Vol.8 No.1 (2024): Journal Scientific MQRInvestigar ISSN: 2588–0659
https://doi.org/10.56048/MQR20225.8.1.2024.204-235

Anonymization and pseudonymization techniques in personal data protection

Técnicas de anonimización y pseudonimización en la protección de datos personales

Autores:

Córdova-Real, Jorge Luis
Ingeniero en Sistemas
Pontificia Universidad Católica del Ecuador
Maestría en Ciberseguridad
Ambato – Ecuador
jcordova@pucesa.edu.ec
https://orcid.org/0009-0009-8915-3831

López-Sevilla, Galo Mauricio
Ingeniero en Sistemas, Máster en Informática
Pontificia Universidad Católica del Ecuador
Docente Escuela de Ingenierías
Ambato – Ecuador
glopez@pucesa.edu.ec
https://orcid.org/0000-0003-4699-4875

Fechas de recepción: 01-DIC-2023 aceptación: 08-ENE-2024 publicación: 15-MAR-2024

https://orcid.org/0000-0002-8695-5005
http://mqrinvestigar.com/

Resumen

Las técnicas de anonimización y pseudonimización ofrecen a los usuarios la protección de sus datos personales para evitar que sean difundidos y utilizados con un propósito ajeno para los cuales fueron recolectados. El presente articulo tiene como objetivo comparar las diversas técnicas de protección de datos personales como la anonimización y pseudonimización. Para ello, se examina a detalle su fundamentación teórica y características inherentes de cada técnica, los procedimientos de aplicación, beneficios y limitaciones que trae consigo la aplicación de estas metodologías. Para investigar acerca de la temática, se emplea la metodología PRISMA que permite buscar, seleccionar y analizar la literatura científica, dando como resultado de la recopilación de documentos mediante el empleo de gestores de búsqueda un total de 32 artículos científicos que fueron desarrollados entre 2018-2023 y contienen información relevante que aporta al desarrollo del presente artículo. Los resultados indican que la técnica de anonimización está enfocada en presentar datos no identificables mediante la adición de ruido, permutaciones o privacidad diferencial que contribuyen a mantener la privacidad de los datos y preservar la utilidad de los mismos. Paralelamente, la pseudonimización tiene como objetivo reemplazar la información inicial identificable con seudónimos que mantengan protegida la identidad de una persona. Como conclusión del estudio, se definen las técnicas de protección de información personal. Estas estrategias son fundamentales para identificar datos considerados como confidenciales para los usuarios, aplicando métodos de privatización que reduzcan los riesgos inherentes al compartir información con terceros; logrando el equilibrio entre utilidad y la reserva de información.

Palabras clave: Anonimización; Pseudonimización; Seguridad; Datos; Información; Protección

Abstract

Anonymization and pseudonymization techniques offer users protection of their personal data from disclosure and use for purposes other than those for which they were collected. The purpose of this article was to compare the various techniques of personal data protection such as anonymization and pseudonymization. For this purpose, the theoretical basis and inherent characteristics of each technique, the application procedures, benefits, and limitations of the application of these methodologies were examined in detail. To recapitulate information about the subject matter, the PRISMA methodology was used to search, select, and analyze the scientific literature, resulting in the collection of documents through the use of search managers a total of 32 scientific articles that were developed between 2018-2023 and contained relevant information that contributed to the development of this article. The results indicate that the anonymization technique is focused on presenting unidentifiable data by adding noise, permutations or differential privacy that help to maintain data privacy and preserve data utility. In parallel, pseudonymization aims to replace initial identifiable information with pseudonyms that keep a person's identity protected. The study concludes by defining personal information protection techniques. These strategies are fundamental to identify data considered confidential for users, applying privatization methods that reduce the risks inherent in sharing information with third parties; achieving a balance between usefulness and confidentiality of information.

Keywords: Anonymization; Pseudonymization; Security; Data; Information; Protection

Introducción

La protección de datos personales es un tema de creciente importancia dentro del contexto actual, debido al aumento de la digitalización y la recopilación masiva de información. Los fenómenos de desarrollo tecnológico han hecho indispensable el uso de herramientas sistemáticas que faciliten el desenvolvimiento diario de las personas, el aumento de su productividad laboral y el entablar relaciones socio económicas; por tal razón, la mayoría de las actividades requieren brindar información de carácter personal (Pratama & Pati, 2021). Según Ozeran la tecnología puede representar una de las herramientas más poderosas que ofrece a los usuarios un mundo de beneficios que ayudan a simplificar el modo de vivir. Sin embargo, también puede generar repercusiones en temas legales al momento de emplear y manejar datos personales propios o de terceras personas (Ozeran et al., 2021).

Actualmente, la preocupación de los seres humanos es el tratamiento y privatización de su información personal; el término big data comprende la recopilación de dicha información para rastrearla y examinarla con el propósito de generar fuentes estratégicas las cuales permitan establecer patrones o tendencias de los seres humanos por satisfacer sus necesidades personales con respecto a sus hábitos y comportamiento habitual (Li et al., 2022; Lutfi et al., 2022). La prestación de información que puede ser fácilmente almacenada y transmitida mediante hardware y software ha generado una mayor atención por parte de la comunidad debido a que para acceder a diversos servicios o aplicaciones a menudo requieren la recopilación de datos informativos cuyo propósito es verificar la identidad y de los cuales en su mayoría no cuentan con una ley de regularización de uso (Batko & Ślęzak, 2022).

La falta de regulaciones ha hecho que los países se vuelvan más propensos a la exposición de información confidencial. El caso de Indonesia un país que no cuenta con ningún tipo de normativa que controle el uso de los datos y su protección ha desencadenado una problemática en donde la información difundida por parte de los usuarios es empleada con fines despectivos como el acoso sexual, persecución, la generación de préstamos a nombre de terceras personas y la falsificación de documentos de identidad que fueron vendidos por millones de rupias. Este caso demuestra el escaso régimen de protección de datos personales

lo que se vuelve ciertamente peligroso porque expone a los usuarios a situaciones desfavorables (Puluhulawa et al., 2020).

En este contexto, el objetivo de este trabajo se centra en comparar las técnicas de anonimización y pseudonimización en la protección de datos personales; las mismas que han surgido como estrategias cruciales para salvaguardar la información contenida en los datos personales. La anonimización implica la eliminación total o parcial de atributos que son de carácter significativo y a partir de los cuales se puede reconocer a un individuo en particular; por el contrario, la pseudonimización sustituye los atributos por seudónimos o códigos que de igual manera precautelan la privacidad de los usuarios. Existe un aspecto muy importante a tomar en cuenta con respecto a estas técnicas y es la predisposición de uso de herramientas tecnologías para revertir el efecto de confidencialidad. Por ende, es importante que se genere mecanismos los cuales no alteren la privacidad de los usuarios (Anant & Prasad, 2022; Murakami et al., 2023) .

Frente al vasto panorama de investigaciones existentes, este documento ha sido elaborado con el propósito de detallar sistemáticamente las técnicas de anonimización y pseudonimización de datos personales, adoptando un enfoque cualitativo para la comprensión de conceptos asociados a las técnicas de protección de información personal, su clasificación, los procedimientos a ejecutar para implementar estos métodos y finalmente reconocer sus beneficios y limitaciones. La revisión se fundamenta en evidencia científica de calidad, empleando fuentes bibliográficas como: Scopus, Web of Sciene, SpringerLink, Proquest, Scielo y Dialnet. De este modo, se contribuirá al conocimiento actualizado acerca de las técnicas de anonimización y pseudonimización dentro del contexto de protección de datos personales, sus etapas, ventajas y desventajas para la comunidad.

Material y métodos

La revisión sistemática es una herramienta importante dentro del ámbito investigativo, permite recopilar información relevante con respecto a diversas temáticas de estudio para responder a interrogantes generadas a partir del conocimiento individual para ser sustentadas

y fundamentadas mediante investigaciones existentes. Es importante que se presente detalladamente el propósito por el cual se ejecuta la revisión, las diversas metodologías empleadas en el desarrollo (identificación y selección de los estudios más trascendentales); así como los hallazgos obtenidos para la formulación de teorías explicativas acerca de la ocurrencia de los fenómenos específicos de interés (Page et al., 2021).

El presente estudio se fundamenta mediante el empleo de la metodología PRISMA. El objetivo es comparar las técnicas de anonimización y pseudonimización por medio de la descripción de sus procedimientos para la protección de datos personales. La finalidad es estructurar de forma ordenada la evidencia recolectada con relación a las diversas técnicas de anonimización y pseudonimización para comprender su fundamentación teórica, principales aplicaciones prácticas, beneficios y limitaciones con relación a la protección de datos personales y seguridad de los usuarios. A través de esta metodología se estructura los autores para evidenciar el proceso en el cual se llevó a cabo las investigaciones y como se obtuvieron los respectivos hallazgos de los estudios (Sarkis-Onofre et al., 2021). El proceso de revisión sistemática inicia con la formulación de preguntas de investigación. Posteriormente, se establecen los métodos de búsqueda, la identificación y selección de investigaciones relacionadas al tema de investigación y se procede a la presentación consolidada de los datos e información clave obtenidos de los estudios analizados (Selcuk, 2019).

Preguntas de investigación

Se establecieron tres preguntas de investigación (Tabla 1). Estas interrogantes se encuentran estrechamente alineadas con el objetivo, funcionando como un marco de orientación durante la fase exploratoria. El estudio abarca tres aspectos principales: descripción de las técnicas de anonimización y pseudonimización, aplicaciones prácticas y los principales beneficios y limitaciones de cada técnica con respecto a la protección de datos personales.

Tabla 1.

Preguntas de investigación

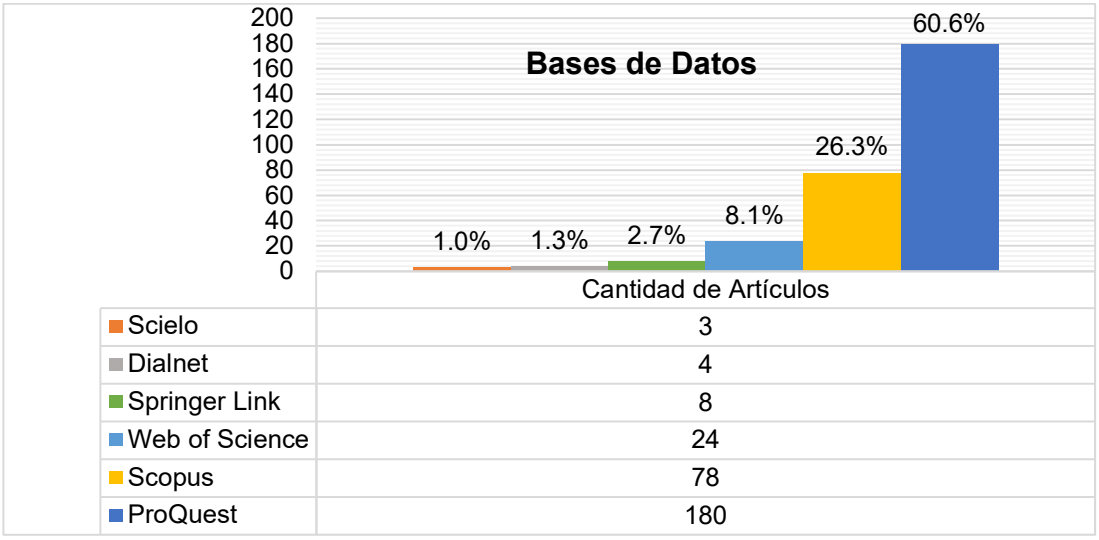
N.	Pregunta	Motivación
RQ1	¿Cuáles son los aportes teóricos acerca de las técnicas de anonimización y pseudonimización en el contexto de protección de los datos personales?	Interpretar la fundamentación teórica y características relacionadas con las técnicas de anonimización y pseudonimización.
RQ2	¿Cuáles son los procedimientos de aplicación de las técnicas de anonimización y pseudonimización para la protección de datos personales?	Describir el conjunto de procedimientos a ejecutar para la aplicación de técnicas de anonimización y pseudonimización.
RQ3	¿Cuáles son los beneficios y limitaciones de las técnicas de anonimización y pseudonimización en la protección de datos personales?	Identificar las ventajas y desventajas propias de utilizar las técnicas de protección de datos personales.

Sistematización de la búsqueda y recolección de datos

Se realizó una búsqueda exhaustiva que abarcó todas las publicaciones entre los años 2018-2023. La investigación se efectuó en diversas bases de datos, entre las que se incluyen Scopus con un porcentaje de representación del 26.3%, Proquest 60.6%, Web of Science 8.1%, Springer Link 2.7%, Scielo 1.0% y Dialnet 1.3 % (Figura 1). Para la selección de artículos se estableció un criterio generalizado el cual determina que los documentos deben contener información asociada al tema de estudio disponibles en idioma español o ingles publicados dentro de los últimos 5 años.

Figura 1.

Número de documentos por base de datos



Nota: El gráfico representa el porcentaje de representación de las diversas bases de datos. Elaboración Propia

Para cribar la información se emplearon conectores lógicos para estructurar palabras clave de búsqueda como: [“anonymization” AND “pseudonymization”] OR [“anonymization” AND “techniques”] OR [“pseudonymization” AND “techniques”] OR [“anonymization techniques” AND “GDPR”] OR [“anonymization procedures” AND “personal data protection”] OR [“pseudonymization techniques” AND “data privacy measures”] OR [“anonymization benefits” AND “personal data protection”] OR [“pseudonymization limitations” AND “data privacy”] OR [“anonymization impact” AND “data utility”]. Los motores de búsqueda permitieron el uso de diversas combinaciones y empleo de términos clave con la finalidad de acceder a aquellos artículos que poseen información relacionada con la temática del artículo que permitan sustentar las preguntas de investigación formuladas.

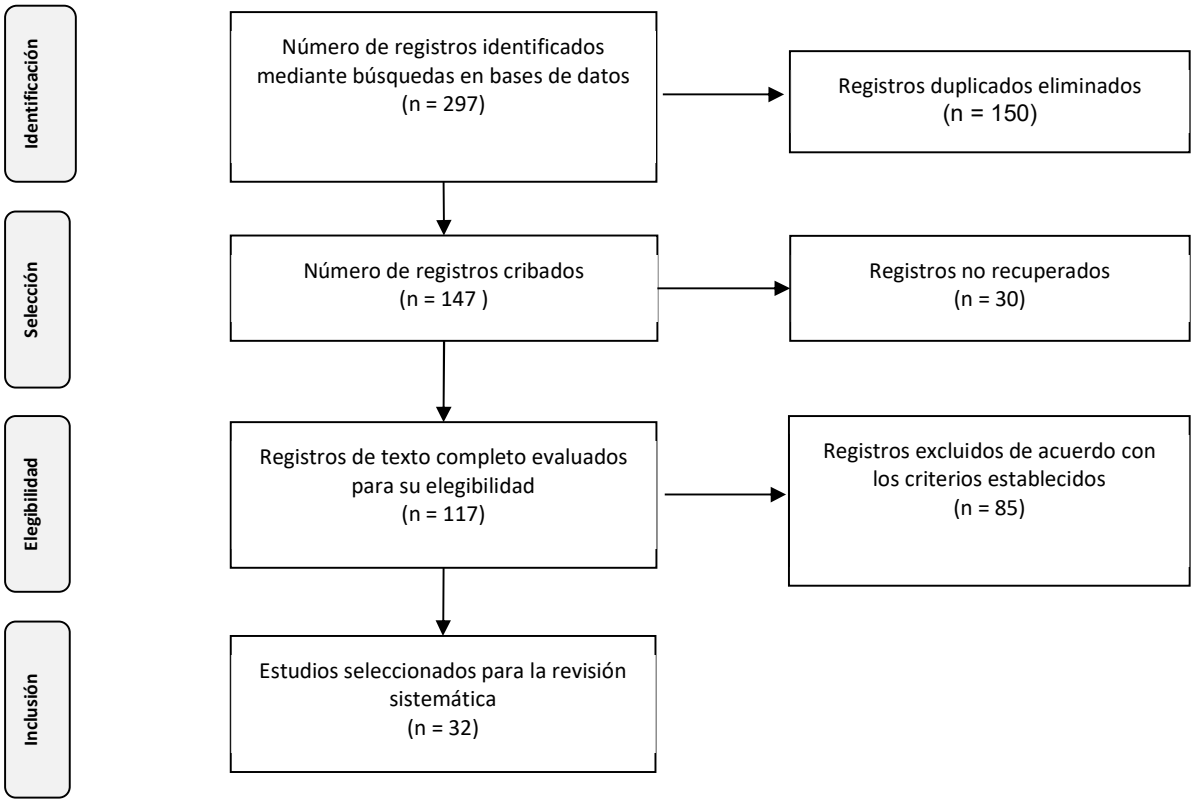
Selección de la información

El proceso de selección de los estudios en el contexto de la revisión sistemática se identificaron 4 etapas primordiales. La primera etapa se define identificación y consiste en la

búsqueda exhaustiva de estudios relevantes en diversas bases de datos y fuentes, para ello se recopiló un total de 297 documentos, la segunda etapa se denomina selección y aplicando el criterio generalizado de elegibilidad se analizó la información contenida de cada artículo para incluir aquellos que cumplen con las especificaciones del estudio. Posteriormente, la elegibilidad es la etapa en donde se examina de manera minuciosa los documentos dando como resultado un total de 32 artículos los cuales forman parte de la última etapa de inclusión y son empleados para el sustento de la investigación (Sánchez et al., 2022). Todas estas etapas se encuentran sintetizadas en un diagrama de flujo (Figura 2).

Figura 2.

Diagrama de flujo de cuatro niveles – PRISMA.



Nota: El gráfico representa el diagrama de flujo representativo de la metodología PRISMA y sus diversas etapas.

Resultados y Discusión

En el Anexo 1, se muestra 32 artículos seleccionados de la búsqueda literaria. La información extraída de cada documento se fundamenta en responder las preguntas de investigación. Se verificaron cada una de las referencias con el objetivo de asegurar que contengan información relevante que aporte a la resolución de las interrogantes de estudio para reconocer las técnicas de privacidad de datos empleadas para asegurar la confidencialidad de información sensible y personal, salvaguardando la dignidad y autonomía de libertad de las personas; teniendo en cuenta que el anonimato de datos previene el abuso de robo de identidad, el fraude y la manipulación a través de la información personal.

RQ1. ¿Cuáles son los aportes teóricos acerca de las técnicas de anonimización y pseudonimización en el contexto de protección de los datos personales?

En la actualidad, la recopilación masiva de datos denominada big data o megadatos recopila conjunto de información que los sistemas tradicionales no tienen la capacidad de procesar. Los datos están conformados por información estructurada y no estructurada y su tamaño varía desde terabytes hasta exabytes; el bigdata no solo comprende el volumen de información, sino que también hace referencia a la velocidad con la que se generan y son procesados (Chahal et al., 2022). Tradicionalmente, la recolección de datos se realiza mediante encuestas u otras fuentes administrativas cuyos resultados se plasman en tablas; según la regulación europea de protección de datos determina que al usar información es necesario tener el consentimiento de los individuos para hacer uso de la misma con fines secundarios (Majeed et al., 2022).

La protección de datos, según la Organización para la Cooperación y Desarrollo Económico, se respalda en los derechos personales para el tratamiento final de los datos proporcionados normados bajo fundamentación legal. El principio de salvaguardar la seguridad comprende la protección de la información ante situaciones como pérdidas, accesos no autorizados, modificaciones o divulgación; además, cuenta con políticas de transparencia que definen la naturaleza de los datos, su uso e identidad del organismo tratante (Angelozzi, 2020).

Actualmente, existen técnicas que contribuyen a la privacidad de las bases de datos denominadas anonimización y pseudonimización (Al-Zubaidie et al., 2019).

La anonimización convierte los datos personales obtenidos en información confidencial mediante la eliminación de detalles para evitar identificar la identidad de los individuos, es una herramienta que prevé riesgos relacionados con el tratamiento y uso de datos (Al-Zubaidie et al., 2019). El objetivo de la anonimización es presentar información de forma reservada, por ejemplo, el analizar la tendencia de la población por aplicar técnicas de sostenibilidad resulta irrelevante que se resguarde información como fechas de nacimiento; por el contrario, el acceso a la información y datos personales mediante direcciones IP requiere de protocolos de seguridad más robustos (Ribeiro & Nakamura, 2019). Las diversas técnicas de anonimización son responsables del tratamiento de valores (ver Tabla 4); la aleatorización, modifica el contenido de los datos para eliminar el vínculo existente entre la persona que suministra la información y la encargada de recolectarla; la finalidad es que los datos se tornen ambiguos y no puedan ser inferidos por terceras personas (Marques & Bernardino, 2020).

La adición de ruido es una metodología empleada para redistribuir los datos; consiste en la alteración de la información recapitulada para tener un margen de probabilidad menor con respecto a los ataques informáticos (Torra & Salas, 2019; Virupaksha & Dondeti, 2021). El tratamiento de valores numéricos con adición de ruido es de carácter aleatorio; en cambio, de forma categórica se alternan para conseguir el equilibrio entre la privacidad y la utilidad de los datos (Bousquet et al., 2018).

Tabla 4.

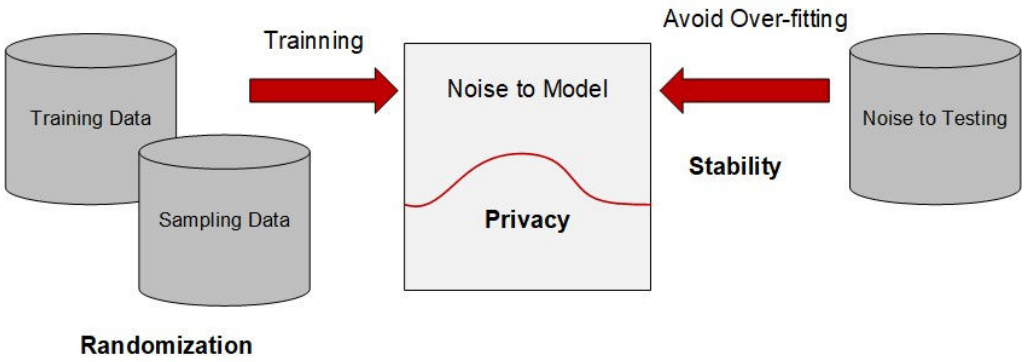
Tipos de técnicas de anonimización de datos

Tipo	Descripción
Adición de Ruido	La adición de ruido es una técnica de anonimización de datos que consiste en la introducción de aleatoriedad y perturbación en el conjunto de información original. El objetivo de esta función es la incorporación de valores aleatorios los cuales tiendan a

	modificar de manera ligera la base de datos con la finalidad de crear patrones complejos de identificar por parte de terceras personas (Torra & Salas, 2019; Virupaksha & Dondeti, 2021). El tratamiento de valores numéricos con adición de ruido es de carácter aleatorio; en cambio, de forma categórica se alternan para conseguir el equilibrio entre la privacidad y la utilidad de los datos (Bousquet et al., 2018).
Permutación	La permutación reordena los elementos presentes dentro de un conjunto de datos. (De Pascale et al., 2023; Majeed & Lee, 2021). Consiste en trasladar los valores desde un registro a otro con la finalidad de suprimir la correlación estadística para que interactúen independiente (Mohammed et al., 2022). El permutar permite reorganizar la distribución de las filas de un conjunto de datos de forma aleatoria suprimiendo la correlación existente entre el individuo y la información proporcionada (Tepe et al., 2023).
Privacidad Diferencial	La privacidad diferencial es una herramienta que preserva la privacidad de los datos a través de algoritmos aleatorizados entregando información modificada que evita la reidentificación del usuario. La idea principal consiste en introducir ruido estadístico o perturbaciones de forma controlada en los valores obtenidos, para aumentar la dificultad de inferir en información específica acerca de una persona (ver Figura 3).

Figura 3.

Estructura de privatización de datos diferencial



Nota: El gráfico representa la estructura de protección de datos mediante la adición de ruido y perturbaciones. Tomado de (Zhu et al., 2022).

La privacidad diferencial se basa en la premisa de no alterar significativamente los valores iniciales o información de una persona para evitar que pierda su sentido (Zhu et al., 2022). Este método ha ido ganando importancia en entornos de tratamiento de información sensible como: pronósticos médicos, investigación y redes sociales (Jiang et al., 2023). La generalización es un grupo de técnicas de anonimización que consiste en normalizar los datos o información obtenida por medio de la modificación de las escalas u órdenes de magnitud las cuales sean superiores dentro de una jerarquía; un ejemplo, sustituir la categoría ciudad por una región agrupando los datos y preservando el anonimato de cada individuo. La generalización se subdivide en un conjunto de técnicas que generan estabilidad y protección de los datos personales que son a) Anonimato-K y b) Diversidad-L y Proximidad-T. El anonimato-k es una técnica cuyo propósito es evitar características de singularidad de los individuos que forman parte de un conjunto k de personas.

La metodología consiste en crear categorías con atributos que identifiquen a los individuos dentro de un conjunto (Slijepčević et al., 2021). Un ejemplo claro se esquematiza en la figura 4., se ha generalizado tres categorías con datos confidenciales como: el código postal, edad

y el salario de cada usuario mediante rangos que contienen los valores proporcionados por cada individuo.

Figura 4.

Modelo de aplicación de técnica de anonimato-k para la privacidad de datos

N°	Código Postal	Edad	Sexo	Salario
1	1001	28	Femenino	550
2	1003	35	Masculino	1000
3	1251	60	Masculino	1520
4	2656	75	Femenino	320
5	3225	54	Femenino	450
6	1323	31	Masculino	750
7	4420	22	Femenino	250

N°	Código Postal	Edad	Sexo	Salario
1	1000-1500	Adultez	Femenino	500-550
2	1000-1500	Adultez	Masculino	1000-1500
3	1000-1500	Persona Mayor	Masculino	1500-2000
4	2500-3000	Persona Mayor	Femenino	300-350
5	3000-3500	Adultez	Femenino	400-450
6	1000-1500	Adultez	Masculino	700-750
7	4000-4500	Juventud	Femenino	200-250

Nota: La técnica de anonimato-k unifica categóricamente el conjunto de datos para proteger información de los individuos. Elaboración Propia.

Un valor k menor reduce significativamente el nivel de protección de los datos dando como resultado una menor cantidad de participantes y una mayor cantidad de información sensible expuesta al público (Mao et al., 2018). La diversidad l está fundamentada en que cada categoría equivalente debe poseer al menos l elementos diferentes entre sí (Kacha et al., 2022). Por otra parte, la proximidad-t optimiza la diversidad-l creando clases equivalentes para demostrar que los atributos de cada conjunto estén distribuidos de forma igualitaria que los valores originales (Mahanan et al., 2021; Sweeney, 2022).

La pseudonimización sustituye atributos por seudónimos; a diferencia de las técnicas de anonimización, esta metodología contiene información que puede ser vinculada a un individuo salvaguardada de manera confidencial (Kohlmayer et al., 2019). El principal problema que se identifica de esta técnica es que las personas aún se mantienen vulnerables de ataques individuales y pueden ser reidentificados debido a la repetibilidad del uso de seudónimos dentro de la base de datos. Entre las principales técnicas de pseudonimización se encuentran las siguientes: a) Cifrado con clave secreta, b) Función Hash, c) Descomposición en tokens.

Tabla 5.

Tipos de técnicas de pseudonimización de datos.

Tipo	Descripción
Cifrado con Clave Secreta	El cifrado con clave secreta es una técnica aplicada para el reconocimiento de información usando contraseñas de seguridad que permite descifrar valores o datos acerca de un usuario, su eficiencia se relaciona con el compromiso del personal autorizado en proteger los datos y almacenarlos en entornos seguros (Neumann et al., 2019).
Función Hash	La función hash es una técnica de cifrado la cual tiene como finalidad transformar un texto de carácter plano en uno de tipo hexadecimal con una extensión fija de 40 caracteres con el objetivo de proteger la información emitida en el texto de entrada u original (Bey et al., 2020).
Descomposición en Tokens	La descomposición en tokens es una técnica utilizada comúnmente en finanzas para proteger información personal como: números de identificación de las tarjetas por caracteres difíciles de descifrar por parte de los atacantes. Su mecanismo se basa en convertir los caracteres de texto en una combinación de números, letras y símbolos especiales utilizando la función hash (Englert et al., 2023).

RQ 2. ¿Cuáles son los procedimientos de aplicación de las técnicas de anonimización y pseudonimización para la protección de datos personales?

La anonimización y pseudonimización de datos es un proceso de protección de información con relación a grandes volúmenes de valores. El procedimiento de aplicación empieza desde el reconocimiento de datos sensibles; posteriormente, aplicación de técnicas; por último, la gestión de riesgos existentes. Cada etapa dificulta el acceso a los datos de terceras personas, precautelando su utilidad sin afectar los derechos de protección, ver Tabla 6.

Tabla 6.

Procedimiento de aplicación de las técnicas de anonimización y pseudonimización

Etapas	Descripción
Reconocimiento de Datos	Para mantener la confidencialidad de los datos, es importante identificar el tipo de información que se está recopilando. Los identificadores directos son datos que pueden utilizarse para identificar a una persona de manera única; suelen ser de conocimiento público, como nombres, direcciones de correo electrónico y redes sociales. Los identificadores indirectos son datos que pueden utilizarse para identificar a una persona, pero no de manera única e incluyen direcciones domiciliarias, edad, género y otros datos demográficos. Finalmente, los atributos objetivo son datos de naturaleza sensible y no deben ser divulgados sin el consentimiento de la persona. Estos datos pueden incluir números de cuentas bancarias, claves e incluso diagnósticos médicos (Meindl & Templ, 2019).
Desidentificar los datos	Etapa que consiste en la eliminación de identificadores directos, indirectos y asociados de información por seudónimos o alias los cuales evitan que mediante deducción o cálculos se reidentifique el origen de los datos (Tahir & Brezillon, 2022; Tomás et al., 2022). **Nombre** / **Alias** Pedro / 1234 Edison / 1525 Samantha / 4897
Aplicar técnicas de anonimización y pseudonimización	La tercera etapa es la aplicación de las técnicas de anonimización y pseudonimización para restringir el acceso de información contenida dentro de una categoría de datos que se considera confidencial, entre las principales técnicas se

encuentra la adición de ruido, permutación, privacidad diferencial, clave secreta, función hash y descomposición en tokens (Kohlmayer et al., 2019).

Calcular el Riesgo

La determinación del riesgo denominado k-anonimidad o de reidentificación de datos determina que si existe una mínima cantidad de registros mayor es la posibilidad de reidentificación; por el contrario, si existe mayor cantidad de información el riesgo disminuye y mayor es el rango de protección (Dubagunta et al., 2022; Mahanan et al., 2021).

Valor k	N° de Identidad	Edad	Sueldo
k=2	180xxxxx01	Adultez	250-300
	023xxxxx54	Juventud	800-850
k=4	175xxxxx58	Persona Mayor	1000-1100
	198xxxxx72	Persona Mayor	450-500
	021xxxxx02	Adultez	300-350
	053xxxxx24	Juventud	500-550
k=2	170xxxxx15	Adultez	600-650
	180xxxxx41	Juventud	700-750

Gestión de los Riesgos

Es importante reconocer los tipos de riesgo existentes al exponer información personal. La revelación de identidad consiste en algoritmos que al no ser tratados correctamente con técnicas de anonimización y pseudonimización tienden a ser reconocidos por otros usuarios debido a su versatilidad como reemplazar “1” por “a” o “2” por “b”. La revelación de atributos son inferencias realizadas a individuos que no forman parte de un conjunto de datos únicamente por presentar atributos similares. Finalmente, la revelación por inferencias conforme a la aplicación de datos estadísticos formula conclusiones de personas que no forman parte de los registros (Meindl & Templ, 2019).

El analizar cada etapa del procedimiento de anonimización y pseudonimización deduce que es importante reconocer cada uno de los riegos y calcularlos con la finalidad de emplear las metodologías más acertadas para precautelar la privacidad de los datos personales. Actualmente, la protección mediante contraseñas es una herramienta empleada

mundialmente y consiste en la asignación de un código de acceso con complejidad alta con una combinación de caracteres tipo mayúsculas, minúsculas y números. El asignar contraseñas robustas a conjuntos de datos sensibles, limita el acceso no autorizado y se reduce el riesgo de revelar información personal. La anonimización implica la eliminación de datos identificables directos, mientras que la pseudonimización implica la sustitución de identificadores por códigos o alias. Sin contraseñas sólidas, estas técnicas pueden resultar vulnerables, comprometiendo la integridad de las mismas (Karunanayake et al., 2021).

RQ3. ¿Cuáles son los beneficios y limitaciones de las técnicas de anonimización y pseudonimización en la protección de datos personales?

La anonimización de datos ofrece una amplia gama de beneficios, desde mejorar la seguridad y la privacidad fomentando la innovación, lo que la convierte en una práctica esencial en la gestión moderna de datos (Bouchelaghem & Omar, 2020; Hintze & El Emam, 2018).

Tabla 7.

Beneficios y limitaciones de la técnica de anonimización de datos.

Beneficios	Limitaciones
Mejora la Privacidad y Seguridad de los Datos: Controla la protección de los datos de los individuos al eliminar o modificar ciertos atributos que lo hacen reconocible por terceras personas lo que contribuye a suprimir el uso de datos con fines malintencionados.	**Compromiso de Utilidad de Datos:** El anonimizar valores o un conjunto de datos supone un desafío entre mantener el equilibrio de protección y su utilidad; por esta razón, al eliminar, suprimir o modificar un valor la calidad de información se ve comprometida para análisis posteriores.
Investigación y Análisis de Datos: El anonimizar datos para el uso exclusivo de personas especializadas con fines investigativos sin comprometer la privacidad individual.	**Riesgo de Reidentificación:** Al emplear técnicas de análisis de datos se presenta el riesgo de identificar aquellos valores que han sido alterados y por ende la información se ve expuesta.
	Costo y Complejidad Técnica: La implementación efectiva de las técnicas de

Facilidad de Compartición de Datos: Dentro del entorno empresarial mantener el anonimato de datos confidenciales es importante ya que asi se puede compartir exclusivamente aquellos que han sido modificados y con esto evitar que terceras personas tengan acceso a información sensible.	anonimización supone un alto costo debido a que se debe contar con especialistas en el manejo y tratamiento de datos los cuales mantengan el compromiso de anonimato efectivo y conforme a los fines pertinentes.
Optimización de los Recursos: El mantener datos protegidos ayuda a optimizar el uso de los recursos debido a que, se suprime la necesidad de controlar de manera rigurosa cada bit de información personal.	**Riesgo de Compartir Datos:** A pesar de tener datos anonimizados persiste el riesgo de que al momento de compartir dicha información con terceras personas no se tiene control sobre su utilidad.

La pseudonimización es una herramienta que difiere de la anonimización, en donde la identificación de la persona se elimina por completo y no puede recuperarse. Como se esquematiza en la Tabla 8 existen beneficios y limitaciones de aplicar este método de protección de datos personales (Gazizov et al., 2020; Hintze & El Emam, 2018).

Tabla 8.

Beneficios y limitaciones de la técnica de pseudonimización de datos

Beneficios	Limitaciones
Mejora la Privacidad y Seguridad de los Datos: Controla la protección de los datos de los individuos al eliminar los identificadores directos por seudónimos; al momento de filtrar la brecha de seguridad se torna más complejo conocer la identidad del individuo al que pertenece la información.	**Claves de Seguridad:** El controlar el manejo de contraseñas o claves de acceso es importante para evitar atracos a la privacidad y por ende los datos sean expuestos a amenazas.

Flexibilidad del Uso de Datos: Los datos pseudonimizados pueden ser recapitulados de forma inversa bajo circunstancias controladas, lo que permite utilizar los valores reales sin ningún tipo de aleatoriedad.	**Complejidad de Implementación:** El utilizar las técnicas de pseudonimización compromete el uso de información técnica y detallada para mantener la coherencia y precisión de los datos en función del tiempo.
Facilidad de Compartición de Datos: La técnica de pseudonimización controla la difusión de los datos a través de identificadores lo que permite que estos sean difundidos sin ningún problema ya que no se expone información confidencial o sensible.	**Limitaciones de la Utilidad de los Datos:** La falta de identificadores específicos que señalen las características propias de un individuo dentro de un conjunto de datos tiende a perjudicar la utilidad de estos.

Conclusiones

Mediante la revisión sistemática con respecto a las diversas técnicas de anonimización y pseudonimización para la protección de datos personales se identificaron, cribaron y evaluaron los estudios en base a los criterios de inclusión y exclusión definidos anticipadamente, dando como resultado un total de 32 artículos científicos. El 83.3% de los documentos sustentan la primera interrogante acerca de las técnicas de anonimización y pseudonimización y sus principales fundamentos; el 6.7% responde a la segunda pregunta y el 10% de los registros contribuyen a establecer los beneficios y limitaciones de las técnicas de protección de datos personales. La investigación se fundamenta en reconocer las técnicas de anonimización y pseudonimización y sus principales métodos de protección para datos personales como: adición de ruido, generalización, cifrado y la descomposición en tokens o alias para garantizar el equilibrio entre utilidad y reserva de información.

Para mantener el equilibrio entre el anonimato y la utilidad del conjunto de datos es necesario seguir una serie de pasos como: identificación y clasificación, desidentificación, aplicación de técnicas de anonimización y pseudonimización, evaluación de los riesgos y planteamiento de medidas de seguridad y políticas de uso de datos. Es importante reconocer los beneficios y limitaciones que ofrece cada una de las técnicas, de forma generalizada una ventaja es potencializar la privacidad y seguridad de los usuarios; por el contrario, las limitaciones están

presentes al momento de compartir los datos con terceras personas ya que no existe control acerca del uso final que se le puede dar a la información. En conclusión, las técnicas de anonimización y pseudonimización constituyen soluciones estratégicas para la privatización de datos personales. Su implementación y la eficacia depende del contexto en el que sean aplicadas ya que son herramientas que no sólo abordan la protección de información desde una perspectiva de privacidad y seguridad, sino que también plantean desafíos con respecto a la integridad de la información y evitar vulnerar los derechos de cada usuario, es decir, el futuro de la protección de datos personales está fundamentado en innovar y adaptar las técnicas conforme a las necesidades personales.

Referencias bibliográficas

Al-Zubaidie, M., Zhang, Z., & Zhang, J. (2019). PAX: Using Pseudonymization and Anonymization to Protect Patients' Identities and Data in the Healthcare System. *International Journal of Environmental Research and Public Health*, *16*(9), 1490. https://doi.org/10.3390/ijerph16091490

Anant, A., & Prasad, R. (2022). Privacy Preservation for Enterprises Data in Edge Devices. *Journal of ICT Standardization*, *10*(1), 85–104. https://doi.org/10.13052/jicts2245-800X.1015

Angelozzi, S. (2020). La gestión de datos de investigación en abierto: introducción al rol emergente para las bibliotecas universitarias y científicas argentinas. *Palabra Clave (La Plata)*, *9*(2). https://doi.org/10.24215/18539912e091

Batko, K., & Ślęzak, A. (2022). The use of Big Data Analytics in healthcare. *Journal of Big Data*, *9*(1). https://doi.org/10.1186/s40537-021-00553-4

Bey, R., Goussault, R., Grolleau, F., Benchoufi, M., & Porcher, R. (2020). Fold-stratified cross-validation for unbiased and privacy-preserving federated learning. *Journal of the American Medical Informatics Association*, *27*(8), 1244–1251. https://doi.org/10.1093/jamia/ocaa096

Bouchelaghem, S., & Omar, M. (2020). Secure and efficient pseudonymization for privacy-preserving vehicular communications in smart cities. *Computers and Electrical Engineering*, *82*. https://doi.org/10.1016/j.compeleceng.2020.106557

Bousquet, J., Agache, I., Almeida, R., Angles, R., Annesi-Maesano, I., Anto, J. M., Arnavielhe, S., Asayag, E., Bacci, E., Bachert, C., Baiardini, I., Baroni, I., Barreto, B. A., Basagana, X., Bedbrook, A., Bedolla-Barajas, M., Bergmann, K. C., Bertorello, L., Bewick, M., … Wittmann, J. (2018). Geolocation with respect to personal privacy for the Allergy Diary app - A MASK study. *World Allergy Organization Journal*, *11*(15). https://doi.org/10.1186/s40413-018-0194-3

Chahal, A., Gulia, P., & Gill, N. (2022). Different analytical frameworks and bigdata model for internet of things. *Indonesian Journal of Electrical Engineering and Computer Science*, *25*(2), 1159–1166. https://doi.org/10.11591/ijeecs.v25.i2.pp1159-1166

De Pascale, D., Cascavilla, G., Tamburri, D. A., & Van Den Heuvel, W. J. (2023). Real-world K-Anonymity applications: The KGEN approach and its evaluation in fraudulent transactions. *Information Systems*, *115*. https://doi.org/10.1016/j.is.2023.102193

Dubagunta, P., Van Son, R., & Magimai, M. (2022). Adjustable deterministic pseudonymization of speech. *Computer Speech & Language*, *72*. https://doi.org/10.1016/j.csl.2021.101284

El Ouazzani, Z., & El Bakkali, H. (2020). A classification of non-cryptographic anonymization techniques ensuring privacy in big data. *International Journal of Communication Networks and Information Security*, *12*(1), 142–152. https://doi.org/10.17762/ijcnis.v12i1.4401

Englert, R., Schedlowski, M., Engler, H., Rief, W., Büchel, C., Bingel, U., & Spisak, T. (2023). ALIIAS: Anonymization/Pseudonymization with LimeSurvey integration and II-factor Authentication for Scientific research. *SoftwareX*, *24*. https://doi.org/10.1016/j.softx.2023.101522

Gazizov, A., Gazizov, E., & Gazizova, S. (2020). Theoretical aspects of the protection of personal data of employees of the enterprise by the method of pseudonymization. *E3S Web of Conferences*, *210*(11001). https://doi.org/10.1051/e3sconf/202021011001

Hintze, M., & El Emam, K. (2018). *Comparing the benefits of pseudonymisation and anonymisation under the GDPR*.

Jiang, H., Pei, J., Yu, D., Yu, J., Gong, B., & Cheng, X. (2023). Applications of Differential Privacy in Social Network Analysis: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, *35*(1), 108–127. https://doi.org/10.1109/TKDE.2021.3073062

Kacha, L., Zitouni, A., & Djoudi, M. (2022). A new k-anonymity approach based on black hole algorithm. *Journal of King Saud University - Computer and Information Sciences*, *34*(7), 4075–4088. https://doi.org/10.1016/j.jksuci.2021.04.014

Karthik, P., & Shanthi Bala, P. (2022). A new design paradigm for provably secure keyless hash function with subsets and two variables polynomial function. *Journal of King Saud University - Computer and Information Sciences*, *34*(5), 1933–1949. https://doi.org/10.1016/j.jksuci.2019.10.003

Karunanayake, I., Ahmed, N., Malaney, R., Islam, R., & Jha, S. K. (2021). De-Anonymisation Attacks on Tor: A Survey. *IEEE Communications Surveys and Tutorials*, *23*(4), 2324–2350. https://doi.org/10.1109/COMST.2021.3093615

Kohlmayer, F., Lautenschläger, R., & Prasser, F. (2019). Pseudonymization for research data collection: is the juice worth the squeeze? *BMC Medical Informatics and Decision Making*, *19*(178), 178. https://doi.org/10.1186/s12911-019-0905-x

Li, C., Chen, Y., & Shang, Y. (2022). A review of industrial big data for decision making in intelligent manufacturing. In *Engineering Science and Technology, an International Journal* (Vol. 29). https://doi.org/10.1016/j.jestch.2021.06.001

Lutfi, A., Alsyouf, A., Almaiah, M. A., Alrawad, M., Abdo, A. A. K., Al-Khasawneh, A. L., Ibrahim, N., & Saad, M. (2022). Factors Influencing the Adoption of Big Data Analytics in

the Digital Transformation Era: Case Study of Jordanian SMEs. *Sustainability (Switzerland)*, *14*(3). https://doi.org/10.3390/su14031802

Mahanan, W., Chaovalitwongse, W. A., & Natwichai, J. (2021). Data privacy preservation algorithm with k-anonymity. *World Wide Web*, *24*(5), 1551–1561. https://doi.org/10.1007/s11280-021-00922-2

Majeed, A., Khan, S., & Hwang, S. (2022). Toward Privacy Preservation Using Clustering Based Anonymization: Recent Advances and Future Research Outlook. *IEEE Access*, *10*, 53066–53097. https://doi.org/10.1109/ACCESS.2022.3175219

Majeed, A., & Lee, S. (2021). Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey. *IEEE Access*, *9*, 8512–8545. https://doi.org/10.1109/ACCESS.2020.3045700

Mao, J., Tian, W., Jiang, J., He, Z., Zhou, Z., & Liu, J. (2018). Understanding structure-based social network de-anonymization techniques via empirical analysis. *EURASIP Journal on Wireless Communications and Networking*, *279*(2018). https://doi.org/10.1186/s13638-018-1291-2

Marques, J., & Bernardino, J. (2020). Analysis of data anonymization techniques. *IC3K 2020 - Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, *2*, 235–241. https://doi.org/10.5220/0010142302350241

Meindl, B., & Templ, M. (2019). Feedback-based integration of the whole process of data anonymization in a graphical interface. *Algorithms*, *12*(191). https://doi.org/10.3390/a12090191

Mohammed, B., Mohd, A., Abdulghani, A., & Ali Safaa, S. (2022). Slicing-Based Enhanced Method for Privacy-Preserving in Publishing Big Data. *Computers, Materials & Continua*, *72*(2), 3665–3686. https://doi.org/10.32604/cmc.2022.024663

Murakami, T., Arai, H., Hamada, K., Hatano, T., Iguchi, M., Kikuchi, H., Kuromasa, A., Nakagawa, H., Nakamura, Y., Nishiyama, K., Nojima, R., Oguri, H., Watanabe, C., Yamada, A., Yamaguchi, T., & Yamaoka, Y. (2023). Designing a Location Trace Anonymization Contest. *Proceedings on Privacy Enhancing Technologies*, *2023*(1), 225–243. https://doi.org/10.56553/popets-2023-0014

Neumann, G., Grace, P., Burns, D., & Surridge, M. (2019). Pseudonymization risk analysis in distributed systems. *Journal of Internet Services and Applications*, *10*(1). https://doi.org/10.1186/s13174-018-0098-z

Ozeran, L., Solomonides, A., & Schreiber, R. (2021). Privacy versus Convenience: A Historical Perspective, Analysis of Risks, and an Informatics Call to Action. In *Applied Clinical Informatics* (Vol. 12, Issue 2, pp. 274–284). Georg Thieme Verlag. https://doi.org/10.1055/s-0041-1727197

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S.,

… Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, *74*(9), 790–799. https://doi.org/10.1136/bmj.n71

Pratama, A. M., & Pati, U. K. (2021). Analysis Principles of Personal Data Protection on COVID-19 Digital Contact Tracing Application: PeduliLindungi Case Study. *Lex Scientia Law Review*, *5*(2), 65–88. https://doi.org/10.15294/lesrev.v5i2.50601

Puluhulawa, F. U., Puluhulawa, J. U., & Katili, G. G. (2020). Legal Weak Protection of Personal Data in the 4.0 Industrial Revolution Era. *Jambura Law Review*, *2*(2), 182–200. https://www.hukumonline.com/berita/baca/lt5d1c3962e01a4/perlindungan-data-pribadi-tersebar-

Ribeiro, S., & Nakamura, E. (2019). Privacy Protection with Pseudonymization and Anonymization in a Health IoT System: Results from OCARIoT. *Proceedings - 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering, BIBE 2019*, 904–908. https://doi.org/10.1109/BIBE.2019.00169

Sánchez, S., Pedraz, I., & Donoso, M. (2022). ¿Cómo hacer una revisión sistemática siguiendo el protocolo PRISMA? *Bordón. Revista de Pedagogía*, *74*(3), 51–66. https://doi.org/10.13042/Bordon.2022.95090

Sarkis-Onofre, R., Catalá-López, F., Aromataris, E., & Lockwood, C. (2021). How to properly use the PRISMA Statement. *Systematic Reviews*, *10*(1), 117. https://doi.org/10.1186/s13643-021-01671-z

Selcuk, A. A. S. AA. A. G. for S. R. PRISMA. T. A. Otorhinolaryngol. 2019;57(1):57–8. (2019). A Guide for Systematic Reviews: PRISMA. *Turkish Archives of Otorhinolaryngology*, *57*(1), 57–58. https://doi.org/10.5152/tao.2019.4058

Slijepčević, D., Henzl, M., Klausner, L., Dam, T., Kieseberg, P., & Zeppelzauer, M. (2021). Anonymity in practice: How generalisation and suppression affect machine learning classifiers. *Computers & Security*, *111*. https://doi.org/10.1016/j.cose.2021.102488

Sweeney, L. (2022). k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(05), 557–570. https://doi.org/10.1142/S0218488502001648

Tahir, H., & Brezillon, P. (2022). A Context Approach to Improve the Data Anonymization Process. *2022 International Conference on Engineering and Emerging Technologies (ICEET)*, 1–6. https://doi.org/10.1109/ICEET56468.2022.10007410

Tepe, B., Haberschusz, D., Figgener, J., Hesse, H., Uwe Sauer, D., & Jossen, A. (2023). Feature-conserving gradual anonymization of load profiles and the impact on battery storage systems. *Applied Energy*, *343*. https://doi.org/10.1016/j.apenergy.2023.121191

Tomás, J., Rasteiro, D., & Bernardino, J. (2022). Data Anonymization: An Experimental Evaluation Using Open-Source Tools. *Future Internet*, *14*(167). https://doi.org/10.3390/fi14060167

Torra, V., & Salas, J. (2019). Graph Perturbation as Noise Graph Addition: A New Perspective for Graph Anonymization. *Data Privacy Management, Cryptocurrencies and Blockchain Technology.*, *11737 LNCS*, 121–137. https://doi.org/10.1007/978-3-030-31500-9_8

Virupaksha, S., & Dondeti, V. (2021). Anonymized noise addition in subspaces for privacy preserved data mining in high dimensional continuous data. *Peer-to-Peer Networking and Applications*, *14*(3), 1608–1628. https://doi.org/10.1007/s12083-021-01080-y

Zhu, T., Ye, D., Wang, W., Zhou, W., & Yu, P. (2022). More Than Privacy: Applying Differential Privacy in Key Areas of Artificial Intelligence. *IEEE Transactions on Knowledge and Data Engineering*, *34*(6), 2824–2843. https://doi.org/10.1109/TKDE.2020.3014246

Conflicto de intereses:
Los autores declaran que no existe conflicto de interés posible.
Financiamiento:
No existió asistencia financiera de partes externas al presente artículo.
Agradecimiento:
N/A
Nota:
El artículo no es producto de una publicación anterior.

Anexos

Tabla de artículos para responder preguntas científicas

N.	Título	Autor (es)	Año	Revista	Cuartil	Resumen
1	Different analytical frameworks and bigdata model for internet of things	Chagal, Ayushi; Culia, Preeti; Gill, Nasib	2022	Indonesian Journal of Electrical Engineering and Computer Science.	Q3	El estudio se centró en cómo los dispositivos IoT generan datos significativos para el análisis de big data, resaltando la creciente importancia del análisis predictivo.
2	Toward Privacy Preservation Using Clustering Based Anonymization: Recent Advances and Future Research Outlook.	Majeed, Abdul; Khan, Safiullah; Hwang, Seong	2022	IEEE Access	Q1	El estudio abordó los mecanismos de anonimización basados en agrupaciones (CAMs), evaluando su efectividad en la preservación de privacidad en diferentes tipos de datos y contextos computacionales.
3	La gestión de datos de investigación en abierto: introducción al rol emergente para las bibliotecas universitarias y científicas argentinas.	Angelozzi, Silvina	2020	La Plata	Q2	El artículo abordó la implementación de políticas de acceso abierto en Argentina, impulsadas por agencias gubernamentales para la publicación y reutilización de datos de investigación.
4	PAX: Using pseudonymization and anonymization to protect patients' identities and data in the healthcare system.	Al-Zubaidie, Mishall; Zhang, Zhongwei; Zhang, Ji.	2019	International Journal of Environmental Research and Public Health	Q2	El estudio abordó las deficiencias de privacidad en los sistemas de registros de salud electrónicos, destacando la necesidad de protección contra accesos no autorizados y privilegios excesivos de usuarios autorizados.
5	Privacy Protection with Pseudonymization and Anonymization in a Health IoT System:	Ribeiro, Sergio; Nakamura, Emilio.	2019	Proceedings - 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering, BIBE 2019	-	El estudio aborda la implementación retrospectiva de un método para proteger la privacidad de los usuarios en un sistema de salud IoT, OCARIoT. La estrategia incluyó capas de seguridad con criptografía y técnicas de anonimato para datos

	Results from OCARIoT					procesados, almacenados y en tránsito, enfatizando el respeto a los derechos de los usuarios.
6	Analysis of data anonymizatio n techniques	Marqués, Joana; Bernardino, Jorge.	2020	IC3K 2020 - Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management	-	El estudio expuso cómo la protección de la privacidad de datos personales, a través de la anonimización y pseudonimización, se alineó con normativas como el GDPR. Se describieron métodos y software que prometen mayor anonimato, evaluando sus fortalezas y debilidades, y la utilidad práctica de cada técnica.
7	A classification of non-cryptographic anonymizatio n techniques ensuring privacy in big data	El Ouazzani, Zakariae; El Bakkali, Hanan.	2020	International Journal of Communicatio n Networks and Information Security	Q3	El artículo trató el auge del procesamiento de Big Data y los consiguientes retos de seguridad y privacidad, proponiendo la anonimización como solución eficaz. Se presentó una clasificación de técnicas de anonimización y se probó su eficacia en un conjunto extenso de datos reales, enfocándose en integridad y confidencialidad.
8	Anonymized noise addition in subspaces for privacy preserved data mining in high dimensional continuous data.	Virupaksha, Shashidhar; Dondeti, Venkatesulu	2021	Peer-to-Peer Networking and Applications	Q2	El estudio demostró un método innovador para la minería de datos que mejora la privacidad y reduce la pérdida de datos e información, facilitando la identificación de clústeres. Al aplicar ruido aleatorio en subespacios, ANAS probó ser más eficaz que métodos previos en conjuntos de datos de referencia.
9	Graph Perturbation as Noise Graph Addition: A New Perspective for Graph Anonymizatio n	Torra, Vicen; Salas, Julián.	2019	Data Privacy Management, Cryptocurrenci es and Blockchain Technology.	-	El artículo propuso un estudio sobre técnicas de perturbación de grafos como adición de ruido en redes sociales y grafos, denominado "noise graph addition", para evaluar su impacto en la protección de datos y analizar sistemáticamente las propiedades de los grafos perturbados.
10	Geolocation with respect to personal privacy for the Allergy Diary app - A MASK study	Bousquet, J; Agache, I; et al.	2018	World Allergy Organization Journal	Q1	El estudio examinó la geolocalización de usuarios del MASK App y evaluó la adición de ruido y k-anonimato para la anonimización. Se descubrió que k-anonimato era más efectivo para proteger la

						privacidad de los usuarios, sugerido como método adecuado para otros bancos de datos.
11	Real-world K-Anonymity applications: The KGEN approach and its evaluation in fraudulent transactions	De Pascale, Daniel; Cascavilla, Giuseppe; Tamburri, Damian; et al.	2023	Information Systems	Q1	El artículo introdujo KGen, una nueva metodología basada en Algoritmos Genéticos para la anonimización K, optimizando el equilibrio entre anonimato y utilidad de datos en grandes conjuntos de datos.
12	Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey	Majeed, Abdul; Lee, Sungchang.	2021	IEEE Access	Q1	El documento proporcionó una revisión detallada de las técnicas de anonimización para la publicación de datos preservando la privacidad, abordando tanto datos tabulares como en grafos.
13	Slicing-Based Enhanced Method for Privacy-Preserving in Publishing Big Data	Mohammed, Jubier; Mohn, Ismail; et al.	2022	Computers, Materials & Continua	Q2	El artículo exploró el uso de técnicas de anonimización de datos en el contexto de la minería de datos, con el objetivo de proteger la privacidad de los individuos.
14	Feature-conserving gradual anonymization of load profiles and the impact on battery storage systems	Tepe, Benedikt; Haberschusz, David; Figgener, Jan¸et al.	2023	Applied Energy	Q1	El estudio abordó la necesidad de anonimizar perfiles de carga eléctrica para proteger la privacidad en la investigación de almacenamiento de baterías.
15	More Than Privacy: Applying Differential Privacy in Key Areas of Artificial Intelligence	Zhu, Tianqing; Ye, Dayong; Wang, Wei; et al.	2021	IEEE Transactions on Knowledge and Data Engineering	Q1	El estudio examinó el uso de la privacidad diferencial en inteligencia artificial para abordar problemas como la privacidad, la seguridad y la equidad de modelos. Se destacó que la privacidad diferencial no solo preserva la privacidad, sino que también mejora la seguridad, estabiliza el aprendizaje y promueve modelos justos.
16	Applications of Differential Privacy in Social Network Analysis: A Survey	Jiang, Honglu; Pei, Jian; Yu, Dongxiao; et al.	2021	IEEE Transactions on Knowledge and Data Engineering	Q1	El estudio exploró el uso de la privacidad diferencial en el análisis de redes sociales para preservar la privacidad y compartir información de manera efectiva. Se examinaron los principios básicos de la

						privacidad diferencial, sus variantes y su aplicación en análisis de redes sociales, incluyendo ataques de privacidad y tareas como el análisis de distribución de grados.
17	Anonymity in practice: How generalisation and suppression affect machine learning classifiers.	Slijepčević, Djordje; Henzl, Maximilian; Klausner, Lukas; et al.	2021	Computers & Security	Q1	El estudio examinó los efectos de diferentes algoritmos de anonimización en los resultados de modelos de aprendizaje automático. Se determinó que, a medida que aumentaba la restricción de anonimización, el rendimiento de clasificación tendía a degradarse, pero variaba según el método y el conjunto de datos.
18	Understanding structure-based social network de-anonymization techniques via empirical analysis	Mao, Jian; Tian, Wenqian; Jiang, Jingbo; et al.	2018	URASIP Journal on Wireless Communications and Networking	Q1	El estudio investigó la desanonimización en redes sociales impulsada por el uso de dispositivos de bienestar. El objetivo fue comprender las técnicas de desanonimización y su impacto en la privacidad de los usuarios en un entorno de redes sociales centradas en el bienestar.
19	A new k-anonymity approach based on black hole algorithm	Kacha, Lynda; Zitouni, Abdelhafid; Djoudi, Mahieddine.	2022	Journal of King Saud University - Computer and Information Sciences	Q1	El estudio abordó el desafío de preservar la privacidad en microdatos utilizando el enfoque de k-anonimato basado en generalización y clustering.
20	k-anonymity: A model for protecting privacy	Sweeney, Latanya	2022	International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems	-	El estudio abordó el desafío de permitir que los titulares de datos privados, como hospitales o bancos, compartan versiones de sus datos con investigadores sin comprometer la privacidad. Se propuso el modelo de protección k-anonimato, donde los datos se modifican de manera que la información de cada individuo en la versión compartida no pueda distinguirse de al menos k-1 individuos más.
21	Data privacy preservation algorithm with k-anonymity	Mahanan, Waranya; Natwichai, Juggapong.	2021	World Wide Web	Q1	El estudio abordó la preservación de la privacidad de datos mediante el método de k-anonimato, especialmente en el contexto de datos con jerarquía de generalización idéntica (IGH).

22	Pseudonymization for research data collection: is the juice worth the squeeze?	Kohlmayer, Florian; Lautenschläger, Ronald; et al.	2019	BMC Medical Informatics and Decision Making	Q1	El estudio abordó la necesidad de proteger los datos sensibles utilizados en la investigación biomédica mediante la pseudonimización. Se destacó que la pseudonimización puede aumentar la complejidad de las soluciones de software, lo que podría crear nuevas vulnerabilidades.
23	Pseudonymization risk analysis in distributed systems	Neumann, Geoffrey; Grace, Paul; et al.	2019	Journal of Internet Services and Applications	Q2	El estudio abordó el desafío de proteger la privacidad de los usuarios en servicios en línea que utilizan datos pseudonimizados. El objetivo era permitir a los desarrolladores explorar estrategias de pseudonimización y optimizarlas en función de métricas como las violaciones de requisitos de privacidad y la utilidad de los datos.
24	Fold-stratified cross-validation for unbiased and privacy-preserving federated learning	Bey, Romain; Goussault, Romain; et al.	2020	Journal of the American Medical Informatics Association	Q1	El estudio abordó la anonimización de los datos de aprendizaje en aspectos médicos.
25	Anonymization/Pseudonymization with LimeSurvey integration and II-factor Authentication for Scientific research	Englert, Robert; Schedlowski, Manfred; et al.	2023	SoftwareX	Q2	El objetivo de estudio era desarrollar un marco de pseudonimización personalizable, llamado ALIIAS, para facilitar el cumplimiento de regulaciones de protección de datos y permitir a los investigadores adaptar la pseudonimización a las necesidades éticas y experimentales de sus proyectos individuales.
26	Feedback-based integration of the whole process of data anonymization in a graphical interface	Meindl, Bernhard; Templ, Matthias.	2019	Algorithms	Q2	Este estudio presentó una aplicación web interactiva que permitía la anonimización de datos sin necesidad de conocimientos en programación. El objetivo era facilitar el proceso completo de anonimización de datos, teniendo en cuenta tanto los aspectos metodológicos como el conocimiento de los especialistas en la materia.
27	A Context Approach to	Tahir, Hassane;	2022	2022 International	-	Se abordó la importancia de proteger los datos personales en

	Improve the Data Anonymization Process	Brezillon, Patrick.		Conference on Engineering and Emerging Technologies (ICEET)		el desarrollo de sistemas de información, especialmente en un contexto de crecientes volúmenes de datos. El objetivo era evitar sanciones, mala publicidad y daños a la imagen de una empresa en caso de una violación de datos.
28	Data Anonymization: An Experimental Evaluation Using Open-Source Tools	Tomás, Joana; Rasteiro, Deolinda;et al.	2022	Future Internet	Q2	En este estudio, se evaluaron dos herramientas de código abierto, ARX Data Anonymization y Amnesia, para la anonimización de datos en un contexto de creciente uso de información personal. El objetivo era determinar cuál de las dos herramientas ofrecía un mejor rendimiento.
29	Adjustable deterministic pseudonymization of speech	Dubagunta,. Pavankumar ; Van Son, Rob; et al.	2022	Computer Speech & Language	Q2	En el estudio presentó un método de pseudonimización reversible para proteger la privacidad de los hablantes en grabaciones de voz sin transcripción. El enfoque alteraba la estructura espectro-temporal del habla para modificar la identidad del hablante mientras preservaba el contenido lingüístico.
30	De-Anonymisation Attacks on Tor: A Survey	Karunanayake, Ishan; Ahmed, Nadeem; et al.	2021	IEEE Communications Surveys and Tutorials	Q1	El estudio examinó ataques conocidos contra la red de anonimato Tor y técnicas para desanonimizar a los usuarios y servicios ocultos. Se evaluó la viabilidad de ejecución de las técnicas y se discutieron posibles mejoras en el marco de Tor para prevenir ataques.
31	Secure and efficient pseudonymization for privacy-preserving vehicular communications in smart cities.	Bouchelaghe, Siham; Omar, Mawloud.	2020	Computers and Electrical Engineering	Q1	El estudio abordó la preocupación por la privacidad de la ubicación en sistemas de transporte inteligente basados en redes vehiculares. Se propuso una estrategia eficiente de cambio de seudónimo que garantiza la privacidad de los usuarios de vehículos, permitiéndoles comunicarse de manera anónima durante sus viajes, pero manteniendo la capacidad de rastrearlos en caso de comportamientos inapropiados.
32	Theoretical aspects of the protection of	Gazizov, Andrey; Gazizov,	2020	E3S Web of Conferences	-	Este estudio señaló que la pseudonimización de datos personales estaba en sus

	personal data of employees of the enterprise by the method of pseudonymization	Evgeny; et al.	primeras etapas de desarrollo en sistemas automatizados. El objetivo era promover la adopción de la pseudonimización como una herramienta importante para la protección de datos a nivel mundial, a pesar de sus limitaciones.