DAV
Deutsche
Aktuarvereinigung e.V.

Ergebnisbericht des Ausschusses Actuarial Data Science

Anonymisierung im aktuariellen Umfeld

Köln, 04.07 2025

Präambel

Die Arbeitsgruppe Anonymisierung und Pseudonymisierung des Ausschuss Actuarial Data Science der Deutschen Aktuarvereinigung e. V. hat den vorliegenden Ergebnisbericht erstellt.[1]

Zusammenfassung

Der Ergebnisbericht behandelt Fragestellungen zu Verfahren der Anonymisierung und Pseudonymisierung und deren Anwendungen und Bewertung im Rahmen von aktuariellen Aufgabenstellungen.

Motiviert durch die rechtlichen Rahmenbedingungen und die Gefahren der Offenlegung von schützenswerten Daten, werden Werkzeuge zur Anonymisierung vorgestellt. Für die Anonymisierung von in der Versicherungswirtschaft meist tabellarischen Daten, stellt der Ergebnisbericht ein Vorgehensmodell vor. Neben den eigentlichen Verfahren zur Anonymisierung, ist es wichtig zu verstehen, wie diese die Anonymität und den damit einhergehenden Informationsverlust und Datennutzen beeinflussen. Für eine Veröffentlichung von Daten soll es potenziellen Angreifern nicht möglich sein, die zu schützenden Informationen offenzulegen. Als weitere Option zur Anonymisierung werden Konzepte zur Erzeugung von synthetischen Daten eingeführt. Neben den in dem Ergebnisbericht einfachen und nicht zusammenhängenden Beispielen, wird anhand von separaten Notebooks in der Statistiksoftware R und in Python die Anwendung vorgestellt.

Der Ergebnisbericht ist an die Mitglieder und Gremien der DAV zur Information über den Stand der Diskussion und die erzielten Erkenntnisse gerichtet und stellt keine berufsständisch legitimierte Position der DAV dar.[2]

Schlagworte

Ergebnisbericht, Anonymisierung, Pseudonymisierung

Verabschiedung

Dieser Ergebnisbericht ist durch den Ausschuss Actuarial Data Science am 04.07 2025 verabschiedet worden.

[1] Der Ausschuss dankt der Arbeitsgruppe Anonymisierung und Pseudonymisierung ausdrücklich für die geleistete Arbeit, namentlich Dr. Christoph Falkenau, Eva Odenkirchen, Sarah Hogekamp, Karten de Braaf Nora Valiente Bauer, Erwin Hetke, Dr. Felix Spangenberg und Dariush Sadeghi-Yam

[2] Die sachgemäße Anwendung des Ergebnisberichts erfordert aktuarielle Fachkenntnisse. Dieser Ergebnisbericht stellt deshalb keinen Ersatz für entsprechende professionelle aktuarielle Dienst-leistungen dar. Aktuarielle Entscheidungen mit Auswirkungen auf persönliche Vorsorge und Absicherung, Kapitalanlage oder geschäftliche Aktivitäten sollten ausschließlich auf Basis der Beurteilung durch eine(n) qualifizierte(n) Aktuar DAV/Aktuarin DAV getroffen werden.

This abstract summarises the report on findings “Ein Überblick zu Anonymisierung und Pseudonymisierung“ which was approved by the DAV committee Actuarial Data Science on 04.07.2025.

The Artificial Intelligence Act in an actuarial context

The report deals with questions relating to anonymization and pseudonymization procedures and their application and evaluation in the context of actuarial tasks.

Motivated by the legal framework and the risks of disclosing sensitive data, tools for anonymization are presented. The report presents a procedural model for the anonymization of data that is usually tabular in the insurance industry. In addition to the actual anonymization procedures, it is important to understand how these methods influence anonymity and the associated loss of information and data utility. When publishing data, it should not be possible for potential attackers to disclose the protected information. Concepts for generating synthetic data are introduced as a further option for anonymization. In addition to the simple and unrelated examples included in the report, separate notebooks in the statistical software R and in Python are presented.

Reports on findings are summaries of the results of work carried out by DAV committees or working groups,

- where their application can be freely decided upon within the framework of the code of conduct,
- that should inform discussion of the current opinion among actuaries or also among the broader public.

As working results of a single committee, they do not, for the time being, represent any recognised position within the DAV and do not comprise any actuarial standards of practice. In this respect they are clearly distinguishable from any standards of practice.

Inhaltsverzeichnis

1. Einleitung

Daten bilden unter anderem die Grundlage für die Entwicklung von Verfahren des Maschinellen Lernens (kurz: ML). Entscheidend ist es, möglichst umfangreiche und qualitativ hochwertige Informationen zu verwenden. Die Anwendungsfälle in der Versicherungswirtschaft kennzeichnen sich vor allem dadurch, dass ein Großteil dieser Daten sensible personenbezogene Daten umfassen. In der Personenversicherung sind dies beispielsweise Angaben zu Vorerkrankungen, die im Rahmen der Risikoprüfung erhoben werden. Im Spannungsfeld zwischen dem Bedarf zur Entwicklung von ML-Modellen und dem Schutz von Daten bietet es sich an, Techniken der Anonymisierung und Pseudonymisierung zur rechtssicheren Nutzung von personenbezogenen Daten anzuwenden. Ziel ist es, Daten nutzbar zu machen, Informationen nicht zu verlieren und gleichzeitig die Gefahr einer Offenlegung des Personenbezugs zu reduzieren.

Dieser Ergebnisbericht gibt einen Überblick zu dem Themengebiet der Anonymisierung im Kontext der aktuellen Gesetzgebung und definiert zentrale Begriffe. Ziel ist es, grundlegende Techniken aufzuzeigen, um Aktuaren einen theoretischen und praktischen Einblick in das Themenfeld zu geben. Sowohl der konkrete Einsatz von bestimmten Verfahren als auch der Grad der Anonymisierung variiert dabei je nach Anwendungsfall. So erfordert es, dass die handelnden Personen sowohl ein tiefes Verständnis der Daten, der Funktionsweise der Modelle und der Nutzung der Ergebnisse haben. Aufgrund der Relevanz für Aktuare werden hier beschriebene Verfahren in die CADS-Ausbildung aufgenommen.

Zunächst werden die rechtlichen Rahmenbedingungen (siehe Kapitel 2) aufgezeigt, allen voran die Datenschutzgrundverordnung und das Bundesdatenschutzgesetz. Die Definition von Begriffen (siehe Kapitel 2) ist Grundlage für die Einführung von Methoden zur Anonymisierung und Pseudonymisierung (siehe Kapitel 3) und von Metriken (siehe Kapitel 4). Die Notwendigkeit des Schutzes personenbezogener Daten wird durch die Gefahren von Angriffen (siehe Kapitel 2.4) motiviert. Eine Alternative zur Anonymisierung von Daten ist die Erzeugung synthetischer Daten (siehe Kapitel 7).

Zusätzlich sind über den GitHub-Account der Deutschen Aktuarvereinigung e.V. (DAV) verschiedene Notebooks zugänglich, die in Teilen die theoretischen Grundlagen dieses Ergebnisberichtes exemplarisch anwenden sollen.

Dieser Ergebnisbericht umfasst keinerlei rechtliche Empfehlungen. Es wird vielmehr empfohlen, die Anwendung der vorgestellten Verfahren gemeinsam mit Datenschutzbeauftragten und Juristen abzustimmen. Im Rahmen der Anonymisierung trifft man zum Teil auf Fragestellungen rund um das Thema Bias und Diskriminierung. Es sei darauf hingewiesen, dass dieses Thema aufgrund der Bedeutung und des Umfangs separat durch eine eigene Arbeitsgruppe innerhalb des Ausschusses Actuarial Data Science behandelt wird.

2. Allgemeine Definitionen und rechtliche Rahmenbedingungen

Im Folgenden werden allgemeine Definitionen vorgestellt, die für die Einführung späterer Konzepte benötigt werden. Zusätzlich wird das regulatorische Umfeld beleuchtet.

2.1. Personenbezogene Daten

Im Mittelpunkt stehen personenbezogene Daten, die vor unerlaubter Veröffentlichung zu schützen sind. So definiert die (Datenschutz-Grundverordnung) (kurz: DSGVO) in Artikel 4 Abs. 1 „personenbezogene Daten“ als alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen (EU-Regulation 2016/679, 2016). Dabei wird eine natürliche Person als identifizierbar angesehen, wenn sie direkt, z.B. über einen Namen, oder indirekt, z.B. über die Kombination von verschiedenen Informationen, wie Postleitzahl, Geburtsdatum oder IP-Adresse, identifiziert werden kann.

Personenbezogene Daten, die für die Pseudonymisierung und Anonymisierung relevant sind, sind beispielsweise: Name, Geburtsdatum, Geschlecht, Ausweisnummer, biometrische Daten, Patientendaten, Finanzdaten, demografische Daten. Die Datenschutzgrundverordnung beschränkt sich dabei auf natürliche Personen, womit der Schutz von juristischen Personen, wie Unternehmen, nicht eingeschlossen ist.

Die Datenschutzgrundverordnung unterscheidet in Artikel 9 zusätzlich besondere Kategorien personenbezogener Daten. So ist nach Art. 9 Abs. 1 DSGVO die Verarbeitung von personenbezogenen Daten, aus denen beispielsweise die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen hervorgehen, untersagt. Art. 9 Abs. 2 DSGVO nennt Ausnahmen wie die Einwilligung der betroffenen Person.

Für die Einführung von Techniken zur Anonymisierung (siehe Kapitel 3) und Metriken zur Messbarkeit von Anonymität (siehe Kapitel 4) ist die Unterteilung von personenbezogenen Daten in die folgenden drei Kategorien relevant (Dalenius, 1986):

- *(Direkte) Identifikatoren*: Attribute, die eine eindeutige Identifikation von Personen ermöglichen, wie z. B. Name oder Ausweisnummer.
- *Quasi-Identifikatoren*: Attribut oder Menge von Attributen, das für sich allein keine Identifizierung ermöglicht, aber durch Hinzunahme anderer Daten zur Identifizierung führen kann. Beispielsweise kann die Kombination aus Geschlecht, Wohnort und Geburtsdatum ausreichen, um eine bestimmte Person zu identifizieren. Datensätze mit identischen Quasi-Identifikatoren bilden sogenannte *Äquivalenzklassen*.
- *Sensible Attribute*: Datenkategorien oder Informationen, die besonders schützenswert sind, weil sie potenziell schädlich oder sensibel für die betroffene Person sein können, wenn sie kompromittiert oder unangemessen verwendet werden. Hierzu zählt beispielsweise die Information zu Gehalt oder Krankheiten.

2.2. Bestimmbarkeit

Um eine Person zu identifizieren, werden Daten herangezogen, die es möglich machen, sie zu bestimmen. Eine Person wird durch die direkten Identifikatoren bestimmt und durch Informationen aus den Quasi-Identifikatoren bestimmbar. Beispielsweise ist eine dynamische IP-Adresse für den Internetanbieter, der eine Zuordnung zum Klarnamen benötigt personenbezogen, für Dritte hat die die Adresse allerdings keinen Personenbezug. Im Allgemeinen werden zwei Arten von Bestimmbarkeit unterschieden:

- *Relative Bestimmbarkeit*: Für die Beurteilung, ob eine Person bestimmt werden kann, werden nur das Wissen und die technischen Mittel, welche der verantwortlichen, datenverarbeitenden Stelle normalerweise zur Verfügung stehen, herangezogen. Die Bestimmbarkeit ist auch dann nicht gegeben, wenn eine Person von der datenverarbeitenden Stelle nur mit unverhältnismäßig hohem Aufwand identifiziert werden kann. Damit können Daten für eine Stelle personenbezogen sein, für eine andere aber nicht.
- *Absolute Bestimmbarkeit*: Ist der Personenbezug allerdings für irgendeine Stelle identifizierbar, die über das zur Identifizierung erforderliche Zusatzwissen verfügt, spricht man von einer absoluten Bestimmbarkeit. Damit wird jedes Zusatzwissen eingeschlossen.

Welcher der beiden Ansätze angewendet werden sollte, ist in der rechtlichen Beurteilung nicht einheitlich und möglicherweise noch nicht final entschieden. So stärkt der Europäische Gerichtshof in verschiedenen Entscheidungen (z. B. in dem „Breyer“-Urteil (Vorlage zur Vorabentscheidung – Verarbeitung personenbezogener Daten – Richtlinie 95/46/EG – Art. 2 Buchst. a – Art. 7 Buchst. f – Begriff ‚personenbezogene Daten‘ – Internetprotokoll-Adressen – Speicherung durch einen Anbieter von Online-Mediendiensten –, 2016)) den absoluten Ansatz, wobei sich in einem jüngeren Urteil („FIN“-Urteil vom 9. November 2023 (Europäischer Gerichtshof, 2023)) vielleicht eine Lockerung des bisherigen Verständnisses andeutet (Keppeler, 2024). In der Rechtsliteratur sprechen sich verschiedene Autoren für den relativen Ansatz aus (zum Beispiel (Gola & Schomerus, 2007),

§ 3, Rn. 10 und Rn. 44; (Tinnefeld, 2003), Kap. 4.1 Rn. 22; (Roßnagel & Scholz, 2000), S. 723. (Gola & Schomerus, 2007) weisen darauf hin, dass die Unverhältnismäßigkeit des Aufwands zur Re-Identifikation letztlich eine Einzelfallentscheidung darstellt. Wenn sich das wirtschaftliche Interesse an der Bestimmbarkeit von Personen in einem Datensatz ändert, oder aufgrund neuer technischer Möglichkeiten die Kosten der Re-Identifikation sinken, dann kann sich die Einschätzung zur Bestimmbarkeit nach dem relativen Ansatz ändern. Es ist des Weiteren zu beachten, dass die genannten Literaturquellen und Urteile zum Teil vor der Verabschiedung der neuesten Fassung des BDSG und der DSGVO veröffentlicht wurden. Für eine abschließende Beurteilung in einem konkret vorliegenden Fall sollte daher immer die dann gültige Rechtslage beachtet werden.

2.3. Anonymisierung und Pseudonymisierung

Der Europäische Datenschutzbeauftragte hat gemeinsam mit der spanischen Datenschutzbehörde 2021 zehn Missverständnisse rund um Anonymisierung vorgestellt, darunter auch das Gleichsetzen von Pseudonymisierung und Anonymisierung (AEPD (Agencia Espanola Proteccion Datos), 2021), S. 3). Beide Begriffe behandeln zwar die Verarbeitung von personenbezogenen Daten, sind aber nicht identisch.

Artikel 4 Nr. 5 der DSGVO (Datenschutz-Grundverordnung) definiert Pseudonymisierung als „die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden“.

Eine Definition von Anonymisierung ist nicht in der Datenschutzgrundverordnung verankert. Hier hilft der Erwägungsgrund 26 (Datenschutz-Grundverordnung), der angibt, dass die DSGVO keine Anwendung auf anonyme Daten hat. Damit sind Informationen gemeint, die sich „nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen oder personenbezogene Daten, die in einer Weise anonymisiert worden sind, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann“. Anonyme Daten hingegen können nicht bestimmten Personen zugeordnet werden. Sobald Daten anonym sind und Personen nicht mehr identifizierbar sind, fallen diese Daten nicht mehr unter den Geltungsbereich der DSGVO ((AEPD (Agencia Espanola Proteccion Datos), 2021), S. 3).

Es wird deutlich, dass beide Begriffe die Verarbeitung von personenbezogenen Daten behandeln, mit dem Unterschied, dass nach einer Anonymisierung kein Personenbezug mehr hergestellt werden kann, wohingegen nach Pseudonymisierung durch Hinzubinden zusätzlicher Informationen dies noch möglich ist.

Ob Daten tatsächlich anonym sind, hängt unter Umständen von den Informationen und Möglichkeiten Dritter ab. So gibt Art. 32 DSGVO (Datenschutz-Grundverordnung) Nr. 1 Anforderungen an die Pseudonymisierung und Verschlüsselung personenbezogener Daten an. Der Stand der Technik und die Implementierungskosten müssen genauso berücksichtigt werden wie die Art, der Umfang, die Umstände und die Zwecke der Verarbeitung. Zudem sind die unterschiedliche Eintrittswahrscheinlichkeit und die Schwere des Risikos für die Rechte und Freiheiten natürlicher Personen einzubeziehen. Die Wirksamkeit dieser technischen und organisatorischen Maßnahmen zur Gewährleistung der Sicherheit der Verarbeitung muss regelmäßigt überprüft, bewertet und evaluiert werden.

Die DSGVO findet zwar keine Anwendung auf anonymisierte Daten, die Anonymisierung von Daten an sich stellt aber aus Sicht der deutschen Datenschutzbeauftragten bereits eine Datenverarbeitung im Sinne der DSGVO dar ((BfDI, 2020), Pkt. 3., S. 6). Art. 4 Nr. 2 der DSGVO (Datenschutz-Grundverordnung) listet die Anonymisierung nicht explizit als Verarbeitung im Zusammenhang mit personenbezogenen Daten auf, dennoch ist davon auszugehen, dass die Auf-

zählung der Vorgänge nicht abschließend anzusehen ist. So werden unter anderem die Anpassung oder Veränderung, das Auslesen oder Abfragen in Art. 4 Nr. 2 DSGVO (Datenschutz-Grundverordnung) genannt.

Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (kurz: BfDI) stellte 2020 klar, dass die Anonymisierung eine Form der Verarbeitung darstellt und einer Rechtsgrundlage bedarf ((BfDI, 2020), S.5). Es ist von den jeweiligen Umständen des Einzelfalls abhängig, welcher Absatz der DSGVO (insbesondere Art. 6 und 17) für eine Anonymisierung herangezogen werden kann. Welche Rechtsgrundlage im konkreten Fall Anwendung finden kann, ist mit den Datenschutzbeauftragten zu klären.

Der Gesamtverband der Deutschen Versicherungswirtschaft hält die Schaffung einer eindeutigen Rechtsgrundlage für die Anonymisierung und Pseudonymisierung für nötig - insbesondere für besondere Kategorien personenbezogener Daten (vergleiche (GDV, 2023), S.3).

Datenschutz-Folgenabschätzung (Vergleiche (BfDI, 2020), S.11)

Im „Positionspapier zur Anonymisierung unter der DSGVO unter besonderer Berücksichtigung der TK-Branche“ stellte der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (kurz: BfDI) 2020 ((BfDI, 2020), S.11) klar, dass vor einer Anonymisierung eine Datenschutz-Folgenabschätzung gemäß Art. 35 Abs. 1 DSGVO (Datenschutz-Grundverordnung) durchzuführen ist. Dies ist dadurch begründet, dass wie in Art. 35 Abs. 1 DSGVO (Datenschutz-Grundverordnung) für Verarbeitungen, die voraussichtlich ein hohes Risiko für die Rechte und Freiheiten natürlicher Personen unter der Erfüllung bestimmter Kriterien haben, eine Datenschutz-Folgeabschätzung notwendig ist. Man kann davon ausgehen, dass zwei dieser Kriterien, die Verarbeitung im großen Umfang und die Verwendung neuer Technologien, bei der Anonymisierung erfüllt sind. Der BfDI stellt zudem klar, dass die Verantwortlichen zusätzlich die Folgen einer möglichen „De-Anonymisierung“ berücksichtigen müssen.

2.4. Angriffsszenarien

Die Anonymisierung von Daten verfolgt das Ziel, personenbezogene Informationen zu schützen und die Identifizierbarkeit natürlicher Personen auszuschließen. Bei der Bewertung der Anonymisierungsmaßnahmen sollten mögliche Mittel, die Dritte zur Re-Identifikation einsetzen könnten, sowie die damit verbundenen Kosten und der erforderliche Zeitaufwand berücksichtigt werden. Zudem ist es entscheidend, die zum Zeitpunkt der Verarbeitung verfügbare Technologie sowie künftige technologische Entwicklungen einzubeziehen, um einen angemessenen und langfristigen Schutz der Daten sicherzustellen.

Um die Robustheit einer Anonymisierungstechnik zu bestimmen und festzustellen, ob eine Anonymisierung tatsächlich stattgefunden hat, verdienen folgende drei Risikokriterien besondere Aufmerksamkeit; nach (Weitzenboeck, Lison, Cyndecka, & Langford, 2022) und (Giomi, Boenisch, Wehmeyer, & Tasnádi, 2022):

- *Singling-out* bezeichnet die Möglichkeit, einzelne oder mehrere Datensätze zu isolieren, die eine Person im Datensatz identifizieren können. Es tritt auf, wenn es möglich ist, im ursprünglichen Datensatz einen Datensatz mit einer einzigartigen Kombination von einem oder mehreren Attributen zu identifizieren. Dabei ist besonders hervorzuheben, dass Singling-out nicht zwangsläufig eine Re-Identifikation bedeutet. Die Fähigkeit, eine Person zu isolieren, kann jedoch ausreichen, um weitere Angriffe auf die Privatsphäre einzuleiten.
- *Linkability* bezeichnet die Möglichkeit, zwei oder mehr Datensätze miteinander zu verknüpfen, die sich auf dieselbe betroffene Person oder eine Gruppe von Personen beziehen, sei es innerhalb derselben oder unterschiedlicher Datenquellen. Wenn ein Angreifer beispielsweise durch Korrelationsanalysen feststellen kann, dass zwei Datensätze derselben Gruppe von Individuen zugeordnet sind, ohne einzelne Personen isolieren zu können, schützt die angewendete Technik zwar vor Singling-out, jedoch nicht gegen Linkability. Da zwischen den generierten Daten und den Originaldaten oft statistische Ähnlichkeiten bestehen, können Linkability-Risiken auch bei synthetischen Datensätzen auftreten.

- *Inference* bezeichnet die Möglichkeit, mit hoher Wahrscheinlichkeit den Wert eines bestimmten Attributs aus den Werten einer Reihe anderer Attribute abzuleiten.

Die Bedeutung diese Risiken für die Anonymisierung ergibt sich aus ihrer Auswirkung auf die Privatsphäre der Individuen: *Singling out* kann einerseits als eine Möglichkeit betrachtet werden, eine Person indirekt in einem Datensatz zu identifizieren und gleichzeitig kann es als Ausgangspunkt für *Linkability-Angriffe* dienen, die zu Re-Identifizierung von Datensätzen führen. Durch *Inference-Angriffe* können hochsensible Informationen über Einzelpersonen offengelegt werden.

Bei der Messung von Datenschutzrisiken ist eine wichtige Unterscheidung zwischen dem, was ein Angreifer auf Bevölkerungsebene (allgemeine Informationen) und auf individueller Ebene (spezifische Informationen) lernen kann, notwendig. Allgemeine Informationen sind das, was den anonymisierten Daten Nutzen verleiht; spezifische Informationen ermöglichen es dem Angreifer, die Privatsphäre einzelner Personen zu verletzen.

Ein Angreifer kann unterschiedliche Ziele verfolgen, wenn er versucht, anonymisierte Daten in einem Datensatz aufzudecken oder zu identifizieren. In den ersten beiden der folgend vorgestellten Szenarien versucht der Angreifer, eine bestimmte Person in einer Datenquelle zu re-identifizieren. Im dritten Szenario ist hingegen das Ziel, so viele Personen wie möglich zu re-identifizieren (nach (Grace, 2016)):

- *Prosecutor Attack*: In diesem Szenario will der Angreifer eine bestimmte Person re-identifizieren, von der er weiß, dass ihre Daten irgendwo im Datensatz vorhanden sind. Der Eindringling hat Hintergrundinformationen über diese Person und nutzt diese, um in der offengelegten Datenbank nach einem passenden Datensatz zu suchen und möglicherweise zusätzliche Informationen zu erfahren. Dieses Risiko ist besonders zu berücksichtigen, wenn der offengelegte Datensatz eine Bevölkerungsgruppe oder ein Teil davon ist (z. B. alle Mitarbeiter/Führungskräfte einer Firma).
- *Journalist Attack*: Ein Angriff, der darauf abzielt, durch den Zugriff auf andere Informationsquellen über eine oder mehrere Personen, diese zu re-identifizieren. Dies ähnelt der Prosecutor Attack, da sie auf einen einzelnen Datensatz abzielt. Im Unterschied dazu weiß der Journalist jedoch nicht sicher, ob eine bestimmte Person im Datensatz enthalten ist.
- *Marketer Attack*: Ein Angriff, der darauf abzielt, so viele Personen wie möglich aus den Daten zu re-identifizieren, auch wenn dies bedeutet, dass einige von ihnen fälschlicherweise identifiziert werden. Hier betrifft das Risiko alle Personen im Datensatz.

Das Ziel des Angreifers könnte lediglich darin bestehen, nachzuweisen, dass die Zielperson im Datensatz enthalten ist, ohne Zusatzangaben zu gewinnen (Powar, 2023).

Für die Bewertung von Anonymisierungsverfahren ist es von zentraler Bedeutung, spezifische Arten von Risiken und Angriffsszenarien zu berücksichtigen. In der Anwendung hängt die Durchführung der Angriffe im Detail stark von der Struktur der Daten und der Wahl des Anonymisierungsverfahrens ab.

(Vassilev, Oprea, Fordyce, & Anderson, 2024)) beschreiben konkrete technische Durchführungen von Angriffsszenarien für sowohl die Veröffentlichung von Datensätzen als auch von Modellergebnissen.

2.5. Die Europäische Datenstrategie

Dass Anonymisierung und Pseudonymisierung in Europa in Zukunft eine stärkere Rolle einnehmen werden, zeigen auch die Gesetzesinitiativen, die im Rahmen der europäischen Datenstrategie bereits in Kraft getreten sind oder aktuell diskutiert werden. Ein Beispiel hierfür ist der (Data Governance Act, 2022). In Art. 5 Nr. 3 a) i) des Data Governance Act wird öffentlichen Stellen die Möglichkeit eingeräumt, bei der Weiterverwendung personenbezogener Daten zu verlangen, dass der Zugang nur gewährt wird, wenn die Daten zuvor anonymisiert wurden. Dies verdeutlicht den

hohen Stellenwert, den die Anonymisierung für den Datenschutz im europäischen Rechtsrahmen einnimmt.

Ein weiteres zentrales Element der europäischen Datenregulierung ist der AI Act, der die Festlegung harmonisierter Vorschriften für KI-Systeme regelt. Der AI Act führt je nach Risikoklassifizierung der KI-Systeme spezifische Anforderungen ein. Da KI-Systeme oft große Mengen personenbezogener Daten verarbeiten, ist es notwendig, dass der AI Act Hand-In Hand mit der DSGVO arbeitet. Im AI Act wird ausnahmsweise, unter bestimmten Voraussetzungen Hochrisiko-KI-Systemen erlaubt, besondere Kategorien personenbezogener Daten zu verarbeiten, um die Erkennung und Korrektur von Verzerrungen sicherzustellen ((AI Act, 2024) Art. 10 Nr. 5).

Diese Beispiele aus dem Data Governance Act und dem AI Act verdeutlichen, dass die Anonymisierung und Pseudonymisierung zentral für den Schutz personenbezogener Daten im Rahmen der europäischen Datenstrategie sind. Sie tragen dazu bei, den Datenschutz zu gewährleisten, während gleichzeitig die Nutzung von Daten in verschiedenen Bereichen – wie Forschung, KI-Entwicklung oder öffentliche Verwaltung – gefördert wird.

2.6. Ergebnisse der „Data Protection Working Party (WP216)“

Eine, wenn auch aus Sicht der Gesetzgebung nicht aktuelle Quelle zu Anonymisierung, stellt die „Stellungnahme der Art. 29-Datenschutzgruppe von 5/2014 zu Anonymisierungstechniken dar (Artikel-29-Datenschutzgruppe / 0829/14/DE WP216, 2014)“. Diese analysiert die Wirksamkeit und Grenzen von bestimmten Anonymisierungstechniken. Es werden Empfehlungen für den Umgang mit verschiedenen Techniken ausgesprochen und das damit einhergehende Risiko einer Identifizierung berücksichtigt. Es gilt zu berücksichtigen, dass die Veröffentlichung vor Inkrafttreten der DSGVO erschienen ist.

Zunächst werden in der Ausarbeitung verschiedene Definitionen von Anonymisierung untersucht und inhaltlich verglichen. Dazu werden die folgenden Quellen verwendet:

- Richtlinie 95/46/EG Absatz 26, (Richtlinie 95/46/EG Absatz 26, 1995)
- Richtlinie 2002/58/EG (Richtlinie 2002/58/EG , 2002) (soll zukünftig ersetzt werden durch die E-Privacy Verordnung),
- Internationalen Normen (ISO).

Der Anonymisierungsbegriff stimmt im Wesentlichen in allen drei Quellen überein. Auffällig ist, dass keine der Quellen vorschreibt, mit welchen Methoden Daten erfolgreich anonymisiert werden und wie die Anonymität überprüft werden kann. Stattdessen werden Eigenschaften festgelegt, die von erfolgreichen Anonymisierungsmethoden und ihren Ergebnissen erfüllt werden sollten:

- Anonymisierung sollte nach dem derzeitigen Stand der Technik dauerhaft sein, ebenso wie die Löschung personenbezogener Daten.
- Anonymisierung soll die Identifizierbarkeit von Personen dauerhaft verhindern, indem sie alle Beteiligten vor den Risiken aus *Singling out*, *Linkability* und *Inference* bewahrt.

Darüber hinaus werden verschiedene Anonymisierungstechniken auf die Einhaltung der vorgestellten Risiken hin untersucht. Das Ergebnis ist, dass keine der Methoden, einzeln angewandt, zuverlässig die oben aufgelisteten Risiken verhindert. Wenn überhaupt, muss ein Kombinationsansatz der unterschiedlichen Methoden verwendet werden, doch dieser ist abhängig vom verwendeten Datensatz.

3. Techniken der Anonymisierung

Im Grunde geht es bei der Anonymisierung von Daten darum, die Attribute zu bestimmen, die es potenziell ermöglichen, eine Person zu identifizieren und die Daten so zu verändern, dass das Risiko einer Identifikation ausreichend reduziert wird. In diesem Kapitel werden Techniken zur Anonymisierung vorgestellt. Die Anwendung einer oder mehrerer Verfahren hängt vom Datentyp ab

und ist fallbezogen zu bewerten. Insbesondere kann es notwendig sein, eine Kombination von mehreren Methoden anzuwenden, um das erwünschte Ergebnis zu erhalten Die hier vorgestellten Techniken fokussieren sich auf tabellarische Daten, wovon ausgewählte in Code-Beispielen vorgestellt werden (siehe Hinweis in Kapitel 1).

3.1. Überblick

Vor Anwendung von Anonymisierungstechniken ist es notwendig, die direkten Identifikatoren zu entfernen. Anschließend können verschiedene Anonymisierungstechniken angewendet werden. Diese Verfahren können in zwei Kategorien aufgeteilt werden (Domingo-Ferrer & Torra, 2001):

- *Nicht-Perturbative* Methoden (siehe Kapitel 3.2),
- *Perturbative* Methoden (siehe Kapitel 3.3).

Bei pertubativen Methoden werden die individuellen Dateneinträge verzerrt, wodurch Kombinationen von Werten im Original-Datensatz verschwinden können und neue einzigartige Kombinationen entstehen können. Nicht-pertubative Methoden verändern die Daten nicht, sondern unterdrücken nur Teile der Information oder reduzieren das Detailniveau der Merkmale.

3.2. Nicht-Perturbative Methoden

Die am häufigsten genutzten Methoden, um das Detailniveau von Daten zu reduzieren sind: Generalisierung und Suppression.

3.2.1. Generalisierung

Eine *Generalisierung* verändert Werte der Klasse „Quasi-Identifikatoren“, um ihre Genauigkeit zu reduzieren (Fredj, Lammari, & Comyn-Wattiau, 2015). Kategorielle Werte können anhand einer Taxonomie durch allgemeinere Werte ersetzt werden. Bei numerischen Attributen werden exakte Angaben gerundet oder durch Intervalle oder Label ersetzt (Benschop, Machingauta, & Welch, 2022).

3.2.2. Suppression

Die *Suppression* unterdrückt bestimmte Werte, bzw. ersetzt sie durch ein spezielles Zeichen (z. B. Asterix *). Dadurch wird angezeigt, dass der zugrunde liegende Wert bewusst unterdrückt wird, im Gegensatz zu einem fehlenden Wert, der schon im Originaldatensatz unbekannt ist (Fung, Wang, Chen, & Yu, 2010).

Mittels Suppression können unter anderem Ausreißer entfernt werden, so dass die anschließende Generalisierung zu geringerem Informationsverlust führt.

- *Record Suppression* unterdrückt komplette Einträge mit all ihren Merkmalen. Dies eignet sich für Einträge mit mehreren seltenen oder ungewöhnlichen Werten. Record Suppression ist operational einfach durchzuführen, birgt aber ein größeres Risiko für Bias.
- *Value Suppression* unterdrückt alle Instanzen eines seltenen Wertes. Dies bewahrt mehr Information als Record Suppression und wird oft mit Generalisierung kombiniert.
- *Cell Suppression* (oder *Local Suppression*) entfernt nur einzelne Instanzen eines Wertes in spezifischen Teilmengen. Dies ermöglicht maximalen Datenerhalt, erfordert aber eine detailliertere Analyse.

3.3. Perturbative Methoden

Unter *Perturbation* werden Methoden zusammengefasst, die die originalen Werte durch synthetische Daten ersetzen, sodass die statistischen Eigenschaften der perturbierten Daten sich nicht signifikant von denen der originalen unterscheiden. Dies kann beispielsweise durch Vertauschen von Werten, Hinzufügen von Rauschen oder durch die Generierung von synthetischen Daten erfolgen (Xu, Jiang, Wang, Yuan, & Ren, 2014).

Die folgende Tabelle stellt dar, welche perturbativen Methoden für welchen Datentyp am besten geeignet sind.

Methode	Numerische Daten	Kategorielle Daten
Random Noise	✓	
PRAM		✓
Data Swapping	✓	✓
Microaggregation	✓	
Shuffling	✓	✓

Tabelle 1 Pertubative Methoden zur Anonymisierung nach Art der zu anonymisierenden Variablen

Im Folgenden werden die genannten Methoden näher erläutert.

3.3.1. Random Noise

Durch Hinzufügen von Rauschen zu den Attributen sollen die originalen Werte maskiert werden. Dies kann entweder durch Addition oder Multiplikation von Werten zu dem zu schützenden Attribut erreicht werden. Diese Methode wird hauptsächlich für numerische Attribute angewendet. Für kategorielle Variablen existiert die Methode „*Post RAndomisation Method (PRAM)*“ (siehe Kapitel 3.3.2). Die gängigsten Konzepte lauten *Additive Noise* und *Multiplicative Noise*. Siehe Für nähere Erläuterungen siehe beispielsweise (Domingo-Ferrer & Torra, 2001) oder (Aggarwal, 2008).

3.3.2. PRAM

Durch Anwendung von *PRAM* werden mithilfe von definierten Übergangswahrscheinlichkeiten kategorielle Variablen randomisiert, d. h. einzelne Ausprägungen in andere transformiert.

Eine Modifikation von *PRAM* ist das invariante *PRAM* (wie beispielsweise in (Ronning & Gnoss, 2003, S. 81)), bei dem die Randverteilungen erhalten bleiben, und somit der Erwartungswert der anonymisierten Variablen, dem der ursprünglichen Variablen entspricht.

Der Informationsverlust und die Risiken zum der Re-Identifikation hängen stark von der Wahl der Übergangswahrscheinlichkeiten (De Wolf, Gouweleeuw, Kooiman, & Willenborg, 1999) ab. Probleme bei der Anwendung von *PRAM* gibt es u.a. (Gouweleeuw, Kooiman, Willenborg, & De Wolf, 1997) bei Abhängigkeiten von Variablen in den originalen Daten, wenn *PRAM* unabhängig auf einzelne Variablen, beispielsweise auf den Wohnort und die Postleitzahl, angewendet wird. Zudem liegt hier die Schwierigkeit in der geeigneten Wahl der Übergangswahrscheinlichkeiten.

3.3.3. Data Swapping

Beim *Data Swapping* werden bestimmte Werte der sensiblen Attribute unter den Einträgen getauscht (Ciriani, 2007, S. 16-17). Die Paare für den Tausch werden dabei anhand von wohldefinierten Kriterien ausgesucht. Diese Technik ist auch für kategorielle Attribute geeignet. Allerdings hat sie den Nachteil, dass gewisse statistischen Eigenschaften der originalen Daten nicht erhalten bleiben.

Das Verfahren kann für stetige und kategorielle Variablen (mit einem Ordnungsverhältnis) verbessert werden, indem Attributwerte nur mit Werten eines ähnlichen Ranges getauscht werden (sogenanntes *Rank Swapping*). Dabei wird der Datensatz anhand der Ausprägungen eines Attributes aufsteigend sortiert und jeder Wert so mit einem anderen Wert vertauscht, der in einem bestimmten (Rang-)Abstand liegt.

3.3.4. Microaggregation

Im Rahmen der *Microaggregation* werden die Datensätze in kleinere Gruppen unterteilt (Domingo-Ferrer & Torra, 2001). Ein Attribut wird maskiert, indem der originale Wert durch den Mittelwert

oder Median der zugehörigen Gruppe ersetzt wird. Für die Gruppengröße wird ein Schwellenwert gewählt. Um den Informationsverlust zu minimieren, sollten die einzelnen Gruppen möglichst homogen sein. Die Gruppengröße kann für alle Gruppen identisch oder variabel sein. Eine Variabilität verstärkt dabei die Möglichkeit, eine Homogenität innerhalb der Gruppen zu erreichen. Damit stellt die *Microaggregation* ein Optimierungsproblem dar, bei dem die Homogenität der Gruppen maximiert werden soll, während die Gruppengröße mindestens einem Schwellenwert entspricht.

Es existieren univariate Methoden, die jede Variable sequenziell und unabhängig betrachten. Dieses Vorgehen wird auch als *Individual Ranking* oder *Blurring* bezeichnet (Defays & Nanopoulos, 1993). Multivariate Methoden bedienen sich beispielsweise der *Principal Component Analysis*.

3.3.5. Shuffling

Eng verwandt mit *Swapping* ist das sogenannte *Shuffling* (Muralidhar & Sarathy, 2006). Beim *Shuffling* wird den Datensätzen ein Rang auf Basis der originalen zu schützenden Variablen zugeordnet. Anschließend wird mit Hilfe der anderen Variablen ein Regressionsmodell auf die zu anonymisierende Variable angewandt. Mit Hilfe des Regressionsmodells werden dann für die zu schützende Variable Werte vorhergesagt, denen ebenfalls ein Rang zugeordnet wird. Auf Basis dieses Rankings werden die originalen Ausprägungen durch die neuen Werte ersetzt. Dies bedeutet, dass alle originalen Werte im Datensatz erhalten bleiben.

3.3.6. Weitere Methoden

Zusätzlich zu den oben genannten Methoden existieren weitere Möglichkeiten zur Anonymisierung von Daten. Hierzu zählt unter anderem Resampling (Ciriani, 2007), Data Data distrot by probability distribution (Domingo-Ferrer & Torra, 2001), MASSC (Singh, Yu, & Dunteman, 2004) oder Camouflage (Gopal, Goes, & Garfinkel, 1999).

4. Metriken der Anonymisierung

4.1. Motivation

Es gibt verschiedene Metriken zur Bestimmung des Grades der Anonymisierung. Welche Metrik geeignet ist, hängt vom Anwendungsfall und den konkreten Angriffsszenarien ab (siehe Kapitel 2.4), die es durch die Anonymisierung der Daten abzuwenden gilt. Im Folgenden stellen wir drei „klassische“ Ansätze vor (*k-Anonymität*, *l-Diversität* und *t-Ähnlichkeit*).

4.2. k-Anonymität

Definition

Unter *k-Anonymität* versteht man eine Definition des Anonymisierungsgrades eines Datensatzes. Ein Datensatz wird als k-anonym bezeichnet, wenn die Werte der Quasi-Identifikatoren eines jeden Individuums in diesem Datensatz mit den Werten der Quasi-Identifikatoren von mindestens *k-1* anderen Individuen im gleichen Datensatz übereinstimmen, sprich wenn jede Wertekombination jener Quasi-Identifikatoren in der Tabelle mindestens k-mal auftritt. Die Daten innerhalb eines Datensatzes, deren Quasi-Identifikatoren dieselben Werte annehmen, werden in einer Äquivalenzklasse zusammengefasst.

Fallbeispiel

Angenommen, die 5-stellige PLZ, das Alter und der Beruf seien Quasi-Identifikatoren.

Quasi-Identifikatoren			Sensitives Attribut
PLZ	Alter	Beruf	Krankheit
13053	28	Kaufmann/-frau	Herzkrankheit
13068	29	Lehrer/in	Herzkrankheit
13068	21	Arzt/Ärztin	Virusinfektion
13053	23	Lehrer/in	Virusinfektion
14853	50	Mechaniker/in	Krebs
14853	55	Kaufmann/-frau	Herzkrankheit
14850	47	Lehrer/in	Virusinfektion
14850	49	Lehrer/in	Virusinfektion
13053	31	Lehrer/in	Krebs
13053	37	Mechaniker/in	Krebs
13068	36	Arzt/Ärztin	Krebs
13068	35	Lehrer/in	Krebs

Quasi-Identifikatoren			Sensitives Attribut
PLZ	Alter	Beruf	Krankheit
1 Äquivalenzklasse			
130**	<30	*	Herzkrankheit
130**	<30	*	Herzkrankheit
130**	<30	*	Virusinfektion
130**	<30	*	Virusinfektion
2 Äquivalenzklasse			
1485*	>40	*	Krebs
1485*	>40	*	Herzkrankheit
1485*	>40	*	Virusinfektion
1485*	>40	*	Virusinfektion
3 Äquivalenzklasse			
130**	3*	*	Krebs
130**	3*	*	Krebs
130**	3*	*	Krebs
130**	3*	*	Krebs

Wenn nun an gewissen Stellen die Werte der Quasi-Identifikatoren (gekennzeichnet mit einem *) unterdrückt oder verrauscht werden, erhält man einen Datensatz, welcher der 4-Anonymität genügt und 3 Äquivalenzklassen besitzt.

Schwachstellen

Unsortiertes Matching bzw. *komplementäre Veröffentlichung:* Diese Problematik wird in (Sweeney, 2002) beschrieben und drückt Folgendes aus. Es werden zu einem Datensatz mehrere verschiedene k-Anonyme Versionen veröffentlicht.

1) Sortierte Reihenfolge: In all diesen Versionen liegt exakt dieselbe Reihenfolge der Daten/Individuen vor. Durch dieses Wissen ist es möglich, die Werte aller Quasi-Identifikatoren zu einem Individuum zu bestimmen.
2) Zeitlich versetzte Veröffentlichung: Sei ein Datensatz zum Zeitpunkt T_0 gegeben und es werde eine k-Anonyme Version veröffentlicht. Diesem Datensatz werden nun im Laufe der Zeit neue Daten hinzugefügt, alte gelöscht, bestehende erweitert etc., und zum Zeitpunkt T_t wird wieder eine k-Anonyme Version veröffentlicht.

In beiden Fällen kann es durch eine Verknüpfung dieser Versionen passieren, dass zu manchen Individuen alle Werte der Quasi-Identifikatoren bekannt sind und somit die Anonymität nicht mehr gewährleistet ist.

Diesem Problem kann zum Teil entgegengewirkt werden, indem

1) Tabellen zufällig sortiert werden,
2) neue Tabellen mit bereits veröffentlichten Versionen verglichen werden,
3) jede Tabelle nur unter einer k-Anonymisierung veröffentlicht wird.

Weiterhin müssen wir für das Konzept der k-Anonymität, wie in (Sweeney, 2002) beschrieben, annehmen, dass der Datenhalter alle Quasi-Identifikatoren kennt. Ansonsten funktioniert dieses Konzept nicht. Hierfür ist es insbesondere notwendig, dass er alle öffentlich zugänglichen Daten kennt, die mit den von ihm gehaltenen Daten in Interaktion stehen und so das Risiko von Re-Identifizierung erhöhen könnten. Es stellt sich daher die Frage, wie realistisch diese Annahme ist.

Geringe Vielfalt der sensiblen Attribute (Homogenität): Diese Problematik wird in (Machanavajjhala, Kiefer, Gehrke, & Venkitasubramaniam, 2007) beschrieben und drückt aus, dass es innerhalb einer Äquivalenzklasse keine oder nur wenige Unterschiede in den Ausprägungen der sensiblen Attribute gibt. Wenn in diesem Fall die Zugehörigkeit eines Individuums zu einer solchen Äquivalenzklasse bekannt ist, dann kennt man auch den Wert des sensiblen Attributes

dieses Individuums. Wir betrachten hierzu folgendes Beispiel angelegt an (Machanavajjhala, Kiefer, Gehrke, & Venkitasubramaniam, 2007):

Man betrachte die 3. Äquivalenzklasse aus obiger Tabelle y. Darin trete folgender Fall auf:

1) Eine Person, Emma, wohnt in einem Ort mit der PLZ 130** beobachtet, dass ihr Nachbar, Bernd, von einem Krankenwagen abgeholt und in das Krankenhaus gebracht wird, dass die obige Tabelle veröffentlicht hat.
2) Emma kennt das Alter von Bernd, das 35 Jahre beträgt. Folglich weiß Emma, dass Bernd in der 3. Äquivalenzklasse auftaucht.
3) Da jedes Individuum aus der 3. Äquivalenzklasse an Krebs leidet, erfährt Emma, dass Bernd Krebs hat.

Angriff durch Hintergrundwissen: Ebenfalls in (Machanavajjhala, Kiefer, Gehrke, & Venkitasubramaniam, 2007) beschrieben, drückt dieses Problem aus, dass es einem Angreifer mittels zusätzlichem Hintergrundwissen möglich ist, ein Individuum einer Äquivalenzklasse und dort einem sensiblen Attribut zuzuordnen (siehe auch Kapitel 2.4). Die *k-Anonymität* schützt nicht gegen solche Angriffe. Dies wird an folgendem Beispiel angelehnt an (Machanavajjhala, Kiefer, Gehrke, & Venkitasubramaniam, 2007) deutlicher:

Wir betrachten wieder unsere obige Tabelle und nehmen an, dass Emma eine japanische Brieffreundin Yumiko hat. Emma liest nun in einem Brief von Yumiko, dass sie in unser Beispielkrankenhaus eingeliefert wurde, das die obige Tabelle veröffentlicht hat. Yumiko hat Emma ihre Krankheit nicht mitgeteilt.

- Da Emma allerdings aufgrund ihrer Brieffreundschaft Yumikos fünfstellige PLZ und ihr Alter kennt, weiß sie, dass Yumiko in der 1. Äquivalenzklasse auftaucht.
- Da Emma zudem weiß, dass Japaner*innen statistisch gesehen sehr selten eine Herzkrankheit erleiden, kann sie also nun aus den veröffentlichten Daten und ihrem Hintergrundwissen schließen, dass Yumiko mit großer Wahrscheinlichkeit an einer Viruserkrankung leidet.

Diese Angriffsszenarien entsprechen dem *Prosecutor-Attack* (siehe Kapitel 2.4).

4.3. l-Diversität

Motivation

Die Schwachstellen der k-Anonymität haben gezeigt, dass der Schutz von Identitäten auf der Ebene von k Individuen nicht gleichwertig ist mit dem Schutz der sensiblen Attribute. Wenn unter den sensiblen Attributen innerhalb einer Gruppe nur eine geringe Variabilität vorliegt, bietet die k-Anonymität keinen ausreichenden Schutz. Um zu verhindern, dass sensible Attribute aufgrund von Homogenität innerhalb einer Gruppe offengelegt werden, wurde in (Machanavajjhala, Kiefer, Gehrke, & Venkitasubramaniam, 2007) das Konzept der l-Diversität ($l\mathrm{D}$) beschrieben.

Definition

$l\mathrm{D}$ stellt die Forderung nach ausreichender Vielfalt der sensiblen Attribute innerhalb einer Äquivalenzklasse. Hierzu müssen mindestens l „*gut repräsentierte*“ Werte für das sensible Attribut innerhalb einer Äquivalenzklasse vorliegen. Eine Tabelle ist l-divers, wenn jede Äquivalenzklasse der Tabelle l-divers ist. $l\mathrm{D}$ ist ein Sammelbegriff für eine Gruppe von Methoden, deren konkrete Umsetzung von der Definition von „*gut repräsentiert*“ abhängt. Wir möchten im folgenden Abschnitt vier Definitionen für „*gut repräsentiert*“ vorstellen. Weitere Methoden werden in (Machanavajjhala, Kiefer, Gehrke, & Venkitasubramaniam, 2007) beschrieben.

Gegeben sei

- S die Menge an sensiblen Attributen in einer Tabelle,
- q^*-Block eine Menge an Tupeln (Äquivalenzklasse), die sich anhand des Quasi-Identifikators nicht unterscheiden,
- $s_1, \dots, s_m$ alle Ausprägungen von S in einem q^*-Block,
- $n(q^*, s_1)$ die Häufigkeit sensibler Attributwerte s in einem q^*-Block.

Distinkte l-Diversität (dlD): Eine Tabelle ist distinkt l-divers, wenn in jeder Äquivalenzklasse q^* mindestens l verschiedene Werte für das sensible Attribut existieren.

Entropie l-Diversität (ElD): Eine Tabelle T^* besitzt ElD, wenn für jeden q^*-Block gilt: $-\sum_{s \in S} P_{q^*,s} * \log(P_{q^*,s}) \geq \log(l)$ wobei $P_{q^*,s} = \frac{n(q^*,s)}{\sum_{s' \in S} n_{q^*,s}}$.

Fallbeispiele dlD und ElD

Tabelle 2 Beispieldaten l-Diversität

Quasi-Identifikatoren (Q)		Sensibles Attribut (S)
PLZ	**Alter**	**Krankheit**
1. Äquivalenzklasse (q_1^*)		
130**	<40	Herzkrankheit
130**	<40	Herzkrankheit
130**	<40	Virusinfektion
130**	<40	Virusinfektion
130**	<40	Krebs
130**	<40	Krebs
130**	<40	Krebs
130**	<40	Krebs
2. Äquivalenzklasse (q_2^*)		
1485*	>40	Krebs
1485*	>40	Herzkrankheit
1485*	>40	Virusinfektion
1485*	>40	Virusinfektion

S_1

S_2

Tabelle 2 ist *distinkt 3-divers*, da in jeder Äquivalenzklasse 3 Ausprägungen des sensiblen Attributs vorliegen: $S = S_1 = S_2 = \{Krebs, Herzkrankheit, Virusinfektion\}$. Durch die Zugehörigkeit eines Individuums zu einer Äquivalenzklasse können keine (eindeutigen) Rückschlüsse über das dazugehörige sensible Attribut geschlossen werden, da in jeder Äquivalenzklasse mindestens 3 Ausprägungen vorliegen.

Die 1. Äquivalenzklasse (q_1^*) in Tabelle 2 erfüllt $l \leq 2.828$ *Entropie l-Diversität* da

$-\frac{4}{8} * \log\left(\frac{4}{8}\right) - \frac{2}{8} * \log\left(\frac{2}{8}\right) - \frac{2}{8} * \log\left(\frac{2}{8}\right) \geq \log(l) \iff l \leq 2.828.$

Schwachstellen von dlD und ElD

Wenn Ausprägungen des sensiblen Attributs über- oder unterrepräsentiert sind, können durch den Ausschluss einzelner Datensätze mit seltenen Ausprägungen Rückschlüsse über die sensiblen Attribute der verbleibenden Datensätze gezogen werden.

Angenommen, in einer Äquivalenzklasse mit 100 Individuen liegen Krebs und Herzkrankheit bei jeweils einer Person vor und eine Virusinfektion bei den verbleibenden 98. Ist diese Verteilung bekannt, so wäre die Äquivalenzklasse zwar distinkt 3-divers, aber es reicht die Identifizierung von nur 2 Individuen – denen mit Herzkrankheit und Krebs - um die sensiblen Attribute aller 100 Individuen offen zu legen.

Weiterentwicklung: Rekursive (c,l)-Diversität

Rekursive (c,l)-Diversität ($RclD$): Durch $RclD$ wird sichergestellt, dass häufige Werte des sensiblen Attributs einer Äquivalenzklasse nicht über- und seltene nicht unterrepräsentiert sind. In jeder Äquivalenzklasse q^* werden für alle Ausprägungen $s_1, \dots, s_m$ von S die Häufigkeiten $n(q^*, s_1), \dots, n(q^*, s_m)$ absteigend sortiert. Sie bilden die Elemente der Sequenz $r_1, \dots, r_m$ (Rang).

Diese Sequenz sagt uns, wie viele Elemente mindestens eliminiert werden müssen, um eine positive Offenlegung innerhalb der Äquivalenzklasse zu ermöglichen. Ein q^*-Block ist *rekursiv* (c, l)-*divers*, wenn gilt: $r_1 < c(r_l + r_{l+1} + \cdots + r_m)$ für eine selbst definierte Konstante c.

Positive Offenlegung-Rekursive (c, l)*-Diversität* $(PORclD)$: Manchmal sind nicht alle Ausprägungen des sensiblen Attributs in gleichem Umfang schützenswert. Zum Beispiel, wenn ein Datensatz „Kranke" und „Gesunde" enthält und die Offenlegung der Ausprägung $\{Gesund\}$ nicht als Verletzung der Privatsphäre gilt. Wenn nur eine Teilmenge der Ausprägungen des sensiblen Merkmals als schützenswert angesehen wird und eine Offenlegung einer bestimmten Ausprägung keine Verletzung der Privatsphäre darstellt, dann kann die $PORcl\mathrm{D}$ angewendet werden.

Schwachstellen

Notwendigkeit / Skewness Attack: Wenn einzelne Merkmale in den Daten deutlich überrepräsentiert sind, kann das Erreichen von *l-Diversität* dem Erhalt eines für Analysen aussagekräftigen Datensatzes entgegenstehen. Nehmen wir an, wir haben einen Test auf ein Virus an 1.000 Personen durchgeführt. Die Testergebnisse sind zu 99 % negativ:

- Wenn wir eine *distinkt 2-diverse* Tabelle erhalten möchten, müssen wir sicherstellen, dass sich in jeder Äquivalenzklasse mindestens ein positiver und ein negativer Test befinden. Da insgesamt nur $1.000 * 1\% = 10$ der Tests positiv ausgefallen sind, können wir maximal 10 Äquivalenzklassen bilden, was zu einem großen Informationsverlust führen kann.
- Hinzu kommt, dass die Identifikation als negativ getestete Person vielleicht nicht von großer Relevanz ist, da es für 99 % der Bevölkerung der Fall ist. In diesem Fall wäre die $PORcl\mathrm{D}$ *als Methode geeigneter als die* $dl\mathrm{D}$.

Ähnlichkeit / Similarity Attack: Die semantische Bedeutung der Attribute wird nicht berücksichtigt. Zum Beispiel: Eine Äquivalenzklasse mit den Werten Schizophrenie, Depression und ADHS erfüllt die *distinkte 3-Diversität*. Es kann trotzdem auf eine psychische Erkrankung geschlossen werden.

4.4. t-Ähnlichkeit

Motivation

Da sowohl die *k-Anonymität* als auch die l-Diversität gewisse Schwachstellen aufweisen, schlagen (Li, Li, & Venkatasubramanian, 2006) mit der t-Ähnlichkeit eine alternative Metrik vor. Diese erfordert, dass die Verteilung des sensiblen Attributs in einer beliebigen Äquivalenzklasse nahe an der Verteilung des Attributs im gesamten Datensatz ist.

Definition

Eine *Äquivalenzklasse* ist *t-ähnlich*, wenn der Abstand der Verteilungen des sensiblen Attributs in der Klasse und die Verteilung des Attributs im gesamten Datenbestand weniger als ein Schwellenwert t ist. Der gesamte Datensatz wird als t-ähnlich bezeichnet, wenn *alle Äquivalenzklassen t-ähnlich* sind.

Folgende praktische Eigenschaften gelten in Bezug auf die *t-Ähnlichkeit* (Li, Li, & Venkatasubramanian, 2006) :

- *Generalisierungs-Eigenschaft*: Generalisierungen eines *t-ähnlichen* Datensatzes sind ebenfalls *t-ähnlich*.
- *Teilmengen-Eigenschaft*: Wenn ein Datensatz gegenüber einer Menge von Attributen *t-Ähnlichkeit* erfüllt, so erfüllt der Datensatz auch *t-Ähnlichkeit* gegenüber einer Teilmenge der Attribute.

Sei *Q* die Verteilung des sensiblen Attributes im gesamten Datenbestand und *P* die Verteilung in einer bestimmten Äquivalenzklasse. Ziel ist es, das Ausmaß, indem Beobachter zusätzliche Informationen über bestimmte Personen erhalten, zu beschränken. Dies wird durch die Begrenzung des Abstands zwischen *P* und *Q* erreicht. Der Abstand der Verteilungen hängt von den Abständen der Attributwerte ab. Zur Messung des Abstands von Verteilungen verwenden (Li, Li, &

Venkatasubramanian, 2006) die „*Earth Mover* Distance“. Diese beschreibt den minimalen Aufwand, der notwendig ist, eine Verteilung in eine andere „umzuwandeln“. Dies erfolgt durch die Verschiebung der Wahrscheinlichkeitsmassen.

Kritische Würdigung

Mit dem Konzept der *t-Ähnlichkeit* soll sichergestellt werden, dass die Verteilung der sensiblen Attribute in jeder Äquivalenzklasse, der der Gesamtverteilung ähnelt. Dadurch werden Risiken durch Homogenitäts- und Skewness-Angriffe, wie sie bei der l*-Diversität* auftreten reduziert. Zudem können semantische Ähnlichkeiten berücksichtigt werden (Li, Li, & Venkatasubramanian, 2006).

Andererseits erhöht die Berechnung und Sicherstellung der *t-Ähnlichkeit* die Komplexität der Implementierung und verringert somit deren Praktikabilität. Zudem müssen die Daten für die Angleichung der Verteilung des sensiblen Attributs an die des gesamten Datensatzes möglicherweise so stark verzerrt werden, dass ihre Nützlichkeit für Analysezwecke erheblich eingeschränkt wird. Diese Effekte verstärken sich, je hochdimensionaler die Daten sind.

Wie auch bei der *k-Anonymität* und der l*-Diversität* stellt die Wahl eines geeigneten Schwellenwerts t eine Herausforderung dar.

5. Differential Privacy

Differential Privacy („differentielle Privatsphäre”) ist ein Konzept, das einen robusten mathematischen Rahmen zur Verfügung stellt, um einen umfassenderen Schutz der Privatsphäre zu gewährleisten. Seit der ersten Veröffentlichung im Jahr 2006 (Dwork, 2006) hat *Differential Privacy* (kurz: *DP*) in der akademischen Welt weite Verbreitung gefunden und wird allmählich auch von der Privatwirtschaft übernommen (Cummings & Desai, 2018). Kurz beschrieben ist *DP* das Versprechen eines Datenhalters an ein Datensubjekt, dass dieses durch die Aufnahme seiner Daten in den Datenpool nicht negativ beeinträchtigt wird, unabhängig davon, welche anderen Studien, Datensätze oder Informationsquellen verfügbar sind (vergleiche (Dwork, 2006)). Dabei bezieht sich Differential Privacy nicht auf die Daten selbst, sondern auf die Antworten von Abfragen auf diese Daten. Der Schutzmechanismus besteht darin, gezielt Rauschen in die Ergebnisse von Berechnungen einzufügen, sodass keine zuverlässigen Rückschlüsse auf das Vorhandensein oder Fehlen einzelner Personen in der Datenbasis gezogen werden können.

DP hat mehrere Vorteile gegenüber klassischen Anonymisierungsmetriken, darunter (Dwork & Roth, 2014):

- **Schutz vor „unvorhersehbaren“ Risiken**: Damit geht *DP* über den reinen Schutz vor Re-Identifikation hinaus und schützt vor jeglicher Art von Inference-Risiken bezüglich Eigenschaften individueller Personen im Datensatz.
- **Automatische Neutralisierung von Verknüpfungsangriffen**: Dies umfasst Versuche mit vergangenen, aktuellen und zukünftigen Datensätzen sowie anderen Formen und Quellen von Hilfsinformationen Rückschlüsse auf die Identität zu erreichen. Insbesondere bietet *DP* damit einen Schutz vor zukünftigen, noch nicht bekannten Datensätzen und Analysemethoden.
- **Quantifizierung des Verlustes an Privatsphäre**: DP ist kein binäres Konzept, sondern beinhaltet ein Maß für den Verlust der Privatsphäre. Dies ermöglicht Vergleiche zwischen verschiedenen Techniken und Parametern und ein Abwägen zwischen Datenschutz und Datennutzen (siehe Kapitel 6).

Definition

Ein randomisierter Algorithmus *A* ist *(ε,δ)-differentiell privat*, wenn für alle Datensätze *D* und *D'*, die sich in höchstens einem Element unterscheiden, und für alle Mengen von Ausgängen *S⊆Bereich(A)*, gilt:

$$P[A(D) \in S] \leq \exp(\varepsilon) \cdot P[A(D') \in S] + \delta$$

Hierbei bedeuten:

ε (Epsilon): Ein nicht-negativer Parameter, der den Grad der Privatsphäre steuert. Kleinere Werte für ε entsprechen stärkeren Datenschutzgarantien.

δ (Delta): Ein Parameter, der die Wahrscheinlichkeit berücksichtigt, dass der Algorithmus die ε -differentielle Privatsphäre nicht erreicht. δ repräsentiert damit, für wie viele Teilnehmer des Datensatzes die Privatsphäre nicht garantiert ist. Idealerweise sollte δ nahe null sein. Im Falle von δ = 0 spricht man auch von einem *ε-differentiell privaten* Algorithmus.

Die Definition beschreibt, dass man von dem Resultat des Algorithmus ausgehend nicht feststellen kann, ob die Information einer Person in dem originalen Datensatz enthalten ist oder nicht (*D* bzw. *D'*).

DP bezieht sich damit immer auf eine spezifische Auswertungsmethode einer Datenbasis, im Gegensatz anderer Anonymisierungsmetriken die auf dem resultierenden Datensatz definiert sind. In (Dwork, 2006) werden mehrere Beispiele gegeben, die die Wichtigkeit dieser Unterscheidung veranschaulichen. So kann in manchen Szenarien das Hinzufügen von etwas Rauschen einen sehr starken, nicht-linearen, Effekt auf den Datenwert haben. Zum Beispiel bei der Auswertung des häufigsten Wertes einer Variable kann schon das Hinzufügen von ein wenig Rauschen den Ausgabewert ändern.

Des Weiteren ist Vorsicht geboten, wenn mehrere Auswertungen von demselben Datensatz erzeugt werden, z.B. separate Histogramme von verschiedener Dimension. Normalerweise summiert sich in diesem Fall das ε_i der einzelnen Auswertungen zu einem Gesamt-Datenverlust ε. Umgekehrt betrachtet startet man normalerweise mit einem Gesamt ε, dem sogenannten *Privacy-Loss Budget*, welches dann auf die verschiedenen gewünschten Auswertungen alloziert wird, siehe z.B. (U.S. Census Bureau, 2023).

DP liefert daher nicht nur eine Anonymisierungsmetrik, sondern auch verschiedene Anonymisierungsmethoden, um *DP* unter diesen Bedingungen zu erreichen.

Anonymisierungsmethoden

Typischerweise wird *DP* durch *pertubative Methoden*, insbesondere das Hinzufügen von Rauschen erreicht, es gibt aber auch andere *nicht-pertubative Ansätze*, siehe z.B. (Bild, Kuhn, & Prasser, 2018). Um sich mit dem Thema vertraut zu machen, wird empfohlen mit den Standardmethoden des Hinzufügens von Rauschen (Laplace und Gauss) anzufangen, die sich besonders für Häufigkeitsauswertungen wie Histogramme und min/max-Abfragen eignen.

Beide Verfahren, Laplace und Gauss, sind *Additive Random Noise* Methoden (siehe Abschnitt 3.3.1), die jedem ursprünglichen Ausgabewert einen Fehlerwert hinzuaddieren (positiv oder negativ). Der Fehlerwert wird zufällig von einer Laplace, bzw. Gauss-Verteilung generiert. Die Parameter beider Verteilungen werden direkt abgeleitet von dem gewünschten Sicherheitsniveaus (ε und δ) und der Sensitivität des zugrundeliegenden Algorithmus.[3] Die Sensitivität entspricht dabei gerade dem Wert, um den sich der Ausgabewert maximal verändert, wenn nur ein Datenelement verändert wird.

Bei der Laplace Methode wird die Sensitivität mit der absoluten Differenz gemessen, bei der Gauss Methode mit der quadrierten Differenz. Dies entspricht den *L1*- und *L2*-Normen wie sie auch in anderen statistischen Kontexten oft benutzt werden. In der Theorie werden gewisse Vor- und Nachteile von Laplace und Gauss genannt, z.B.:

- Der Gauss-Methode liegt die Normalverteilung zugrunde, die einfach zu interpretieren sein kann und sich einfacher kombinieren lässt, z.B. bei der Komposition von zwei Algorithmen (Dwork & Roth, 2014)

[3] Laplace: μ = 0, b = 1/ε * L1 Sensitivität
Gauss: μ = 0, σ = c/ε * L2 Sensitivität, mit $c^2 > 2 * \ln(1{,}25/\delta)$

- Die Laplace-Methode ermöglicht ε-differentielle Privacy (δ = 0), wohingegen Gauss nur δ > 0 erlaubt.

In der Praxis sind diese Einschränkungen jedoch nicht immer von Relevanz, und es wird empfohlen beide Optionen am gegebenen Anonymisierungsfall auszuprobieren. So ist z.B. das Erreichen von δ = 0 in der Praxis oft eh nicht realistisch oder nur unter hohem Verlust des Datennutzens zu erreichen.

Ein zu hoher Verlust des Datennutzens bei Anwendung der Standardmethoden scheint ein häufiges Problem in der Praxis zu sein (Aircloak, 2019) und es existieren bereits mehrere Weiterentwicklungen mit dem Ziel einer höheren Genauigkeit, bzw. geringerem Verlust von Datennutzen. In dem folgenden Beispiel des US Census wurde z.B. eine verbesserte Version der Gauss-Methode genutzt. Da dies ein aktiver Forschungsbereich ist, wird empfohlen bei Interesse die aktuelle Literatur zu recherchieren und sich mit den letzten Ergebnissen vertraut zu machen.

Fallbeispiele

Im Folgenden werden zwei reale Anwendungsbeispiele von DP gegeben, die sich gut eignen, um sich tiefer mit dem Thema vertraut zu machen.

Das erste Beispiel, der US-Zensus, ist am ehesten vergleichbar mit den typischen Problemstellungen im Versicherungsbereich. Ein zentraler Datenhalter hat Zugriff auf die Rohdaten aller Datensubjekte und wünscht die Ergebnisse einer Auswertung zu anonymisieren, um diese mit anderen Nutzern zu teilen. Dieses Modell wird als *Central Differential Privacy*, oder *Trusted-Curator Modell* von *Differential Privacy* bezeichnet.

Die Situation im zweiten Beispiel, Apple, betrifft viele Unternehmen, die sensible oder detaillierte Informationen ihrer Privatkunden sammeln und nutzen möchten. Um Datenschutzvorbehalten der Kunden entgegenzuwirken, können deren Daten mittels *Local Differential Privacy* schon im Vorfeld anonymisiert werden, bevor sie vom Unternehmen gespeichert und aggregiert werden.

1) Central Differential Privacy: US Zensus

Das US Census Bureau führt alle zehn Jahre eine umfassende Volkszählung durch, die neben der Einwohnerzahl auch demographische und ökonomische Informationen umfasst. Die publizierten Statistiken werden von vielen staatlichen Institutionen unter anderem zur Budgetallokation und Planung genutzt, und erfordern eine hohe Genauigkeit. Gleichzeitig ist das Census Bureau gesetzlich verpflichtet, den Datenschutz der teilnehmenden Individuen (und Haushalte) zu gewährleisten (U.S. Census Bureau, Disclosure Avoidance for the 2020 Census, 2021a). Die Herausforderung für das US Census Bureau ist es daher, diese zwei Anforderungen in Einklang zu bringen.

In früheren Zensus Publikationen (zuletzt 2010) benutzte das Bureau klassische Anonymisierungsmethoden, wie Generalisierung, Suppression und Data (U.S. Census Bureau, 2021a). In Folge der zunehmenden Verfügbarkeit von detaillierten Bevölkerungsdaten wurde diese Anonymisierungsstrategie als nicht mehr ausreichend eingestuft. Das Bureau simulierte selber verschiedene Angriffsszenarien auf die 2010er Zensus Daten mit erfolgreicher Rekonstruktion einer Vielzahl an Originaldaten. Das Dokument „Disclosure Avoidance for the 2020 Census: An Introduction“ (U.S. Census Bureau, 2021a) bietet eine gute Einführung in die aktuelle Anonymisierungsstrategie und deren Hintergründe.

Für den 2020 Zensus wurde Differential Privacy angewendet. Anstelle der klassischen Definition wurde eine Weiterentwicklung, die „*zero-Concentrated Differential Privacy*“ (Bun & Steinke, 2016), genutzt. Diese hat gegenüber der klassischen Definition gewisse Vorteile, insbesondere wenn mehrere Auswertungen auf dem gleichen Datensatz kombiniert werden. Sie lässt sich leichter auf eine diskrete Gauss-Verteilung aufsetzen, welche für den Zensus anstelle der klassischen stetigen Gauss-Verteilung genutzt wurde (U.S. Census Bureau, 2023).

Nach verschiedenen Konsultationsrunden wurde das Privacy-Loss Budget ε = 19,61 gewählt (U.S. Census Bureau, 2021b). Der Findungsprozess des US Zensus Bureau zeigt anschaulich

die Herausforderung bei der Wahl eines geeigneten ε. Die in der Theorie postulierten niedrigen Werte von ε führen in der Praxis meist zu sehr starkem Verlust des Datennutzens. Dies veranschaulicht den Vorteil des DP- Konzeptes, welches ein objektiveres Abwägen von Datensicherheit und Datennutzen erlaubt.

Im finalen Algorithmus des US Zensus, dem „TopDown Algorithm“ werden neben dem Hinzufügen von Rauschen auch weitere besondere Anforderungen an den Zensus modelliert (U.S. Census Bureau, 2023). So ist z.B. eine Nebenbedingung, dass gewisse Statistiken, wie die Einwohnerzahl pro Bundesstaat, ihren Originalwert beibehalten.

Die Quelle (U.S. Census Bureau, 2023) liefert einen guten tieferen Einblick in die Anonymisierungsstrategie und deren technische Umsetzung.

2) Local Differential Privacy: Apple

Im Jahr 2016 kündigte Apple die geplante Nutzung von Differential Privacy für die Sammlung von Nutzerstatistiken an. Mit Release von MacOS und iOS 10 wurde DP selektiv für die Anonymisierung von gewissen Nutzerinformationen, wie Emoji-Beliebtheit und Schreibkorrekturen, eingesetzt (Apple Inc, 2017). Für einen verstärkten Datenschutz werden die Original-Nutzerdaten nicht zentral gesammelt, sondern schon lokal auf dem Endgerät des Nutzers anonymisiert.

Die Definition von Local DP ähnelt dabei der Definition von Central DP (Bassily & Smith, 2015), nur dass der Anonymisierungsalgorithmus A_i auf den individuellen Datensatz eines Nutzers i angewandt wird. Die anonymisierten Einträge werden anschließend durch einen geeigneten Algorithmus in die gewünschte Statistik transformiert. (Apple Inc, 2017) liefert eine detaillierte Beschreibung des gesamten Prozesses.

Der Trade-off zwischen dem Schutz der Privatsphäre und Erhalt von Datennutzen scheint auch für Apple eine kontinuierliche Herausforderung. Eine strikte Anwendung von Local DP erlaubt oft nur die Erkenntnis von groben Trends und limitiert den Anwendungsbereich. Apple arbeitet daher an verfeinerten Modellen, die es erlauben mehr Informationen zu erhalten, siehe z.B. (Apple Inc, 2023).

6. Messung von Datennutzen und Informationsverlust

6.1. Motivation

Durch die Anonymisierung von Daten kommt es zu einem Informationsverlust gegenüber den Originaldaten, was zu einer Verringerung des Datennutzens führen kann. Um im Sinne des Datenschutzes das Risiko einer Offenlegung von sensiblen Daten zu minimieren und trotzdem die anonymisierten Daten in ihren statistischen Eigenschaften so nah wie möglich an den Originaldaten zu behalten, ist es notwendig einen Kompromiss zu schaffen.

In diesem Kapitel werden Maße vorgestellt, mit denen der Datennutzen sowohl vor als auch nach der Anonymisierung bewertet werden kann. Zudem werden Ansätze zur Quantifizierung des durch Anonymisierung entstehenden Informationsverlustes aufgezeigt. Wichtig ist dabei zu unterscheiden, wie die anonymisierten Daten genutzt werden sollen und welche Anforderungen an die Daten nach Anonymisierung gestellt werden.

Eine praxisnahe Anleitung zur Messung von Datennutzen und Informationsverlust bietet das Tutorial https://sdcpractice.readthedocs.io/.

6.2. Allgemeine Maße für kategorielle und stetige Variablen

Der Verlust an Datennutzen hängt von den Anwendungsszenarien des Datensatzes ab und idealerweise werden diese in der Berechnung berücksichtigt. Dies ist in der Praxis jedoch oft nicht möglich da Datensätze diverse Anwendungen unterstützen, bzw. diese nicht alle im Vorfeld bekannt sind (Domingo-Ferrer & Torra, 2001). Es ist daher von Vorteil, wenn Verlust von Datennutzen allgemein bestimmt werden kann, ohne Spezifizierung der konkreten Anwendungsszenarien.

(Domingo-Ferrer & Torra, 2001) beschreibt beispielsweise, dass ein geringer Informationsverlust vorliegt, wenn die analytische Struktur des maskierten Datensatzes sehr ähnlich zu den Originaldaten ist. Möglichkeiten einen solchen Vergleich zu erstellen, werden als nächstes vorgestellt.

- Anzahl fehlender Werte beispielsweise nach der Anwendung von Suppression: Umso höher die Anzahl fehlender Werte, umso größer der Informationsverlust. Vergleich der fehlenden Werte vor und nach Anonymisierung, bzw. deren relativen Anteil
- Anzahl geänderter Werte nach Anonymisierung gibt Indikation für die Auswirkung der Anonymisierung. Hierzu schlägt (Domingo-Ferrer & Torra, 2001) Maße für nominale und ordinale Variablen vor.
- Messung von Informationsverlust durch Vergleich der Kontingenztabellen des originalen und des maskierten Datensatzes. Weitere Maße in der Literatur auf Basis von Kontingenztabellen ist beispielsweise µ-Argus (Hundepool & Willenborg, 1998), das den Informationsverlust für lokale Suppression misst. Um einen möglichst analytisch validen Datensatz nach Anonymisierung zu erhalten, sollten die Kontingenztabellen sehr ähnlich sein.

6.3. Allgemeine Maße für stetige Variablen

Speziell für stetige Variablen bietet es sich an statistische Maße vor und nach der Anonymisierung zu vergleichen. Idealerweise sollten sich die statischen Eigenschaften für die relevanten Variablen, wie Mittelwert, Kovarianz und Korrelation, vor und nach Anonymisierung möglichst wenig ändern. Das gleiche gilt für Gütemasse von Modellen, wie *Mean Square Error*, *Mean Absolute Error* und *Mean Variation* (Domingo-Ferrer & Torra, 2001).

Um den Grad der Änderung besser bewerten zu können wird empfohlen zusätzlich zu den Statistiken auch deren Konfidenzintervalle mitzuberechnen. Dies erlaubt z.B. zu überprüfen ob die neue Metrik nach Anonymisierung in dem Konfidenzintervall der Originalmetrik enthalten ist. Der Grad von Überlappung zwischen den beiden Konfidenzintervallen, vor und nach Anonymisierung, ist auch ein Indikator des Grad der Veränderung (Benschop, Machingauta, & Welch, 2022).

(Yancey, 2002) schlägt verschiedene Normen zur Messung des Informationsverlusts („*Information Loss*“) nach *Pertubation* (z.B. *Rank Swapping* oder *Additive Noise*) vor. Die Idee ist einen *Penalty Score* zu definieren, der angibt, inwieweit die maskierten Daten von den originalen abweichen. Die IL1-Norm misst beispielsweise den durchschnittlichen, absoluten Abstand der originalen zu den anonymisierten Daten. Weitere Maße basieren auf den Mittelwerten, Kovarianzen und Korrelationen, diese bergen allerdings Probleme bei Werten nahe oder gleich Null.

In (Meindl, 2008) wird der Vergleich der Eigenwerte vor und nach Anonymisierung beschrieben. Als Maß für die Abweichung wird die Differenz zwischen den Eigenwerten der Kovarianzen der standardisierten originalen und den pertubierten Daten der stetigen Variablen herangezogen.

6.4. Visualisierungsmethoden

Neben der Berechnung von bestimmten Maßen bietet es sich an, grafische Darstellungen für einzelne Variablen vor und nach Anonymisierung zu vergleichen.

- Mit Hilfe von Histogrammen lassen sich beispielsweise einfach Abweichungen in der Verteilung einzelner Variablen feststellen.
- Dichtefunktionen hängen zwar von der Wahl des Schätzers und deren Anpassungsgüte ab, bieten aber die Möglichkeit Unterschiede vor und nach Anonymisierung von stetigen Variablen zu visualisieren.
- Für stetige Variablen geben Boxplots einen Einblick in mögliche Änderungen in der Streuung und vorhandene Ausreißer.
- Für kategorielle Variablen eignen sich besonders Mosaic Plots, mit denen der Einfluss von Anonymisierungstechniken gut gegenübergestellt werden kann. Beispielsweise die Wahl des Parameters *k* bei der Herstellung von *k-Anonymität*.

6.5. Maße für Nutzen des Endnutzers

Sind die Endnutzer oder der Verwendungszweck des Datensatzes bekannt, so können sie in der Bewertung des Informationsverlustes auf den anonymisierten Daten Berücksichtigung finden.

Ein Anwendungsbeispiel ist die Berechnung des *GINI-Koeffizienten* basierend auf der Variable Einkommen, um die Ungleichverteilung von Einkommen zu analysieren. Ein Vergleich des GINI-Koeffizienten vor und nach Anonymisierung verdeutlicht den Informationsverlust.

Wird der Datensatz mittels Regressionsmodellen analysiert, dann lassen sich diese auch gut nutzen, um sicherzugehen, dass die Struktur in den Daten durch die Anonymisierung erhalten bleibt. Dieses Vorgehen ist sowohl für kategorielle als auch für stetige Variablen geeignet. Z.B. können die Regressionsparameter direkt verglichen werden.

Ein konkretes Beispiel ist die Nutzung eines Datensatzes zur Modellierung des Einkommens mittels der *Mincer Gleichung*. Die *Mincer Gleichung* (Mincer, 1974) modelliert das Einkommen als lineare Regression von Geschlecht, Bildung und Arbeitserfahrung. Nach Anonymisierung des Datensatzes (z.B. von Geschlecht) und Neuberechnung der linearen Regression können die Regressionsparameter und deren Konfidenzintervalle vor und nach Anonymisierung verglichen werden.

6.6. Wahl geeigneter Maße

Welche Maße sich für die Bewertung von Anonymisierungstechniken am besten eignen, hängt von der Wahl der Anonymisierungsmethode und der zu Grunde liegenden Datentypen ab.

Bei Perturbationsmethoden ist es beispielsweise nicht sinnvoll, einzelne Werte direkt zu vergleichen. Stattdessen sollten statistische Größen vor und nach der Anonymisierung untersucht werden. Zudem sollten nicht nur einzelne Variablen betrachten werden, sondern auch deren Interaktionen. Hierzu können Kreuztabellen oder Regressionsmodelle genutzt werden.

7. Erzeugung Synthetischer Daten

7.1. Motivation

Die bisher vorgestellten Anonymisierungstechniken basieren alle auf dem originalen Datensatz, deren Datenelemente angepasst oder modifiziert werden, mit dem Ziel einen geeigneten Kompromiss zwischen Anonymisierung und Datennutzen zu erreichen. Bei der Generierung synthetischer Daten bleibt das Ziel das gleiche, allerdings werden die statistischen Eigenschaften der Originaldaten genutzt, um ein Modell zu trainieren, das neue Daten erzeugt mit den gleichen Eigenschaften. Eine Definition von synthetischen Daten gibt der Europäische Datenbeauftragte (kurz EDSB, siehe (Riemann, abgerufen am 26.03.2025)):

Synthetische Daten sind künstliche Daten, die aus Originaldaten generiert werden, und ein Modell, das trainiert wird, die Eigenschaften und die Struktur der Originaldaten zu reproduzieren. Dies bedeutet, dass synthetische Daten und Originaldaten sehr ähnliche Ergebnisse liefern sollten, wenn sie derselben statistischen Analyse unterzogen werden. Der Grad, in dem synthetische Daten ein genauer Proxy für die ursprünglichen Daten sind, ist ein Maß für den Nutzen der Methode und des Modells.

Der Vorteil synthetischer Daten ist, dass nicht auf individuelle Originaldaten zurückgegriffen werden muss, wodurch das Risiko der Re-Identifizierung deutlich sinkt. Die Herausforderung besteht dann darin die wesentlichen Eigenschaften des Datensatzes in dem statistischen Modell abzubilden und den Verlust an Datennutzen möglichst gering zu halten.

Um synthetische Daten zu erzeugen, gibt es sehr viele verschiedene Methoden. Grob kann man unterscheiden zwischen Modellen mit Verteilungsannahme und ohne Verteilungsannahme. Zu den Methoden ohne Verteilungsannahme gehören Machine Learning Modelle sowie Deep-Learning

Verfahren. Im Folgenden werden drei prominente Modelle detaillierter vorgestellt, die bereits in bekannten R- und Python-Bibliotheken implementiert sind. Diese Modelle dienen als Grundlage für viele komplexere Modelle.

7.2. Modelle mit Verteilungsannahme: Parametrische Modelle

Mit Hilfe eines parametrischen Modells wird die *Posterior predictive distribution* berechnet, aus der anschließend Stichproben gezogen werden, die die synthetischen Daten darstellen. (Raab, Nowok, & Dibben, 2016) präsentiert ein Verfahren unter der Annahme, dass die beobachteten Daten eine Stichprobe aus einer Verteilung mit Parametervektor θ sind, die vom Datensynthesizer geschätzt werden kann. Dabei wird eine Variable X_i aus $\{X_1, \dots, X_p\}$ gewählt und die Verteilung bedingt X_{-i} (Menge aller Variablen außer der i-ten Variablen) geschätzt. Mit Hilfe der Posteriori Verteilung für θ und dem vorher festgelegten parametrischen Modell kann die posterior predictive Verteilung simuliert werden. Die synthetischen Daten sind anschließend Zufallsstichproben aus der Posterior Predictive distribution. Dieses Vorgehen wird für alle Spalten wiederholt, sodass am Ende die gemeinsame Verteilung aus einer Serie von bedingten Verteilungen definiert ist. Diese Methodik ist im R-Paket *synthpop* implementiert.

Weitere Methoden sind Markov-Chain-Monte-Carlo-Methoden, Gaußsche Mischverteilungsmodelle, Copula-Modelle oder Bayes'sche Netze, die mit dem Differential-Privacy-Ansatz kombiniert werden können, (Zhang, Cormode, Procopiuc, Srivastava, & Xiao, 2017).

7.3. Modelle ohne Verteilungsannahme: Nicht-parametrische Machine Learning Modelle

Genutzt werden können verschiedene nicht parametrische Regressionsmodelle wie zum Beispiel *CART*, *Random Forest* oder *Support Vector Machines*. Gegeben seien die Variablen X_i, $i = 1, \dots, p$ und deren *n* Beobachtungen, sowie ein nicht-parametrisches Modell. Der Algorithmus zum Generieren von synthetischen Daten kann dann wie folgt beschrieben werden:

1. Regressiere X_1 auf X_{-1} mit einem nicht-parametrischen Modell der Wahl. Diese Regression kann als $X_1 | X_{-1}$ geschrieben werden. Seien Y_1 die synthetischen Daten für X_1.
2. Regressiere X_2 auf X_{-2} und nutze dazu $(Y_1, X_3, \dots, X_p)$ um neue Werte für X_2 vorherzusagen. Seien dann Y_2 die synthetischen Werte für X_2.
3. Folge diesem Ansatz weiterhin, allgemein lässt sich somit schreiben: Für jedes $i = 3, \dots, p$ wird X_i auf $\{X_{i+1:p}\} \cup \{Y_{1:i-1}\}$ regressiert.

Somit erhält man für jede Variable X_i die zugehörigen synthetischen Daten Y_i . Diese Methodik ist in dem R Paket synthpop implementiert.

7.4. Deep Learning Methoden

7.4.1. Nutzung von GANs

Die *Generative Adversarial Nets* (kurz: *GAN*) wurden erstmals von (Goodfellow, et al., 2014) vorgestellt. Es handelt sich um einen Rahmen für die Schätzung generativer Modelle innerhalb eines adversen Prozesses. Dabei werden gleichzeitig zwei Modelle trainiert: ein generatives Modell, das die Datenverteilung *G* erfasst, und ein diskriminatives Modell *D*, das die Wahrscheinlichkeit schätzt, dass eine Stichprobe aus den realen Daten und nicht aus *G* stammt. Das Trainingsverfahren entspricht einem Minimax-Spiel mit zwei Spielern, da das Ziel von *G* darin besteht, die Wahrscheinlichkeit zu maximieren, dass *D* einen Fehler macht. Die mehrschichtigen Perzeptrons werden für *G* und *D* so gewählt, dass das gesamte System durch Backpropagation trainiert werden

kann. Das Training eines GANs ist in der nachfolgenden Grafik beispielhaft abgebildet.

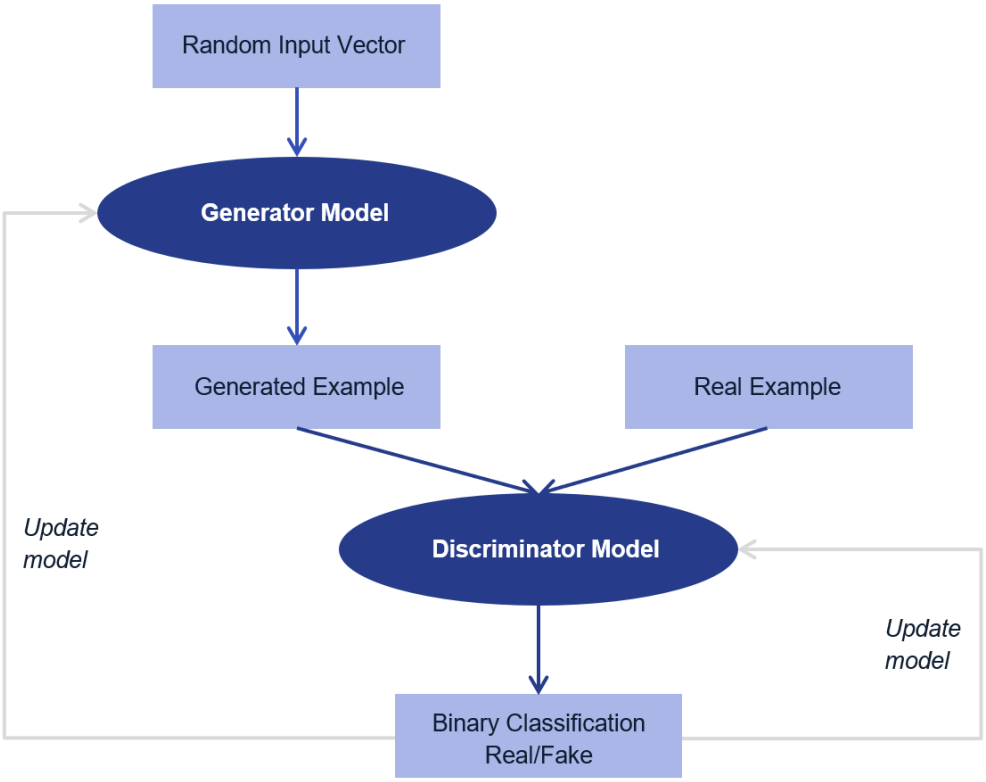

Abbildung 1 Vereinfachte Darstellung des Trainings eines GANs

Für die Erzeugung synthetischer Daten *Y* lernt der Datensynthetisierer *G* aus einer Tabelle *X*, die sowohl stetige als auch diskrete Spalten enthalten kann. Dabei wird jede Spalte als eine Zufallsvariable aufgefasst und angenommen, dass die Zufallsvariablen einer gemeinsamen Verteilung folgen. Um *G* zu trainieren, werden die Daten in Trainings- und Testdaten unterteilt, wobei ausschließlich die Trainingsdaten zum Trainieren von *G* genutzt werden. Anschließend werden die synthetischen Daten erzeugt, indem sie zufällig und unabhängig aus *G* gezogen werden.

7.4.2. GAN Erweiterung

GANs haben einige Probleme beim Erzeugen synthetischer Daten, unter anderem:

1. Nicht-Gaußsche Verteilungen: Gauß-ähnliche Verteilungen können durch eine *Min-Max-Transformation* normalisiert werden. Kontinuierliche Daten sind jedoch in der Regel nicht gaußförmig, sodass eine *Min-Max-Transformation* zum Problem des verschwindenden Gradienten führt.
2. Multinomiale Verteilungen: (Srivastava, Valkov, Russel, Gutmann, & Sutton, 2017) zeigen, dass ein *„Vanilla"-GAN* bei einem einfachen zweidimensionalen Datensatz nicht alle Modi modellieren kann. Daher würde es auch Schwierigkeiten mit der Modellierung multimodaler Verteilungen für zusammenhängende Spalten geben.
3. Stark unausgewogene kategorielle Spalten: Stark unausgewogene Spalten verursachen einen schweren Modus-Kollaps. Das Fehlen einer seltenen Kategorie führt nur zu minimalen Veränderungen in der Datenverteilung, die vom Diskriminator nur schwer zu erkennen sind.

In (Xu, Skoularidou, Cuesta-Infante, & Veeramachaneni, 2019) wird das *CTGAN*-Modell vorgestellt, eine Erweiterung des klassischen *GAN*-Ansatzes. Es löst die oben genannten Herausforde-

rungen durch die Einführung einer modus-spezifischen Normalisierung (*mode-specific normalization*) und durch den Einsatz eines bedingten Generators (*conditional generator*) sowie durch *Training-by-Sampling*. CTGAN sowie viele Trainings- und Evaluationsmethoden für synthetischen Daten sind in der *sdv-Library* in Python implementiert.

7.4.3. GAN mit Differential Privacy

Die Erzeugung von synthetischen Daten verringert das Re-Identifizierungsrisiko, jedoch ist oft unklar inwieweit. Eine mögliche Antwort bietet Differential Privacy (siehe Kapitel 5).

Der *PATE-GAN* Ansatz (Jordon, Yoon, & Van Der Schaar, 2018) integriert Differential Privacy durch die Kombination von zwei Techniken: den *Generative Adversial Nets* sowie der *Private Aggregation of Teacher Ensembles (PATE)*.

Die Hauptkomponenten eines *PATE-GANs* sind die eines herkömmlichen *GANs*: Generator und Diskriminator. Der Generator erzeugt synthetische Daten, während der Diskriminator lernt, zwischen echten und synthetischen Daten zu unterscheiden. Das Besondere ist, dass *PATE-GAN* als Diskriminator ein *PATE*-Modell verwendet, sodass der Diskriminator differentiell privat ist. Das *PATE*-Framework geht auf (Papernot, Abadi, Erlingsson, Goodfellow, & Talwar, 2016) zurück bei dem ein „Student"-Modell lernt Ergebnisse vorherzusagen durch einen Mehrheitsentscheid vieler „Teacher"-Modelle. Dabei kann das „Student"-Modell nicht auf die zugrunde liegenden Daten oder Parameter der „Teacher"-Modelle zurückgreifen, wodurch Differential Privacy gewährleistet wird.

Im *PATE-GAN* Modell funktioniert der *PATE*-Ansatz wie folgt: Um eine neue Stichprobe zu klassifizieren, werden die Ergebnisse jedes „Teacher"-Modells auf der Stichprobe ausgewertet und dann werden alle Ergebnisse verrauscht aggregiert. Diese verrauschte Aggregation erzeugt jedoch einen Klassifikator, der in Bezug auf die Parameter des Generators nicht mehr differenzierbar ist.

Um diesem Problem entgegenzuwirken, wird die Idee aus (Papernot, Abadi, Erlingsson, Goodfellow, & Talwar, 2016) verfolgt:

- Öffentlich zugängliche, ungelabelte Daten werden genutzt und mit Hilfe des Standard-*PATE*-Mechanismus gelabelt.
- Diese werden anschließend dazu genutzt, um das „Student"-Modell zu trainieren.

Da öffentliche Daten zur Generierung synthetischer Daten oft nicht verfügbar oder praktikabel sind, wird die Trainingsmethode entsprechend angepasst. Statt öffentlicher Daten werden ausschließlich die Ergebnisse des differentiell privaten Generators für das Training des „Student"-Modells verwendet.

8. Anonymisierungsprozess

In diesem Kapitel wird ein Prozess vorgestellt, der als Leitfaden zur Anonymisierung eines konkreten Datenbestandes genutzt werden kann. Er integriert die verschiedenen Anonymisierungskomponenten die in den vorherigen Kapiteln vorgestellt wurden. Der vorgestellte Prozess entspricht der Best Practice zur statistischen Offenlegungskontrolle, wie er typischerweise von Forschern, Statistikämtern und ähnlichen Organisationen angewendet wird (Benschop, Machingauta, & Welch, 2022).

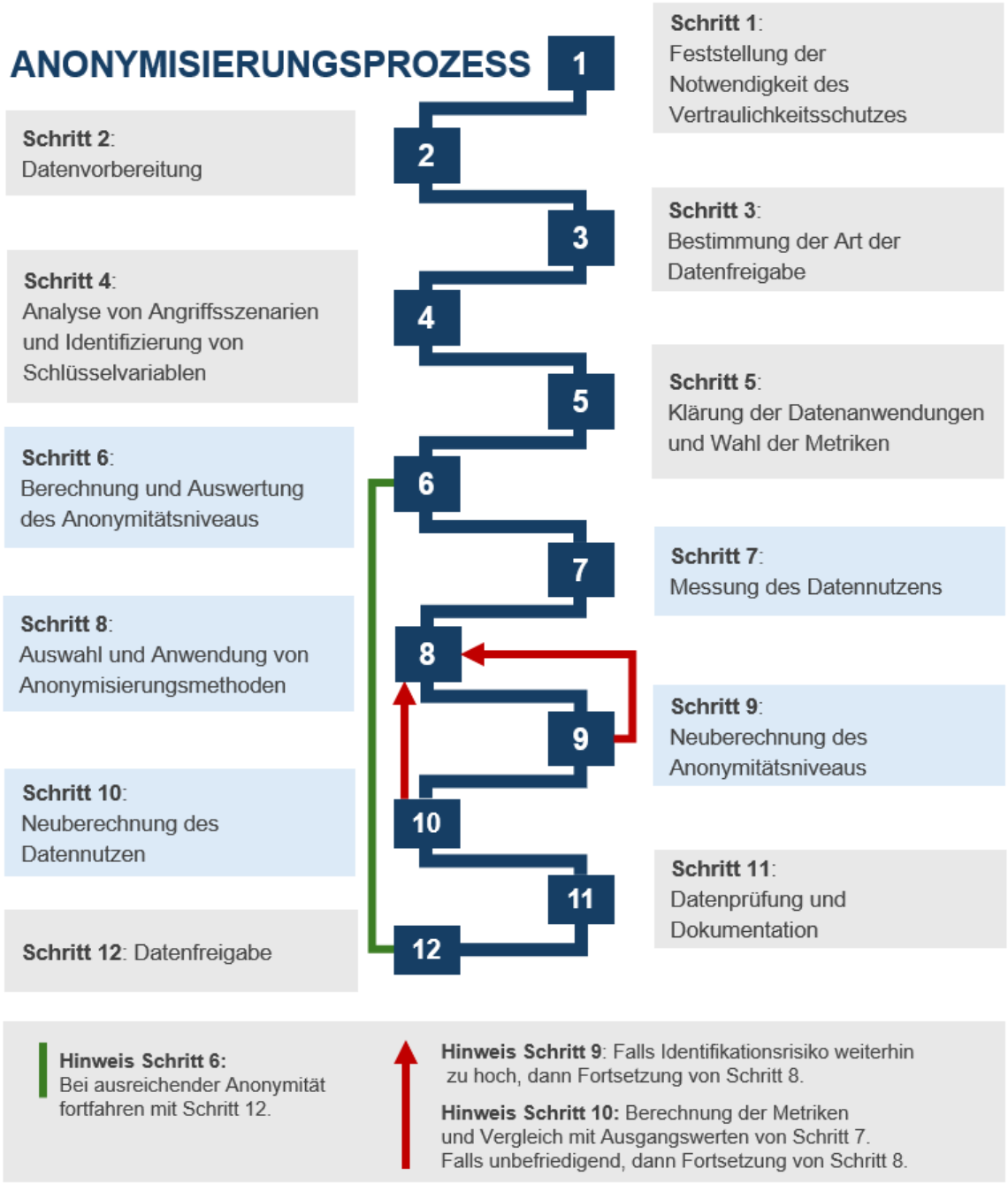

Abbildung 2 Übersicht des Anonymisierungsprozesses

Die Abbildung 2 präsentiert den kompletten Prozess, schematisch. Eine iterative Anwendung des Prozesses ist dabei meist erforderlich, insbesondere was die schrittweise Durchführung von Anonymisierungsmethoden und Messung deren Effektivität betrifft.

Schritt 1 – Notwendigkeit des Vertraulichkeitsschutzes

Im ersten Schritt ist es sinnvoll, die rechtlichen Rahmenbedingungen zu bestätigen, die für den Datensatz und seine gewünschte Anwendung zutreffen. Außerdem wird empfohlen, die vorhandenen Variablen nach Identifikatoren und sensiblen Attributen zu kategorisieren (siehe Kapitel 2.1).

Schritt 2 – Datenvorbereitung

Für die Bestimmung einer angemessenen Anonymisierungsstrategie ist es erforderlich, die Daten in die gewünschte Datenstruktur zu transformieren, zu bereinigen und erste deskriptive Analysen durchzuführen.

Die Vorgehensweise und Best Practices unterscheiden sich dabei nicht wesentlich von der Datenvorbereitung bei klassischen aktuariellen oder Data Science Analysen. Besondere Beachtung verdienen dabei:

- Konsolidierung von Variablen mit ähnlicher Information.
  Die Entfernung oder Konsolidierung von Variablen mit ähnlicher Information verringert das spätere Risiko von Inkonsistenzen bei der Anwendung von Anonymisierungstechniken. Außerdem verringert man das Risiko der Re-Identifikation.
- Analyse von Attributsabhängigkeiten. Es ist wichtig strukturelle Abhängigkeiten zwischen Attributen zu identifizieren (z. B. zwischen Alter und Berufsstatus), da diese das Identifizierungsrisiko erhöhen können. Die angewandten Anonymisierungstechniken müssen auch sicherstellen, dass diese Abhängigkeiten erhalten bleiben und keine Inkonsistenzen in den anonymisierten Daten entstehen.

Außerdem werden direkte Identifikatoren entfernt.

Schritt 3 – Art der Freigabe

Die Anonymisierungsstrategie hängt auch von der Art der Datenfreigabe, bzw. -veröffentlichung ab. Insbesondere sollte hier zwischen einer unternehmensinternen und einer öffentlichen Datenfreigabe unterschieden werden.
Die Art der Freigabe beeinflusst das Risiko der Verknüpfung mit anderen Datenquellen und, allgemeiner, das Risiko von Re-Identifikations-Angriffen.

Als Teil dieses Schrittes können auch Möglichkeiten eruiert werden, das Risiko zu reduzieren. Z. B.

- Beim Teilen des Datensatzes innerhalb des Unternehmens den Zugriff auf Teams begrenzen die keinen Zugriff auf andere sensible personenbezogene Datensätze besitzen
- Anstelle einer Datenübermittlung an ein anderes Unternehmen (z. B. Berater oder Rückversicherer) könnte der Zugriff in einem Data Center innerhalb des eigenen Unternehmens erfolgen. Dies würde das Kombinieren mit anderen Datenquellen verhindern.
- Das Teilen mit externen Firmen oder Organisationen könnte auf berechtigte Personen begrenzt werden, die vor dem Datenzugriff eine Vereinbarung unterschreiben müssen. Diese könnte zum Beispiel die Anwendungen des Datensatzes limitieren und den Versuch einer Re-Identifikation verbieten

Es empfiehlt sich die definierte Zielgruppe der Datenfreigabe zu dokumentieren, um zu verhindern, dass der Datensatz später an andere Gruppen weitergeleitet wird.

Schritt 4 – Angriffsszenarien

Nach Festlegung der Art der Datenfreigabe ist es hilfreich, mögliche Angriffsszenarien zu identifizieren. Eine Übersicht von Angriffsszenarien hilft bei der Ausarbeitung einer Anonymisierungsstrategie. Diese unterstützt bei der Festlegung und Priorisierung der *Quasi-Identifikatoren* und der Bestimmung des gewünschten Sicherheitsniveaus. Die Priorisierung basiert auf der Anfälligkeit der Attribute für Angriffe und einhergehende Risiken.

Insbesondere bei einer extern Datenveröffentlichung ist es wichtig, den Einsatz von externen Datenquellen durch einen Angreifer zu berücksichtigen. Der Aufwand eines Angriffes hängt dabei auch von den nationalen Gegebenheiten ab und wie einfach verschiedene Datensätze zugänglich sind.

Eine umfangreiche Übersicht möglicher Angriffsszenarien findet sich in Kapitel 2.4.

Schritt 5 – Datenverwendung

Eine Anonymisierungsstrategie ist immer ein Kompromiss zwischen einem geringen Risiko einer Re-Identifikation auf der einen Seite und einem hohen Nutzen der anonymisierten Daten auf der anderen Seite. Es ist wichtig, die geplante Anwendung der Daten durch ihre Nutzer zu verstehen, um einen größtmöglichen Nutzen durch die Auswahl von geeigneten Anonymisierungstechniken

zu gewährleisten. Idealerweise erhält man vom Nutzer eine Definition der gewünschten Datengranularität pro Variable zusammen mit einer Minimum-Granularität die nicht unterschritten werden sollte.

Bei der Generalisierung von Quasi-Identifikatoren ist es wichtig, sich bestehenden Konventionen anzulehnen. Bei aktuariellen Anwendungen ist es zum Beispiel üblich, Alter in 5-Jahres Bänder zu gruppieren. Eine Anonymisierung in 3 Jahres Bänder würde den Datennutzen stark mindern, da trotz größerem Informationsdetail die Ergebnisse nicht für weitere Prozesse verwendbar wären.

Ähnlich kann die geplante Verwendung der Daten nicht-prioritäre Datenelemente identifizieren, die als Teil einer *Suppression* entfernt, werden können. Oft werden Datensätze in regelmäßigen Abständen produziert. Dann ist es wichtig, sicherzustellen, dass die Anonymisierungsmethoden konsistent zu den früheren Datensätzen sind.

Kapitel 6 stellt verschiedene Masse vor, mit denen der Nutzen für den Anwender gemessen werden kann.

Schritt 6 – Messung von Anonymität

Basierend auf den gesammelten Informationen der vorhergehenden Schritte wird nun mit Hilfe von geeigneten Anonymisierungskriterien (vgl. Kapitel 4) das Risiko einer Re-Identifikation berechnet. Der Zielwert des Anonymitätsniveaus, z. B. bei *k-Anonymität*, sollte dabei im Verhältnis zu der Art der Datenfreigabe (Schritt 3) und den identifizierten Angriffsszenarien (Schritt 4) stehen.

Es ist auch wichtig zu berücksichtigen, dass die Metriken auf dem Datensatz berechnet werden und nicht bezogen auf die Gesamtbevölkerung. Sie stellen daher oft ein Worst-Case-Szenario dar mit einem tatsächlichen Risikoniveau unter dem berechneten.

Schritt 7 – Messung des Datennutzen

Um den Informationsverlust der Anonymisierung berechnen zu können, wird erst der Datennutzen der Originaldaten berechnet. Dafür werden die im Schritt 5 gewählten Maße verwendet.

Schritt 8 – Wahl und Anwendung von Anonymisierungsverfahren

Die Wahl der Anonymisierungsverfahren hängt von mehreren Faktoren ab, u. a. von

- dem Grad des Identifikationsrisikos (Schritt 6) und wie stark der Datensatz anonymisiert werden muss,
- der Datenstruktur und dem Typ der Quasi-Identifikatoren (kategoriell/ numerisch),
- dem Verlust von Datennutzen durch die Methoden.

Es empfiehlt sich, verschiedene Verfahren auszuprobieren und deren Auswirkungen auf die Anonymisierungsmetriken und Nutzenmaße zu messen und zu vergleichen. Meist bietet es sich an, unnötige Variablen zu entfernen und kategorielle Variablen zu generalisieren, bevor die *Suppression-Verfahren* angewendet werden.

Perturbative Methoden benötigen oft keine vorherige *Generalisierung* oder *Suppression* und können als alternative Methoden getestet werden. Diese Methoden führen zu einer stärkeren Verzerrung der ursprünglichen Datenstruktur, sind aber manchmal die einzige Lösung, insbesondere bei Datensätzen mit vielen *Quasi-Identifikatoren*.

(Templ, 2017, S. 182) gibt in einem vereinfachten Workflow praktische Hinweise zur geeigneten Wahl und Reihenfolge von Anonymisierungstechniken.

Schritt 9 – Neubewertung der Anonymität

Der Grad der Anonymität wird wie in Schritt 6 neu berechnet. Spätestens nun empfiehlt es sich, den Datensatz auf Teilmengen oder einzelne Individuen mit seltenen Eigenschaften zu untersuchen. Diese stellen oft ein erhöhtes Risiko der Re-Identifikation dar, welches durch gezielte Anwendung von lokalen („cell“) Suppressions- oder Generalisierungsmethoden reduziert werden kann.

Falls das Risiko das gewünschte Sicherheitsniveau übersteigt, sollen Schritte 8 -10 mit anderen Parametern oder Methoden wiederholt werden.

Bei der Anwendung von perturbativen Methoden muss überprüft werden, ob die gewählte Anonymitätsmetrik weiterhin angebracht ist. Das Ziel von perturbativen Methoden ist es, die Unsicherheit des Datensatzes zu erhöhen und nicht die Datengranularität zu reduzieren. Metriken die die Frequenz von Quasi-Identifikatoren messen, wie z. B. k-Anonymität, können daher das Risiko überschätzen und sind nur eingeschränkt nutzbar.

Schritt 10 – Neubewertung des Datennutzen

Die Metriken von Schritt 7 werden neu berechnet und mit den Werten der Originaldaten verglichen. Abgesehen von den Durchschnittswerten ist es wichtig, die zugehörigen Konfidenzintervalle zu vergleichen.

Falls der Informationsverlust zu hoch ist und die Minimumanforderungen der Nutzer nicht erfüllt, dann sollten Schritte 8–10 mit anderen Parametern oder Methoden wiederholt werden.

Schritt 11 – Datenprüfung und Dokumentation

Nach Anwendung der Anonymitätsmethoden ist es wichtig, die Datenstruktur zu überprüfen und sicherzustellen, dass die Beziehungen zwischen den verschiedenen Variablen (siehe Schritt 2) weiterhin konsistent sind.

Bei Anwendung von perturbativen Methoden ist es außerdem wichtig, die Verteilung der Variablen auf Anomalien hin zu untersuchen, wie z. B. negative Werte bei Gehalt.

Falls frühere Versionen des Datensatzes freigegeben wurden, ist es geraten, Abweichungen zwischen den beiden zu minimieren. Z. B. bei einer jährlichen Freigabe der letzten 5 Jahre könnte man überprüfen, dass die Durchschnittswerte der überschneidenden Jahre vergleichbar sind.

Eine ausreichende Dokumentation für interne und externe Zwecke ist geboten.
Die interne Dokumentation sollte den angewendeten Anonymisierungsprozess abdecken, inklusive der Anonymisierungsmethoden, zugehörigen Parameter und den Risikometriken vor und nach Anonymisierung. Dies erlaubt ein späteres Reproduzieren der Anonymisierung und kann bei einer Überprüfung durch Aufsichtsbehörden als Beleg der ausreichenden Anonymisierung dienen.

Die externe Dokumentation informiert Nutzer darüber, dass die Daten anonymisiert wurden und beschreibt die einhergehenden Einschränkungen in der Nutzung und Analyse der Daten. Eine grobe Beschreibung der Anonymisierungsmethoden kann hilfreich sein, insbesondere bei Anwendung von perturbativen Methoden. Es sollte sichergestellt werden, dass Metadaten des Datensatzes (Variablenbeschreibung, Labels) weiterhin aktuell sind.

Schritt 12 - Datenfreigabe

Als letzter Schritt erfolgt die tatsächliche Datenfreigabe, im Einklang mit der gewählten Art der Freigabe in Schritt 3.

9. Zusammenfassung und Fazit

Die Anonymisierung von Daten stellt trotz der Prominenz des Themas, welches sich nicht zuletzt aus der DSGVO ergibt, nach wie vor eine große Herausforderung für Unternehmen der Versicherungsbranche dar. So fehlt es bisher an Branchenstandards, die einen einheitlichen Rahmen für den Umgang mit Daten in Versicherungsprozessen definieren.

Gleichzeitig steigt der Bedarf an Datenaustausch sowohl abteilungsübergreifend und konzernweit als auch in Zusammenarbeit mit externen Partnern und Dienstleistern kontinuierlich. Um diesen Anforderungen gerecht zu werden und gleichzeitig ein hohes Maß an Datenschutz zu gewährleisten, sind effektive und praktisch umsetzbare Lösungen gefragt.

Die Analyse verschiedener Anonymisierungstechniken zeigt deutlich: Es gibt keine universelle Lösung für alle Anwendungsfälle. Vielmehr bedarf es meist einer durchdachten Kombination verschiedener Techniken und Metriken. Grundlegend ist dabei das Verständnis allgemeiner Risikoszenarien sowie die sorgfältige Prüfung der spezifischen Anfälligkeit der gegebenen Anwendung. Dies ermöglicht die Entwicklung einer maßgeschneiderten Anonymisierungsstrategie.

In einem iterativen Prozess sollten verschiedene Metriken getestet und die jeweils geeigneten Anonymisierungsmethoden angewandt werden. Nicht-perturbative Methoden stellen in den meisten Fällen eine erste "Verteidigungslinie" dar, sind aber nicht immer ausreichend. Häufig ist eine Kombination mit perturbativen Verfahren notwendig.

Es handelt sich bei der Datenanonymisierung nicht um eine binäre Fragestellung mit einer eindeutigen Lösung. Vielmehr gilt es, auf einem Kontinuum zwischen Datensicherheit und Datennutzen einen optimalen Kompromiss zu finden. Dieser sollte in eine ganzheitliche Strategie eingebettet sein, die auf den identifizierten Risiken basiert und regelmäßig aktualisiert wird.
Die Anonymisierungsstrategie sollte nicht nur regelmäßig auf neue Datenquellen und Angriffstechniken hin überprüft werden, sondern auch auf Veränderungen in der Rechtsauffassung.

Besonders wichtig ist dabei die holistische Betrachtung der beiden Aspekte Datenschutz und Datennutzen. Dies reduziert potenzielle unternehmensinterne Reibungen, die entstehen können, wenn verschiedene Parteien jeweils nur für eine Seite verantwortlich sind. Der vorgestellte SDC-Prozess kann hier als Basis für einen firmeninternen Leitfaden dienen.

Einen vielversprechenden Lösungsansatz für die Zukunft stellen synthetische Daten dar. Diese bieten gerade beim Teilen von Daten mit externen Parteien ein großes Potenzial, da hier oftmals besonders strenge Anforderungen an den Datenschutz gestellt werden. Das Thema wird daher in einer dedizierten neuen Arbeitsgruppe weiter vertieft.
Auch das Konzept der Differential Privacy ist dank eines holistischen und sicheren Lösungsansatzes vielversprechend. Die praktische Umsetzung ist jedoch sehr anspruchsvoll und die Anwendung derzeit begrenzt.

Diese Entwicklungen verdeutlichen, dass das Feld der Datenanonymisierung sich weiter dynamisch entwickelt und kontinuierlich neue Möglichkeiten für den Ausgleich zwischen Datenschutz und Datennutzen entstehen.

Als praktischen Einstieg in die Materie empfehlen wir den GitHub-Account der DAV, wo verschiedene Notebooks mit Beispielen zu Anonymisierungsmetriken und -techniken als auch zur Erstellung von synthetischen Daten zur Verfügung stehen.

10. Literaturverzeichnis

Alle online verfügbaren Quellen wurden zuletzt geprüft am 31.10.2024.

AEPD (Agencia Espanola Proteccion Datos). (27. 04 2021). 10 misunderstandings related to anonymisation. *10 misunderstandings related to anonymisation*. Von https://www.edps.europa.eu/data-protection/our-work/publications/papers/aepd-edps-joint-paper-10-misunderstandings-related_en abgerufen

Aggarwal, C. C. (2008). *A General Survey of Privacy-Preserving Data Mining Models and Algorithms.* In C. C. Aggarwal, & P. S. Yu, Privacy-Preserving Data Mining Models and Algorithms (pp. 11 - 52): Springer.

AI Act. (13. 07 2024). *http://data.europa.eu/eli/reg/2024/1689/oj.* Von Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 an: http://data.europa.eu/eli/reg/2024/1689/oj abgerufen

Aircloak. (2019). Differential Privacy in drei Schwiedigkeitsgraden. Von https://aircloak.com/de/wie-funktioniert-differential-privacy/: https://aircloak.com/de/wie-funktioniert-differential-privacy/ abgerufen.

Apple Inc. (2017). Learning with Privacy at Scale. Von https://machinelearning.apple.com/research/learning-with-privacy-at-scale: https://machinelearning.apple.com/research/learning-with-privacy-at-scale abgerufen.

Apple Inc. (2023). Learning Iconic Scenes with Differential Privacy. Von https://machinelearning.apple.com/research/scenes-differential-privacy: https://machinelearning.apple.com/research/scenes-differential-privacy abgerufen.

Artikel-29-Datenschutzgruppe / 0829/14/DE WP216. (04 2014). *https://www.bing.com/ck/a?!&&p=0de803d97370f64081bff6d93121fbab575a17ada775276b5ad33d63aa177932JmltdHM9MTczNjY0MDAwMA&ptn=3&ver=2&hsh=4&fclid=2bcde95d-b982-62f2-2ccc-fd89b8b96308&psq=Stellungnahme+der+Art.+29-Datenschutzgruppe+von+5%2f2014+zu+Anonymisieru.* Von Stellungnahme 5/2014 zu Anonymisierungstechniken. abgerufen

Bassily, R., & Smith, A. (2015). Local, private, efficient protocols for succinct histograms. Proceedings of the forty-seventh annual ACM symposium on Theory of computing, 127-135.

Benschop, T., Machingauta, C., & Welch, M. (2022). Statistical Disclosure Control for Microdata: A Practice Guide. *International Household Survey Network and the World Bank, Tech. Rep.*

BfDI. (29. 06 2020). *BfDI (Der Bundesbeauftragte für Datenschutz und Informationssicherheit).* Von Positionspapier zur Anonymisierung unter der DSGVO unter berücksichtigung der TK-Branche: https://www.bfdi.bund.de/SharedDocs/Downloads/DE/Konsultationsverfahren/1_Anonymisierung/Positionspapier-Anonymisierung.pdf?__blob=publicationFile&v=6 abgerufen

Bild, R., Kuhn, K. A., & Prasser, F. (2018). Safepub: A truthful data anonymization algorithm with strong privacy guarantees. Proceedings on Privacy Enhancing Technologies.

Bun, M., & Steinke, T. (2016). Concentrated differential privacy: Simplifications, extensions, and lower bounds. *Theory of cryptography conference*, 635-658.

Ciriani, V. D. (2007). *Microdata protection. Secure data management in decentralized systems, S. 291 - 321.*

Cummings, R., & Desai, D. (2018). The role of differential privacy in gdpr compliance. *FAT'18: Proceedings of the Conference on Fairness, Accountability, and Transparency, 20*.

Dalenius, T. (1986). *Finding a Needle In a Haystack or Identifying Anonymous Census Records.* Journal of Official Statistics, Vol.2, No.3, pp. 329–336.

Data Governance Act. (30. 05 2022). *http://data.europa.eu/eli/reg/2022/868/oj.* Von Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA relevance): http://data.europa.eu/eli/reg/2022/868/oj abgerufen

Datenschutz-Grundverordnung. (kein Datum). *Verordnung (EU) 2016/679*. Von https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32016R0679 abgerufen

De Wolf, P.-P., Gouweleeuw, J. M., Kooiman, P., & Willenborg, L. (1999). Reflections on PRAM. *Statistical Data Protection*, 337-349.

Defays, D., & Nanopoulos, P. (1993). Panels of enterprises and confidentiality: the small aggregates method. *Proceedings of the 1992 symposium on design and analysis of longitudinal surveys*, 195-204.

Domingo-Ferrer, J., & Torra, V. (2001). Disclosure control methods and information loss for microdata. *Confidentiality, disclosure, and data access: theory and practical applications for statistical agencies*, 91-110.

Drechsler, J., & Reiter, J. P. (2011). *An empirical evaluation of easily implemented, nonparametric methods for generating synthetic datasets.* Computational Statistics & Data Analysis, 55(12), 3232-3243.

Duchi, J. C., Jordan, M. I., & Wainwright, M. J. (2013). Local privacy, data processing inequalities, and minimax rates. arXiv preprint arXiv:1302.3203.

Dwork, C. (2006). Differential privacy. *International colloquium on automata, languages, and programming*, 1-12.

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science, 9*(3-4), 211-407.

Dwork, C., & Rothblum, G. N. (2016). Concentrated Differential Privacy. arXiv preprint arXiv:1603.01887.

EU-Regulation 2016/679. (2016). *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*.

Europäischer Gerichtshof. (2023). ECLI:EU:C:2023:837. https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX:62022CJ0319.

Fredj, F. B., Lammari, N., & Comyn-Wattiau, I. (2015). Abstracting anonymization techniques: a prerequisite for selecting a generalization algorithm. *Procedia computer science*, 206-215.

Fung, B. C., Wang, K., Chen, R., & Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys (Csur), 42*(4), 1-53.

GDV, G. d. (09. 08 2023). *https://www.bundestag.de/resource/blob/1009184/19d6b96f0b1d92bb12bf57216d8a1c01/20-4-448.pdf.* Von Stellungnahme des Gesamtverbandes der Deutschen Versicherungswirtschaft Lobbyregister-Nr. R000774 zum Regierungsentwurf eines Ersten Gesetzes zur Änderung des Bundesdatenschutzgesetzes: https://www.bundestag.de/resource/blob/1009184/19d6b96f0b1d92bb12bf57216d8a1c01/20-4-448.pdf abgerufen

Giomi, M., Boenisch, F., Wehmeyer, C., & Tasnádi, B. (2022). *A unified framework for quantifying privacy risk in synthetic data.* arXiv preprint arXiv:2211.10459.

Gola, P., & Schomerus, R. (2007). *Bundesdatenschutzgesetz-Kommentar, 9. Auflage.*

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y. (2014). Generative adversarial nets. *dvances in neural information processing systems, 27*.

Gopal, R., Goes, P., & Garfinkel, R. (1999). Confidentiality via camouflage: the CVC approach to database query management. *Proceedings of the Conference on Statistical Data Protection*, 25-27.

Gouweleeuw, J. M., Kooiman, P., Willenborg, L., & De Wolf, P.-P. (1997). *Post Randomisation for Statistical Disclosure Control: Theory and Implementation.* Voorburg: Statistics Netherlands: Research Paper No. 9731.

Grace, P. Z. (2016). *D4.3 - Guidelines for data anonymization report.* OPERANDO.

Hundepool, A., & Willenborg, L. (1998). ARGUS, software packages for statistical disclosure control. *In COMPSTAT: Proceedings in Computational Statistics 13th Symposium held in Bristol, Great Britain*, pp. 341-345.

Inc, A. (2017). Learning with Privacy at Scale. Von https://machinelearning.apple.com/research/learning-with-privacy-at-scale: https://machinelearning.apple.com/research/learning-with-privacy-at-scale abgerufen.

Jordon, J., Yoon, J., & Van Der Schaar, M. (2018). PATE-GAN: Generating synthetic data with differential privacy guarantees. *International conference on learning representations*.

Keppeler, L. P. (2024). *Lockert der EuGH durch sein FIN-Urteil den strengen „Personenbezug "?—Eine kritische Analyse der EuGH-Rechtsprechung anlässlich EuGH, Urt. v. 9.11. 2023–C-319/22, CR 2023, 798.* Computer und Recht, 40(1), 18-22.

Li, N., Li, T., & Venkatasubramanian, S. (2006). t-closeness: Privacy beyond k-anonymity and l-diversity. *IEEE 23rd international conference on data engineering*, 106-115.

Machanavajjhala, A., Kiefer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). l-diversity: Privacy beyond k-anonymity. *Acm transactions on knowledge discovery from data (tkdd), 1*(1), 3-es.

Meindl, M. T. (2008). *Robustification of microdata masking methods and the comparison with existing methods.* Lecture Notes in Computer Science: Springer.

Mincer, J. A. (1974). The human capital earnings function. *Schooling, experience, and earnings*, 83-96.

Muralidhar, K., & Sarathy, R. (2006). Data shuffling—A new masking approach for numerical data. *Management Science, 52*(5), 658-670.

Papernot, N., Abadi, M., Erlingsson, U., Goodfellow, I., & Talwar, K. (2016). *Semi-supervised knowledge transfer for deep learning from private training data.* arXiv preprint arXiv:1610.05755.

Powar, J. &. (2023). *SoK: Managing risks of linkage attacks on data privacy.* Proceedings on Privacy Enhancing Technologies.

Raab, G. M., Nowok, B., & Dibben, C. (2016). Practical data synthesis for large samples. *Journal of Privacy and Confidentiality, 7*(3), 67-97.

Richtlinie 2002/58/EG . (12. 07 2002). *Richtlinie 2002/58/EG des Europäischen Parlaments und des Rates vom 12. Juli 2002 über die Verarbeitung personenbezogener Daten und den Schutz der Privatsphäre in der elektronischen Kommunikation (Datenschutzrichtlinie für elektronische*

Kommunikation). Von http://data.europa.eu/eli/dir/2002/58/oj: http://data.europa.eu/eli/dir/2002/58/oj abgerufen

Richtlinie 95/46/EG Absatz 26. (24. 10 1995). *Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr.* Von http://data.europa.eu/eli/dir/1995/46/oj: http://data.europa.eu/eli/dir/1995/46/oj abgerufen

Riemann, R. (abgerufen am 26.03.2025). *European Data Protection Supervisor*. Von https://www.edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en?etrans=de abgerufen

Ronning, G., & Gnoss, R. (2003). Anonymisierung wirtschaftsstatistischer Einzeldaten: Beiträge zum Workshop am 20./21. März 2003 in Tübingen.

Roßnagel, A., & Scholz, P. (2000). *Datenschutz durch Anonymität und Pseudonymität, Rechtsfolgen der Verwendung anonymer und pseudonymer Daten.* Multimedia und Recht (MMR) 3 (12), 721-729.

Singh, A., Yu, F., & Dunteman, G. (2004). MASSC: A new data mask for limiting statistical information loss and disclosure. *Proceedings of the Joint UNECE/EUROSTAT Work Session on Statistical Data Confidentiality*, 373-394.

Srivastava, A., Valkov, L., Russel, C., Gutmann, M. U., & Sutton, C. (2017). Veegan: Reducing mode collapse in gans using implicit variational learning. *Advances in neural information processing systems, 30*.

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems, 10*(5), 557-570.

Templ, M. (2017). *Statistical disclosure control for microdata.* Cham: Springer.

Tinnefeld. (2003). *in: Roßnagel, Handbuch Datenschutzrecht.*

U.S. Census Bureau. (2021a). Disclosure Avoidance for the 2020 Census. Von https://www2.census.gov/library/publications/decennial/2020/2020-census-disclosure-avoidance-handbook.pdf: https://www2.census.gov/library/publications/decennial/2020/2020-census-disclosure-avoidance-handbook.pdf abgerufen.

U.S. Census Bureau. (2021b). Census Bureau Sets Key Parameters to Protect Privacy in 2020 Census Results. Von https://www.census.gov/newsroom/press-releases/2021/2020-census-key-parameters.html: https://www.census.gov/newsroom/press-releases/2021/2020-census-key-parameters.html abgerufen.

U.S. Census Bureau. (2023). Disclosure Avoidance and the 2020 Census. Von https://www2.census.gov/library/publications/decennial/2020/census-briefs/c2020br-04.pdf:

https://www2.census.gov/library/publications/decennial/2020/census-briefs/c2020br-04.pdf abgerufen.

Vassilev, A., Oprea, A., Fordyce, A., & Anderson, H. (2024). *Adversarial machine learning: A taxonomy and terminology of attacks and mitigations.* o. NIST Artificial Intelligence (AI) 100-2 E2023: National Institute of Standards and Technology. doi:doi.org/10.6028/NIST.AI.100-2e2023

Vorlage zur Vorabentscheidung – Verarbeitung personenbezogener Daten – Richtlinie 95/46/EG – Art. 2 Buchst. a – Art. 7 Buchst. f – Begriff ‚personenbezogene Daten' – Internetprotokoll-Adressen – Speicherung durch einen Anbieter von Online-Mediendiensten –, C-582/14 (URTEIL DES GERICHTSHOFS 19. 10 2016).

Weitzenboeck, E. M., Lison, P., Cyndecka, M., & Langford, M. (2022). The GDPR and unstructured data: is anonymization possible? *nternational Data Privacy Law, 12*(3), 184-206. doi:https://doi.org/10.1093/idpl/ipac008

Xu, L., Jiang, C., Wang, J., Yuan, Y., & Ren, Y. (2014). Information security in big data: privacy and data mining. *Ieee Access, 2*, 1149-1176.

Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling tabular data using conditional gan. *Advances in neural information processing systems, 32*.

Yancey, W. E. (2002). *Disclosure risk assessment in perturba-tive microdata protection (pp. 135-152).* Springer Berlin Heidelberg.

Zhang, J., Cormode, G., Procopiuc, C. M., Srivastava, D., & Xiao, X. (2017). Privbayes: Private data release via bayesian networks. *ACM Transactions on Database Systems (TODS), 42*(4), 1-41.