

I Big Data non “parlano da soli”.

Il ruolo dei modelli nella diffusione degli analytics per il management accounting

Roberto Del Gobbo *

Big Data doesn't “speak for themselves”. The role of models in the dissemination of analytics for management accounting

Abstract

Big Data and *analytics* have the potential to significantly influence decision-making processes and all management accounting activities, from reporting to strategic planning and control. However, the diffusion of *Big Data* applications in companies still remains low, for a variety of reasons, relating to people, technology and cultural factors, which the literature has documented.

This article intends to propose a reflection on the different “philosophical” views that managers assume, more or less consciously, on *Big Data*. These views affect expectations on the way *Big Data* should generate knowledge and their effective use in practice. Those who adopt the *representational view* of data are led to believe that *Big Data* can help generate knowledge quickly and automatically, only by using advanced statistical-computational techniques. Conversely, according to the *relational view*, the central role in the knowledge generation process is played by models, understood as ways of organizing data, which act as a “bridge” between *Big Data* and their applications to decision-making processes. The article compares these two opposing views and highlights, through theoretical and practical considerations, how relying excessively on the automatic analysis of *Big Data* can fuel unrealistic expectations, which are then disregarded. The emphasis is placed on the function of the models, built on the decision-maker's knowledge of the application domain, which determine what counts as “data” and guide the technical choices in the elaboration phase and in the interpretation of results, favoring an effective use of *Big Data* for the improvement of business processes.

Keywords: Big Data, Analytics, Models, Management accounting

* University of Macerata, Department of Economic and Law, roberto.delgobbo@unimc.it.

1. Introduzione

“Questo è un mondo in cui enormi quantità di dati e la matematica applicata sostituiscono ogni altro strumento che potrebbe essere utilizzato. [...] Con dati sufficienti, i numeri parlano da soli”¹. Così scriveva nel 2008 Chris Anderson, che a quel tempo era *editor-in-chief* di *Wired*, una delle riviste americane più famose tra quelle che trattano tematiche di carattere tecnologico. L’assunto di fondo del suo editoriale era che la disponibilità di grandi quantità di dati e di strumenti di analisi sempre più potenti avrebbe garantito la produzione di conoscenza in maniera automatica e veloce, senza la necessità di presupposti teorici di riferimento.

Da allora sono passati 15 anni ed in effetti abbiamo assistito ad una continua e vertiginosa innovazione tecnologica nella produzione, disseminazione e analisi dei dati, che è stata accompagnata da un’enfasi crescente sul concetto di *Big Data*. Con questa espressione ci si riferisce comunemente a grandi volumi di dati, di formati diversi (strutturati e non), che vengono generati da varie fonti ad una grande velocità² (McAfee *et al.*, 2012). Le opportunità insite nei *Big Data* sono strettamente legate alla possibilità di elaborarli utilizzando strumenti statistico-computazionali, noti come *Big Data Analytics*, che permettono, attraverso algoritmi “intelligenti” potenzialmente in grado di “imparare” da ogni interazione con nuovi dati, di realizzare analisi autonome o semi-autonome, con finalità predittive e prescrittive (Rusom, 2011). Gli *analytics* hanno suscitato grande interesse anche tra i manager, oltre che nelle comunità scientifiche, creando l’aspettativa di poter utilizzare proficuamente i *Big Data* nei processi decisionali, sia di tipo strategico che operativo (Merendino *et al.*, 2018; Choi *et al.*, 2018).

Ma è proprio vero che basta disporre di strumenti di elaborazione potenti per far sì che i *Big Data* “parlino da soli”, come scriveva Anderson? Quale risposta si può dare, anche alla luce dell’evidenza empirica in riferimento alle applicazioni realizzate sul campo?

¹ Anderson C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, 16 (7), p.2

² I *Big Data* vengono generalmente associati a tre caratteristiche principali, chiamate le tre “V”: il *Volume*, che si riferisce alle quantità elevate di dati e alle dimensioni dei file usati per archivarli e trasferirli, la *Velocità* che indica la rapidità con cui i dati sono prodotti e la *Varietà*, che riguarda la diversità dei formati, delle fonti e delle strutture (McAfee *et al.*, 2012). Questa schematizzazione ha subito nel tempo delle evoluzioni, con l’aggiunta di altre “V”, per identificare ulteriori caratteristiche che possono essere associate ai *Big Data* (Verma, 2017).

È indubbio che i *Big Data* e gli *analytics* abbiano il potenziale per influire significativamente sui processi decisionali aziendali e su tutte le attività tipiche del management accounting, dal reporting ai processi di pianificazione e controllo strategico, come ampiamente documentato dalla letteratura (Agostini *et al.*, 2020; Lombardi & Secundo, 2021; Moller *et al.*, 2020; De Santis & D’Onza, 2021). A motivo del crescente numero di articoli pubblicati sul tema negli ultimi anni, gli studiosi iniziano a considerare “maturo” il filone di ricerca che indaga le possibili applicazioni dei *Big Data* in ambito economico-aziendale, evidenziandone l’ampio ventaglio di benefici a tutti i livelli manageriali (Lombardi *et al.*, 2020; Caserio, 2021).

Al tempo stesso, però, si riconosce che nel contesto attuale la vera sfida sia l’effettiva introduzione dei *Big Data* e degli *analytics* all’interno del sistema delle decisioni per ottenerne quanti più vantaggi possibile (Lombardi *et al.*, 2021). Infatti, il loro grado di utilizzo nella pratica è relativamente scarso, sia nell’aziende italiane, soprattutto di piccole e medie dimensioni (Paolini, 2022), che in quelle europee (Moller *et al.*, 2020). Se guardiamo in particolare ai processi di management accounting, molte aziende non hanno neppure definito una strategia di digitalizzazione per la funzione “finanza e controllo di gestione” e gli stessi controller giudicano ancora troppo limitati gli investimenti aziendali per la trasformazione digitale delle loro attività (Schäffer & Weber, 2019). Da un punto di vista più generale, il rapporto *Digital Economy and Society Index (DESI) 2022*, pubblicato dalla Commissione europea, mostra che nei fatti i *Big Data* rimangono tuttora poco impiegati dalle imprese europee. La diffusione è ancor più limitata in Italia, con solo il 9% delle aziende che usano i *Big Data*, contro una media UE del 14%.

Chiaramente fino ad oggi i cambiamenti indotti dai *Big Data* nei processi decisionali aziendali e nelle attività di management accounting sono stati più contenuti rispetto alle aspettative e all’enfasi che il tema ha ricevuto sia tra i professionisti che tra gli accademici. Le ragioni sono molteplici e possono essere approfondite da diverse prospettive e mediante varie chiavi di lettura.

Come hanno sostenuto alcuni studiosi, il processo di diffusione delle innovazioni “radicali” richiede tempo, soprattutto nei casi in cui le trasformazioni che le accompagnano devono “scendere in profondità” ed essere sostenibili nel lungo termine (Frey & Osborne, 2017). C’è sicuramente bisogno che i controller sviluppino nuove competenze in termini di *Big Data* e *analytics*, competenze che tradizionalmente chi ricopre questo ruolo in azienda non possiede (Castellano *et al.*, 2017). Di recente, tra i professionisti del settore si sta parlando sempre più spesso di *Augmented Chief Financial Officer*, proprio per indicare la direzione verso cui dovrebbero evolvere le competenze del responsabile della funzione “finanza e controllo di gestione”, che

risulterebbero “aumentate” grazie al supporto dei *Big Data* e ad un uso intensivo degli *analytics*.

Ma le aspettative sul modo in cui i *Big Data* e gli *analytics* producono conoscenza dipende molto anche dalla visione “epistemologica” che i manager assumono, più o meno consapevolmente, sui dati stessi. I *Big Data* sono spesso associati all’approccio *data-driven*, secondo cui l’apprendimento si realizza grazie alla raccolta di grandi volumi di dati e all’applicazione di tecniche induttive per estrarre da essi correlazioni significative. Questa idea alimenta l’aspettativa che i *Big Data* consentano di generare conoscenza in maniera automatica e veloce. È chiaro però che la possibilità di applicare con successo una data tecnologia ai dati disponibili dipende da fattori che vanno oltre le capacità computazionali degli algoritmi e chiamano in causa l’“intelligenza umana” (Presti, 2022) e quindi valutazioni soggettive legate al modello concettuale che determina tipologia e formato dei dati da analizzare, significatività dei *pattern* “scoperti” e confini decisionali. In questo senso il ruolo della modellizzazione, in quanto momento in cui i dati vengono organizzati e successivamente interpretati, è fondamentale per la credibilità dell’approccio *data-driven*.

L’obiettivo di questo editoriale è proporre una riflessione critica riguardo alle diverse visioni che i manager possono adottare sui *Big Data*, e alle conseguenti aspettative sul modo in cui essi dovrebbero produrre conoscenza, enfatizzando il ruolo dei modelli nel favorire il processo di diffusione delle applicazioni dei *Big Data* e degli *analytics* anche in ambito economico-aziendale.

2. La diverse visioni dei *Big Data* e il ruolo dei modelli operativo-decisionali

2.1 La visione rappresentativa e la visione relazionale dei *Big Data*

Da un punto di vista epistemologico, si possono distinguere due visioni contrapposte dei dati, e quindi anche dei *Big Data*, che definiscono il modo in cui essi contribuiscono alla generazione di conoscenza (Leonelli, 2020).

La prima visione è quella “rappresentativa”, che intende i dati come rappresentazioni della realtà, generate dalle interazioni tra l’uomo e il mondo. In altre parole ciò che assume lo *status* di “dato” è l’oggetto o l’insieme di oggetti creati nel processo di descrizione e misurazione del mondo. Questi oggetti, dal momento che rispecchiano i fenomeni di cui rappresentano una “traccia”, costituiscono il punto di partenza necessario per studiare la realtà

ed estrarre da essa nuova conoscenza. Nella visione rappresentativa, che è molto diffusa perché è spesso accettata in maniera implicita, i dati in sé stessi sono il fondamento “oggettivo” dei processi di generazione della conoscenza, ed è la loro oggettività che rende affidabile la conoscenza estratta. Secondo questa visione produrre dati significa riuscire a cogliere caratteristiche “oggettive” del mondo, che possono essere analizzate e comprese attraverso i metodi dell’inferenza induttiva, ottenendo così un’istantanea affidabile dei fenomeni di interesse. Chi adotta questa visione considera i *Big Data* un patrimonio assolutamente prezioso proprio perché essi facilitano la generazione di conoscenza: produrre o raccogliere molti dati significa avere a disposizione un’enorme quantità di “fatti” che sono pronti per essere analizzati e quindi più fatti vengono raccolti e collegati tra loro più conoscenza può essere estratta.

Un approccio alternativo e contrapposto alla visione “rappresentativa” dei dati è la cosiddetta visione “relazionale”. Secondo questa visione i dati assumono valore non in sé stessi ma solo nel momento in cui vengono messi in relazione ad uno specifico quesito irrisolto secondo modalità e per motivi che dipendono dalla situazione in cui il quesito viene posto.

La visione relazionale ha delle implicazioni importanti sul modo in cui deve intendersi l’intero processo di generazione della conoscenza. Come nella visione rappresentativa, il processo inizia con le interazioni tra l’uomo e il mondo, che producono oggetti di vario genere, tra cui misurazioni numeriche. Alcuni di questi oggetti sono poi selezionati ed organizzati per poter essere potenzialmente utilizzabili come fonte di conoscenza. È il risultato della manipolazione di questi oggetti, cioè il modo in cui vengono organizzati, a svolgere la funzione rappresentativa di una parte della realtà. Quindi non sono i dati in sé stessi il vero elemento chiave del processo di generazione di conoscenza, “i dati non parlano da soli” semplicemente utilizzando tecniche di apprendimento induttivo. Sono rilevanti piuttosto le motivazioni dell’analisi, i presupposti concettuali e il contesto decisionale in cui si trovano coloro che decidono le modalità per organizzare i dati e ne valutano il potenziale significato. Secondo la visione relazionale, per utilizzare i dati come fonte di conoscenza sono quindi necessari due passaggi ulteriori rispetto a quanto previsto dalla visione rappresentativa: la creazione di modi di organizzare i dati – chiamati “modelli” in ambito scientifico – attraverso cui si manifesta la funzione rappresentativa e l’uso di questi modelli come fondamento per la generazione di conoscenza. In definitiva, è il modello, cioè il modo in cui i dati vengono organizzati e resi rilevanti per un certo tipo

di analisi, che rappresenta una parte della realtà e la rende suscettibile di essere analizzata. Dare un ordine ai dati è ciò che permette ai dati stessi di diventare trattabili come fonte di nuova conoscenza.

Nella visione relazionale, il processo di generazione della conoscenza può essere così descritto nei suoi tratti essenziali: l'analista organizza i dati generati dall'interazione tra l'uomo e il modo, attraverso un modello che lo aiuti a rappresentare il fenomeno che vuole studiare. Questo modello viene poi testato per verificarne l'affidabilità e la rilevanza. Se, a seguito della verifica empirica, il modello è considerato adeguato, viene usato come fonte di conoscenza, altrimenti l'analista torna ad esaminare i dati e prova a organizzarli in modi diversi, secondo modelli alternativi, il che a volte può portare a cambiare radicalmente il tipo di oggetto che viene considerato come "dato" oppure l'aspetto della realtà che si sta studiando. Va da sé che, considerando la fase di modellizzazione determinante per attribuire ai dati uno specifico valore rappresentativo, la visione relazionale riconosce un ruolo fondamentale all'esperienza di coloro che trattano e analizzano i dati ai fini del loro efficace utilizzo.

Alla luce di quanto detto fin qui, è evidente che in base alla visione dei dati che si adotta, le aspettative sul modo in cui i *Big Data* possono essere usati per generare conoscenza varieranno notevolmente. La visione rappresentativa accoglie l'idea che i *Big Data* forniscano un fondamento di conoscenza completo e affidabile, in virtù della sua vastità ed eterogeneità. Secondo questa visione, mettere insieme tanti dati costituisce di per sé un incremento della base empirica della conoscenza. L'accumulo di dati equivale all'accumulo di tanti fatti, un vero e proprio tesoro da cui estrarre nuove scoperte semplicemente tramite tecniche induttive. È facile notare come chi adotta questa visione dei dati possa essere facile preda delle false promesse legate all'uso dei *Big Data*. La visione rappresentativa può generare infatti aspettative sbagliate che poi nella pratica vengono disattese, perché, considerando i dati come "fatti" indipendenti da un modello concettuale di riferimento, non si tiene conto del contesto decisionale in cui essi devono essere utilizzati e interpretati e questo può limitarne l'effettivo impiego a supporto dei processi aziendali.

La visione relazionale presuppone invece un approccio più cauto, e una concezione meno utopistica del ruolo dei *Big Data* come fonti attendibili ed efficaci di conoscenza. Nella visione relazionale l'estrazione di conoscenza comporta che gli oggetti scelti per svolgere la funzione di dati vengano posti in relazione ad altri elementi cruciali per l'interpretazione, come per esempio gli scopi dell'analisi, i presupposti concettuali su cui l'analisi è basata e il tipo di decisioni che si devono prendere. Le modalità attraverso cui i dati

vengono organizzati sono fondamentali per poterli effettivamente utilizzare come fonte di conoscenza, e per giudicarne le potenzialità rappresentative nei confronti della realtà. In questo senso la visione relazionale ha una conseguenza importantissima per la diffusione degli *analytics* che si basano sui *Big Data*, quella cioè di sottolineare lo stretto legame che c'è tra le decisioni prese nell'organizzare i dati e il modo in cui essi vengono effettivamente utilizzati.

Le tecniche avanzate di analisi dei *Big Data* hanno sicuramente il potenziale per supportare le decisioni aziendali, favorendo il miglioramento di alcune attività (Jeble *et al.*, 2018). Ma per quanto queste tecniche siano in grado di gestire enormi moli di dati, in tempo reale e a sostegno di numerose funzioni aziendali, sono pur sempre le doti intellettuali dei soggetti decisori che determinano il modo in cui i dati vengono organizzati e successivamente interpretati e quindi l'efficacia e l'efficienza dei processi decisionali che ne derivano.

2.2 Il ruolo dei modelli nella diffusione degli analytics basati sui Big Data

Nelle letteratura economico-aziendale con il termine modello si intende un'astrazione teorica elaborata al termine di un processo analitico e sintetico per rappresentare una parte della realtà nei suoi tratti dominanti, che può avere finalità sia descrittivo-esplicative sia operativo-decisionali (Marchi & Ciaramella, 2009).

Nel processo di modellizzazione si compie quindi innanzitutto un'operazione di semplificazione³, per dare rilevanza alle caratteristiche principali della realtà indagata, trascurando tutti i dettagli ritenuti non rispondenti alle esigenze conoscitive ed operative per le quali il modello è stato elaborato.

Quando la semplificazione della realtà, formalizzata secondo un certo linguaggio, ha lo scopo di fornire una guida per la decisione e l'azione, si parla di modelli operativo-decisionali (Marchi & Ciaramella, 2009). Un modello operativo-decisionale è in grado di fornire elementi spendibili per la risoluzione di uno specifico problema, perché attraverso una rappresentazione schematica delle variabili principali e delle relazioni causali che le legano permette di supportare efficacemente il soggetto decisore nell'individuazione della decisione-azione migliore. È stato giustamente osservato che il

³ Già Zappa nella sua opera "*Le produzioni nell'economia delle imprese*" affermava: "La semplificazione per astrazione è lo strumento efficace per mezzo del quale la nostra mente [...] sa percepire nella realtà troppo ricca e diversa lo svolgersi dei fatti, innalzandosi a comprensioni più ampie e alla visione di correlazioni più profonde" (1956, p.10)

modello è la “struttura portante” di ogni processo decisionale, dal momento che già la sua formulazione, prima ancora del suo utilizzo, risulta fondamentale per analizzare ed affrontare con rigore logico il fenomeno che si sta indagando (Ferraris Franceschi, 1984).

Un modello operativo-decisionale può essere inteso quindi come un modo per dare un ordine ai dati e metterli in relazione tra loro, a partire da un processo di concettualizzazione basato sulla conoscenza del fenomeno in esame, a cui fa seguito una fase in cui le cognizioni e le ipotesi sulla realtà, espresse nell’astrazione teorica, si traducono in termini formali, matematici o informatici, in modo che gli oggetti scelti come dati possano essere efficacemente utilizzati per prendere decisioni. In questo senso i modelli svolgono un ruolo fondamentale per la diffusione delle applicazioni dei *Big Data* in azienda, perché costituiscono il “ponte” tra i dati e il loro effettivo utilizzo nei processi decisionali.

Da tempo gli studiosi sono concordi nel ritenere che tra le caratteristiche del sistema informativo aziendale un’importanza fondamentale è rivestita dall’idoneità delle informazioni non solo di soddisfare le esigenze conoscitive, interne ed esterne, ma anche di ridurre le incertezze allo scopo di favorire un processo decisionale efficiente ed efficace (Marchi, 2003; Bruni, 1990). Anche i *Big Data* acquistano valore in funzione di questa caratteristica, cioè se possono essere utilizzati concretamente per gestire la complessità e suggerire una soluzione efficace ad un problema decisionale. Ma perché ciò accada non si ci può affidare esclusivamente all’analisi automatica dei *Big Data*, attraverso gli strumenti dell’apprendimento induttivo. È fondamentale ad esempio che il problema oggetto di analisi venga identificato correttamente. Se infatti il problema viene definito in maniera errata, incompleta o distorta, senza comprendere pienamente quali siano gli elementi rilevanti che intervengono e le relazioni che legano tra loro le corrispondenti variabili, le possibili soluzioni possono essere altrettanto distorte. Anche in questo senso la fase di modellizzazione acquista rilievo, perché il modello è il punto di partenza per la definizione del problema stesso, nel senso che la sua costruzione e la sua successiva esplicitazione in termini formali rendono possibile la visualizzazione delle conoscenze sul problema, permettendo eventualmente di focalizzare l’attenzione solo sugli elementi ritenuti più importanti a fini decisionali (Marchi & Caserio, 2010; Favotto, 1990).

Soprattutto di fronte a problemi caratterizzati da un elevato grado di complessità, la modellizzazione permette di fare chiarezza e di esplicitare anche elementi poco visibili, sfruttando la conoscenza del dominio di appartenenza del soggetto decisore. In questo modo la modellizzazione produce la cosiddetta *Situation Awareness* (Niu *et al.*, 2013) che permette di comprendere la

situazione decisionale in cui il manager si trova e di prendere decisioni sempre più consapevoli. I *Big Data* offrono sicuramente delle opportunità ai manager per arricchire di elementi nuovi i loro modelli mentali e possono fornire un contributo importante per ridurre gli errori cognitivi (*biases*) sempre presenti in ogni processo decisionale, soprattutto in quei contesti in cui si avverte la necessità di gestire situazioni di incertezza. Ma senza l'uso dei modelli sarebbe difficile utilizzare efficacemente i *Big Data* per supportare in modo sistematico e continuo le decisioni aziendali.

Da questo punto di vista non è condivisibile l'idea, oggi accettata anche in modo implicito, secondo cui più numerosi sono i dati e migliori gli strumenti per elaborarli, maggiori saranno le possibilità di estrarre nuova conoscenza. Affidarsi eccessivamente, nel trattamento dei *Big Data*, alle tecniche per l'inferenza induttiva può produrre risultati fuorvianti. Nel campo del *machine learning* ci si riferisce a questa situazione con l'espressione *no free lunch theorem*, ad indicare l'importanza dei presupposti concettuali e del contesto decisionale per la selezione e l'efficace implementazione degli algoritmi di elaborazione dei dati (Chehbi Gamoura *et al.*, 2020).

La conoscenza umana, esplicitabile attraverso i modelli, rimane quindi fondamentale perché i *Big Data* possano effettivamente supportare le decisioni e per la diffusione delle loro applicazioni anche in ambito economico-aziendale. In questo senso assumono rilevanza gli schemi mentali, l'esperienza maturata, le competenze e capacità professionali del soggetto decisore, come anche la sua comprensione delle relazioni intercorrenti tra il problema effettivo da risolvere e i meccanismi operativi o le procedure che regolano il funzionamento dell'azienda (Caserio, 2015).

Emerge così l'importanza del ruolo che al tal fine può svolgere il controller, nella veste di *business advisor* (Lean *et al.*, 1999), essendo in grado di effettuare un'analisi dei processi interni e di schematizzarli in modo da favorirne una migliore comprensione, individuando i punti critici in cui l'utilizzo dei *Big Data* può essere conveniente e portare a reali vantaggi per l'azienda.

È vero che negli ultimi anni si sta ponendo un'enfasi crescente sulla figura del *data analyst* o del *data scientist*, come soggetto che all'interno dell'azienda è deputato alla gestione e al trattamento dei *Big Data*. Ma per utilizzare efficacemente i *Big Data* non bisogna soltanto conoscere i metodi statistici e computazionali ed essere in grado di applicare algoritmi "intelligenti". La *data expertise* comprende anche la conoscenza del dominio di applicazione e quindi delle condizioni in cui i dati sono stati generati e delle implicazioni del loro uso. Questa *expertise* svolge un ruolo fondamentale nell'indirizzare le scelte tecniche. Infatti non si tratta semplicemente di scegliere l'algoritmo più adatto al problema, per la cui individuazione interverrà

il *data analyst*. Nella traduzione in termini formali di un modello concettuale occorre determinare l'orizzonte temporale, le variabili di decisione, le relazioni matematiche. Tutti questi elementi, che permettono di dare un ordine ai dati, chiamano in causa il soggetto che ha ideato il modello, che è in grado di esplicitarne gli aspetti più nel dettaglio, di testarlo nella pratica ed infine utilizzarlo come base di partenza per miglioramenti successivi. È sempre il soggetto decisore che interpreta il feedback ricevuto sul risultato dell'applicazione, per sviluppare ed arricchire la conoscenza sulla realtà in esame, che porterà ad un affinamento del modello concettuale, e quindi in definitiva a ridefinire ciò che conta come "dato".

2.3 Il legame tra i modelli e le scelte tecniche per l'analisi dei Big Data: un esempio applicativo

Per evidenziare questi concetti, possiamo considerare a titolo esemplificativo un caso di applicazione dei *Big Data* ad un problema di controllo strategico del portafoglio clienti. Il caso riguarda un operatore di telefonia mobile le cui performance economico-finanziarie non risultano in linea con le aspettative del top management, a causa di un tasso di abbandono della clientela giudicato troppo elevato. L'azienda vuole sviluppare un modello predittivo che consenta di stimare per ogni cliente la probabilità di disattivazione del servizio, così da individuare i segmenti di clientela più a rischio, su cui concentrare le azioni di *retention*, aumentandone l'efficacia. L'azienda è interessata anche a valutare l'efficienza delle iniziative di marketing, confrontando attraverso un modello di simulazione i risultati previsti con i costi corrispondenti.

Si tratta di un tipico problema che rientra nel campo di applicazione dei *Big Data*, dal momento che per la natura del servizio offerto le aziende di telefonia dispongono di una vasta quantità di dati relativi ai loro clienti, di tipo anagrafico e comportamentale, che provengono da fonti interne ed esterne e che vengono aggiornati continuamente.

Per poter ricavare il massimo beneficio da questi dati, l'utilizzo di un modello concettuale di riferimento è fondamentale, perché influisce su "ciò che conta come dato" ai fini dell'analisi. In primo luogo la conoscenza degli esperti del dominio può suggerire di sviluppare più di un modello predittivo, distinguendo a priori alcuni macro-segmenti di mercato che possono essere considerati tra loro eterogenei. Ad esempio, i possessori di carte telefoniche prepagate potrebbero essere analizzati con un modello predittivo differente rispetto a quello dei clienti in abbonamento, visto che per le schede prepagate non è prevista una comunicazione di disdetta all'azienda e quindi il cambio

di operatore risulta agevole e poco costoso. Ne consegue che per le due tipologie di clienti i comportamenti segnalati di un possibile abbandono potrebbero essere diversi, il che si riflette sull'insieme delle variabili esplicative da utilizzare come input negli algoritmi. Quindi costruire due modelli predittivi, che considerano dimensioni di analisi differenti, potrebbe produrre risultati più accurati rispetto a quelli ottenibili con un unico modello riferito all'intera *customer base*. Si tratta di un'ipotesi di lavoro che dipende ovviamente dal grado di conoscenza del dominio da parte del soggetto decisore e dalla sua rappresentazione della realtà.

Il modello concettuale influisce anche sulla scelta dell'orizzonte temporale dell'analisi. Ad esempio, il traffico telefonico in ingresso e in uscita di ogni cliente, che viene spesso considerato dagli operatori del settore una delle tante variabili esplicative della probabilità di abbandono, potrebbe essere misurato alternativamente in termini di settimane o di mesi. In altre parole l'ampiezza degli intervalli di base può variare in funzione delle percezioni e degli schemi mentali dell'esperto del dominio. Anche l'individuazione del periodo di latenza del modello predittivo rispecchia la conoscenza del fenomeno da parte del soggetto decisore. La variabile target, cioè l'abbandono dei clienti, deve essere predetta con n periodi di anticipo e quindi le variabili esplicative, relative all'utilizzo del servizio, dovranno essere riferite al periodo $t-n$ (non tenendo conto per semplicità di esposizione della cosiddetta "latenza di training" legata alla necessità di testare l'accuratezza predittiva dell'algoritmo).

Come accennato sopra, la formulazione del modello prevede anche la selezione delle variabili rilevanti per l'analisi. Per le aziende di telefonia il numero di attributi disponibili su ogni cliente è veramente elevato, nell'ordine di alcune centinaia, con diversi livelli di granularità temporale. Un errore che può essere indotto dalla visione rappresentativa dei *Big Data* è quello di inserire in input nell'algoritmo tutte le variabili disponibili, sperando in questo modo di ottenere risultati più soddisfacenti. Tuttavia questo approccio difficilmente consentirà di raggiungere gli obiettivi desiderati ed è per questo che è importante avvalersi delle conoscenze di chi ha esperienza nel settore ed è in grado di determinare l'insieme delle variabili più significative per il problema in esame. Gli attributi che si considerano di scarsa utilità in relazione agli obiettivi di analisi dovranno essere rimossi, in modo da depurare il dataset dalle informazioni "concettualmente" ininfluenti. Inoltre, sulla base del modello mentale di riferimento, si potrà procedere all'aggiunta di nuovi attributi, ottenuti a partire dalle variabili originarie mediante opportune trasformazioni. Ad esempio potrebbe essere utile introdurre nuovi attributi che

esprimono le tendenze presenti nei dati, mediante il calcolo di rapporti e differenze tra le variabili originarie. Questa attività di manipolazione dei dati, guidata dagli schemi mentali e dalle conoscenze “storiche” del soggetto decisore, produce nuovi oggetti che hanno potenzialità rappresentative nei confronti della realtà che si sta indagando. In altre parole, le variabili “derivate” potrebbero avere un contenuto informativo superiore rispetto a quello delle variabili originarie ed essere più utili agli scopi dell’analisi.

Nel momento in cui i risultati dell’algoritmo predittivo dovranno essere inseriti all’interno di un modello di simulazione, per stimare l’efficienza delle iniziative di marketing rivolte alla *retention*, sarà poi necessario identificare i nessi causali che legano tra loro le variabili di decisione e gli indicatori di performance. Anche per questa attività è fondamentale il modello concettuale dell’esperto del dominio, inteso come modalità di organizzazione dei dati che esplicita, attraverso la formalizzazione di ipotesi, il sistema di relazioni inerenti il fenomeno indagato.

Il modello concettuale del soggetto decisore svolge un ruolo cruciale anche nella fase conclusiva dell’analisi. Alcuni metodi predittivi presentano un’accuratezza più elevata rispetto ad altri, ma producono regole di classificazione poco comprensibili. Quindi nella scelta finale del metodo per stimare la *retention* si dovrà tener conto sia degli indicatori di accuratezza sia dell’interpretabilità dei risultati, che dipende in buona parte dagli schemi mentali del soggetto decisore. Ovviamente i metodi predittivi che producono regole semplici e intuitive saranno da preferire, perché arricchiranno la conoscenza sul fenomeno, migliorando la comprensione delle determinanti della *loyalty* ed evidenziando i segmenti di mercato più significativi. I risultati del modello di simulazione, che riceve in input le previsioni generate dall’algoritmo selezionato, dovranno inoltre essere vagliati in termini di plausibilità, verosimiglianza e consistenza.

Questo esempio dimostra il ruolo che il modello svolge nel favorire l’effettivo utilizzo dei *Big Data*, fungendo da “ponte” tra i dati e le decisioni. Molte delle scelte tecniche compiute nell’analisi dipendono dal modello, cioè dal modo in cui i dati vengono organizzati e posti in relazione tra di loro, che ne influenza le potenzialità rappresentative nei confronti della realtà e li rende interpretabili e suscettibili di produrre nuova conoscenza.

È evidente che per dare la giusta importanza al ruolo dei modelli nel favorire le applicazioni dei *Big Data* ci deve essere una precisa visione di ciò che i *Big Data* sono e del modo in cui possono essere usati per generare conoscenza. Dalle considerazioni esemplificative svolte sopra, dovrebbe essere chiaro che la visione relazionale, a differenza di quella rappresentativa,

riconosce una funzione centrale alla fase di organizzazione dei dati, dal momento che qualsiasi oggetto prodotto nelle interazioni tra l'uomo e il mondo può essere utilizzato come dato o smettere di essere tale, a seconda delle finalità perseguite dall'analista, dei suoi presupposti concettuali e del contesto decisionale in cui si trova ad operare.

4. Conclusioni

È indubbio che i *Big Data* possano avere un impatto significativo nell'ambito del management accounting, sia nell'evoluzione dei modelli di business e nelle attività di pianificazione e controllo strategico, sia nel miglioramento dei singoli processi aziendali in un'ottica di efficacia ed efficienza. La diffusione delle applicazioni dei *Big Data* nelle realtà aziendali rimane però ancora limitata, per una molteplicità di ragioni, attinenti a persone, tecnologia e fattori culturali, che la letteratura ha documentato (Agostini *et al.*, 2020).

In questo articolo si è voluta proporre una riflessione su un elemento che è alla base delle sfide e delle barriere aziendali all'utilizzo dei *Big Data*, e cioè la visione epistemologica dei dati che si riflette sul ruolo dei modelli come "ponte" tra i dati e le decisioni.

Il modo in cui oggi i *Big Data* vengono gestiti sta creando un crescente divario tra lo sviluppo sempre più rapido di algoritmi "intelligenti" e la mancanza di procedure e principi che ne permettano un efficace utilizzo nella pratica aziendale. Troppo spesso i *Big Data* sono visti come un mezzo per accelerare enormemente i tempi di produzione di nuova conoscenza. Chi adotta la visione "rappresentativa" dei *Big Data* è però facile preda delle false promesse legate al loro uso. Molti degli esperti nel trattamento dei *Big Data* osservano che nella pratica esiste un evidente contrasto tra la complessità della gestione dei dati e l'aspettativa che forniscano in maniera semplice e veloce conoscenza spendibile nei processi decisionali. Anche alcuni studiosi hanno criticato l'assunzione che più dati producano necessariamente più informazioni, sostenendo che in database molto grandi possono emergere correlazioni anche semplicemente a causa delle dimensioni e non della struttura dei dati e che l'analisi automatica è incapace di distinguere le correlazioni spurie da quelle significative (Calude & Longo, 2017). È necessario quindi l'intervento dell'intelligenza umana. L'attenzione alla fase di modellizzazione, anche se richiede tempo, può aiutare a migliorare l'attendibilità e la qualità della conoscenza prodotta, e l'utilizzo dei dati per scopi realmente innovativi.

Quindi i *Big Data* non “parlano da soli”, c’è bisogno di un modo per organizzarli e successivamente per interpretarli, processi che dipendono dal tipo di abilità e conoscenze in possesso di chi li analizza. La validità delle interpretazioni dipende dalle circostanze specifiche del processo di analisi, incluso l’impiego di modelli mentali su cui basare l’identificazione di ciò che conta come “dato” in quel particolare contesto. La scelta degli oggetti da selezionare perché svolgano la funzione di “dati” nel processo di estrazione di conoscenza richiede un’adeguata giustificazione, che è inevitabilmente legata alla concettualizzazione della realtà elaborata dal soggetto decisore.

Secondo questa visione, l’analisi dei dati non è un processo puramente “oggettivo” in cui il significato dei dati stessi viene progressivamente rivelato attraverso tecniche di apprendimento induttivo, come supposto dalla visione rappresentativa. Non c’è dato senza relazione. I dati, e quindi anche i *Big Data*, sono interpretabili solo a partire da una rete di relazioni concettuali che devono essere esplicitate in modo da inquadrare correttamente i risultati dell’analisi. È infatti nella fase della modellazione dei dati che viene attribuito ad essi uno specifico valore rappresentativo. Il ruolo dei modelli, enfatizzato dalla visione relazionale, spiega anche il crescente riconoscimento dell’esperienza di coloro che selezionano e analizzano i dati come indispensabile per l’efficace utilizzo dei *Big Data*.

In questo senso il controller, come “coscienza economica” dell’azienda, può svolgere una funzione fondamentale nel processo di diffusione delle applicazioni dei *Big Data*. Essendo in grado di analizzare le relazioni che legano i processi interni e di schematizzarle in modo da fornirne una migliore comprensione, il controller può individuare con più efficacia i punti critici in cui l’utilizzo dei *Big Data* è potenzialmente suscettibile di produrre reali vantaggi per l’azienda. La sua conoscenza del dominio di applicazione e del contesto in cui i dati sono stati generati, come anche la sua comprensione dei legami esistenti tra il problema da risolvere e i meccanismi operativi che regolano il funzionamento dell’azienda, risultano fondamentali nella costruzione dei modelli concettuali e nella loro successiva traduzione in termini formali.

Ovviamente per assolvere pienamente questa funzione il controller dovrebbero sviluppare nuove competenze e maggiore esperienza nel campo dei *Big Data* e degli *analytics*, senza mettere in secondo piano il *business acumen*, il pensiero critico e altre competenze più tradizionali del suo profilo professionale che di fronte alle sfide della trasformazione digitale non dovrebbero ridimensionarsi, ma al contrario diventare più importanti (Moller *et al.*, 2020). Una corretta visione dell’epistemologia dei dati risulta quindi cruciale per proporre modalità di utilizzo dei *Big Data*, che attraverso l’uso

dei modelli, consentano di ottimizzare i processi di estrazione di conoscenza effettivamente spendibile per le decisioni aziendali.

Bibliografia

- Agostini M., Chiuochi M.S., Giuliani M. (2020). Big Data e Analytics: impatto sul sistema aziendale e performance. In: Lombardi R., Mancini D., Chiuochi M.S. (2020), *Smart technologies, digitalizzazione e capitale intellettuale: Sinergie e opportunità*. Milano, FrancoAngeli, pp. 132-156.
- Anderson C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, 16(7), pp. 16-07.
- Bruni, G. (1990). *Contabilità per l'alta direzione. Il processo informativo funzionale alle decisioni di governo dell'impresa*, Milano, Etas Libri.
- Calude C.S., Longo, G. (2017), The Deluge of Spurious Correlations in Big Data. *Foundations of Science*, 22(3), pp. 595-612. Doi: 10.1007/s10699-016-9489-4.
- Caserio, C. (2021). I sistemi informativi aziendali per il supporto decisionale: profili evolutivi e prospettive di ricerca. In: Mancini, D., D'Onza, G., & Caserio, C. (20219). *Scritti in onore di Luciano Marchi, Volume II, Sistemi informativi e revisione aziendale*. Torino, G. Giappichelli Editore, pp. 65-73.
- Caserio C. (2015). *Modelli d'azienda per il supporto decisionale e la generazione della conoscenza*, Torino, G. Giappichelli Editore.
- Castellano N., Presti, C., Del Gobbo R. (2017). Employing big data & analytics in decision-making: Factors affecting managers' trustworthiness. In: *Proceedings of the 11th European Conference on Information Systems Management, ECISM 2017*, Academic Conferences and Publishing Limited, pp. 37-46.
- Chehbi Gamoura S., Koruca H.İ., Gülmez E., Kocaer E.R., Khelil I. (2020). The Evidence of the "No Free Lunch" Theorems and the Theory of Complexity in Business Artificial Intelligence. In: *The International Conference on Artificial Intelligence and Applied Mathematics in Engineering*, April, Springer, Cham, pp. 325-343. Doi: 10.1007/978-3-030-79357-9_32.
- Choi T.M., Wallace S.W., Wang Y. (2018). Big data analytics in operations management. *Production and Operations Management*, 27(10), pp. 1868-1883.
- De Santis F. D'Onza G. (2021). Big data and data analytics in auditing: in search of legitimacy, *Meditari Accountancy Research*, 29(5), pp. 1088-1112. Doi: 10.1108/MEDAR-03-2020-0838.
- European Commission (2022). *Digital Economy and Society Index (DESI)*, -- <https://digital-strategy.ec.europa.eu/en/policies/desi>.
- Favotto F. (1990). Simulazione e modello del bilancio di esercizio. In: Brunetti, G., Coda, V., Favotto, F. (a cura di), *Analisi, previsioni, simulazioni economico-finanziarie d'impresa*, Milano, EtasLibri.
- Ferraris Franceschi R. (1984). *Finalità dell'azienda e condizioni di funzionamento*, Pisa, SEU.
- Frey C.B., Osborne M.A. (2017). The future of employment: How susceptible are jobs to computerisation?. *Technological Forecasting and Social Change*, 114, 254-280. DOI: 10.1016/j.techfore.2016.08.019.
- Jeble S., Kumari S., Patil Y. (2017). Role of big data in decision-making. *Operations and Supply Chain Management: An International Journal*, 11(1), pp. 36-44.
- Lean J., Down S., Sadler-Smith E. (1999). The nature of the client-personal business advisor

- relationship within Business Link, *Journal of small business and enterprise development*, 6(1), pp. 80-88. DOI: 10.1108/EUM0000000006666.
- Leonelli, S. (2020). Scientific Research and Big Data. In: Zalta E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*, Summer Edition, Metaphysics Research Lab, Stanford University, -- <https://plato.stanford.edu/archives/sum2020/entries/science-big-data>.
- Lombardi R., Secundo G. (2020). The digital transformation of corporate reporting – a systematic literature review and avenues for future research, *Meditari Accountancy Research*, 29(5), pp. 1179-1208. Doi: 10.1108/MEDAR-04-2020-0870.
- Lombardi R., Cano-Rubio M., Schimperna F., Trequattrini R. (2021), The impact of smart technologies on the management and strategic control: a structured literature review. *Management Control*, Special Issue 1, pp. 11-30. Doi: 10.3280/MACO2021-001-S1002.
- Lombardi R., Mancini D., Chiucci M.S. (2020). *Smart technologies, digitalizzazione e capitale intellettuale: Sinergie e opportunità*, Milano, FrancoAngeli.
- Marchi L. (2003). *I sistemi informativi aziendali*, Milano, Giuffrè.
- Marchi L., Caserio C. (2010). Generating knowledge by combining prediction models with information technology. In: D’Atri A., De Marco M., Braccini A., Cabiddu F. (Eds.), *Management of the Interconnected World*, Physica-Verlag HD, pp. 307-314. DOI: 10.1007/978-3-7908-2404-9_36.
- Marchi L., Ciaramella N. (2009), Prime considerazioni sui caratteri dei modelli in campo economico-aziendale. In Marchi L., Mancini D. (a cura di), *Gestione informatica dei dati aziendali*, Milano, FrancoAngeli
- McAfee A., Brynjolfsson E., Davenport T.H., Patil D.J., Barton D. (2012). Big data: the management revolution, *Harvard business review*, 90(10), pp. 60-68.
- Merendino A., Dibb S., Meadows M., Quinn L., Wilson D., Simkin L., Canhoto A. (2018). Big data, big decisions: The impact of big data on board level decision-making. *Journal of Business Research*, 93, pp. 67-78. Doi: 10.1016/j.jbusres.2018.08.029.
- Möller K., Schäffer U., Verbeeten F. (2020). Digitalization in management accounting and control: an editorial. *Journal of Management Control*, 31(1), pp. 1-8. Doi: 10.1007/s00187-020-00300-5.
- Niu L., Lu J., Zhang G., Wu D. (2013), FACETS: A cognitive business intelligence system. *Information Systems*, 38(6), pp. 835-862. Doi: 10.1016/j.is.2013.02.002.
- Paolini A. (2022), Integrated data management: New perspectives for management control. *Management control*, Special Issue, 2, pp. 5-14. Doi: 10.3280/MACO2022-002-S1001.
- Presti C. (2022). L’azienda intelligente: opportunità e minacce per la creazione di valore. *Management Control*, 3, pp. 5-12. Doi: 10.3280/MACO2022-003001.
- Russom P. (2011). Big data analytics. *TDWI best practices report*, 19(4), pp. 1-34.
- Schäffer U., Weber J. (2019). Digitalization ante portas. The transformation of controlling as reflected in the third WHU study on the future of controlling. In U. Schäffer (Ed.), *Behavioral Controlling: Anniversary Volume in Honor of Jürgen Weber*, pp. 179-190, Wiesbaden. Doi: 10.1007/978-3-658-25983-9_12.
- Verma S. (2017). Big Data and advance analytics: Architecture, techniques, applications, and challenges. *International Journal of Business Analytics (IJBAN)*, 4(4), pp. 21-47. Doi: 10.4018/IJBAN.2017100102.
- Zappa G. (1956). *Le produzioni nell’economia delle imprese*, Milano, Giuffrè.