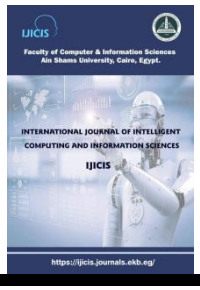
IJICIS, Vol.24, No.4, 43-59
DOI: 10.21608/ijicis.2025.332711.1361

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/



EVALUATING PRIVACY-LEVEL METRICS IN PRIVACY-PRESERVING DATA MINING

Saad A. Abdelhameed*

Planning Techniques Center,
Institute of National Planning,
Cairo, Egypt
saad.abd.elhameed@inp.edu.eg

Alshaimaa Abo-Alian

Information Systems Department,
Faculty of Computer and Information Sciences, Ain Shams University,
Cairo, Egypt
a_alian@cis.asu.edu.eg

Tamer Abdelkader

Faculty of Computer Science and Engineering, Galala University,
Suez, Egypt
Tamer.abdelkader@gu.edu.eg

Nagwa L. Badr

Information Systems Department,
Faculty of Computer and Information Sciences, Ain Shams University,
Cairo, Egypt
nagwabadr@cis.asu.edu.eg

Received 2024-10-31; Revised 2024-10-31; Accepted 2025-01-07

Abstract: *The increasing collection and analysis of sensitive personal data necessitates the development of robust Privacy-Preserving Data Mining (PPDM) methods. PPDM techniques are essential for extracting valuable insights from sensitive data while ensuring the maintenance of individuals' privacy. A critical aspect of implementing PPDM involves assessing the efficacy of these techniques in safeguarding privacy. However, despite the growing significance of PPDM, there remains a limited comprehensive understanding of the metrics used to evaluate their effectiveness, particularly concerning privacy preservation. This paper addresses this research gap by presenting an extensive study of privacy-level metrics for PPDM methods. The study examines data privacy metrics, which quantify the uncertainty faced by adversaries attempting to infer original sensitive data from transformed datasets. In addition, the paper analyzes results privacy metrics, which assess the risk of sensitive information disclosure from data mining outputs. Besides, the paper presents a new classification for privacy-level metrics based on the phase of PPDM processes in which they can be utilized. Moreover, the study provides a detailed analytical discussion of privacy-level metrics used in PPDM, examining their strengths and limitations while demonstrating their implications for practical applications. Furthermore, the paper highlights several considerations and challenges associated with measuring privacy within different PPDM methods in the absence of a universally accepted definition. By providing a comprehensive overview of existing privacy-level metrics, the proposed study establishes a vital foundation for the evaluation of PPDM methods and contributes to the advancement of responsible and trustworthy data mining practices.*

***Corresponding Author**: Saad A. Abdelhameed

Planning Techniques Center, Institute of National Planning, Cairo, Egypt

Email address: saad.abd.elhameed@inp.edu.eg

Keywords: *Privacy-Preserving Data Mining, Privacy-level metrics, Data privacy, Results privacy, Quantifying privacy.*

1. Introduction

In today's data-driven world, data mining has become a major source of essential approaches and tools for extracting valuable insights and knowledge from large datasets [1,2]. In addition, the discovery of valuable information from extensive datasets is crucial for advancement across various domains, including healthcare, banking and finance, and marketing [3-6]. However, this advancement must be carefully balanced with the critical need to protect individuals' privacy, particularly when handling sensitive information. Privacy-Preserving Data Mining (PPDM) methods have emerged as a vital solution to this challenge, enabling researchers and practitioners to identify hidden patterns and knowledge, facilitate data analysis, and protect sensitive information from being disclosed [7-9]. Besides, PPDM encompasses a wide range of methods that operate across different phases of the data lifecycle, from data collection and publishing to data mining output analysis [7]. In addition, these methods employ various techniques to mitigate privacy risks and restrict the disclosure of sensitive information during and after the data mining process. These include data randomization and perturbation, data anonymization, cryptographic techniques, association rule hiding, and query auditing [10-13]. The primary objective of these techniques is to transform the data in a way that prevents the disclosure of sensitive information while maintaining its usefulness for further analysis tasks [14-16].

Several PPDM algorithms have been proposed in the literature [10-13]. However, evaluating their effectiveness, which necessitates well-defined metrics capable of quantifying the level of privacy protection they offer, remains challenging. Therefore, a comprehensive understanding of privacy-level metrics is crucial for assessing the effectiveness of PPDM techniques and guiding the development of more robust privacy-preserving solutions. Besides, this understanding is vital to ensure that PPDM methods achieve an optimal balance between data utility and privacy preservation [17-18]. This paper focuses on a crucial aspect of PPDM methods' evaluation: privacy-level metrics. Privacy-level metrics are critical for assessing the effectiveness of PPDM techniques in safeguarding against the disclosure of sensitive information while preserving data utility. Furthermore, these metrics quantify the degree of uncertainty associated with predicting hidden sensitive information, with higher levels indicating stronger privacy protection. In the existing literature, privacy-level metrics are categorized into two primary types: data privacy metrics and results privacy metrics [11], [19]. Data privacy metrics focus on measuring the extent to which original data values are protected from unauthorized disclosure. Additionally, these metrics quantify the risk that an adversary may infer sensitive information about individuals' records from the transformed data. On the other hand, results privacy metrics concentrate on assessing the privacy of the aggregated knowledge extracted from the data, such as patterns, models, and rules. Moreover, these metrics evaluate the potential for an adversary to uncover sensitive insights or knowledge from the data mining results, even if the individuals' records are adequately protected. Despite the significance of privacy-level metrics, a thorough and systematic study of these metrics remains lacking in existing literature. Aside from that, some studies have addressed specific aspects of PPDM evaluation; however, a comprehensive and unified treatment of privacy metrics is still required. This paper aims to consider this gap by presenting an extensive evaluation study of privacy-level metrics for PPDM methods. This underscores the necessity of the current study, which aims to deliver a comprehensive analysis of privacy-level metrics and their applicability to various PPDM methods, thereby contributing to a more robust and standardized evaluation framework for PPDM.

The main contributions of this paper can be summarized as follows: (1) The paper addresses the existing research gap by introducing a comprehensive examination of present privacy-level metrics for PPDM methods, including data privacy and results privacy metrics. (2) We introduce a new classification for privacy-level metrics grounded in the specific phase of PPDM processes in which they are most applicable. (3) The study will go beyond just enumerating privacy-level metrics by providing an in-depth analysis of their strengths and limitations, and implications for practical applications. (4) The paper emphasizes the critical consideration and challenges of measuring privacy within PPDM methods, as well as the selection of appropriate privacy metrics. The rest of the paper is structured as follows: Section 2 introduces a foundational understanding of privacy in the context of data mining. Section 3 defines privacy-level metrics and explains how they are used to evaluate different PPDM techniques. Section 4 provides a detailed analysis of privacy-level metrics of the data collection phase. Section 5 presents an in-depth discussion of privacy-level metrics of the data publishing phase. Section 6 examines privacy-level metrics of the data mining phase. Section 7 analyzes and discusses various privacy-level metrics used in PPDM, examining their strengths, limitations, and implications for practical applications. Section 8 examines the key challenges and considerations in measuring privacy within PPDM methods. Finally, Section 9 concludes the paper.

2. Background: Privacy in Data Mining Context

Defining privacy presents inherent challenges due to its subjective nature, which relies on individual perceptions, cultural norms, and societal values [11], [20]. Although a single, universally accepted definition remains elusive, a common theme among the diverse interpretations emphasizes an individual's right to control the collection, storage, access, and utilization of their personal information [4], [19], [21]. However, in the context of data mining, privacy concerns emerge when analyzing datasets that contain sensitive personal information, which may result in the unintended disclosure of confidential details or the identification of individuals. Thus, the escalating volume and variety of collected data, combined with the advancing sophistication of data mining techniques, further intensify these concerns. PPDM methods seek to balance the advantages of data-driven knowledge discovery with the critical necessity of safeguarding individual privacy [19]. Therefore, these techniques are designed to extract valuable insights from vast datasets while concurrently ensuring that the disclosure or inference of private information is prevented during the mining processes and in the resultant outputs. PPDM methods categorize the attributes of input microdata tables into three distinct types: Explicit Identifiers (EIs), Quasi-Identifiers (QIDs), and Sensitive Attributes (SAs) [7], [22,23]. EIs are attributes that can directly identify individuals, thereby posing a significant risk to data privacy. QIDs are attributes that may identify individuals when combined or linked with other publicly available datasets. On the other hand, SAs represent attributes containing personal information that individuals typically wish to keep private and prefer not to have inferred by unauthorized parties. For example, consider a medical dataset containing the patient's records and having the following attributes (name, age, gender, zip-code, disease). In this dataset, the attribute (name) represents the EI, the attributes (age, gender, zip-code) represent the QIDs, and the attribute (disease) is the SA. Another instance is a dataset of a certain business organization containing the employees' records and having the following attributes (name, NID, age, gender, address, salary). In this dataset, the attributes (name, NID) are the EIs, the attributes (age, gender, address) are the QIDs, and the attribute (salary) is considered the SA. In addition, PPDM methods often modify or transform the original data to obscure sensitive information [7], [22,23]. However, these modifications can inadvertently reduce the data's utility for analysis, potentially leading to less accurate or less meaningful results. Hence, finding the optimal balance between privacy preservation and maintaining data utility represents a central challenge in PPDM.

Besides, privacy vulnerabilities exist at the various phases involved in the data lifecycle, from the initial collection, sharing, or publishing to conducting the desired data mining processes [7]. Consequently, PPDM methods are specifically designed to mitigate these vulnerabilities across these various stages. Various PPDM methods, such as randomization or perturbation, can be applied during the data collection phase to protect against privacy breaches, ensure secure transmission, and anonymize data at the centralized data source. In distributed data scenarios, where multiple parties collaborate to analyze data without fully sharing their individual datasets, privacy concerns emerge regarding the prevention of local information leakage during computation [10-12]. This necessitates the use of cryptographic protocols that enable joint computation on private data without disclosing the underlying data to other parties. These protocols encompass secure multiparty computation, homomorphic encryption, and secret sharing [7], [11], [24]. In the data publishing phase, when releasing datasets publicly or sharing them with third parties for analysis, it is essential to anonymize records to prevent the re-identification of individuals [2,3]. This process involves employing anonymization techniques such as k-anonymity, l-diversity, and t-closeness, which aim to prevent the disclosure of sensitive information by ensuring that individuals are indistinguishable within groups [2,3]. Even after implementing privacy-preserving techniques in earlier stages, the outputs of data mining algorithms may inadvertently reveal sensitive information [7], [10-13]. For example, certain association rules or classification models could unintentionally disclose private details about individuals or groups. Consequently, result-oriented PPDM methods are designed to address these vulnerabilities by modifying data mining algorithms or carefully analyzing and modifying outputs to prevent the disclosure of sensitive patterns. These methods include query auditing, downgrading classifier effectiveness, and association rule hiding [7], [11], [24].

3. Privacy-Level Metrics

Due to the absence of a standardized definition of privacy, quantifying the provided privacy-preserving level presents a challenge [11], [21]. Privacy-level metrics address this issue by offering a means to assess the effectiveness of data protection measures in relation to potential privacy breaches, particularly within the context of PPDM [19]. These metrics are essential for several reasons: (1) They facilitate the evaluation of PPDM techniques by allowing for the comparison of various methods and the selection of the most effective approach for a specific application and desired level of privacy. (2) They help quantify privacy risks by assessing the likelihood of various types of privacy breaches, such as identity disclosure or sensitive attribute disclosure. (3) They inform the design of newly proposed PPDM methods by identifying the strengths and limitations of existing techniques as highlighted by privacy metrics, thereby enabling researchers to develop more robust and effective methodologies. Aside from that, evaluating the effectiveness of PPDM techniques requires an accurate understanding of the various dimensions of privacy. A single metric is often insufficient to capture the complex trade-off between data utility and privacy protection. Consequently, a structured categorization of privacy-level metrics is essential to establishing a comprehensive evaluation framework. Thus, in the literature, privacy-level metrics are broadly categorized into two main types based on the aspect of the privacy-preserving process they measure: data privacy metrics and results privacy metrics [11], [19]. The primary objective of data privacy metrics is to assess the effectiveness of PPDM techniques in protecting the transformed individuals' records from any potential unauthorized disclosure. Besides, these metrics focus on evaluating the extent to which sensitive information can be inferred from the modified data. They measure how effectively the PPDM technique conceals sensitive information within the dataset after alterations for privacy protection and quantify the uncertainty or potential vulnerability of individual records following the application of specific privacy-preserving transformations. In contrast, results privacy metrics shift the focus to the outputs generated by applying data mining algorithms to the transformed data. Hence, the main objective

of these metrics is to evaluate whether the outputs inadvertently disclose sensitive information about individuals' records included in the original dataset. More specifically, these metrics aim to determine whether an attacker can leverage the data mining outputs, such as association rules or classification models, to infer sensitive information about the original data. Indeed, it is essential to consider both data privacy and result privacy metrics when evaluating the effectiveness of a PPDM technique. A technique may perform exceptionally well in safeguarding privacy at the data level; however, it may still be susceptible to privacy breaches when examining the information disclosed by the results of data mining. Evaluating the effectiveness of a PPDM method in relation to the provided privacy level should be considered across all phases, including data collection, data publishing, and the data mining phase. Consequently, various privacy-level metrics are employed in each phase to quantify the degree of privacy protections offered. Accordingly, we classify the privacy-level metrics examined based on the associated phase of PPDM in which they are most applicable. Figure 1 illustrates the new classification of privacy-level metrics of PPDM methods. In the following sections, we provide a comprehensive discussion of various privacy-level metrics, building upon the definitions and categories previously outlined.

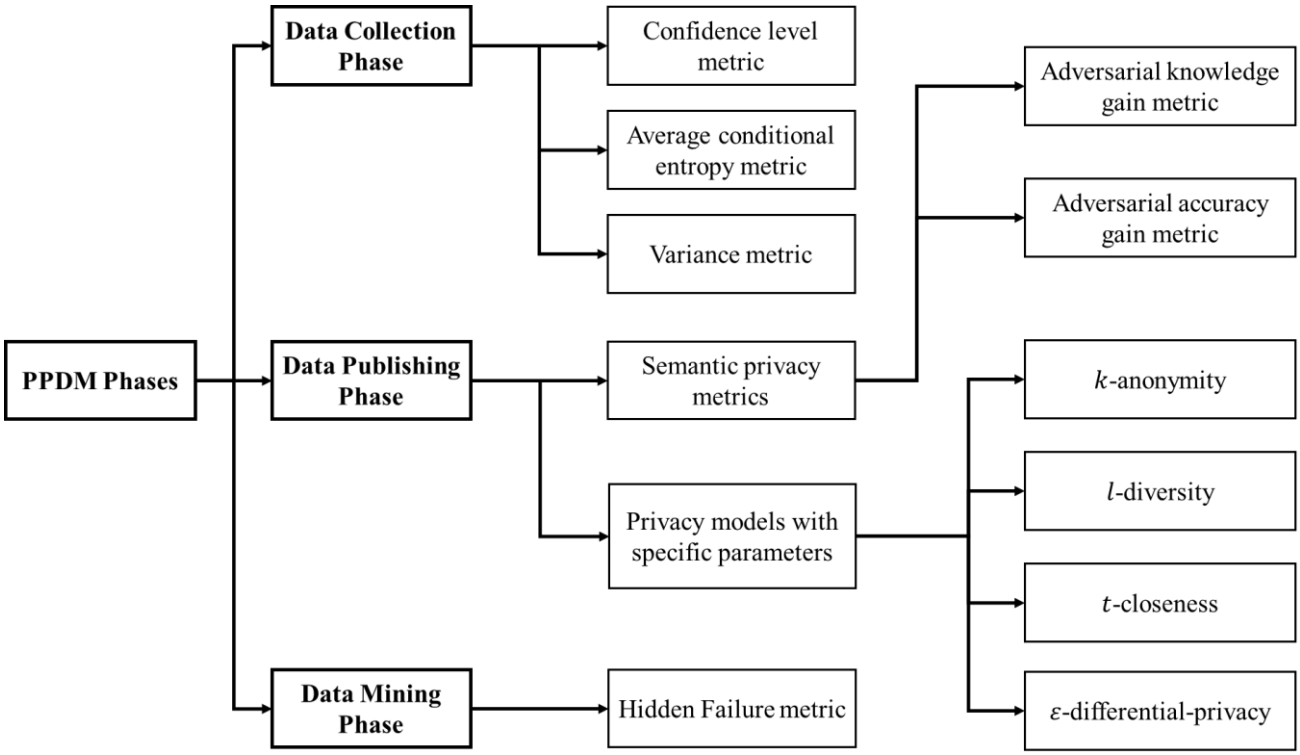


Figure. 1: The new classification of privacy-level metrics of PPDM methods.

4. Privacy-Level Metrics of the Data Collection Phase

4.1. Confidence level

The confidence level metric [25], commonly employed with additive-noise-based randomization techniques, evaluates the accuracy with which an attacker can estimate an individual's original data value from the randomized data. This metric quantifies the uncertainty an attacker encounters when attempting to identify the original value. A high confidence level signifies a lower degree of privacy protection, as it suggests an enhanced ability to approximate the original value. For instance, if an attacker can estimate

an original value to fall within the interval $[x_1, x_2]$ with 95% confidence, the interval $(x_2 - x_1)$ denotes the extent of privacy at that confidence level. However, a notable limitation of this metric is its inability to account for the distribution of the original data. As a result, an attacker may be able to refine the potential range of the original value to a narrower interval than $[x_1, x_2]$, while maintaining the same confidence level, by leveraging knowledge of the original data distribution.

4.2 Average Conditional Entropy

Average conditional entropy metric addresses the limitations of the confidence level metric by incorporating the original data distribution [26]. It utilizes the concept of information entropy to quantify the remaining uncertainty regarding the original data (A) in light of the transformed, perturbed data (B). Information entropy, a fundamental measure of uncertainty in information theory, is denoted by $H(A)$ and quantifies the average amount of information contained in a random variable A. In the context of PPDM, a higher average conditional entropy indicates greater uncertainty for an attacker attempting to infer the original values, thereby enhancing privacy preservation. Given a random variable A that represents the original data, along with the knowledge of another random variable B, which denotes the perturbed data, the average conditional entropy metric, denoted by $H(A|B)$, can be described as follows.

$$H(A|B) = 2^{h(a|b)} \tag{1}$$

Where $h(A|B)$ represents the conditional differential entropy of X, and can be defined as follows.

$$h(A|B) = -\int_{\Omega_{a,b}} f_{A,B}(a,b) \log_2 f_{A|B=b}(a)\, da\, db \tag{2}$$

Where $f_A(.)$ and $f_B(.)$ represent the density functions of A and B, respectively. Therefore, this metric evaluates potential information leakage from protected data to assess privacy. A higher conditional entropy indicates a greater difficulty for an attacker to infer the original values from the perturbed data. However, calculating the average conditional entropy necessitates an understanding of the perturbing distribution, which may not always be readily available or easily estimated, particularly in complex PPDM scenarios.

4.3 Variance

The variance metric [27], commonly employed with multiplicative noise randomization techniques, evaluates privacy by the variance between the original data value (A) and the respective perturbed data value (B). This metric can be expressed as follows.

$$Sec = \frac{Var\,(A - B)}{Var\,(A)} \tag{3}$$

Where Sec represents the maintained security level. This variance quantifies the extent to which the perturbed data deviates from the original data. A higher variance indicates greater difficulty in estimating the original values from the modified values. Consequently, a higher Sec value, indicating a larger variance difference, signifies a higher level of privacy protection. Additionally, this metric quantitatively measures the extent of distortion introduced by the applied randomization process. Therefore, a higher variance suggests a greater disparity between the original and perturbed data, thereby increasing the difficulty for an attacker to reverse-engineer the original values from the perturbed ones, indicating enhanced privacy.

5. Privacy-Level Metrics of the Data Publishing Phase

5.1 Semantic Privacy Metrics

The authors in [28] examined semantic metrics for measuring the privacy of anonymized databases through generalization and suppression methods commonly employed during the data publishing phase. The introduced analysis in the article is particularly relevant in the context of preventing sensitive attribute disclosure privacy attack. This attack occurs when an adversary obtains information regarding an individual's SAs from an anonymized database [3], [29-30]. Semantic privacy definitions are designed to quantify the adversary's knowledge gain about SAs derived from observing published anonymized databases. This methodology emphasizes the disparity between the adversary's baseline knowledge and the knowledge acquired from the anonymized data. Baseline knowledge refers to the minimum information about SAs that an adversary can discern following any form of sanitization process, even trivial sanitization in which all QIDs or SAs are removed. Conversely, posterior knowledge encompasses the information the adversary acquires about the SAs of a targeted individual from the sanitized table, considering the generalized and suppressed quasi-identifiers. The authors introduced formulas for two privacy metrics that evaluate the effectiveness of PPDM anonymization methods, utilizing generalization and suppression operations to safeguard sensitive information: adversarial knowledge gain (A_{know}) and adversarial accuracy gain (A_{acc}). Both metrics aim to quantify the extent of information that an adversary can obtain about sensitive attributes from a sanitized database. Furthermore, these metrics are designed to capture the potential for sensitive attribute disclosure, wherein an adversary exploits QIDs to infer sensitive information about individuals. We discuss each of the two metrics in the following subsections.

5.1.1 Adversarial Knowledge Gain Metric

This metric measures the average amount of information an adversary gains about SAs by identifying an individual's equivalence class within the sanitized database. It's based on the concept of attribute disclosure distance (A_{diff}), which quantifies the difference between the distribution of SAs within an equivalence class and the overall distribution in the entire table. The adversarial knowledge gain metric (A_{know}) is mathematically expressed as follows.

$$A_{know} = \frac{1}{|T|}\sum_{t\epsilon\varepsilon_Q} |\langle t\rangle|. A_{diff}(\langle t\rangle) \tag{4}$$

Where $|T|$ represents the total number of records in the original table, ε_Q donates the set of representative records for each quasi-identifier equivalence class, $\langle t\rangle$ represents the quasi-identifier equivalence class containing the sanitized record of individual t, $|\langle t\rangle|$ indicates the number of records within the equivalence class $\langle t\rangle$, and $A_{diff}(\langle t\rangle)$ is the attribute disclosure distance for the equivalence class $\langle t\rangle$. Besides, the mathematical formula of $A_{diff}(\langle t\rangle$ is given as:

$$A_{diff}(\langle t\rangle) = \frac{1}{2}\sum_{i=1}^{l} |p(T, s_i) - p(\langle t\rangle, s_i)| \tag{5}$$

Where l is the number of possible values for the sensitive attribute S, $p(T, s_i)$ denotes the probability of the ith sensitive attribute value s_i in the entire table T, and $p(\langle t\rangle, s_i)$ is the probability of the ith sensitive attribute value s_i within the equivalence class $\langle t\rangle$. Essentially, A_{know} metric calculates the weighted average of attribute disclosure distances for all equivalence classes in the anonymized table, with the weights determined by the size of each class relative to the total number of records. A lower value of the

adversarial knowledge gain metric (A_{know}) indicates better privacy, as it suggests the adversary gains less information from anonymized data.

5.1.2 Adversarial Accuracy Gain Metric

This metric quantifies the adversary's ability to predict a target individual's sensitive attribute by guessing the most common sensitive attribute within their equivalence class. It measures the increase in the adversary's prediction accuracy compared to a baseline scenario where they only have access to a trivially sanitized database. The adversarial accuracy gain (A_{acc}) metric is mathematically expressed as follows.

$$A_{acc} = (\frac{1}{|T|}\sum_{t\epsilon\varepsilon_Q} |\langle t\rangle|.\, p(\langle t\rangle, s_{max}(\langle t\rangle))) - p\big(T, s_{max}(T)\big) \quad (6)$$

Where $|T|$ is the total number of records in the original table, ε_Q denotes the set of representative records for each quasi-identifier equivalence class, $\langle t\rangle$ represents the quasi-identifier equivalence class containing the sanitized record of individual t, $|\langle t\rangle|$ indicates the number of records within the equivalence class $\langle t\rangle$, $s_{max}(\langle t\rangle)$ is the most common sensitive attribute value within the equivalence class $\langle t\rangle$, and $p\big(T, s_{max}(T)\big)$ is the probability of the most common sensitive attribute value within the equivalence class $\langle t\rangle$. Basically, the adversarial accuracy gain (A_{acc}) calculates the ratio between the adversary's average prediction accuracy when using the sanitized data and their baseline accuracy when using a trivially sanitized database, subtracting 1 to represent the gain. A lower A_{acc} value indicates better privacy, as it suggests the adversary's ability to predict sensitive attributes doesn't improve significantly after observing the anonymized data. In addition, it is important to note that this metric may underestimate the actual information leakage, as it exclusively considers the adversary's ability to predict the most common sensitive attribute within each equivalence class. Moreover, it does not account for potential shifts in the probabilities of other sensitive attribute values, which could still yield valuable information to the adversary.

5.2 Privacy Models with Specific Parameters

In addition to the previously mentioned metrics, various privacy-preserving models, such as k-anonymity [31], l-diversity [32], t-closeness [33], and ε-differential-privacy [34], utilize distinct inherent parameters as privacy metrics. These parameters (k, l, t, ε) do not represent specific formulas, but they serve as privacy models that facilitate privacy control by establishing respective parameter values. Besides, these parameters directly influence the level of privacy assured by the techniques, as they are predetermined to define the desired level of privacy protection. Generally, higher values of these parameters indicate stronger privacy protection; however, this may occur at the expense of data utility. For example, in k-anonymity, the parameter k specifies the minimum number of records in the dataset that must be indistinguishable from one another based on the generalized QIDs attributes. Therefore, a higher k value typically corresponds to enhanced privacy protection, but it may come at the cost of reduced data utility [3], [29]. Similarly, l-diversity, t-closeness, and ε-differential-privacy utilize their respective parameters to regulate the level of privacy protection. The l-diversity model employs the parameter l to quantify the diversity of SAs values within each one of the constructed equivalence classes [32]. Furthermore, t-closeness utilizes the parameter t to limit the difference between the distribution of a sensitive attribute in the anonymized data and its distribution in the original data [33]. Moreover, the differential-privacy model employs the privacy parameter ε to regulate the amount of noise added, thereby controlling the

level of privacy protection [34]. Thus, these metrics are specific to their corresponding techniques and provide a direct means of quantifying the privacy guarantees offered.

6. Privacy-Level Metrics of the Data Mining Phase

While data privacy metrics emphasize the protection of sensitive information within the transformed data itself, the results privacy metrics address the potential for privacy violations that may arise from the outcomes of data mining processes. Thus, results privacy metrics are designed to evaluate the degree to which sensitive information may be inferred or disclosed from the outputs generated by data mining algorithms, such as classifiers, association rules, or cluster analysis, applied to transformed data.

6.1 Hidden Failure Metric

In the context of pattern recognition data mining techniques, such as classification and association rule mining, the objective of various PPDM algorithms is to achieve zero hiding failure [19]. This means that all sensitive patterns are effectively hidden from unauthorized discovery. However, striving for absolute hiding failure may inadvertently compromise valuable, non-sensitive information. As more sensitive patterns are hidden, there is an increased risk of obscuring or distorting non-sensitive patterns that could be valuable for knowledge discovery. Consequently, the primary measure to evaluate is Hiding Failure (HF). HF can be quantified by calculating the percentage of sensitive information that remains discoverable following the data anonymization process. Ideally, this percentage should be as close to 0 as possible. One approach for assessing hiding failure is to compare the number of restrictive patterns discovered in the original database to those identified in the sanitized database. As such, the HF metric is widely recognized for this purpose [35]. The HF metric quantifies the balance between privacy and knowledge discovery, calculating the ratio of sensitive patterns successfully hidden by the PPDM algorithm to the total number of sensitive patterns in the original data [19]. While an HF of 0 indicates that all sensitive patterns are effectively hidden, this may come at the cost of losing non-sensitive information [35]. Accordingly, the HF metric can be described as follows.

$$HF = \frac{\#Rp(D')}{\#Rp(D)} \tag{7}$$

Where $\#Rp(D')$ represents the number of sensitive patterns found in the anonymized dataset (D'), and $\#Rp(D)$ represents the number of sensitive patterns present in the original dataset (D). A lower HF value indicates that the PPDM method is more effective at hiding sensitive patterns and preventing their disclosure through data mining results. However, low HF value does not necessarily imply that no sensitive information is leaked. It is essential to recognize that some non-sensitive patterns may also be hidden during the privacy preservation process.

7. Analysis and Discussion

PPDM has gained significant importance in the context of big data and growing privacy concerns. PDM emphasizes the protection of sensitive information within datasets during and after data analysis processes [7]. A critical aspect of PPDM is the assessment of privacy levels achieved through various techniques. Several metrics have been proposed to quantify the extent of privacy preservation achieved by these methods [19]. These privacy-level metrics serve as essential tools for evaluating the effectiveness of PPDM approaches, facilitating a balance between privacy and data utility. This section analyzes and discusses several privacy-level metrics used in PPDM as addressed in our study, examining their strengths, limitations, and implications for practical applications.

7.1 Privacy-Level Metrics of the Data Collection Phase

As a data privacy metric, the confidence level metric offers a straightforward approach to quantifying uncertainty in estimating original values from randomized data [25]. Its primary strength lies in its simplicity and ease of interpretation. However, this metric fails to consider the distribution of the original data, which can lead to inaccurate privacy assessments in certain scenarios. To address the limitations of the confidence level metric, researchers have introduced the average conditional entropy metric [26]. This metric integrates the original data distribution into the privacy calculation, offering a more nuanced assessment of privacy levels. However, it increased computational complexity and required knowledge of the joint density function of both the original and randomized data, which may not always be readily available. Furthermore, in the context of multiplicative noise randomization, variance metric serves as a straightforward measure of the extent to which the original data can be estimated from the perturbed data [27]. While this measure is effective for this specific technique, its applicability is limited to multiplicative noise randomization and may not be suitable for other privacy-preserving methods.

7.2 Privacy-Level Metrics of the Data Publishing Phase

As for semantic privacy metrics, the adversarial knowledge gain metric quantifies the average amount of information an adversary acquires regarding sensitive attributes across all individuals in a database [28]. Its strengths include its ability to capture average information gain, as well as its foundation in the concept of the attribute disclosure distance. This semantic grounding makes the metric more informative than purely syntactic metrics, such as k-anonymity. However, the calculation of adversarial knowledge gain can be computationally intensive for large datasets with numerous equivalence classes, and it may be sensitive to the specific distribution of sensitive attribute values within the dataset. In the same context, the adversarial accuracy gain metric directly measures the improvement in an adversary's ability to predict a target individual's sensitive attribute after observing the sanitized data [28]. Compared to the adversarial knowledge gain metric, its strengths include intuitive interpretation and computational efficiency. Nevertheless, the adversarial accuracy gain metric is primarily concentrated on the adversary's ability to predict the most common sensitive attribute value, potentially underestimating the actual privacy risks involved. Additionally, its scope is somewhat restricted, primarily representing the adversary's ability to perform a specific type of attack.

Regarding the privacy models with specific parameters, the k-anonymity metric introduced the concept of group-based anonymization [3]. This model complicates the process of linking records to individuals when at least $k - 1$ other records share identical quasi-identifiers. Its strengths include its straightforward concept and the availability of various implementation algorithms. However, k-anonymity is vulnerable to several types of privacy disclosure attacks [3] [29]. Furthermore, it does not address scenarios in which a single individual may have multiple records within the dataset. To address the limitations of k-anonymity, researchers have proposed the l-diversity model. This approach requires the presence of at least l "well-represented" values for SAs within each equivalence class. While l-diversity provides enhanced protection against certain privacy disclosure attacks, the determination of 'well-represented' values can be subjective and may vary depending on the specific instantiation of l-diversity. Moreover, it does not consider the distribution of sensitive attribute values, making it vulnerable to skewness attacks [3] [29]. In the same context, the t-closeness metric further enhances the concept of l-diversity by ensuring that the distribution of sensitive attributes within each equivalence class closely resembles the distribution in the original dataset. This approach effectively mitigates the skewness attack limitation associated with l-diversity. However, selecting an appropriate t value involves a trade-off between

privacy and data utility. In addition, a very small value of t may necessitate significant data generalization, which could reduce the dataset's utility for analysis. Besides, ε-differential-privacy model provides a robust and formal privacy guarantee, ensuring that the outcome of any analysis on the dataset is not significantly influenced by the presence or absence of a single record. This metric offers strong protection against a variety of potential attacks. However, selecting an appropriate ε value is critical and involves trade-offs between privacy and utility. A smaller ε enhances privacy but may substantially affect the accuracy of the analysis.

7.3 Privacy-Level Metrics of the Data Mining Phase

As a common results privacy metric, the HF metric provides a direct measurement of the privacy-utility trade-off in pattern recognition data mining techniques [35]. Its strength lies in quantifying the proportion of hidden sensitive patterns. However, the HF metric exclusively addresses this aspect and does not directly quantify the loss of non-sensitive information during the privacy-preserving process. In addition, a low HF value indicates a higher level of privacy protection, suggesting that a significant proportion of sensitive rules are effectively concealed within the anonymized dataset. Conversely, a high HF value implies that the PPDM technique may struggle to hide sensitive patterns, potentially exposing the data to privacy breaches. Therefore, an ideal PPDM technique should aim for a low HF value to enhance the protection of sensitive patterns. However, it is crucial to achieve a balance, as an overly aggressive approach to concealing sensitive information may inadvertently suppress valuable non-sensitive patterns, thereby reducing the data's overall utility for analysis. Consideration of the potential trade-off with data utility is essential, as excessively hiding patterns could render the data less useful for analytical purposes. Furthermore, this metric is particularly relevant for rule-based data mining techniques, such as association rule mining, where the objective is to discover relationships between variables. Table 1 summarizes a comparison between various privacy-level metrics.

In summary, data privacy metrics, such as confidence level and average conditional entropy, provide simplicity but may not fully capture the complexity of privacy risks in modern datasets. In contrast, semantic privacy metrics, including adversarial knowledge gain and adversarial accuracy gain, offer more precise assessments of privacy risks, although they can be computationally intensive or limited in scope. However, utilizing both metrics in conjunction within their relevant context can provide a more balanced and insightful understanding of the privacy implications associated with various anonymization techniques. On the other hand, anonymization model metrics such as k-anonymity, l-diversity, and t-closeness deliver increasingly sophisticated protection against specific types of attacks; however, they may be computationally expensive and still possess privacy vulnerabilities. Additionally, metrics like ε-differential privacy offer strong theoretical guarantees but present challenges in practical implementation and parameter tuning.

The analysis of these privacy metrics indicates that no single metric is universally superior for all PPDM applications or comprehensively captures all aspects of privacy preservation. Each metric has its own strengths and limitations, making it essential for researchers and practitioners to carefully evaluate the specific requirements and constraints of their data mining tasks when selecting appropriate privacy metrics. Hence, it is essential to consider privacy-level metrics when evaluating the effectiveness of a PPDM technique. By quantifying the potential for information leakage through both the transformed data and data mining outputs, privacy-level metrics provide valuable insights for developing and deploying PPDM solutions that balance the extraction of meaningful knowledge with the protection of sensitive information. Furthermore, understanding the limitations of each metric is crucial for selecting the

appropriate PPDM method, interpreting the results, and making informed decisions regarding the trade-off between privacy and utility.

Table 1 .A comparison between various privacy-level metrics.

Metric	Key Characteristics	Strengths	Limitations
Confidence level	· Quantifies uncertainty in estimating original values from randomized data	· Simple and straightforward · Easy to interpret.	· Ignores the original data distribution. · Can lead to inaccurate privacy assessments.
Average conditional entropy	· Incorporates original data distribution into privacy calculation.	· More nuanced assessment than Confidence level. · Considers data distribution	· Computationally complex. · Requires knowledge of joint density function.
Variance	· Measures estimation accuracy in multiplicative noise randomization.	· Simple measure for specific technique	· Limited to multiplicative noise randomization. · Not applicable to other techniques.
k-anonymity	· Ensure each record is indistinguishable from at least $k-1$ others.	· Introduces group-based anonymization concept. · Easy to implement. · Several implementation algorithms available	· Susceptible to privacy disclosure attacks. · Doesn't account for multiple records per individual.
l-diversity	· Requires l well-represented values for sensitive attributes in each group.	· Addresses the limitations of k-anonymity model.	· Determining 'well-represented' values can be subjective. · Vulnerable to skewness attacks.
t-closeness	· Ensures distribution of sensitive attributes in each group is close to original distribution.	· Addresses skewness attack limitation of l-diversity.	· Choosing appropriate t involves privacy-utility trade-off. · May require excessive data generalizations.
ε-differential privacy	· Ensures analysis outcomes are not significantly affected by single record presence/absence.	· Provides strong, formal privacy guarantees. · Robust against various attacks.	· Setting appropriate ε involves privacy-utility trade-off. · Can significantly impact analysis accuracy.
Adversarial knowledge gain	· Quantifies average information gain about sensitive attributes.	· Captures average information gain. · Based on semantic privacy concept.	· Computationally expensive for large datasets. · Sensitive to data distribution.
Adversarial accuracy gain	· Measures improvement in adversary's prediction ability.	· Intuitive interpretation. · Computationally efficient	· May underestimate information leakage. · Limited to specific attack scenario.
Hidden failure	· Measures privacy-utility trade-off in pattern recognition techniques.	· Direct measure of hidden sensitive patterns.	· Focuses only on hidden patterns. · Doesn't quantify loss of non-sensitive information.

8. Challenges and Considerations in Measuring Privacy

Measuring privacy in the context of PPDM presents a significant challenge. The inherent subjectivity of privacy, along with the technical complexities of data mining and the evolving nature of privacy threats, necessitates a nuanced and multifaceted approach. This section examines the key challenges and considerations in measuring privacy within PPDM methods.

8.1 The Lack of a Universal Privacy Definition

A fundamental challenge arises from the absence of a universally accepted definition of privacy. What constitutes a privacy violation can vary significantly based on individual perceptions, cultural norms, legal frameworks, and the specific context of data usage. As a result, there is no "one size fits all" privacy metric capable of adequately capturing the diverse range of privacy concerns across all applications. This lack of standardization makes it difficult to objectively compare different PPDM techniques and assess their relative effectiveness in protecting privacy.

8.2 Balancing Privacy and Utility: A Delicate Trade-off

A recurring theme in the literature is the inherent tension between privacy and utility in data management. PPDM techniques frequently modify or transform original data to protect privacy while maintaining its utility for analysis. However, these objectives often conflict. Techniques that offer robust privacy protection frequently result in significantly reduced data utility. For instance, excessive noise added to a dataset can render it unsuitable for specific data mining tasks. Achieving the optimal balance between these competing objectives presents a persistent challenge, necessitating careful consideration of the specific requirements of the data mining application and acceptable levels of information loss.

8.3 Contextual Factors: Data, Tasks, and Adversaries

The selection of appropriate privacy metrics is significantly influenced by contextual factors. The nature and characteristics of the underlying data, whether numerical or categorical, determine the suitability of specific metrics. Additionally, the particular data mining task, such as association rule mining, classification, or clustering, introduces distinct privacy considerations. For instance, metrics that are relevant for evaluating privacy risks in association rule mining may not be suitable for assessing privacy breaches in a classification context. Hence, different data mining tasks present varying levels of privacy risk and necessitate distinct forms of privacy protection. Therefore, the selection of privacy metrics should correspond with the specific data mining task and its objectives. Furthermore, it is crucial to understand the capabilities and motivations of potential adversaries when selecting relevant metrics. Metrics should consider the attacker's background knowledge, access to auxiliary information, and possible attack strategies. By thoroughly evaluating these factors, data miners can select metrics that provide meaningful insights into the privacy risks and effectiveness of PPDM techniques for their specific applications.

8.4 Utilizing Data Privacy and Results Privacy Metrics.

It is essential to consider privacy-level metrics when evaluating various PPDM methods. Data privacy metrics assess the extent to which sensitive information can be inferred from the transformed data. These metrics may account for privacy-related factors, such as the amount of noise added to numerical data, or the level of generalization applied to categorical data. On the other hand, the results privacy metrics focus on the potential for privacy breaches that may arise from the outputs of data mining algorithms applied to transformed data. Therefore, privacy-level metrics provide valuable insights for the development and

deployment of PPDM solutions that balance the extraction of meaningful knowledge with the protection of sensitive information.

8.5 The Quest for Standardized and Interpretable Metrics.

The absence of standardized privacy metrics presents a considerable challenge in comparing various PPDM techniques and assessing their effectiveness. Although numerous metrics have been proposed, there is currently no consensus on the most suitable metrics for specific scenarios, complicating the objective evaluation and comparison of different PPDM approaches. Thus, future research should prioritize the development of standardized and widely accepted metrics that provide meaningful insights into the privacy risks associated with different PPDM techniques. Moreover, these metrics should be easily interpreted, enabling data practitioners to understand the practical implications of different privacy-utility trade-offs.

8.6 Measuring Hiding Failure: Addressing a Key Challenge

Accurately measuring hiding failure of PPDM techniques is crucial for evaluating their effectiveness in protecting sensitive information. As previously clarified, hiding failure refers to the degree to which sensitive patterns remain detectable after the data has been sanitized. However, quantifying hiding failure poses challenges, as it necessitates the identification of all sensitive patterns within a dataset, a task that can be both computationally intensive and context dependent. Furthermore, the effectiveness of a PPDM algorithm in preventing disclosure can vary based on the specific data mining techniques utilized by an adversary. Consequently, evaluating hiding failure requires consideration of a diverse range of potential attack strategies and data mining algorithms. The development of robust methodologies and standardized frameworks for assessing hiding failure against various attacks remains an active area of research.
In summary, measuring privacy in PPDM presents a complex set of challenges. The lack of a universal definition of privacy, the inherent trade-off between privacy and utility, and the necessity to consider various contextual factors all contribute to this complexity. Addressing these challenges requires moving beyond simplistic metrics and adopting a multifaceted approach that encompasses both data and results privacy, incorporates adversary models, and acknowledges the dynamic nature of privacy threats. Moreover, a deeper understanding of privacy-level metrics and their limitations is essential for guiding the development and application of more effective and trustworthy PPDM techniques that ultimately contribute to the advancement of responsible data mining practices. Consequently, future research should prioritize the development of standardized, interpretable, and context-aware privacy metrics that can effectively quantify and mitigate evolving privacy risks in PPDM.

9. Conclusion

Privacy-Preserving Data Mining (PPDM) techniques are crucial for extracting valuable insights from sensitive data while ensuring the protection of individuals' privacy. Evaluating the effectiveness of these techniques necessitates robust metrics that accurately quantify the level of privacy protection provided. Despite the increasing importance of PPDM methods, there remains a limited number of studies that address the metrics used to evaluate their effectiveness, particularly in relation to privacy preservation. This paper addressed this research gap by presenting an extensive evaluation study of privacy-level metrics for PPDM methods. In addition, a new classification for the privacy-level metrics examined based on the associated phase of PPDM in which they are most applicable is introduced. Besides, this paper presented a detailed analysis of various privacy-level metrics, exploring their strengths, limitations, and

applicability to various data mining scenarios. The study investigated data privacy metrics such as confidence level, average conditional entropy, and variance between original and perturbed data, which measure the uncertainty an attacker encounters when attempting to infer original data from the protected dataset. Furthermore, the introduced analysis included results privacy metrics like hidden failure, which assesses the risk of sensitive information disclosure from data mining outputs. Moreover, the paper offered a comprehensive discussion of various considerations and challenges associated with measuring privacy across different PPDM methods in the absence of a universally accepted definition. This discussion highlighted the importance of several contextual factors in selecting appropriate privacy metrics. Ultimately, this work aims to assist researchers and practitioners in the appropriate selection and interpretation of these metrics while also contributing to the development and deployment of robust PPDM techniques that balance the extraction of valuable insights from data with the maintenance of individuals' privacy.

References

1. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. Amsterdam, The Netherlands: Elsevier, 2012.
2. Siraj, Maheyzah Md, Nurul Adibah Rahmat, and Mazura Mat Din. "A survey on privacy preserving data mining approaches and techniques." In Proceedings of the 2019 8th international conference on software and computer applications, pp. 65-69. 2019.
3. Abdelhameed, Saad A., Sherin M. Moussa, and Mohamed E. Khalifa. "Privacy-preserving tabular data publishing: a comprehensive evaluation from web to cloud." Computers & Security 72 (2018): 74-95.
4. Yu, Shui. "Big privacy: Challenges and opportunities of privacy study in the age of big data." IEEE access 4 (2016): 2751-2763.
5. Amer, Abeer Abdel H. "-Unveiling the Power of Big Data: A Comprehensive Review of its Role in the Banking Sector." International Journal of Intelligent Computing and Information Sciences (2024).
6. Naidu, P. Annan, and M. Vamsi Krishna. "Comprehensive Review on Privacy Preserving Data Mining Techniques and Methods." International Journal of Engineering and Management Research (IJEMR) 7, no. 1 (2017): 121-126.
7. Abdelhameed, Saad A., Sherin M. Moussa, Nagwa L. Badr, and M. Essam Khalifa. "The generic framework of privacy preserving data mining phases: challenges & future directions." In 2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS), pp. 341-347. IEEE, 2021.
8. Patel, Darshana, and Radhika Kotecha. "Privacy preserving data mining: A parametric analysis." In Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications: FICTA 2016, Volume 2, pp. 139-149. Springer Singapore, 2017.
9. L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, “Information security in big data: privacy and data mining,” Access, IEEE, vol. 2, pp. 1149–1176, 2014.
10. Nasiri, Negar, and MohammadReza Keyvanpour. "Classification and evaluation of privacy preserving data mining methods." In 2020 11th International Conference on Information and Knowledge Technology (IKT), pp. 17-22. IEEE, 2020.
11. Mendes, Ricardo, and João P. Vilela. "Privacy-preserving data mining: methods, metrics, and applications." IEEE Access 5 (2017): 10562-10582.
12. Senosi, Aobakwe, and George Sibiya. "Classification and evaluation of privacy preserving data mining: a review." In 2017 IEEE AFRICON, pp. 849-855. IEEE, 2017.

13. Kiran, Ajmeera, and D. Vasumathi. "A comprehensive survey on privacy preservation algorithms in data mining." In 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1-7. IEEE, 2017.
14. Ali, Mohamed Ashraf, Sherine Rady, Tamer Abdelkader, and Tarek F. Gharib. "An efficient hiding method for privacy preserving utility mining." International Journal of Intelligent Computing and Information Sciences 23, no. 1 (2023): 69-83.
15. Basha, M. John, T. Satyanarayana Murthy, A. S. Valarmathy, Ahmed Radie Abbas, Djuraeva Gavhar, R. Rajavarman, and N. Parkunam. "Privacy-Preserving Data Mining and Analytics in Big Data." In E3S Web of Conferences, vol. 399, p. 04033. EDP Sciences, 2023.
16. Prasanthi, Kundeti Naga, and M. V. P. Chandra Sekhara Rao. "A comprehensive assessment of privacy preserving data mining techniques." In Proceedings of Second International Conference on Sustainable Expert Systems: ICSES 2021, pp. 833-842. Singapore: Springer Nature Singapore, 2022.
17. Zhang, Nan, Wei Zhao, and Jianer Chen. "Performance measurements for privacy preserving data mining." In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 43-49. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.
18. Bertino, Elisa, Igor Nai Fovino, and Loredana Parasiliti Provenza. "A framework for evaluating privacy preserving data mining algorithms." Data Mining and Knowledge Discovery 11 (2005): 121-154.
19. Bertino, Elisa, Dan Lin, and Wei Jiang. "A survey of quantification of privacy preserving data mining algorithms." Privacy-preserving data mining: Models and Algorithms (2008): 183-205.
20. Langheinrich, Marc. " Privacy in Ubiquitous Computing." In Ubiquitous computing fundamentals, pp. 109-174. Chapman and Hall/CRC, 2018.
21. Kumar, Atul, and Manasi Gyanchandani. "A comparative survey on privacy preservation and privacy measuring techniques in data publishing." In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1902-1906. IEEE, 2018.
22. Shah, Alpa, and Ravi Gulati. "Privacy preserving data mining: techniques, classification and implications-a survey." Int. J. Comput. Appl 137, no. 12 (2016): 40-46.
23. Vaghashia, Hina, and Amit Ganatra. "A survey: privacy preservation techniques in data mining." International Journal of Computer Applications 119, no. 4 (2015).
24. Aggarwal, Charu C., and Philip S. Yu. A general survey of privacy-preserving data mining models and algorithms. Springer US, 2008.
25. Agrawal, Rakesh, and Ramakrishnan Srikant. "Privacy-preserving data mining." In Proceedings of the 2000 ACM SIGMOD international conference on Management of data, pp. 439-450. 2000.
26. Agrawal, Dakshi, and Charu C. Aggarwal. "On the design and quantification of privacy preserving data mining algorithms." In Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pp. 247-255. 2001.
27. Oliveira, Stanley RM, and Osmar R. Zaiane. "Privacy preserving clustering by data transformation." Journal of Information and Data Management 1, no. 1 (2010): 37-37.
28. Brickell, Justin, and Vitaly Shmatikov. "The cost of privacy: destruction of data-mining utility in anonymized data publishing." In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 70-78. 2008.
29. Fung, Benjamin CM, Ke Wang, Rui Chen, and Philip S. Yu. "Privacy-preserving data publishing: A survey of recent developments." ACM Computing Surveys (Csur) 42, no. 4 (2010): 1-53.
30. Abdelhameed, Saad A., Sherin M. Moussa, and Mohamed E. Khalifa. "Enhanced additive noise approach for privacy-preserving tabular data publishing." In 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), pp. 284-291. IEEE, 2017.

31. Sweeney, Latanya. "k-anonymity: A model for protecting privacy." International journal of uncertainty, fuzziness and knowledge-based systems 10, no. 05 (2002): 557-570.
32. Machanavajjhala, Ashwin, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. "l-diversity: Privacy beyond k-anonymity." Acm transactions on knowledge discovery from data (tkdd) 1, no. 1 (2007).
33. Li, Ninghui, Tiancheng Li, and Suresh Venkatasubramanian. "t-closeness: Privacy beyond k-anonymity and l-diversity." In 2007 IEEE 23rd international conference on data engineering, pp. 106-115. IEEE, 2006.
34. Dwork, Cynthia. "Differential privacy." In International colloquium on automata, languages, and programming, pp. 1-12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.
35. Oliveira, Stanley RM, and Osmar R. Zaiane. "Privacy preserving frequent itemset mining." In Proceedings of the IEEE ICDM workshop on privacy, security and data mining, vol. 43, p. 54. 2002.