UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO
DOCTORADO EN COMPUTACIÓN E INFORMÁTICA

# BLOQUEO DE INFERENCIAS A TRAVÉS DE PREFERENCIAS INDIVIDUALES DE PRIVACIDAD

Tesis sometida a la consideración de la Comisión del Doctorado en Computación e Informática para optar al grado y título de Doctorado Académico en Computación e Informática

**LUIS GUSTAVO ESQUIVEL QUIRÓS**

Ciudad Universitaria Rodrigo Facio, Costa Rica

2019

# Dedicatoria

Dios es el guía de mi vida aunque a veces me le salgo del canasto, por ello, a él quiero dedicar la oportunidad de alcanzar este objetivo. Objetivo que también dedico a mi fallecido amigo René Francisco Vargas Gamboa. Le cedo la palabra:

*Te envío este pensamiento desde mi cama, donde me encuentro postrado por la enfermedad del cáncer. Soy una persona joven como tú, tengo muchas ilusiones y proyectos, me acabo de graduar en Ingeniería en Sistemas, pero por Gracia de Dios, obligado por esta enfermedad, he tenido que hacer un alto en el camino de mi vida.*

*Este alto, sumamente doloroso, ha desgastado mi cuerpo y lo ha reducido a tal debilidad, que ni tan siquiera puedo movilizarme sin ayuda hasta para ir al baño. Sin embargo, mi mente se fortalece cada día más y mi alma está sintiendo muy cerca el amor verdadero que viene de nuestro Padre del Cielo.*

*Y éste es el mensaje que te quiero dar a ti hoy: Vive, ama, aprovecha hoy al máximo tu vida. No la desperdicies, no la malgastes, tú que tienes en este momento un cuerpo sano, que te responde y te moviliza donde tú quieres, úsalo para realizar buenos proyectos, que sea un instrumento para construirte cada día, y no para destruirte a ti mismo. Emplea todo lo que tienes y lo que eres en hacer solamente el bien.*

*Yo quisiera tener toda la fuerza física que tú tienes, para hablar personalmente contigo y decirte con mis palabras lo que siento, pero te lo digo así para que sepas que estoy ofreciendo mi enfermedad como oración para que tú ames aún más tu propia vida, y seas cada día una persona superada y totalmente realizada.*

*Gracias por sacar un rato para leer estas líneas, en las cuales pido a Dios que te llene de su bendición y que ésta llegue a toda tu familia, a tus amigos y a todos aquellos que te rodean.*

*Orando, siempre por ti, me despido...*

# Agradecimientos

Mi más sincero reconocimiento a mi familia, amigos y seres queridos, que su fuerza siempre me impulse a seguir adelante. Esta tesis fue posible gracias al apoyo financiero recibido por parte del Ministerio de Ciencia, Tecnología y Telecomunicaciones (MICITT) a través del Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICIT) de la República de Costa Rica. También quiero agradecer a todos mis estudiantes que en múltiples ocasiones y a veces sin saberlo me motivaron durante está investigación. Es importante destacar a mis compañeros profesores, ex-compañeros de estudios, asistentes y ex-asistentes; así como demás personas que de una u otra forma me acompañaron y motivaron durante este proceso. Agradezco mucho el inconmensurable apoyo y colaboración brindado por la Dra. Ileana Castillo Arias quien como coordinadora del programa de doctorado y amiga, siempre demostró un gran interés por la calidad del programa y por mi investigación. Finalmente, agradezco a los tres profesores que me guiaron durante este proceso: el Dr. Fernando Esponda Darlington, el Dr. Ricardo Villalón Fonseca y la Dra. Gabriela Barrantes Sliesarieva, directora de tesis. Gracias por sus increíbles aportes y todo el tiempo invertido para el desarrollo de este trabajo.

"Esta Tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Computación e Informática de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Doctorado Académico en Computación e Informática"

Dr. Luis Alberto Guerrero Blanco
**Representante del Decano**
**Sistema de Estudios de Posgrado**

Dra. Elena Gabriela Barrantes Sliesarieva
**Directora de tesis**

Dr. Ricardo Villalón Fonseca
**Asesor**

Dr. Carlos Fernando Esponda Darlington
**Asesor**

Dra. Gabriela Marín Raventós
**Directora**
**Programa de Posgrado en Computación e Informática**

Luis Gustavo Esquivel Quirós
**Candidato**

# Índice general


Portada . . . . . i
Dedicatoria . . . . . ii
Agradecimientos . . . . . iii
Hoja de Aprobación . . . . . iv
Índice general . . . . . v
Resumen . . . . . ix
Abstract . . . . . x
Índice de Cuadros . . . . . xi
Índice de figuras . . . . . xiii

**1. Introducción** **1**
1.1. Motivación . . . . . 3
1.2. Objetivos . . . . . 7
1.2.1. Objetivo General . . . . . 7
1.2.2. Objetivos específicos . . . . . 7
1.3. Justificación . . . . . 7
1.4. Alcances y limitaciones . . . . . 11
1.5. Antecedentes . . . . . 12
1.5.1. Modelos de preservación de la privacidad . . . . . 12
1.5.2. Preservación de la privacidad personalizada . . . . . 13
1.5.3. Negociación de la privacidad . . . . . 13
1.5.4. Métricas de preservación de la privacidad . . . . . 14
1.5.5. Inferencias siendo utilizadas con fines maliciosos . . . . . 14
1.6. Trabajo relacionado . . . . . 15
1.7. Organización de la tesis . . . . . 17

**2. Marco teórico** **19**
2.1. Privacidad . . . . . 19
2.1.1. Autodeterminación informativa . . . . . 21
2.1.2. Consentimiento informado . . . . . 21
2.2. Preservación de la privacidad de datos a publicar . . . . . 24
2.2.1. Operaciones de preservación de la privacidad . . . . . 26
2.2.2. El k-anonimato . . . . . 29
2.2.3. Métricas de privacidad . . . . . 33

**3. Metodología** **41**
3.1. Definición de un modelo conceptual de expresión de preferencias individuales de privacidad . . . . . 42
3.2. Instrumentación de un bloqueo de inferencia de datos en un algoritmo de desidentificación . . . . . 42
3.3. Definición y evaluación de una métrica de bloqueo de inferencias . . . . 42
3.4. Desarrollar un caso de estudio . . . . . 43
3.5. Evaluación del modelo de expresión de inferencias . . . . . 43

**4. Modelo de expresión de inferencias** **44**
4.1. Selección de inferencias aplicables a una base de datos . . . . . 45
4.2. Traducción de inferencias a reglas y elementos interpretables en una base de datos . . . . . 47
4.3. Representación y uso de las preferencias de privacidad de los dueños de los datos . . . . . 54
4.4. Resumen . . . . . 58

**5. Bloqueo de inferencia de datos** **60**
5.1. Aplicación de bloqueo de preferencias individuales a métodos de desidentificación . . . . . 60
5.2. Implementación de modificación al k-anonimato . . . . . 62
5.3. Resumen . . . . . 64

**6. Métrica de bloqueo de inferencia de datos** **65**
6.1. Métrica de privacidad propuesta . . . . . 65

6.1.1. Selección del algoritmo de aprendizaje y su aplicación . . . . . 67
6.1.2. Interpretación de la predicción . . . . . 67
6.1.3. Ejemplo de aplicación . . . . . 68
6.1.4. Análisis de resultados . . . . . 69
6.2. Resumen . . . . . 75

**7. Caso de estudio** **77**
7.1. Definición del contexto de estudio . . . . . 77
7.2. Consultas típicas . . . . . 78
7.3. Selección de inferencias posibles en la base de datos y su representación en lenguaje natural . . . . . 79
7.4. Traducción de inferencias a reglas y elementos interpretables en la base de datos estudiantil . . . . . 82
7.5. Consulta a publicar y conjunto de datos a desidentificar . . . . . 84
7.6. Simulación de las preferencias individuales sobre el bloqueo de inferencia 86
7.7. Resumen . . . . . 87

**8. Evaluación del modelo de expresión de inferencias** **88**
8.1. Metodología de evaluación . . . . . 88
8.1.1. Preguntas de investigación que guían la evaluación . . . . . 89
8.1.2. Variables de respuesta . . . . . 89
8.1.3. Factores de diseño . . . . . 90
8.1.4. Factores fijos . . . . . 92
8.1.5. Otros aspectos de diseño . . . . . 92
8.1.6. Instrumentación de k-anonimato . . . . . 93
8.2. Resultados y Análisis . . . . . 94
8.3. Resumen . . . . . 99

**9. Conclusiones** **101**
9.1. Contribuciones . . . . . 103
9.1.1. Trabajo futuro . . . . . 104

**Bibliografía** **106**

**Apéndices** **114**

**A. Encuesta de grado de privacidad** **116**

# Resumen

Esta tesis presenta el uso de inferencias como expresión de una posible afectación de la privacidad, permitiendo expresar en lenguaje natural la posible violación a la misma. Al mismo tiempo, permite presentar al dueño de los datos los riesgos de privacidad a los que se ve expuesto desde una perspectiva simple, informativa y consistente, permitiendo una elección informada para restringir el uso de sus datos. Este uso se llevó a la aplicación práctica mediante la definición de un modelo de expresión de preferencias individuales de privacidad sobre inferencias posibles en una base de datos de una organización. Además se definió la instrumentación de un bloqueo de inferencia aplicado a un algoritmo de desidentificación y se definió una métrica que identifico la eficacia del bloqueo de inferencia. Para probar lo anteriormente descrito se desarrolló un caso de estudio que demostró su aplicación práctica. Los resultados evidencian que se logró identificar claramente el efecto positivo de tomar en cuenta las preferencias mediante la evaluación del modelo de expresión de inferencias aplicado al caso de estudio por medio de un experimento.

# Abstract

This thesis presents the use of inferences as an expression of a possible affectation of privacy, allowing to express in natural language the possible violation of it. At the same time, it allows to present to the data owner the privacy risks to which he is exposed from a simple, informative and consistent perspective, allowing an informed choice to restrict the use of his data. This use was brought to practical application by defining a model of expressing individual privacy preferences about possible inferences in a database of an organization. In addition, the instrumentation of an inference block applied to a deidentification algorithm was defined and a metric that identified the effectiveness of the inference block was defined. To test the practical application of the model, a case study was developed and evaluated trough an experiment. The results show the positive effect of applying user preferences to de-identification processes.

# Índice de cuadros

4.1. Matriz de preferencias de privacidad sobre bloqueos a inferencias. . . . 55
4.2. Marcado de tuplas por la inferencia "Deducir su horario de clases en este semestre" . . . . . 58

7.1. Caracterización de base de datos de la UCR . . . . . 78
7.2. Caracterización de consultas SQL utilizadas por la ORI para producir los conjuntos de datos respuesta . . . . . 79
7.3. Taxonomía de datos de estudiantes . . . . . 80
7.4. Columnas del conjunto de datos a desidentificar y su descripción. . . . 86

8.1. Resumen de factores de diseño . . . . . 91
8.2. Valores utilizados de $k$ y sus respectivos alfas de Lodha. . . . . 94

# Índice de figuras

1.1. Escenario de publicación de datos . . . . . 3
1.2. Bloqueo de inferencias. . . . . 9
1.3. Reidentificación a un estudiante. . . . . 10

2.1. Fases y actores en la PPDP . . . . . 25
2.2. Jerarquía de generalización de valores (JGV) con las clases de equivalencia para números de 0 a 99. . . . . 27
2.3. Jerarquía de generalización de valores (JGV) con las clases de equivalencia para el atributo categórico estado civil . . . . . 28
2.4. Clasificación de modelos de preservación de la privacidad (Fuente: [Goswami and Madan, 2017]) . . . . . 29

3.1. Diagrama de pasos descritos en la metodología . . . . . 41

4.1. Selección de inferencias. . . . . 46
4.2. Traducción de inferencias de datos a reglas en la base de datos de estudio 48
4.3. Una de las consultas SQL que permite ejecutar la inferencia “Deducir su horario de clases en este semestre”. . . . . 49
4.4. Regla para la inferencia: “Deducir su horario de clases en este semestre” 51
4.5. Traducción de inferencias a reglas. . . . . 52
4.6. Activación de bloqueos de inferencias. . . . . 53
4.7. Ejemplo de una interfaz de usuario para activación de bloqueo de inferencias. . . . . 54
4.8. Activación de bloqueos de inferencias utilizando la consulta SQL y las preferencias de privacidad. . . . . 56
4.9. Consulta ejemplo de publicación. . . . . 57

5.1. Árbol de generalización de datos que permite construir los grupos que cumplen con las restricciones de privacidad. . . . . 63

6.1. Evaluación de la precisión en la predicción de la variable “race”. . . . . 70
6.2. Evaluación de la precisión en la predicción de la variable “income”. . . 71
6.3. Relación entre el nivel de privacidad k, el número de cuasi-identificadores y el valor obtenido por la métrica GCP. . . . . 72
6.4. Evaluación de la precisión en la predicción de la variable “age”. . . . . 74

7.1. Frecuencia de palabras en sentencia SQL . . . . . 79
7.2. Respuestas a encuesta . . . . . 81
7.3. Consultas SQL que permiten ejecutar la inferencia “Deducir el lugar de residencia de un estudiante”. . . . . 83
7.4. Reglas para la inferencia: “Deducir el lugar de residencia de un estudiante” . . . . . 84
7.5. Consulta SQL que permiten ejecutar sobre los datos de admisión. . . . 85

8.1. Porcentaje promedio de estudiantes re-identificados para cada valor de $k$ base. Las curvas azul y rojo muestran los porcentajes de reidentificación antes y después de la aplicación de preferencias. . . . . 95
8.2. Porcentaje promedio de estudiantes correctamente reidentificados que escogieron tener más privacidad para cada valor de $k$. . . . . 96
8.3. Diferencia apareada de las medias, antes y después de la aplicación de las preferencias, intervalo de confianza a 99 %. . . . . 98
8.4. Costo en pérdida de utilidad de la aplicación de desidentificación instrumentada y sin instrumentar, con 5 % y 10 % de datos de entrenamiento. Importante: el eje Y va de 0 a 0.25 para poder visualizar las diferencias entre las curvas, que son mínimas. La notación “con/sin PA” se refiere al uso o no de $k$-anonimato instrumentado. Los números 5 y 10 corresponden a los porcentajes de datos de entrenamiento . . . . . 99

UNIVERSIDAD DE COSTA RICA

SEP Sistema de Estudios de Posgrado

**Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.**

Yo, Luis Gustavo Esquivel Quirós, con cédula de identidad 1-1059-0024, en mi condición de autor del TFG titulado Bloqueo de Inferencias a través de preferencias individuales de privacidad

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. **SI** [X] **NO** * [ ]

***En caso de la negativa favor indicar el tiempo de restricción:** _____ **año (s).**

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

**INFORMACIÓN DEL ESTUDIANTE:**

Nombre Completo: Luis Gustavo Esquivel Quirós

Número de Carné: _____ Número de cédula: 1-1059-0024

Correo Electrónico: luis.esquivel@ucr.ac.cr

Fecha: 30/08/2019 Número de teléfono: 8817-8072

Nombre del Director (a) de Tesis o Tutor (a): Elena Gabriela Barrantes



**FIRMA ESTUDIANTE**

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

# Capítulo 1

# Introducción

El uso de Internet en el mundo ha modificado la forma de recopilar, intercambiar y manipular datos. Las velocidades de procesamiento, los volúmenes de información y la relevancia de los contenidos han cambiado rápidamente hasta niveles que la persona promedio no visualiza correctamente [Henriksen-Bulmer and Jeary, 2016]. Esto se ve asociado al interés en crear servicios personalizados basados en la información disponible. Diferentes organizaciones tienen datos sobre las personas y estos datos son un elemento vital en la creación de nuevos servicios. Se estima que hasta el 80 % de todos los datos almacenados en las organizaciones pueden clasificarse como grandes datos (*Big Data*) [Khan et al., 2014], esto es, conjuntos de datos cuyo tamaño supera la capacidad de las herramientas disponibles para llevar a cabo su adquisición, acceso, análisis o uso en un período de tiempo razonable [Tien, 2013]. Esto repercute directamente en las opciones disponibles de protección y seguridad de parte de las organizaciones que los almacenan.

Las organizaciones que almacenan los datos los utilizan de muchas formas, por ejemplo, para promover sus negocios e intereses o unir información de otras fuentes para generar nuevo contenido [Chaudhuri et al., 2011]. En Costa Rica, al igual que en otras partes del mundo, existe toda una revolución alrededor de la publicación de datos abiertos (*Open Data*). Los llamados a la transparencia, el control democrático y la participación ciudadana, han generado que individuos y organizaciones liberen sus datos y los pongan a disposición de terceros, en muchos casos sin restricciones de derechos u otros mecanismos de control que permitan delimitar su uso [Open Knowledge, 2016, República de Costa Rica, 2015].

Una publicación de datos en la mayoría de los casos contiene información de múltiples individuos o dueños. Esto es aprovechado por organizaciones que utilizan análisis de datos (*data mining*) para descubrir aspectos sobre clientes e identificar tendencias [Henriksen-Bulmer and Jeary, 2016]. Desde una perspectiva individual, esto plantea preguntas acerca de cuánto conocimiento puede ser recabado acerca de la vida de

un individuo, por ejemplo, sobre su situación actual o su paradero, [Ohm, 2010]. así como el uso que le dan las organizaciones a ese conocimiento, por ejemplo, el uso no autorizado a información personal identificable (PII, por sus siglas en inglés) de más de 87 millones de usuarios desprevenidos de Facebook en el caso de Cambridge Analytica [Isaak and Hanna, 2018].

La publicación de datos, una vez realizada, puede tener consecuencias muy variadas para el dueño o productor de los datos, exponiendo al individuo a discriminación o escarnio público, entre otros [Di Caterino, 2016]. Existen ejemplos muy renombrados al respecto, como el trabajo donde investigadores lograron identificar usuarios conocidos de forma exitosa de los registros de Netflix publicados para un concurso de optimización de recomendaciones. Estos investigadores descubrieron aparentes preferencias políticas y otra información potencialmente confidencial [Narayanan and Shmatikov, 2006]. También es común escuchar del caso AOL, dónde luego de una publicación de búsquedas de usuarios, un periodista logró identificar a una anciana y generó gran escandalo social [Barbaro and Zeller, 2006, Leoni, 2012]. Estos aspectos plantean serias preocupaciones sobre como proteger a cada individuo y el potencial efecto de la falta de protección.

La acción mediante la cual se revela información de una persona sin su consentimiento es conocida como violación de la privacidad y en muchos casos provoca una afectación de índole individual, aún cuando la publicación sea grupal [Solove, 2006]. La afectación como se indicó anteriormente, puede incluir desde la molestia hasta la persecución sobre los dueños de los datos [Greenwald et al., 2013, Aimeur et al., 2012, Griffin, 2012]. William Parent [Parent, 1983] plantea que la privacidad corresponde al escenario dónde el dueño de los datos puede determinar qué información desea revelar. En este escenario lo privado corresponde a la información que no se desea revelar. Surge entonces la pregunta de cómo atender los deseos de cada persona respecto a la Preservación de la Privacidad en Datos a Publicar (PPDP, por sus siglas en inglés). Una descripción completa de PPDP se desarrolla en el capítulo 2.2, pero debe quedar claro que los deseos respecto a qué hacer público o no, es en el fondo una preferencia de privacidad de los dueños de los datos.

La búsqueda de una protección adecuada para la privacidad de los dueños de los datos constituye un trabajo continuo. Este trabajo continuo se ve reflejado en la diversidad de investigación científica [Agrawal and Srikant, 2000, LeFevre et al., 2006b, Machanavajjhala et al., 2007, Shah, 2012, Garfinkel, 2015] y

en el desarrollo de jurisprudencia a nivel mundial [República de Costa Rica, 2011, República de Costa Rica, 2013, El Parlamento Europeo, 2016]. Como consecuencia de estas regulaciones, se debe potenciar el desarrollo de mecanismos, técnicas y modelos para proteger a los individuos respecto a lo que se puede o no revelar de ellos. En este contexto general es que se desarrolla el trabajo de la presente tesis. A continuación, se presentan los elementos que llevan a la definición de los objetivos.

## 1.1. Motivación

Un escenario de publicación de datos tiene diversos actores y roles que pueden ser explicados a partir de la figura 1.1. Por ejemplo, supóngase que en una universidad (**organización que custodia los datos**), se almacenan los datos académicos y personales de sus estudiantes (**dueños de los datos**). La universidad responde frecuentemente solicitudes de distintos tipos de información (en su rol de **publicador de datos**), de parte de entes internos y externos (**solicitantes de datos**). Para responder a estas solicitudes, en lo que concierne a los estudiantes, la universidad consulta sus bases de datos y, primero consigue los datos solicitados, para luego filtrarlos en función de las preferencias de privacidad de los estudiantes, armando una respuesta para ser publicada ("entregando" la información solicitada al solicitante o **receptor de los datos**).

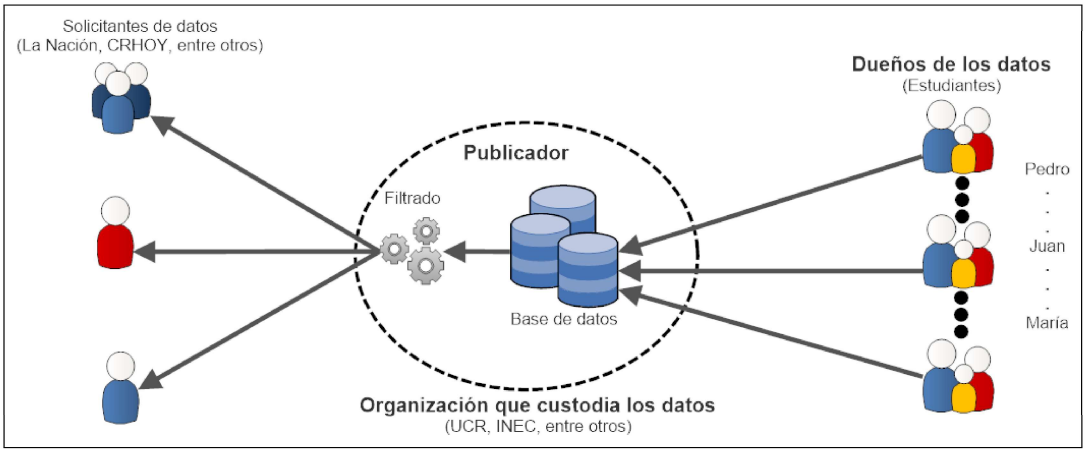



Figura 1.1: Escenario de publicación de datos

Las preferencias de privacidad deben ser recopiladas mucho antes de su utilización

en el filtrado. Esto se debe a que la mayoría de los datos almacenados se refieren a algún estudiante, y todo estudiante tiene varias diferentes piezas de datos almacenadas en la base de datos. Los estudiantes son en este caso a quienes se les deben respetar las preferencias de privacidad (en su rol de **usuarios**). Como se explica en el marco teórico (capítulo 2), existe normativa, tanto nacional como internacional que define los alcances de la posesión de datos por parte de los dueños en bases de datos, tanto públicas como privadas. Asimismo, existe normativa que regula a los publicadores de datos respecto a sus responsabilidades.

Las publicaciones de datos deben ser cuidadosamente filtradas para evitar revelar información sensible (religión, filiación política, ingresos, entre otros). Para esto, la universidad en el caso de la figura 1.1, debe utilizar algún método de Preservación de la Privacidad en Datos a Publicar (PPDP), que puede ir desde algo tan básico como ocultar los nombres o números de cédula de los estudiantes, hasta enfoques mucho más complejos, como los que se describen en el capítulo 2.

En la búsqueda de proveer un filtrado al dueño de los datos, se desea construir una forma de pedirle, preferiblemente en lenguaje natural, sus preferencias de privacidad de una forma que tenga sentido para una persona sin conocimientos especializados. Esta forma debe permitir vislumbrar los alcances de una posible publicación, sin abrumar con una multitud de datos técnicos, ni redundar en la clásica opción única de "acepte todo" para poder continuar. Se debe tener presente que las preferencias de privacidad son algo muy personal, y se deben solicitar a los dueños de los datos a través del consentimiento informado (como se explicará en el capítulo 2).

Sin embargo, un consentimiento (efectivamente) informado es realmente difícil de construir [Geest et al., 2005]. Por un lado están las formas más descaradas de (no) capturarlo, tales como las que vienen con algunos tipos de software que se limitan a decir "sí" o "no" a todo, que son simplemente una forma de no permitirle al usuario el control sobre sus datos. Por otro lado, existen propuestas que literalmente le permiten al usuario elegir cuáles campos de la base de datos prefiere "compartir" y cuáles no. Dichas propuestas resultan poco útiles, dada la complejidad de cualquier base de datos actual y la poca intuición que un dueño de datos común puede tener sobre el posible impacto de bloquear o no la publicación de la mayoría de los campos de la base de datos. Todo esto sin tomar en cuenta cuáles de los datos son sensibles y que probablemente el publicador de los datos los puede identificar fácilmente.

El método de filtrado más utilizado para la preservación de la privacidad es conocido como desidentificación, que comprende una colección de enfoques, algoritmos y herramientas que se aplican a diferentes tipos de datos para proteger la privacidad en los individuos o entidades involucradas en esos datos [Garfinkel, 2015]. La desidentificación puede ser vulnerada al identificar a un dueño de datos y/o enlazarlo con su información sensible, produciendo una violación de la privacidad. En general, al proceso que permite revertir la desidentificación se le conoce como **reidentificación**. Existen múltiples ocurrencias reales de reidentificación, con casos publicados principalmente sobre datos de personas de los Estados Unidos [Narayanan and Shmatikov, 2008, El Emam et al., 2011]. En términos de protección de la privacidad, la reidentificación es precisamente lo que se debe evitar.

Entre otros métodos, la reidentificación ocurre típicamente por medio de inferencias sobre múltiples conjuntos de datos [Malin et al., 2003]. Una inferencia, se compone de un conjunto de técnicas de extracción de información, que mediante el análisis de datos obtiene en algunos casos sin autorización conocimiento sobre un tema, individuo o base de datos [Krumm, 2007]. Por lo tanto, la motivación del presente trabajo es tratar de bloquear la construcción de inferencias. La inferencia es de especial relevancia, dado que el conocimiento extraído puede permitir relacionar la identidad del dueño de los datos con un conjunto o subconjunto de datos publicado. Por ejemplo, la inferencia puede permitir relacionar y revelar información sensible respecto a preferencias sexuales, religión o enfermedades. El tipo de afectación no depende de la identificación unívoca del dueño original de los datos o del exponer una correlación con sólo algunos de los registros. La afectación proviene del posible uso sin consentimiento que se da sobre los datos publicados y por lo tanto se presenta una violación de la privacidad.

La propuesta es utilizar la presentación de bloqueos a inferencias sobre información de los dueños de los datos, como el medio para expresar las preferencias de privacidad de los mismos. Esto tiene la ventaja de que al mismo tiempo que se está especificando la preferencia, se le está advirtiendo al dueño de los datos sobre las posibles consecuencias (realizar la inferencia). Las inferencias van a estar expresadas en lenguaje natural, por ejemplo: “Descubrir a qué hora salgo hacia la universidad”, “Encontrar cuántos años duré sacando la carrera”, “Localizar mi lugar de residencia” y el permitirlas o no, constituye la expresión del consentimiento informado.

Es importante señalar que una especificación de preferencias de privacidad por me-

dio de la presentación de bloqueos de inferencias en lenguaje natural no significa nada para efectos de una base de datos, es simplemente una forma de representación que intenta mejorar la comprensión sobre los peligros inherentes a una publicación. Esta representación en lenguaje natural del peligro latente, se debe mapear por medio de un método (manual) a una especificación en términos de reglas sobre los elementos la base de datos del publicador. Estas reglas sobre elementos de la base de datos (BD) son requeridas únicamente cuando ingresa una solicitud de publicación de datos y constituyen parte de los insumos utilizados en el filtrado. En ese momento se debe construir una consulta para obtener los datos de la BD respectiva, y de paso, se coteja esta consulta contra las reglas almacenadas que mapean a las inferencias. Si un porcentaje "grande", definido por un umbral, calza contra las reglas, se asume que la publicación podría llevar a facilitar la (o las) inferencia(s) que se han definido. Asimismo, si en el grupo de datos a ser publicado, se encuentran tuplas de alguno(s) de los dueños de datos con preferencias de privacidad que indiquen el bloqueo de la inferencia, entonces se debe realizar el filtrado respectivo para respetar la preferencia de privacidad del dueños de los datos.

En la búsqueda por respetar las preferencias o deseos de privacidad del dueño de los datos, si el dueño indica que el bloqueo debe ser activado, la inferencia se debe dificultar o evitar (indicando el grado de cumplimiento), pero en caso contrario, no debería realizarse ninguna alteración de los datos a publicar. La preservación de la privacidad requiere del estudio del equilibrio entre el respeto a los deseos o preferencias de privacidad de múltiples dueños de datos ante una posible inferencia autorizada o no, y la posibilidad de la reidentificación de cada dueño o el enlace de información sensible sobre el conjunto de datos publicado. El estudio de ese equilibrio permite que surjan diversas preguntas sobre la PPDP y el respeto a los deseos o preferencias de los dueños de los datos. Una de esas posibles preguntas y que dirige esta investigación, expresa: ¿Es viable utilizar las preferencias de privacidad individuales sobre inferencias posibles a los datos de un individuo, de forma que permita mejorar procesos de desidentificación?

La respuesta a esta pregunta es significativamente difícil de encontrar, como se verá en el capítulo 2. Sin embargo, para guiar la construcción de una respuesta se desarrollaron los objetivos que son presentados a continuación.

## 1.2. Objetivos

Esta sección presenta el objetivo de esta tesis doctoral. En la sub-sección 1.2.1 se describe el objetivo general de la tesis, seguido por la descripción de cada uno de los objetivos específicos que se presentan en la sub-sección 1.2.2.

### 1.2.1. Objetivo General

El objetivo general de esta tesis es incorporar preferencias individuales de privacidad basadas en inferencias, a procesos de desidentificación de datos de consultas en bases de datos relacionales.

### 1.2.2. Objetivos específicos

Los objetivos específicos que se definieron se presentan a continuación.

1. Definir un modelo para expresar posibles inferencias sobre consultas a una base de datos relacional que permitan tomar en cuenta las preferencias individuales de privacidad.

2. Desarrollar un caso de estudio que incorpore un conjunto de inferencias especificadas con el modelo de expresión definido, en un modelo de desidentificación conocido.

3. Definir una métrica que identifique la eficacia de un bloqueo de inferencia de datos para un contexto particular.

4. Evaluar la eficacia del proceso de desidentificación del caso de estudio desarrollado utilizando la métrica definida.

En la siguiente sección 1.3, se desarrollan las razones que motivan la realización de este trabajo de tesis y su pertinencia en el contexto de nuestro país.

## 1.3. Justificación

La publicación de información en empresas públicas y semipúblicas es cada vez más necesaria cuando estas desean cumplir con estrategias de Gobierno Abierto

[Conradie and Choenni, 2012a]. En este sentido, toda organización debe solventar estas necesidades y hacer frente a nueva jurisprudencia que protege a los dueños de los datos que almacena [República de Costa Rica, 2011].

En el caso de Costa Rica se ha adquirido jurisprudencia y experiencia en temas de privacidad de los datos con el fin de procurar garantías y mantener la equidad de los derechos y responsabilidades de los dueños de los datos. La jurisprudencia creada en Costa Rica incluye la Ley de Delitos Informáticos [República de Costa Rica, 2001], la Ley de protección de la persona frente al tratamiento de sus datos personales [República de Costa Rica, 2011] y el Reglamento de la Ley de Protección de la Persona frente al tratamiento de sus datos personales [República de Costa Rica, 2013].

Estos avances jurídicos, entre otros elementos, como la Agencia de la Protección de los Datos de los Habitantes (PRODHAB), fomentan que los desarrollos científicos alrededor del tema de la preservación de la privacidad y el ejercicio de los deseos o preferencias de los dueños de los datos, sean de gran importancia en el país y a nivel mundial, donde este tipo de legislaciones han sido tema de estudio y desarrollo [Scholl et al., 2008, Gopalan et al., 2012, Wu et al., 2012].

Las regulaciones en Costa Rica dan énfasis a los derechos de los dueños de los datos y por lo tanto son los dueños quienes deberían definir lo que se realiza con sus datos, preferiblemente a través de instrumentos de fácil uso. Por tanto, existe una gran necesidad de trabajar en formas para ejercer las preferencias de privacidad de los dueños de los datos, que sean utilizables mediante un proceso de desidentificación.

Se plantea el uso de inferencias en la expresión de las preferencias de privacidad de los dueños de los datos utilizando como fundamento el almacenar solo el complemento de la información requerida, esto como una forma de presentación negativa [Esponda et al., 2007] del potencial uso de los datos facilitados. Para ello, como se muestra en la figura 1.2 se busca bloquear las inferencias no deseadas por los dueños de los datos, en lugar de intentar almacenar todos los usos con los que el dueño de los datos este de acuerdo.

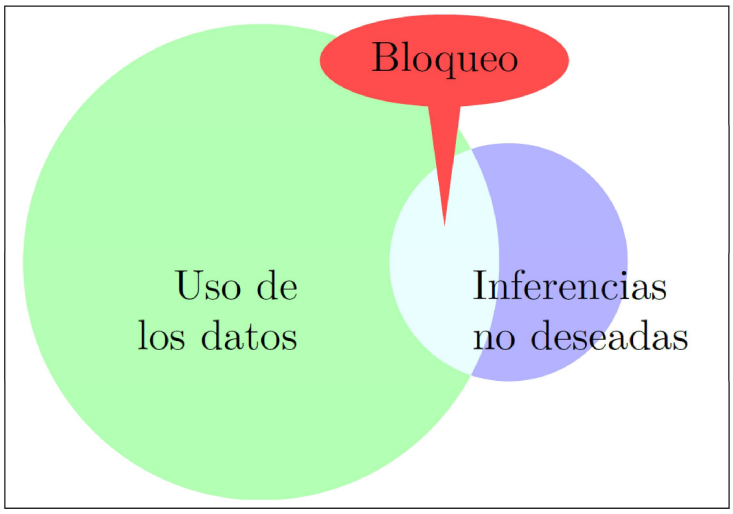



Figura 1.2: Bloqueo de inferencias.

Esto permite al dueño de los datos expresar lo que no desea se realice con sus datos. En otras palabras, permite expresar la negación al potencial uso de los datos. La negación facilita determinar qué usos son considerados una violación de la privacidad. También, permite considerar autorizado cualquier otro uso. Un ejemplo de este tipo de inferencias se muestra en la figura 1.3, que presenta los pasos de un caso hipotético de inferencia realizada sobre un estudiante de la Universidad de Costa Rica (UCR). En este caso una persona atacante desea localizar los lugares donde acostumbra estar en la UCR el estudiante Vladimir Aguilar. Inicialmente el atacante solo cuenta con el primer nombre, un apellido y un rango de edad posible. La secuencia de pasos posibles para lograr la inferencia de la figura 1.3 son:

(A) Tomar el nombre, apellido y rango de edad para realizar una búsqueda en la Base de Datos del Tribunal Supremo de Elecciones. Al realizar esta búsqueda se encuentra que solo tres personas cumplen con todas las características solicitadas.

(B) Una vez obtenida la lista de nombres completos procedente del paso (A), se puede realizar una búsqueda sobre el Padrón Electoral Estudiantil de la UCR.

(C) La búsqueda del paso (B) permite identificar que solo una persona de la lista del paso (A) es estudiante activo de la UCR. Producto de esta asociación se obtiene la carrera y carné del estudiante.

(D) Utilizando la carrera y el carné del estudiante se puede deducir el semestre que cursa.

(E) Una vez deducido el semestre, se realiza una búsqueda en el plan de estudios de la carrera, producto de ello se encuentran los cursos en los que el estudiante podría estar matriculado.

(F) La lista de cursos del paso (E) permite buscar en la guía de horarios de la UCR las posibles ubicaciones (edificio, aula y horario) donde se puede encontrar al estudiante.

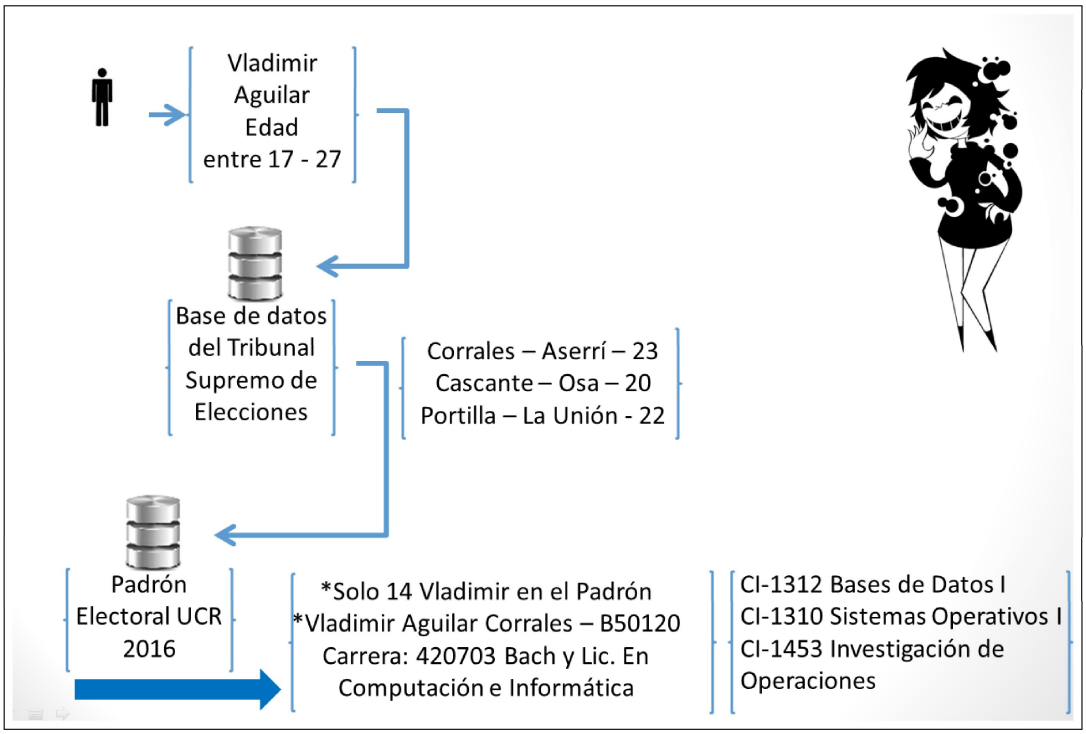



Figura 1.3: Reidentificación a un estudiante.

En el ejemplo anterior se utiliza información adicional para realizar la inferencia, pero en algunos casos, con tan sólo alcanzar una disminución en la cantidad de opciones posibles es suficiente para que se realice una afectación a la privacidad. Por ejemplo, note que para ejecutar el paso (F) donde se logran obtener las posibles ubicaciones, no

es requerido conocer el nombre del estudiante. Tan solo es necesario el nombre del curso o cursos. Esta inferencia puede resultar positiva si la persona afectada tiene interés en ser localizada, pero puede resultar muy delicada si facilita un secuestro o algún otro tipo de afectación.

Utilizar como expresión de una posible afectación de la privacidad posibles inferencias, permite expresar en lenguaje natural la posible violación a la privacidad y a la vez, presenta al dueño de los datos una forma simple y consistente de autorizar o no su realización. Los posibles usos de un conjunto de datos son difíciles de identificar, pero se asume que los posibles usos nocivos son un conjunto más reducido.

## 1.4. Alcances y limitaciones

Esta tesis presenta una forma de utilizar preferencias individuales de privacidad, esto mediante el uso de la presentación de inferencias no deseadas a los dueños de los datos. Las inferencias deben crearse en función de los datos disponibles en el contexto de la organización publicadora de datos. El bloqueo de las inferencias de datos se propone mediante la modificación de un modelo de desidentificación conocido. Todo esto se conduce bajo la búsqueda de dificultar o no permitir la reidentificación del dueño de los datos o la asociación de datos sensibles con datos proporcionados por un usuario.

Aspectos tales como la prueba de un instrumento de autorización o recolección de las preferencias de privacidad del dueño de los datos exceden los alcances de esta tesis. Esto se debe a que el enfoque es probar la factibilidad de aplicación y no evaluar aspectos como la interacción humano computador de la solución, aún cuando la finalidad es presentar una forma para que los dueños de los datos realicen decisiones informadas, responsables y con resolución variable sobre el futuro uso de sus datos. Se presenta una forma para respetar las decisiones individuales de los dueños de los datos mediante una representación menos compleja que la autorización dato por dato y al mismo tiempo permitir que los datos mantengan alguna semejanza estadística cercana a las distribuciones de los datos originales. También excede esta tesis el estudiar el problema de privacidad generado al realizar consultas secuenciales.

## 1.5. Antecedentes

La privacidad es un tema complejo, puede ser abordado desde diferentes perspectivas y utilizando diferentes herramientas. En está tesis se integran la preservación de la privacidad personalizada, la negociación de la privacidad, el uso de inferencias con fines maliciosos, así como el uso de modelos y métricas de preservación de la privacidad existentes. Estos problemas no son nuevos. Existen trabajos en la comunidad científica que se deben considerar. Sin embargo, ninguno de estos trabajos unifica y propone soluciones a los problemas más amplios experimentados en el entorno actual y que son el tema de este trabajo. Antes de examinar estas temáticas, establezco un vocabulario común adoptando las siguientes definiciones. El término datos se refiere a información específica de una organización que se estructura conceptualmente como una tabla de filas (o registros) y columnas (o campos). Cada fila se denomina tupla.

Normalmente el objetivo de la preservación de la privacidad intenta ejercer un control con respecto a los datos específicos de un individuo y ese control consiste en garantizar que los datos publicados respeten la privacidad de los dueños.

### 1.5.1. Modelos de preservación de la privacidad

Existen muchas formas de preservación de la privacidad, que van desde la simple eliminación de los identificadores directos (cédula de identidad, pasaporte, nombre completo, entre otros)[Shah, 2012] hasta permitir sólo la ejecución de consultas específicas (bases de datos cifradas) [Popa et al., 2012]. Sin embargo, dado que el propósito en esta tesis es la preservación de la privacidad en datos que deben publicarse, se decidió enfocar el texto en estudios relacionados a esta temática.

La preservación de la privacidad depende principalmente de cuánta privacidad se quiere aplicar sobre los datos. Se han propuesto diferentes modelos de preservación de la privacidad que permiten indicar está cuantificación. El avance de estos modelos ha permitido un desarrollo constante de la investigación sobre la preservación de la privacidad y ha concedido modelos de gran renombre como k-anonimato, l-diversidad, t-cercanía y otros [Sweeney, 2002b, Sweeney, 2002a, Li et al., 2007a, Machanavajjhala et al., 2006, Chi-Wing et al., 2006]. Entre los muchos modelos de privacidad, el k-anonimato y sus variantes son los más utilizados [Xu et al., 2014a]. Una presentación amplia de esta temática se realiza en el capítulo 2.

### 1.5.2. Preservación de la privacidad personalizada

La mayoría de los modelos de preservación de la privacidad se centran en un enfoque universal de la privacidad donde se busca la misma cantidad de preservación para todos los dueños de datos. Como consecuencia, dichos modelos pueden ofrecer una protección insuficiente para algunos dueños de datos al tiempo que aplican un control de privacidad excesivo a otros. La preservación de la privacidad personalizada elimina estos problemas al ofrecer formas para satisfacer los requisitos personales de cada dueño de datos [Xiao and Tao, 2006b].

La idea de la preservación de la privacidad personalizada fue propuesta por primera vez por Xiao y Tao. Como en la mayoría de los casos documentados en la literatura, esta personalización se logra mediante la adición de restricciones adicionales sobre los modelos de privacidad existentes. Ellos establecen diferentes restricciones de generalización o agrupamiento de valores sensibles, para que cada individuo logre el requisito de preservación de privacidad personalizado [Xiao and Tao, 2006a]. Ye *et al* [Ye et al., 2008] extendieron esta idea y presentaron el modelo ($\alpha$, k)-anonimato personalizado que le permite a cada individuo establecer diferentes restricciones de frecuencia para la generalización de cada uno de los valores sensibles en un campo. Se enfocan en trabajar registro a registro. Sin embargo, causa mucho trabajo y muchos datos redundantes al mantener diferentes frecuencias para un mismo valor.

Xu *et al* [Xu et al., 2013] proponen el uso de diferentes valores de k y rutas de generalización personales flexibles para generalizar diferentes tuplas. Se basa en la selección de un valor k mínimo como un valor a cumplir en las generalizaciones de datos. En general, cada implementación de la preservación de la privacidad personalizada es dependiente del modelo de privacidad base (k-anonimato, l-diversidad, entre otros) que se utilice, esto a fin de aprovechar los mecanismos de cuantificación de la privacidad que provean.

### 1.5.3. Negociación de la privacidad

El uso de tecnologías de mejora de la privacidad puede ayudar a diseñar sistemas y servicios de información y comunicación de una manera que minimice la recopilación y el uso de datos personales, así como facilitar el cumplimiento de las normas de protección de datos existentes [Padyab and Ståhlbröst, 2015].

Se identifican tres temas alrededor del ejercicio del derecho a la privacidad de las personas [Acquisti et al., 2015]:

1. El primer tema es la incertidumbre del dueño de los datos acerca de la naturaleza de privacidad (ventajas y desventajas), y sus propias preferencias sobre ellas.

2. El segundo es la poderosa dependencia del contexto de las preferencias de privacidad: La misma persona puede en algunos casos ser ajeno a un tema de privacidad, pero en otras ocasiones estará claramente preocupado.

3. El tercer tema es la maleabilidad de las preferencias de privacidad, donde se observa que las preferencias de privacidad están sujetas a la influencia de los que poseen un mayor conocimiento de sus factores determinantes.

Dadas estas temáticas, se obtiene que la forma adecuada de negociación de la privacidad de los dueños de los datos es a través de la autodeterminación informativa, la cual se explica con detalle en el capítulo 2.

### 1.5.4. Métricas de preservación de la privacidad

El garantizar que los datos publicados respeten la privacidad de los dueños implica algunas modificaciones de los datos originales, lo que disminuye la utilidad de los datos protegidos; por lo tanto, es importante poder evaluar la calidad de los datos protegidos, así como la protección de la privacidad. Medir la preservación de los datos publicados es una tarea difícil. Actualmente, ninguna medida única es ampliamente aceptada y se ha demostrado que carecen de comportamiento monotónico [Bertino et al., 2008, Wagner and Eckhoff, 2018].

En este trabajo se desarrolla una explicación detallada sobre los aspectos más relevantes de métricas de privacidad en el capítulo 2 y se presenta una propuesta de uso de métrica en el capitulo 6 que es aplicada en el caso de estudio (capítulo 7) sobre el modelo presentado en el capítulo 4.

### 1.5.5. Inferencias siendo utilizadas con fines maliciosos

El uso de inferencias en la literatura sobre preservación de la privacidad es frecuente [Wang et al., 2005, LeFevre et al., 2006b, Rao et al., 2013, Tian et al., 2017]. Una inferencia, en el contexto de esta tesis consiste en la obtención de información sobre un

tema sin autorización del dueño, utilizando como medio el aprendizaje máquina. La principal razón de su constante uso es el intentar emular escenarios dónde se conozca con anterioridad los posibles usos de datos, pero también se a utilizado para simular un ataque real [LeFevre et al., 2006b, Cormode et al., 2013].

La literatura demuestra que el $72,7\,\%$ de todos los ataques a la privacidad de datos documentados del 2009 al 2016 habían tenido éxito. La mayoría de los ataques utilizaban múltiples conjuntos de datos. La mayoría de ellos habían tenido lugar en conjuntos de datos globales, como los datos de redes sociales, y habían sido conducidos por investigadores con sede en los Estados Unidos [Henriksen-Bulmer and Jeary, 2016]. Debido a que las violaciones de la privacidad tienen implicaciones de seguridad, políticas y jurídicas es de especial importancia el validar toda propuesta de preservación de la privacidad desde la perspectiva de un atacante. Mayores detalles sobre el uso de inferencias son desarrollados en el capítulo 2 y el capítulo 6.

## 1.6. Trabajo relacionado

En esta sección se discuten investigaciones con objetivos similares a los propuestos en esta tesis, y su relación con el trabajo que se desarrolló en la misma.

La preservación de la privacidad utilizando las preferencias de los dueños de los datos o preservación de la privacidad personalizada, en la mayoría de los casos se logra aumentando las restricciones de los modelos de preservación de la privacidad generales, como K-anonymity, L-diversity, y otros [Sweeney, 2002b, Sweeney, 2002a, Li et al., 2007a, Machanavajjhala et al., 2006, Chi-Wing et al., 2006]. Estos modelos se centran en un enfoque universal que ejerce la misma cantidad de protección para todos los individuos incluidos en el conjunto a publicar, y no tienen en cuenta las preferencias individuales de privacidad de los dueños de los datos.

Shen *et al* [Shen et al., 2013] presentan un modelo de k-anonimato granular personalizado, que evalúa los parámetros de restricción con respecto a los grados de protección de la privacidad especificados por los individuos. Cada individuo impone una mayor sensibilidad al valor del atributo sensible, clasificando los datos originales en diferentes grados de decisión para preservar la privacidad de forma personalizada en términos de atributos sensibles. Este modelo puede llevar a la inexactitud de algunos parámetros relacionados, ya que solo se basa en los datos especificados individualmente.

Liu *et al* [Liu et al., 2015] extendieron el modelo de ($\alpha$, k)-anonimato personalizado. Crean grupos de valores sensibles que utilizan un nivel de preservación de la privacidad preestablecido, y los niveles de preservación de la privacidad cambian según la frecuencia de los valores en los diferentes grupos. Una vez hecho esto, el grado de generalización se establece en el árbol de jerarquía de generalización [Liu et al., 2017].

En el modelo de Nakagawa [Nakagawa et al., 2017], cada individuo puede designar libremente los elementos que desea proteger y la preservación de la privacidad se realiza para reducir la pérdida innecesaria de información. El problema para todos los modelos es el mismo. Todos los modelos asumen que los propietarios de los datos definen los datos confidenciales o el nivel de privacidad. Estas conjeturas son muy difíciles de argumentar, principalmente si se toma en consideración la gran cantidad de datos que se generan y la falta de conocimiento sobre los riesgos por parte de los dueños de los datos.

Romero y Tris, en 2018 [Romero-Tris and Megías, 2018], argumentan que las propuestas actuales de preservación de la privacidad personalizada se centran en la preservación de la privacidad que ya está almacenada en una base de datos de terceros, lo que no le da al usuario ningún control real sobre la información que comparte. En su lugar, proponen centrarse en preservar la privacidad de los datos antes de que los datos se envíen a un tercero que los recopila y almacena. La idea consiste en proteger los datos del lado del dueño de los datos, permitiéndoles seleccionar la cantidad y la precisión con que se comparten.

Un enfoque considerado adecuado para recopilar las preferencias de privacidad es la negociación de la privacidad, que consiste en el proceso mediante el cual el dueño de los datos puede expresar una preferencia de privacidad que establece cómo se deben utilizar sus datos [Rahman et al., 2013]. Este enfoque es abordado por varias investigaciones, pero se limita a la presentación de incentivos a recibir por compartir los datos, la autorización de elementos particulares o grupos de datos y todo ello sin presentar riesgos asociados a la autorización para los dueños de los datos. [El-Khatib, 2003, Pearson and Elahi, 2011, Ukil et al., 2012, Probst, 2016, Aydoğan et al., 2017, Alanezi and Mishra, 2018].

Ninguna de estas investigaciones indica cómo extraer y representar el requisito personalizado para preservar la privacidad de los dueños de los datos desde la perspectiva de los riesgos de privacidad. El acercamiento sistemático propuesto en esta tesis incluye el uso de la negociación con el dueño de los datos por medio de la presentación de

riesgos de privacidad, materializados en inferencias posibles. Los expertos o encargados de la BD en el publicador de datos deben identificar los elementos o conjuntos de datos vulnerables. También, deben definir la aplicación y la definición de modelos de privacidad adecuados para apoyar o brindar alguna garantía científica de cumplimiento de las preferencias de privacidad.

Una explicación amplia del modelo de expresión de posibles inferencias sobre consultas a una base de datos relacional es presentado en el capítulo 4.

## 1.7. Organización de la tesis

El resto de esta tesis está organizado de la siguiente manera:

- En el capítulo 2 se desarrolla una revisión de la literatura relevante al tema de investigación. En esta revisión se estudian los aspectos generales de la preservación de la privacidad de datos a publicar, la preservación de la privacidad personalizada, la negociación con el usuario y la recopilación de preferencias.

- El capítulo 3 provee una visión general de la metodología utilizada en el desarrollo de la tesis. La descripción detallada de métodos, técnicas y procedimientos utilizados se encuentran en cada capítulo.

- El capítulo 4 describe el modelo propuesto de expresión de posibles inferencias sobre consultas a una base de datos relacional, que permite tomar en cuenta las preferencias individuales de privacidad de los dueños de los datos. Parte de los resultados presentados en este capítulo fueron publicados en [Esquivel-Quirós and Barrantes, 2015] y [Esquivel-Quirós and Barrantes, 2016].

- En el capítulo 5 se presentan mecanismos para instrumentar el método de preservación de la privacidad aplicado por el publicador se para tomar en cuenta las preferencias de los usuarios sobre bloqueos de inferencias.

- El capítulo 6 describe la métrica desarrollada para la medición de la preservación de la privacidad sobre inferencias a datos y su evaluación. Los resultados de este capítulo fueron publicados en [Esquivel-Quiros et al., 2018] y [Esquivel-Quirós et al., 2019].

- El capítulo 7 presenta el desarrollo de un caso de estudio que permite demostrar el uso del modelo propuesto.
- En el capítulo 8 se realiza el análisis de los datos y la discusión de los resultados producidos en por el caso de estudio desarrollado en el capítulo 7.
- El capítulo 9 concluye la tesis y sugiere algunas direcciones para futuras investigaciones.

# Capítulo 2

# Marco teórico

En este capítulo se explican los conceptos más relevantes para el trabajo realizado en esta tesis. Se trata sobre conceptos y teorías en el campo de la preservación de la privacidad, la expresión de preferencias de privacidad y la preservación de la privacidad de datos a publicar (PPDP, por sus siglas en inglés). El capítulo se estructura presentando el concepto de privacidad y las teorías que permiten ejercer el derecho a la privacidad en la sección 2.1. La sección 2.2 detalla los modelos y métricas utilizados en la PPDP.

## 2.1. Privacidad

El concepto de privacidad no tiene un significado único, claro y definido, y varía ampliamente entre las diferentes culturas. Es un concepto discutido en filosofía, política, derecho, sociología, antropología y otras áreas [Warren and Brandeis, 1890, Bates, 1964, Posner, 1981, Solove, 2006, Shklovski and Vertesi, 2013]. Las diferencias entre las esferas de acción públicas y privadas se reconocen en todas las sociedades humanas [Solove, 2006]. Es por ello que se debe distinguir entre lo que se define como privado y el grado de protección de esa privacidad. Algunos autores son críticos del concepto, por ejemplo al cuestionar la dificultad en la definición de fronteras entre la privacidad de distintos entes [Warren and Brandeis, 1890], por ineficacia económica [Posner, 1981] o por el potencial abuso (impunidad) [MacKinnon, 1989]. Sin embargo, la mayoría de los autores consideran que es un concepto significativo y valioso. Los debates alrededor del tema cobraron mayor importancia a partir de la segunda mitad del siglo 20, en especial en lo que concierne a la información sobre las personas [DeCew, 2018].

Desde una perspectiva ética, las empresas, personas, universidades, centros de investigación o cualquier otra organización que puede ser o es tutelar de información, debe velar porque no se expongan individuos o grupos en las publicaciones de datos que realizan [Shklovski and Vertesi, 2013]. En 1977, Dalenius proporcionó una defini-

ción muy estricta, donde especifica que el acceso a los datos publicados no debe permitir que un atacante aprenda algo adicional sobre cualquier víctima objetivo, en comparación con la información que obtendría si no contara con el acceso a la base de datos publicada, incluso con la presencia de conocimiento previo obtenido de otras fuentes [Dalenius, 1977].

Los individuos y las organizaciones se exponen a la divulgación no autorizada de sus datos cuando se publica información privada en plataformas públicas o semi-públicas. La publicación puede ser intencional o no, pero una vez realizada, el productor de los datos se ve expuesto a diversas consecuencias, en ocasiones negativas. Esto lleva a que en algunos casos se efectúen procesos legales contra individuos y empresas por divulgación no autorizada [Griffin, 2012]. Por lo general los grupos vulnerables son fácilmente identificables dadas sus particularidades, lo que permite una influencia directa de la privacidad sobre las inclinaciones de honestidad en las revelaciones de información y, en algunos estudios, encuentran que además de la calidad de la información provista, aumenta la cantidad de la misma [Bodle, 2013]. Ante estos detalles, es común que existan leyes e instituciones para preservar la privacidad en la mayoría de los países. En Costa Rica existe por ejemplo la "Ley de protección de la persona frente al tratamiento de sus datos personales" [República de Costa Rica, 2011] y la Agencia de Protección de Datos de los Habitantes (PRODHAB).

La privacidad ofrece seguridad contra represalias y escarnio público, lo que permite a grupos vulnerables o marginales, participar, realizar transacciones y en muchos casos, buscar ayuda, consejo y hasta consuelo por la situación que viven [Bodle, 2013]. Los desarrollos legales y tecnológicos creados alrededor de la privacidad buscan en el fondo un control similar al planteado por William Parent [Parent, 1983], donde se permita al usuario determinar cuál información desea revelar o no, sin que ese ejercicio signifique un proceso exhaustivo y complejo. En la práctica la expresión de las preferencias de los usuarios sobre la privacidad de sus datos es realizada a través de la autodeterminación informativa. La sub-sección 2.1.1 describe la forma en que un usuario expresa sus determinaciones o preferencias respecto a sus datos por medio de la autodeterminación informativa.

### 2.1.1. Autodeterminación informativa

La autodeterminación informativa incluye el derecho a participar libremente, el derecho a no ser seguido, el derecho a acceder a la información sobre sí mismo y el derecho a borrarla si se desea. Se puede afirmar que, para equilibrar estos derechos, deben equilibrarse los derechos de los titulares de la propiedad intelectual y los derechos de los intermediarios con respecto a principios de necesidad y proporcionalidad, para obtener garantías contra abusos y mantener una equidad entre los derechos y las responsabilidades [Bodle, 2013]. En el caso de Costa Rica la autodeterminación informativa se define como el derecho al control que toda persona tiene sobre sus datos de carácter personal, que conlleva saber quién los tiene y para qué los va a usar [República de Costa Rica, 2011].

De esta forma, toda persona tiene derecho a la autodeterminación informativa, la cual abarca el conjunto de principios y garantías relativas al legítimo tratamiento de sus datos personales. También se reconoce a la autodeterminación informativa como un derecho fundamental, con el objeto de controlar el flujo de datos que concierne a cada persona, este derecho es derivado del derecho a la privacidad y busca evitar que se propicien acciones discriminatorias [Cordoba Ortega, 2003, República de Costa Rica, 2011, Troncoso Reigada, 2006]. La autodeterminación informativa se fundamenta en la protección de la privacidad y en el análisis de los datos asociados al dueño de los mismos. Se indica en la literatura que estas protecciones se basan en el uso del "consentimiento informado", debido a que este básicamente tiene como objetivo proteger a una persona de interferencias externas con sus propias elecciones y la toma de decisiones. Este enfoque es particularmente relevante cuando se busca un acceso equitativo [Gefenas, 2012]. En general, las leyes le dan énfasis a los derechos de las personas, y por lo tanto son las personas quienes deben definir lo que se realiza con sus datos a través de la autodeterminación informativa. En la sub-sección 2.1.2 se aborda el consentimiento informado y su relación en la expresión de preferencias individuales de privacidad.

### 2.1.2. Consentimiento informado

La literatura presenta al consentimiento informado como el principal medio para expresar la autodeterminación informativa. Esto principalmente por ser un medio para abordar los obstáculos de la confianza, el control y la privacidad de los dueños de los

datos. El consentimiento informado es el medio por el cual se garantiza que el dueño de los datos exprese voluntariamente su autorización, esto después de haber comprendido las implicaciones (buenas y malas) de la misma. Se concibe como un ritual humano, altamente social y dependiente del contexto en el que se realiza. Su estudio debe abordar los aspectos formales y los aspectos técnicos que lo describen [Geest et al., 2005].

**Aspectos formales**

De forma conceptual, el consentimiento informado se basa en cuatro principios consagrados en el Código de Nuremberg [Ghooi, 2011]. Estos principios indican que debe ser voluntario, competente, informado y comprendido por los individuos que lo ejercen [Luger et al., 2014]. El Código de Nuremberg fue el primer documento jurídico internacional para introducir el requisito de "consentimiento voluntario" como una condición necesaria para la participación en una investigación, y se colocó como la primera en una lista de los diez principios básicos de la experimentación [Gefenas, 2012]. Los hechos que dieron lugar a la aplicación de los principios del consentimiento informado se remontan a los experimentos realizados en la Alemania nazi, los cuales llevaron a la creación del Código de Nuremberg una vez que la guerra había terminado [Ghooi, 2011].

El consentimiento informado se puede definir como una doctrina ética y legal basada en la suposición de que todo acto se puede efectuar si el participante ha sido informado sobre el propósito, la naturaleza, los riesgos involucrados y las consecuencias, y este ha dado su consentimiento libremente a ello, así como que conserva la potestad y el derecho a retirar este consentimiento [Gefenas, 2012]. Los participantes de un consentimiento informado deben dar su consentimiento a través de una acción afirmativa explícita, como responder a una pregunta específica, en forma escrita u oral, sobre su disposición a participar. Por otro lado, un consentimiento informado amplio implica acordar un extenso conjunto de usos futuros, muchos de los cuales son potenciales o secundarios bajo un marco de gobernanza particular [Politou et al., 2018].

La constitución de un adecuado consentimiento informado es una de las características clave de la ética moderna. Esta constitución se realiza sobre la base de la autonomía personal y está estrechamente vinculada a otros conceptos éticos y legales importantes, tales como la privacidad y la confidencialidad [Gefenas, 2012]. Actualmente se busca que las organizaciones que almacenan datos otorguen mayor control y protección a los dueños de los datos que almacenan [U.S. Federal Trade Commission FTC, 2014]. En

términos informáticos este consentimiento informado requiere la abstracción de cada componente que lo conforma y esta abstracción define los aspectos técnicos a implementar.

### Aspectos técnicos

Friedman, Millet y Felten [Friedman et al., 2000] afirman que el consentimiento informado en las políticas de privacidad informáticas comprende los siguientes elementos: divulgación, comprensión, voluntariado, competencia y acuerdo.

- La divulgación se refiere a cómo proporcionar al dueño de los datos, información precisa sobre los beneficios y los daños que se podrían esperar. Lo que se divulga o presenta al productor de los datos debe abordar los valores, necesidades e intereses importantes de éste.

- La comprensión se refiere a la interpretación exacta del productor de los datos respecto a lo que se le está divulgando o presentando. Este componente plantea la pregunta: ¿Qué criterios se deben cumplir para decir que algo ha sido adecuadamente comprendido? Por ejemplo: ¿entiende el productor de los datos la declaración de privacidad de la organización?

- La voluntariedad implica que un productor de datos sólo debe participar voluntariamente sobre lo que se consulta, no puede haber control sobre las acciones de éste y la acción no puede ser coaccionada.

- La competencia se refiere a poseer las capacidades mentales, emocionales y físicas necesarias para ser capaz de dar el consentimiento informado. Por ejemplo, se debe analizar si los productores de los datos están mental y emocionalmente capacitados para juzgar si deben o no autorizar la publicación de sus datos.

- El acuerdo se refiere a una oportunidad razonablemente clara de aceptar o no mediante el consentimiento informado. Esto no sólo implica la oportunidad de elegir, sino que también debe otorgarse la oportunidad de optar por cambiar de opinión en un momento dado.

Estos elementos indican que el productor de los datos debe poseer el control total en todo momento [Friedman et al., 2000, Geest et al., 2005]. Lamentablemente los usua-

rios no suelen leer ni entender los consentimientos informados [Heeger, 2015]. Así como es difícil argumentar que la sociedad tiene una comprensión suficiente para tomar decisiones sobre posibles riesgos respecto a sus datos [Hodges and Creese, 2013]. Es un tema importante el "¿cómo mostrar a los productores de datos los riesgos de privacidad?". La práctica actual en la publicación de los datos se basa principalmente en políticas y directrices que indican los tipos de datos que se pueden publicar, así como acuerdos sobre el uso de los datos publicados. Este enfoque solo puede conducir a la excesiva distorsión de datos o insuficiente protección [Fung et al., 2010b]. El ideal es respetar la autodeterminación y la autonomía personal que va más allá de la norma de "negación" minimalista donde tan solo se determina que dato se quiere compartir y cuál no [Gefenas, 2012]. Esto convierte en todo un reto preservar la privacidad en datos a publicar y este detalle es afrontado en la siguiente sección.

## 2.2. Preservación de la privacidad de datos a publicar

El proceso de preservación de la privacidad de datos a publicar (PPDP, por sus siglas en inglés), como se muestra en la figura 2.1 puede visualizarse en dos fases, la recolección de datos y la publicación de datos. La PPDP requiere de tres actores, el dueño de los datos, el publicador de datos y el destinatario de datos. En la fase de recolección de datos, el publicador de datos recolecta datos pertenecientes al dueño de los datos (productor). Luego, en la fase de publicación de datos, el publicador de datos envía todos o una parte de los datos recolectados al destinatario o consumidor de datos. Este envío se realiza previo procesamiento de los datos a publicar [Xu et al., 2014b].

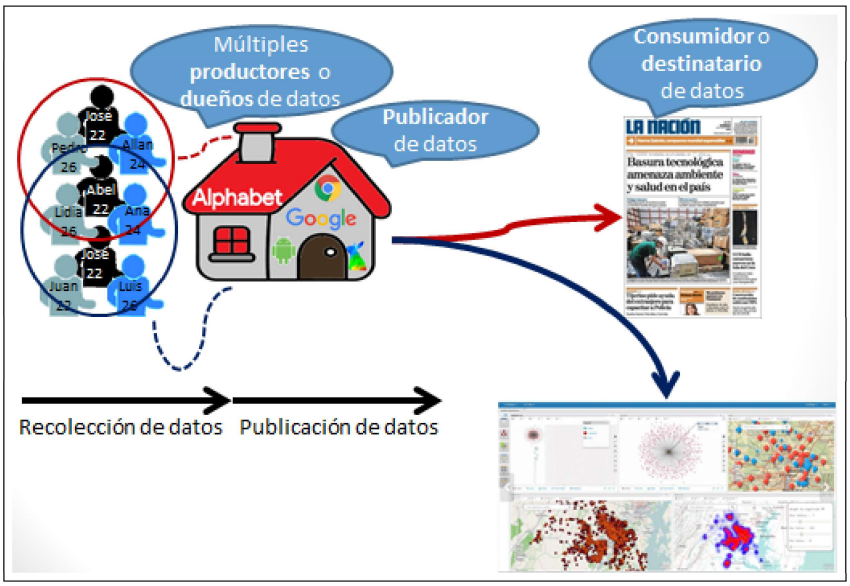



Figura 2.1: Fases y actores en la PPDP

La PPDP proporciona los métodos y herramientas para procesar los datos a publicar, de forma que se mantenga la utilidad y se preserve la privacidad de los datos [Fung et al., 2010b, Gkoulalas-Divanis and Aonghusa, 2014]. Esto lo proporciona a través del estudio de cómo transformar los datos en bruto a una versión que está inmunizada contra ataques a la privacidad, pero que todavía apoya la realización de tareas con ellos [Hamza and Hefny, 2013]. El objetivo de las técnicas de PPDP es modificar los datos haciéndolos menos específicos, de modo que la privacidad de los individuos esté protegida y a su vez se retenga la utilidad de los datos anónimos. La esencia de la PPDP es producir conjuntos de datos donde se oculte la identidad del dueño y que a la vez tengan una buena utilidad para una variedad de tareas, principalmente debido a que comúnmente, se desconocen todos los posibles escenarios de uso para los datos a publicar [Hamza and Hefny, 2013, Ayala-Rivera et al., 2014]. Las técnicas provistas por la PPDP deben permitir asegurar una utilidad práctica de uso sobre el conjunto de datos publicado y un grado de dificultad al intentar realizar una identificación de los dueños de los datos a partir de la información provista.

En otras palabras, la PPDP se refiere al proceso mediante el cual un administrador de datos tiene un conjunto de datos T, compuesto por identificador(es) directo(s), cuasi-identificador(es), atributo(s) sensible(s), atributo(s) no sensible(s), donde uno o más identificadores directos permiten identificar explícitamente a los dueños de los da-

tos; el cuasi-identificador es uno o más atributos que podrían identificar a los dueños de los datos, esto por medio de la vinculación a información externa que permita reidentificar los dueños originales. Para un atributo cuasi-identificador la característica más importante es qué tan disponibles estén datos externos con una variable que corresponda al potencial valor del cuasi-identificador. Los atributos sensibles consisten en información susceptible a crear algún perjuicio o discriminación específica hacía el dueño de los datos, como por ejemplo la religión, una enfermedad, el salario o su estado de discapacidad, entre otras; y los atributos no sensibles contienen todos los atributos que no entran en las tres categorías anteriores [Fung et al., 2010a].

Dado esto, la PPDP toma la tabla $T(ID : identificadores, CID : cuasi-identificadores, S : Columnas\ sensibles, NS : Columnas\ no\ sensibles)$ y la lleva a un estado $T'(CID', S', NS)$. Donde $CID'$ y $S'$ es una versión desidentificada del $CID$ y $S$ originales. Esto al aplicar operaciones de preservación de la privacidad (anonimización) a las columnas $ID$, $CID$ y $S$ de la tabla original $T$, según el modelo de privacidad seleccionado. El conjunto de operaciones de preservación de la privacidad que permiten la desidentifiación requiere un trabajo detallado en la subsección 2.2.1 dada su variedad y funcionalidad.

### 2.2.1. Operaciones de preservación de la privacidad

Los datos recolectados originalmente por el publicador de los datos generalmente no cumplen con un requisito de privacidad específico y requieren modificarse antes de publicarse. La modificación se realiza aplicando una secuencia de operaciones de preservación de la privacidad a los datos originales. Las operaciones típicas en la preservación de la privacidad son:

- *Supresión*: se eliminan valores de registros individuales o atributos completos, implica no liberar un valor en absoluto. En algunos casos se utiliza el reemplazo de los datos originales con un valor especial (por ejemplo, "*") para indicar que esos datos no se revelan [Sweeney, 2002a].

- *Generación de grupos o cubetas (Bucketization)*: consiste en agrupar los datos en pequeños segmentos con el fin de ocultar la asociación a nivel de registro entre los datos sensibles y los datos de identificación. Al segmentarlos se permite publicar

todos los valores cuasi-identificadores y sensibles directamente en dos tablas separadas, pero sin presentar directamente la relación entre estos. No requiere una taxonomía de los valores de los atributos [Xiao and Tao, 2006a, Wang et al., 2016].

- *Microagregación*: los registros se agrupan según una medida de proximidad sobre los valores. En lugar de publicar una variable con el valor original para un registro dado, se publica el promedio de los valores del grupo al que pertenece el registro. El fundamento de la microagregación es que las reglas de confidencialidad permitan la publicación de conjuntos de datos si estos corresponden a grupos de k o más individuos, donde ningún individuo domina (es decir, contribuye demasiado) al grupo y k es un valor umbral. Para minimizar la pérdida de información, los grupos deben ser lo más homogéneos posible. Otra forma de explicar la microagregación, es entenderla como la creación iterativa de cúmulos (clusters) de al menos k elementos y los cuales remplazan los valores originales para que se vuelvan k-indistinguibles (anónimos) [Domingo-ferrer and Torra, 2001].

- *Generalización*: (también llamada recodificación), consiste en reemplazar los valores de un atributo con valores menos específicos pero consistentes. Es común emplear una jerarquía de generalización de valores (JGV). En la figura 2.2 se presenta un ejemplo de una JGV para valores numéricos entre 0 y 99.

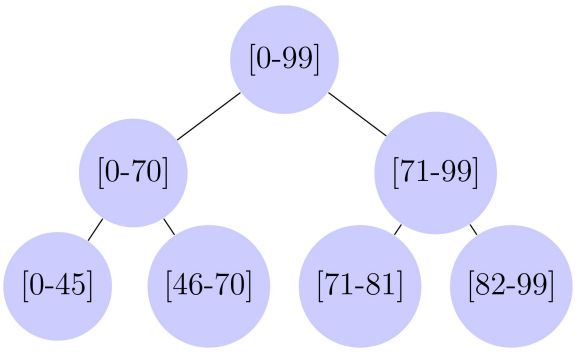



Figura 2.2: Jerarquía de generalización de valores (JGV) con las clases de equivalencia para números de 0 a 99.

También se puede emplear una jerarquía de generalización de valores para atributos categóricos o de otros tipos (que no tienen un orden claramente definido).

Estas jerarquías intentan conservar la semántica de los valores que incluyen. Un ejemplo de una JGV para el atributo categórico estado civil se muestra en la figura 2.3. Los valores en el nivel más bajo están en el dominio de valores original del atributo y corresponden a los valores más específicos (valores originales). Cada uno de los nodos del árbol es llamado clase de equivalencia y permite agrupar un número determinado de tuplas. El nivel más alto que muestra el valor "*" en el caso de los atributos categóricos, corresponde a la máxima generalización o a la supresión total del valor [Ayala-Rivera et al., 2014].

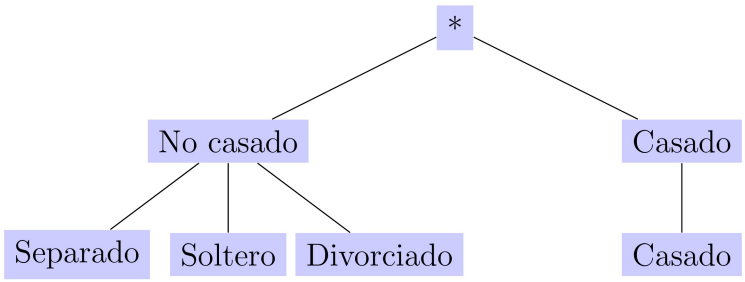



Figura 2.3: Jerarquía de generalización de valores (JGV) con las clases de equivalencia para el atributo categórico estado civil

La mayor parte de los modelos de preservación de la privacidad aplican una o varias combinaciones de operaciones de privacidad con el fin generar conjuntos de datos desidentificados [Pierangela Samarati, 2001, Sweeney, 2002b, Aggarwal et al., 2005, LeFevre et al., 2006b, Li et al., 2007b, Machanavajjhala et al., 2007]. El formato original, es decir, la posición, el tamaño y el tipo de datos no deben alterarse durante el proceso de preservación de la privacidad, por lo que los datos deben parecer realistas incluso para los entornos de datos de prueba y deben ser útiles para procesarse. La figura 2.4 presenta la clasificación de modelos de preservación de la privacidad propuesta por Goswami en [Goswami and Madan, 2017]. En ella se presentan dos grandes categorías, la privacidad diferencial que se preocupa por la protección estadística de la información [Dwork, 2006] y la preservación de la privacidad sintáctica que modifica el conjunto de datos para lograr alguna dificultad en la reidentificación. El desarrollo de este trabajo se centra en las posibles modificaciones que pueden realizarse a los datos para respetar las preferencias de privacidad de los dueños de los datos y por lo tanto es de mayor interés el trabajo en modelos de preservación de

la privacidad sintáctica. Los modelos de preservación de la privacidad sintáctica más conocidos incluyen el k-anonimato [Samarati and Sweeney, 1998, Sweeney, 2002b], la l-diversidad [Machanavajjhala et al., 2006, Machanavajjhala et al., 2007] y la t-cercanía [Li et al., 2007b].

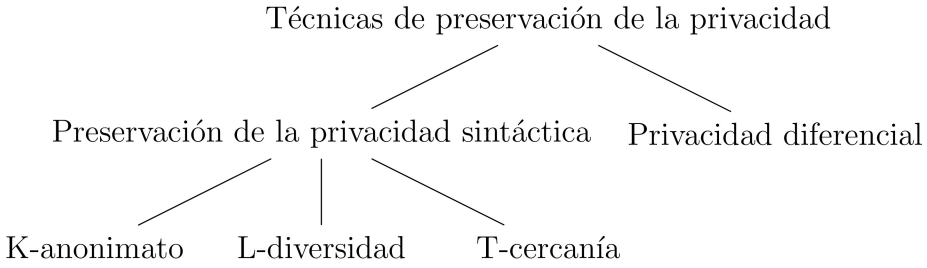



Figura 2.4: Clasificación de modelos de preservación de la privacidad (Fuente: [Goswami and Madan, 2017])

El k-anonimato, uno de los modelos más conocidos para preservar la privacidad de datos, fue propuesto por Samarati y Sweeney en [Samarati and Sweeney, 1998] y extendido por Sweeney en [Sweeney, 2002b]. El aporte del trabajo de Sweeney es definir una propiedad de un grupo de datos tal que si es cumplida, dichos datos se pueden publicar con una reducción significativa en la posibilidad de realizar una reidentificación a partir de ellos.

### 2.2.2. El k-anonimato

El k-anonimato es una técnica que permite proteger los datos a publicar contra la vinculación de datos externos y la posterior identificación de los dueños de los datos. En una tabla de datos k-anónima, cada tupla obtenida sobre las columnas cuasi-identificadoras ocurre al menos $k$ veces [Pierangela Samarati, 2001]. El k-anonimato, puede ser visto como una propiedad deseable de una tabla $T$. Consiste en el compromiso de preservar la privacidad de todo valor en los campos (cuasi-identificadores) que permitan reidentificar una tupla, de forma que los valores de estos campos estén ocultos en al menos otras $k$ tuplas. De esta forma, dada una tupla $t$ en la tabla $T$, el valor de la i-ésima tupla es $t_i[c]$ y $T$ satisface la k-anonimidad si para toda tupla $t_0 \in T, \exists\, k-1$ tuplas $t_1, t_2, ...t_{(k-1)} \in T$ tales que $t_0[c] = t_1[c] = ... = t_{(k-1)}[c]$, $\forall\ c \in CID$ [Sweeney, 2002b].

Las clases de equivalencia definen los grupos de valores que se establecen para ocultar o agrupar las tuplas y funcionan de forma similar a una consulta en $SQL$ con un $GROUPBY$ en una columna o columnas $CID$. De esa forma cada nivel en la jerarquía de generalización provee un conjunto de clases de equivalencia que no comparten elementos con otra y que permiten agrupar más o menos registros para lograr el valor de $k$ solicitado.

Las operaciones de k-anonimato ocultan información detallada para que los múltiples registros se vuelvan indistinguibles con respecto a los valores en el $CID$. En consecuencia, si el dueño de los datos está vinculado a un registro a través de un valor en el $CID$, el dueño de los datos también está vinculado a todos los demás registros que tienen el mismo valor para el $CID$, lo que hace que el enlace sea ambiguo con los demás dueños de los datos. De esta forma, podemos entender que el problema del k-anonimato es producir una $T$ anónima ($T'$) y que la misma satisfaga un requisito de privacidad determinado por el nivel del valor del $k$ en el modelo de privacidad, de forma que retenga la mayor utilidad de datos posible.

Un desarrollo posterior del trabajo de Samarati y Sweeney es la versión multidimensional del k-anonimato conocida como Mondrian Básico, propuesta por Kristen LeFevre en [LeFevre et al., 2006b] y modificada en [LeFevre et al., 2006a].

### K-anonimato multidimensional - Mondrian Básico

Mondrian es un algoritmo de preservación de la privacidad de datos voraz descendente que permite la generalización de atributos $CID$ de forma multidimensional. La versión inicial del algoritmo está diseñada para trabajar solo con atributos numéricos [LeFevre et al., 2006b]. En un trabajo posterior LeFevre propuso una modificación a su algoritmo para que aceptara atributos categóricos, además de numéricos [LeFevre et al., 2006a].

Para satisfacer el requerimiento del k-anonimato sobre la tabla $T$, se debe validar que para cada valor distinto en una columna $CID$, este valor se encuentra $k$ veces. En muchos casos esto no se cumple en la tabla original y mucho menos si se cuenta con varias columnas $CID$. Por este motivo la visión multidimensional de los datos ayuda a satisfacer el requerimiento de forma más rápida. Permitiendo que si tenemos como $CID$ a las columnas $c_1, ..., c_n$, y si por cada tupla única $(t_1, .., t_n)$ al seleccionar estas columnas sobre $T$ se producen al menos $k$ tuplas iguales, se dice que cumple el

principio de k-anonimato en el grado del $k$ reportado. Es decir, el tamaño de cada clase de equivalencia en $T$ con respecto a las columnas $c_1, ..., c_n$ es al menos de tamaño $k$ o en otras palabras el tamaño de la clase de equivalencia se determina contando en número de tuplas que son agrupadas mediante ella [Fung et al., 2010a].

Una propuesta de pseudocódigo de Mondrian Básico se puede observar en el algoritmo 1, en este se puede observar que Mondrian Básico se basa en una partición recursiva voraz del espacio de dominio de los atributos $CID$. Para obtener particiones aproximadamente uniformes, elige recursivamente la columna de división con el mayor rango normalizado de valores, y (para columnas continuos u ordinales) crea la partición de los datos alrededor de la mediana de la columna. Este proceso se repite hasta que no se pueda realizar ninguna división, lo que significa que una región particular no puede dividirse más sin violar la restricción de preservación de la privacidad o las restricciones impuestas por las jerarquías de generalización de las columnas [LeFevre et al., 2006a].

**Algorithm 1** Mondrian Básico

**Require:** $tuplas = \{t_1, t_2, ...t_n\}$ , $atributos = \{A_1, A_2, ...A_n\}$

**Ensure:** conjunto de tuplas anonimizado

**if** no hay división permitida para las tuplas **then**
  **return** $\phi : t \in tuplas \rightarrow$ región límite $(tuplas)$
**else**
  $mejorAtributo \leftarrow elijeAtributo(atributos, tuplas)$
  **if** $continuo(mejorAtributo)$ *o* $ordinal(mejorAtributo)$ **then**
    $umbral \leftarrow elijeUmbral(mejorAtributo)$
    $lhs \leftarrow \{t \in tuplas : t.mejorAtributo \leq umbral\}$
    $rhs \leftarrow \{t \in tuplas : t.mejorAtributo > umbral\}$
    **return** $MondrianBasico(rhs, atributos) \cup MondrianBasico(lhs, atributos)$
  **else**
    **if** $nominal(mejorAtributo)$ **then**
      $recodificaciones \leftarrow \{\}$
      **for cada** hijo $v_i$ de $origen(mejorAtributo.jerarquia)$ **do**
        $tuplas_i \leftarrow \{t \in tuplas : t.mejorAtributo \preceq v_i\}$
        $atributos'$ ← reemplace $origen(mejorAtributo.jerarquia)$ con $v_i$ en $atributos$
        $recodificaciones \leftarrow recodificaciones \cup MondrianBasico(tuplas_i, atributos')$
      **end for**
      **return** recodificaciones
    **end if**
  **end if**
**end if**

La utilidad de datos de Mondrian se ve significativamente afectada por la distribución de datos y el mecanismo utilizado para la creación de particiones. El enfoque de partición mediana es adecuado para distribuciones uniformes porque la mayoría de los atributos se pueden usar para dividir los datos. Por el contrario, cuando los datos están sesgados, los datos no se pueden dividir entre todos los atributos, lo que afecta la medición de la pérdida de información. Por lo tanto, Mondrian tiene un bajo rendi-

miento en este caso. Además, al aplicarse sobre valores categóricos, el orden en que se procesan los valores para calcular la mediana juega un papel importante en la ejecución del algoritmo, llegando a provocar que no sea posible crear la partición. Por lo tanto, esto también afecta la medición de la utilidad de datos. Mondrian termina su ejecución una vez que no se permiten más particiones [Ayala-Rivera et al., 2014].

El contar con alguna métrica que verifique la utilidad de los datos y la preservación de la privacidad es fundamental. En general se crea una pérdida de información cada vez que un valor en una columna $CID$ se generaliza a una clase de equivalencia mayor en la jerarquía de generalización, esto debido a que cada clase de equivalencia es más genérica que la anterior. El objetivo de un buen algoritmo de PPDP es encontrar una transformación de los datos originales, de forma que satisfaga un requerimiento de privacidad al tiempo que minimiza la pérdida de información y maximiza la utilidad de los datos resultantes. Por lo tanto, es necesaria una medida para indicar la calidad y la preservación de la privacidad de los datos.

### 2.2.3. Métricas de privacidad

En la literatura se presentan varias métricas que evalúan la calidad de los datos utilizando medidas simples basadas en el tamaño de la clase de equivalencia o el número total de generalizaciones [Pierangela Samarati, 2001, Sweeney, 2002b, Meyerson and Williams, 2004, Aggarwal et al., 2005, Bayardo and Agrawal, 2005, LeFevre et al., 2006b]. Dado el efecto del uso de jerarquías de generalización, en este trabajo es de interés la métrica de **Penalización de Certeza Normalizada (NCP, por sus siglas en inglés)** [Xu et al., 2006]. Está métrica intenta capturar la incertidumbre causada por la generalización de cada clase de equivalencia en el espacio de columnas $CID$. Para un valor en una columna $CID$ continua u ordinal ($c_i$), el NCP en una clase de equivalencia G se define como:

$$NCP_{A_{Num}}(G) = \frac{max^{G}_{A_{Num}} - \ min^{G}_{A_{Num}}}{max_{A_{Num}} - \ min_{A_{Num}}} \tag{2.1}$$

El numerador y el denominador de la ecuación 2.1 representan los rangos de valores del atributo $CID$ en los que se agrupa el $A_{Num}$ en la clase de equivalencia G y el dominio completo del atributo $CID$, respectivamente. En el caso de los valores de atributos categóricos ($A_{Cat}$), donde no exista un orden total o alguna función de distancia, el

NCP se define con respecto a la jerarquía de generalización del atributo $CID$. Podemos expresarla como:

$$NCP_{A_{Cat}}(G) = \begin{cases} \text{si } card(u) = 1 & \rightarrow & 0 \\ \text{si no} & \rightarrow & card(u)/|A_{Cat}| \end{cases} \tag{2.2}$$

Donde $u$ en ecuación 2.2 es el ancestro común más bajo de todos los valores de $A_{Cat}$ que están incluidos en la clase de equivalencia G, $card(u)$ es el número de hojas (es decir, el número de valores agrupados del atributo) en el subárbol de $u$ y $|A_{Cat}|$ es el número total de valores de $A_{Cat}$ distintos. El NCP de la clase de equivalencia G sobre todos los atributos $CID$ es:

$$NCP(G) = \sum_{i=1}^{d} w_i * NCP_{A_i}(G) \tag{2.3}$$

Donde en la ecuación 2.3, $d$ es el número de atributos en $CID$ (es decir, la dimensionalidad). $A_i$ es un atributo continuo, ordinal o categórico y tiene un peso $w_i$, donde $\sum_{i=1}^{d} w_i = 1$. NCP mide la pérdida de información para una única clase de equivalencia y caracteriza la pérdida de información de una partición completa al sumar el NCP de todas las tuplas de cada grupo. Para este trabajo adoptamos la formulación normalizada de la versión agregada de NCP, llamada **Penalización de Certeza Global (Global Certainty Penalty - GCP)**) [Ghinita et al., 2009]. Está mide la pérdida de información de toda la tabla anonimizada, tomando $P$ como el conjunto de todas las clases de equivalencia en la tabla anonimizada y de se define como:

$$GCP(P) = \frac{\sum_{G \in P} |G| * NCP(G)}{d * N} \tag{2.4}$$

En la ecuación 2.4 tenemos que $N$ representa el número de registros en la tabla original $T$, $|G|$ es la cardinalidad de la clase de equivalencia $G$, y $d$ es la dimensionalidad de los $CID$. La ventaja de esta formulación es su capacidad para medir la pérdida de información entre tablas con cardinalidad y dimensionalidad variables. Además, GCP está entre 0 y 1, donde 0 significa que no existe pérdida de información (es decir, los datos originales) y 1 corresponde a la pérdida total de información (es decir, solo hay una clase de equivalencia que cubre todos los registros en la tabla). Además de adoptar GCP, en este trabajo lo transformamos en porcentaje. Este porcentaje de GCP se

calcula dividiendo el valor de GCP con el número de valores en el conjunto de datos y lo hemos llamado **GCP** por facilidad de uso.

Las métricas de privacidad permiten verificar de forma cuantificada el porcentaje de éxito de un proceso de desidentificación [Kelly et al., 2008, El Emam et al., 2011]. En general, se cuantifica la incertidumbre de poder inferir los dueños de los datos [Byun et al., 2007]. El grado de incertidumbre es la cantidad de información privada que puede ser descubierta a partir de la publicación de datos desidentificados [Shu et al., 2015]. Por ejemplo, el grado de incertidumbre para el k-anonimato es al menos $1/k$, esto debido a que no hay menos de $k$ valores únicos en la tabla. Cuanto mayor es el grado de incertidumbre alcanzado por un algoritmo, mejor se protege la privacidad de los datos [Bertino et al., 2008].

En cuanto a las métricas de utilidad, la literatura generalmente las define en forma negativa, como *pérdida de información* [Bayardo and Agrawal, 2005]. Cualquiera de estas métricas asigna algún valor a cada cambio en el conjunto de datos realizado en la búsqueda de la preservación de la privacidad (la "penalización", debido a que refleja la pérdida de datos causada por el proceso), y luego genera algún tipo de promedio. En este documento también se utilizara la métrica de clasificación (CM) propuesta por Iyengar [Iyengar, 2002]. El CM determina la pérdida de información asignando un valor de 1 a cada tupla que está generalizada o suprimida, y 0 a todos los valores de los elementos que cumplen los requisitos de $k$ por sí mismos. Cuando un valor se generaliza o se suprime, automáticamente pierde información. Sin embargo, si el valor no es afectado por el proceso de desidentificación, su pérdida de información es de cero. La suma de todos los pesos es un conteo de los registros afectados. Luego se divide por el número total de filas para obtener el valor de CM. Por lo tanto, esta métrica se puede usar como una aproximación para medir la pérdida de información. El CM es una buena medida de utilidad porque considera un posible uso para los datos [Kifer and Gehrke, 2006] .

La importancia de una métrica de privacidad no solo se basa en la capacidad para medir la utilidad de un conjunto de datos. Por esto varios investigadores en PPDP se han visto motivados en intentar medir la posibilidad de vinculación de datos [LeFevre et al., 2006a, Machanavajjhala et al., 2007, Chen et al., 2008, Tian and Zhang, 2013].

### Vinculación de datos

La vinculación de datos representa el proceso mediante el cual se identifican datos que se refieren a la misma entidad del mundo real y puede ser realizada a través de múltiples fuentes. La gran explosión de datos de las organizaciones e individuos ha creado retos críticos en el proceso de vinculación de datos, lo que lleva a la comunidad científica para desarrollar una rica serie de técnicas para lograr asociar, inferir e identificar cuáles datos están relacionados [Bonomi et al., 2015]. Una forma para alcanzar esta vinculación es la inferencia, la cual se produce cuando un usuario es capaz de deducir, a partir de información trivial, información más sólida sobre una base de datos sin tener acceso directo a ella [Agrawal et al., 1993]. El objeto de la inferencia en términos contrarios a la ética consiste en juntar la información en un nivel de seguridad para determinar un hecho que debe ser protegido a un nivel de seguridad más alto.

La posibilidad de vincular la información publicada con otros datos de dominio público se ve potenciada por el aumento de métodos sofisticados para la vinculación de registros [Iyengar, 2002], métodos de análisis de enlaces [Bilenko et al., 2003, McCallum and Wellner, 2003] y la mayor disponibilidad de datos públicos de alta calidad. En muchos casos, los datos publicados se pueden emparejar con otros, utilizando métodos puramente manuales o automáticos [Malin et al., 2003, Esquivel-Quirós and Barrantes, 2015].

Para casos en los que se desea trabajar con datos relacionales, el problema de la extracción, emparejamiento y resolución de la entidad o individuo mencionado en los datos es un desafío desde hace mucho tiempo en la gestión de bases de datos, recuperación de información, aprendizaje automático, procesamiento del lenguaje natural y estadística [Getoor and Machanavajjhala, 2012]. Irónicamente, diferentes disciplinas se refieren a esta labor por una variedad de nombres, incluyendo el registro de vinculación, la de-duplicación, la resolución de la correspondencia, la reconciliación de referencia, la consolidación de entidades, la incertidumbre de identidad, el endurecimiento de la base de datos y ataques de inferencia [Kiran and Kavya, 2012, Henriksen-Bulmer and Jeary, 2016, Hamza and Hefny, 2013]. La resolución de una entidad o individuo tiene enormes implicaciones prácticas en una amplia variedad de dominios comerciales, científicos y de seguridad. Existe una diversidad de enfoques - incluyendo métodos basados en reglas de clasificación, enfoques de agrupación y formas de inferencia probabilística - y una gran falta de teoría que permita

una guía ordenada y eficiente para realizarla [Getoor and Machanavajjhala, 2012].

### Minería de datos

La minería de datos es la ciencia de extraer conocimiento útil de repositorios de datos. El proceso de descubrimiento de conocimiento y minería de datos (KDD, por sus siglas en inglés) consiste en la selección de datos, limpieza de datos, transformación y reducción de datos, minería, interpretación evaluación e incorporación del conocimiento generado en algún otro proceso [Chaudhuri et al., 2002, Intensive Working Group of ACM SIGKDD Curriculum Committee, 2017].

La minería de datos está basada en los siguientes pasos:

1. Selección del conjunto de datos: tanto en lo que se refiere a las variables objetivo.

2. Análisis de las propiedades de los datos: en especial los histogramas, diagramas de dispersión, presencia de valores atípicos y ausencia de datos.

3. Transformación del conjunto de datos de entrada (preprocesado y recodificación): se realizará de diversas formas en función del análisis previo. Estas formas incluyen la limpieza de datos, la cual incluye el proceso orientado a eliminar datos con ruido o incorrectos, la integración de datos, que trata de integrar diferentes fuentes de datos en un solo repositorio coherente y homogéneo. Así como la transformación de los datos como, por ejemplo, una normalización y la reducción de los datos, orientada a reducir el tamaño de los datos mediante la agregación y/o eliminación de características redundantes.

4. Seleccionar y aplicar un algoritmo de aprendizaje máquina: se construye el modelo predictivo, de clasificación o segmentación.

5. Extracción de conocimiento: mediante un algoritmo de aprendizaje máquina, se obtiene un modelo de conocimiento, que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables.

6. Interpretación y evaluación de datos: una vez obtenido el modelo, se debe proceder a su validación comprobando que las conclusiones que arroja son válidas y suficientemente satisfactorias.

El conocimiento generado por la minería puede ser utilizado en el proceso de asociación de los datos con sus dueños a partir de diferentes enfoques de riesgo. Dado que el riesgo de identificación es el riesgo de relacionar los identificadores suprimidos por medio de los datos desidentificados, se presentan tres escenarios de riesgo [El Emam and Dankar, 2008]:

1. Escenario del fiscal [acusador] *(Prosecutor Scenario)*: el riesgo de que una persona específica pueda ser identificada cuando el atacante sabe que están en el conjunto de datos.

2. Escenario del periodista *(Journalist Scenario)*: el riesgo de que al menos una persona pueda ser identificada.

3. Escenario del comerciante *(Prosecutor Scenario)*: El porcentaje de identidades que pueden ser identificadas correctamente.

Las técnicas de minería de datos permiten hacer ese riesgo tangible, para ello se apoyan constantemente en estadística descriptiva y algoritmos que permiten facilitar, así como automatizar, la extracción del conocimiento. Estos algoritmos generalmente son llamados algoritmos de aprendizaje máquina.

### El aprendizaje máquina

El aprendizaje máquina permite en muchos casos extraer información útil, interesante y previamente desconocida de grandes conjuntos de datos. El éxito siempre se basa en la disponibilidad de datos de alta calidad y el intercambio efectivo de estos. La recopilación de datos por parte de gobiernos, empresas e individuos ha creado un entorno que facilita el trabajo con algoritmos de aprendizaje máquina que nos permiten tomar múltiples decisiones. Asimismo, se ha logrado un impulso en la publicación de datos tanto por beneficio mutuo, como por regulaciones que obligan a la publicación de ciertos datos [Conradie and Choenni, 2012b]. La publicación de datos es omnipresente en muchos dominios. Por ejemplo, en el 2006, el proveedor de Internet AOL lanzó un conjunto de datos que contenían 3 meses de búsquedas de 650,000 usuarios. Los nombres fueron enmascarados con identificadores aleatorios y aún así, en cuestión de días, un reportero del New York Times identificó a Thelma Arnold, una viuda de 62 años como

el usuario 4,417,749 [Barbaro and Zeller, 2006]. Esto trajo muchas consecuencias a la empresa y a varios de sus empleados [Leoni, 2012]. Situaciones como está han impulsado a los investigadores en PPDP a utilizar algoritmos de aprendizaje máquina en sus investigaciones [LeFevre et al., 2006a, Machanavajjhala et al., 2007, Chen et al., 2008, Tian and Zhang, 2013].

En general, los algoritmos de aprendizaje máquina se pueden clasificar en algoritmos de aprendizaje supervisado o no supervisado (o una combinación de ambos). El aprendizaje supervisado es muy utilizado por investigadores en privacidad y utiliza datos etiquetados, donde cada conjunto de atributos está asociado con un atributo de salida que podría ser una clasificación (etiqueta de clase) o un valor continuo (regresión) [Al-Rubaie and Chang, 2019]. En estos modelos de aprendizaje máquina, los atributos se caracterizan típicamente en al menos uno de los siguientes tipos [Han et al., 2012]:

- *Atributo objetivo o de interés :* es el atributo nominal cuyo valor busca predecir con precisión el modelo de clasificación construido. En el caso de la regresión, es el atributo numérico cuyo valor tiene como propósito predecir el modelo de regresión construido.

- *Atributos de predicción:* son el conjunto de atributos (discretos o continuos, según el algoritmo) que se utilizan como entradas para construir el modelo que intenta predecir el atributo objetivo.

Algunos de los algoritmos supervisados más utilizados en trabajos de preservación de la privacidad son:

- *Regresión logística :* es un tipo de análisis de regresión utilizado para predecir el resultado de un atributo en función de los atributos de predicción. La regresión logística puede ser binaria o multinomial dependiendo de si la clasificación se realiza en dos o más grupos, respectivamente. La regresión logística es útil en situaciones en las que se desee predecir la ocurrencia de un suceso. Tiene una amplia aplicación en el área médica y social [Verma, 2013]. En general, la regresión logística es adecuada para describir y probar hipótesis sobre las relaciones entre un atributo objetivo y una o más atributos de predicción categóricos o continuos [Hastie et al., 2002].

- *Árbol de decisión:* es un modelo predictivo que se puede usar para representar clasificadores y modelos de regresión. Además los árboles de decisión refieren a un modelo jerárquico de decisiones y sus consecuencias. Se pueden emplear árboles de decisión para identificar la mejor estrategia para alcanzar el objetivo. Cuando se utiliza un árbol de decisión para las tareas de clasificación, se le puede denominar árbol de clasificación y cuando se utiliza para tareas de regresión, se le puede llamar árbol de regresión. Los árboles de clasificación se usan para clasificar un objeto o una instancia en función de sus atributos [Rokach and Maimon, 2014].

- *Bosques aleatorios:* es una modificación sustancial del embolsado (*bootstrap*), que es una técnica para reducir la varianza de una función de predicción estimada. Los bosques aleatorios construyen una gran colección de árboles no correlacionados, y luego los promedia. La idea esencial es promediar muchos modelos ruidosos pero aproximadamente no sesgados, y por lo tanto reducir la varianza. Como consecuencia, los bosques aleatorios son populares y se implementan en una variedad de paquetes [Hastie et al., 2009].

La utilidad de los datos requiere potenciar el uso de la información disponible y por ello se debe tomar en cuenta que los atributos no sensibles se publican si son importantes para la tarea de minería de datos. El atributo objetivo en la minería es definido como no sensible en el proceso de desidentificación para que conserve su semántica. Por lo general cuando se considera un algoritmo de clasificación o de regresión junto con la PPDP, cada atributo tiene solo dos caracterizaciones (atributo de predicción y atributo objetivo) [Esquivel-Quiros et al., 2018].

En el siguiente capítulo se describe la estructura de la metodología utilizada, con el propósito de proveer una visión general de los pasos seguidos en el desarrollo de la tesis.

# Capítulo 3

# Metodología

Este capítulo presenta una visión global de la metodología que se siguió para cumplir con los objetivos de tesis propuestos. Los detalles específicos se encuentran en los capítulos respectivos.

Para cumplir los objetivos propuestos, se siguieron varias etapas. En la primera etapa se definió un modelo de expresión de preferencias individuales de privacidad sobre inferencias posibles en una base de datos de una organización (sección 3.1). En la segunda se definió la instrumentación de un bloqueo de inferencia aplicada a un algoritmo de desidentificación (sección 3.2). La tercera etapa fue la de definir una métrica que identificara la eficacia del bloqueo de inferencia y realizar su evaluación (sección 3.3). La cuarta etapa trató con la definición de un caso de estudio para la aplicación de los productos de los pasos anteriores (sección 3.4). La última etapa consistió en diseñar y efectuar la evaluación del modelo de expresión de inferencias (sección 3.5). En la figura 3.1 se presenta un diagrama que ilustra la secuencia de etapas realizadas. En las siguientes secciones se describen generalidades de cada etapa.

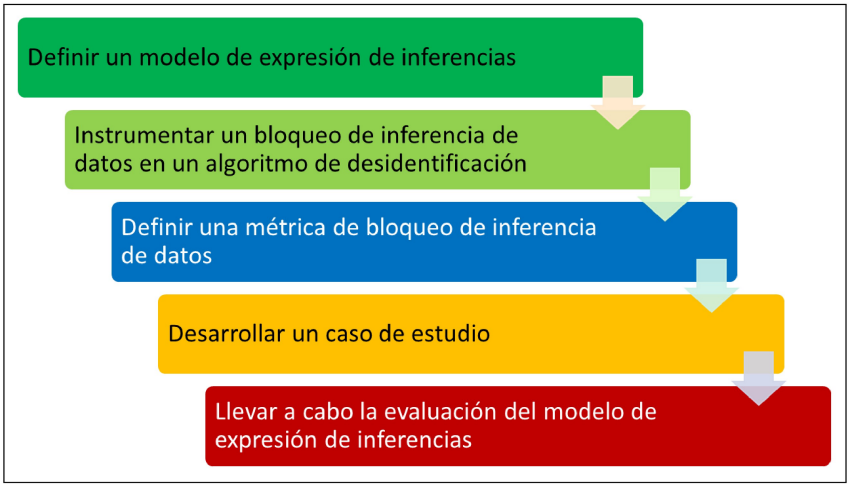



Figura 3.1: Diagrama de pasos descritos en la metodología

## 3.1. Definición de un modelo conceptual de expresión de preferencias individuales de privacidad

El primer paso consistió en definir y probar una forma de expresar las preferencias de privacidad de los dueños de los datos por medio de ejemplos de inferencias comunes que fueran realizables con datos almacenados en la base de datos relacional de una organización seleccionada. Este paso corresponde al cumplimiento del primer objetivo específico, por lo que se propone un modelo para traducir inferencias expresadas en lenguaje natural a una representación en la base de datos relacional de referencia. Dicha representación relaciona elementos de la base de datos correspondientes a una inferencia dada, y permite decidir (de forma probabilística) si una consulta determinada puede llegar a violar la preferencia de privacidad de cada dueño de datos. La metodología completa correspondiente a este paso se presenta en el capítulo 4.

## 3.2. Instrumentación de un bloqueo de inferencia de datos en un algoritmo de desidentificación

En el segundo paso se definió una metodología para instrumentar implementaciones de modelos de preservación de la privacidad, y se ejemplificó con el caso del modelo de $k$-anonimato. Esta instrumentalización se puede realizar de múltiples formas, pero siempre debe implicar la capacidad de aplicar un grado de privacidad diferente a cada dueño de datos. La metodología de este paso se describe a completitud en el capítulo 5.

## 3.3. Definición y evaluación de una métrica de bloqueo de inferencias

El tercer paso consistió en la definición y evaluación de una métrica que permitiera identificar la eficacia de un bloqueo de inferencia de datos, en otras palabras, una forma de medir si es posible realizar una inferencia específica sobre un conjunto de datos publicado. La definición de la métrica corresponde al objetivo específico 3.

La métrica se seleccionó para medir la tasa de éxito del bloqueo de inferencias solicitado por los dueños de los datos en un conjunto de datos publicado. Para esto se utilizó un algoritmo de aprendizaje máquina que permitió indicar el porcentaje de exactitud al predecir la información a proteger. La explicación sobre los pasos seguidos en la definición de la métrica se desarrollan en el capítulo 6.

## 3.4. Desarrollar un caso de estudio

El cuarto paso consistió en desarrollar un caso de estudio desde la identificación de inferencias de interés hasta la extracción de los datos que responden a una solicitud de publicación de datos, y corresponde al objetivo específico 2. La metodología de este paso se explica en el capítulo 7.

## 3.5. Evaluación del modelo de expresión de inferencias

El quinto paso fue evaluar la eficacia del proceso de desidentificación instrumentado para el caso de estudio desarrollado. Este paso corresponde al cuarto objetivo específico. Se desarrolló un análisis desde la perspectiva de dos poblaciones (con preferencias de privacidad individuales y sin preferencias de privacidad individuales). Esto permitió validar la eficacia del cumplimiento del consentimiento informado de los dueños de los datos y además se verificaron las diferencias en la utilidad de los datos al aplicar o no las preferencias de privacidad. La metodología detallada se presenta en el capítulo 8.

En el siguiente capítulo se presenta un modelo que permite expresar posibles inferencias sobre consultas a una base de datos relacional, de forma que permitan tomar en cuenta las preferencias individuales de privacidad de los dueños de los datos, en cumplimiento del primer objetivo.

# Capítulo 4

# Modelo de expresión de inferencias

Como se indicó en el capítulo 1, las inferencias con las que se trabaja en esta tesis están expresadas en lenguaje natural para poder presentárselas a un dueño de datos. Esta forma de presentación es muy útil para expresar de forma clara una preferencia de privacidad, ya que no requiere conocimiento técnico, y permite visibilizar preocupaciones de privacidad sobre las que el dueño de los datos no estuviera consciente. Sin embargo, esta expresión en lenguaje natural no significa absolutamente nada en términos de los componentes de una base de datos relacional. Debido a ello, se requiere sistematizar una manera de traducir la inferencia en términos de estos componentes. Precisamente, este capítulo presenta un modelo de expresión de posibles inferencias sobre una base de datos relacional.

En la sección 4.1 se describe el proceso de selección de inferencias posibles en la base de datos y su representación en lenguaje natural. De forma análoga a la labor del experto en la preservación de la privacidad básica, donde el mismo analiza y clasifica los datos en identificadores, sensibles, y cuasi-identificadores (ver capítulo 2), o decide sobre parámetros de los algoritmos de desidentificación, al manipular inferencias se debe utilizar conocimiento experto para determinar un conjunto de inferencias que se pueden realizar con los datos disponibles. Este experto será típicamente el administrador de la base de datos de la organización que almacena los datos.

Una vez identificado un grupo de inferencias, será también rol del mismo experto determinar qué elementos de la base de datos facilitan la realización de la inferencia, pero de una manera sistematizada. La metodología de traducción de las inferencias en lenguaje natural a reglas interpretables en una base de datos se presenta en la sección 4.2. Esta metodología permite mapear las inferencias creadas en lenguaje natural a elementos reconocibles en la base de datos.

Finalmente, la sección 4.3 trata sobre cómo capturar las preferencias de bloqueo de inferencias de los dueños de datos, cómo representar sus escogencias en la base de datos, y «lo más importante» cómo determinar cuáles deben ser bloqueadas en un conjunto

de datos a publicar. Se muestra una posible interfaz para capturar las preferencias de bloqueo de los dueños de los datos con respecto al conjunto de inferencias planteado. Un dueño de los datos puede realizar esta activación, inferencia por inferencia, de forma similar a un consentimiento por columnas. La diferencia con la autorización convencional es que una inferencia es semánticamente más clara para el dueño de los datos y permite que reciba más información sobre los peligros a los que se expone al compartir los datos, sin requerir una comprensión de la estructura de la base de datos.

Se inicia con la descripción del proceso de selección de inferencias, a continuación.

## 4.1. Selección de inferencias aplicables a una base de datos

Una inferencia es un concepto abstracto y dependiente en gran medida del conocimiento base utilizado para realizarla [Shokri et al., 2017]. Como se indicó en el capítulo 2, una inferencia se produce cuando un atacante es capaz de deducir a partir de información poco sensible, otra que si lo es (sensible) para el dueño de los datos, sin tener que acceder directamente a ella. En esta tesis se supone que las inferencias se refieren al contexto particular de la base de datos. Un ejemplo de inferencia referida a una base de datos estudiantil es "Deducir su horario de clases en este semestre" (dado que el dueño de los datos en este caso seria un estudiante).

El proceso requerido para seleccionar un grupo de inferencias relevantes para los dueños de los datos es una tarea compleja. Como se indicó en la sección 1.4 (alcances), escapa al ámbito de esta tesis definir una metodología o proceso automatizado para realizar dicha selección. En lugar de esto, se plantea que el proceso permita la selección de un grupo de inferencias de forma manual, con base a estudios previos de preferencias de privacidad de una población (uso de cuestionario o encuestas), y utilizando el conocimiento experto sobre la base de datos.

Es importante destacar que una inferencia puede surgir del conocimiento experto del encargado de la base de datos. Por ejemplo, la inferencia anterior se puede llevar a cabo uniendo información de los horarios de los cursos del semestre, la matricula del estudiante o su plan de estudios y su avance en dicho plan. Este mapeo requiere conocimiento sobre la base de datos original. De la misma forma, se pueden crear otras

inferencias, que en todos los casos son dependientes del contexto y del conocimiento base para su realización. Estas inferencias se pueden extraer asimismo por medio de entrevistas o encuestas a los dueños de los datos con el fin de conocer sus preocupaciones de privacidad.

Para este propósito se tiene que utilizar una taxonomía de privacidad, como la planteada por Song et al. [Song et al., 2018], la cual incluye atributos personales, relaciones, actividades, ubicación, emoción, atención médica, hitos de la vida y declaraciones neutrales o alguna propia. Dada la finalidad genérica de este modelo se recomienda crear una taxonomía propia para la base de datos en estudio.

La taxonomía permite agrupar formalmente la información relacionada y esto colabora en la creación de inferencias menos específicas. Como se muestra en la figura 4.1, producto del uso del conocimiento de la base de datos, de la taxonomía de privacidad y del conocimiento experto (conocimiento del contexto y arquitectura de la BD) se obtiene un conjunto de inferencias en lenguaje natural que permiten al dueño de los datos conocer los riesgos a los que se ve expuesto en la publicación.

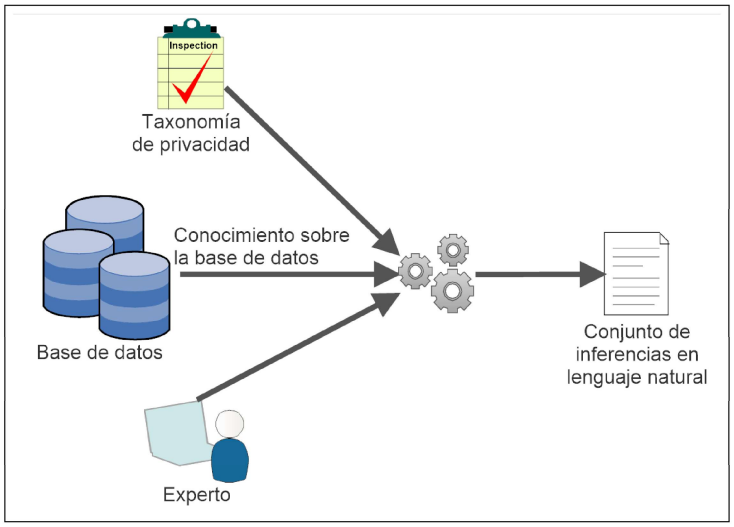



Figura 4.1: Selección de inferencias.

## 4.2. Traducción de inferencias a reglas y elementos interpretables en una base de datos

En la sección anterior se presentó como definir las inferencias a presentar al dueño de los datos. Cada inferencia tiene múltiples formas de realizarse con los datos disponibles en la base de datos. Tal como se indicó en la sección 1.4 (alcances) escapa al ámbito de esta tesis la generación automática de las reglas que permiten traducir una inferencia a elementos de la base de datos.

El modelo que se describe en esta tesis utiliza conocimiento experto sobre la base de datos para definir, de forma manual, sobre cada inferencia, una lista de reglas relacionadas a atributos, valores, tablas, relaciones y otros aspectos de la base de datos. Estas reglas intentan identificar los elementos necesarios que podrían ser utilizados por un atacante, para que, en conjunto con información externa, dicho atacante pueda realizar una inferencia, en caso que alguno de esos elementos sea publicado.

Es importante destacar que el publicador de datos no requiere realizar ninguna inferencia debido a que posee toda la información. El publicador de datos debe realizar el ejercicio de buscar con qué reglas generar las inferencias identificadas sobre su propia base de datos, con el objetivo de identificar y marcar elementos que pueden ser utilizados por un atacante en el contexto de una publicación futura. Si una publicación alerta algunos de estos elementos, es entonces que se deben aplicar las preferencias indicadas por el o los dueños de los datos.

En la figura 4.2 se ilustra este concepto, donde podemos observar a la izquierda varias esferas que representan diferentes inferencias posibles, las cuales están asociadas con consultas SQL que aparecen al centro de la figura. Cada una de las consultas SQL representa una forma diferente de llevar a cabo una inferencia específica y de ahí la asociación. El conjunto de consultas SQL provee diferentes formas de mapear en la base de datos las características y valores que permiten realizar una inferencia sobre la base de datos. Para cada consulta se debe generar un listado con los componentes asociados a cada inferencia.

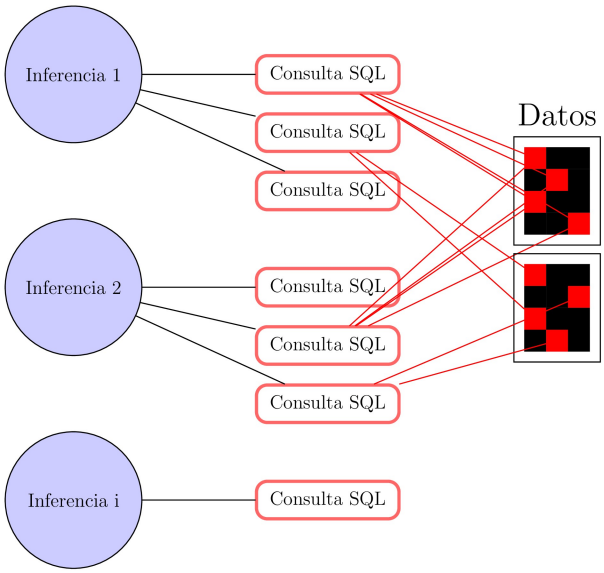



Figura 4.2: Traducción de inferencias de datos a reglas en la base de datos de estudio

Para ilustrar el concepto de la traducción de inferencias a reglas interpretables en la base de datos, se presenta un pequeño ejemplo. Supongamos que la inferencia es "Deducir su horario de clases en este semestre". El publicador ahora debe proponer varias consultas SQL alternativas que permitan obtener esa información, dado su conocimiento de la base de datos completa. Una de las posibles consultas SQL se presenta en la figura 4.3. La consulta extrae los días, horas, cursos, aulas y edificios que están relacionados con cada estudiante, permitiendo reconstruir el horario de cualquiera de estos estudiantes.

```
1 SELECT MA.CARNE,MA.ANNO,MA.PERIODO,MA.SIGLA,MA.GRUPO,HO.DIA, HO.
      HORA_INICIO, HO.HORA_FIN ,HO.EDIFICIO,HO.AULA
2 FROM MA_MATRICULA MA, HO_HORARIOS HO , MA_CONDICION_CURSO CO
3 WHERE MA.ANNO = HO.ANNO
4 AND MA.PERIODO = HO.PERIODO
5 AND MA.RECINTO = HO.RECINTO
6 AND MA.SIGLA = HO.SIGLA
7 AND MA.GRUPO = HO.GRUPO
8 AND MA.MODALIDAD_ESTUDIO = 1
9 AND MA.CONDICION_CURSO = CO.CONDICION_CURSO
10 AND CO.INDICADOR_ASIGNADO = 'A'
11 AND MA.ANNO = 2019
12 AND MA.PERIODO = '1'
```

Figura 4.3: Una de las consultas SQL que permite ejecutar la inferencia “Deducir su horario de clases en este semestre”.

Esta consulta SQL permite identificar elementos de la base de datos como las tablas MATRICULA y HORARIOS, así como los atributos utilizados, como SIGLA, GRUPO, HORA_INICIO, HORA_FINAL, AULA y EDIFICIO. También permite identificar valores de atributos que compartirán todas las tuplas de la consulta, como el año y el período. Todos estos elementos deben ser representados de forma abstracta, para que en una posterior publicación de datos sean fácilmente identificables. Este proceso debe realizarse para cada consulta SQL que se propuso como forma de lograr la inferencia.

Recapitulando, a cada inferencia se le asocian varias consultas SQL que permiten su realización. Para cada consulta SQL se extraen sus componentes principales, según lo explicado anteriormente. A continuación se explica como se representan los componentes asociados a cada inferencia.

Los componentes asociados a cada inferencia se representan mediante una estructura XML. La estructura se organizó de la siguiente manera: “DESCRIPTION” donde se escribe una descripción en lenguaje natural que describe la inferencia a bloquear, “RULE” donde se indica el inicio de una regla y su comportamiento (activo/inactivo) en el contexto de la base de datos. “TABLE” donde se indica el nombre de la tabla que debe disparar la alerta. Cada sección “TABLE” posee una o más secciones “FIELD” que permiten indicar los atributos específicos que deben tomarse en cuenta. También se

cuenta con una sección “VALUE” donde se presentan los atributos y valores que deben disparar alertas especificas por criterios de privacidad del publicador de los datos.

Siguiendo el ejemplo de la inferencia realizada a partir de la consulta SQL de la figura 4.3 se obtiene la estructura XML de la figura 4.4. En esta, se puede observar una etiqueta “DESCRIPTION” que contiene una descripción en lenguaje natural. También se observa una regla presentada por la etiqueta “RULE” que indica si el comportamiento de la regla es activo o inactivo. Esta regla posee dos etiquetas “TABLE”, una para cada tabla (MA_MATRICULA y HO_HORARIOS) de la base de datos. Cada etiqueta “TABLE” indica todos los atributos que se deben tomar en cuenta por medio de las etiquetas “FIELD”. La etiqueta “VALUE” en este caso indica que el atributo “PERIODO” debe poseer el valor 1 y el atributo “ANNO” el valor 2019, para que se active la regla. Esta representación es fácilmente extensible para incluir nuevos componentes de la base de datos conforme se mejora el entendimiento del problema y se puede almacenar en la misma base de datos del publicador de los datos.

```
<INFERENCE_DEDUCIR_HORARIO>
  <DESCRIPTION>
    Deducir el lugar de residencia de un estudiante
  <DESCRIPTION>
  <RULE behavior="active">
   <TABLE name = MA_MATRICULA>
        <FIELD>ANNO</FIELD>
        <FIELD>PERIODO</FIELD>
        <FIELD>RECINTO</FIELD>
        <FIELD>SIGLA</FIELD>
        <FIELD>GRUPO</FIELD>
   </TABLE>
   <TABLE name = HO_HORARIOS>
        <FIELD>ANNO</FIELD>
        <FIELD>PERIODO</FIELD>
        <FIELD>RECINTO</FIELD>
        <FIELD>SIGLA</FIELD>
        <FIELD>GRUPO</FIELD>
        <FIELD>DIA</FIELD>
        <FIELD>HORA_INICIO</FIELD>
        <FIELD>HORA_FINAL</FIELD>
   </TABLE>
   <VALUE>
        <PERIODO>1</PERIODO>
        <ANNO>2019</ANNO>
   </VALUE>
  </RULE>
</INFERENCE_DEDUCIR_HORARIO>
```

Figura 4.4: Regla para la inferencia: “Deducir su horario de clases en este semestre”

Los pasos descritos en la sección 4.1 y en está sección ilustran el proceso presentado en la figura 4.5. Este proceso detalla el camino para crear las reglas asociadas a la inferencia. Partiendo desde la base de datos del publicador, mediante el uso de la taxonomía de privacidad, el conocimiento de la base de datos y el experto, se logran seleccionar las inferencias. Luego para cada inferencia se escriben consultas SQL que permiten realizarlas a partir del conocimiento de la base de datos. Estas consultas son procesadas para extraer sus componentes principales y finalmente generar reglas en

formato XML que se utilizarán para detectar la inferencia.

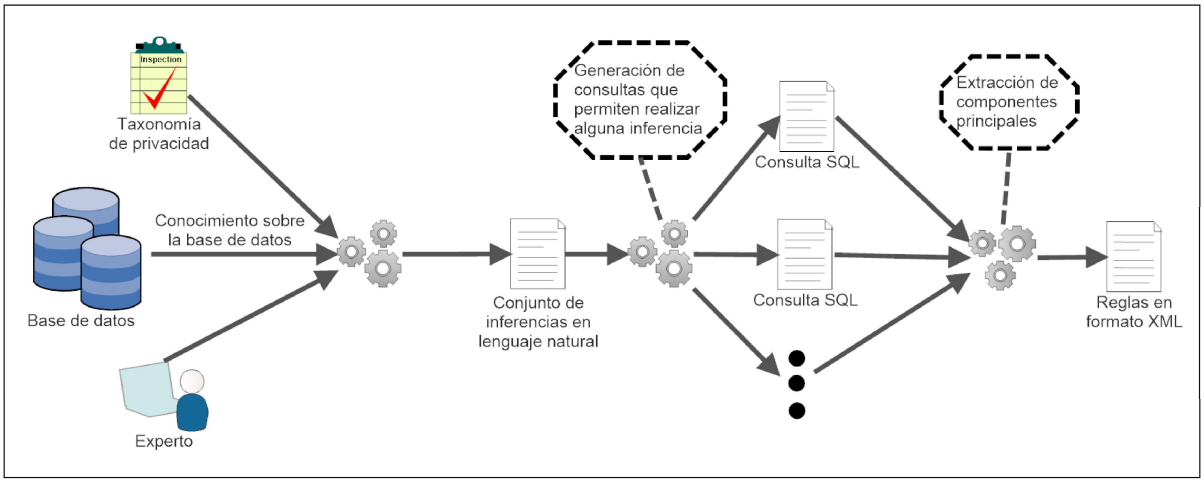



Figura 4.5: Traducción de inferencias a reglas.

Dado que ya están definidas las reglas asociadas a la inferencia, se procede a explicar la aplicación de dichas reglas. La aplicación de las reglas se debe realizar cada vez que se requiera generar una publicación de datos, por cuanto las reglas se refieren a qué debe protegerse al publicar los datos.

Para realizar una publicación de datos se requiere realizar una consulta SQL que permita obtener los datos a publicar. A esta consulta se le extraen los elementos de la base de datos relevantes. Estos elementos se comparan con los elementos definidos en las reglas de cada inferencia. Cada inferencia contiene una o varias reglas que deben ser verificadas. Como es muy posible que la regla sea activada parcialmente, se deben definir umbrales de activación. Por ejemplo, si dos de cinco atributos que están presentes en la regla de la inferencia no están presentes en la consulta a publicar, se debe tomar la decisión de activar o no, bajo algún criterio. Los umbrales que definen este criterio pueden ser mínimos (con solo un elemento de la base de datos, se activa la inferencia) o se pueden definir con conocimiento experto.

El paso anterior permite realizar la activación de las reglas que permiten realizar el bloqueo de la inferencia sin ejecutar la consulta SQL. Sin embargo, en algunos casos es requerido analizar la totalidad de los datos generados por la consulta. Por ejemplo,

si una regla indica que se activa cuando el valor del atributo “ENFERMO_VIH” es positivo, se requiere el acceso a los datos que genera la consulta para identificar si existen valores que cumplen la regla. Por lo tanto, para completar el proceso de activación de bloqueo de inferencias se requiere generar los datos y en algunos casos analizar valores de atributos específicos. Una optimización del proceso consiste en solo realizar el análisis de los datos cuándo existan reglas que lo requieran.

El conjunto de inferencias activadas o conjunto de activaciones de bloqueo de inferencias se debe almacenar por dos razones. La primera responde a que es el insumo para establecer el bloqueo de inferencia (descrito en el capítulo 5) y la segunda corresponde a una optimización del procesado de la consulta SQL, suponiendo que se pueden realizar múltiples consultas similares.

En la figura 4.6 se representa de forma abstracta el proceso de activación de bloqueo inferencias que se presentó anteriormente.

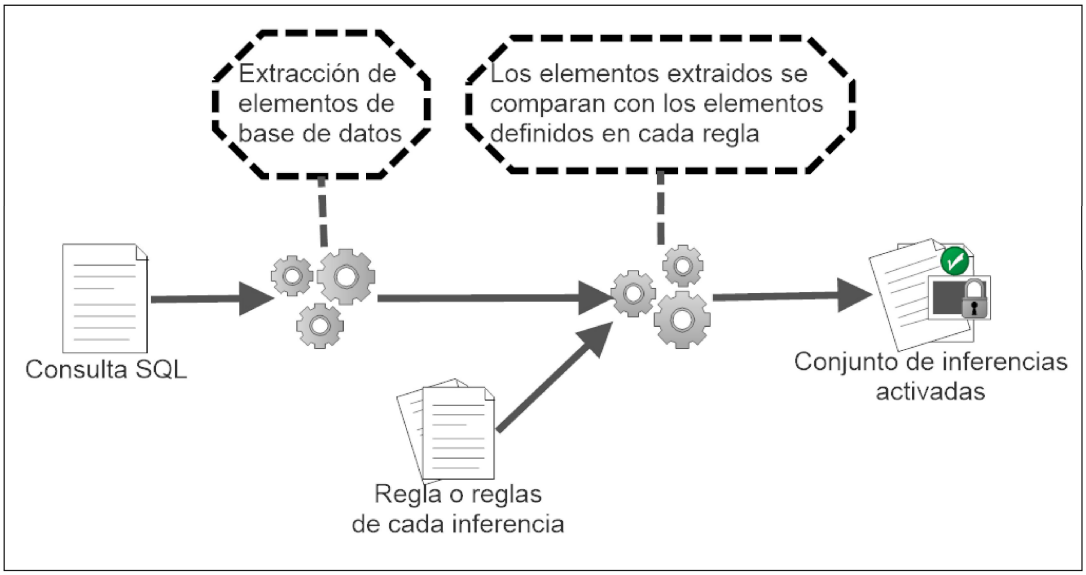



Figura 4.6: Activación de bloqueos de inferencias.

En la sección 4.3 se aborda como representar y usar la selección de inferencias para los dueños de los datos.

## 4.3. Representación y uso de las preferencias de privacidad de los dueños de los datos

En esta sección se describe cómo se obtienen las inferencias a ser bloqueadas según las preferencias de los dueños de los datos. En primer lugar se explica cómo los dueños de los datos (por ejemplo, estudiantes) seleccionan cuáles de las inferencias identificadas en la sección 4.1 deben ser bloqueadas según sus preferencias. Luego se explica cómo se almacena esa selección. Finalmente se describe el proceso que determina cuáles de las inferencias seleccionadas para su bloqueo aplican para un conjunto especifico de datos a publicar.

La activación o selección de las inferencias a bloquear se realiza mediante el consentimiento informado, donde por medio de una interfaz de usuario se presentan un conjunto de inferencias que se puedan realizar sobre los datos del usuario (dueño de los datos). Para cada inferencia el usuario debe decidir si desea activar o desactivar su bloqueo. El proceso por medio del cuál se definieron estas inferencias fue explicado en la sección 4.1. En la figura 4.7 se muestra una posible interfaz de activación de bloqueos de inferencias. La activación implica que se debe bloquear dicha inferencia.




Figura 4.7: Ejemplo de una interfaz de usuario para activación de bloqueo de inferencias.

Para cada usuario, la selección de inferencias a bloquear se representa como una estructura que almacena el identificador de cada uno de los dueños de los datos, una referencia a cada una de las inferencias y un indicador de la selección de cada usuario sobre la activación del bloqueo sobre cada inferencia. A nivel abstracto, esta estructura es una matriz de preferencias de privacidad como la que se muestra en la figura 4.1.

Cuadro 4.1: Matriz de preferencias de privacidad sobre bloqueos a inferencias.

| Productor / Inferencias | Estudiante #1 | Estudiante #2 | ... | Estudiante #N |
|---|---|---|---|---|
| Deducir su horario de clases en este semestre | Si | No | ... | No |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| Inferencia #N | Si | No | ... | Si |

Una vez almacenadas las preferencias de bloqueo de inferencias de los usuarios dueños de los datos, se explicará cómo se usan estas preferencias para informar los procesos de bloqueo de inferencias. Es importante destacar que los pasos anteriores consistieron en la recolección del conocimiento necesario para poder identificar y caracterizar una inferencia de datos. Así como la selección de preferencias de los dueños de los datos. Todo este conocimiento se mantiene almacenado en la BD del publicador y solo será requerido cuando se intente realizar una publicación de datos.

Como se puede observar en la figura 4.8, cuándo se solicita una publicación de datos se genera una consulta SQL. Esta consulta SQL permite identificar los elementos de base de datos que se utilizan para ejecutarla. Una vez identificados estos elementos se realiza un emparejamiento con los elementos correspondientes en las reglas especificadas para cada inferencia (identificadas en el capítulo 4.2). Esto permite conocer el conjunto de bloqueos de inferencia activados por elementos de la consulta SQL y cuáles reglas los activaron. La segunda parte del proceso requiere identificar los dueños de cada una de las tuplas de la publicación y determinar que bloqueos de inferencia tiene activos cada dueño de datos. Una vez realizados ambos procesos se revisan las coincidencias entre el conjunto de bloqueos de inferencia activados por elementos de la consulta SQL y los bloqueos de inferencia activados por cada dueño de datos con tuplas en la publicación,

así se determina que bloqueos de inferencia se deben realizar para cada dueño de datos en la publicación de datos a realizar (ver figura 4.8). El conjunto resultado contiene las únicas inferencias que se deben bloquear y es este resultado el que se pasa como insumo al proceso de bloqueo que se desarrolla en el capítulo 5. La figura 4.8 describe el uso de las reglas bloqueo de inferencias que se generaron en la sección 4.2 y los pasos descritos en está sección.

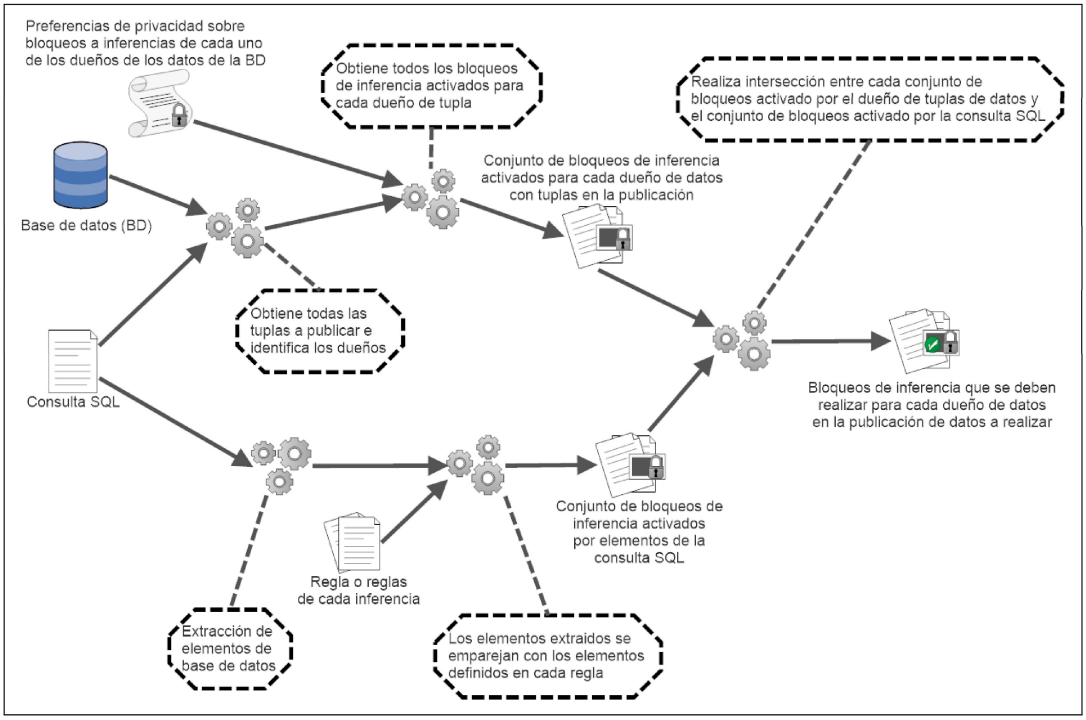



Figura 4.8: Activación de bloqueos de inferencias utilizando la consulta SQL y las preferencias de privacidad.

Para comprender mediante un ejemplo el uso del modelo, se puede tomar la inferencia “Deducir su horario de clases en este semestre” que se ha desarrollado a lo largo de este capítulo y suponer que se intenta realizar una publicación generada con la consulta SQL de la figura 4.9. Las reglas descritas anteriormente en la figura 4.4 logran activarse al coincidir en el uso de las tablas “MA_MATRICULA” y los campos “ANNO”, “PERIODO”, “SIGLA” y “GRUPO”. Se puede notar que no coinciden en el uso de todas

las tablas, ni todos los campos. También, respecto a los valores identificados en la regla, estos si logran coincidir totalmente. Ahora, si suponemos que existe o se definió un umbral de activación respecto a la existencia de al menos una coincidencia, este umbral de activación del bloqueo de la inferencia se encontraría aprobado y por lo tanto el bloqueo de inferencia estaría activo para la consulta SQL de la figura 4.9.

```
1  SELECT MA.CARNE,MA.ANNO,MA.PERIODO,MA.SIGLA,MA.GRUPO,
2  CASE WHEN E.GENERO = '1' THEN 'MASCULINO' WHEN E.GENERO = '2' THEN '
       FEMENINO' END AS GENERO,
3  P.DESCRIPCION AS PROVINCIA
4  FROM MA_MATRICULA MA, MA_CONDICION_CURSO CO, ESTUDIANTE E,
       INFORMACION_PERSONAL I, PROVINCIA P
5  WHERE E.CARNE = MA.CARNE AND E.CARNE = I.CARNE AND I.PROVINCIA = P.
       PROVINCIA (+)
6  AND I.SECUENCIA_VIGENCIA = (SELECT MAX (I.SECUENCIA_VIGENCIA) FROM
       INFORMACION_PERSONAL I WHERE I.CARNE = E.CARNE)
7  AND MA.MODALIDAD_ESTUDIO = 1
8  AND MA.CONDICION_CURSO = CO.CONDICION_CURSO
9  AND CO.INDICADOR_ASIGNADO = 'A'
10 AND MA.ANNO = 2019
11 AND MA.PERIODO = '1'
```

Figura 4.9: Consulta ejemplo de publicación.

Una vez identificados los bloqueos de inferencia activados por la consulta SQL, se debe determinar cuales tuplas corresponden a dueños de datos con preferencias de bloqueo de alguna de esas inferencias. Este proceso identifica a que dueño de datos que corresponde cada tupla y verifica cuáles de las preferencias de bloqueo activadas por la consulta SQL le corresponden. Esto facilita el marcar o identificar cada tupla específica a la que se deben aplicar los bloqueos de inferencia. Un ejemplo de este marcado se puede observar en la figura 4.2, donde se puede ver en otro color los atributos y tuplas identificados por las reglas de bloqueo de inferencias. Se asume que solo un dueño de datos preseleccionó activar el bloqueo de la inferencia en cuestión (se oculta intencionalmente en la muestra el atributo identificador carné para no exponer a ningún estudiante).

| CARNE | AÑO | PERIODO | SIGLA | GRUPO | GENERO | PROVINCIA |
|---|---|---|---|---|---|---|
| ****** | 2019 | 1 | LM5032 | 1 | 2 | PUNTARENAS |
| ****** | 2019 | 1 | LM5032 | 1 | 2 | PUNTARENAS |
| ****** | 2019 | 1 | LM1030 | 3 | 1 | SAN JOSÉ |
| ****** | 2019 | 1 | LM1030 | 3 | 1 | SAN JOSÉ |
| ****** | 2019 | 1 | QU0100 | 2 | 1 | SAN JOSÉ |
| ****** | 2019 | 1 | QU0100 | 2 | 1 | SAN JOSÉ |

Cuadro 4.2: Marcado de tuplas por la inferencia "Deducir su horario de clases en este semestre"

## 4.4. Resumen

En este capítulo se inició con la definición de un modelo de representación de inferencias, para lo cual se definieron procesos para:

- Seleccionar inferencias que tengan sentido en el contexto de una base de datos específica
- Identificar los elementos de la base de datos que inciden en la realización de una inferencia y se propuso un método para almacenar dicho conocimiento para su uso fácil posteriormente
- Almacenar las preferencias de bloqueo de los dueños de los datos.
- Relacionar consultas de publicación de datos con los elementos asociados a las inferencias y con las preferencias de bloqueo de los dueños de los datos

Con el modelo que se presentó, se cumple el objetivo específico 1: *"Definir un modelo para expresar posibles inferencias sobre consultas a una base de datos relacional que permitan tomar en cuenta las preferencias individuales de privacidad"*.

Una vez que se han determinado los datos afectados por bloqueos a inferencias que preseleccionaron los dueños de los datos, se debe proceder a aplicar los bloqueos

en sí, esto es, a instrumentar una técnica de desidentificación para que se respeten las preferencias de los dueños de las tuplas marcadas. En el capítulo 5, que sigue, se describe la aplicación del proceso de bloqueo de inferencias con base en el insumo que se provee de la selección de bloqueos de inferencia por parte de los dueños de los datos y de los bloqueos de inferencias activados por la publicación que se explicaron en este capítulo.

# Capítulo 5

# Bloqueo de inferencia de datos

En el capítulo anterior se explicó cómo identificar inferencias y cómo mapearlas a elementos de la base de datos. Una vez identificados los componentes de base de datos que pueden llevar a una inferencia, almacenadas las preferencias de bloqueo de inferencias de los dueños de los datos y marcadas las tuplas afectadas por estas preferencias en una posible publicación, se debe definir la forma en que el método de preservación de la privacidad utilizado por el publicador se verá afectado por las preferencias de privacidad de los usuarios dueños de los datos.

## 5.1. Aplicación de bloqueo de preferencias individuales a métodos de desidentificación

Como se ha explicado en diferentes momentos, el modelo que se presenta en esta tesis supone que el publicador de datos aplica algún método de preservación de la privacidad estándar antes de realizar la publicación final. El aporte de esta tesis es mostrar cómo se puede aumentar el método de preservación de la privacidad utilizado por el publicador para tomar en cuenta preferencias individuales de privacidad.

No se puede suponer nada sobre cuál método de preservación de la privacidad va a estar utilizando un publicador en particular. Sin embargo, existen una serie de pasos básicos que todos los métodos de PPDP que utilizan modelos de preservación de la privacidad sintáctica deben cumplir, tal como la eliminación de las columnas que contienen identificadores directos. En esta tesis se asume que los datos con los que se trabajará ya fueron procesados con estos pasos y en cualquier experimento o ejemplo que se reporte aquí se trabajará con datos que ya fueron sanitizados de esta manera [Shah, 2012]. Esto aplica para todos los datos que revele un publicador.

La implementación de un modelo de preservación de la privacidad, para el problema en cuestión, debe tomar como línea base estándar de preservación de la privacidad

los pasos básicos de PPDP y ampliar esta línea base. La ampliación implica proveer garantías extra, que se requieren para cumplir las preferencias de privacidad de los dueños de los datos que desean bloquear inferencias.

En general, el método de PPDP estándar que se esté utilizando se puede aumentar haciendo más estrictos algunos parámetros o modificando algunas operaciones de privacidad o incluso agregando algunas operaciones extra. En este trabajo se explota esta característica para aumentar la capacidad del PPDP, pero especialmente para los dueños de los datos que posean preferencias de privacidad. Para ejemplificar esto, se utilizó el modelo de privacidad conocido como k-anonimato que fue explicado en el capítulo 2. El parámetro más sencillo de modificar en dicho modelo es el valor de k, debido a que se conoce que, para valores de k más altos, se obtiene una mayor protección de la privacidad. Como acercamiento preliminar se puede asumir que simplemente se aplica un valor de k mayor para los datos relacionados a los dueños con preferencias de privacidad sobre bloqueos de inferencias. Sin embargo, esta modificación en la práctica es compleja, dado que requiere modificar el comportamiento del algoritmo de k-anonimato que se encuentre en uso para que en lugar de crear grupos de datos de tamaño **k**, cree algunos grupos de datos de tamaño **k** y otros grupos de datos de tamaño $\mathbf{k + X}$, donde **X** es un grado de aumento definido por medio de algún conocimiento experto. Para complicar aún más las cosas los dueños de los datos con preferencias pueden no estar distribuidos uniformemente en los datos y por lo tanto se complica aún más la creación de la segregación de los grupos del k-anonimato.

Lo descrito en el párrafo anterior ilustra algunas de las complejidades de la modificación a un método de PPDP para que respete las preferencias de privacidad de los dueños de los datos. Por lo tanto la modificación del PPDP estándar debe realizarse verificando qué partes del algoritmo subyacente se verán afectadas y cómo se realizaran esas afectaciones. En otras palabras, para cada implementación de PPDP se debe realizar un análisis particular. En esta tesis se presenta el análisis para el modelo de k-anonimato, con la modificación específica del parámetro k, aunque existen otras posibilidades de modificación.

## 5.2. Implementación de modificación al k-anonimato

Como se explicó en el capítulo 2, el k-anonimato es un modelo de preservación de la privacidad con un objetivo de privacidad definido por un parámetro k sobre el conjunto de columnas cuasi-dentificadoras [Pierangela Samarati, 2001]. El k-anonimato se puede ver como una propiedad de una tabla T. Se dice que una tabla T es k-anónima si, para cualquier combinación de valores que los campos cuasi-identificadores tienen presentes en la tabla, hay al menos k tuplas con la misma combinación [Sweeney, 2002b]. En general, las tablas no poseen esta propiedad de forma natural. Por lo tanto, se debe aplicar un proceso para forzar la propiedad sobre ellas. La forma más común de lograr esto es mediante el uso de operaciones de generalización y/o supresión (explicadas en el capítulo 2).

En el caso de esta tesis, para generalizar una tabla dada, el procedimiento seguido fue iterar sobre las tuplas asignando los elementos de cada cuasi-identificador en la tupla a una clase de equivalencia temporal que debe llegar a un valor **k base**. Esto permite que si se desea modificar el tamaño de esta clase, sea más sencillo. Así, si hay una fila que requiere mayor privacidad (posee preferencia de bloqueo de inferencia), lo cuál en el contexto del k-anonimato significa un aumento del valor de $k$, el tamaño de destino de la clase cambiará a **k base + X**, lo que en esta tesis llamaremos **k privado**. La iteración se detiene cuando la clase de equivalencia temporal ha alcanzado su tamaño objetivo o no comparte valores con una nueva fila, esto se explica con más detalle en los siguientes párrafos.

El resultado de cada iteración se debe guardar en un diccionario y crear una nueva clase de equivalencia temporal cada vez que se logré el valor de $k$ objetivo. Si al guardar una clase de equivalencia no se ha alcanzado su valor de $k$ objetivo, se fusiona con la o las clases de equivalencia anteriores de la misma jerarquía de generalización hasta alcanzar un tamaño igual o superior al $k$ objetivo. Si todos los elementos de una determinada jerarquía no son suficientes para alcanzar el tamaño objetivo, la jerarquía de generalización de la clase subirá un nivel dentro del árbol de generalización y el proceso se repetirá. A continuación se presentará un pequeño ejemplo para ilustrar esta modificación.

El problema ejemplo consiste en generalizar, un conjunto de tuplas con los cuasi-

identificadores Provincia, Cantón y Distrito, y las preferencias de privacidad (preferencias de bloqueo de inferencias) de los dueños de cada tupla. Esta generalización estará guiada por la semántica jerárquica de la división administrativa geográfica de Costa Rica. Por lo tanto, el triplete Provincia, Cantón y Distrito se generaliza como un único atributo cuasi-identificador. La generalización preserva la composición geográfica, de modo que una provincia se compone de cantones únicos y cada cantón se compone de distritos únicos. Por lo tanto, los distritos son las hojas del árbol de generalización y las tuplas originales se asocian con ellos. El árbol de generalización que permite realizar el proceso de creación de cada clase de equivalencia se muestra en la figura 5.1.

Para esta generalización se define el valor **k base** en 2 y el valor para el **k privado** en 3 (lo cuál significa que X es 1). Los elementos azules en las hojas representan a las tuplas con requisitos de privacidad básicos y los elementos rojos a las tuplas que requieren mayor protección de privacidad. En la rama izquierda, el nodo Nicoya tiene 5 registros, por lo que cumple con el requerimiento de **k base**, lo mismo ocurre con los nodos Nacascolo y Mayorga. En el caso del nodo Barrantes, cumple con el requisito **k base**, pero el nodo Llorente no lo hace, por lo que el nodo Llorente se fusiona con el nodo Barrantes en una sola clase de equivalencia. Dado que el nodo de San Pablo requiere mayor privacidad, este debe cumplir con **k privado**, se puede notar que este cumple con ello. El nodo de San Roque también debe cumplir con **k privado**, pero no tiene suficientes elementos, por lo que tiene que fusionarse con otros nodos, en este caso, el nodo de San Pablo. Finalmente, se termina con 5 clases de equivalencia, 4 con valor **k base** ([Nicoya], [Nacascolo], [Mayorga], [Barrantes, Llorente] = [Flores]) y 1 con el **k privado** ([San Pablo, San Roque] = [Barva]).

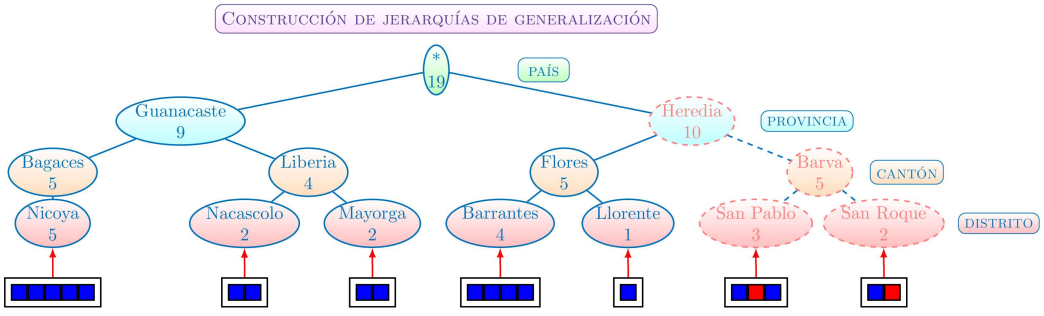



Figura 5.1: Árbol de generalización de datos que permite construir los grupos que cumplen con las restricciones de privacidad.

## 5.3. Resumen

En este capítulo 5 se definió una metodología para instrumentar procesos de desidentificación para que considere las preferencias de los dueños de los datos sobre posibles inferencias que se podrían realizar con los mismos. Asimismo, se instrumentó una implemententación de método de desidentificación de $k$-anonimato, para que considerara preferencias individuales sobre inferencias de datos.

El objetivo de un buen algoritmo de preservación de la privacidad es encontrar una transformación de los datos originales, de modo que satisfaga el requisito de privacidad al tiempo que minimiza la pérdida de información y maximiza la utilidad de los datos resultantes [Sun et al., 2012]. Por lo tanto, es necesaria una medida para indicar la calidad y preservación de la privacidad de los datos. En el capítulo 6 se presenta la definición de un marco de medición de la privacidad que permite evaluar el nivel de preservación de la privacidad en función de las preferencias de privacidad de cada dueño de los datos.

# Capítulo 6

# Métrica de bloqueo de inferencia de datos

Este capítulo presenta una métrica de privacidad basada en algoritmos de aprendizaje de máquina que permite identificar la eficacia del bloqueo de inferencias sobre una base de datos relacional. Primero se describe la métrica, luego la selección del algoritmo de aprendizaje, y su aplicación. Finalmente, se presenta un pequeño caso que utiliza y compara distintos algoritmos, para evaluar su aplicabilidad.

## 6.1. Métrica de privacidad propuesta

La preservación de privacidad en un conjunto de datos desidentificado usualmente se juzga con dos criterios contradictorios: (a) la privacidad que le da a cada individuo representado en la base de datos y (b) la utilidad dada a quien ocupe usar los datos publicados [Ohm, 2010]. Estos criterios son contradictorios debido a que conforme la privacidad se incrementa, la utilidad de los datos decrece. Inversamente, para obtener mayor utilidad de los datos, la granularidad (contraria a generalizar) debe incrementarse. Dentro de estas limitaciones, la práctica usual es establecer un umbral en uno de los factores (dependiendo del sesgo de la tarea), y luego maximizar el otro.

Una métrica de preservación de la privacidad depende de los objetivos de privacidad requeridos, así que las métricas pueden cambiar para reflejar objetivos específicos. Estas métricas usualmente no son ni siquiera comparables [Wagner and Eckhoff, 2018]. Dado que el objetivo de privacidad en este trabajo es que los registros sean “menos inferibles” para los dueños de datos con requerimientos de privacidad más estrictos, la medida de privacidad será la probabilidad de reidentificación luego de aplicar un algoritmo de aprendizaje máquina, la cual no es una métrica estándar, pero es una métrica razonable para el objetivo de privacidad perseguido [Agrawal and Srikant, 2000].

En este trabajo se propone recolectar preferencias individuales de bloqueo a inferencias. Dichas preferencias son tomadas en cuenta al modificar y aplicar a un método de desidentificación, tal como se describe en el capítulo 5. Lo mínimo que se debe lograr es que las personas que expresaron el deseo de que ciertas inferencias se bloquearan, obtengan algún efecto diferenciado en los datos a publicar. En otras palabras, en promedio, deber ser más difícil realizar correctamente la inferencia sobre tuplas de sujetos con mayor preferencia de privacidad.

La métrica de privacidad que se propone aquí, responde justamente a la afirmación del párrafo anterior. Se pretende medir la probabilidad de reidentificación de los usuarios cuando se realiza la inferencia que se debía bloquear. Se utiliza el término "reidentificación" para designar el efecto de predecir correctamente el dato que originalmente se manipuló en el proceso de desidentificación, y no al hecho de encontrar los identificadores personales del dueño de los datos.

El enfoque que se propone toma el punto de vista de un hipotético atacante. Se supone que el atacante tiene acceso a los datos publicados, a un conjunto de datos adicionales y a algún algoritmo de aprendizaje de máquina. Usando estos componentes genera una predicción con respecto a los dueños de los datos representados en el conjunto. El publicador de datos en su rol de defensor de las preferencias de privacidad puede verificar que tan buena es esta predicción. La métrica propuesta es precisamente un reflejo del grado de aproximación de la predicción a la verdad.

Como el dueño de la base de datos no tiene forma de saber con qué información adicional cuenta el atacante, se ha construido una métrica que puede calcularse con los recursos disponibles para el publicador de datos.

Si se cuenta con un conjunto de datos a publicar, la métrica se define como la probabilidad de que la predicción (la reidentificación) haya sido correcta y se calcula como el número de predicciones correctas dividido por el número total de filas a predecir. La predicción busca encontrar algún campo o campos originalmente modificados en el proceso de desidentificación que luego pueden permitir completar la inferencia objeto.

En términos de métricas de utilidad se explica únicamente para efecto de establecer umbrales de funcionamiento. La literatura usualmente define las métricas de utilidad en forma negativa, como *pérdida de información*. [Bayardo and Agrawal, 2005]. Estas métricas asignan algún valor a cada cambio que se realice en el conjunto de datos en la búsqueda de la preservación de la privacidad (la "penalidad", dado que refleja la

pérdida de datos causada por el proceso), y luego construyen algún tipo de promedio.

Para este trabajo se adoptó la formulación normalizada de la versión agregada de NCP, llamada Penalización de Certeza Global (GCP, *Global Certainty Penalty* en inglés) propuesta por Ghinita [Ghinita et al., 2009] y la Métrica de Clasificación (CM, *Classification Metric* en inglés) propuesta por Iyengar [Iyengar, 2002]. Ambas explicadas en el capítulo 2.

### 6.1.1. Selección del algoritmo de aprendizaje y su aplicación

Existen muchos métodos y algoritmos de aprendizaje de máquina. Se describe la aplicación de árboles de decisión, bosques aleatorios y regresión logística, para ilustrar su uso.

Dependiendo del algoritmo de aprendizaje máquina el subconjunto de datos debe preprocesarse (modificar los datos para ser utilizables por el algoritmo) de diferentes formas. El proceso completo para aplicar el algoritmo de aprendizaje máquina se describe a continuación:

1. El conjunto de datos adicional se preprocesa y recodifica.
2. Los atributos se categorizan como de predicción y objetivo.
3. Los datos se dividen en conjuntos de entrenamiento y prueba (o en conjuntos de validación cruzada), $T_{entrenamiento}$ y $T_{prueba}$.
4. Con el conjunto de entrenamiento se genera un modelo de predicción mediante el algoritmo de aprendizaje seleccionado
5. Se realiza la evaluación sobre los datos a publicar.

### 6.1.2. Interpretación de la predicción

En el modelo abstracto se obtiene una predicción, esta predicción se realiza calculando cuáles son los casos donde el algoritmo de aprendizaje máquina realiza la predicción correctamente. Si esa predicción se realiza correctamente quiere decir que el algoritmo logra reidentificar o recuperar los datos a su estado anterior y por lo tanto se produce una vulneración de la protección preestablecida. Es importante indicar que se califica

el nivel de inferencia como el nivel de exactitud que ofrezca el algoritmo al ejecutar la predicción.

### 6.1.3. Ejemplo de aplicación

Para probar el uso de la probabilidad de reidentificación usando algoritmos de aprendizaje de máquina se usaron las implementaciones de los siguientes algoritmos de aprendizaje proporcionados por el paquete de software scikit-learn [Pedregosa et al., 2011]:

- **Árbol de decisión**: Se usaron las configuraciones predeterminadas.
- **Bosques aleatorios**: Cada clasificador estaba compuesto por 10 árboles aleatorios, y se utilizaron todas las demás configuraciones predeterminadas.
- **Regresión logística**: Se usaron las configuraciones predeterminadas.

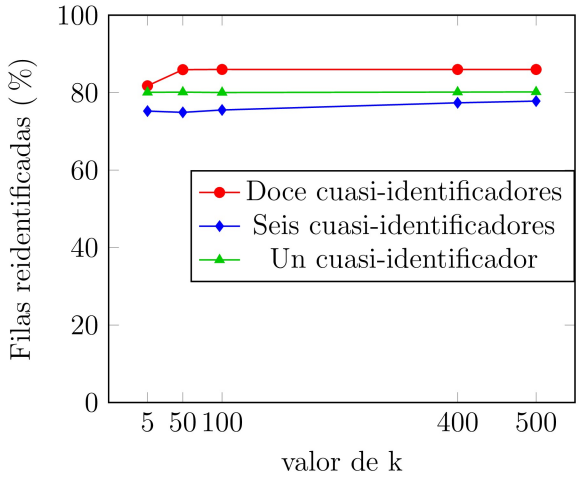
Se utilizó el conjunto AdultDatabase de UC Irvine [Dua and Graff, 2017] dado que es el estándar en uso para comparar métricas y algoritmos de preservación de la privacidad [Iyengar, 2002, Bayardo and Agrawal, 2005, Machanavajjhala et al., 2006, Ye et al., 2008, Liu et al., 2015]. Se tenía que utilizar ese conjunto de datos para poder realizar las correspondientes comparaciones con los resultados previamente publicados y así minimizar inconsistencias producto de la implementación. Esto para poder efectuar una comparación que tenga sentido y poder evaluar la métrica.

El conjunto de datos original contiene 32,561 filas con 15 atributos, una vez finalizado el preprocesado quedaron 30,162 filas. Se definió como atributo sensible (no cuasi-identificador): “race”, como atributo de interés (no cuasi-identificador): “income” y se define como atributo no sensible: “education” (para asegurar un grado de utilidad constante, esto aunque es equivalente a el atributo “education-num”). Los demás atributos se utilizan como posibles cuasi-identificadores. En nuestros experimentos, variamos el tamaño del conjunto de atributos cuasi-identificadores en un rango de 1, 6 y 12. Cuando se modifica el número de atributos cuasi-identificadores, el atributo que deja de ser cuasi-identificador automáticamente pasa a ser no sensible para mantener un mismo número de entradas en los modelos de aprendizaje máquina. Las jerarquías de generalización para los atributos categóricos están disponibles en desarrollos de código abierto GitHub (https://github.com/lgesquivel/Basic_Mondrian/). También, se definieron los niveles de privacidad k en rango de 5, 50, 100, 400 y 500.

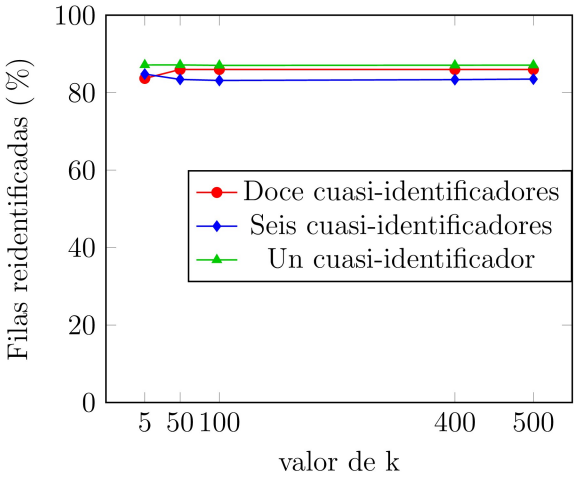
### 6.1.4. Análisis de resultados

El primer paso de la experimentación incluye evaluar la preservación de la privacidad de los datos. Con este propósito se construyen modelos de predicción utilizando los algoritmos de aprendizaje máquina seleccionados y se realiza la medición de la exactitud en la predicción del atributo sensible "race".

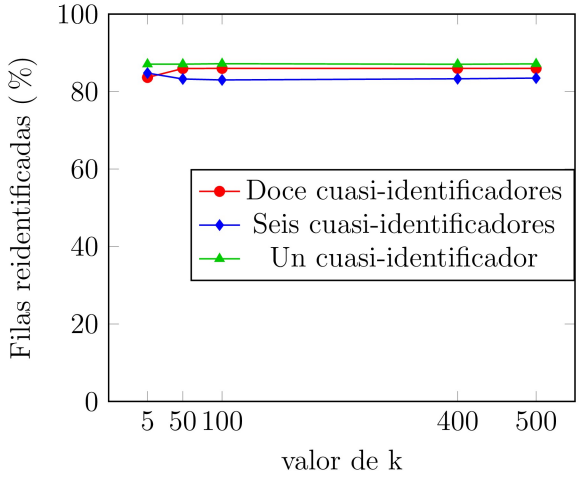
Como se observa en la figura 6.1, los modelos creados con todos los algoritmos mantienen una precisión superior al 70 %. Podemos observar que en general la precisión de la predicción sufre variaciones conforme se cambia el nivel de privacidad en el algoritmo cuando se tienen doce cuasi-identificadores. Esto se debe a que las clases de equivalencia son más grandes y ayudan a la generación de reglas más simples. Esto concluye en un sobreajuste (overfitting) producido por el alto porcentaje de tuplas con un valor específico para la variable sensible. Esto implica que los algoritmos realizan la inferencia deseada en la mayoría de los casos.




*Algoritmo de árboles de decisión.*



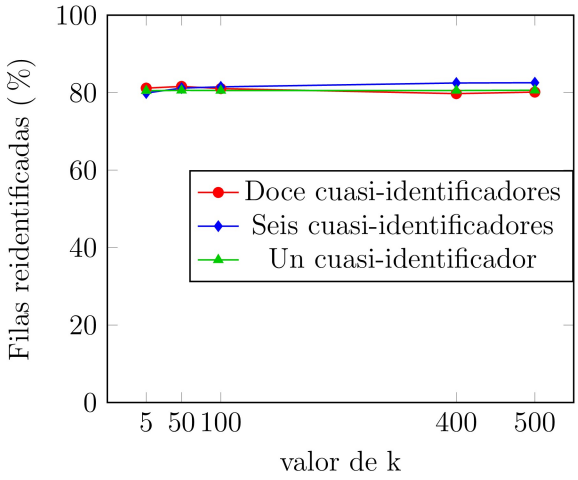

*Algoritmo de regresión logística.*



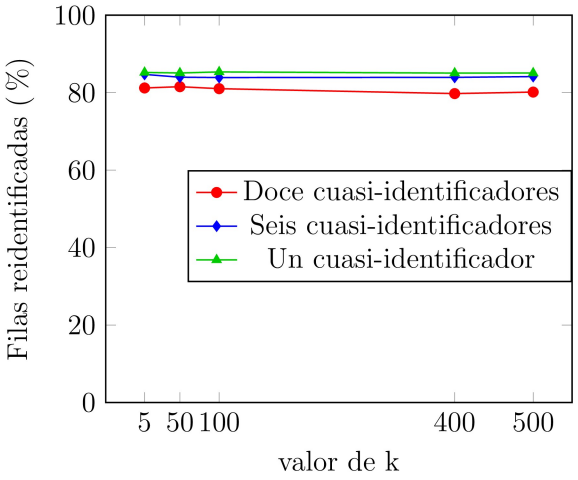

*Algoritmo de bosques aleatorios.*

Figura 6.1: Evaluación de la precisión en la predicción de la variable “race”.

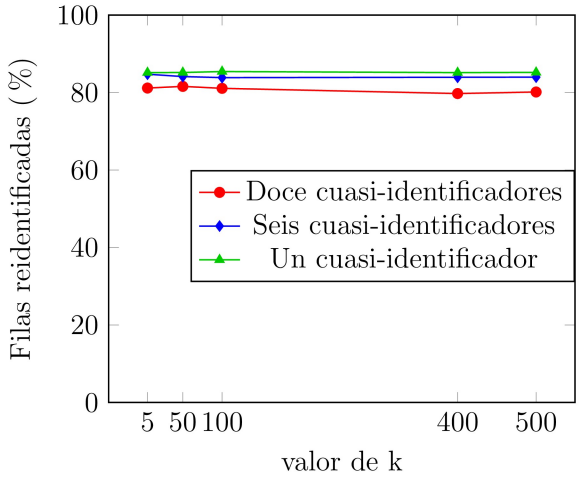
Al analizar los modelos de predicción creados para la variable de interés “income”, cuyos resultados se presentan en la figura 6.2, se observa que la precisión de la predicción no varía. Los resultados en la precisión de la predicción no varían en función del nivel de privacidad y la diferencia al cambiar el número de cuasi-identificadores no es significativa. También se observa que dicho comportamiento es similar para la variable “race” cuando solo se tiene un atributo cuasi-identificador.




*Algoritmo de árboles de decisión.*




*Algoritmo de regresión logística.*




*Algoritmo de bosques aleatorios.*

Figura 6.2: Evaluación de la precisión en la predicción de la variable “income”.

Los resultados de los modelos de predicción creados indican que la precisión de la predicción no varía una vez aplicado el primer nivel de preservación de la privacidad. Destaca que ni el número de cuasi-identificadores utilizado, ni el nivel de preservación de la privacidad influyen en el resultado de la predicción para el dato que se indicó como sensible.

Existen varios aspectos a destacar de los resultados, el primero es que la preservación de la privacidad no cambia significativamente al variar el nivel de privacidad k o el número de cuasi-identificadores en el algoritmo de k-anonimato utilizado en la

preservación de la privacidad. Los resultados son consecuentes dado que el algoritmo implementado no toma en cuenta el atributo sensible para el proceso de preservación de la privacidad. Al tomar como métrica de privacidad los resultados obtenidos por los algoritmos de aprendizaje máquina, podemos indicar que la precisión en la predicción sobre el atributo sensible es muy alta. Dados los resultados obtenidos se puede indicar que existe una preservación de la privacidad muy pobre.

En particular, si evaluamos el impacto de la pérdida de información reportada por una métrica como GCP en la figura 6.3 observamos que está reporta pérdidas de información de hasta 60 % y esto no es consecuente con los resultados en las predicciones de los algoritmos de aprendizaje máquina.

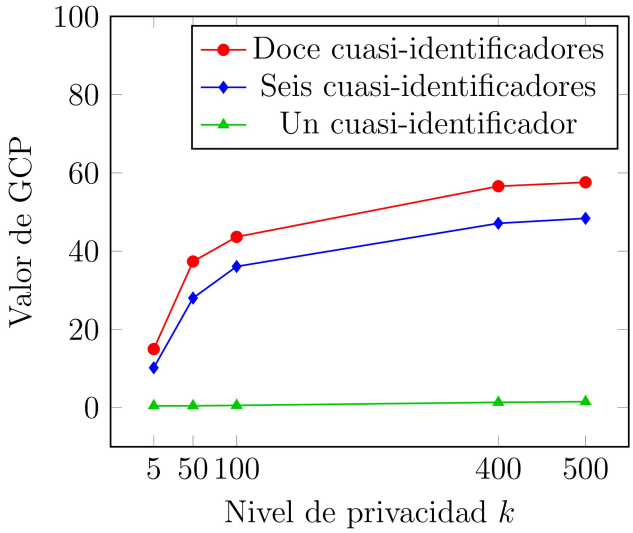
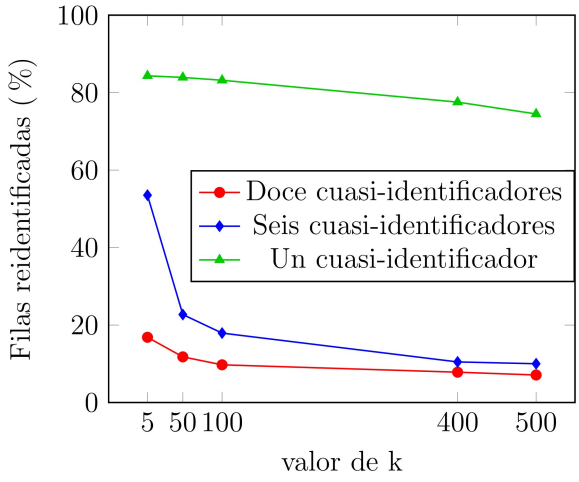



Figura 6.3: Relación entre el nivel de privacidad k, el número de cuasi-identificadores y el valor obtenido por la métrica GCP.

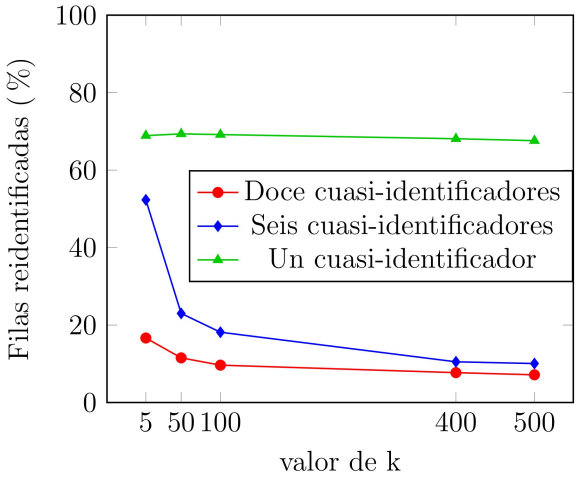
De hecho se reporta un incremento en la pérdida de información al modificar el nivel de privacidad k, situación que no se ve reflejada en la precisión de la predicción de los algoritmos de aprendizaje máquina. Resulta probable que la reducción de los dominios de los atributos influya en la conservación de la precisión en las predicciones, así como la falta de verificación de la diversidad de valores en el atributo sensible por parte tanto de la métrica, como del algoritmo implementado para la preservación de la privacidad. Como el fin último de una garantía científica sobre la preservación de la

privacidad es medir el peligro de reidentificación del atributo sensible o el dueño de los datos, está claro que los algoritmos de aprendizaje máquina logran este propósito.

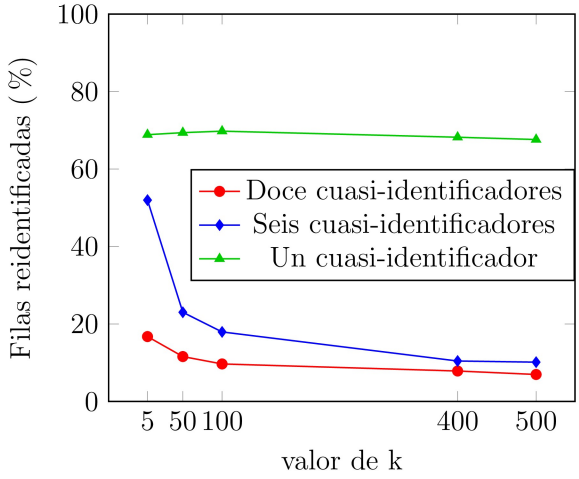
Saliendo del ámbito de la replicación de experimentos similares para experimentar con la métrica propuesta, se pasa a elegir, ahora sí, un cuasi-identificador en el que la aplicación de $k$-anonimato genere grupos razonables que no sean susceptibles al *overfitting* por parte de los algoritmos de aprendizaje-máquina. El comportamiento de la precisión en la predicción varía si se analizan los modelos de predicción creados para la variable de "age", la cual siempre es tomada como atributo cuasi-identificador en el algoritmo de preservación de la privacidad y cuyos resultados se presentan en la figura 6.4. Se puede observar que, tal como se espera, la precisión de la predicción disminuye conforme aumenta el $k$, y por lo tanto se produce una mejora en la preservación de la privacidad para la variable "age".




*Algoritmo de árboles de decisión.*




*Algoritmo de regresión logística.*




*Algoritmo de bosques aleatorios.*

Figura 6.4: Evaluación de la precisión en la predicción de la variable “age”.

Los comportamientos no monotónicos por parte de las métricas de privacidad están documentados en la literatura [Bertino et al., 2008, Wagner and Eckhoff, 2018] y por lo tanto es esperable un comportamiento similar de los algoritmos de aprendizaje máquina al ser utilizados como métricas de privacidad.

El uso de algoritmos de aprendizaje máquina para asegurar al dueño de los datos qué se puede realizar con su información es algo primordial y esto hace que su uso sea base para un aseguramiento de la privacidad más riguroso. Estos resultados demuestran que utilizar algoritmos de aprendizaje máquina son una medida simple de calidad de

la preservación de la privacidad. De forma similar se plantea utilizar algoritmos de aprendizaje máquina para verificar que se cumplan los deseos de un dueño de datos sobre un conjunto de datos a publicar. Donde estos deseos sean autorizar o no un análisis especifico.

## 6.2. Resumen

En este capítulo se presentó una métrica que permite mostrar la aplicabilidad del modelo planteado en los capítulos 4 y 5.

Más concretamente:

- Se definió una métrica que mide la probabilidad de reidentificación utilizando aprendizaje máquina.
- Se evaluó el uso de la métrica utilizando los siguientes algoritmos de aprendizaje máquina:
  - Árboles de decisión
  - Bosques aleatorios
  - Regresión lineal

La evaluación de la métrica fue un aporte adicional al trabajo de tesis, ya que esto no estaba contemplado entre los objetivos específicos. De la evaluación de la métrica se descubrió:

- Evidencia de la efectividad de la métrica propuesta
- Evidencia sobre la poca efectividad de las métricas de privacidad descritas en la literatura
- La existencia de serias dificultades para establecer buenos parámetros cuando se evalúan y comparan diferentes algoritmos de desidentificación

La métrica descrita cumple el objetivo específico 3: *Definir una métrica que identifique la eficacia de un bloqueo de inferencia de datos para un contexto particular.*

El uso de la métrica propuesta requiere su aplicación por medio de con un caso de estudio concreto. En el capítulo 7 se presenta un caso de estudio donde se describe el

proceso seguido para asignar las preferencias de bloqueo, y la descripción del conjunto de datos a utilizar en la evaluación que posteriormente se realizó por medio de la métrica propuesta en este capítulo.

# Capítulo 7

# Caso de estudio

En este capítulo se desarrolla un caso de estudio que incorpora un conjunto de inferencias especificadas con el modelo definido en el capítulo 4 y las aplica con el modelo de desidentificación discutido en el capítulo 5. El capítulo está organizado de la siguiente manera:

1. En la sección 7.1 se describe el contexto de base datos utilizada en el caso de estudio.

2. Un análisis de las consultas típicas realizadas a la base de datos se presenta en la sección 7.2.

3. La extracción de posibles inferencias se presenta en 7.3.

## 7.1. Definición del contexto de estudio

La primera labor que se realizó en el desarrollo del caso de estudio fue encontrar una organización donde existiera interés en la protección de la privacidad y donde un subconjunto de los dueños de los datos pudiera ser relativamente fácil de consultar en caso de necesidad. También se tomó en cuenta la posibilidad de contar con algún subconjunto de solicitudes realizadas por los consumidores de los datos. Para este efecto se logró conversar con la Vicerrectoría de Vida Estudiantil de la Universidad de Costa Rica (VVE), Vicerrectoría encargada de salvaguardar los datos de los estudiantes de la Universidad a través de la Oficina de Registro e Información (ORI). La VVE autorizó el acceso a la base de datos de estudiantes utilizada para desarrollo de sistemas (copia de la original), el uso de esta base de datos fue aprobado por la señora Vicerrectora de Vida Estudiantil, la Mag. Ruth de la Asunción Romero. La base de datos facilitada es una copia de la base de datos original y los datos que contiene conservan gran similitud con la información de producción. Además, se facilitó un conjunto de consultas de datos

realizadas a la ORI a lo largo de tres años, así como la asesoría de los encargados de dicha oficina. En el cuadro 7.1 se presenta una caracterización a grandes rasgos de la base de datos utilizada. Esto representa la información que corresponde a los dueños de los datos (los estudiantes) de la organización universitaria.

Cuadro 7.1: Caracterización de base de datos de la UCR

| Características | Valor |
|---|---|
| Número de tablas | 459 |
| Número total de atributos | 4,051 |
| Número total de filas | 84,859,746 |
| Estudiantes registrados en la base de datos | 315.089 |
| Años representados en la base de datos | 2008-2016 |

## 7.2. Consultas típicas

En cuanto a la caracterización de posibles consultas de consumidores de datos, la ORI facilitó tres años de solicitudes recibidas por esa oficina tanto vía correo electrónico, como vía oficio. Todas las solicitudes fueron digitalizadas y mediante técnicas de recuperación de información e intervención humana se extrajo la información relevante. La información relevante incluye una descripción de la solicitud en lenguaje natural y la consulta SQL que produjo el conjunto de datos respuesta. Como se puede observar en la figura 7.1 el análisis de la información recuperada permitió identificar nombres de tablas y atributos utilizados en las consultas SQL. Además permitió identificar frecuencias de uso que pueden ser de interés en trabajos futuros.




Figura 7.1: Frecuencia de palabras en sentencia SQL

La digitalización y análisis de la información recopilada permitió además determinar las características generales de las consultas SQL de la ORI. Este análisis se encuentra resumido en la tabla 7.2.

Cuadro 7.2: Caracterización de consultas SQL utilizadas por la ORI para producir los conjuntos de datos respuesta

| Características | Valor |
|---|---|
| Número de consultas SQL | 642 |
| Número total de tablas consultadas | 59 |
| Número total de atributos consultados | 220 |
| Años representados en las consultas SQL | 2013-2016 |

## 7.3. Selección de inferencias posibles en la base de datos y su representación en lenguaje natural

El primer paso de la metodología presentada en el modelo de expresión de inferencias es el proceso para extraer inferencias (sección 4.1) y requiere conocimiento sobre las preocupaciones de privacidad de los dueños de los datos. Este conocimiento se puede adquirir por medio de entrevistas o encuestas a los dueños de los datos, como se indicó en la sección indicada.

Las entrevistas o encuestas deben utilizar una taxonomía de privacidad, como la planteada por Song *et al* [Song et al., 2018]. Dada la finalidad genérica del modelo de expresión de inferencias, se creó una taxonomía propia utilizando el conocimiento de los publicadores de datos. Para esto se utilizó el conjunto de consultas facilitado por la ORI, se trabajó en un análisis de las consultas SQL realizadas a la base de datos por parte del publicador de datos y se extrajeron los elementos utilizados para generar la respuesta a cada una de dichas consultas. Esta información permitió crear la taxonomía que se muestra en el cuadro 7.3.

Cuadro 7.3: Taxonomía de datos de estudiantes

| Clasificación |
|---|
| Información de horarios |
| Información de educación secundaria |
| Información de admisión |
| Información geográfica |
| Información de carreras |
| Información socioeconómica |
| Información de calificaciones |
| Información títulos |
| Información financiera |
| Información personal |

Esta lista de elementos se utilizó como insumo para la creación de una encuesta, la cual se aplicó a un conjunto de 54 de dueños de datos (estudiantes). Estos 54 dueños de datos son formados por dos grupos de estudiantes, uno de 31 estudiantes del año 2016 (Grupo A) y otro de 23 estudiantes de año 2017 (Grupo B). El instrumento utilizado como encuesta se encuentra en el apéndice A. Los estudiantes clasificaron diferentes datos mediante una escala *Likert* de cinco puntos, donde el valor menor correspondía a “Nada privada” (valor de 0) y el valor mayor a “Muy privada” (valor de 4). Como se muestra en la figura 7.2 la clasificación de privacidad que asignaron los estudiantes se puede considerar equilibrada, dado que las respuestas no varían de forma sorpresiva.

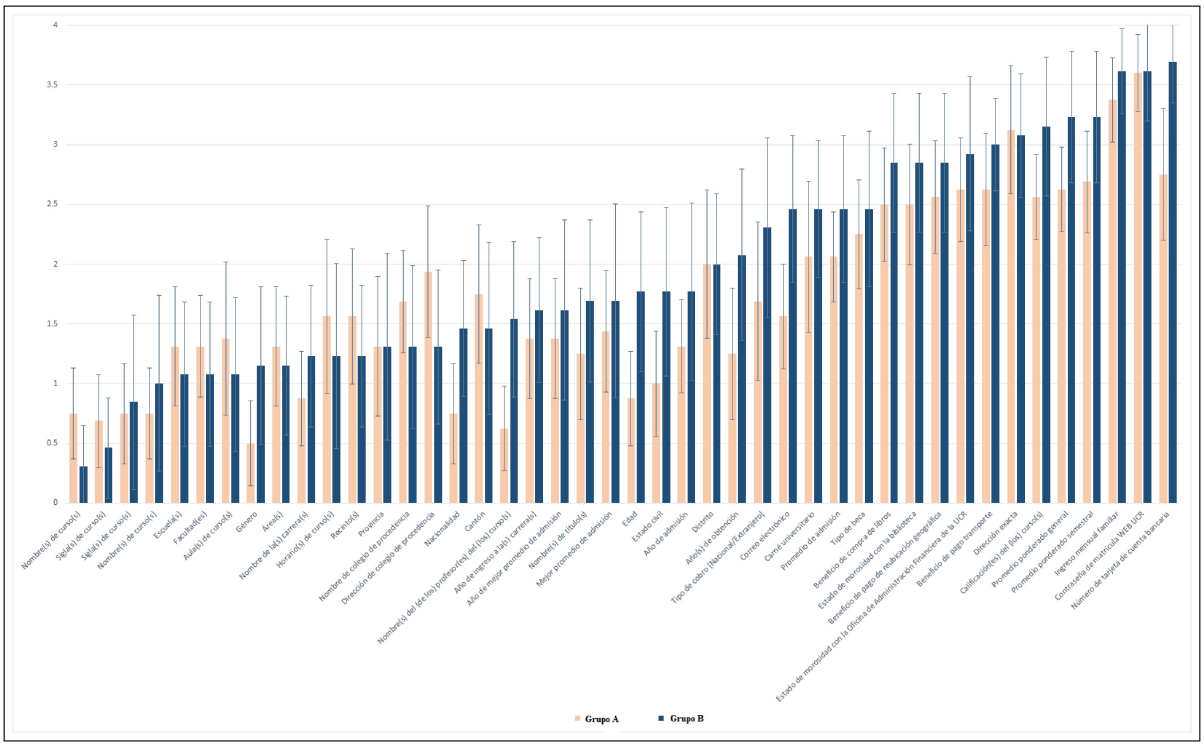



Figura 7.2: Respuestas a encuesta

Los resultados de la encuesta no intentan ser representativos de toda la población de la base de datos, pero muestra que es factible recopilar información respecto a las preferencias de privacidad de los dueños de los datos utilizando una encuesta. Una vez tabulados los resultados y mediante conocimiento experto de la base de datos se definieron las siguientes inferencias:

1. Deducir el horario de clases de un estudiante.
2. Deducir el lugar de residencia de un estudiante.
3. Deducir la calificación obtenida en el examen de admisión.
4. Deducir la ruta de autobús para llegar a la UCR de un estudiante.
5. Deducir el colegio de procedencia de un estudiante.

Una vez que se identificó un grupo de inferencias interesantes para los estudiantes y realizables sobre la base de datos, se escogió la inferencia “Deducir el lugar de residencia

de un estudiante". Esto se debe a que gran parte de las inferencias seleccionadas están relacionadas directa o indirectamente con el lugar de residencia. Por ejemplo, "la ruta de autobús" requiere identificar en qué lugar se tomará el autobús o "deducir el colegio de procedencia" dónde es muy probable que exista una cercanía geográfica entre el colegio y la residencia del estudiante.

Una vez escogidas las inferencias, el segundo paso de la metodología es la traducción de dichas inferencias a varias consultas SQL en la base de datos objetivo. El objetivo de este paso es extraer de las consultas SQL los elementos de la base de datos que se usarán para formar las reglas. La ejecución de este paso para el caso de estudio se describe en la siguiente sección.

## 7.4. Traducción de inferencias a reglas y elementos interpretables en la base de datos estudiantil

La traducción de inferencias a reglas y elementos interpretables en la base de datos estudiantil requiere la extracción de conocimiento. Esto implica encontrar varias formas de recuperar los datos requeridos por la inferencia. Para definir estas formas se utilizó el conocimiento experto sobre el sistema de bases de datos de estudiantes de la UCR de la Lic. Lilliana Rojas Madrigal, del Área de Desarrollo de Sistemas de Información del Centro de Informática.

Dada la inferencia "Deducir el lugar de residencia de un estudiante", se presentan tres consultas SQL que permiten recuperar los datos requeridos. Este número de consultas no pretende ser exhaustivo, solo busca ejemplificar el uso del modelo. En la figura 7.3 se muestran estas consultas SQL.

```
1 SELECT INS.IDENTIFICACION_ESTUDIANTE,
2 (SELECT DESCRIPCION FROM PROVINCIA WHERE PROVINCIA = INS.PROVINCIA) AS
      PROVINCIA,
3 (SELECT DESCRIPCION FROM CANTON WHERE PROVINCIA = INS.PROVINCIA AND
      CANTON = INS.CANTON) AS CANTON,
4 (SELECT DESCRIPCION FROM DISTRITO WHERE PROVINCIA = INS.PROVINCIA AND
      CANTON = INS.CANTON AND INS.DISTRITO = DISTRITO) AS DISTRITO
5 FROM ADM_INSCRIPCION INS
6
7 //
8
9 SELECT E.CARNE, INF.PROVINCIA, INF.CANTON, INF.DISTRITO
10 FROM ESTUDIANTE E, INFORMACION_PERSONAL INF
11 WHERE INF.SECUENCIA_VIGENCIA = (SELECT MIN (INF.SECUENCIA_VIGENCIA)
12                                 FROM INFORMACION_PERSONAL INF
13                                 WHERE INF.CARNE = E.CARNE)
14       AND E.CARNE = INF.CARNE
15
16 //
17
18 SELECT INS.IDENTIFICACION_ESTUDIANTE,(SELECT NOMBRE_COLEGIO FROM
      COLEGIO WHERE COLEGIO = INS.COLEGIO) AS COLEGIO,
19 (SELECT DESCRIPCION FROM PROVINCIA WHERE PROVINCIA = INS.PROVINCIA) AS
      PROVINCIA,
20 (SELECT DESCRIPCION FROM CANTON WHERE PROVINCIA = INS.PROVINCIA AND
      CANTON = INS.CANTON) AS CANTON,
21 (SELECT DESCRIPCION FROM DISTRITO WHERE PROVINCIA = INS.PROVINCIA AND
      CANTON = INS.CANTON AND INS.DISTRITO = DISTRITO) AS DISTRITO
22 FROM ADM_INSCRIPCION INS
```

Figura 7.3: Consultas SQL que permiten ejecutar la inferencia “Deducir el lugar de residencia de un estudiante”.

El siguiente paso en el caso de estudio es obtener la estructura XML que contiene las reglas que identifican los elementos de la base de datos relacionados a la inferencia.

```
<INFERENCE_DEDUCIR_RESIDENCIA>
  <RULE behavior="active">
    <DESCRIPTION>
     Deducir el lugar de residencia de un estudiante
    <DESCRIPTION>
   <TABLE name = INFORMACION_PERSONAL>
        <FIELD>PROVINCIA</FIELD>
        <FIELD>CANTON</FIELD>
        <FIELD>DISTRITO</FIELD>
        <FIELD>DIRECCION</FIELD>
   </TABLE>
   <TABLE name = ADM_INSCRIPCION>
        <FIELD>PROVINCIA</FIELD>
        <FIELD>CANTON</FIELD>
        <FIELD>DISTRITO</FIELD>
        <FIELD>COLEGIO</FIELD>
   </TABLE>
  </RULE>
</INFERENCE_DEDUCIR_RESIDENCIA>
```

Figura 7.4: Reglas para la inferencia: “Deducir el lugar de residencia de un estudiante”

## 7.5. Consulta a publicar y conjunto de datos a desidentificar

Como se indicó anteriormente, se definió la inferencia como “Deducir el lugar de residencia de un estudiante”. Para continuar con la aplicación de la metodología propuesta, ante una consulta correspondiente a una solicitud de publicación, se deben cruzar las reglas de base de datos correspondientes a la inferencia con el SQL de la solicitud de publicación.

En el caso de estudio se escogió la consulta de notas del Examen de Admisión de la UCR para el año 2015, que se puede observar en la figura 7.5. Esta es una consulta relativamente común en la base de datos de estudiantes según el análisis de consultas realizado en la sección 7.2. Esta consulta es realizada en algunos casos por los medios de comunicación y tiene la característica que facilita la realización de pruebas sobre los

mecanismos de bloqueo de inferencias.

```
1 SELECT INS.IDENTIFICACION_ESTUDIANTE,(SELECT NOMBRE_COLEGIO FROM
      COLEGIO WHERE COLEGIO = INS.COLEGIO) AS COLEGIO,
2 (SELECT DESCRIPCION FROM PROVINCIA WHERE PROVINCIA = INS.PROVINCIA) AS
       PROVINCIA,
3 (SELECT DESCRIPCION FROM CANTON WHERE PROVINCIA = INS.PROVINCIA AND
      CANTON = INS.CANTON) AS CANTON,
4 (SELECT DESCRIPCION FROM DISTRITO WHERE PROVINCIA = INS.PROVINCIA AND
      CANTON = INS.CANTON AND INS.DISTRITO = DISTRITO) AS DISTRITO,
5 TRUNC(MONTHS_BETWEEN(SYSDATE,INS.FECHA_NACIMIENTO)/12) AS EDAD,
6 CASE
7                 WHEN (INS.ADECUACION_EN_SECUNDARIA = 1) THEN 'SI'
8                 WHEN (INS.ADECUACION_EN_SECUNDARIA = 0) THEN 'NO'
9                 ELSE 'DESCONOCIDO'
10              END AS ADECUACION_EN_SECUNDARIA,
11  CASE
12                WHEN (INS.SEXO = 2) THEN 'F'
13                WHEN (INS.SEXO = 1) THEN 'M'
14                ELSE 'DESCONOCIDO'
15              END AS SEXO, (SELECT GENTILICIO FROM NACION WHERE
                    CODIGO_NACION = INS.CODIGO_NACION) AS NACIONALIDAD,
16              INS.PUNTAJE_500 AS PROMEDIO_ADMISION
17 FROM ADM_INSCRIPCION INS
18 WHERE INS.ANNO = '2015'
```

Figura 7.5: Consulta SQL que permiten ejecutar sobre los datos de admisión.

El conjunto fue filtrado para eliminar inconsistencias como valores nulos o repetidos. El conjunto final utilizado para la evaluación en el capítulo 8 contiene 42,538 filas. En el experimento se usaron únicamente las seis columnas resumidas en la tabla 7.4. La reidentificación en este caso se definió como la correcta identificación, por parte de un atacante, de la tripleta (desidentificada) de (Provincia, Cantón, Distrito) de un estudiante.

Cuadro 7.4: Columnas del conjunto de datos a desidentificar y su descripción.

| Columna | Descripción | No. de elementos únicos |
|---|---|---|
| Sexo | Sexo | 2 |
| PromedioAdmisión | Calificación en prueba de admisión de la UCR | 531 |
| Provincia | Primer nivel división administrativa CR | 7 |
| Cantón | Segundo nivel división administrativa CR | 81 |
| Distrito | Tercer nivel división administrativa CR | 467 |
| Colegio | Nombre de centro de educación secundaria en CR | 980 |

Las columnas Provincia, Cantón y Distrito, que juntas describen el lugar de origen del estudiante, fueron seleccionadas como cuasi-identificadores. La columna "PromedioAdmisión" es la nota del examen y el dato sensitivo. Las otras columnas representan datos crudos.

## 7.6. Simulación de las preferencias individuales sobre el bloqueo de inferencia

Como aproximación a las preferencias de bloqueo de inferencias de los estudiantes se usó la clasificación de Taylor sobre preocupaciones de privacidad en la población general [Taylor, 2003]. Taylor concluyó que el 26 % de las personas son "fundamentalistas de la privacidad", 10 % eran "despreocupados de la privacidad" (*privacy unconcerned*), y el 64 % eran "pragmáticos de la privacidad". Se supuso que únicamente los fundamentalistas de la privacidad iban a desear bloquear las inferencias, y con base en esto, se seleccionaron aleatoriamente el 26 % de las entradas del conjunto de datos, a las que se les asignó la elección de "bloquear la inferencia".

Se destaca que la asignación de tuplas a ese 26 % de bloqueo de inferencias se realizó una única vez. Por lo tanto, cuando se determinaron los subconjuntos para los respectivos experimentos, el porcentaje de fundamentalistas entre subconjuntos podía variar levemente.

## 7.7. Resumen

En este capítulo se siguieron los pasos definidos en las secciones 4.1 y 4.2 del capítulo 4 con un caso de estudio concreto, para mostrar la aplicabilidad del modelo. Se describió asimismo el proceso seguido para asignar las preferencias de bloqueo, y la descripción del conjunto de datos a utilizar en la evaluación. Más específicamente:

- Se seleccionaron inferencias aplicables a la base de datos
- Se generaron las reglas que permiten identificar los elementos que componen una inferencia
- Se simuló la captura de preferencias de los dueños de los datos sobre el bloqueo de inferencias de datos

Con esto, se cumple el objetivo específico 2: *Desarrollar un caso de estudio que incorpore un conjunto de inferencias especificadas con el modelo de expresión definido, en un modelo de desidentificación conocido.*

En el capítulo 8, a continuación, se presenta la evaluación del caso de estudio, utilizando una instanciación del bloqueo de inferencia de datos presentado en el capítulo 5.

# Capítulo 8

# Evaluación del modelo de expresión de inferencias

En el capítulo 7 se propuso una inferencia y una solicitud de publicación específica. Para la inferencia se indicó su representación como reglas de la base de datos. Para la solicitud de publicación, se especificó la consulta SQL necesaria para obtener los datos requeridos, y se ilustró el proceso de mapeo entre las reglas de base de datos asociadas a la inferencia y la consulta SQL de publicación. Finalmente, se obtuvo el conjunto de datos totales que correspondían a la consulta, para preparar el proceso de evaluación de las acciones de desidentificación que se le deben a realizar a dichos datos. En este capítulo se parte de ese punto para describir el experimento de evaluación realizado sobre estos datos, utilizando el método de desidentificación propuesto en el capítulo 5.

El capítulo está organizado de la siguiente manera: Se inicia con la explicación de la metodología seguida para realizar la evaluación (sección 8.1), seguido de la presentación de los resultados y su análisis (sección 8.2), y una pequeña sección de resumen (sección 8.3).

## 8.1. Metodología de evaluación

El propósito principal de este capítulo es evaluar, sobre los datos del caso de estudio, si la instrumentación que se aplique al algoritmo de desidentificación se refleja como una mejora en la privacidad de los dueños de datos involucrados. Esto corresponde al objetivo específico *"Evaluar la eficacia del proceso de desidentificación del caso de estudio desarrollado utilizando la métrica definida"*. El diseño del experimento se guía por este propósito. Se empieza por especificar las preguntas locales de investigación asociadas a la evaluación de la eficacia de la instrumentación (sección 8.1.1), se explican luego las variables de respuesta utilizadas (sección 8.1.2), los factores de diseño (sección

8.1.3) y fijos (sección 8.1.4), la descripción de otros aspectos adicionales de diseño (sección 8.1.5 y se concluye con una explicación del método seguido para instrumentar el algoritmo de desidentificación: el $k$-anonimato (sección 8.1.6).

### 8.1.1. Preguntas de investigación que guían la evaluación

Las preguntas de investigación que se plantearon para la evaluación de la eficacia fueron las siguientes:

1. ¿Cuál es el efecto en la población total de usar el algoritmo instrumentado de desidentificación?
2. Específicamente para el caso de los estudiantes con preocupaciones de privacidad ¿Disminuye significativamente la probabilidad de reidentificación al usar el algoritmo de desidentificación instrumentado?
3. ¿Cuál es la relación entre la meta de privacidad de $1/k$ de $k$-anonimato y los porcentajes de reidentificación conseguidos?
4. ¿Qué efecto tiene usar diferentes valores de k?
5. ¿Qué efecto tiene en el algoritmo de aprendizaje máquina el uso de más o menos datos de entrenamiento?
6. ¿Cuánta es la magnitud del efecto en la población fundamentalista de la aplicación del algoritmo instrumentado de desidentificación?
7. ¿Qué tanto es afectada la utilidad del conjunto de datos al aplicar la desidentificación instrumentada?

Para poder responder estas preguntas, se describe el planteamiento del experimento en las siguientes secciones, empezando por la definición de las variables de respuesta

### 8.1.2. Variables de respuesta

Como se indicó en el capítulo 7, en el escenario estudiado, la UCR es la administradora de la base de datos, y los estudiantes son los dueños de los datos. Las consultas

provienen de fuentes externas, y los datos a entregar deben ser desidentificados antes de ser entregados (publicados).

Para realizar la evaluación, se utilizaron como variables de respuesta las dos métricas principales que se han discutido en la tesis: la **métrica de probabilidad de reidentificación basada en aprendizaje máquina** propuesta y evaluada en el capítulo 6, y una métrica estándar de evaluación de pérdida de información, la **Métrica de Clasificación (CM)**, explicada en el capítulo 2.

Siguiendo lo explicado en el capítulo 7, en el caso de la métrica de reidentificación por aprendizaje máquina, la reidentificación se definió como la correcta identificación, por parte de un atacante, de la tripleta (desidentificada) de (Provincia, Cantón, Distrito) de un estudiante.

El algoritmo de aprendizaje máquina utilizado para la métrica de reidentificación fue el Árbol de Clasificación y Regresión (CART por sus iniciales en inglés) [Breiman et al., 1984], tal como está implementado en el módulo scikit-learn de Python [Pedregosa et al., 2011]. Luego de entrenar el modelo con un porcentaje del conjunto de datos se reidentifican los elementos restantes, y se mide la proporción de los correctamente reidentificados. CART fue seleccionado por su relativa facilidad de uso, poca demanda de poder computacional, y capacidad de trabajar con conjuntos de entrenamiento relativamente pequeños. Adicionalmente, el módulo de Python genera salidas fácilmente procesables[Berk, 2016].

Con respecto a la métrica de utilidad, se utiliza CM y no GCP (utilizada en el capítulo 6) para poder explorar el comportamiento de una métrica adicional de pérdida de información estándar ante la aplicación de la instrumentación propuesta.

### 8.1.3. Factores de diseño

Para responder las preguntas planteadas con respecto a la eficacia de la instrumentación realizada al algoritmo de desidentificación, se decidió trabajar con tres factores de diseño: la desidentificación usada, los k base, y el tamaño del conjunto de entrenamiento para el algoritmo de aprendizaje máquina, todos los cuales se explican a continuación

El factor de diseño principal es claramente el que se refiere al efecto de la aplicación del $k$-anonimato instrumentado en contraposición a la aplicación del $k$-anonimato estándar. Para este efecto, se definió el factor **Desidentifición** con dos valores posi-

bles: "$k$-anonimato estándar" y "$k$-anonimato instrumentado". Este factor de diseño se requiere para las siete preguntas planteadas.

El segundo factor de diseño en importancia son los distintos valores de $k$ utilizados en $k$-anonimato (instrumentado o sin instrumentar). Este factor tiene impacto directo en las preguntas 3 y 4, y se utiliza para comparar respuestas en todas las demás preguntas. Este factor fue denominado $k$ **base**. Si se usa el $k$-anonimato instrumentado, el $k$ base se le aplica a los usuarios "despreocupados" pero si en un grupo hay usuarios "fundamentalistas" se usa el $k$ base más un factor de incremento. Los valores utilizados fueron 60, 90, 120 y 150, según se explica en la sección 8.1.6.

La pregunta de investigación 5 es más compleja de lo que aparenta, dado que contiene un supuesto muy importante con respecto a los datos que el atacante tiene a su disposición para realizar el aprendizaje de máquina. En esta investigación esto se hizo equivalente al subconjunto de datos de aprendizaje de cada bloque. Esto es una forma cerrada de describir el conocimiento del atacante, pero de hecho es un mejor caso, dado que se supone que el atacante tiene acceso a todos los datos completos de un subconjunto de la población. Es por esto que este supuesto se considera válido. Para la evaluación se optó por agregar un factor llamado **Porcentaje de información externa**, que equivale al porcentaje de datos dentro de cada bloque utilizados para entrenar el algoritmo de aprendizaje máquina. Se utilizaron únicamente dos valores, 5 % y 10 %. Por lo tanto, el conjunto de datos contenía 95 % y 90 % de elementos previamente desconocidos por el modelo. Este factor se utiliza, asimismo, para responder las preguntas 2, 6 y 7.

Los factores de diseño se resumen en el cuadro 8.1.

Cuadro 8.1: Resumen de factores de diseño

| **Nombre del factor** | **Valores** |
|---|---|
| $k$ | 60, 90, 120, 150 |
| Porcentaje de información "externa" del atacante | 5 %, 10 % |
| Desidentificación | Generalización estándar, Generalización con preferencias de usuario |

### 8.1.4. Factores fijos

Como se indicó en el capítulo 7, el conjunto de datos resultado de la consulta para cumplir con la solicitud externa, ya filtrado para remover inconsistencias, contiene 42.538 filas, y se usaron las seis columnas resumidas en la tabla 7.4.

La instrumentación del *k-anonimato* se realizó siguiendo lo descrito en la sección 5.2. Se definió que el factor de incremento utilizado sobre el $k$ para los grupos que contienen estudiantes con preferencias de bloqueo es de 30. El valor de 30, bastante modesto en sí, se utilizó porque los valores de $k$ explorados en el factor de diseño $k$ base avanzaban de 30 en 30, y se consideró que los "más privados" básicamente se pasaban a la siguiente categoría.

### 8.1.5. Otros aspectos de diseño

Para poder contar con varias iteraciones, dado que el conjunto de datos era limitado, se crearon 30 bloques de 10,634 filas cada uno. Las filas que constituían cada bloque fueron elegidas aleatoriamente del conjunto total de 42,538 filas. Cada uno de estos bloques constituye la **unidad experimental** a la que se le aplicaba el proceso de (a) desidentificación, y (b) de reidentificación mediante aprendizaje máquina. Los bloques fueron creados una única vez y utilizados consistentemente para todos los procesos. Debido a este diseño de bloqueo, se obtuvieron resultados más robustos en el análisis estadístico.

Como se explicó en la sección 8.1.6, las preferencias de cada estudiante en el conjunto fueron asignadas aleatoriamente con una probabilidad de 0.26 de ser "bloquear", y de 0.74 de ser "permitir". Esta asignación se realizó una única vez sobre el conjunto total de 42.538 filas. Esto significa que a lo interno de cada bloque de 10.634 filas, las proporciones de estudiantes "fundamentalistas" y "despreocupados" varían ligeramente. Este factor se denomina **Población**. Dado que este es un factor permitido a variar medible, se incluyen resultados donde se miden efectos sobre la población de fundamentalistas.

Todas la combinaciones de factores fueron probadas en cada subconjunto, y lo que se reporta en los resultados corresponde a distintos promedios de la aplicación de cada combinación de factores a los 30 subconjuntos. La ejecución del experimento primero aleatorizó el orden de aplicación de cada combinación de factores a los subconjuntos.

La técnica estadística utilizada para el análisis fue ANOVA factorial completo, aun-

que los resultados se presentan únicamente para las combinaciones que corresponden a las preguntas de investigación planteadas.

Luego de explicar los aspectos importantes del diseño y organización del experimento, se presenta una breve explicación de la instrumentación que se realizó al algoritmo de desidentificación utilizado en la evaluación: el k-anonimato.

### 8.1.6. Instrumentación de k-anonimato

Se implementó una versión de k-anonimato para el problema dado, según los principios indicados en el capítulo 5. La generalización se guió por semántica jerárquica de la división territorial administrativa de Costa Rica. La tripleta (Provincia, Cantón, Distrito) se generalizó como un único atributo. La generalización preserva la composición geográfica.

La asignación de preferencias explicadas en la sección 7.6 se aplicó siguiendo la modificación a k-anonimato discutida en la sección 5.2.

El k-anonimato depende críticamente del valor de $k$. Desde el trabajo original de Sweeney [Sweeney, 2002b], la selección de un $k$ correcto, se ha dejado principalmente a practicantes basados en su experiencia. A pesar de que varios artículos se han publicado sobre cómo juzgar la elección de $k$, o cómo seleccionarlo (por ejemplo Lodha y Thomas [Lodha and Thomas, 2007]), no existe una respuesta definitiva. En este trabajo se supone que un experto seleccionó los valores de $k$ usados, pero adicionalmente se aplica el método de Lodha para evaluar las escogencias de $k$. Lodha y Thomas [Lodha and Thomas, 2007] definen un parámetro $\alpha$ que mide la proporción de tuplas únicas (singletones) en el conjunto de cuasi-identificadores. Cualquier $k$ que logre un $\alpha$ debajo de 0.5 se considera una elección razonable. Intuitivamente, tanto más pequeño el $\alpha$, tanto mejor es la preservación de la privacidad de los datos. Sin embargo, la medición de la calidad de la preservación de la privacidad no es simple, y requiere analizar los datos a ser publicados.

Utilizando únicamente la operación de generalización en el k-anonimato, se trabajó con cuatro diferentes valores de $k$, seleccionados por conocimiento experto. Estos valores fueron confirmados como válidos aplicando el criterio del alfa de Lodha [Lodha and Thomas, 2007], esto para aumentar la rigurosidad del caso de estudio.

El cuadro 8.2 presenta los cuatro valores de $k$ y sus respectivos valores del alfa

de Lodha. Como se puede ver, todos están por debajo del máximo recomendado por Lodha(0,5).

Cuadro 8.2: Valores utilizados de $k$ y sus respectivos alfas de Lodha.

| Valor de $k$ | Valor del alfa de Lodha |
|---|---|
| 60 | $0{,}4 * 10^{-2}$ |
| 90 | $0{,}3 * 10^{-}2$ |
| 120 | $0{,}2 * 10^{-2}$ |
| 150 | $0{,}1 * 10^{-2}$ |

## 8.2. Resultados y Análisis

De previo, se realizaron todos los chequeos de normalidad, aleatoriedad y varianza constante necesarios para el uso de ANOVA factorial completo. El modelo es significativo, al igual que cada uno de los factores de diseño, y prácticamente todas las combinaciones de factores.

Los resultados se van a analizar de acuerdo con las preguntas de investigación planteadas.

La pregunta 1 era: *¿Cuál es el efecto en la población total de usar el algoritmo instrumentado de desidentificación?*. En la figura 8.1 se presenta el porcentaje promedio de estudiantes correctamente reidentificados para cada valor de $k$. La curva azul muestra las proporciones antes de la aplicación de preferencias, y la curva roja representa las proporciones luego de la aplicación de preferencias. Todos los promedios son significativamente diferentes ($p < 0.0001$). Los porcentajes fueron calculados para todo el conjunto, que contiene estudiantes con y sin preferencias. El intervalo de confianza del 99 % está graficado pero es difícil de apreciar debido a lo cerca que está del valor promedio. Por lo tanto, el efecto en la población total en estudio es que al usar el $k$-anonimato instrumentado, la probabilidad de reidentificación disminuye.

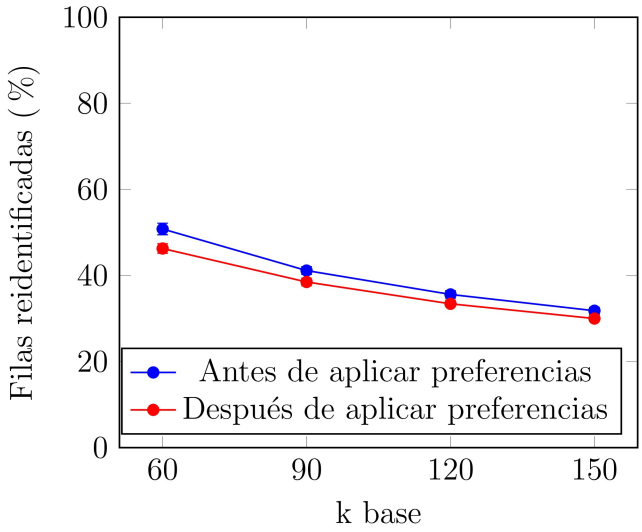
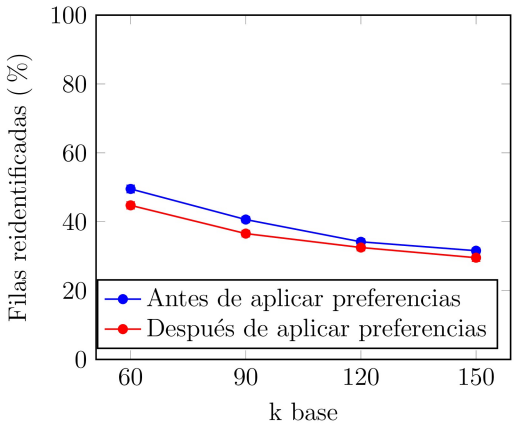
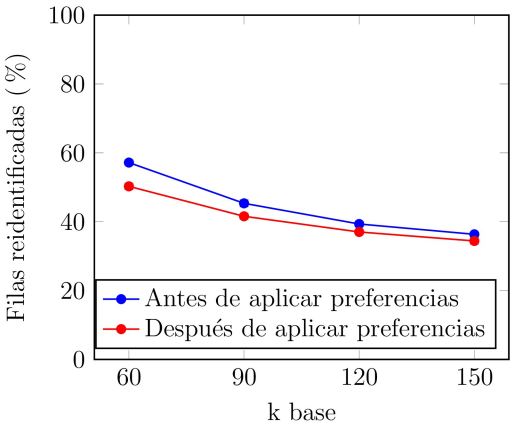



Figura 8.1: Porcentaje promedio de estudiantes re-identificados para cada valor de $k$ base. Las curvas azul y rojo muestran los porcentajes de reidentificación antes y después de la aplicación de preferencias.

La segunda pregunta se planteaba: *Específicamente para el caso de los estudiantes con preocupaciones de privacidad ¿Disminuye significativamente la probabilidad de reidentificación al usar el algoritmo de desidentificación instrumentado?*. Para responderla, se muestra la figura 8.2 en la que se observa que el **porcentaje promedio** de estudiantes correctamente reidentificados de esta sub-población luego de aplicar las preferencias **es menor**. Aunque es difícil de apreciarlo en el gráfico, el efecto es estadísticamente significativo, excepto para el caso de k=120 cuando la información del atacante es del 5 % de los datos. Dado que los porcentajes de reidentificación son menores con menos información disponible, no es de extrañar que en algunos casos, con $k$ más grandes, no fueran diferentes. Para el 10 % de información disponible, en todos los casos el porcentaje de reidentificación es menor, o sea que el efecto de esconder gente mejor es más notorio cuando el atacante tiene más información. Los intervalos de confianza del 99 % están graficados pero casi no se observan debido a la cercanía de los promedios. Por lo tanto, la respuesta a la pregunta 2 es afirmativa.




*Información de atacante: 5 % información no desidentificada.*




*Información de atacante: 10 % de información no desidentificada.*

Figura 8.2: Porcentaje promedio de estudiantes correctamente reidentificados que escogieron tener más privacidad para cada valor de $k$.

Con respecto a la pregunta 3: *¿Cuál es la relación entre la meta de privacidad de $1/k$ de $k$-anonimato y los porcentajes de reidentificación conseguidos?* se debe referir de nuevo a la figura 8.1. El porcentaje de estudiantes re-identificados es consistentemente mayor que lo que la meta de privacidad de $1/k$ nos llevaría a pensar, dado que sobrepasa el 25 % para todos los $k$ y tipos de privacidad aplicados. El hecho de que la reidentificación usando $k$-anonimato puede ser mayor a la meta de $1/k$ esperado es un fenómeno bien documentado (véase, por ejemplo, [Ohm, 2010]). Los resultados presentados únicamente contribuyen a añadir evidencia al hecho de que los algoritmos de desidentificación solos son insuficientes para garantizar la privacidad.

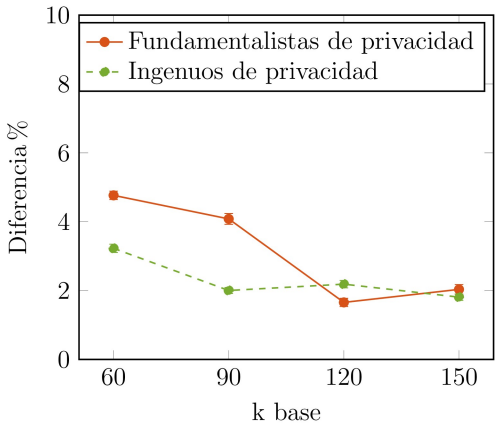
La pregunta 4 plantea: *¿Qué efecto tiene usar diferentes valores de $k$?*. Tal como era de esperar, la proporción de elementos reidentificados disminuye a como el valor de $k$ se incrementa (dado que se crean grupos más grandes). Este efecto se nota en todos los gráficos presentados en este capítulo.

La pregunta 5, refiere a los datos disponibles al atacante: *¿Qué efecto tiene en el algoritmo de aprendizaje máquina el uso de más o menos datos de entrenamiento?*. Volviendo ahora al gráfico 8.2 se puede distinguir claramente que para cada $k$ específico, los porcentajes de reidentificación correcta son consistemente menores para el 5 % y mayores para el 10 % de datos disponibles para el entrenamiento. Si bien esto no es sorprendente en sí mismo, al igual que para el caso de la pregunta 4, es una indicación

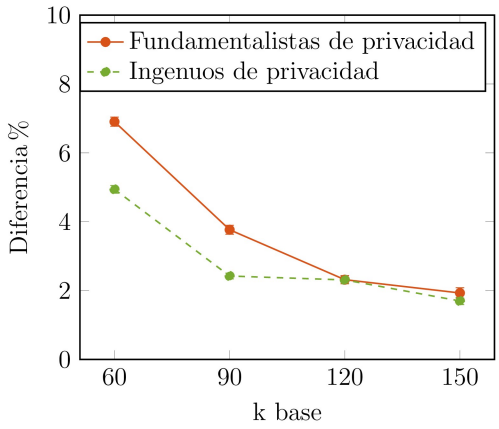
de que los valores elegidos de $k$ base, las columnas elegidas como cuasi-identificadores, y la caracterización de los datos fueron escogidos correctamente, y que no existen efectos de *overfitting*.

En la pregunta 6 se cuestiona la magnitud del efecto: *¿Cuánta es la magnitud del efecto en la población fundamentalista de la aplicación del algoritmo instrumentado de desidentificación?*. Se definió por magnitud de efecto la diferencia, por bloque y combinación de valores de factores, entre los promedios de reidentificación, lo que genera una tercer variable (secundaria) de respuesta. En la figura 8.3 se muestran los promedios de diferencias con el 99 % de intervalo de confianza. Las curvas corresponden al 5 % y 10 % de niveles de entrenamiento para el algoritmo de aprendizaje máquina. Las diferencias se encuentran normalizadas a porcentajes, que es lo que se muestra en el eje Y. Excepto para $k = 60$ y $k = 90$ hay diferencias significativas entre las curvas, lo que significa que el algoritmo de aprendizaje tuvo poco efecto en la magnitud de las diferencias con y sin el uso del algoritmo instrumentado. Sin embargo, lo importante es que, en todos los niveles de $k$, la aplicación de la preferencias de privacidad sobre el bloqueo de inferencias redujo el porcentaje de elementos reidentificados en al menos un 3 %, que sobre la población en cada bloque de 10,634 filas, significa que más de 300 estudiantes adicionales ya no fueron reidentificados gracias al uso del algoritmo instrumentado de reidentificación. Esto proporciona evidencia adicional con respecto a que la aplicación de las preferencias fue valiosa para mejorar la privacidad de los estudiantes fundamentalistas.

Es importante anotar asimismo que el porcentaje tan pequeño de diferencia se puede deber al incremento tan conservador de únicamente 30 unidades de $k$ adicionales para los usuarios con preferencias de bloqueo que se usó como factor fijo en la instrumentación. Sería interesante realizar pruebas para determinar un valor de $k$ que haga un mayor impacto en la protección de los que solicitan más privacidad.




*Información del atacante: 5 % de información no desidentificada.*




*Información del atacante: 10 % de información no desidentificada.*

Figura 8.3: Diferencia apareada de las medias, antes y después de la aplicación de las preferencias, intervalo de confianza a 99 %.

Finalmente, se llega a la pregunta 6: *¿Qué tanto es afectada la utilidad del conjunto de datos al aplicar la desidentificación instrumentada?*. La figura 8.4 presenta la pérdida de utilidad (medido usando la métrica CM) para toda la población comparando los resultados usando $k$-anonimato estándar y $k$-anonimato instrumentado, para 5 % y 10 % de datos de entrenamiento disponibles para el atacante. En primer lugar, la tendencia de pérdida de información es casi lineal asociado a mayores valores de $k$ base, lo cual es esperable, llegando a casi 0,25 (o sea, una cuarta parte de las filas han sido "dañadas" por la desidentificación) para el caso de k=150. Esto es completamente esperable y documentado en la literatura que usa CM. Por otra parte, es interesante ver que el mayor salto entre las curvas corresponde al uso o no del algoritmo instrumentado, mientras que las curvas difieren muy poco entre sí para las dos con y las dos sin instrumentación, pero sin nunca pasar del 0,25. De este comportamiento se puede afirmar que el impacto en la utilidad de los datos es muy limitado, lo cuál es positivo para poder recomendar el uso de las instrumentaciones basadas en preferencias de privacidad.

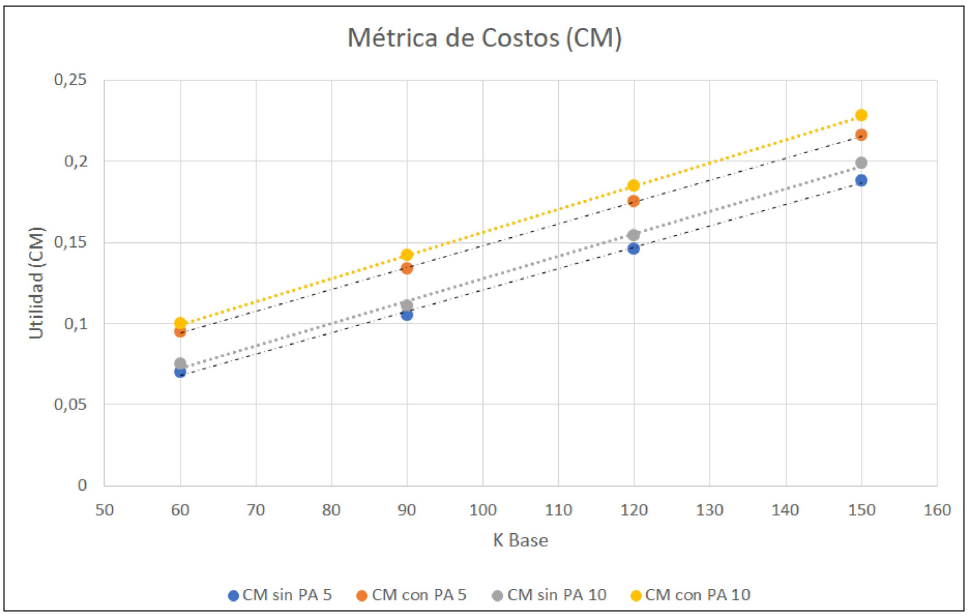



Figura 8.4: Costo en pérdida de utilidad de la aplicación de desidentificación instrumentada y sin instrumentar, con 5 % y 10 % de datos de entrenamiento. Importante: el eje Y va de 0 a 0.25 para poder visualizar las diferencias entre las curvas, que son mínimas. La notación "con/sin PA" se refiere al uso o no de $k$-anonimato instrumentado. Los números 5 y 10 corresponden a los porcentajes de datos de entrenamiento

Con base a las respuestas a las preguntas asociadas a la evaluación de la eficacia, se puede afirmar que el objetivo 4 fue cumplido.

## 8.3. Resumen

Se realizó un experimento guiado por siete preguntas de investigación con respecto a la eficacia de la instrumentación propuesta del algoritmo de $k$-anonimato. El diseño experimental fue de tipo factorial completo con tres variables de diseño explícitas y una implícita (el factor permitido a variar de porcentaje de estudiantes con selección fundamentalista).

Con respecto a cada pregunta de investigación, se determinó que:

- Aún para la población total, el uso del $k$-anonimato instrumentado conduce a un menor porcentaje de reidentificación de los individuos

- En el caso de los estudiantes fundamentalistas de la privacidad, el uso del $k$-anonimato instrumentado conduce a porcentajes significativamente menores de reidentificación
- En el caso de $k$-anonimato se evidencia que la promesas de que la probabilidad de reidentificación es de $1/k$ crea expectativas falsas, ya que la probabilidad de reidentificación para cada $k$ dado es mucho más alta
- Para $k$-anonimato, se deben explorar varios valores válidos de $k$ antes de comprometerse con uno definitivo para la publicación, ya que los efectos en la probabilidad de reidentificación varían significativamente
- Los atacantes siempre tienen la ventaja, dado que aún con cantidades de datos externos relativamente pequeñas es posible reidentificar proporciones grandes de la población. Sin embargo, el efecto del $k$-anonimato instrumentado es mayor cuando el atacante tiene más datos, agregando valor a la solución.
- La magnitud del cambio en la población fundamentalista de la aplicación del $k$-anonimato instrumentado es significativo, pero con los parámetros utilizados en este experimento particular, es limitado (alrededor del 3 % de diferencia)
- La utilidad de los datos a entregar es afectada solo marginalmente por la aplicación del algoritmo instrumentado de $k$-anonimato.

Todas las respuestas anteriores permiten concluir que el uso de una instrumentación de un algoritmo de desidentificación guiada por preferencias individuales de privacidad, es eficaz en su cometido de conseguir (probabilísticamente) una mayor protección de la privacidad a los individuos que así lo deseen. Sin embargo, dado que el efecto detectado es relativamente pequeño, abre posibilidades de investigación futura en la determinación de mejores parámetros a usar para la instrumentación, que lleven a efectos de protección de la privacidad más grandes.

En resumen, los dueños de los datos que solicitaron más protección, obtuvieron mayor protección. La utilidad del conjunto de datos a publicar disminuyó de forma poco significativa, y el porcentaje de datos reidentificados fue siempre superior a la promesa de $1/k$.

En el siguiente capítulo se describen las conclusiones y trabajo futuro de esta tesis.

# Capítulo 9

# Conclusiones

La preservación de la privacidad en datos a publicar es un problema ampliamente estudiado en la búsqueda de proteger la privacidad individual de los dueños de los datos. Pero de la mano con los estudios de privacidad se requiere crear medios simples, consistentes e informativos para que los dueños de los datos ejerzan sus derechos sobre los datos. Esta tesis desarrolló el uso de inferencias como expresión de una posible afectación de la privacidad, permitiendo expresar en lenguaje natural la posible violación a la privacidad y a la vez, presentar al dueño de los datos una forma simple, informativa y consistente de restringir el uso de sus datos.

Los posibles usos de un conjunto de datos son difíciles de identificar, por ese motivo en el capítulo 4 se inició con la definición de un modelo de representación de inferencias, para lo cual se definieron procesos que permiten realizar esta labor y aprovecharla en el contexto de la preservación de la privacidad.

El primer proceso presentó como seleccionar inferencias que tengan sentido en el contexto de una base de datos específica. Este proceso requiere un amplio conocimiento sobre la base de datos y aún requiere trabajo para un desarrollo automatizado. El siguiente proceso permite identificar los elementos de la base de datos que inciden en la realización de una inferencia. Este proceso se apoya en el conocimiento de la base de datos y mecanismos de consulta de la base de datos como el lenguaje SQL. Para poder agregar utilidad a toda esta extracción de conocimiento se indica una propuesta para almacenar las preferencias de bloqueo de los dueños de los datos. Esto permite que se tengan agrupados los conocimientos sobre la preferencias de los dueños de los datos, así como las inferencias y su mapeo en la base de datos. El ultimo paso consistió en relacionar consultas de publicación de datos con los elementos asociados a las inferencias y con las preferencias de bloqueo de los dueños de los datos. Esta relación permitió identificar que tuplas se verían afectadas en una posterior publicación de datos.

Una vez presentado el modelo que permite la extracción del conocimiento que identifica y define las inferencias se definió una metodología para instrumentar procesos de

desidentificación en el capítulo 5. Esta metodología considera las preferencias de los dueños de los datos sobre posibles inferencias que se podrían realizar sobre los mismos. Asimismo, se instrumentó una implemententación de método de desidentificación del modelo $k$-anonimato, para que considerara las preferencias individuales sobre inferencias de datos.

Todo esto permitió cumplir el objetivo específico 1: *Definir un modelo para expresar posibles inferencias sobre consultas a una base de datos relacional que permitan tomar en cuenta las preferencias individuales de privacidad.*

En el capítulo 6 se plantea una métrica que permite evaluar si los bloqueos a las inferencias pueden tener efecto sobre la protección a los usuarios con mayores expectativas de preservación de la privacidad. Concretamente, se definió una métrica que mide la probabilidad de reidentificación utilizando la exactitud registrada mediante un algoritmo de aprendizaje máquina. Además, se evaluó el uso de la métrica utilizando los algoritmos de árboles de decisión, bosques aleatorios y regresión lineal. Esto permitió observar que la efectividad de su uso y sus resultados distaban de los presentados por otras métricas de privacidad.

La evaluación de la métrica fue un aporte adicional al trabajo de tesis, debido a que esto no estaba contemplado entre los objetivos específicos. La evaluación de la métrica permitió descubrir evidencia sobre la efectividad de la métrica propuesta y como se convierte en un mecanismo natural para medir la privacidad si se conoce de antemano los posibles riesgos a los que se vera expuesto el dueño de los datos. Asimismo, se encontró evidencia sobre la poca efectividad de las métricas de privacidad descritas en la literatura y el poco trabajo existente en vías a tener una métrica de privacidad universal. Esto redundó en la existencia de serias dificultades para establecer buenos parámetros cuando se evalúan y comparan diferentes algoritmos de desidentificación. La métrica descrita permitió cumplir el objetivo específico 3: *Definir una métrica que identifique la eficacia de un bloqueo de inferencia de datos para un contexto particular.*

En el capítulo 7 se muestra un caso de estudio sobre el que se aplicó el modelo antes mencionado y se siguieron todos los pasos enunciados para la extracción del conocimiento de la base de datos y de la preferencias de los dueños de los datos. Se logró seleccionar inferencias aplicables a la base de datos. También generar las reglas que permiten identificar los elementos que componen una inferencia. Así como simular la captura de preferencias de los dueños de los datos sobre el bloqueo de inferencias de

datos. Con esto, se cumplió el objetivo específico 2: *Desarrollar un caso de estudio que incorpore un conjunto de inferencias especificadas con el modelo de expresión definido, en un modelo de desidentificación conocido.*

Finalmente, se logro identificar claramente el efecto positivo de tomar en cuenta las preferencias mediante la evaluación del modelo de expresión de inferencias aplicado al caso de estudio por medio de un experimento con diseño de ANOVA factorial completo y que se presentó en el capítulo 8.

El experimento permitió mostrar que los dueños de los datos que solicitaron más protección, recibieron mayor protección. Además, que la utilidad del conjunto de datos a publicar disminuyó de forma poco importante y que el porcentaje de reidentificación fue muy superior de lo que la promesa del valor $k$ permitía suponer. Esto da evidencia relevante para continuar trabajando en la búsqueda de vanzar en el desarrollo de una preservación de la privacidad real y más informativa para los dueños de los datos. Por otra parte, la evidencia presentada en este capítulo permite el cumplimiento del objetivo especifico 4: *Evaluar la eficacia del proceso de desidentificación del caso de estudio desarrollado utilizando la métrica definida.*

Tomando en cuenta que se cumplen cada uno de los objetivos específicos, se dío por cumplido el objetivo general de la tesis que buscaba incorporar preferencias individuales de privacidad basadas en inferencias, a procesos de desidentificación de datos de consultas en bases de datos relacionales.

## 9.1. Contribuciones

Las contribuciones del trabajo realizado se resumen de la siguiente manera:

- Se desarrolló una nueva forma de informar a la persona sobre los riesgos de privacidad
- Se propuso un nuevo método de obtener el consentimiento informado
- Se diseñó una forma de desidentificar en función de las preferencias de los dueños de los datos
- Se desarrolló una nueva métrica de privacidad, asociada con riesgo reales, con base en la aplicación de algoritmos de aprendizaje máquina para reidentificación

de individuos. Dicha métrica no solo se presentó (tal como lo indicaba el objetivo específico respectivo), sino que se probó exitosamente en conjuntos de datos estándar, y en competencia con otras métricas de privacidad.

- Se presentó un modelo de preservación de la privacidad que se puede adoptar a otros requisitos de información con un criterio de selección diferente.
- Se demostró el uso de inferencias a datos para presentar una elección negativa del tratamiento autorizado por un usuario.

La finalización de esta tesis hace evidente que existen muchos aspectos por trabajar en la búsqueda de una mejor preservación de la privacidad. La variedad de aspectos e interacciones disciplinarias es amplia por lo que a continuación se mencionan posibles rutas de trabajo futuro.

### 9.1.1. Trabajo futuro

Aunque existen muchas direcciones de trabajo futuro generadas por esta tesis, algunas de las más interesantes se citan a continuación:

- Profundizar en la metodología de identificación de preferencias relevantes de bloqueo de inferencias, preferiblemente con el apoyo de profesionales en psicología
- Crear una guía que describa en detalle la metodología de adaptación de las preferencias de bloqueo en lenguaje natural a elementos de la base de datos, y evaluar la guía
- Extender el alcance de la instrumentación de algoritmos de desidentificación a otros modelos más complejos, como L-diversidad.
- Realizar un análisis de sensitividad con respecto al alcance de las instrumentaciones de algoritmos de desidentificación con conjuntos de datos con diversas características
- Explorar otras técnicas de aprendizaje máquina para reidentificar individuos
- Desarrollar diferentes métodos para simular el conocimiento del atacante, para entrenar los algoritmos de aprendizaje máquina

- Aplicar metodologías de interacción humano computador para probar y obtener las preferencias de bloqueo de inferencias de parte de los usuarios, para proveer un mecanismo adecuado para la expresión del consentimiento informado.
- Validar empíricamente todas las métricas de privacidad existentes con el fin de estandarizar el uso de aprendizaje máquina en la medición de la preservación de la privacidad.
- Cuantificar el conocimiento qué puede ser recabado acerca de la vida de un individuo
- Cuantificar las consecuencias qué puede tener una publicación de datos

# Bibliografía

[Acquisti et al., 2015] Acquisti, A., Brandimarte, L., and Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221):509–514.

[Aggarwal et al., 2005] Aggarwal, G., Feder, T., Kenthapadi, K., Motwani, R., Panigrahy, R., Thomas, D., and Zhu, A. (2005). Anonymizing tables. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3363 LNCS:246–258.

[Agrawal et al., 1993] Agrawal, R., Imieliński, T., Swami, A., Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93*, volume 22, pages 207–216, New York, New York, USA. ACM Press.

[Agrawal and Srikant, 2000] Agrawal, R. and Srikant, R. (2000). *Privacy-preserving data mining*, volume 29 of *Advances in Database Systems*. Springer US.

[Aimeur et al., 2012] Aimeur, E., Brassard, G., and Molins, P. (2012). Reconstructing profiles from information disseminated on the internet. In *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*, pages 875–883. Ieee.

[Al-Rubaie and Chang, 2019] Al-Rubaie, M. and Chang, J. M. (2019). Privacy-Preserving Machine Learning: Threats and Solutions. *IEEE Security and Privacy*, 17(2):49–58.

[Alanezi and Mishra, 2018] Alanezi, K. and Mishra, S. (2018). A privacy negotiation mechanism for the internet of things. *Proceedings - IEEE 16th International Conference on Dependable, Autonomic and Secure Computing, IEEE 16th International Conference on Pervasive Intelligence and Computing, IEEE 4th International Conference on Big Data Intelligence and Computing and IEEE 3*, pages 512–519.

[Ayala-Rivera et al., 2014] Ayala-Rivera, V., McDonagh, P., Cerqueus, T., and Murphy, L. (2014). A Systematic Comparison and Evaluation of k -Anonymization Algorithms for Practitioners. *Transactions on Data Privacy*, 7(3):337–370.

[Aydoğan et al., 2017] Aydoğan, R., Øzturk, P., and Razeghi, Y. (2017). Negotiation for Incentive Driven Privacy-Preserving Information Sharing. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10621 LNAI, pages 486–494. Springer, Cham.

[Barbaro and Zeller, 2006] Barbaro, M. and Zeller, T. (2006). A Face Is Exposed for AOL Searcher No. 4417749. *New York Times*, 1(4417749):1–3.

[Bates, 1964] Bates, A. P. (1964). Privacy-a Useful Concept? *Social Forces*, 42(4):429–434.

[Bayardo and Agrawal, 2005] Bayardo, R. J. and Agrawal, R. (2005). Data privacy through optimal k-anonymization. In *Proceedings - International Conference on Data Engineering*, pages 217–228.

[Berk, 2016] Berk, R. A. (2016). Classification and Regression Trees (CART). In *Statistical Learning from a Regression Perspective*, pages 129–186. Springer, Cham.

[Bertino et al., 2008] Bertino, E., Lin, D., and Jiang, W. (2008). A Survey of Quantification of Privacy Preserving Data Mining Algorithms. *PrivacyPreserving Data Mining*, 34:183–205.

[Bilenko et al., 2003] Bilenko, M., Mooney, R., Cohen, W., Ravikumar, P., and Fienberg, S. (2003). Adaptive Name Matching in Information Integration. *IEEE Intelligent Systems*, 18(5):16–23.

[Bodle, 2013] Bodle, R. (2013). The ethics of online anonymity or Zuckerberg vs. Moot. *ACM SIGCAS Computers and Society*, 43(1):22–35.

[Bonomi et al., 2015] Bonomi, L., Fan, L., and Xiong, L. (2015). *A Review of Privacy Preserving Mechanisms for Record Linkage*, pages 233–265. Springer International Publishing, Cham.

[Breiman et al., 1984] Breiman, L., Friedman, J., Stone, C., and Olshen, R. (1984). *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis.

[Byun et al., 2007] Byun, J.-W., Kamra, A., Bertino, E., and Li, N. (2007). Efficient k-Anonymization Using Clustering Techniques. In *Advances in Databases: Concepts, Systems and Applications*, pages 188–200. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Chaudhuri et al., 2011] Chaudhuri, S., Dayal, U., and Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8):88.

[Chaudhuri et al., 2002] Chaudhuri, S., Sarawagi, S., and Narasayya, V. (2002). Efficient evaluation of queries with mining predicates. In *Proceedings of 18th International Conference on Data Engineering*. IEEE Computer Society.

[Chen et al., 2008] Chen, B.-C., LeFevre, K., and Ramakrishnan, R. (2008). Adversarial-knowledge dimensions in data privacy. *The VLDB Journal*, 18(2):429–467.

[Chi-Wing et al., 2006] Chi-Wing, R., Li, J., Fu, A. W.-C., and Wang, K. (2006). ($\alpha$, k)-anonymity: An Enhanced k-Anonymity Model for Privacy-Preserving Data Publishing. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06*, page 754, New York, New York, USA. ACM Press.

[Conradie and Choenni, 2012a] Conradie, P. and Choenni, S. (2012a). Exploring process barriers to release public sector information in local government. In *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance - ICEGOV '12*, page 5, New York, New York, USA. ACM Press.

[Conradie and Choenni, 2012b] Conradie, P. and Choenni, S. (2012b). Exploring process barriers to release public sector information in local government. In *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance - ICEGOV '12*, page 5, New York, New York, USA. ACM Press.

[Cordoba Ortega, 2003] Cordoba Ortega, J. (2003). La legislación costarricense y el derecho de acceso a la información pública. *Revista de Ciencias Jurídicas*.

[Cormode et al., 2013] Cormode, G., Procopiuc, C. M., Entong Shen, Srivastava, D., and Ting Yu (2013). Empirical privacy and empirical utility of anonymized data. In *2013 IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, pages 77–82. IEEE.

[Dalenius, 1977] Dalenius, T. (1977). Towards a methodology for statistical disclosure control. *Statistik Tidskrift*, 15:429–444.

[DeCew, 2018] DeCew, J. (2018). Privacy. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2018 edition.

[Di Caterino, 2016] Di Caterino, M. (2016). S'impicca a 31 anni dopo video hot — Il Mattino.

[Domingo-ferrer and Torra, 2001] Domingo-ferrer, J. and Torra, V. (2001). Disclosure Control Methods and Information Loss for Microdata. *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*, pages 93–112.

[Dua and Graff, 2017] Dua, D. and Graff, C. (2017). UCI machine learning repository. http://archive.ics.uci.edu/ml. Accedido: 2017-02-07.

[Dwork, 2006] Dwork, C. (2006). Differential privacy. *Automata, languages and programming*, 4052:1–12.

[El Emam and Dankar, 2008] El Emam, K. and Dankar, F. K. (2008). Protecting Privacy Using k-Anonymity. *Journal of the American Medical Informatics Association*, 15(5):627–637.

[El Emam et al., 2011] El Emam, K., Jonker, E., Arbuckle, L., and Malin, B. (2011). A systematic review of re-identification attacks on health data. *PloS one*, 6(12):e28071.

[El-Khatib, 2003] El-Khatib, K. (2003). A Privacy Negotiation Protocol for Web Services. *Workshop on Collaboration Agents: Autonomous Agents for Collaborative Environments*, NRC 46518(October):85–92.

[El Parlamento Europeo, 2016] El Parlamento Europeo, C. d. l. U. E. (2016). Reglamento (UE) 2016/679 del parlamento europeo y del consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la D. *Diario Oficial de la Unión Europea. Serie L*, 59(119):1–88.

[Esponda et al., 2007] Esponda, F., Ackley, E. S., Helman, P., Jia, H., and Forrest, S. (2007). Protecting data privacy through hard-to-reverse negative databases. In *International Journal of Information Security*, pages 403–415.

[Esquivel-Quirós and Barrantes, 2015] Esquivel-Quirós, L. G. and Barrantes, E. G. (2015). Construyendo conciencia sobre amenazas a la privacidad. In *II Jornadas Costarricenses de Investigación en Computación e Informática*.

[Esquivel-Quirós and Barrantes, 2016] Esquivel-Quirós, L. G. and Barrantes, E. G. (2016). *An experience with a de-identifying task to inform about privacy issues*, volume 476, pages 49–60. Springer International Publishing, Cham.

[Esquivel-Quiros et al., 2018] Esquivel-Quiros, L. G., Barrantes, E. G., and Darlington, F. E. (2018). Measuring data privacy preserving and machine learning. *Applications in Software Engineering - Proceedings of the 7th International Conference on Software Process Improvement, CIMPS 2018*, pages 85–94.

[Esquivel-Quirós et al., 2019] Esquivel-Quirós, L. G., Elena, G. E., and Esponda Dalington, F. (2019). Marco de medición de la privacidad. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, N°31(31):66–81.

[Friedman et al., 2000] Friedman, B., Felten, E., and Millett, L. I. (2000). Informed Consent Online. *DisClosure*, pages 1–8.

[Fung et al., 2010a] Fung, B., Wang, K., Fu, A., and Yu, P. (2010a). *Introduction to Privacy-Preserving Data Publishing*. CRC Press.

[Fung et al., 2010b] Fung, B. C. M., Wang, K., Chen, R., and Yu, P. S. (2010b). Privacy-preserving data publishing. *ACM Computing Surveys*, 42(4):1–53.

[Garfinkel, 2015] Garfinkel, S. L. (2015). DRAFT NISTIR 8053, De-Identification of Personally Identifiable Information. Technical report, National Institute of Standards & Technology, Gaithersburg, MD, United States.

[Geest et al., 2005] Geest, T. V. D., Pieterson, W., and Vries, P. D. (2005). Informed consent to address trust, control, and privacy concerns in user profiling. *Privacy Enhanced . . .*, pages 1–10.

[Gefenas, 2012] Gefenas, E. (2012). *Encyclopedia of Applied Ethics*. Elsevier.

[Getoor and Machanavajjhala, 2012] Getoor, L. and Machanavajjhala, A. (2012). Entity resolution: Theory, practice #38; open challenges. *Proc. VLDB Endow.*, 5(12):2018–2019.

[Ghinita et al., 2009] Ghinita, G., Karras, P., Kalnis, P., and Mamoulis, N. (2009). A framework for efficient data anonymization under privacy and accuracy constraints. *ACM Transactions on Database Systems*, 34(2):1–47.

[Ghooi, 2011] Ghooi, R. B. (2011). The Nuremberg Code. *Perspectives in clinical research*, 2(2):72–76.

[Gkoulalas-Divanis and Aonghusa, 2014] Gkoulalas-Divanis, A. and Aonghusa, P. M. (2014). Privacy protection in open information management platforms. *IBM Journal of Research & Development*, 58(1):1–11.

[Gopalan et al., 2012] Gopalan, R., Antón, A., and Doyle, J. (2012). UCON LEGAL. In *Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12*, page 227, New York, New York, USA. ACM Press.

[Goswami and Madan, 2017] Goswami, P. and Madan, S. (2017). Privacy preserving data publishing and data anonymization approaches: A review. In *2017 International Conference on Computing, Communication and Automation (ICCCA)*, pages 139–142. IEEE.

[Greenwald et al., 2013] Greenwald, G., MacAskill, E., and Poitras, L. (2013). Edward Snowden: the whistleblower behind the NSA surveillance revelations. *The Guardian*, page 1.

[Griffin, 2012] Griffin, M. (2012). Twitter sued over Hardy tweet. http://www.smh.com.au/technology/technology-news/twitter-sued-over-hardy-tweet-20120216-1tbxz.html. Accedido: 2012-02-16.

[Hamza and Hefny, 2013] Hamza, N. and Hefny, H. (2013). Attacks on Anonymization-Based Privacy-Preserving: A Survey for Data Mining and Data Publishing. *Journal of Information Security*, 2013(April):101–112.

[Han et al., 2012] Han, J., Kamber, M., and Pei, J. (2012). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers.

[Hastie et al., 2002] Hastie, T., Tibshirani, R., and Friedman, J. (2002). An introduction to logistic regression analysis and reporting. *Journal of Educational Research*, 96(1):3–14.

[Hastie et al., 2009] Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Number 3 in Springer Series in Statistics. Springer, New York, NY, New York, NY.

[Heeger, 2015] Heeger, E. (2015). Controlling Your Online Profile: Reality or an Illusion? A Research into Informed Consent as a Mechanism to Regulate Commercial Profiling. *SSRN Electronic Journal*.

[Henriksen-Bulmer and Jeary, 2016] Henriksen-Bulmer, J. and Jeary, S. (2016). Re-identification attacks - A systematic literature review. *International Journal of Information Management*, 36(6):1184–1192.

[Hodges and Creese, 2013] Hodges, D. and Creese, S. (2013). Breaking the Arc: Risk control for Big Data. In *2013 IEEE International Conference on Big Data*, pages 613–621. IEEE.

[Intensive Working Group of ACM SIGKDD Curriculum Committee, 2017] Intensive Working Group of ACM SIGKDD Curriculum Committee (2017). Curriculum Design: Introduction.

[Isaak and Hanna, 2018] Isaak, J. and Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*, 51(8):56–59.

[Iyengar, 2002] Iyengar, V. S. (2002). Transforming data to satisfy privacy constraints. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '02*, page 279, New York, New York, USA. ACM Press.

[Kelly et al., 2008] Kelly, D. J., Raines, R. A., Grimaila, M. R., Baldwin, R. O., and Mullins, B. E. (2008). A survey of state-of-the-art in anonymity metrics. In *Proceedings of the 1st ACM workshop on Network data anonymization - NDA '08*, page 31, New York, New York, USA. ACM Press.

[Khan et al., 2014] Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., Shiraz, M., and Gani, A. (2014). Big Data: Survey, Technologies, Opportunities, and Challenges. *The Scientific World Journal*, 2014:1–18.

[Kifer and Gehrke, 2006] Kifer, D. and Gehrke, J. (2006). Injecting utility into anonymized datasets. *Proceedings of the 2006 ACM SIGMOD international conference on Management of data - SIGMOD '06*, page 217.

[Kiran and Kavya, 2012] Kiran, P. and Kavya, N. (2012). A Survey on Methods, Attacks and Metric for Privacy Preserving Data Publishing. *International Journal of Computer Applications*, 53:20–28Kiran, P., & Kavya, N. (2012). A Survey on Me.

[Krumm, 2007] Krumm, J. (2007). Inference attacks on location tracks. In LaMarca, A., Langheinrich, M., and Truong, K. N., editors, *Pervasive Computing*, pages 127–143, Berlin, Heidelberg. Springer Berlin Heidelberg.

[LeFevre et al., 2006a] LeFevre, K., DeWitt, D., and Ramakrishnan, R. (2006a). Workload-aware anonymization. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06*, page 277, New York, New York, USA. ACM Press.

[LeFevre et al., 2006b] LeFevre, K., DeWitt, D. J., and Ramakrishnan, R. (2006b). Mondrian multidimensional K-anonymity. In *Proceedings - International Conference on Data Engineering*, volume 2006, page 25. IEEE.

[Leoni, 2012] Leoni, D. (2012). Non-interactive differential privacy: a survey. In *WOD '12 Proceedings of the First International Workshop on Open Data*, pages 40–52.

[Li et al., 2007a] Li, C., Shirani-Mehr, H., and Yang, X. (2007a). Protecting Individual Information Against Inference Attacks in Data Publishing. *Lecture Notes in Computer Science*, 4443(60503036):422–433.

[Li et al., 2007b] Li, N., Li, T., and Venkatasubramanian, S. (2007b). t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In *2007 IEEE 23rd International Conference on Data Engineering*, pages 106–115. IEEE.

[Liu et al., 2015] Liu, X., Xie, Q., and Wang, L. (2015). A Personalized Extended (a, k)-Anonymity Model. In *2015 Third International Conference on Advanced Cloud and Big Data*, pages 234–240. IEEE.

[Liu et al., 2017] Liu, X., Xie, Q., and Wang, L. (2017). Personalized extended ($\alpha$, k)-anonymity model for privacy-preserving data publishing. *Concurrency and Computation: Practice and Experience*, 29(6):e3886.

[Lodha and Thomas, 2007] Lodha, S. and Thomas, D. (2007). Probabilistic Anonymity. In *Privacy, Security, and Trust in KDD*, pages 56–79. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Luger et al., 2014] Luger, E., Rodden, T., Jirotka, M., and Edwards, L. (2014). How do you solve a problem like consent?: The workshop. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, pages 613–619, New York, NY, USA. ACM.

[Machanavajjhala et al., 2006] Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. (2006). L-diversity: privacy beyond k-anonymity. In *22nd International Conference on Data Engineering (ICDE'06)*, pages 24–24. IEEE.

[Machanavajjhala et al., 2007] Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007). L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, 1(1).

[MacKinnon, 1989] MacKinnon, C. A. (1989). *Toward a feminist theory of the state*. Harvard University Press.

[Malin et al., 2003] Malin, B., Sweeney, L., and Newton, E. (2003). Trail re-identification: learning who you are from where you have been. *Proc. LIDAP-WP12*, pages 1–10.

[McCallum and Wellner, 2003] McCallum, A. and Wellner, B. (2003). Object Consolidation by Graph Partitioning with a Conditionally Trained Distance Metric.

[Meyerson and Williams, 2004] Meyerson, A. and Williams, R. (2004). On the complexity of optimal K-anonymity. *Proceedings of the twentythird ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems PODS 04*, pages 223–228.

[Nakagawa et al., 2017] Nakagawa, T., Arai, H., and Nakagawa, H. (2017). Personalized Anonymization for Set-Valued Data by Partial Suppression. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 1003–1010. IEEE.

[Narayanan and Shmatikov, 2006] Narayanan, A. and Shmatikov, V. (2006). How to break anonymity of the netflix prize dataset. *CoRR*, abs/cs/0610105.

[Narayanan and Shmatikov, 2008] Narayanan, A. and Shmatikov, V. (2008). Robust Deanonymization of Large Sparse Datasets. In *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, SP '08, pages 111–125, Washington, DC, USA. IEEE Computer Society.

[Ohm, 2010] Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, 57:1701, 2010.

[Open Knowledge, 2016] Open Knowledge, F. (2016). The Open Data Handbook. Technical report, Open Knowledge, Foundation.

[Padyab and Ståhlbröst, 2015] Padyab, A. and Ståhlbröst, A. (2015). Privacy Enhancing Tools : A Literature Review on End-User Role and Evaluation. In *15th Seminar within the Framework of a Swedish IT Security Network for PhD students*.

[Parent, 1983] Parent, W. A. (1983). Privacy, morality, and the law. *Philosophy & Public Affairs*, 12(4):269–288.

[Pearson and Elahi, 2011] Pearson, S. and Elahi, T. (2011). Privacy assurance checking. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6545:427–456.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python . *Journal of Machine Learning Research*, 12:2825–2830.

[Pierangela Samarati, 2001] Pierangela Samarati (2001). Protecting respondents' identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*, 13(6):1010–1027.

[Politou et al., 2018] Politou, E., Alepis, E., and Patsakis, C. (2018). Forgetting personal data and revoking consent under the GDPR: Challenges and Proposed Solutions. *Journal of Cybersecurity*, April(April):1–20.

[Popa et al., 2012] Popa, R. A., Redfield, C. M. S., Zeldovich, N., and Balakrishnan, H. (2012). CryptDB. *Communications of the ACM*, 55(9):103.

[Posner, 1981] Posner, R. A. (1981). The Economics of Privacy. *The American Economic Review*, 71(2):405–409.

[Probst, 2016] Probst, C. W. (2016). Guaranteeing privacy-observing data exchange. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9952 LNCS, pages 958–969. Springer, Cham.

[Rahman et al., 2013] Rahman, F., Hoque, M. E., Ahamed, S. I., and Alam, M. A. U. (2013). Preserving User Privacy in Pervasive Environments with a Collaborative Model. In *2013 IEEE Seventh International Conference on Software Security and Reliability Companion*, pages 84–93. IEEE.

[Rao et al., 2013] Rao, W., Chen, L., and Tarkoma, S. (2013). Toward Efficient Filter Privacy-Aware Content-Based Pub/Sub Systems. *IEEE Transactions on Knowledge and Data Engineering*, 25(11):2644–2657.

[República de Costa Rica, 2001] República de Costa Rica, A. L. (2001). Ley 8148-Código Penal - Ley 4573 - Delitos Informáticos.

[República de Costa Rica, 2011] República de Costa Rica, A. L. (2011). Ley de protección de la persona frente al tratamiento de sus datos personales.

[República de Costa Rica, 2013] República de Costa Rica, P. (2013). Reglamento a la Ley de protección de la persona frente al tratamiento de sus datos personales.

[República de Costa Rica, 2015] República de Costa Rica, P. (2015). Declaratoria de interés público de la estragtegia nacional para un gobierno abierto.

[Rokach and Maimon, 2014] Rokach, L. and Maimon, O. (2014). *Data Mining with Decision Trees*, volume 81. World Scientific Publishing Co. Pte. Ltd.

[Romero-Tris and Megías, 2018] Romero-Tris, C. and Megías, D. (2018). Protecting Privacy in Trajectories with a User-Centric Approach. *ACM Transactions on Knowledge Discovery from Data*, 12(6):1–27.

[Samarati and Sweeney, 1998] Samarati, P. and Sweeney, L. (1998). Protecting Privacy when Disclosing Information: k-Anonymity and Its Enforcement through Generalization and Suppression. Technical report, SRI Computer Science Laboratory, Palo Alto.

[Scholl et al., 2008] Scholl, M. A., Stine, K. M., Hash, J., Bowen, P., Johnson, L. A., Smith, C. D., and Steinberg, D. I. (2008). Sp 800-66 rev. 1. an introductory resource guide for implementing the health insurance portability and accountability act (hipaa) security rule. Technical report, National Institute of Standards And Technology, Gaithersburg, MD, United States.

[Shah, 2012] Shah, A. (2012). State-of-art in Statistical Anonymization Techniques for Privacy Preserving Data Mining. *International Journal of Computer Science & Engineering Technology*, 3(7):258–262.

[Shen et al., 2013] Shen, Y., Guo, G., Wu, D., and Fan, Y. (2013). A novel algorithm of personalized-granular k-anonymity. *Proceedings - 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer, MEC 2013*, pages 1860–1866.

[Shklovski and Vertesi, 2013] Shklovski, I. and Vertesi, J. (2013). "Un-Googling" Publications: The Ethics and Problems of Anonymization. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, page 2169, New York, New York, USA. ACM Press.

[Shokri et al., 2017] Shokri, R., Stronati, M., Song, C., and Shmatikov, V. (2017). Membership Inference Attacks Against Machine Learning Models. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 3–18. IEEE.

[Shu et al., 2015] Shu, X., Yao, D., and Bertino, E. (2015). Privacy-preserving detection of sensitive data exposure. *IEEE Transactions on Information Forensics and Security*, 10(5):1092–1103.

[Solove, 2006] Solove, D. J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review*, 154(3):477–560.

[Song et al., 2018] Song, X., Wang, X., Nie, L., He, X., Chen, Z., and Liu, W. (2018). A Personal Privacy Preserving Framework. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR '18*, pages 295–304, New York, New York, USA. ACM Press.

[Sun et al., 2012] Sun, L., Wang, Y., and Wu, Y. (2012). A survey of transaction dada anonymous publication. In *Proceedings - 2012 IEEE Symposium on Robotics and Applications, ISRA 2012*, pages 239–243. IEEE.

[Sweeney, 2002a] Sweeney, L. (2002a). Achieving k-anonymity Privacy Protection using Generalization and Suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):571–588.

[Sweeney, 2002b] Sweeney, L. (2002b). k -ANONYMITY: A MODEL FOR PROTECTING PRIVACY. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570.

[Taylor, 2003] Taylor, H. (2003). Most People Are "Privacy Pragmatists" Who , While Concerned about Privacy , Will Sometimes Trade It Off for Other Benefits. *The Harris Poll*, pages 1–6.

[Tian et al., 2017] Tian, F., Zhang, S., Lu, L., Liu, H., and Gui, X. (2017). A Novel Personalized Differential Privacy Mechanism for Trajectory Data Publication. In *2017 International Conference on Networking and Network Applications (NaNA)*, pages 61–68. IEEE.

[Tian and Zhang, 2013] Tian, H. and Zhang, W. (2013). Privacy-Preserving Data Publishing Based on Utility Specification. In *2013 International Conference on Social Computing*, pages 114–121. IEEE.

[Tien, 2013] Tien, J. M. (2013). Big data: Unleashing information. *Journal of Systems Science and Systems Engineering*, 22(2):127–151.

[Troncoso Reigada, 2006] Troncoso Reigada, A. (2006). La publicación de datos de profesores y alumnos y la privacidad personal: acerca de la protección de datos personales en las Universidades. *Revista de Derecho Político*, 0(67).

[Ukil et al., 2012] Ukil, A., Bandyopadhyay, S., Joseph, J., Banahatti, V., and Lodha, S. (2012). Negotiation-based privacy preservation scheme in internet of things platform. In *Proceedings of the First International Conference on Security of Internet of Things - SecurIT '12*, pages 75–84, New York, New York, USA. ACM Press.

[U.S. Federal Trade Commission FTC, 2014] U.S. Federal Trade Commission FTC (2014). Data Brokers. A Call for Transparency and Accountability.

[Verma, 2013] Verma, J. (2013). *Data Analysis in Management with SPSS Software*. Springer India, India.

[Wagner and Eckhoff, 2018] Wagner, I. and Eckhoff, D. (2018). Technical Privacy Metrics. *ACM Computing Surveys*, 51(3):1–38.

[Wang et al., 2005] Wang, K., Fung, B. C., and Yu, P. S. (2005). Template-based privacy preservation in classification problems KeWang. In *Proceedings - IEEE International Conference on Data Mining, ICDM*, pages 466–473.

[Wang et al., 2016] Wang, K., Wang, P., Fu, A. W., and Wong, R. C.-W. (2016). Generalized bucketization scheme for flexible privacy settings. *Information Sciences*, 348:377–393.

[Warren and Brandeis, 1890] Warren, S. D. and Brandeis, L. D. (1890). The Right to Privacy. *Harv. L. Rev.*, 4(5):193–220.

[Wu et al., 2012] Wu, R., Ahn, G.-J., and Hu, H. (2012). Towards HIPAA-compliant healthcare systems. In *Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12*, page 593, New York, New York, USA. ACM Press.

[Xiao and Tao, 2006a] Xiao, X. and Tao, Y. (2006a). Anatomy: Simple and effective privacy preservation. *Proceedings of the 32nd international conference on Very Large Database, ACM*, pages 150, 139.

[Xiao and Tao, 2006b] Xiao, X. and Tao, Y. (2006b). Personalized privacy preservation. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data - SIGMOD '06*, page 229, New York, New York, USA. ACM Press.

[Xu et al., 2006] Xu, J., Wang, W., Pei, J., Wang, X., Shi, B., and Fu, A. W.-C. (2006). Utility-based anonymization using local recoding. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06*, page 785, New York, New York, USA. ACM Press.

[Xu et al., 2014a] Xu, L., Jiang, C., Wang, J., Yuan, J., and Ren, Y. (2014a). Information security in big data: Privacy and data mining. *IEEE Access*, 2:1151–1178.

[Xu et al., 2014b] Xu, Y., Ma, T., Tang, M., and Tian, W. (2014b). A survey of privacy preserving data publishing using generalization and suppression. *Applied Mathematics and Information Sciences*, 8(3):1103–1116.

[Xu et al., 2013] Xu, Y., Qin, X., Yang, Z., Yang, Y., and Li, K. (2013). A personalized k-anonymity privacy preserving method. *Journal of Information and Computational Science*, 10(1):139–155.

[Ye et al., 2008] Ye, X., Zhang, Y., and Liu, M. (2008). A Personalized (a,k)-Anonymity Model. In *2008 The Ninth International Conference on Web-Age Information Management*, pages 341–348. IEEE.

# A P É N D I C E S

# Apéndice A

# Encuesta de grado de privacidad

1. Indique la confianza que le tiene a las siguientes entidades a la hora de compartir información personal:

| | Ninguna | Poca | Suficiente | Mucha |
|---|---|---|---|---|
| Amigos y familia | O | O | O | O |
| Investigadores de la UCR | O | O | O | O |
| Entidades financieras (p.ej. bancos) | O | O | O | O |
| Medios de comunicación (p.ej. periódicos) | O | O | O | O |
| Redes sociales (p. ej. Facebook) | O | O | O | O |

2. Indique qué tan privada considera usted la siguiente información:

| | Nada privada | Poco privada | Medianamente privada | Bastante privada | Muy privada |
|---|---|---|---|---|---|
| **Información de horarios** | | | | | |
| Sigla(s) de curso(s) | O | O | O | O | O |
| Nombre(s) de curso(s) | O | O | O | O | O |
| Aula(s) de curso(s) | O | O | O | O | O |
| Horario(s) de curso(s) | O | O | O | O | O |
| **Información de educación secundaria** | | | | | |
| Nombre de colegio de procedencia | O | O | O | O | O |
| Dirección de colegio de procedencia | O | O | O | O | O |
| **Información de admisión** | | | | | |
| Año de admisión | O | O | O | O | O |
| Promedio de admisión | O | O | O | O | O |
| Mejor promedio de admisión | O | O | O | O | O |
| Año de mejor promedio de admisión | O | O | O | O | O |
| Carné universitario | O | O | O | O | O |
| Contraseña de matrícula WEB UCR | O | O | O | O | O |
| **Información geográfica** | | | | | |
| Provincia | O | O | O | O | O |
| Cantón | O | O | O | O | O |
| Distrito | O | O | O | O | O |
| Dirección exacta | O | O | O | O | O |

| | Nada privada | Poco privada | Medianamente privada | Bastante privada | Muy privada |
|---|---|---|---|---|---|
| **Información de carreras** | | | | | |
| Nombre de la(s) carrera(s) | O | O | O | O | O |
| Año de ingreso a la(s) carrera(s) | O | O | O | O | O |
| Recinto(s) | O | O | O | O | O |
| Área(s) | O | O | O | O | O |
| Facultad(es) | O | O | O | O | O |
| Escuela(s) | O | O | O | O | O |
| **Información socioeconómica** | | | | | |
| Tipo de beca | O | O | O | O | O |
| Beneficio de pago de reubicación geográfica | O | O | O | O | O |
| Beneficio de pago transporte | O | O | O | O | O |
| Beneficio de compra de libros | O | O | O | O | O |
| Ingreso mensual familiar | O | O | O | O | O |
| **Información de calificaciones** | | | | | |
| Sigla(s) de curso(s) | O | O | O | O | O |
| Nombre(s) de curso(s) | O | O | O | O | O |
| Nombre(s) del (de los) profesor(es) del (los) curso(s) | O | O | O | O | O |
| Calificación(es) del (los) curso(s) | O | O | O | O | O |
| Promedio ponderado semestral | O | O | O | O | O |
| Promedio ponderado general | O | O | O | O | O |
| **Información de títulos** | | | | | |
| Nombre(s) de título(s) | O | O | O | O | O |
| Año(s) de obtención | O | O | O | O | O |

| | Nada privada | Poco privada | Medianamente privada | Bastante privada | Muy privada |
|---|---|---|---|---|---|
| **Información financiera** | | | | | |
| Estado de morosidad con la Oficina de Administración Financiera de la UCR | O | O | O | O | O |
| Estado de morosidad con la biblioteca | O | O | O | O | O |
| Tipo de cobro [Nacional/Extranjero] | O | O | O | O | O |
| Número de tarjeta de cuenta bancaria | O | O | O | O | O |
| **Información personal** | | | | | |
| Edad | O | O | O | O | O |
| Nacionalidad | O | O | O | O | O |
| Estado civil | O | O | O | O | O |
| Género | O | O | O | O | O |
| Correo electrónico | O | O | O | O | O |

3. En pocas palabras, describa de qué se trata este estudio.

_____