51–66, 2011, doi: 10.1007/s00236-010-0131-6.

[9] X. Sun, L. Sun, and H. Wang, "Extended k-anonymity models against sensitive attribute disclosure," Comput. Commun., vol. 34, no. 4, pp. 526–535, 2011, doi: 10.1016/j.comcom.2010.03.020.

[10] G. Loukides, A. Gkoulalas-Divanis, and J. Shao, "Efficient and flexible anonymization of transaction data," Knowl. Inf. Syst., vol. 36, no. 1, pp. 153–210, 2013, doi: 10.1007/s10115-012-0544-3.

[11] F. Ben Fredj, N. Lammari, and I. Comyn-Wattiau, "Abstracting anonymization techniques: A prerequisite for selecting a generalization algorithm," Procedia Comput. Sci., vol. 60, no. 1, pp. 206–215, 2015, doi: 10.1016/j.procs.2015.08.120.

[12] Z. G. Chen, H. S. Kang, S. N. Yin, and S. R. Kim, "An efficient privacy protection in mobility social network services with novel clustering-based anonymization," Eurasip J. Wirel. Commun. Netw., vol. 2016, no. 1, pp. 1–9, 2016, doi: 10.1186/s13638-016-0767-1.

[13] H. Lee, S. Kim, J. W. Kim, and Y. D. Chung, "Utility-preserving anonymization for health data publishing," BMC Med. Inform. Decis. Mak., vol. 17, no. 1, pp. 1–12, 2017, doi: 10.1186/s12911-017-0499-0.

[14] W. L. Croft, W. Shi, J. R. Sack, and J. P. Corriveau, "Comparison of approaches of geographic partitioning for data anonymization," J. Geogr. Syst., vol. 19, no. 3, pp. 221–248, 2017, doi: 10.1007/s10109-017-0251-4.

[15] J. Salas and J. Domingo-Ferrer, "Some Basics on Privacy Techniques, Anonymization and their Big Data Challenges," Math. Comput. Sci., vol. 12, no. 3, pp. 263–274, 2018, doi: 10.1007/s11786-018-0344-6.

[16] R. Wei, H. Tian, and H. Shen, "Improving k-anonymity based privacy preservation for collaborative filtering," Comput. Electr. Eng., vol. 67, pp. 509–519, 2018, doi: 10.1016/j.compeleceng.2018.02.017.

[17] Z. El Ouazzani and H. El Bakkali, "A new technique ensuring privacy in big data: K-Anonymity without prior value of the threshold k," Procedia Comput. Sci., vol. 127, pp. 52–59, 2018, doi: 10.1016/j.procs.2018.01.097.

[18] A. Majeed, S. Khan, and S. O. Hwang, "A Comprehensive Analysis of Privacy-Preserving Solutions Developed for Online Social Networks," Electron., vol. 11, no. 13, 2022, doi: 10.3390/electronics11131931.

[19] O. Takaki, N. Hamamoto, A. Takefusa, S. Yokoyama, and K. Aida, "Reasonable Setting Values for Anonymization Algorithms for Online Educational Data Analysis Support System," Procedia Comput. Sci., vol. 207, no. Kes, pp. 2556–2566, 2022, doi: 10.1016/j.procs.2022.09.314.

[20] A. Kiran and N. Shirisha, "K-Anonymization approach for privacy preservation using data perturbation techniques in data mining," Mater. Today Proc., vol. 64, pp. 578–584, 2022, doi: 10.1016/j.matpr.2022.05.117.

[21] L. Caruccio, D. Desiato, G. Polese, G. Tortora, and N. Zannone, "A decision-support framework for data anonymization with application to machine learning processes," Inf. Sci. (Ny)., vol. 613, pp. 1–32, 2022, doi: 10.1016/j.ins.2022.09.004.

[22] R. C. J. Neto, P. Mérindol, and F. Theoleyre, "Enabling privacy by anonymization in the collection of similar data in multi-domain IoT," Comput. Commun., vol. 203, no. January 2022, pp. 60–76, 2023, doi: 10.1016/j.comcom.2023.02.022.

[23] [23] Y. T. Tsou et al., "(k, ε, δ)-Anonymization: privacy-preserving data release based on k-anonymity and differential privacy," Serv. Oriented Comput. Appl., vol. 15, no. 3, pp. 175–185, 2021, doi: 10.1007/s11761-021-00324-2.

[24] U. Sopaoglu and O. Abul, "Classification utility aware data stream anonymization," Appl. Soft Comput., vol. 110, p. 107743, 2021, doi: 10.1016/j.asoc.2021.107743.

[25] A. Girka, V. Terziyan, M. Gavriushenko, and A. Gontarenko, "Anonymization as homeomorphic data space transformation for privacy-preserving deep learning," Procedia Comput. Sci., vol. 180, pp. 867–876, 2021, doi: 10.1016/j.procs.2021.01.337.

[26] B. B. Mehta and U. P. Rao, "Improved l-diversity: Scalable anonymization approach for Privacy Preserving Big Data Publishing," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 4, pp. 1423–1430, 2022, doi: 10.1016/j.jksuci.2019.08.006.

[27] S. Chakraborty and B. K. Tripathy, "Alpha-anonymization techniques for privacy preservation in social networks," Soc. Netw. Anal. Min., vol. 6, no. 1, pp. 1–11, 2016, doi: 10.1007/s13278-016-0337-x.

[28] W. Y. Lin, D. C. Yang, and J. T. Wang, "Privacy preserving data anonymization of spontaneous ADE reporting system dataset," BMC Med. Inform. Decis. Mak., vol. 16, no. Suppl 1, 2016, doi: 10.1186/s12911-016-0293-4.

[29] "Adult - UCI Machine Learning Repository." .

[30] Nair, K. S. S. . (2023). Rapidly Convergent Series from Positive Term Series. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3), 79–86. https://doi.org/10.17762/ijritcc.v11i3.6204

[31] Kshirsagar, P. R., Yadav, R. K., Patil, N. N., & Makarand L, M. (2022). Intrusion Detection System Attack Detection and Classification Model with Feed-Forward LSTM Gate in Conventional Dataset. Machine Learning Applications in Engineering Education and Management, 2(1), 20–29. Retrieved from http://yashikajournals.com/index.php/mlaeem/article/view/21