MANUAL DE ANOMIZACIÓN DE DATOS

TABLA DE CONTENIDO

Pág.

1. INTRODUCCIÓN

La información es un activo fundamental de la Escuela Tecnológica Instituto Técnico Central (ETITC) además se debe velar por su seguridad. Como premisa básica desde la seguridad de la información se debe proveer su uso malintencionado. Cuando la información se malversa es necesaria de su control, es ahí cuando recae el concepto de seguridad, por tal motivo la capacidad de recopilar y almacenar información sobre las personas y sus acciones y hábitos es más fácil que nunca. Los avances en la tecnología de la información hacen que el almacenamiento, la catalogación y el uso de dicha información sean triviales. Muchas instituciones educativas han almacenado datos tanto en papel como electrónicos sobre las personas, ya sea mediante la recopilación directa de dichos datos con fines organizativos o datos almacenados como resultado de la prestación de servicios a las personas. Debido a preocupaciones sobre la privacidad, a menudo dichos datos deben anonimizarse antes de ser utilizados o estudiados.

Las instituciones educativas pueden tener varias razones para usar datos anónimos para funciones comerciales, académicas u operativas. Por ejemplo, los datos pueden estar disponibles para uso institucional, sin identificar a los interesados subyacentes, con fines de investigación, estudios de eficacia institucional, estudios operativos y de desempeño, revisiones operativas y de seguridad de la tecnología de la información.
Otros usos de datos no identificados pueden requerir la capacidad de retener identificadores únicos para las personas en el conjunto de datos, sin identificar la identidad real de las personas. Por ejemplo, un investigador puede necesitar saber que ciertas acciones fueron realizadas por la misma persona, para poder llegar a conclusiones sobre cómo las personas usan los datos o el servicio. Un diseñador de sitios web puede querer determinar cuánto tiempo permanecen las personas en el sitio o cómo las personas atraviesan el sitio para encontrar la información buscada.

Los entornos de desarrollo, prueba y capacitación de sistemas pueden requerir el uso de datos que simulen datos de producción reales, mientras que en realidad no consisten en elementos de datos reales, como números de Seguro Social. Si bien la anonimizacion de los datos es un paso útil para proteger la privacidad, los datos des identificados aún pueden conllevar una serie de riesgos de privacidad. Por ejemplo, en algunas situaciones, es posible que las instituciones deban asegurarse de que los datos anónimos o no identificados no se puedan rediseñar para identificar a los interesados subyacentes.

El presente manual se realiza en el marco de la Política de Gestión Estadística de la ETITC, en el proceso de implementación de la NTC PE 1000:2020, dando cumplimiento a los elementos obligatorios, según el Modelo Integrado de Planeación y Gestión (MIPG), y la política estadística, de carácter vinculante para todas las entidades públicas, cuya medición se realiza mediante el Formulario Único de Reporte de Avance en la Gestión (FURAG).

2. OBJETIVOS DEL MANUAL

- Preservar información privada o confidencial mediante la eliminación o codificación de identificadores que vinculan a las personas y los datos almacenados.
- Garantizar que la entidad comprenda y cumpla su deber de proteger los datos sensibles, personales y confidenciales.
- Supresión de identidades de la base de datos restringiría la capacidad de extraer información significativa de los resultados.

3. ALCANCE DEL MANUAL

La razón principal para emprender la anonimización es proteger la privacidad de las personas al poner a disposición los recursos de datos que actividades como la investigación y la planificación dependen. Es legítimo utilizar datos personales para determinados fines, por ejemplo, cuando la intención es informar decisiones sobre individuos en particular, o prestarles servicios. Gran parte de la investigación educativa implica el acceso a los datos personales de los estudiantes y se lleva a cabo sobre la base de consentimiento e implicación. Sin embargo, cuando el uso de datos personales no es necesario, entonces el objetivo debería ser generalmente utilizar en su lugar, datos anonimizados.

4. PÚBLICO OBJETIVO

La intención de este manual es proporcionar información sobre técnicas que podrían aplicarse para anonimizar datos. Esta manual se dirige principalmente a la ETITC que no tiene la intención de divulgar los datos anonimizados al dominio público, pero que comparten datos con otras organizaciones o entidades, donde se pueden imponer controles administrativos y/o técnicos adicionales para reducir el riesgo de divulgación no autorizada de datos personales y que deban dar cumplimiento a la normatividad de protección de datos personales.

5. ¿QUÉ SON LOS DATOS PERSONALES?

Según la Ley 1581 de 2012, un dato personal se define como cualquier información que pueda asociarse a una o varias personas naturales determinadas o determinables. Una persona o individuo puede ser identificado directa o indirectamente a través de su nombre, número de identificación, datos de ubicación, información laboral, entre otros.

6. ¿QUÉ ES ANONIMIZAR DATOS?

La anonimización de datos es el proceso de proteger información privada o confidencial borrando o encriptando identificadores que conectan a una persona con los datos almacenados. Por ejemplo, puede ejecutar información de identificación personal como nombres, números de seguridad social y direcciones a través de un proceso de anonimización de datos que retiene los datos, pero mantiene la fuente en el anonimato.

7. ¿POR QUÉ ES IMPORTANTE EL ANONIMATO DE DATOS?

A través de las amenazas modernas de robo de identidad, fraude con tarjetas de crédito y similares, el anonimato de datos es una forma de proteger la identidad y la privacidad de las personas. Así como proteger la información privada y sensible de las organizaciones.
La anonimización de datos le permite seguir las numerosas leyes de protección de datos que protegen la privacidad del usuario. Estas leyes brindan salvaguardas en torno a la recopilación de datos personales o información de identificación personal, por lo que el anonimato de datos es una buena solución para garantizar que no esté procesando dicha información confidencial.

8. TERMINOLOGÍA

Término	Definición
Anonimización	La conversión de datos personales en "datos anonimizados" aplicando una serie de técnicas de anonimización.
Atributo	También denominado campo de datos, columna de datos o variable. Una información que se puede encontrar en los registros de datos de un conjunto de datos. El nombre, el sexo y la dirección son ejemplos de atributos.
Adversario	Una parte que intenta volver a identificar a una (s) persona (s) de un conjunto de datos que se supone debe ser anonimizado.
Conjunto de datos anonimizados	El conjunto de datos resultante después de la técnica o técnicas de anonimización se ha aplicado en combinación con una evaluación de riesgos adecuada.
Conjunto de datos	Un conjunto de registros de datos. Conceptualmente similar a una tabla en una base de datos relacional típica u hoja de cálculo, que tiene registros (filas) y atributos (columnas).
Conjunto de datos original	El conjunto de datos antes de que se aplique cualquier técnica de anonimización.
Reidentificación	Identificar a una persona a partir de un conjunto de datos anonimizados. La reidentificación espontánea se refiere a la reidentificación no intencionada debido a un conocimiento especial de las personas.
Seudonimización	La técnica de reemplazar un identificador con un valor no relacionado, pero típicamente todavía único

9. TÉCNICAS DE ANONIMIZACIÓN

9.1 Supresión de atributos

Descripción:
La supresión de atributos se refiere a la eliminación de una parte completa de los datos (también denominada "columna" en bases de datos y hojas de cálculo) en un conjunto de datos.

¿Cuándo usarlo?: Cuando no se requiere un atributo en el conjunto de datos anonimizados, o cuando de otro modo, el atributo no se puede anonimizar adecuadamente con otra técnica. Esta técnica debe aplicarse al inicio del proceso de anonimización, ya que es una manera fácil de disminuir la identificabilidad en este punto.

¿Cómo usarlo?: Elimine (p. Ej., Elimine) los atributos o si la estructura del conjunto de datos debe mantenerse, borre los datos (y posiblemente el encabezado). Tenga en cuenta que la supresión debe ser una eliminación real (es decir, permanente) y no solo "ocultar la columna" 5. Del mismo modo, "redactar" puede no ser suficiente si los datos subyacentes siguen siendo algo accesibles.

Otros consejos: Este es el tipo más fuerte de técnica de anonimización, porque no hay forma de recuperar información de tal atributo.
En ciertos escenarios, puede ser posible crear un "atributo derivado" que proporcione utilidad y, sin embargo, sea menos sensible que los atributos originales que, por lo tanto, pueden suprimirse. Por ejemplo, crear un atributo de "duración en las instalaciones" basado en los atributos "fecha y hora de entrada" y "fecha y hora de salida".

Ejemplo:
En este ejemplo, el conjunto de datos consta de puntuaciones de evaluaciones de diferentes materias. Como el destinatario solo necesita analizar las puntuaciones de las pruebas obtenidas por los estudiantes con respecto a sus diversos formadores, pero sin analizar a los propios estudiantes, se eliminó el atributo “estudiante”.

Antes de la anonimización:

Estudiante	Profesor	Resultado
Luis	Milena	4.5
Eduardo	Milena	3.5
Tatiana	Milena	3.8
Pedro	José	2.5
Amparo	José	5.0

Después de suprimir el atributo “Estudiante”

Profesor	Resultado
Milena	4.5
Milena	3.5
Milena	3.8
José	2.5
José	5.0

9.2 Supresión de registros

Descripción: La supresión de registros se refiere a la eliminación de un registro completo en un conjunto de datos. A diferencia de la mayoría de las otras técnicas, esta técnica afecta a múltiples atributos al mismo tiempo.

¿Cuándo usarlo?: Para eliminar registros atípicos que son únicos o no cumplen con otros criterios y no mantener en el conjunto de datos anonimizados. Los valores atípicos pueden llevar a una fácil reidentificación. Puede aplicarse antes o después de que se hayan aplicado otras técnicas (por ejemplo, generalización).

¿**Cómo usarlo?:** Elimina todo el registro. Tenga en cuenta que la supresión debe ser permanente, y no solo una "fila oculta”. Del mismo modo, "redactar" puede no ser suficiente si los datos subyacentes siguen siendo accesibles.

9.3 Enmascaramiento de caracteres

Descripción: El enmascaramiento de caracteres es el cambio de los caracteres de un valor de datos, por ejemplo, por utilizando un símbolo constante (por ejemplo, "*" o "x"). El enmascaramiento es típicamente parcial, es decir, se aplica solo a algunos caracteres del atributo.

¿Cuándo usarlo?: Cuando el valor de los datos es una cadena de caracteres y ocultar parte de él es suficiente para proporcionar el grado de anonimato requerido.

¿Cómo usarlo?: Dependiendo de la naturaleza del atributo, reemplace los caracteres apropiados con un símbolo elegido. Dependiendo del tipo de atributo, puede decidir reemplazar un número fijo de caracteres (por ejemplo, para números de tarjetas de crédito) o un número variable de caracteres (por ejemplo, para direcciones de correo electrónico).

Otros consejos: Tenga en cuenta que es posible que el enmascaramiento deba tener en cuenta si la longitud de los datos originales proporciona información sobre los datos

originales. El conocimiento de la materia es fundamental, especialmente para el enmascaramiento parcial, a fin de garantizar que los personajes correctos estén enmascarados. También se puede aplicar una consideración especial a las sumas de comprobación dentro de los datos; a veces, la suma de comprobación se puede utilizar para recuperar (otras partes de) los datos enmascarados. En cuanto al enmascaramiento completo, el atributo podría suprimirse alternativamente a menos que la longitud de los datos sea de alguna relevancia.

Ejemplo:

Este ejemplo muestra el número de identificación los estudiantes que han sido promovidos al siguiente curso:

Antes de la anonimización

Identificación	Promedio	Curso
1026457389	3.5	8
3647382919	3.8	9
1028374622	4.2	8
2845143827	4.5	10
2647281923	3.9	11

Después de enmascarar el número de identificación del estudiante

Identificación	Promedio	Curso
10XXXXXXXX	3.5	8
36XXXXXXXX	3.8	9
10XXXXXXXX	4.2	8
28XXXXXXXX	4.5	10
26XXXXXXXX	3.9	11

9.4 Seudonimización

Descripción:

La sustitución de datos identificativos por valores inventados. La seudonimización también se denomina codificación. Los seudónimos pueden ser irreversibles, cuando los valores originales se eliminan correctamente y la seudonimización se realizó de manera no repetible, o reversibles (por el propietario de los datos originales), donde los valores originales se guardan de forma segura, pero se pueden recuperar y vincular al seudónimo, si surge la necesidad.

Los seudónimos persistentes permiten la vinculación mediante el uso de los mismos valores de seudónimo para representar al mismo individuo en diferentes conjuntos de datos. Por otro lado, se pueden utilizar diferentes seudónimos para representar al mismo individuo en diferentes conjuntos de datos para evitar la vinculación de los diferentes conjuntos de datos. Los seudónimos también se pueden generar de forma aleatoria o determinista.

¿Cuándo usarlo?: Cuando los valores de los datos deben distinguirse de manera única y donde no Se conservará el carácter o cualquier otra información implícita del atributo original.

¿Cómo usarlo?: Reemplace los valores de atributo respectivos con valores inventados. De una sola mano Para hacer esto, es necesario generar previamente una lista de valores inventados y seleccionar aleatoriamente de esta lista para reemplazar cada uno de los valores originales. Los valores inventados deben ser únicos y no deben tener relación con los valores originales (de manera que se puedan derivar los valores originales de los seudónimos).

Otros consejos: Al asignar seudónimos, asegúrese de no volver a utilizar seudónimos que ya se hayan utilizado (especialmente cuando se generaron al azar). También evite usar exactamente el mismo generador de seudónimos sobre varios atributos, sin un cambio (por ejemplo, al menos use una semilla aleatoria diferente). Los seudónimos persistentes generalmente brindan una mejor utilidad al mantener la integridad referencial en todos los conjuntos de datos.

Para los seudónimos reversibles, la base de datos de identidad no se puede compartir con el destinatario; debe guardarse de forma segura y solo puede ser utilizado por la organización para resolver cualquier consulta específica (sin embargo, el número de dichas consultas debe ser controlado, de lo contrario se pueden utilizar para "decodificar" la seudonimización completa).

De manera similar, si se utiliza cifrado, la clave de cifrado no se puede compartir y, de hecho, debe protegerse de forma segura contra el acceso no autorizado, ya que una fuga de dicha clave podría resultar en una filtración de datos al permitir la reversión del cifrado.

Lo mismo se aplica a los generadores de números pseudo aleatorios, que requieren una semilla. La seguridad de cualquier clave utilizada debe garantizarse como con cualquier otro tipo de cifrado o proceso reversible.

Ejemplo:

Este ejemplo muestra la seudonimización que se aplica a los nombres de los estudiantes que obtuvieron diploma de excelencia y cierta información sobre ellos. En este ejemplo, los nombres se reemplazaron con seudónimos en lugar de suprimir el atributo.

Antes de la anonimización

Estudiante	Grado	Resultado
Luis	8	4.5
Eduardo	9	3.5
Tatiana	9	3.8
Pedro	10	2.5
Amparo	11	5.0

Después de seudonomizar el atributo estudiante

Estudiante	Grado	Resultado
543456	8	4.5
125214	9	3.5
984213	9	3.8
674832	10	2.5
094532	11	5.0

9.5 Generalización

Descripción: una reducción deliberada de la precisión de los datos. Por ejemplo, convertir el de una persona edad en un rango de edad, o una ubicación precisa en una ubicación menos precisa. Esta técnica también se conoce como recodificación.

¿**Cuándo usarlo?:** para valores que pueden generalizarse y seguir siendo útiles para el propósito.

¿Cómo usarlo?: Diseñe categorías de datos y reglas apropiadas para traducir datos. Considere la posibilidad de suprimir cualquier registro que aún se destaque después de la traducción (es decir, la generalización).

Otros consejos: Diseñe los rangos de datos con tamaños apropiados. Los rangos de datos que son demasiado grandes pueden significar que los datos se pueden "modificar" mucho, mientras que los rangos de datos que son demasiado pequeños pueden significar que los datos apenas se modifican y, por lo tanto, aún son fáciles de volver a identificar. Si N-se utiliza el anonimato, el N El valor elegido afectará también a los rangos de datos. Tenga en cuenta que el primero y el último rango pueden ser un rango mayor para acomodar el número típicamente menor de registros en estos extremos; esto a menudo se denomina codificación superior / inferior.

Ejemplo:

En este ejemplo, este conjunto de datos contiene el nombre de la persona (que ya se ha seudonimizado), el curso y el promedio del estudiante.

Estudiante	Grado	Promedio
543456	8	4.5
125214	9	3.5
984213	9	3.8
674832	10	2.5
094532	11	5.0

Para el curso, el enfoque adoptado es generalizar en los siguientes rangos del grado.

<6
6-7
8-9
9-10
10-11
>11

Una vez después de generalización del curso

Estudiante	Grado	Promedio
543456	8-9	4.5
125214	9-10	3.5
984213	9-10	3.8
674832	10-11	2.5
094532	10-11	5.0

9.6 Intercambio

Descripción: El propósito del intercambio es reorganizar los datos en el conjunto de datos de manera que el valor de los atributos individuales todavía está representado en el conjunto de datos, pero generalmente no corresponden a los registros originales. Esta técnica también se conoce como mezcla y permutación.

¿Cuándo usarlo?: Cuando el análisis posterior solo necesita mirar datos agregados, o el análisis está en el nivel del atributo; en otras palabras, no es necesario analizar las relaciones entre atributos a nivel de registro.

¿Cómo usarlo?: Primero, identifique qué atributos intercambiar. Luego, para cada uno, intercambie o reasigne los valores de atributo a cualquier registro en el conjunto de datos.

Otros consejos: Evalúe y decida qué atributos (columnas) deben intercambiarse. Dependiendo de la situación, las organizaciones pueden decidir que, por ejemplo, solo los atributos (columnas) que contienen valores que son relativamente identificativos, necesitan ser intercambiados.

Ejemplo:

En este ejemplo, el conjunto de datos contiene información sobre los registros de estudiantes de una entidad.

Antes de la anonimización

Estudiante	Grado	Resultado
Luis	8	4.5
Eduardo	9	3.5
Tatiana	9	3.8
Pedro	10	2.5
Amparo	11	5.0

En este ejemplo se han intercambiado todos los valores del atributo estudiante:

Estudiante	Grado	Resultado
Eduardo	9	3.5
Tatiana	9	3.8
Luis	8	4.5
Amparo	11	5.0
Pedro	10	2.5

10. METODOLOGÍA DE ANONIMIZACIÓN

La siguiente es una metodología sugerida para realizar la anonimización:

1) Determine el modelo de lanzamiento.

Esto se refiere a cómo se publicará el conjunto de datos anonimizados. Público se refiere a ponerlo a disposición básicamente de cualquier persona. No público se refiere a una liberación controlada a destinatarios conocidos limitados (y a menudo, un número fijo de). El modelo de divulgación pública plantea inherentemente más desafíos a las técnicas de anonimización.

2) Determine el umbral de riesgo de re identificación aceptable, así como la utilidad esperada y el umbral de riesgo previsto o requerido.

Tenga en cuenta que el umbral de riesgo establecido en esta etapa debe distinguirse claramente si los controles adicionales se toman en consideración o solo reflejan el riesgo de los datos.

3) Clasifique los atributos de los datos.

Se trata de clasificar los atributos en el conjunto de datos como identificadores directos, identificadores indirectos o no identificadores, lo que afecta cómo se procesarán posteriormente los atributos.

4) Elimine los atributos de datos no utilizados.

En el proceso de anonimización, generalmente la mayoría de los atributos, ya sean identificadores directos o indirectos, requieren procesamiento o al menos consideración, para que sean menos identificativos. Por lo tanto, debe suprimirse cualquier atributo que claramente no sea necesario en el conjunto de datos anonimizados.

5) Anonimizar identificadores directos e indirectos.

Esto se hace aplicando técnicas como las que se describen en este manual. Se pueden aplicar diferentes técnicas para los tipos de identificadores. Algunas técnicas pueden (y, a menudo, deben) usarse en combinación. Los registros atípicos deben considerarse para la supresión de registros.

6) Realice más anonimización, si es necesario.

Si el riesgo real es mayor que el umbral, se requiere una anonimización “más fuerte” y los pasos 5 a deben realizarse nuevamente con los ajustes necesarios, hasta que el riesgo real sea menor que el umbral.

7) Evalúe la solución.

Esto incluye examinar el conjunto de datos anonimizados para evaluar si la utilidad cumple con el objetivo. Si la utilidad es insuficiente, es posible que sea necesario rediseñar el proceso de anonimización o se puede considerar si la anonimización es factible para este conjunto de datos.

8) Documentar el proceso de anonimización.

Los detalles del proceso de anonimización, los parámetros utilizados y los controles deben registrarse claramente para referencia futura. Dicha documentación facilita la revisión, el mantenimiento, el ajuste y las auditorías. Tenga en cuenta que dicha documentación debe mantenerse de forma segura ya que la publicación de los parámetros puede facilitar el re identificación.

11. EVALUACION DEL RIESGO DE REIDENTIFICACIÓN

Hay varias formas de evaluar el riesgo de reidentificación, y estas pueden implicar cálculos bastante complejos que implican el cálculo de probabilidades.

1) Esta sección describe un modelo simplificado asegurándose supuestos. Uno de los supuestos es que el modelo de publicación no es público. El segundo supuesto es que el ataque intenta vincular a una persona con el conjunto de datos anonimizados. El tercer supuesto es que el contenido de los datos anonimizados no se tiene en cuenta y que el riesgo se calcula independientemente del tipo de información que el atacante tenga realmente disponible.

2) Primero, se debe establecer el umbral de riesgo. Este valor, que refleja una probabilidad, varía entre 0 y 2. Refleja el nivel de riesgo que la ETITC está dispuesta a aceptar. Los principales factores que afectan esto deben incluir el daño que podría causar al interesado, así como el daño a la entidad, en caso de que se produzca una reidentificación; pero también tiene en cuenta qué otros controles se han establecido para mitigar el riesgo en otras formas distintas de la anonimización. Cuanto mayor sea el daño potencial, mayor debe ser el umbral de riesgo. No existen reglas estrictas y rápidas sobre qué valores de umbral de riesgo deben usarse; los siguientes son solo ejemplos:

Daño Potencial	Umbral de Riesgo
Bajo	0.2
Medio	0.1
Alto	0.01

3) Al calcular el riesgo real, este manual explica cómo se analiza el riesgo que supone que el adversario conoce a una persona específica en el conjunto de datos y está tratando de establecer qué registro en el conjunto de datos se refiere a esa persona.

4) La regla simple para calcular la probabilidad de reidentificación de un solo registro en un conjunto de datos es tomar el inverso del tamaño de clase de equivalencia del registro, es decir

P (vincular individuo a un solo registro) = 1 / tamaño de clase de equivalencia del registro

5) Ahora, para calcular la probabilidad de re-identificación de cualquier registro en todo el conjunto de datos, nuevamente, dado que hay un intento de reidentificación, un enfoque conservador sería equipararlo a la probabilidad máxima de reidentificación entre todos los registros del conjunto de datos.

   P (re-ID de cualquier registro en el conjunto de datos) = 1 / Min. tamaño de clase de equivalencia en el conjunto de datos

   Nota: si el conjunto de datos ha sido k- anonimizado, P (re-ID de cualquier registro en el conjunto de datos) <= 1 / k

6) Se puede considerar 3 escenarios de ataque de reidentificación: (1) el ataque interno deliberado; (2) reconocimiento involuntario por parte de un conocido y (3) violación de datos.

   P (re-ID) = P (re-ID | intento de re-ID) x P (intento de re-ID) Donde P (re-ID | intento de re-
   ID)

   Se refiere a la probabilidad de una reidentificación exitosa, dado que hay un intento de reidentificación. Como se observa anteriormente, se puede tomar P (re-ID | intento de re-ID) como (1 / Min. Tamaño de clase de equivalencia en el conjunto de datos) Por lo tanto, P (re-ID) = (1 / Min. Tamaño de clase de equivalencia en el conjunto de datos) x P (intento de re-ID)

7) <u>Para el escenario 1:</u> el ataque interno deliberado, se asume que una parte que recibe el conjunto de datos intenta reidentificarse. Para estimar P (intento de reidentificación), es decir, la probabilidad de un intento de reidentificación, los factores que se pueden considerar incluyen el alcance de los controles de mitigación establecidos, así como los motivos y la capacidad del adversario. La siguiente tabla presenta valores de ejemplo; una vez más, corresponde a la parte que anonimiza el conjunto de datos decidir los valores adecuados a utilizar.

P (intento de re-identificación) para el escenario # 1 - el ataque interno deliberado		Baja motivación y recursos del adversario		
		Bajo	Medio	Alto
Alcance de mitigar controles	Alto	0,03	0,05	0,1
	Medio	0,2	0,25	0,3
	Bajo	0,4	0,5	0,6
	Ninguna	1.0	1.0	1.0

Los factores que afectan la motivación y los recursos del adversario pueden incluir:

- La voluntad de violar el contrato (asumiendo que el contrato previene la re-identificación) está en su lugar
- Limitaciones financieras y de tiempo
- Inclusión de personalidades de alto perfil (por ejemplo, celebridades) en el conjunto de datos Los factores que afectan el alcance de los controles de mitigación incluyen:
- Estructuras organizativas
- Controles administrativos (por ejemplo, contratos)
- Medidas técnicas y físicas

8) Para el escenario n. ° 2: reconocimiento inadvertido por parte de un conocido, asumimos que una parte que recibe el conjunto de datos reidentifica inadvertidamente a un sujeto de datos mientras examina el conjunto de datos. Esto es posible porque la parte tiene algún conocimiento adicional sobre el sujeto de los datos debido a su relación (por ejemplo, amigo, vecino, pariente, colega, etc.). Para estimar P (intento de re-identificación), es decir, la probabilidad de un intento de re-identificación, el factor principal a considerar es la probabilidad de que el receptor de datos conozca a alguien en el conjunto de datos

9) Para el escenario n. ° 3: una violación de datos que se produce en el sistema de TIC del destinatario de los datos, la probabilidad se puede estimar en función de las estadísticas disponibles sobre la prevalencia de las violaciones de datos en la industria del destinatario de los datos. Esto se basa en la suposición de que los atacantes que obtuvieron el conjunto de datos intentarán reidentificarse.

12 CONTROL DE CAMBIOS

FECHA	VERSIÓN	CAMBIOS
18/12/2020	1	Adopción de Documento

ELABORÓ	REVISÓ	APROBÓ
JUAN SEBASTIAN RUIZ BOTIA Líder del proceso de Gestión de Seguridad de la Información	**YANETH JIMENA PIMIENTO C.** Administrador de la documentación	**DORA AMANDA MESA C.** Representante de la Dirección