



[PRIVATEZZA PROTEZIONE DATI]

VOCABOLARIO

AFFERISCE: riguarda

SUSSUMERE: ricondurre il concetto in uno più grande

Esempio: la generalizzazione sussume la soppressione (eliminazione tuple)

DISCLOSURE: divulgazione: Identity o Attribute

SAMPLING: campionamento

SURVEY: sondaggio

IL RAZIONALE: la ragione

INFERENZA: deduzione logica

INFERENZA CORRELAZIONE: il sito capisce che sei incinta xk guardi sempre pannolini

MACRODATI: aggregati di tipo **COUNT e MAGNITUDE (avg,sum...)**

MICRODATI: singoli rispondenti

RISPONDENTE: persona in questione

OUTLIER: soggetti interessati con possibili info. sensibili

QUASI-IDENTIFIER quasi identificato tramite attributi linkabili deducibili

LINKING: correlazione di informazioni

VERSIONE SANITIZZATA: dati pubblicati "sanificati" protetti spesso **PACIUGATI =**

PERTURBATI = SPORCATI con del **RUMORE** (dati non veri)

STICK TO THE STORY = coerenza tra le bugie

K-ANONYMITY: nascondere il soggetto in K persone (1 tupla 1 soggetto diverso)

L-diversity: diverse categorie

T-closeness: è l-diversity con campi semantici diversi

ASSOCIATION RULE: SUPPORTO = X/TOT CONFIDENZA = $X/X1$

DIFFERENTIAL PRIVACY: non so se è presente nella lista il sogg. ?

Generalizzare troppo = sopprimere = **OVER DOING IT**

Tupla = **CLUSTER**

PRUNING: potatura, eliminazione del ramo che non mi interessa

RECAP: resumè

UNDO: tornare indietro

POOL: insieme

DATI POLARIZZATI: DATI DIVERSI



DATI MAGNITUDE: MACRODATI POLARIZZATI (alta variabilità) analisi + significativa

EPURARE: separare il valore del magnitude

COALIZIONE: Unione temporanea

K-ANONIMITY punta alla K minimale **USO: MATRIX DISTANCE O RICERCA BINARIA**

DISTANCE VECTOR: la distanza di generalizzazione tra i rispondenti

DEIDENTIFICARE (non certo) I DATI E' DIVERSO DA ANONIMIZZAZIONE (sicuro)

SEMANTICA DI SENSIBILIZZAZIONE: L'identità è sensibile SOLO se appartiene ad un set che è sensibile.

TRUST BOUNDARY: linea di fiducia, di sicurezza

ATTACCHI

OMOGENEITA'
(L diveristy)

SIMILARITY
(T closeness)

BACKGROUND
KNOWLEDGE

SKEWNESS

PRIMA DEL RILASCIO DATI

1) **SAMPLING:** "Il soggetto è nella lista?"

2) **SPECIAL RULE:** x evitare inferenze

TECNICHE MICRODATI

PERTURBATIVE: **SWAPPING**, **BLURRING** (aggregazione tra colonne)

NON PERTURBATIVE: SAMPLE, SUPPRESSIONE, **GLOBAL RECORDING** (raggruppare in categorie)

TECNICHE MICRODATI

ROUNDING (random, controlled), **CONFIDENTIAL EDIT** (micro sporchi)



UNIVERSITÀ
DEGLI STUDI
DI MILANO

LAUREA MAGISTRALE
Dr. EMANUELE MERONI

PROBLEMI RUN-TIME

WINDOW HISTORY: tenere in memoria tutte le domande della Query

COLLUSIONE: coalizione di soggetti "piu rilevanti" per fare inferenze