

CAPITOLO 7

Le potenzialità dei Big Data nel Turismo

di Andrea Ciccarelli, Eleonora Scarsella

Nel corso degli ultimi anni, i progressi tecnologici hanno portato a un aumento smisurato del materiale presente nei dataset, tanto che, sempre più frequentemente, per descrivere questa enorme mole di dati si fa riferimento al termine *Big Data*. I Big Data irrompono sulla scena nel primo decennio del XXI secolo.

Il concetto sta a indicare un insieme di dati di grandi dimensioni, anche non strutturati e in rapido movimento, tali da non poter essere gestiti con i tradizionali approcci. La quantità di dati oggi generata è, effettivamente, abnorme: dai telefoni cellulari alle carte di credito usate per gli acquisti, dalla televisione agli *storage* necessari per le applicazioni dei computer, dalle infrastrutture intelligenti delle città fino ai sensori montati sugli edifici, sui mezzi di trasporto pubblici e privati e via discorrendo. Le prime organizzazioni che hanno iniziato a trattare simili tipi di dati sono state quelle online: aziende come Google, eBay, LinkedIn e Facebook sono state costruite, fin dall'inizio, intorno ai Big Data, che per esse hanno rappresentato non un fine quanto, piuttosto, un importante strumento. Con ogni probabilità, ciò è accaduto perché, avendo a che fare con una grande quantità di dati in formati meno strutturati, esse hanno dovuto adattarsi all'utilizzo di nuove tecnologie e tentare approcci innovativi di gestione.

I Big Data rappresentano anche l'interrelazione di dati provenienti da fonti eterogenee, quindi non soltanto i dati strutturati, come i database, ma anche quelli non strutturati, come immagini, email, dati GPS, informazioni estratte dai social network. Inoltre, il termine Big Data non è utilizzato soltanto per descrivere una raccolta di dati estesa in termini di volume, ma serve anche per comprendere la velocità e la varietà dei dati stessi, quando queste siano tali da richiedere tecnologie e metodi analitici specifici per l'estrazione di un valore significativo.

Il progressivo aumento della dimensione dei dataset pone quindi al ricercatore un obiettivo – e un'opportunità – di livello superiore, quello cioè di estrarre una maggiore quantità di informazioni rispetto a quelle che si potrebbero ottenere analizzando piccole serie. Ad esempio, le numerose informazioni che viaggiano e transitano attraverso internet potrebbero rivelarsi basilari per capire quali siano effettivamente gli

“umori” dei mercati e quindi le tendenze che governano la domanda dei consumatori su tantissimi fronti.

Per spiegare nel migliore dei modi cosa siano i Big Data, possiamo forse far nostra la definizione di un ricercatore di Yahoo Research, Alexander Jaimes, che in una recente conferenza ha affermato: «I dati siamo noi». E se i dati raffigurano quello che noi siamo, di sicuro essi rappresentano la base migliore per conoscere e comprendere i gusti, i bisogni e i desideri dei consumatori e creare prodotti sempre più *tailor made*, in grado di fidelizzare i clienti. Questi prodotti hanno la caratteristica di tenere conto delle esigenze dell'utente, che oggi ricerca la massima personalizzazione. E la personalizzazione comporta la possibilità di intervenire su più fattori e conseguentemente può rendere il prodotto «X» più concorrenziale rispetto agli altri prodotti sul mercato.

Caratteristiche dei Big Data

Come anticipato, con il termine “Big Data” si identifica l'enorme volume di dati (strutturati o meno) noti per essere, da un lato, facilmente reperibili sul web, dall'altro difficili da lavorare utilizzando le tradizionali tecniche per i database comuni o ricorrendo a metodi statistici di base.

In realtà, il problema non risiede tanto (o solo) nel volume dei dati, come in passato è stato per alcuni settori (vedi l'astrofisica, la fisica, la genetica, ecc.), quanto, piuttosto, nella loro frammentazione e variabilità, che determinano la necessità di combinare tecniche di analisi diverse (strutturate e non strutturate, per estrarre risultati significativi).

Sempre più di frequente, oggi, si sente parlare di Big Data; non sempre, tuttavia, chi usa questa espressione ne comprende appieno il significato. Soprattutto è importante comprendere l'importanza del ruolo che tale mole di dati, se correttamente lavorata, può giocare in termini di evoluzione e opportunità per molti settori economici, e non solo.

Con il termine Big data si è soliti indicare una raccolta di dati talmente estesa – quanto a volume, velocità e varietà¹ – da richiedere tec-

¹ Il primo a suggerire questa definizione fu Doug Laney nel 2001; cfr. D. Laney, *3D Data Management: Controlling Data Volume, Velocity and Variety*, February 2001, disponibile all'indirizzo internet blogs.gartner.com

nologie e metodi analitici specifici per l'estrazione di valore². Porre l'accento sul termine "Big", a dire il vero, potrebbe risultare fuorviante: la traduzione "grandi dati" o "grossi dati", infatti, fa pensare all'enorme quantità di dati oggi disponibili in diversi settori e, in automatico, porta a concludere che per "rivoluzione Big Data" si intenda esclusivamente lo scenario che oggi scaturisce dall'opportunità di avere a disposizione così tante informazioni. D'altra parte, enfatizzare la categoria del *volume* è pressoché inevitabile: stiamo parlando di una mole di dati – non lavorabile in tempi ragionevoli con i software e gli hardware normalmente utilizzati per i comuni dataset – dell'ordine degli Zettabyte, ovvero miliardi di Terabyte³; ciò richiede una potenza di calcolo parallela, massiva e distribuita su decine, centinaia o anche migliaia di server⁴.

Ma il volume non è l'unica caratteristica da considerare. Un'indagine del NewVantage Partner⁵ ha evidenziato che, per le grandi aziende, trattare l'assenza di struttura nei dati è cruciale, ancor più che trattarne semplicemente la grande mole. Il 30% degli intervistati, infatti, ha affermato che la principale ragione del loro interesse per i Big Data è rappresentata dalla possibilità di analizzare dati provenienti da fonti diverse, mentre il 22% ha sottolineato la necessità di analizzare nuovi tipi di dati; infine, per il 12% si tratta dell'opportunità di analizzare dati in streaming.

Altro parametro decisivo è dunque quello relativo alla *varietà* dei contenuti. Se i comuni database organizzano dati omogenei (o comunque facilmente standardizzabili) provenienti da indagini statistiche o da

² A. De Mauro, M. Greco e M. Grimaldi, *A Formal definition of Big Data based on its essential Features*, in *Library Review*, vol. 65, n° 3, 2016, pp. 122-135, DOI:10.1108/LR-06-2015-0061.

³ Come noto, il byte è l'unità di misura della capacità di memoria; il Megabyte indica una quantità di 10^6 byte, il Gigabyte ne indica 10^9 , il Terabyte 10^{12} , il Petabyte 10^{15} , l'Exabyte 10^{18} , lo Zettabyte 10^{21} , lo Yottabyte 10^{24} . Per farsi un'idea della mole di dati prodotta, basti pensare ad alcuni aspetti ormai di routine nella vita quotidiana: un buon hard disk di un moderno computer ha una capacità di circa 1 Terabyte (ossia circa 1.000 Gigabyte); il solo motore di un aereo civile genera circa 10 TB di dati ogni trenta minuti di volo, in questo modo un volo di un'ora su un normale aereo di linea genera circa 40 TB di dati, mentre un volo Milano-New York ne genera circa 640.

⁴ In un recente studio si stima in 4.4 Zettabyte il volume dei dati attualmente impegnati nel mondo ogni anno (dato del 2013), con una crescita stimata fino ad oltre 40 ZB annuali nel 2020 (ossia circa dieci volte tanto, con una velocità di raddoppio inferiore ai due anni); cfr. IDC (2014), *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*, disponibile all'indirizzo www.emc.com.

⁵ Cfr. NewVantage Partners (2016), *Big Data Executive Survey 2016*, disponibile al sito internet www.newvantage.com.

ricerche ben strutturate, generalmente raccolti in tabelle o in database relazionali, nel caso dei Big Data, invece, l'elemento centrale diventa la ricerca su più fronti delle informazioni, senza che queste debbano necessariamente concretizzarsi in un dataset strutturato, potendo essere rappresentate, oltre che dai "classici" dati statistici, da immagini, video, documenti, conversazioni, ecc⁶.

Vi è poi, a definire i Big Data, la prerogativa della *velocità*, intesa non tanto (o non solo) come il ritmo con cui tali dati tendono ad aumentare, ma soprattutto come la velocità necessaria per analizzarli prima che diventino obsoleti (proprio a causa della velocità con cui vengono generati).

Oltre alle "3V" appena enunciate, e che derivano dalla già citata definizione fornita da Gartner, alcuni suggeriscono di aggiungere ulteriori dimensioni, quali la variabilità e la complessità⁷. Oltre alla velocità di produzione e all'estrema varietà dei dati disponibili, un ulteriore elemento che complica la gestione di tali dati è infatti proprio la loro *variabilità*, ovvero la possibilità che si alternino periodi di "picco" e di "inconsistenza" nella loro generazione (si pensi ad esempio ai trend dei social media, all'interno dei quali l'accadere di determinati eventi può causare un'esplosione nella produzione di informazioni, in aggiunta a quelle quotidianamente prodotte, le quali diventano ancor più difficilmente gestibili se non strutturate).

Quanto alla *complessità*: dal momento che i dati provengono da fonti estremamente diverse, anche le più semplici operazioni di raccolta e trasformazione in un database analizzabile implicano sforzi notevoli, e ingenti risorse dal punto di vista della dotazione software e hardware.

Infine, ci sembra opportuno, come già proposto da altri autori, considerare una ulteriore V, quella di *valore*, facendo riferimento alle possibilità di creare valore aggiunto all'interno delle imprese proprio grazie all'utilizzo e all'interpretazione dei Big Data; basti pensare, a tal proposito, che la stessa Amazon stima in circa il 30% il fatturato derivante dalle proposte di vendita complementare generate automaticamente dal sistema attraverso suggerimenti del tipo "ti potrebbe piacere

⁶ Basti pensare che i dati non strutturati sono stimati rappresentare circa l'80% delle informazioni prodotte nel mondo (si veda A. Nadkarni, N. Yezhkova, 2014: *Structured versus unstructured data: The balance of power continues to shift*, IDC – Industry Development and Models).

⁷ Si veda, tra le altre, l'impostazione fornita dal SAS Institute, azienda leader a livello mondiale nella produzione di software nel campo del Data Mining (cfr. www.sas.com).

anche...”⁸.

Abbiamo accennato alla questione della varietà delle informazioni; ma come possono presentarsi i Big Data agli analisti? Fondamentalmente, sono due le tipologie di dati disponibili: la prima può essere fatta risalire ai tradizionali modi di collezionare informazioni statistiche; la seconda, invece, prende forma a partire dai dati generati dalla rete. A mero titolo di esempio, possiamo qui riportare alcune fonti di dati tipiche della Big Data Analytics⁹:

- Dati strutturati in tabelle (relazionali): sono quelli prodotti, ad esempio, dai sistemi gestionali e rappresentano una delle principali fonti di dati per gli analisti;
- Dati semi-strutturati: si tratta generalmente di dati del tipo business-to-business, organizzabili gerarchicamente, e creati in linguaggio XML;
- Dati di eventi e macchinari: messaggi, sensori, Mfid, periferiche che grazie alle accresciute capacità di storage possono essere collezionati anche per intervalli di tempo non brevi;
- Dati non strutturati: linguaggio umano, audio, video – sono dati generalmente memorizzati sul web e dai quali è possibile estrarre informazioni strutturate attraverso complesse tecniche di analisi semantica;
- Dati non strutturati da social media: tweet, blog, social network – possono essere “estratti” dalla rete e trasformati in dati strutturati attraverso avanzate tecniche quali *crawling*, *parsing*, *entity extraction*;
- Dati dalla navigazione web: web Logs, tag javascript, packet sniffing consentono di ottenere una enorme mole di informazioni su consumi e preferenze degli utenti;
- Dati GIS: dati geospaziali generati ormai da numerosissime applicazioni, sempre più diffuse; consentono di ottenere informazioni di business, sociali o sulla sicurezza;
- Dati scientifici: dati astronomici, dalla fisica, dalla genetica – si

⁸ McKinsey Global Institute (2011), *Big data: The next frontier for innovation, competition, and productivity*, maggio 2011, disponibile sul sito <http://www.mckinsey.com>. Si rammenta in questa sede che le “raccomandazioni di acquisto” proposte dal sito di Amazon derivano dall’utilizzo di particolari algoritmi matematico-statistici di natura predittiva chiamati *collaborative filtering*.

⁹ La presente classificazione è stata ripresa da: Dedagroup, *Dedagroup Highlights – Big Data: riconoscerli, gestirli, analizzarli*, disponibile al sito internet <http://www.ecos2k.it>.

tratta di dati generati da “eventi” e costituiscono i Big Data per definizione; abbisognano delle tecniche di analisi computazionale più innovative sia nella fase di collezionamento/trattamento sia in quella di analisi.

La rivoluzione Big Data

Il termine Big Data si riferisce, dunque, anche agli algoritmi capaci di trattare così tante variabili in poco tempo e con risorse computazionali il più possibile limitate. Fino a pochi anni fa, uno scienziato che avesse voluto analizzare una mole di dati che oggi definiremmo “Small” o “Medium” Data, avrebbe impiegato molto tempo e si sarebbe servito di computer mainframe dal costo molto elevato (oltre il milione di euro); oggi, con l’avanzamento della tecnologia e attraverso l’utilizzo di innovativi algoritmi, quelle stesse informazioni possono essere elaborate nel giro di poche ore.

Questa è la vera rivoluzione dei Big Data: l’emergere di nuovi strumenti e operazioni, capaci di collegare fra loro le informazioni per fornire un approccio visuale ai dati più ampio, suggerendo strutture e modelli di interpretazione fino ad ora impensabili.

Si potrebbe ritenere che i Big Data costituiscano un campo di interesse solo per il settore IT. Ma se, indubbiamente, l’Information Technology rappresenta per i Big Data il primo passo da cui partire – grazie a strumenti fondamentali come il *cloud computing*, gli algoritmi di ricerca, ecc. –, d’altra parte i Big Data si rivelano utili – quando non indispensabili – nei mercati e nei settori più disparati, dall’automotive alla medicina, dal commercio all’astronomia, dalla biologia alla chimica farmaceutica, dalla finanza al mercato assicurativo. Nessun settore che si fondi sul marketing e in cui vi siano dei dati da analizzare può davvero dirsi indenne dalla rivoluzione Big Data.

Tale rivoluzione tocca le vite di ogni singola persona senza che nessuno se ne accorga direttamente: tutti i dati provenienti dalla navigazione di un utente – ricerche, acquisti, valutazioni... – permettono ai colossi del commercio (elettronico e non) di individuare e proporre i prodotti più adatti alle esigenze e ai desideri del cliente, quelli che sollecitano la sua curiosità e lo spingono a comprare per necessità momentanea, permanente o sull’onda di un semplice impulso. Appartenenti ai Big Data sono gli algoritmi che riescono a determinare, ad esempio, se una shopper è incinta, tracciando le sue ricerche sul web e gli oggetti

acquistati in precedenza: a quel punto, il passo successivo sarà proporre offerte speciali e coupon relativi ai prodotti che maggiormente potrebbero interessarla, in base al suo stato.

Per quanto, almeno a parole, possa apparire tutto sommato semplice e certamente allettante, la “rivoluzione Big Data” non rappresenta una prospettiva del tutto priva di difficoltà, per i potenziali utilizzatori. Il primo ostacolo da superare – oltre, naturalmente, alle problematiche di tipo tecnologico – è la resistenza delle aziende, dei centri di ricerca e di taluni scienziati a condividere le informazioni necessarie a un approccio di questo tipo.

Del resto, la grande mole di dati provenienti da fonti online è spesso inaffidabile, e la loro dinamicità rende complesso ogni tentativo di studio o indagine sistematica per fini di garanzia e consolidamento. Inoltre, gli errori e le eventuali lacune possono risultare ingranditi nel momento in cui più insiemi di dati diversi tra loro siano utilizzati congiuntamente. In assenza di un obiettivo di ricerca chiaro e di un piano di raccolta dei dati altrettanto rigoroso, il rischio di ottenere risultati poco significativi o ingannevoli è piuttosto elevato.

Inoltre, le difficoltà e i costi per accedere ai Big Data rischiano di produrre un nuovo tipo di divario digitale tra coloro che possono permettersi lo sforzo economico necessario e ottenere vantaggi importanti e coloro che, invece, non possono avvalersi delle indicazioni ricavabili da tali studi. Quando si ha a disposizione un insieme di dati potenzialmente sconfinato, infatti, la probabilità di trovare una correlazione significativa tra due “serie” qualsiasi – anche se di numeri completamente casuali – può arrivare fino al 90%.

In sintesi, studiosi e operatori dei Big Data concordano sul fatto che avere accesso a una grande quantità di dati, che copre praticamente ogni aspetto della vita umana, comporti notevoli vantaggi. Ma per godere di questi vantaggi è necessario curare minuziosamente la fase di elaborazione dei dati ricavati dalla movimentazione di qualsiasi indagine che utilizzi le informazioni provenienti da un mondo online enormemente popolato.

Superata, quindi, una comprensibile diffidenza iniziale, i Big Data possono supportare al meglio la raccolta, la classificazione, l'analisi e la sintesi dei dati di un determinato settore, offrendo informazioni preziose che vanno al di là del semplice dato grezzo.

I Big Data nel turismo

La rivoluzione che l'utilizzo dei Big Data rappresenta per ogni settore, risulta forse ancora più significativa per l'industria turistica.

Infatti, tra le aziende che per prime si sono avvalse di questa risorsa ci sono soprattutto società aeroportuali e compagnie aeree. La British Airways, ad esempio, per contrastare la concorrenza, ha deciso di investire sulla conoscenza approfondita della propria clientela attraverso la raccolta di informazioni online e offline derivanti dai programmi fedeltà. In questo modo si riescono a comprendere le esigenze e le problematiche più frequenti dei viaggiatori e a sviluppare proposte e soluzioni di maggiore efficacia. Altre compagnie, come Swiss Air, Air France-KLM e Lufthansa, utilizzano i Big Data allo scopo di migliorare le strategie di *revenue management*, e anche diverse catene alberghiere hanno iniziato a implementare interventi basati sull'impiego di Big Data. Hilton, per esempio, ha introdotto l'uso di una *Balanced Scorecard* finalizzata a comprendere quali fattori guidano le performance organizzative. Grazie a questa attività, è riuscita a individuare delle correlazioni tra il grado di soddisfazione dei clienti e i loro comportamenti. Alcuni hotel, invece, utilizzano piattaforme della Stem capaci di favorire – sulla base di algoritmi di analisi dei Big Data che analizzano continuamente il clima, l'edificio e l'utilizzo di energia elettrica – una efficiente gestione dell'energia stessa, arrivando a ridurre i costi di almeno il 10-15%. Anche le grandi OTA (*Online Travel Agencies*) non trascurano questo aspetto: Expedia, ad esempio, sta effettuando investimenti significativi in questo ambito, ritenuto la chiave di volta per il futuro del *travel*.

Il potenziale per il turismo, quindi, è particolarmente grande: raccogliere, omogeneizzare, estrapolare e interpretare correttamente il set di dati che rappresenta la “traccia” del comportamento, delle scelte e persino del “sentimento” dei turisti che visitano l'Italia, saranno operazioni basate non più solo sui commenti spontanei (o pilotati) dei viaggiatori sulle piattaforme di recensioni (un pool inevitabilmente ristretto e che ormai non rispecchia più il “turista medio” *vero*), ma orientate all'analisi di ulteriori dati non condizionati sulle abitudini e gli stili di vita, sulle preferenze, sui flussi reali del turismo. Informazioni dal valore inestimabile, per chi saprà sfruttarle.

I Big Data per il turismo, inoltre, offrono importanti informazioni non solo sui comportamenti collettivi, ma anche sul rapporto tra luoghi, cose e persone. Secondo quanto recentemente confermato dal TDLAB (Laboratorio di Turismo Digitale), i comportamenti quotidiani

delle persone sono ormai sempre caratterizzati da una qualche forma di intermediazione digitale che, di fatto, alimenta flussi di dati di dimensioni enormi: i Big Data, appunto. Se analizzati con algoritmi più complessi, questi dati articolati e diversificati permettono di implementare in maniera sostanziale i processi decisionali delle aziende turistiche, ma anche di migliorare l'offerta rispondendo adeguatamente alla complessità della domanda.

Le Reti Sociali Online (OSN), per esempio, non sono solo un potente strumento per la promozione e la commercializzazione di offerte turistiche, ma, data la loro incredibile diffusione, sono anche (e forse soprattutto) una straordinaria fonte di informazioni sulle preferenze dei turisti, sulle loro attività, o su come diano valore a ciò che viene loro offerto. Basti pensare alla risorsa rappresentata dall'analisi di tutte quelle piattaforme online che ospitano commenti e discussioni degli utenti sulle proprie impressioni di viaggio, o da una implicita indagine delle "tracce" che questi lasciano durante le loro vacanze. Le attitudini e i comportamenti dei viaggiatori sono, infatti, sempre più "social" e "digital". Tra tutti gli utenti che dispongono di un accesso a Internet, il 91% ha prenotato online almeno un prodotto o un servizio negli ultimi 12 mesi e utilizza i motori di ricerca come principale fonte attraverso cui cercare o pianificare una vacanza; il 42% utilizza un *device* mobile (smartphone, tablet, ecc.) per pianificare, prenotare, informarsi (il 33% nel 2012); il 68% ricerca online prima di decidere luogo e modalità del suo viaggio. Non solo: l'utilizzo di internet si conferma essenziale per il turista non solo nella fase di ispirazione (il 61% si informa attraverso Internet), ma soprattutto nella fase di pianificazione (l'80% utilizza Internet) e in quella di fruizione, una volta a destinazione (il 58% utilizza fonti online per valutare attività e servizi, mentre il 40% crea direttamente nuovi contenuti e li condivide).

Inoltre, grazie anche alle informazioni disponibili online, è possibile fornire con molta più precisione una valutazione sulla effettiva consistenza dei flussi turistici, analizzando la loro attività nello spazio dei social media. Non è difficile capire che si tratta di un'opportunità di cruciale importanza, se si pensa – ad esempio – che oggi, per misurare il flusso turistico di un paese, ancora ci si basa su un conteggio di tipo "tradizionale", ovvero sul numero di visitatori ospitati dalle strutture ricettive "classiche", mentre il fenomeno della sistemazioni alternative (case private, *couchsurfing*, aziende agricole, stabilimenti religiosi, ecc.) si sta sviluppando rapidamente, e ciò fa sì che una parte anche consistente dei turisti che sceglie queste strutture ricettive alternative non venga

adeguatamente contabilizzata. Poiché In Italia le statistiche sull'andamento del turismo vengono messe a disposizione con mesi di ritardo¹⁰, averle a disposizione in tempo quasi reale permetterebbe di agire tempestivamente, di trovare correttivi, di studiare le serie storiche e di sapere cosa potrebbe accadere nell'immediato futuro. I turisti, come tutti, sono produttori consapevoli e inconsapevoli di Big Data e di tracce digitali: un'analisi strutturata di questi dati potrebbe quindi rappresentare uno strumento predittivo di grande utilità.

Ovviamente si tratta di una pratica analitica non esente da criticità; da più parti è stata messa in evidenza la necessità di un approccio rigoroso per l'analisi dei dati provenienti dai social media al fine di evitare errori. Tuttavia, appare evidente l'enorme valore aggiunto rappresentato, in termini di conoscenza, dalla corretta trasformazione di una tale mole di dati in indicazioni utili: questo significa che attraverso l'analisi dei Big Data si potrebbero spiegare fenomeni complessi combinando tutte le informazioni che scaturiscono da tutte le fonti disponibili, ed è evidente come ciò si traduca in un vantaggio straordinario per le aziende e i mercati di riferimento.

Tutto ciò è ancor più evidente nel settore dei viaggi: ogni prenotazione in un hotel, ogni volo acquistato, ogni noleggio auto, ogni transazione eseguita o ogni treno prenotato – in sostanza ogni attività che includa uno smartphone, un GPS, una carta di credito, ecc. – lascia dietro di sé una scia di dati di notevole importanza. Sempre più spesso, poi, l'organizzazione di un viaggio viene discussa in aree dedicate nei siti di blog, in cui i turisti raccontano le proprie esperienze, mettendone in luce gli aspetti positivi o negativi all'interno di contenitori web che sono visibili a tutti.

I Big Data sono considerati quindi da molti una incredibile opportunità per prevedere o influenzare comportamenti, opinioni e sentimenti; del resto, comprendere l'esperienza di viaggio di un cliente è fondamentale per capire ciò che, in un'offerta turistica, si deve aggiungere, migliorare o eliminare. Il vantaggio più grande si concretizzerà nella possibilità di poter prendere decisioni in tempo reale, risorsa che può rivelarsi decisiva in un settore come quello dell'industria turistica, in cui il fattore tempo è spesso determinante.

In sintesi, quindi, i vantaggi offerti dall'analisi di tale tipo di dati sono, da una parte, di carattere strategico, perché i Big Data permettono

¹⁰ Le statistiche sugli arrivi dell'anno in corso saranno noti solamente l'anno successivo.

di conoscere la reputazione di una determinata struttura, di un territorio, di un servizio o di un itinerario; dall'altra, di carattere operativo, perché tutte le informazioni raccolte e analizzate possono portare alla massimizzazione della soddisfazione del turista, tramite una personalizzazione della sua esperienza di viaggio e dell'offerta. Queste informazioni apparentemente semplici portano con sé un valore incalcolabile, rappresentato dalla possibilità di ottimizzare la propria politica finalizzandola a un miglioramento della *reputation*.

I Recommender Systems (RSS)

La crescita esplosiva delle informazioni presenti online può rendere complesse le operazioni di ricerca e selezione; gli utenti spesso sono sopraffatti da una quantità di opzioni che non possono valutare, per mancanza di tempo o di competenze. In questo senso, un meccanismo importante per l'industria turistica sono i *Recommender Systems* (RSS), che si rivelano un prezioso strumento, al servizio degli utenti online, per gestire il sovraccarico di informazioni.

Un Recommender Engine è infatti una *utility* di *front-end* che suggerisce all'utente quali elementi potrebbe trovare interessanti e con quali altri utenti potrebbe volersi mettere in relazione. Perché siano davvero personalizzati, questi suggerimenti devono essere generati tenendo in considerazione i gusti dell'utente così come vengono desunti: i) dalle sue abitudini di navigazione sul sito; ii) dalle sue interazioni con altri elementi/utenti; iii) dalle informazioni che compongono il suo profilo (account).

Lo scopo di un Recommender System è di convincere l'utente a intraprendere determinate *azioni* su uno o più *elementi* di un catalogo (leggere contenuti; comprare articoli; recensire libri...) o a *relazionarsi* con altri *utenti* di un social network. Questo tentativo, tuttavia, non è percepito dall'utente come coercitivo bensì come servizio *personalizzato* sui propri gusti.

I RSS usano i dettagli dei profili degli utenti registrati e le abitudini di tutta la comunità per confrontare le informazioni disponibili rispetto ai parametri di riferimento. Tipicamente, un RS confronta un profilo utente in base ad alcuni criteri di riferimento e cerca di prevedere il "rating" o la "preferenza" che un utente potrebbe dare a un elemento di cui non si è ancora occupato. In prima battuta, gli RSS si sono affermati con successo su siti web di e-commerce, offrendo all'utente una sele-

zione dei dati nella forma del “suggerimento” di oggetti e prodotti che potrebbero rispondere ai suoi interessi, sulla base del suo profilo e delle sue ricerche precedenti (film, libri, notizie, pagine web, ecc). Ultimamente, sempre più spesso sono impiegati nel campo del turismo elettronico (e-tourism), per fornire servizi di consulenza al viaggio e ricavare suggerimenti e preferenze in base alla “profilazione” dell’utente. In questo senso, sono pensati per emulare i servizi generalmente offerti da un agente turistico a cui gli utenti si rivolgono in cerca di consigli sulle destinazioni turistiche; il vantaggio degli RSS in termini di tempo e costi è evidente.

In questi sistemi, l’utente dichiara quali sono i suoi bisogni, interessi e vincoli, e il sistema correla le scelte degli utenti con le destinazioni catalogate.

Una recente innovazione, nell’ambito dell’e-tourism, consiste nell’uso di dispositivi mobili come piattaforma principale per l’accesso alle informazioni, che hanno dato origine al settore del turismo *mobile*.

Il settore turistico presenta poi caratteristiche peculiari, che possono rappresentare nuove sfide e nuove opportunità per l’evoluzione di servizi personalizzati e innovativi: la possibilità di conoscere l’esatta posizione dell’utente, ad esempio, apre la strada a servizi specifici basati sulla geolocalizzazione. Inoltre, poter disporre di informazioni sulla mobilità dell’utente consente di sfruttare tutti i servizi esistenti nelle vicinanze geografiche.

La maggior parte di questi sistemi, lungi dall’essere versioni elettroniche delle tradizionali guide turistiche, incorporano dunque funzioni di personalizzazione che traggono vantaggio dalle funzionalità di rilevamento dei dispositivi mobili moderni, offrendo all’utente la possibilità di avere servizi *context-aware* avanzati.

Le applicazioni RSS mobile possono limitare il sovraccarico di informazioni andando a selezionare solo quelle mirate rispetto a un determinato cliente. Esse forniscono consigli personalizzati attraverso l’impiego di sofisticate metodologie, fornendo raccomandazioni utili sui servizi turistici e suggerimenti modellati sul profilo dell’utente.

Considerazioni sull’utilizzo dei Big Data: opportunità e limiti

Dal quadro complessivo fin qui tracciato, emerge dunque con chiarezza che un uso corretto dei Big Data richiede, oltre a una formazione intensiva del personale, anche la pianificazione di uno sforzo di inve-

stimento atto a colmare il gap generazionale, tecnologico e mentale, ancora presente nei potenziali utilizzatori. Conoscere e saper interpretare questa enorme mole di dati disponibili sul web porta con sé, lo si è visto, notevoli vantaggi in termini di competitività e di efficacia gestionale per le imprese del settore turistico; si potrebbe addirittura immaginare, nei prossimi anni, di essere in grado di disporre in tempo quasi reale, oltre che delle previsioni del tempo, anche delle previsioni degli arrivi.

Tuttavia, avere a disposizione questi *dataset* non equivale automaticamente ad avere in mano dati statistici affidabili ed utilizzabili in campo aziendale, per questo è necessario che il loro trattamento avvenga secondo metodi rigorosi e dal valore accertato, sotto il profilo sia tecnico sia etico. Solo attraverso analisi statistiche rigorose e complesse, infatti, è possibile scoprire modelli nascosti, mettere in luce correlazioni sconosciute e ottenere così altre informazioni potenzialmente strategiche non immediatamente ricavabili dall'insieme dei dati.

Per dirla con le parole del Chief Economist di Google, Hal Varian, le aziende «più ancora che avere la capacità di raccogliere le informazioni, devono avere quella di utilizzarle per conoscere più a fondo la clientela. Sono *le capacità di analisi che fanno la differenza*, e sono quelle di cui si avverte il gap. Probabilmente il lavoro più richiesto e appagante dei prossimi dieci anni sarà quello dello statistico».

Esiste tutta una serie di idee fuorvianti che circolano intorno ai Big Data¹¹: che qualsiasi analisi dei dati possa produrre risultati accurati; che ogni singolo dato possa essere catturato rendendo obsolete le tradizionali tecniche di campionamento; che non sia necessario riconoscere il nesso di causa-effetto tra le variabili, perché le correlazioni tra queste ci raccontano tutto ciò di cui abbiamo bisogno; che non necessitiamo di sofisticati modelli statistici in quanto, se disponiamo di una grande mole di dati, i numeri parlano da soli. Soprattutto quest'ultimo convincimento può condurre a errori clamorosi, ed ebbero modo di sperimentarlo quegli analisti del «Literary Digest» che, nel 1936, produssero forse il più grande “buco” sondaggistico della storia, proponendo come vincitore delle elezioni alla presidenza degli Stati Uniti il Governatore Alfred Landon, mentre la storia ci ha raccontato che Franklin D. Roosevelt venne eletto con oltre il 60% dei voti (risultato al quale si avvicinò molto di più George Gallup, che negli anni a venire avrebbe avuto un

¹¹ Si veda l'ottimo articolo di Tim Harford: *Big data: are we making a big mistake?*, apparso sul «Financial Times» il 28 Marzo 2014 e disponibile sul sito www.ft.com.

certo ruolo nel campo dei sondaggi politici e d'opinione...). Come ben sanno gli statistici (bravi...), infatti, esistono, tra gli altri, due elementi di fondamentale importanza quando si ha a che fare con delle indagini campionarie, ossia l'errore di campionamento e la distorsione; è senza dubbio meglio avere tra le mani un campione non distorto, piuttosto che un campione di grandi dimensioni. Il grande vantaggio che abbiamo è che sia il primo che il secondo possono essere misurati.

Troppo spesso, dunque, ci si dimentica che, se anche potessimo estrarre tutte le informazioni utili riferibili a un certo "oggetto" da un social network come Facebook, non è poi detto che queste, per quanto grande sia la mole di dati, possano ritenersi effettivamente riferibili a tutti, proprio perché l'insieme degli utenti di Facebook è ben lungi dal poter essere considerato un campione davvero realistico popolazione di riferimento, e, pertanto, mal si adatta a rappresentarla in modo corretto.

Il vero grande problema, forse, sta nel fatto che le attuali conoscenze metodologiche non sembrano ancora in grado di consentire un uso efficace ed efficiente dei Big Data. Siamo ancora lontani, insomma, dall'abbandono dei tradizionali metodi di raccolta e creazione di dati statistici, che, allo stato attuale, permettono di soddisfare quei requisiti di pertinenza, affidabilità e replicabilità indispensabili per ottenere informazioni di buona qualità, utilizzabili in campo decisionale.

Questo, naturalmente, non significa che le analisi sui Big Data non vadano tenute nel giusto conto; da un lato proprio perché, con gli auspicabili progressi metodologici, aumenteranno in modo esponenziale le potenzialità di utilizzo; dall'altro in quanto, in ogni caso, sarebbe un grande errore ignorare una tale mole di informazioni, soprattutto in relazione ad alcuni aspetti che mal si prestano ad essere quantificati con le rilevazioni di tipo tradizionale (si pensi, tra tutte, alla *sentiment analysis*). Diventa fondamentale, tuttavia, conoscerne non solo le potenzialità, ma anche e soprattutto i limiti, poiché solo a queste condizioni si rende possibile quell'indispensabile salto qualitativo dalla mera raccolta di informazioni alla conoscenza dei fenomeni che stiamo investigando.