

Il data mining alla prova della privacy

Alcuni sostengono che per contrastare l'evasione non sia necessario rivedere la disciplina della privacy.

Costoro partono da un assunto apparentemente molto semplice: esistono già gli strumenti normativi che consentono all'Agenzia delle entrate di utilizzare i dati necessari.

L'esempio che viene normalmente portato è quello dell'anagrafe dei conti correnti.

È infatti noto che, se l'Agenzia delle entrate ha un ragionevole sospetto di evasione su un contribuente, a determinate condizioni, può ottenere l'accesso ai dati del conto corrente del contribuente.

Tale accesso è tuttavia limitato per la necessità di rispettare una serie di vincoli che evitano che dati sensibili siano trattati in mancanza di ragioni obiettive o, peggio, per ragioni che nulla hanno a che fare con l'interesse pubblico.

Quindi, secondo questo ragionamento, un intervento legislativo si configurerebbe come un tentativo di ridurre o eliminare tali vincoli e di eliminare fondamentali garanzie individuali.

Questo ragionamento è in realtà viziato da un errore di fondo.

Le moderne tecniche di analisi dei dati, cosiddetto data mining, non servono ad analizzare i dati di un contribuente che sia già stato individuato come sospetto.

Queste tecniche servono, invece, per capire quali sono i contribuenti che, eventualmente, dovrebbero essere controllati.

Questo implica che, se si vogliono veramente utilizzare le potenzialità del machine learning, si deve consentire l'accesso massivo ai dati individuali,

sebbene anonimizzati.

Per tornare all'esempio dell'anagrafe dei conti correnti: un modello di machine learning potrebbe, accedendo a tutti i dati presenti in quell'anagrafe, e incrociandoli con altri dati relativi alla pericolosità fiscale accertata, calcolare per ciascun contribuente la probabilità che gli sia un evasore.

In questa fase non è affatto necessario che l'identità del contribuente sia rivelata, anzi è decisamente meglio che non lo sia.

La mappatura del rischio viene affidata a degli algoritmi numerici e gli esiti di questa mappatura sono poi affidati al fondamentale controllo umano.

Esaminando i risultati alla luce dell'esperienza e del contesto, l'Agenzia potrà poi procedere a scegliere chi effettivamente controllare. Solo a quel punto sarà necessario accedere all'identità del contribuente.

In sintesi: si tratta di integrare la tradizionale prospettiva deduttiva con una logica induttiva.

Nella prospettiva deduttiva, l'esperienza umana del singolo funzionario consente di definire, sulla base dell'esperienza e delle conoscenze acquisite, le regole generali per procedere ai controlli e l'accesso ai dati serve per confermare o meno queste regole.

Nella prospettiva induttiva, non ci sono regole predefinite, perché è l'accesso ai dati che consente di identificare i criteri di rischio. L'intervento dell'uomo rimane fondamentale nella fase di interpretazione, applicazione e arricchimento di questi criteri di rischio.

Alessandro Santoro, ordinario di scienza delle finanze, università Milano Bicocca

— © Riproduzione riservata —

