



Il soggetto dei big data: da individuo a utente



Profilazione dei dati

È il metodo attraverso cui si esaminano i dati provenienti da fonti come database, sistemi e utenti. L'attività di **profiling** è utilizzata per **raccogliere statistiche** e informazioni sui dati e determinarne l'accuratezza e la completezza. Tali statistiche aiutano a **ottimizzare l'uso e la qualità dei dati e dei metadati**.

«L'attività di Profiling dati è un'attività che consente di configurare i profili che si desidera calcolare.

Un profilo-dati è una raccolta di statistiche aggregate sui dati.»

<https://docs.microsoft.com/it-it/sql/integration-services/control-flow/data-profiling-task-and-viewer?view=sql-server-ver15>



Profilazione degli individui

La profilazione si riferisce, quindi, al processo di elaborazione e applicazione di profili generati dall'analisi di un insieme di dati. Questa avviene attraverso l'uso di **algoritmi** e **altri strumenti matematici**, che consentono di trovare **schemi** e **correlazioni statistiche** in grandi quantità di dati, aggregati nei database.

Quando tali schemi e correlazioni vengono utilizzati per **identificare** o **rappresentare persone**, sia individui che gruppi, si parla di creazione di **profili (utente)**.

La nozione di profilazione diventa problematica nel momento in cui profili di gruppo vengono applicati al singolo individuo, come nei casi relativi all'affidabilità creditizia o all'identificazione dei rischi per la sicurezza.



Proxy data → dati vicarianti o **indiretti**

Utilizzati per inferire la presenza di alcune caratteristiche quando il **dato diretto non è disponibile**.

Ad esempio nei casi in cui

- Si tratta di informazioni sensibili, quindi protette o molto costose
- L'informazione di interesse non può essere espressa attraverso un'unica variabile


L'uso dei proxy data è problematico nel momento in cui questi vengono usati per inferire delle correlazioni causa-effetto.



Profilazione degli utenti in quanto consumatori

Lo studio delle **correlazioni tra gli acquisti e l'inclinazione a essere più o meno responsabili**, serve alle società di carte di credito ad avere delle informazioni che permettano di **prevedere il comportamento dell'utente** riguardo al saldo del debito.

Comprendere in anticipo chi è più o meno responsabile permette di **ridurre il rischio** di distribuire carte a potenziali insolventi e di massimizzare i profitti nella gestione di chi è già titolare di una carta.




Esempio: catena di negozi canadese, vendita di articoli sportivi, accessori per automobili, oggettistica per casa e tempo libero

Articolo acquistato	Grado di responsabilità	Previsione saldo
Accessorio per auto a forma di teschio	Molto basso	Alta probabilità inadempienza
Sistema di scarico modificato	Molto basso	Alta probabilità inadempienza
Feltrini per sedie	Alto	Prevista puntualità nei pagamenti
Becchime di alta qualità	Molto alto	Garanzia del pagamento del saldo della carta



Profilazione dello stato di salute degli utenti


La raccolta e l'analisi dei **dati** riguardanti abitudini **alimentari**, **comportamenti**, **acquisti**, ecc. possono portare a **identificare lo stato di salute degli utenti**. Così come il **comportamento adottato sui social network** consente di individuare persone che soffrono di **depressione** o **stress post traumatico**, sulla base di dati quali: la frequenza di determinate parole, la distanza tra la pubblicazione di un contenuto e quello successivo, la fascia oraria in cui viene maggiormente utilizzata la piattaforma social, ecc.



Esempio: reclutamento di pazienti per testare un farmaco per cura intensiva per la perdita di peso

Un'azienda farmaceutica californiana ha contattato e reclutato pazienti in sovrappeso o con problemi di obesità sulla base di dati come

- Frequenza dei pasti consumati nei fast food
- Possesso di un certo tipo di automobili come minivan e simili
- Acquisto di abiti online
- Abbonamenti “premium” a servizi di intrattenimento come la tv via cavo o simili




«Chiunque può essere etichettato in un database come "inaffidabile", "alto costo medico", "reddito in declino" o con qualche altro termine dispregiativo. I sistemi di reputazione stanno creando nuove (e in gran parte invisibili) minoranze, sfavorite a causa di errori o ingiustizie. Gli algoritmi non sono immuni dal problema fondamentale della discriminazione, in cui assunzioni negative e infondate si ingigantiscono in pregiudizio»

Pasquale F., *The Black Box Society. The Secret Algorithms that Control Money and Information*, Londra: Harvard University Press, 2015

Non possiamo sapere quali informazioni sono raccolte, quali utilizzate, quali determinanti, quello che sappiamo è che contribuiscono a comporre un profilo caleidoscopico che è il referente unico e ultimo delle decisioni prese.




Un altro elemento che crea notevoli difficoltà è riconducibile alla composizione del **dataset**, il quale spesso contiene dei *bias* dovuti a delle situazioni di discriminazione sedimentate socialmente e storicamente.



Le relazioni tra i dati vengono scoperte dagli algoritmi di *machine learning*, i quali svolgono il lavoro in due momenti distinti:

- in un primo momento l'algoritmo lavora su una parte di dati chiamata **training set**, in cui viene esposto a dati selezionati da esempi di interesse per i programmatori
- successivamente il modello individuato nel *training set* viene verificato applicando l'algoritmo ad un altro insieme di dati, ovvero il **test set**.

Le relazioni ricorrenti scoperte dall'algoritmo costituiscono un **modello**. Quest'ultimo, creato durante il *training set*, viene successivamente utilizzato per analizzare altri dataset e classificare gli elementi al suo interno o per prevedere comportamenti e risultati futuri.



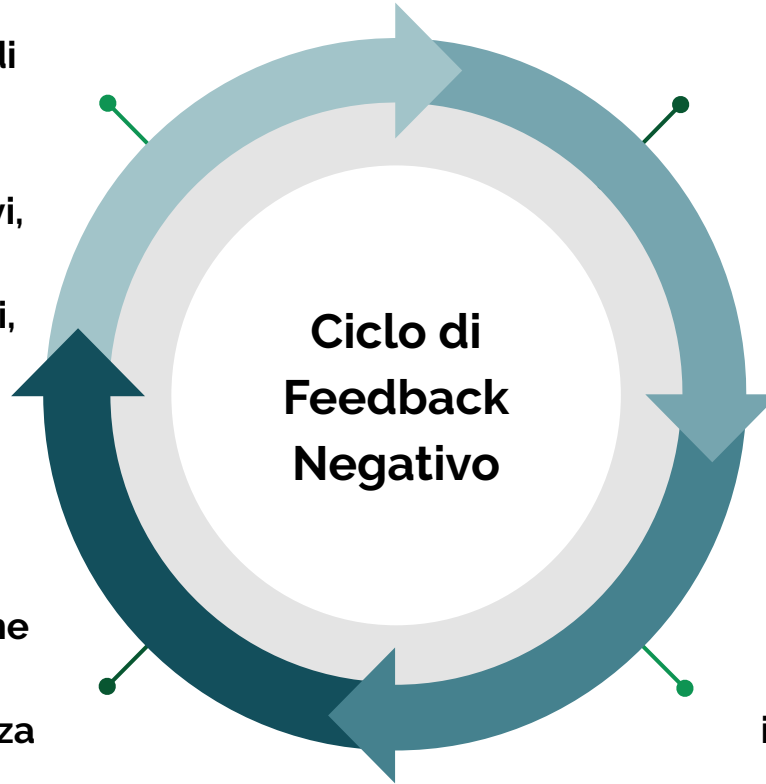
Esempio: PredPol, software utilizzato negli Stati Uniti per analizzare i dati sulla criminalità e calcolare le zone che hanno la maggiore probabilità di subire dei crimini

- **definizione di “crimine”** per la scelta dei dati da elaborare → inclusione di **reati minori**
- **analisi dei dati** per stabilire le **zone** della città **più a rischio** → nelle **periferie** e nelle **zone svantaggiate** socio-economicamente la piccola criminalità è endemica, quindi sono classificate come ad alto rischio
- **ottimizzazione delle risorse** del corpo di polizia, attraverso la concentrazione delle pattuglie nelle zone indicate come ad “alto rischio” → **conferma delle previsioni** e creazione di nuovi dati a supporto



aumento costi di gestione della carta, aumento costi assicurativi, difficoltà a ottenere prestiti, ecc.

situazione economica instabile



Ciclo di Feedback Negativo

classificato come irresponsabile/ rischio insolvenza

mancanza di un pagamento per improvvisa spesa (es. medica)



Realtà algoritmica

Dimensione completamente immanente ai database → **dato = fatto**

Non vi è alcuna mediazione → il **profilo** utente **non *sta per*** l'individuo, ma **è l'individuo**

Data behaviourism

1. **raccolta** massiva e automatica delle **informazioni**. Il soggetto diventa l'oggetto (osservato)
2. analisi dei dati aggregati per trovare **correlazioni statistiche** e **creazione del profilo**
3. associazione del profilo a **classi** che indicano la probabilità di assumere un determinato comportamento



Quadrivio ontologico di P. Lévy →

possibile	reale
virtuale	attuale

La coppia **possibile-reale** esprime uno stato di **stabilità, prevedibilità, ripetizione**, mentre la coppia **virtuale-attuale** esprime il **divenire, la situazione problematica e creativa**

«Ogni situazione vivente mette in atto una sorta di motore ontologico a quattro tempi e non deve essere “disposta” in blocco in una delle quattro caselle» Lévy P., *Il virtuale*, tr. it. M. Colò, M. Di Sopra, Milano: Cortina, 1997.


Virtuale = processo di problematizzazione, come passaggio all’interrogazione, che mette in discussione la staticità del reale e del possibile. Modalità **capace di svincolarsi dal qui e ora**, consentendo al soggetto di sospendere qualsiasi definizione di se stesso e di **sottrarre il proprio agire a qualsiasi tipo di previsione**



Le operazioni di *profiling* e gli algoritmi predittivi **chiudono la dimensione del virtuale**, lasciando spazio unicamente ai **poli di possibilità e realtà**. La realtà dei dati cristallizza in un **eterno presente** la condizione del soggetto.

Possibile = **serie finita di** candidate **soluzioni** la cui la realizzazione avviene a scapito di tutte le altre e senza possibilità di ritorno.

La **collocazione di un profilo in una determinata classe** implica la previsione del comportamento individuale secondo un **insieme finito di possibilità**, relative alla classe in cui si è collocati.



Si potrebbe dire che uno degli obiettivi ultimi della creazione di profili e delle previsioni sia la **gestione dell'incertezza**, per mezzo di un sistema in cui «**ciascuno paga per il proprio rischio**» Rouvroy A., Stiegler B., *Il regime di verità digitale. Dalla governamentalità algoritmica a un nuovo Stato di diritto*, «La Deleuziana» n. 3, 2016


Che porta all'affermarsi di un «darwinismo nuova maniera, che elogia e premia i «vincitori» per la loro forza e la loro intelligenza e fustiga gli «sconfitti» nella «lotta per l'esistenza economica» sottolineandone le mancanze caratteriali o comportamentali» Wacquant L., *Punire i poveri. Il nuovo governo dell'insicurezza sociale*, trad. it. Ménard M., Roma: DeriveApprodi, 2006.



I **punteggi statistici** diventano quindi lo strumento per individuare e **marginalizzare** gli elementi che costituiscono le cifre passive nel bilancio del sistema sociale.

L'intrecciarsi della **realtà algoritmica** con la **razionalità economica**, quindi, fa sì che la grande varietà di classi prodotte statisticamente sia riconducibile in ultimo a un **ordine dicotomico**

- da una parte i **meritevoli** ed **economicamente attivi**
- dall'altra gli **inadempienti che non producono** ai ritmi richiesti, i quali appartengono sistematicamente alle **fasce più svantaggiate** della popolazione



Le diseguaglianze perpetrate da questo sistema rischiano di essere aggravate dalla creazione di una realtà algoritmica che riproduce **soggetti catalogati e giudicati senza possibilità di appello in base alla situazione socio-economica di provenienza.**

Possibili direzioni da prendere per cercare delle soluzioni

- implementare i diversi tipi di ***fairness*** nei processi algoritmici
- utilizzare gli strumenti di analisi dei dati all'interno di progetti come quello dell'**Hudson Valley Current, Economic Resilience Plan**




Bibliografia minima

Beckett L., Everything We Know about What Data Brokers Know about You, «ProPublica», 7 Marzo 2013.

Berg N., Predicting Crime, LAPD-Style, «The Guardian», 25 giugno 2014

Cagan D., Hewitt C., Nachbar R., Ussher L., Network Analysis and Management of the Hudson Valley Current, 5th Biennial RAMICS International Congress, September 11th-15th, 2019 Hida-Takayama, Japan

Donaggio E., Direi di no. Desideri di migliori libertà, Milano: Feltrinelli, 2016.




Duhigg C., How Companies Learn Your Secrets, «New York Times Magazine», 16 febbraio 2012

Guntuku S. C., Yaden D. B., Kern M. L., Ungar L. H., Eichstaedt J. C., Detecting depression and mental illness on social media: an integrative review, «Current Opinion in Behavioral Sciences», vol. 18, 2017, pp. 43-49

Lévy P., Il virtuale, tr. it. M. Colò, M. Di Sopra, Milano: Cortina, 1997.

O'Neil C., Armi di distruzione matematica. Come i Big data aumentano la disuguaglianza e minacciano la democrazia, trad. it. D. Cavallini, Milano: Bompiani, 2017

Pasquale F., The Black Box Society. The Secret Algorithms that Control Money and Information, Londra: Harvard University Press, 2015.



Perry W. L., McInns B., Price C. C., Smith S. C., Hollywood J. S., Predictive Policing. The Role of Crime Forecasting in Law Enforcement Operations, Santa Monica: RAND Corporation, 2013

Rouvroy A., Berns T., Algorithmic Governmentality and Prospects of Emancipation. Disparateness as precondition for individuation through relationships?, «Réseaux», vol. 177, n. 1, 2013, pp. 163-196

Rouvroy A., Privacy, Data Protection, and the Unprecedented Challenges of the Ambient Intelligence, Studies in Ethics, Law and Technology, Berkeley Electronic Press, 2008

Wacquant L. Punire i poveri. Il nuovo governo dell'insicurezza sociale, trad. it. Ménard M., Roma: DeriveApprodi, 2006