ANNO ACCADEMICO 2024/2025

Etica, Società e Privacy

Privacy

Altair's Notes

DIPARTIMENTO DI INFORMATICA

Premessa

Licenza

Formato utilizzato

Box di "Corollario":

Corollario 0.0.1 Nome del corollario

Testo del corollario. Per corollario si intende una definizione minore, legata a un'altra definizione.

Box di "Definizione":

Definizione 0.0.1: Nome delle definizione

Testo della definizione.

Box di "Domanda":

Domanda 0.1

Testo della domanda. Le domande sono spesso utilizzate per far riflettere sulle definizioni o sui concetti.

Box di "Note":

Note:-

Testo della nota. Le note sono spesso utilizzate per chiarire concetti o per dare informazioni aggiuntive.

Box di "Osservazioni":

Osservazioni 0.0.1

Testo delle osservazioni. Le osservazioni sono spesso utilizzate per chiarire concetti o per dare informazioni aggiuntive. A differenza delle note le osservazioni sono più specifiche.

"Ultimately, arguing that you don't care about the right to privacy because you have nothing to hide is no different than saying you don't care about free speech because you have nothing to say."
- Edward Snowden

1 Introduzione

1.1 Il Corso in Breve...

Obiettivi:

- Riconoscere problemi di privacy nella modellazione e nell'analisi dei dati.
- Conoscenza di base su metodi per preservare la privacy.

Syllabus:

1. Il concetto di privacy e le leggi sulla privacy in differenti paesi.
2. Le sfide della privacy nell'era dei Big Data.
3. Sistemi di informazione.
4. Modelli statistici.
5. Attacchi alla privacy e modelli di anonimizzazione in database statistici.
6. Privacy differenziale.
7. Separazione dei dati.

1.2 Privacy e Leggi sulla Privacy

1.2.1 Che Cos'è la Privacy?

Domanda 1.1

Che cos'è la privacy?

Definizione 1.2.1: Privacy 1

La *privacy* può essere definita come il diritto a stare da soli.

Warren and Brandeis (1890), *The Right to Privacy*:

- Una delle tesi più influenti nella storia americana.
- GLi autori tentarono di trovare un modo di descrivere legalmente la privacy.

Note:-

Esempio: il diritto di una persona a scegliere la seclusione dalle attenzioni altrui, il diritto di non essere osservati nella sfera privata.

Definizione 1.2.2: Privacy 2

La *privacy* può essere definita come accesso limitato alle informazioni.

- Una persona deve essere libera di scegliere in che misura partecipare alla società senza che gli altri debbano sapere.
- Godkin (1880): “nothing is better worthy of legal protection than private life, or, in other words, the right of every man to keep his affairs to himself, and to decide for himself to what extent they shall be the subject of public observation and discussion.”.
- Bok (1989): la privacy è "the condition of being protected from unwanted access by others—either physical access, personal information, or attention.".

Definizione 1.2.3: Privacy 3

La *privacy* può essere vista come controllo sull’informazione.

- Westin and Blom-Cooper (1970): "privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.“.
- Fried (1968): "Privacy is not simply an absence of information about us in the minds of others; rather it is the control we have over information about ourselves.”.

Definizione 1.2.4: Privacy FIPS PUB 41

Il diritto di un’entità ... a determinare il grado con il quale interagire con il proprio ambiente, compreso il grado con cui un’entità voglia condividere informazioni personali con gli altri.

Definizione 1.2.5: Privacy ISO

Il diritto di un individuo a controllare o influenzare quali informazioni collegate a loro possono essere collezionate e salvate e da chi e a chi queste informazioni possono essere accedute.

Osservazioni 1.2.1

Definizioni derivabili:

- La privacy è l’abilità di una persona di controllare la disponibilità di *informaziioni* e la sua *esposizione*.
- È collegata a essere abili a funzionare in una società *anonimamente*.

Edward Snowden files: viene pubblicato il file contente le informazioni riguardo i dati raccolti dal NSA riguardo le chiamate di milioni di privati cittadini.

1.2.2 Perché la Privacy è Così Importante?

Domanda 1.2

Perché la privacy è così importante?

La privacy è importante perché:

- *Sentimenti individuali:*
  - Non confortabile: possesso dell'informazione.
  - Non sicuro: l'informazione può essere usata in modo improprio (furto d'identità).
- *Le aziende hanno bisogno di:*
  - Far si che i loro clienti si sentano al sicuro.
  - Mantenere una buona reputazione[1].
  - Proteggere sé stesse da ogni disputa legale.
  - Obbedire alle leggi.

Tipi di privacy:

- Political privacy.
- Consumer Privacy.
- Medical privacy.
- Private property.
- Information/Data privacy.

Definizione 1.2.6: Data Privacy

Il problema della *data privacy* emerge quando dei dati unicamente associabili a un individuo sono collezionati e salvati.

Le fonti più comuni di dati affetti da data privacy:

- Record sanitari.
- Investigazioni e processi criminali.
- Transazioni finanziarie.
- Tratti biologici (e. g. materiale genetico).
- Residenza e posizione geografica.
- Geolocalizzazione.
- Utilizzo del web.

Osservazioni 1.2.2

- La sfida della data privacy è trovare un modo per *condividere dati* proteggendo le informazioni che permettono di identificare una determinata persona.
  - Per esempio negli ospedali i dati vengono trasferiti in maniera aggregata.
  - L'idea di condividere i dati in forma aggregata garantisce che solo dati non identificabili sono condivisi.
- La protezione legale deò diritto alla privacy *cambia drasticamente a seconda della nazione*.

[1] Beh, visto che merda sono Twitter e META non credo gliene freghi qualcosa.

1.2.3 Privacy negli Stati Uniti

La data privacy *non è molto regolata* negli USA:

- Non c'è una legge che controlli tutti gli aspetti dei dati (acquisizione, collezione, trattamento e uso).
- Se un'azienda colleziona dei dati (anche senza permesso) ha il diritto di utilizzarli.
- Gli istituti possono informarsi sulle condizioni finanziarie di una persona (banche, assicurazioni, etc.) chiedendo report a terze parti.
- Ci sono alcune eccezioni:
  - Dati sanitari (HIPAA).
  - Dati dei bambini sotto i 13 anni online (COPPA).
  - Richieste di prestito (FCRA).
  - Sicurezza informatica (ECPA, PATRIOT[2], etc.).

1.2.4 Privacy nell'Unione Europea

La data privacy nell'UE *è pesantemente regolata*:

- l'articolo 8 della convenzione europea sui diritti umani (ECHR) prevede il diritto al rispetto della privacy di una persona (poter disporre di una sfera privata, una vita familiare, un domicilio e della propria corrispondenza).
- La corte dei diritti umani ha dato a quest'articolo varie interpretazioni, con le seguenti eccezioni:
  - Ottenere informazioni per censimenti ufficiali.
  - Raccogliere impronte digitali e fotografie per attività di polizia.
  - Collezionare dati medici e spese personali.
  - Implementare sistemi di identificazione personale (un documento di identificazione).

Note:-

Spesso questa maggiore tutela della privacy viene criticata dalle aziende che la vedono come un freno al progresso.

Ci sono state due ere della privacy:

- 1995-2018: Data Protection Directive (DPD).
- 2018-presente: General Data Protection Regulation (GDPR[3]).

Definizione 1.2.7: Data Protection Directive

La Data Protection Directive (DPD) fu adottata nel 1995 dal parlamento europeo:

- Come tentativo di armonizzare la protezione dei dati nell'unione europea.
- Doveva essere regolamentata da leggi nazionali nel 1998.

[2] Nel 2001, in seguito a un certo evento terroristico.
[3] Promulgato nel 2016, ma diventato effettivo dal maggio del 2018

Gli 8 princìpi di base del DPD:

1. I dati personali devono essere processati a norma di legge e in maniera corretta.
2. I dati personali devono essere processati solo per certi scopi limitati.
3. I dati personali devono essere processati in modo adeguato, rilevante e non eccessivo.
4. I dati personali devono essere processati accuratamente.
5. I dati personali devono essere trattenuti solo per il tempo strettamente necessario.
6. I dati personali devono essere processati in accordo con i diritti del soggetto interessato.
7. I dati personali devono essere processati in modo sicuro.
8. I dati personali devono essere trasferiti solo a nazioni con una protezione adeguata.

Definizione 1.2.8: Personal Data - DPD

Ogni informazione relativa a una persona identificata i identificabile. Una persona identificabile è una persona che può essere identificata, direttamente o indirettamente, in particolare identificando un particolare numero di identificazione o uno o più fattori fisici, psicologici, mentali, economici, culturali o sociali.

Note:-

Questa definizione viene rafforzata nel GDPR.

Definizione 1.2.9: Data Processing - DPD

Ogni operazione o insieme di operazioni che viene effettuata su dati personali, tramite un mezzo automatizzato o meno, come la collezione, registrazione, organizzazione, immagazzinamento, adattamento o alterazione, recupero, consultazione, uso, trasmissione, disseminazione o altra distribuzione, allineamento o combinazione, blocco, cancellazione o distruzione.

Note:-

Vengono effettuati cambiamenti minori nel GDPR.

Definizione 1.2.10: Responsabile - DPD

Il responsabile può essere una persona fisica o giuridica, un autorità pubblica o un ente che deve garantire l'integrità dei dati trattati.

Corollario 1.2.1 Processore

Il processore dei dati è una persona fisica o giuridica, un'autorità pubblica, un'agenzia o qualunque altro ente che processa i dati personali per conto del responsabile.

Note:-

Non ci sono cambiamenti nel GDPR.

Princìpi della gestione dei dati. I dati personali non dovrebbero essere processati a meno che non riguardino queste categorie:

- *Trasparenza*: il soggetto dei dati ha il diritto di essere informato da un'azienda che sta elaborando i suoi dati personali.
- *Scopo legittimo*: i dati personali possono essere processati solo per scopi legittimi e non usati per altri scopi non pertinenti.
- *Proporzionalità*: i dati personali processati devono essere adeguati, rilevanti e non eccessivi in relazione allo scopo per quale i dati vengono collezionati e ulteriolmente processati.

Trasparenza:

- Il soggetto dei dati deve dare il *consenso*.
- Il processamento era necessario per l'*esecuzione di un contratto*.
- Il processamento era necessario per *assolvere un obbligo legale*.
- Il processamento era necessario per *proteggere gli interessi vitali* del soggetto dei dati.
- Il processamento era necessario per un'operazione di *interesse pubblico*.
- Il processamento era necessario per *scopi di interesse legittimato dal controllore, eccetto quando quegli interessi sono sovrascritti dagli interessi dei diritti fondamentali e della libertà del soggetto*.

Osservazioni 1.2.3

Il soggetto ha il diritto a:

- Accedere a tutti i dati su di lui/lei/loro.
- Domandare la rettifica, cancellazione o blocco dei dati se sono incompleti, inaccurati o non sono processati come previsto dalle regole sulla protezione dei dati.

Proporzionalità:

- I dati devono essere accurati e, quando necessario, aggiornati.
- I dati non devono essere mantenuti in una forma che permette l'identificazione del soggetto per più tempo del necessario.
- Gli stati membri hanno la possibilità di memorizzare i dati personali per fini statistici o scientifici.
- Vengono applicate tutele aggiuntive per la gestione dei *dati sensibili* (credenze religiose, opinioni politiche, salute, orientamento sessuale, gruppo etico, appartenenza a organizzazioni).
- Il soggetto può sempre chiedere la cancellazione dei suoi dati usati per fini pubblicitari[4].
- Qualunque decisione automatizzata sulla persona non deve essere fatta in maniera completamente automatica.
- L'individuo può *fare ricorso* su qualsiasi decisione automatica in cui vengono processati i propri dati.

Autorità della privacy:

- Ogni stato membro deve eleggere un'*autorità di supervisione* che:
  - Deve *monitorare la protezione dei dati* in quello stato membro.
  - Dare *avvisi al governo riguardo le misure amministrative e i regolamenti*.
  - *Iniziare procedure legali* quando il regolamento sulla protezione dei dati viene violatao.
- Il *controllore* deve notificare all'autorità di supervisione le seguenti informazioni:
  - Il nome e l'indirizzo del controllore.
  - Lo scopo del processamento.
  - Una descrizione delle categorie dei dati del soggetto.
  - Il recipiente a cui i dati possono essere inotrati.
  - Proposte di trasferimento dei dati a stati terzi.
  - Una descrizione generale delle misure prese per garantire la sicurezza dei dati processati.
- Le informazioni sono tenute in un *registro pubblico*.

[4] Contrariamente agli stati uniti.

Leggi della privacy in Italia:

- Il DPD è stato implementato con il decreto legislativo 196/2003 (Codice in materia di protezione dei dati personali).
- Inoltre il Garante per la protezione dei dati limitati doveva applicare misure appropriate in determinati campi (video sorveglianza, dati biometrici, dati sanitari, notifiche di data breach, informazioni bancarie, profili online, processamenti fatti da amministratori di sistema, processamenti a fini di marketing e profiling, pagamenti elettronici, cookies).

Definizione 1.2.11: General Data Protection Regulation

La General Data Protection Regulation (GDPR) è un regolamento attraverso il quale l'unione europea intende aumentare la forza e unificare la protezione dei dati per tutti gli individui all'interno dell'unione europea.

Note:-

L'obiettivo principale del GDPR è quello di offrire un controllo semplificato agli appartenenti ai paesi membri dell'unione europea.

Definizione 1.2.12: Personal Data - GDPR

Uguale al DPD, ma aggiunge come caratteristiche di identificazione i dati di locazion, gli identificatori online, lo stato genetico, economico e culturale.

Definizione 1.2.13: Data Processing - GDPR

Uguale al DPD, ma viene inclusa la "strutturazione" dei dati processati.

Note:-

Inoltre il GDPR include altre definizioni.

Definizione 1.2.14: Pseudoanimizzazione

La pseudoanimizzazione è il processamento in cui i dati personali non possono più essere attribuiti univocamente a un soggetto senza informazioni aggiuntive. Le informazioni aggiuntive sono tenute separate e soggette a misure tecniche e organizzative per assicurare che i dati siano anonimi.

La pseudoanimizzazione:

- Se viene effettuata con politiche adeguate non è soggetto a controlli e penalità.
- La regolamentazione non coinvolge dati usati per statistiche o ricerche.
- Le politiche e le misure che raggiungono la privacy by Design e la privacy by Default sono adeguati.

Definizione 1.2.15: Personal Data Breach

Un leak accidentale nella sicurezza per cui il dato personale viene distrutto, perso, alterato, rubato o acceduto.

Diritti del soggetto:

- L'individuo deve dare un *consenso chiaro* per il trattamento dei propri dati.
- Il soggetto ha un *facile accesso* ai propri dati.
- Viene evidenziato il diritto alla rettifica, alla cancellazione e all'oblio[5].

[5] Cosa non facile, soprattutto con i social.

- Il diritto di obiezione, incluso l'utilizzo dei propri dati per "profiling".
- Il diritto alla portabilità dei dati da un servizio a un altro.

Privacy *by Design* e *by Default*:

- Viene richiesto che la protezione dati sia presente fin dall'inizio nel sistema informativo.
- Le impostazioni della privacy devono essere di alto livello.
- La privacy deve essere presente per tutto il ciclo di vita del processamento dei dati.
- Come già detto i dati personali devono essere processati solo quando è necessario per ogni specifico scopo.

Definizione 1.2.16: Privacy by Design - Ann Cavoukian

Un approccio all'ingegneria dei sistemi che tiene in considerazione la privacy durante tutto il ciclo di vita. Si basa su 7 princìpi fondamentali:

1. Proattività, non reattività (prevenzione, non rimedio).
2. Privacy come impostazione di Default.
3. Privacy integrata nel Design.
4. Completamente funzionale.
5. End-to-end security.
6. Visibilità e trasparenza.
7. Rispetto per la privacy degli utenti.

Responsabilità:

- Il titolare del trattamento del dato deve dimostrare che tutte le operazioni messe in atto per garantire la sicurezza del dato siano aderenti alla legge.
- È responsabilità del controllore dei dati di implementare misure effettive e di dimostrare la *compliance* alle attività di processamento.
- L'utente deve essere chiaramente informato riguardo le finalità del trattamento, alla legge usata come base, all'intervallo temporale del trattamento, se i dati vengono trasferiti a terze parti e se avvengono delle decisioni automatizzate.
- Viene introdotto il *Data Protection Officer*, un esperto tecnico del dato e legale del GDPR.
- Se avvengono eventi di rischio bisogna fare un *Data Protection Impact Assessments* (DPIA) per valutarne l'estensione.
- La valutazione e la mitigazione del rischio sono necessarie e approvazione delle autorità nazionali di protezione dei dati (DPA) è necessario per rischi elevati.
- I records delle attività di trattamento devono essere conservate in modo includere le finalità del trattamento, le categorie coinvolte e Termini previsti.
- I records devono essere disponibili all'autorità di supervisione, su richiesta.

Definizione 1.2.17: Data Protection Officer

Il GDPR stabilisce la figura del Data Protection Officer (DPO), una persona con conoscenze specialistiche in materia di protezione dei dati e pratiche che dovrebbero aiutare il titolare del trattamento o l'incaricato del trattamento a monitorare conformità interna al presente regolamento. The DPO are also expected to be proficient at managing IT processes, data security (including dealing with cyber-attacks) and other critical business continuity issues around the holding and processing of personal and sensitive data.

Note:-

Autorità pubbliche e imprese le cui attività principali sono trattamento regolare o sistematico dei dati personali, sono necessario per assumere un DPO.

Definizione 1.2.18: Data Breach

Il controllore dei dati ha l'obbligo legale di notificare l'autorità di supervisione senza ritardi non necessari, a meno che sia improbabile che il data breach causi rischi alla libertà e ai diritti dell'individuo. C'è un massimo di 72 ore dopo aver scoperto il data breach per notificare.

Le seguenti sanzioni possono essere imposte:

- Nel primo e non intenzionale caso di noncompliance viene emesso un avviso.
- Una multa fino a 10 milioni di euro o fino al 2% dell'annualità mondiale fatturato dell'esercizio finanziario precedente nel caso di un'impresa, se vi è stata una violazione di alcuni obblighi.
- Una multa fino a 20 milioni di euro o fino al 4% dell'annuo in tutto il mondo fatturato dell'esercizio finanziario precedente nel caso di un'impresa, se vi sono state gravi violazioni dei princìpi.

1.2.5 Regolamenti Transazionali

- Safer Harbor privacy principles (fino al 2015).
- EU-US Privacy Shield (2016-2020[6]).
- EU-US Data Privacy Framework (2023-presente).

Definizione 1.2.19: Safe Harbor

Sviluppato tra il 1998 e il 2000 al fine di impedire alle organizzazioni private di l'UE o gli Stati Uniti dalla divulgazione accidentale o dalla perdita di informazioni personali.

Le compagnie USA che tengono dati devono aderirea questi 7 princìpi:

1. *Notifica*: gli individui devono essere informati che i loro dati stanno venendo collezionati e come verranno usati.
2. *Scelta*: gli individui possono scegliere di rimuovere i propri dati rinunciando ai servizi[7].
3. *Trasferimento*: la trasmissione dei dati a terzi può avvenire solo ad altri organizzazioni che seguono principi di protezione dei dati adeguati.
4. *Sicurezza*: bisogna evitare la perdità di informazioni collezionate.
5. *Data Integrity*: i dati devono essere rilevanti e affidabili per lo scopo per cui sono raccolti.
6. *Accesso*: gli individui devono essere in grado di accedere, correggere e cancellare i propri dati.
7. *Applicazione*: devono esistere mezzi efficaci per far rispettare queste norme.

[6] Caduto a causa del primo governo Trump.
[7] Google merda.

Breve storia di Safe Harbor:

- Nel 2000 la Commissione europea ha deciso che il principi degli Stati Uniti erano conformi alla direttiva dell'UE (la cosiddetta "Decisione Safe Harbor").
- Dopo che un cliente si è lamentato del fatto che i suoi dati di Facebook non erano sufficientemente protetti la corte di giustizia della commissione europea, nel 2015, dichiara la decisione di Safe Harbor invalida.
- La Commissione ha tenuto ulteriori colloqui con gli Stati Uniti autorità competenti verso "un quadro rinnovato e solido per flussi di dati transatlantici".
- La Commissione europea e gli Stati Uniti hanno convenuto di istituire un nuovo quadro per i flussi transatlantici di dati il 2 febbraio 2016, noto come "EU-US Privacy Shield".

Definizione 1.2.20: EU-US Privacy Shield

Lo scudo UE-USA per la privacy è un quadro per gli scambi transatlantici di dati per finalità commerciali tra l'Unione Europea e gli Stati Uniti. Uno dei suoi scopi è quello di consentire alle aziende statunitensi di ricevere più facilmente dati provenienti da entità dell'UE ai sensi delle leggi sulla privacy dell'UE volte a proteggere i cittadini dell'unione europea.

Osservazioni 1.2.4

- La Commissione europea ha adottato il quadro il 12 luglio 2016 ed è entrato in vigore lo stesso giorno.
- Il presidente degli Stati Uniti Donald Trump ha firmato un ordine esecutivo intitolato "Enhancing Public Safety", in cui si afferma che le protezioni della privacy degli Stati Uniti non saranno estese oltre i cittadini o residenti statunitensi.
- Nel luglio 2020 lo scudo UE-USA per la privacy è stato abrogato dalla Corte europea giustizia in quanto non forniva tutele adeguate ai cittadini dell'UE rispetto allo Snooping governativo americano[a].

[a] In poche parole gli americani si comportano da americani.

Definizione 1.2.21: AI Act

La legge sull'IA mira a garantire che i sistemi di IA utilizzati nell'UE siano sicuri, trasparenti e rispettosi diritti fondamentali.

L'AI Act classifica i sistemi di IA in base al livello di rischio che rappresentano:

- *Rischio inaccettabile*: i sistemi di IA che rappresentano una chiara minaccia per la sicurezza o i diritti fondamentali sono vietato. Ciò include i sistemi che implementano tecniche subliminali per manipolare il comportamento, sfruttare le vulnerabilità di gruppi specifici o abilitare il punteggio sociale da parte dei governi.
- *Alto rischio*: sistemi di intelligenza artificiale utilizzati in aree critiche come la sanità, l'istruzione, l'occupazione, il diritto l'applicazione delle norme e i servizi essenziali sono soggetti a obblighi rigorosi. Questi includono rigorosi valutazioni dei rischi, misure di governance dei dati, supervisione umana e monitoraggio continuo per garantire la conformità.
- *Rischio limitato*: i sistemi di intelligenza artificiale a rischio limitato, come i chatbot e i deepfake, sono soggetti a obblighi di trasparenza. I fornitori e i distributori devono informare gli utenti che stanno interagendo con un sistema di intelligenza artificiale.
- *Rischio minimo*: la maggior parte dei sistemi di intelligenza artificiale, come i videogiochi abilitati all'intelligenza artificiale o i filtri antispam, cadono rientrano in questa categoria e sono in gran parte non regolamentate, incoraggiando l'innovazione e lo sviluppo.

Tutti i providers di modelli di AI devono:

- Fornire documentazione tecnica e istruzioni.
- Rispettare la direttiva europea sul copyright.
- Pubblicare un sommario dei dati usati per il training.
- Condurre valutazioni e test (anche di tipi "adversarial"[8].
- Assicurare misure di cybersecurity adeguate.

Note:-

Questo copre la parte legale di assegnamento delle responsabilità in campo giuridico.

Obblighi:

- *Sviluppatori:* entità che sviluppano sistemi ad alto rischio, devono assicurarsi che i datasets siano di alta qualità, mantenere documentazione tecnica e sviluppare sistemi con livelli di sicurezza, robustezza e sicurezza.
- *Utilizzatori:* chi utilizza sistemi ad alto rischio deve monitorare le performance e ridurre/mitigare i rischi con la presenza umana.

Osservazioni 1.2.5

- L'AI act prevede un board europeo che monitori l'intelligenza artificiale.
- Controlla le implementazioni negli stati membri.
- Forzano l'aderenza alla legge.

Definizione 1.2.22: Digital Service Act (2018)

Regola i servizi online. Sono 44 articoli e 128 misure che coprono le seguenti aree:

- Demonetizzazione: riduzione degli incentivi finanziari per leakers e disinformatori.
- Trasparenza nella pubblicità politica.
- Integrità dei servizi: riduzione di account fake, deepfake, amplificazione dovuta a bots, disinformazione, etc.
- Fact-checking[a].

[a]Non se il tuo cognome è Musk.

1.3 La Privacy al Tempo dei "Big Data"

1.3.1 La Privacy sotto Attacco

Nel 1996, il 24% degli americani ha subito un'invasione della sua privacy. Questo tema rientra nel più ampio spettro del tema: privacy vs. convenienza.

$$+ \text{PRIVACY} \Leftrightarrow - \text{COMODITÀ}$$
$$+ \text{COMODITÀ} \Leftrightarrow - \text{PRIVACY}$$

Note:-

Questo trade-off è il risultato di ignoranza e di una politica di conformismo e indifferenza.

[8]Per portare il sistema in errore.

Violazione della privacy, secondo alcune persone:

- Chiamate di basso livello per ottenere l'identificatore univoco nei chip intel.
- Database delle immagini dei guidatori negli USA.
- Utilizzo dei dati di Facebook.
- Scandali del NSA.
- Cambridge Analytica e scandali di Facebook.

Note:-

Negli anni si sta verificando un sempre maggiore raccoglimento di dati personali: carte di credito, registri di chiamate, dati sulla sanità, emails, account social, DNA, etc.

Definizione 1.3.1: Big Data

Il termine Big Data si riferisce all'acquisizione e all'analisi di enormi collezioni di informazioni, talmente grandi che fino a poco tempo fa la tecnologia per analizzarli non esisteva.

Corollario 1.3.1 Legge delle Conseguenze Inattese

La legge delle conseguenze inattese è un fenomeno frequentemente osservato per cui ogni azione ha risultati che non fanno parte delle intenzioni dell'attore. Le conseguenze possono essere positive, neutrali o negative.

1.3.2 Deanonimizzazione

I dati vengono raccolti in modo che siano anonimizzati (vedremo in un successivo capitolo come). Però a volte si può bypassare questa anonimizzazione, com'è successo nel caso di Netflix nel 2006:

- Si potevano incrociare due datasets (uno anonimizzato per una challenge di netflix e l'altro non anonimizzato riguardante i rating dei film).
- Così facendo, conoscendo 6-8 valutazioni di film e date, si potevano identificare unicamente gli utenti con una probabilità del 90%.
- Non si può sempre anonimizzare i dati semplicemente rimuovendo gli identificatori.
- Si ha una vulnerabilità aggregando dati da sorgenti diverse (fig: 1.1).

Examples	*Health-specific and General Examples of Re-identification*
AOL search data	Researchers were capable of revealing sensitive details of the participant's private lives, such as Social Security numbers, credit-card numbers, addresses etc. from the anonymized AOL Internet search data that contains health related searches as well
Chicago homicide database	A large percentage of individuals were re-identified easily by linking the Chicago homicide database with the social security death index
Netflix movie recommendations	Several individuals were re-identified from the publicly available anonymized Netflix movie recommendations database by linking their anonymized movie ratings with ratings in a publicly available Internet movie rating web site
Re-identification of the medical record	Massachusetts governor's sensitive medical records was re-identified by linking the anonymized data of the Group Insurance Commission, which purchases health insurance for state employees, with the voter list for Cambridge
Southern Illinoisan vs. The Department of Public Health	Individuals in a neuroblastoma data set from the Illinois cancer registry was re-identified with a very high accuracy
Canadian Adverse Event Database	An unfortunate death of a 26 year-old student by taking a particular drug was re-identified from the publicly released adverse drug reaction database of Health Canada

Figure 1.1: Esempi di re-identificazione.

1.3.3 Data Breach

Cause comuni di data breaches:

- Credenziali compromesse.
- Servizi malconfigurati.
- Vulnerabilità software.

Note:-

Un caso emblematico italiano è il data breach avvenuto a danni dell'INPS durante il governo Renzi (fig: 1.2).

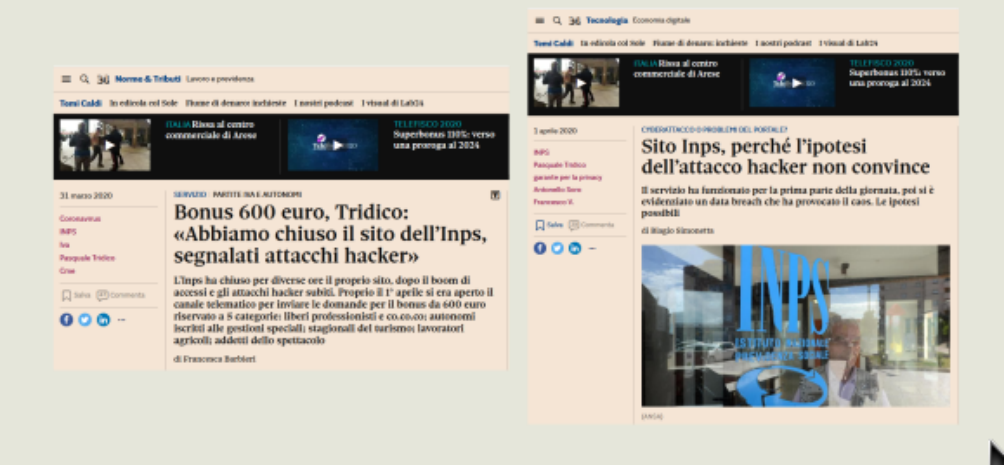


Figure 1.2: INPS data breach.

Domanda 1.3

Ma qual è il costo di un data breach?

- I data breach sono molto costosi, sia in termini di valore dei dati che di credibilità.
- Si stima che il costo per i cittadini dei propri dati sia tra 1000$ e 3000$
- Stefano Rodotà lega il concetto di privacy alla dignità attribuendo un costo immenso alla perdità dei propri dati personali.

2
Sistemi Informativi e Privacy

2.1 Sistemi Informativi

Domanda 2.1

Che cos'è un sistema informativo?

Definizione 2.1.1: Sistema Informativo

Un sistema informativo è un sistema formale, sociotecnico e orgazzativo per collezionare, processare, raccogliere e distribuire informazioni usate da organizzazioni.

Corollario 2.1.1 Sistema Informativo Computazionale

Un sistema informativo computazionale è un sistema composto da persone e computer ce processano o interpretano informazioni.

Lo scopo di un sistema informativo è quello di supportare:

- Le attività (automazione).
- Le decisioni prese dagli alti ranghi di organizzazioni (analisi, controllo, coordinazione, statistica).

I sistemi informativi funzionano grazie agli scambi di dati e informazioni:

- Dati modellati come insiemi di processi interconnessi dove l'output di un processo è l'input di un altro processo.
- I dati appartengono a diverse entità:
  - Impiegati.
  - Clienti.
  - Fornitori.
- Differenti tipi di dati:
  - Identificativi personali.
  - Emails.
  - Dati di processi, transazioni, etc.

Un sistema informativo, per essere conforme al GDPR, deve assicurare:

- *Autorizzazione basata su attributi:* meccanismi per accedere a tutti i dati, sotto determinate circostanze.
- *Anonimizzazione* e *Pseudoanonimizzazione* dei dati: meccanismi per garantire l'anonimato o lo pseudoanonimato.
- *Tracciabilità:* un registro di chi ha creato, modificato o cancellato informazioni, quando e per quale scopo.
- *Cancellazione dei dati:* meccanismi per il diritto all'oblio.

2.1.1 Tecniche di Controllo degli Accessi

Definizione 2.1.2: Controllo degli Accessi

Si restringe l'accesso alle risorse computazionali, specialmente in sistemi multi-utente.

Note:-

I requisiti di privacy e sicurezza devono essere mantenuti nel sistema in modo efficiente. Tuttavia in situazioni di emergenza si possono fare delle eccezioni (e.g. in un ospedale con un paziente in pericolo di vita).

Corollario 2.1.2 Role Based Access Control (RBAC)

Formalizzato da NIST nel 1992. Si gestisce l'accesso in base al ruolo invece che a un identificatore. Il ruolo fornisce un livello di astrazione (come collezione di permessi). Ogni ruolo può essere assegnato a un numero arbitrario di utenti.

Domanda 2.2

Come funziona RBAC (fig: 2.1)?

- Gli *amministratori* assegnano permessi a ogni ruolo.
- I ruoli possono essere assegnati a utenti individuali (e ogni utente può avere più ruoli).
- Gli amministratori possono aggiornare i ruoli aggiungendoli o rimuovendoli da determinati utenti.

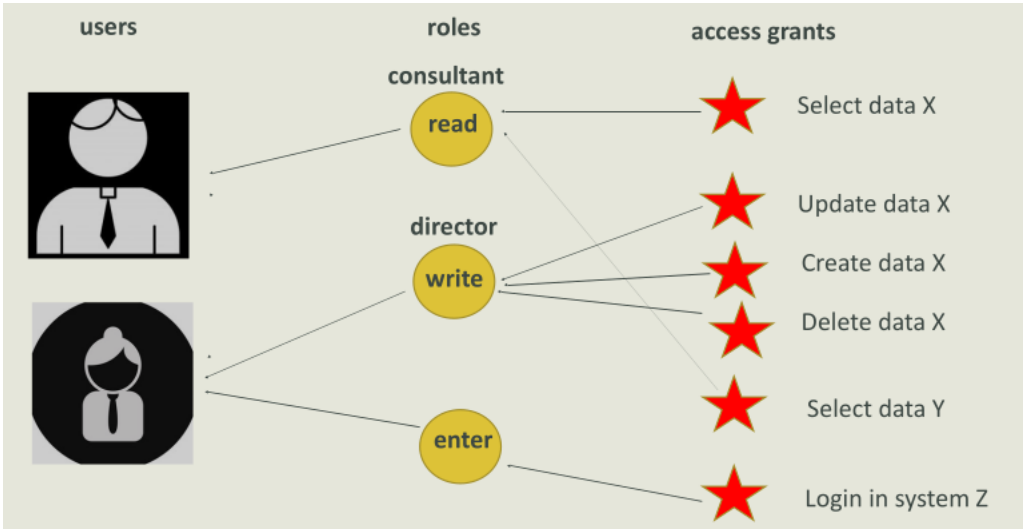


Figure 2.1: RBAC.

Limiti di RBAC:

- Non fornisce un meccanismo flessibili per cui i clienti possano esprimere dei loro requisiti.
- Non cattura lo scopo per cui i dati vengono rilasciati.
- Per cui RBAC si evolve in Attribute Base Access Control (ABAC).

Aggiungendo contesto, le decisioni autorizzative possono essere basate su:

- Ruolo.
- Persone od oggetti collegati all'utente.
- Di che cosa si ha bisogno.
- Dove l'utente vuole accedere.
- Quando l'utente vuole accedere.
- Com'è l'utente vuole accedere a quelle informazioni.

Corollario 2.1.3 Attribute Based Access Control (ABAC)

Definendo un contesto si possono aggiungere adeguate politiche di accesso che definiscono in maniera dichiarativa come debba avvenire l'accesso.

Note:-

Conoscere il ruolo di un utente non è abbastanza per assicurare la sicurezza. Si richiede il contesto e le relazioni tra le varie entità presenti nel sistema.

Caratteristiche di ABAC(fig: 2.2):

- Adotta un approccio *policy driven*.
- Usa gli attributi *soggetto*, *oggetto* e *ambiente*.
- Rimuove la necessità di dover essere registrati in un sistema per essere in grado di accedere a risorse condivise.
- Utilizza un *authorization engine* (fig: 2.3).

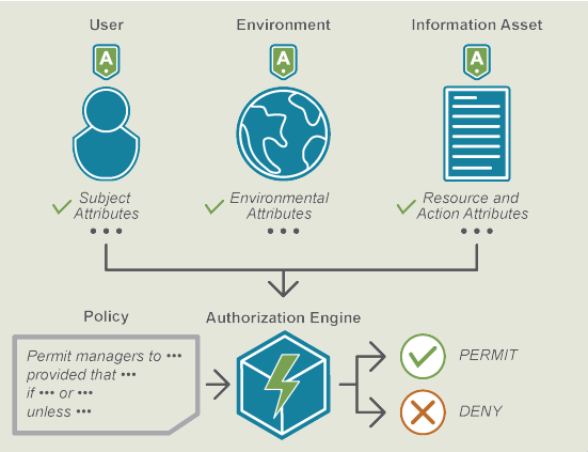


Figure 2.2: ABAC.

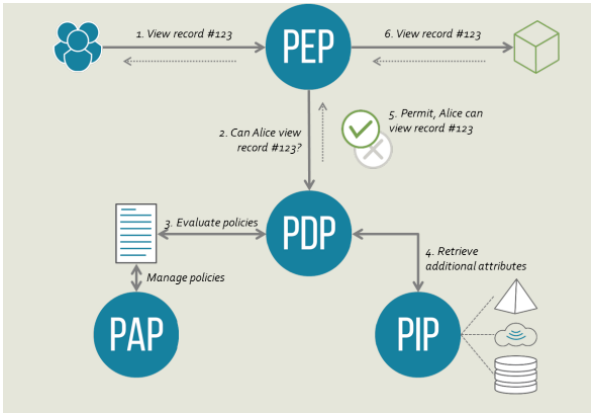


Figure 2.3: Dettaglio dell'authorization engine.

Note:-
Le policies sono espresse in XACML (eXtensible Access Control Markup Language).

2.2 Pseudoanonimizzazione con Crittografia

Definizione 2.2.1: Pseudoanonimizzazione

La pseudoanonimizzazione è una procedura di data management e di deidentificazione per cui informazioni identificabili in un record sono rimpiazzate da uno o più identificatori artificiali (pseudonimi).

Note:-
Un singolo pseudonimo viene utilizzato in modo consistente per permettere un'analisi accurata. Tuttavia questi dati pseudoanonimizzati possono essere riportati al loro stato originale perché viene salvata l'associazione con lo pseudonimo.

Corollario 2.2.1 Crittografia su Colonne

La Crittografia su colonne è una feature per proteggere dati sensibili nei databases. Permette di crittografare dati sensibili senza rivelare la chiave all'engine del database. Questo fornisce una separazione tra:

- Chi possiede i dati e può visualizzarli.
- Chi gestisce i dati, ma non ha il diritto di accedervi.

Ci sono due tipi di chiavi:

- *Column encryption keys:* per crittografare il dato.
- *Column master keys:* per crittografare la column encryption key.

Note:-
Le chiavi non sono mai memorizzate nei meta dati, ma vengono salvate in un repository esterno.

Crittografia deterministica vs. randomizzata:

- *Crittografia deterministica:* genera sempre gli stessi valori per ogni testo.
- *Crittografia randomizzata:* utilizza valori casuali. È più sicura, ma impedisce le ricerche sul database.

Rivest-Shamir-Adleman (RSA):

- È un sistema crittografico a chiave pubblica.
- Gli utenti creano e pubblicano una chiave pubblica basata su due numeri primi sufficientemente grandi segreti.
- I messaggi possono essere crittati da chiunque usando la chiave pubblica.
- Possono essere decrittati solo da chi conosce la chiave privata.
- Utilizza anche un sistema di padding.

TDE (Oracle):

- Si usa sempre una sola chiave di crittografia.
- Nessuna chiave è salvata in chiaro.
- Il repository esterno di Oracle è chiamato Oracle wallet.
- Vengono separate le responsabilità per prevenire accessi illeciti.

2.2.1 Tracciabilità mediante Auditing

Definizione 2.2.2: Auditing

L'auditing è il processo di esaminazione e validazione di documenti, dati, processi, procedure, sistemi.

Corollario 2.2.2 Audit Log

Documento in cui si memorizzano in ordine cronologico tutte le attività di interesse.

Corollario 2.2.3 Audit Objectives

Insieme di business rules, controlli di sistema, regolazioni di governi, policies di sicurezza.

Altre definizioni:

- *Auditor:* persona autolizzata a fare audit.
- *Audit procedure:* insieme di istruzioni per il processo di auditing.
- *Audit report:* documento contenente tutti i risultati dell'audit.
- *Audit trail:* record cronologico di cambiamenti su documenti e dati, attività di sistema od operazioni.

Tipi di audit:

- Interno: esaminazione di attività condotte da membri dello staff.
- Esterno: esaminazione di attività condotte da terze parti.

Attività di auditing (fig: 2.4):

- Identificare problemi di sicurezza.
- Stabilire piani, politiche e procedure.
- Assicurarsi che gli elementi contrattuali siano riscontrati.
- Organizzare incontri tra team di verifica.

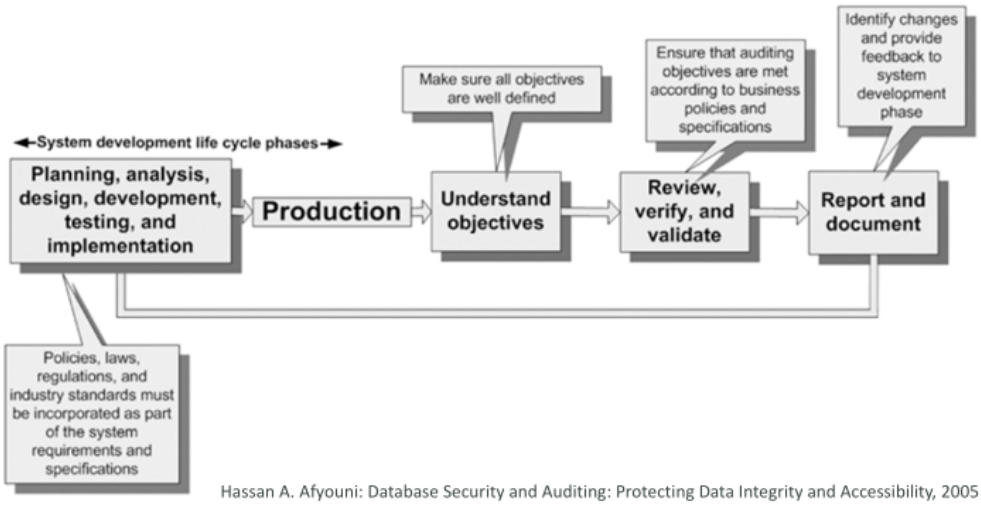


Figure 2.4: Processo di auditing.

Modelli di auditing:

- *Modello 1:*
  - Registra le entità in un repository.
  - Tiene traccia delle attività performate.
- *Modello 2:*
  - Memorizza solo i valori delle colonne.
  - C'è un meccanismo di archiviazione.
  - Non registra le azioni performate sui dati.
  - Ideale per fare auditing di una o due colonne di una tabella.

2.3 Controllo del Rilascio di Informazioni Statistiche

I databases sono risultati del collezionamento di dati di diversa provenienza:

- Web.
- Social media.
- Smartphones.
- etc.

Note:-

Molte di queste informazioni sensibili potrebbero essere utili per un'analisi dei dati.

Spesso i dati per scopi statistici sono rilasciati come:

- Open data: uso del web per rappresentare i legami tra le pagine.
- Datasets per riproducibilità degli esperimenti.
- Data challenge.
- Politiche/Regolamenti.

Note:-

In alcuni casi (e.g. Netflix nel 2006) i dati sono stati usati per inferire informazioni oltre lo scopo per cui sono stati rilasciati.

2.3.1 Disclosure

Definizione 2.3.1: Disclosure

Un disclosure può:

- Occorrere basandosi solamente sui dati rilasciati.
- Risultare da combinazioni dei dati rilasciati con dati esterni che potrebbero o meno essere disponibili al pubblico.

Macrodata e Microdata:

- *Macrodata:* spesso ottenuta con aggregazioni di tipo statistico:
  - *Conti/Frequenza:* ogni cella di una tabella contiene il conto o la frequenza delle situazioni.
  - *Magnitudo:* ogni cella di una tabella contiene un valore aggregato di una quantità di interesse.
- *Microdata:* grana fine, su singole persone (rischio maggiore di data breach).

Domanda 2.3

Come può avvenire questo rilascio non voluto di informazioni?

- Inferenza su sondaggi statistici (inferential disclosure):
  - Avviene quando le informazioni possono essere inferite con alta confidenza per via di proprietà statistiche dei dati.
  - Questo tipo di disclosure è difficilmente gestibile perché in linea teorica si può fare inferenza su qualsiasi dato e quindi nessun dato potrebbe essere rilasciato.
  - L'inferenza serve per predirre aggregati, non singoli valori.
- Informazioni sensibili (attribute disclosure):
  - Avviene se si riescono a stimare le informazioni di un attributo.
- Direttamente sui dati rilasciati (identity disclosure):
  - Avviene quando una terza parte può identificare il soggetto.
  - Nel caso dei macrodata non è particolarmente pericoloso.
  - Per i microdata è un problema perché sono più personali.

Note:-

I microdata contengono specifici identificatori: nome, numero telefonico, email, codice fiscale, etc. Il primo step per trattare i microdata è quello di eliminare o crittografare quelle colonne.

2.3.2 Statistical Disclosure Control

Definizione 2.3.2: Statistical Disclosure Control

È una collezione di metodi che sono usati come parte del processo di anonimizzazione per controllare l'accesso ai dati.

Note:-

Quest'attività non è strettamente considerata a favore della privacy (fig: 2.5).

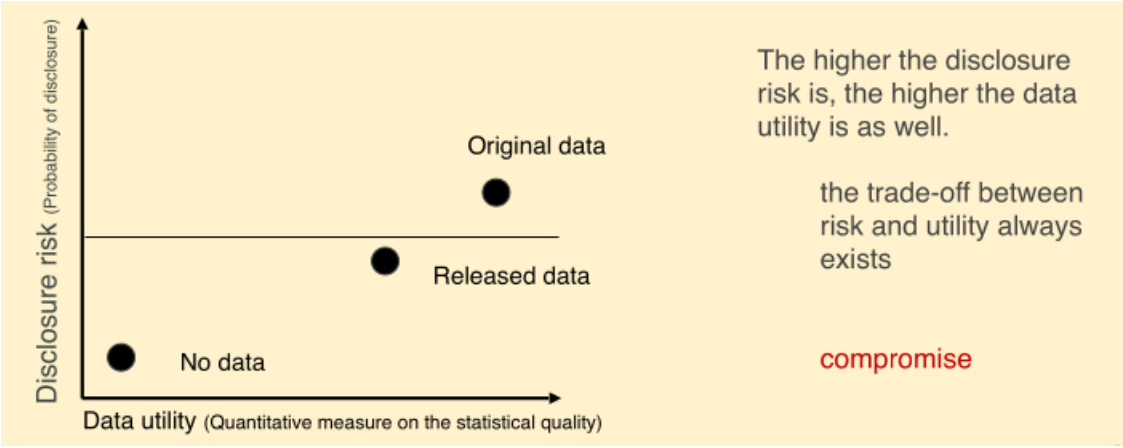


Figure 2.5: Utilità vs. Privacy.

Figure 2.6: Ciclo di Statistical Disclosure Control (SDC).

I passi per proteggere i dati sono:

1. Definire i *quasi identificatori*.
2. Una piccola percentuale di quasi identificatori (generalmente meno del 5%) possono essere identificati.
3. Calcolare le frequenze per le combinazioni di quasi identificatori.
4. Effettuare il *recoding* dei valori degli attributi (e.g. discretizzazione[1]).
5. Applicare metodi SDC.
6. Effettuare una soppressione locale delle variabili di identificazione se ci sono ancora dei records che richiedono protezione.

Metodi SDC per i macrodata:

- *Non-perturbativi:* non modificano i valori delle celle:
  - Soppressione della cella (CS, Cell Suppression).
- *Perturbativi:* modificano i valori delle celle:
  - Arrotondamento casuale dei valori nelle celle (RR, Random Rounding).
  - Arrotondamento controllato dei valori nelle celle (CR, Controlled Rounding).
  - Se con tabelle statistiche si possono modificare le frequenze (CTA, Controlled Tabular Adjustment).

Definizione 2.3.3: Cell Suppression

In una tabella si possono sopprimere delle celle (fig: 2.7):

- Soppressione primaria: non si pubblicano le celle ritenute a rischio.
- Soppressione secondaria: non si pubblicano celle non a rischio per aumentare la protezione delle celle a rischio.

[1] Associare un valore costante di un intervallo associato a un valore continuo.

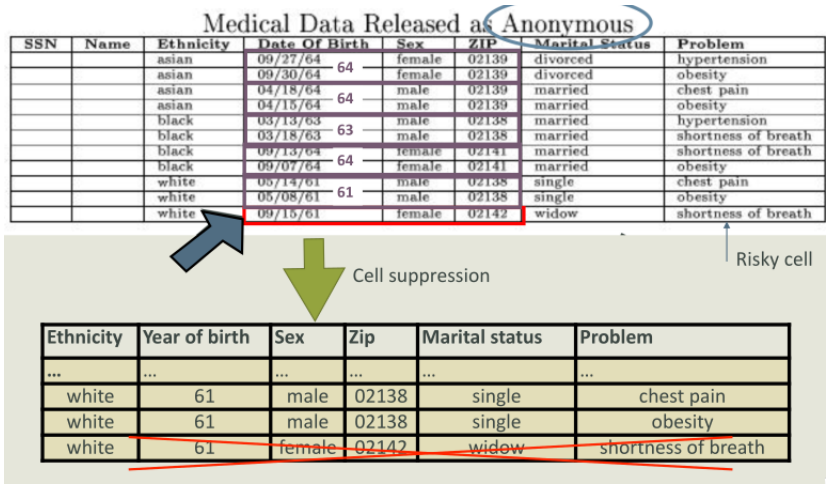


Figure 2.7: Esempio di Cell Suppression.

Definizione 2.3.4: Rounding e Tabular

- Random Rounding: viene deciso in modo randomico se effettuare un arrotondamento per difetto o per eccesso.
- Controlled Rounding: arrotondamento classico.
- Controlled Tabular Adjustment (fig: 2.8): i valori delle celle sensibili sono rimpiazzati da valori simili e le altre celle sono modificate di conseguenza per correggere i valori totali delle righe e delle colonne.

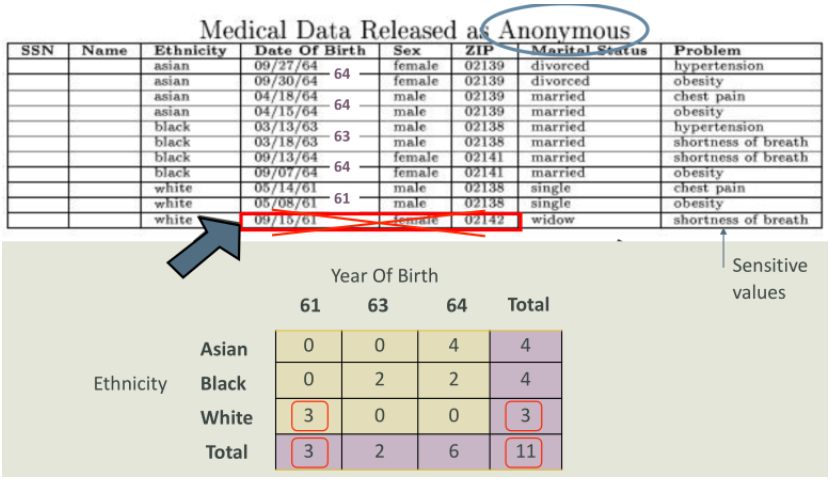


Figure 2.8: Esempio di CTA.

Metodi SDC per query:

- *Perturbativi:* si può applicare una perturbazione sia all'input della query che all'output.
- *Restrittivi:* si rifiuta di rispondere a certe query.

Metodi SDC per i microdata:

- *Data Masking:* genera versioni modificate dei microdata.

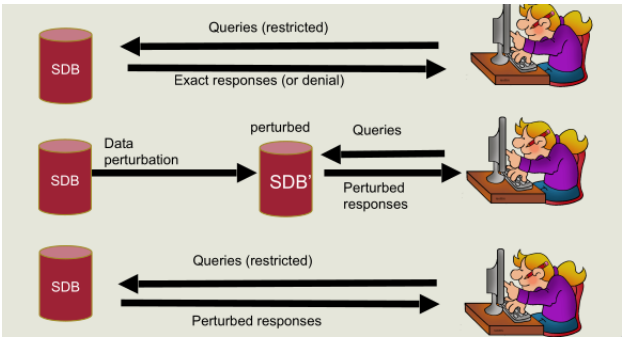


Figure 2.9: Sicurezza di databases statistici (SDB).

- *Data Synthesis:* genera versioni sintetiche dei microdata.

Definizione 2.3.5: Masking Perturbativo

- Aggiunta di rumore:
  - Aggiungere a ogni record un rumore.
  - Media e correlazioni possono essere preservate.
- Microaggressioni:
  - Partizionare i records in gruppi (basati su similarità).
  - Solo la media per ogni gruppo è pubblicata.
- Data swapping:
  - I valori per ogni attributo sono ordinati in ordine crescente.
  - I valori sono scambiati casualmente in un range ristretto.
- Post randomizzazione:
  - Lavora su attributi categorici.
  - Gli attributi sensibili sono cambiati secondo determinate matrici.

Definizione 2.3.6: Masking non Perturbativo

- Sampling: pubblicare un campione dei dati originali.
- Generalizzazione:
  - Gli attributi categorici sono combinati in categorie meno specifiche.
  - Gli attributi numerici sono sostituiti da intervalli.
- Top/Bottom coding: valori sopra/sotto un threshold sono sostituiti da un estremo (simile all'arrotondamento, ma adattativo).
- Soppressione locale:
  - Alcuni attributi sono sostituiti.
  - Aumentare la dimensione dei gruppi.

Sintesi dei dati per i microdata:

- *Goal:* rilasciare un dataset mantenendo i dati confidenziali.

- *Metodi:*
  - *Fully Synthetic Data:* sintetizzare tutto il dataset.
  - *Partially Synthetic Data:* sintetizzare solo le variabili sensibili.
- *Problemi:*
  - I dati sintetizzati devono avere validità per analisi statistiche.
  - La sintesi dei dati dipende dal modello usato.
  - *Overfitting:* records sintetici troppo simili ai records originali potrebbero favorire la reidentificazione.

Note:-

Spesso è difficile misurare l'utilità perché non si sa cosa si voglia fare sui dati. Si potrebbe avere *perdita di informazioni* (più è grande peggio è).

Perdita di informazioni per macrodata:

- *Cell Suppression:* misura del numero delle soppressioni.
- *Rounding and Tabular:* somma delle distanza tra i valori veri e quelli perturbati.

Perdita di informazioni per query:

- *Perturbazione:* differenza tra query non perturbata e query effettiva.
- *Restrizione:* il numero di query rifiutate.

Perdita di informazioni per microdata:

- Valutazione di quanto SDC cambi l'output di un'analisi.
- Altre misure base (statistiche, punteggi, distanze, etc.). Si vanno a calcolare le divergenze di Kullback-Leibler (KL) e di Jensen-Shannon (JS).

2.3.3 Divergenze di Kullback-Leibler e Jensen-Shannon

Definizione 2.3.7: Kullback-Leibler

La divergenza misura quanto una distribuzione di probabilità perturbata $P(x)$ sia diversa dalla distribuzione di probabilità originaria $Q(x)$ con x il valore di una varibile casuale. È definita come l'entropia relativa di P dato Q:

$$KL(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log \left(\frac{P(x)}{Q(x)} \right)$$

Note:-

Però KL non è simmetrica, per cui si introduce JS.

Definizione 2.3.8: Jensen-Shannon

La divergenza misura la similarità mediante KL. Tuttavia JS la simmetrizza ed effettua smoothing:

$$JS(P \parallel Q) = \frac{1}{2} KL(P \parallel M) + \frac{1}{2} KL(Q \parallel M) \text{with } M = \frac{1}{2}(P + Q)$$

3 Framework di Anonimizzazione

Note:-

Molti di questi Framework sono superati, ai giorni nostri *differential privacy* ha soppiantato tutti gli altri.

3.1 Il Problema dell'Anonimità

Problema:

- Molte agenzie, istituzioni, organizzazioni, etc. rendono pubblicamente disponibili dati sensibili riguardanti persone:
  - Molti microdata per le analisi.
  - Spesso la legge richiede la loro anonimizzazione.
- I microdata vengono *sanificati* (rimossi gli id espliciti).
- Non è sufficiente per preservare la privacy:
  - Suscettibili a *linking attack*.
  - Databases pubblici possono rilevare identità "segrete".

3.1.1 Linking Attack

Nel 2001, Latanya Sweeney riuscì a reidentificare i record medico del governatore del Massachussetts:

- Il Massachussetts colleziona e pubblica dati medici sanificati per gli impiegati statali.
- I dati dei votanti registrati sono pubblicamente disponibili.

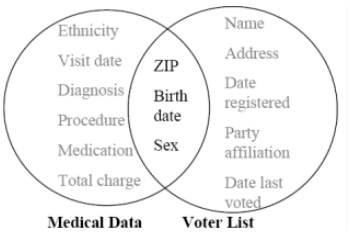


Figure 3.1: Linking Attack del 2001.

Ruoli degli attributi nei microdata:

- Identificatori espliciti: sono rimossi.
- Quasi identificatori: possono essere usati per reidentificare individui.
- Attributi sensibili: portano informazioni sensibili.

Note:-
Il goal della preservazione della privacy è quello di *deassociare* individui da informazioni sensibili.

Definizione 3.1.1: Quasi Identificatori (Tore Dalenius, 1986)

I quasi identificatori sono attributi che non sono univoci di per sé, ma che se combinati con altri quasi identificatori possono creare un identificatore univoco.

I quasi identificatori sono facilmente attaccabili:

- Linking attack di Sweeney.
- Arvind Narayanan e Vitaly Shmatikov hanno usato i quasi identificatori per deanonimizzare dati di Netflix.

3.2 K-Anonimity

Definizione 3.2.1: Quasi Identificatori

Assumiamo $A = a_1, \ldots, a_n$ come insieme di n attributi e D un dataset definito su A. Un quasi identificatore di D è un sottoinsieme di attributi $QI \subseteq S$ che deve essere controllato prima della pubblicazione.

Note:-
Per risolvere il problema Sweeney e Samarati proposero, nel 1998, la *k-anonimity*.

Definizione 3.2.2: K-Anonimity

Assumiamo T come dataset su un insieme $A = a_1, \ldots, a_n$ di attributi e QI_T come insieme di quasi identificatori di T. T *soddisfa* la k-anonimity se e solo se per ogni quasi identificatore $QI \in QI_T$ ogni sequenza esistente di valori attribuiti a QI compare almeno con k occorrenze in T.

Osservazioni 3.2.1
Ogni pubblicazione dei dati deve essere controllata in modo che la combinazione dei valori dei quasi identificatori può essere associata ad almeno k tuple:

- Si nasconde ogni individuo in $k-1$ altri individui.
- I linking attack non possono essere effettuati con una confidenza superiore a $\frac{1}{k}$.

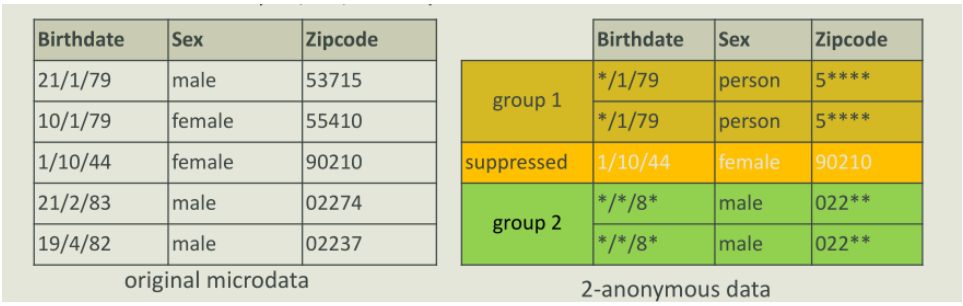
Condizioni sufficienti per la k-anonimity:

- Ogni insieme di valori associati a un quasi identificatore deve avere almeno k occorrenze.
- Gli attributi sensibili non sono considerati.

3.2.1 Come Ottenere la K-Anonimity?

Esistono diversi modi per ottenere la k-anonimity:

- *Generalizzazione:* il valore di un dato attributo è rimpiazzato da uno più generale.
  - Negli ZIP code si nascondono le ultime due cifre ($10149 \rightarrow 101**, 10126 \rightarrow 101**$).
  - La data di nascita rimpiazzata dall'anno di nascita ($27/09/1964 \rightarrow 1964, 30/09/1964 \rightarrow 1964$).
- *Soppressione:* proteggere informazioni sensibili rimuovendole.



Birthdate	Sex	Zipcode
21/1/79	male	53715
10/1/79	female	55410
1/10/44	female	90210
21/2/83	male	02274
19/4/82	male	02237

original microdata

	Birthdate	Sex	Zipcode
group 1	*/1/79	person	5****
	*/1/79	person	5****
suppressed	1/10/44	female	90210
group 2	*/*/8*	male	022**
	//8*	male	022**

2-anonymous data

Figure 3.2: Esempio di applicazione di k-anonimity.

Privacy vs. Utility:

- Se i dati vengono troppo anonimizzati diventano inutili.
- Per cui non si deve anonimizzare più del necessario.

Definizione 3.2.3: Domain Generalization Hierarchy

Una gerarchia di generalizzazione su un dominio (DGH_D) di un attributo A è un ordine parziale su un insieme di domini $Dom_A = D_0, \ldots, D_n$ che soddisfa le seguenti condizioni:

- Ogni dominio D_i ha al più una generalizzazione diretta.
- Tutti gli elementi massimi di Dom sono singletons (per far sì che tutti i valori in ogni dominio possano essere generalizzati in un singolo valore).

Osservazioni 3.2.2

Esempio:

- Z0: 53715, 53710, 53706, 53703.
- Z1: 5371*, 5370*.
- Z2: 537**.

Corollario 3.2.1 Value Generalization Relationship

Una relazione di valori generalizzati associa a ogni valore v_i del dominio D_i un valore univoco v_j in un dominio D_j dove D_j è una diretta generalizzazione di D_i.

Note:-

La relazioni implicano, per ogni dominio D, l'esistenza di un Value Generalization Relationship VGH_D. VGH_D può essere rappresentato come un albero con il valore più generale come radice e i valori più specifici come foglie.

Corollario 3.2.2 Generalization Lattice

Dato un dominio di tuple $DT =< D_{A1}, \ldots, D_{An}$ tale che D_{Ai} sia in Dom_{Ai}, la gerarchia di generalizzazione del dominio di DT è

$$DGH_{DT} = DGH_{D_{A1}} x \ldots x DGH_{D_{An}}$$

DGH_{DT} definisce un reticolo il cui elemento minimo è DT.

Definizione 3.2.4: Generalized Table with Suppression

Dato un insieme di attributi $A = A_1, \ldots, A_n$ e due tabelle T_i e T_j definiti su A, la tabella T_j è una generalizzazione della tabella T_i ($T_i \leqslant T_j$) se e solo se:

- Il dominio di ogni attributo A_x in T_j è uguale o una generalizzazione del dominio di A_x in T_i.
- Ogni tupla t_j in T_j ha una tupla t_i corrispondente in T_i tale che per ogni attributo A_x, $t_j[A_x]$ è uguale o una generalizzazione di $t_i[A_x]$.

Note:-

Non tutte le generalizzazioni hanno lo stesso valore.

Corollario 3.2.3 Vettore Distanza

Date due tabelle T_i e T_j definite sullo stesso insieme di elementi $A = A_1, \ldots, A_n$ tale che $T_i \leqslant T_j$, il vettore distanza di T_j da T_i è il vettore $DV_{ij} = [d_1, \ldots, d_n]$ dove ogni d_x è la lunghezza del percorso univoco tra $dom(A_x, T_i)$ e $dom(A_x, T_j)$ in DGH_{DX}.

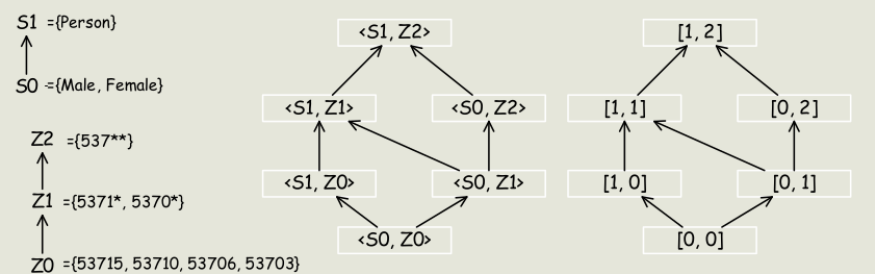


Figure 3.3: Percorsi di generalizzazione.

Note:-

L'obiettivo è trovare la minima generalizzazione che soddisfi la k-anonimity. In questo modo si ha sia anonimizzazione che perdità di informazioni limitata (massimizzazione dell'utilità). Ciò si può fare trovando il vettore distanza minimo.

Definizione 3.2.5: K-Minimal Generalization with Suppression

Date due tabelle T_i e T_j tali che $T_i \leqslant T_j$ e MaxSup lo specifico threshold di soppressione accettabile, T_j è una generalizzazione k-minima di T_i se e solo se:

- T_j soddisfa la k-anonimity.
- La soppressione è minimale.
- $|T_j| - |T_i| \leqslant$ MaxSup.
- $\forall T_x : T_i \leqslant T_x, T_x$ soddisfa gli altri tre punti $\Rightarrow DV_{ix} \geqslant DV_{ij}$.

Livelli di granularità delle tecniche di k-anonimity:

- Generalizzazione:
  - Su singola colonna: passi di generalizzazione su tutti i valori di una colonna.
  - Su singola cella: per una specifica colonna.
- Soppressione:
  - Su singola riga: si sopprime tutta la tupla.
  - Su singola colonna: si oscurano tutti i valori di una colonna.
  - Su singola cella: solo alcune celle sono cancellate.

3.2.2 Algoritmi per la K-Anonimity

Il problema di trovare le tabelle con la minima k-anonimity, con generalizzazione di attributi e soppressione di tuple, è *NP-hard*. La grande maggioranza degli algoritmi proposti hanno un tempo computazionale esponenziale con il numero dei valori degli attributi (approccio greedy). Quando il numero $|QI|$ degli attributi nei quasi identificatori è piccolo rispetto al numero n di tuple nelle tabelle di k-anonimity questi algoritmi sono effettivamente praticabili.

Definizione 3.2.6: Samarati's Algorithm

Ogni percorso in DGH_{DT} rappresenta una strategia di generalizzazione. L'algoritmo si basa su una generalizzazione minima locale (il nodo minore di ogni percorso che soddisfi la k-anonimity). L'algoritmo effettua una ricerca binaria sul reticolo dei vettori distanza:

1. Valuta tutte le soluzioni ad altezza $\frac{h}{2}$.
2. Se esiste almeno una soluzione che soddisfi la k-anonimity:
   - Valuta tutte le soluzioni ad altezza $\frac{h}{4}$.
   - Altrimenti valuta tutte le soluzioni ad altezza $\frac{3h}{4}$.
3. Ripeti fino a ché l'algoritmo non raggiunge l'altezza minore per cui esiste un vettore distanza che soddisfi la k-anonimity.

Note:-

Se non c'è nessuna soluzione che garantisce la k-anonimity ssopprimendo meno di MaxSup tiple ad altezza h allora non può esistere una soluzione con altezza minore a h che garantisce la k-anonimity (approccio branch and bound).

Altre proprietà:

- Ogni generalizzazione k-minima è localmente minima rispetto a un percorso.
- Salendo nella gerarchia il numero di tuple che devono essere rimosse dalla k-anonimity decresce.

Definizione 3.2.7: Incognito Algorithm

Adotta un approccio bottom-up per la visita di $DGHs$. La proprietà di k-anonimity rispetto a un sottoinsieme QI è una condizione *necessaria* per la k-anonimity rispetto a QI.

- Iterazione 1: controlla la k-anonimity per ogni attributo in QI scartando le generalizzazioni che non soddisfano la k-anonimity.
- Iterazione 2: combina le rimanenti generalizzazioni in coppie e controlla la k-anonimity per tutte.
- Iterazione i: combina tutte le i-uple e controlla nuovamente la k-anonimity.
- Iterazione QI: risultato.

Note:-

La proprietà sfruttata dal k-anonimity è la monoticità del reticolo. In realtà l'Incognito Algorithm potrebbe andare sia dal basso verso l'alto che dall'alto verso il basso.

3.3 Oltre la K-Anonimity

3.4 Differential Privacy

4
Fairness

Domanda 4.1

Che cos'è l'IA?

Un sistema IA può essere inteso come:

- *Dati:* esperienza di input presa in considerazione.
- *Algoritmi:* mezzo per elaborare i dati.
- *Decisioni:* che possiamo impiegare direttamente o utilizzare come assistenza.

Note:-

Per esempio, per Russell e Norvig, il termostato è un esempio di IA.

Domanda 4.2

Come si può sapere se questo sistema è *affidabile*?

4.1 Introduzione alla Fairness

Si può concordare che uno strumento è affidabile se commette un numero limitato di errori, ossia quando è in grado di seguire in modo accurato alle nostre istruzioni. Spesso non è semplice capirlo, esistono due possibili approcci:

- *Metodo Clinico:* il decisore combina o elabora le informazioni nella sua mente. Un metodo IA clinico si basa su una serie di regole prestabilite fornite da un esperto del settore.
- *Metodo Attuariale o Statistico:* le conclusioni si basano esclusivamente su relazioni empiricamente stabilite tra i dati e la condizione o l'evento di interesse. Un metodo IA attuariale esami i dati passati per prevedere i risultati futuri.

Note:-

Grosso modo si sono effettuati metodi clinici dal 1956 al 1990, mentre dal 1980 a oggi si sono sviluppati sistemi attuariali.

4.1.1 Metodo Clinico vs. Metodo Attuariale

Definizione 4.1.1: Sistemi Esperti

I sistemi esperti hanno rappresentato un'applicazione incredibilmente risucita dell'IA (metodo clinico): se si raccoglie una quantità sufficiente di conoscenze, è possibile costruire una catena di regole che offrono buone prestazioni per un determinato compito.

Figure 4.1: Immagine di bambino con mattoncini.

Domanda 4.3

Cosa si vede nell'immagine 4.1?

- Per anni i ricercatori IA hanno considerato impossibile risolvere questo task con il metodo clinico, ossia definire cosa sia un "bambino" o un "mattoncino" in ogni circostanza.
- Ai giorni nostri è possibile farlo.
- Le tecniche di *apprendimento automatico* (IA attuarie) hanno compiuto progressi teorici e tecnologici.
- Per questo è nata la necessità di sviluppare enormi quantità di dati.
- Negli anni dal 2010 in avanti le tecniche di *deep learning* hanno contribuito a migliorare questo processo.
- Le reti neurali hanno maggiore scalabilità rispetto a Hidden Markov Model, Modelli Gaussiani, Reti Bayesiane, etc.

Note:-

CONTENT WARNING (TW: razzismo, sessismo, autolesionismo, suicidio, ideazione suicida). Saltare a 4.2.

4.1.2 Esempi del Machine Learing

- Classificazione automatica delle immagini.
- Google Photos tagga due afro-americani come gorilla con il suo software di riconoscimento facciale.
- Un algoritmo IA utilizzato per concorsi di bellezza[1]: ai modelli non piacciono le persone di pelle scura.
- Modelli di predizione dei crimini: per valutare un "livello di rischio".
- I modelli di scraping AI mostrano una discriminazione nei confronti delle donne[2] (fig: 4.2).
- I modelli di Amazon avrebbero dovuto essere addestrati a osservare i modelli ricorrenti nei curriculum inviati all'azienda (che erano appunto uomini).

[1] Well i concorsi di bellezza sono intrinsecamente razzisti quindi non era effettivamente un'idea intelligente.
[2] Che sorpresa...

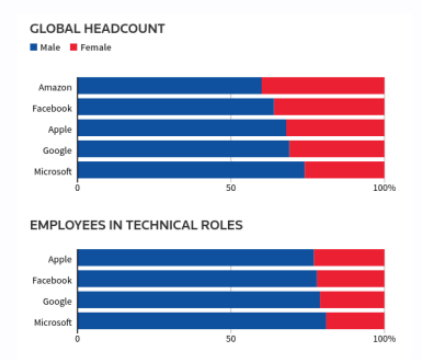


Figure 4.2: Impiegati per sesso (rip NB representation).

- Negli stati uniti le persone con un nome tipicamente bianco hanno più possibilità di essere chiamati per colloqui.
- Un uomo belga si è suicidato dopo sei settimane di scambi con un chatbot.

Domanda 4.4

Ma come è potuto succedere?

- Gli algoritmi sono sessisti/razzisti?
- Chi è il responsabile?
- Come hanno fatto gli ingegneri a non notare questi problemi prima dell'implementazione?
- È illegale?

Problema:

- L'IA attuariale è spesso *opaca*: è difficile capire perché un determinato sistema di IA ha preso una particolare decisione.
- Nonostante sia possibile ispezionare i parametri di un modello (il suo algoritmo) collegarli a delle ragioni per cui sono state prese effettivamente delle decisioni.

4.2 Machine Learning e Fairness

Definizione 4.2.1: Apprendimento

Si dice che un programma per computer apprende dall'esperienza E rispetto a una classe di compiti T e a una misura di prestazione P se la sua prestazione nei compiti in T, misurata da P, migliora con l'esperienza E.

Tasks comuni nel ML:

- *Classificazione:* dati alcuni dati su entità del mondo reale e un insieme di etichette a essi applicabili, imparare ad assegnare tali etichette a entità non viste.
- *Regressione:* dati alcuni dati su entità del mondo reale e un insieme di numeri reali a essi applicabili, imparare ad assegnare numeri a entità non viste.

4.2.1 Classificazione

Definizione 4.2.2: Classificazione

- X: dati, covariati.
- Y: verità fondamentali, etichette, target variabili.

La classificazione è il processo che consiste nel determinare un valore plausibile per Y dato X. Si cerca di apprendere i parametri θ di una funzione f_θ che mappa la variabile casuale X su una stima $\hat{Y}$ di Y:

$$\hat{Y} = f_\theta(X)$$

Corollario 4.2.1 Variabile Casuale

Una variabile casuale è un dispositivo matematico utilizzato per descrivere quantità che dipendono da eventi casuali o che sono in generale incerte.

Note:-

Esempi di variabili casuali: lancio di un dado o di una moneta.

Figure 4.3: Esempio di classificazione.

Osservazioni 4.2.1

- Il processo di apprendimento dei parametri per f_θ dipende dall'algoritmo e dalla scelta della rappresentazione.
- Per la regressione logistica θ è un vettore di valori reali la cui lunghezza è uguale $X + 1$ dove X è il numero di colonne.
- La rappresentazione più naturale per θ può essere diversa se cerchiamo di apprendere un albero decisionale.

Definizione 4.2.3: Accuratezza di un Classificatore

Se $P(Y = f_\theta(X)) = P(Y = \hat{Y}) = 1$ si ha il classificatore perfetto. In generale $P(Y = \hat{Y})$ è l'accuratezza del classificatore.

In alcuni casi l'accuratezza potrebbe non essere utile:

- Si vuole prevedere se il prossimo anno si verificherà una pandemia.
- Negli ultimi 2000 anni si sarebbe potuta prevedere una probabilità del 99% di no.

- Questo mostra che è necessario avere altri modi per descrivere le *prestazioni del classificatore*.

Corollario 4.2.2 Costo

Il costo è il numero reale $l(\hat{y}, y)$ che otteniamo quando classifichiamo un esempio con etichetta y come $\hat{y}$.

Definizione 4.2.4: Classificatore Ottimale

Il classificatore ottimale è il classificatore che minimizza

$$\mathbb{E}(l(Y, \hat{Y}))$$

dove $\mathbb{E}$ è il valore atteso.

Note:-

Due classificatori possono avere la stessa accuratezza ma avere costi differenti.

Corollario 4.2.3 Errore di Classificazione

Il classificatore ottimale che riduce al minimo l'errore di classificazione soddisfa la seguente proprietà:

$$\hat{Y} = f(X), \quad \text{dove } f(X) = \begin{cases} 1 & \text{se } \mathbb{P}(Y = 1 \mid X = x) > \frac{1}{2} \\ 0 & \text{altrimenti} \end{cases}$$

Note:-

Si può ottenere un classificatore ottimale da alcuni valori di probabilità scegliendo la soglia giusta.

Problemi:

- Tasso di falsi positivi sbilanciato.
- Disparità nei risultati positivi.

4.2.2 Il Ciclo ML

Domanda 4.5

Cosa succede nell'apprendimento automatico prima e dopo l'addestramento di un classificatore?

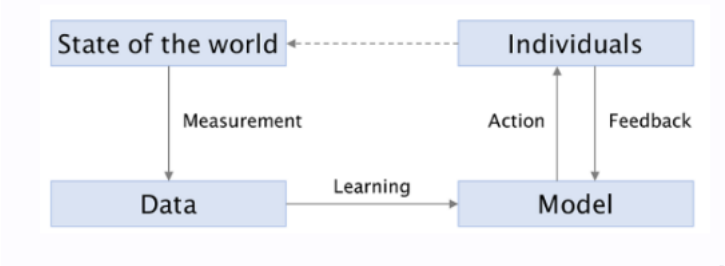


Figure 4.4: Ciclo del ML.

- *Apprendimento:* il processo di ricerca di buoni parametri θ per f e, possibilmente, di una soglia ottimale.
- *Azioni:* come vengono impiegate le decisioni del nostro modello su input nuovi e mai visti prima.
- *Esempi:* eliminare tutte le mail classificate come spam, decidere se una persona con un determinato punteggio di rischio può essere rilasciata su cauzione.

- *Feedback:* facoltativamente registrare come gli utenti hanno reagito alle azioni intraprese.
- *Misurazione:* il processo in cui lo stato del mondo viene ridotto a tabelle di dati.

Il processo di misurazione:

- Spesso nel ML è dato per scontato.
- Vengono effettuate N misurazioni su una proprietà osservabile L.
- Il valore convergerà alla lunghezza vera quando $N \to \infty$.
- Definizione alternativa: la n-esima misurazione $\hat{l}_n$ è correlata a L tramite un errore additivo.

$$\hat{l}_n = L + \epsilon_n$$

- Se gli errori ϵ_n sono:
  - Normalmente distribuiti.
  - Indipendenti.
  - Con una varianza sufficientemente piccola.
- Allora $\frac{1}{N}\sum_{n=1}^{N} \hat{i}_n \to K$ con probabilità 1 quando $N \to \infty$.

Note:-

Alcune caratteristiche possono essere *non osservabili*, ma ciò non significa che siano *impossibili da misurare*. Si deve ipotizzare un *modello di misurazione*, ossia un modello statistico che colleghi il costrutto teorico non osservabile a una proprietà osservabile.

Osservazioni 4.2.2

- Abbiamo abbastanza dati per certi gruppi di persone?
- Siamo sicuri che le nostre misurazioni non siano di parte e non contengono bias nei confronti di certi gruppi?
- L'assunzione di indipendenza è probabilmente falsa: gli errori non sono distribuiti equamente tra i gruppi.

Domanda 4.6

Che fare?

- Finora ci si è basati su una comprensione intuitiva del pregiudizio/discriminazione.
- Lo studio dell'equità nel ML propone una serie di metodi per misurare l'equità di un modello di classificazione.
- Ma come si misura il costrutto teorico di *equità*? Quali ipotesi si vanno a formulare?

4.3 Discriminazione e Misure di Discriminazione

Domanda 4.7

Che cos'è la discriminazione? Perché è sbagliata? La *discriminazione algoritmica* è diversa da quella umana?

4.3.1 Che Cos'è la Discriminazione?

Definizione 4.3.1: Discriminazione Diretta

La discriminazione diretta è la pratica di trattare in modo diverso le persone in base alla loro appartenenza a un gruppo sociale rilevante.

Note:-

Nell'UE è disciplinata dall'articolo 21 della Carta dei Diritti Fondamentali.

Definizione 4.3.2: Discriminazione Indiretta

La discriminazione indiretta è la pratica di offrire lo stesso trattamento a persone appartenenti a gruppi sociali distinti se ciò comporta che un gruppo di persone sia posto in una situazione di particolare svantaggio.

Esempio 4.3.1 (Hilde Schönheit contro Stadt Frankfurt am Main)

Le pensioni dei dipendenti a tempo parziale erano calcolate utilizzando un tasso diverso da quello dei dipendenti a tempo pieno. Questo tasso diverso non era basato sulle differenze di tempo trascorso al lavoro. Pertanto, i dipendenti a tempo parziale ricevevano una pensione inferiore a quella dei dipendenti a tempo pieno, anche tenendo conto della diversa anzianità di servizio, il che significava, in pratica, che i lavoratori a tempo parziale erano retribuiti in misura inferiore. Questa norma neutra sul calcolo delle pensioni si applicava in modo uguale a tutti i lavoratori a tempo parziale. Tuttavia, poiché circa l'88% dei lavoratori a tempo parziale erano donne, *'effetto della norma era sproporzionatamente negativo per le donne rispetto agli uomini*.

Esempio 4.3.2 (D.H. e altri contro Repubblica Ceca)

Una serie di test era stata utilizzata per valutare l'intelligenza e l'idoneità degli alunni al fine di determinare se dovessero essere trasferiti dall'istruzione ordinaria a scuole speciali. Queste scuole speciali erano destinate a persone con disabilità intellettive e altre difficoltà di apprendimento. Lo stesso test è stato applicato a tutti gli alunni che erano stati presi in considerazione per l'inserimento in scuole speciali. Tuttavia, nella pratica il test era stato concepito per la popolazione ceca generale, con la conseguenza che gli studenti rom erano intrinsecamente più inclini a ottenere risultati scarsi, cosa che effettivamente è avvenuta, *con la conseguenza che tra il 50% e il 90% dei bambini rom sono stati istruiti al di fuori del sistema scolastico tradizionale*.

La discriminazione indiretta ha alcuni risvolti interessanti:

- Si potrebbe non essere d'accordo che costituisca vera discriminazione.
- Un'argomentazione etica: nessuno *aveva intenzione* di danneggiare determinati gruppi.

4.3.2 Cosa Rende Sbagliata la Discriminazione?

- L'esistenza di una *systematic animosity* (ostilità sistematica) a favore o contro determinati gruppi sociali rilevanti da parte di chi prende le decisioni è stata tradizionalmente la base per definire la discriminazione come sbagliata.
- Se chi prende le decisioni ha un *intento* negativo allora sta commettendo discriminazione.
- Ma questo funziona per gli algoritmi?

Intenzione e Discriminazione:

- Se il possesso di stati mentali ostili sia necessario per definire la disriminazione allora i casi di studio riportati sopra non costituiscono discriminazione.
- Però gli algoritmi non possiedono stati mentali e gli scienziati che li progettano possono semplicemente affermare che non avevano determinate intenzioni.
- Inoltre la spiegazione basata su stati mentali non copre la discriminazione indiretta.
- Un altro possibile motivo: fare inferenze sugli individui basate sui gruppi di cui fanno parte. Ossia non trattare le persone come individui.

Sono presenti diversi problemi:

- La definizione di generalizzazione.
- I mezzo di generalizzazione possono essere *insufficientemente precisi*.
- Il processo decisionale algoritmico è ammissibile in queste condizioni? Un sistema di apprendimento automatico di solito esegue un'induzione (che è una generalizzazione).

4.3.3 Egalitarismo

Definizione 4.3.3: Egalitarismo

L'egalitarismo è l'idea che le persone debbano essere trattate in modo uguale e che alcune cose di valore debbano essere distribuite equamente.

Note:-

I sistemi osservati fin'ora non sono egalitari.

Domanda 4.8

Cosa deve essere distibuito equamente?

- Cohen: piacere o soddisfazione delle preferenze.
- Rawls e Dworkin: reddito e beni.
- Sen: capacità e risorse necessarie per fare determinate cose.

Note:-

Gli algoritmi che non riescono a distribuire equamente questi beni possono causare danni alle persone.

Tipi di danni:

- *Allocativi:* a determinati gruppi vengono negate risorse in modo disparato (e.g. COMPAS, sistema di assunzione di Amazon).
- *Rappresentativi:* viene rafforzata la subordinazione di alcuni gruppi (beauty.ai, traduzione automatica).

Fortuna e Merito:

- Il *merito* è la condizione di essere meritevoli di qualcosa grazie a scelte, talento, duro lavoro, etc.
- Per l'egalitarismo le disuguaglianze dovute alla *fortuna* dovrebbero essere corrette.
- Tracciare una linea precisa di divisione tra fortuna e merito è un'idea un po' utopica.

Definizione 4.3.4: Inconsapevolezza

L'inconsapevolezza di solito non esiste nei dati del mondo reale ed è difficile da sostenere se si osservano le correlazioni o se le osservano gli algoritmi.

Corollario 4.3.1 Equità per Inconsapevolezza

Una cosa di cui si deve essere consapevoli è che gli algoritmi basati sui dati sono in grado di recuperare informazioni sensibili.

Criteri algoritmici di non discriminazione:

- È interessante cercare di *quantificare* lo squilibrio nelle decisioni di un algoritmo.
- Presi X covariate, Y variabile target, $\hat{Y}$ la stima del target tramite il classificatore f_θ, R il punteggio di rischio, A l'attributo sensibile:
  - Se $\hat{Y}$ è binario l'indipendenza è semplicemente:
$$\mathbb{P}(\hat{Y} = 1 \mid A = a) = \mathbb{P}(\hat{Y} = 1 \mid A = b)$$
  - L'indipendenza viene rilassata (con $\epsilon \geqslant 0$) come:
$$\frac{\mathbb{P}(\hat{Y} = 1 \mid A = a)}{\mathbb{P}(\hat{Y} = 1 \mid A = b)} \geqslant 1 - \epsilon$$
- L'indipendenza può riflettere la convinzione che *tutti i gruppi abbiano pari diritto all'accettazione*. Inoltre non tiene conto di Y.
- Però anche l'indipendenza ha dei problemi:
  - Un decisore malintenzionato che assume persone del gruppo a con un processo rigoroso e una probabilità $p > 0$.
  - Le persone del gruppo b vengono assunte in modo casuale con la stessa probabilità p.
  - In questo caso l'indipendenza è soddisfatta, ma le persone del gruppo b, scelte casualmente, avranno risultati scadenti.
- *Separazione:* le variabili casuali $(\hat{Y}, A, Y)$ soddisfano la separazione se $\hat{Y} \perp A | Y$.
- La separazione può essere intesa come *parità del tasso di errore* tra i gruppi.
- *Sufficienza:* le variabili casuali $(\hat{Y}, A, Y)$ soddisfano la sufficienza se $Y \perp A | \hat{Y}$.
- Se A e Y non sono indipendenti allora sufficienza e indipendenza non possono essere entrambe vere.

Interventi di fairness:

- *Pre-processing:* cambiare i dati, X o Y o entrambi.
- *In-processing:* cambiare il processo di learning f_θ.
- *Post-processing:* cambiare $\hat{Y}$.

Note:-

Non c'è un consenso univoco su quale sia il migliore.

Definizione 4.3.5: Soppressione

Trovare delle caratteristiche in X che sono correlate con le informazioni sensibili X.

Note:-

Una nozione più generale è la mutua informazione sviluppata da Shannon.

White Miku vi augura buona fortuna.