

# I rischi di una prossima “data-base society”

**L'innovazione legata ai dati si manifesta nelle aziende, piccole e grandi, dalle start-up alle multinazionali, e nelle istituzioni, che li possono valorizzare per migliorare la propria offerta di servizi e per aumentare la trasparenza dei processi decisionali e più in generale delle attività e dei relativi costi.**

Articolo di

Andrea Granelli

andrea.granelli@kanso.it

**P**aolo Atzeni, Prorettore alla Didattica dell'Università Roma Tre, insegna “Basi di dati” mentre **Paolo Merialdo** - anche lui di Roma Tre - è professore di Ingegneria Informatica. Entrambi sono tra gli organizzatori del summit “Data Driven Innovation”, giunto alla sua terza edizione, si è tenuto il 18 e il 19 maggio scorso.

## **Come potremmo descrivere il contributo dei dati all'innovazione?**

Per usare un termine forse abusato, ma certamente corretto, possiamo dire che i dati sono pervasivi. Ogni attività umana, anche se semplice, si basa sui dati, e questa affermazione era vera anche prima dello sviluppo della tecnologia: i nostri padri, per decidere che cosa fare, verificavano l'agenda, per organizzare un viaggio consultavano l'orario ferroviario o aereo, per analizzare spese della famiglia esaminavano l'estratto conto della banca, e così via. Lo sviluppo tumultuoso degli ultimi anni ha causato una incredibile crescita nella disponibilità di dati, generati anche da strumenti e dispositivi, oltre che dagli

esseri umani. Contemporaneamente, è cresciuta anche la capacità di elaborazione, con architetture parallele e distribuite in grado di utilizzare i dati per risolvere problemi con approcci completamente nuovi. Un esempio per tutti è quello della traduzione da una lingua all'altra: i traduttori automatici, invece di analizzare in dettaglio la sintassi delle frasi, ricercano porzioni di testo già tradotte e disponibili in archivio, per proporre la traduzione. È evidente che questa soluzione è possibile solo se si dispone di un archivio immenso di testi tradotti e si è in grado di ricercare velocemente in esso il testo da tradurre. I domini di utilizzo dei dati sono tantissimi, perché quasi ovunque vengono generati dati interessanti: quelli dai contesti più noti e disponibili da tempo, relativi ad esempio al settore finanziario, a quelli della grande distribuzione (che con le campagne promozionali acquisisce informazioni sui comportamenti dei clienti) o dei viaggi aerei, a quelli di più recente sviluppo, legati alle grandi quantità di dati generati automaticamente, da macchine di ogni tipo, ad esempio

i sensori sulle automobili oppure i telefonini che si muovono con i loro proprietari. Le possibilità di utilizzo dei dati sono così ampie e variegate da generare opportunità di impresa e di innovazione, che a loro volta producono dati, inducendo così un circolo virtuoso. L'innovazione legata ai dati si manifesta nelle aziende, piccole e grandi, dalle start-up alle multinazionali, e nelle istituzioni, che li possono valorizzare per migliorare la propria offerta di servizi e per aumentare la trasparenza dei processi decisionali e più in generale delle attività e dei relativi costi.

## **Quali sono i rischi di un uso massiccio dei dati, anche a valle di Cambridge Analytica e delle recenti riflessioni sui pericoli dell'AI?**

Ci sono vari tipi di questioni da considerare. Il caso di Cambridge Analytica evidenzia diversi aspetti importanti. Il primo è che le aziende attive su internet raccolgono molti dati sui comportamenti degli utenti, usati inizialmente e prevalentemente a beneficio degli utenti stessi, al fine di fornire loro servizi migliori. Un

primo problema è che spesso gli utenti non sono consapevoli di questo fatto, che magari è citato nelle condizioni d'uso, ma la maggior parte degli utenti non se ne rende conto. Il secondo problema è il fatto che questi dati vengono poi usati anche per finalità diverse rispetto a quelle che l'utente intuisce a prima vista: anche questo talvolta va a beneficio dell'utente, perché può fruire di servizi diversi (pensiamo ai dati raccolti dalle scatole nere degli autoveicoli utilizzati per generare le informazioni sul traffico) ma in altri casi vengono usati per sfruttare gli utenti stessi (ad esempio sommergendoli di messaggi pubblicitari). Infine, e, per quanto ancora da verificare in dettaglio, questo è l'aspetto più grave del caso Cambridge Analytica, le informazioni sui comportamenti degli utenti vengono raccolte per valutare le loro inclinazioni politiche e anche le loro reazioni alle informazioni comunicate sulla rete e, di conseguenza, per generare informazioni false o almeno distorte, volte a influenzare le loro opinioni e il loro comportamento elettorale. Passando a un discorso generale e forse più semplice, una tendenza che si sta diffondendo al giorno d'oggi è quella di cercare di misurare e valutare ogni fenomeno o attività, con il risultato di cercare di ottenere classifiche di ogni genere, spacciandole come oggettive, in quanto ottenute sulla base di grandi quantità di dati. Il problema, in molti casi, è che è difficile costruire misure che descrivano davvero il fenomeno di interesse. Ad esempio, si diffondono in molti contesti le classifiche delle università o addirittura delle scuole superiori; ma la vera qualità di una istituzione educativa è costituita dalla sua capacità di formare i giovani per la vita e questo potrà essere verificato solo dopo venti o trent'anni. D'altra parte ci sono modalità diverse anche di definire gli obiettivi di una tale istituzione: interessa la didattica o la ricerca? La qualità della didattica si misura con i tassi di occupazione dei laureati o con il numero degli abbandoni o con la durata



degli studi? Alcune delle classifiche che vengono pubblicate si basano sulla reputazione, cioè sull'opinione che hanno di esse persone ritenute competenti. Come si può facilmente immaginare, ognuno di questi indicatori può avere senso, ma non descrive certo la qualità e nemmeno una media ponderata degli indicatori può farlo, anche se spesso si procede così, con la motivazione che usando tanti criteri non ci si polarizza su un aspetto. Ovviamente, rilevazioni e misure sono importanti, ma lo spirito critico è comunque necessario. Il problema diventa ancora più serio se si pensa di poter utilizzare dati e formule per prendere decisioni. Ci sono esempi che vengono spesso citati come quello di un sistema utilizzato in alcuni stati degli USA per decidere sulla concessione della libertà vigilata (che, basandosi sulle esperienze precedenti, libera sempre i bianchi benestanti e quasi mai i detenuti di altre etnie, soprattutto se residenti in quartieri poveri), o per la valutazione degli insegnanti (basandosi sul presunto profitto, misurato con test schematici di fine anno, facilmente influenzabile alterando i test: tutto il mondo è paese, anche negli USA ci sono stati casi di insegnanti che, per migliorare la propria valutazione, hanno aiutato i bambini nei test, mettendo così in seria difficoltà i docenti degli anni successivi).

#### **Come è nata l'idea del summit?**

L'idea è nata come multidisciplinare, alcuni di noi, docenti di ingegneria informatica, si sono incontrati con un

collega di giurisprudenza, interessato a queste tematiche e abbiamo pensato a realizzare un evento che favorisse la discussione e il confronto di esperienze diverse, con la partecipazione del mondo accademico, delle istituzioni, delle aziende e anche di singoli cittadini. Il tutto con riferimento sia alla tecnologia sia alla più grande varietà possibile di domini applicativi. Dopo tre edizioni possiamo senz'altro dirci soddisfatti, con più di cento oratori e mille partecipanti, ogni volta, sessioni plenarie molto qualificate e cinque o sei sessioni parallele.

#### **Come pensate di creare ricadute a valle dell'iniziativa? (nuovi progetti di ricerca, nuove alleanze pubblico-privato...)**

Il risultato più importante è la conferma della capacità del nostro Ateneo e dei nostri gruppi di costituire un punto di riferimento sul tema a livello nazionale e in particolare nell'area romana. Ci siamo sempre impegnati in collaborazioni e progetti in tante direzioni e certamente ne nasceranno altri. La difficoltà che forse abbiamo è quella di rispondere a tutte le sollecitazioni che riceviamo, perché il tempo disponibile per ciascuno di noi prima o poi finisce.

#### **E nel futuro, che cosa bolle in pentola?**

Il successo delle prime tre edizioni ci spinge a continuare: siamo già al lavoro per organizzare l'edizione 2019. Abbiamo diverse idee su come far evolvere l'evento. Ma è ancora presto per le anticipazioni.