

# Introduzione

## Gli advanced analytics a supporto delle decisioni manageriali

di *Emanuele Borgonovo*

---

### Disponibilità di dati e intelligenza tecnologica: la rivoluzione tecnologica dei nostri giorni

---

Siamo da qualche anno ormai nella data driven economy. La rivoluzione digitale, l'aumento della connettività, l'istantaneità della diffusione delle informazioni stanno rivoluzionando il modo in cui aziende e istituzioni operano.

È ormai assodato come l'umanità stia creando quantità di dati sempre maggiori, lasciando sempre più tracce digitali di sé, dei suoi gusti, bisogni e consumi. Non solo, ma macchine opportunamente addestrate, robot, sono in grado di svolgere compiti sempre più sofisticati che in molti casi aiutano l'uomo e ne amplificano capacità e risultati. Infatti, anche se quella parte di intelligenza artificiale che colpisce di più l'immaginazione può essere ritenuta ancora lontana dallo scaricare a terra tutte le proprie potenzialità, tutta la parte di intelligenza artificiale basata sul machine learning è molto vicina a noi e sta già avendo un impatto non trascurabile sulla vita di tutti i giorni. La digitalizzazione delle informazioni ha quasi completamente sostituito le informazioni su carta, rendendo disponibili grandi quantità di dati che risiedono nei sistemi formativi delle imprese. I nuovi sistemi di analisi dei dati stanno operando un grosso cambiamento nel modo in cui vengono prese le decisioni manageriali. Sempre più imprese, piccole e grandi, si rendono conto della potenzialità che risiedono nei dati a propria disposizione. Scoprire trend, patterns, comportamenti in anticipo è fattore di vantaggio competitivo.

Anche dal punto di vista istituzionale ci si è resi conto da qualche anno del potenziale impatto delle nuove metodologie di analisi. Nel 2012 è nata infatti l'agenzia italiana del digitale; nel 2014 un documento presentato al Parlamento Europeo da un'apposita Commissione parla esplicitamente di *Digital Economy*. Il rapporto della commissione evidenzia come l'economia stava (e stia ancora oggi) per essere travolta e stravolta dall'enorme massa di dati generati dall'umanità e dall'altrettanto veloce digitalizzazione dei processi di decisione. La Commissione sottolinea come diventino essenziali nuovi tipi di competenze: in particolare occorrono professionisti che sappiano trasformare le informazioni prodotte dai calcolatori e dagli algoritmi di advanced analytics in indicazioni rilevanti ai decisori, istituzionali o

di business. Il report della Commissione sottolinea anche come sia necessario uno sforzo formativo congiunto dell'università e dell'industria per riuscire a colmare il gap professionale, già rilvenate all'epoca e considerato in ulteriore crescita prospettica nel successivo decennio. In tal senso possiamo sottolineare la nascita di numerosi nuovi programmi a livello universitario creati appositamente per formare le figure professionali capaci di vincere le sfide della data science.

### **Gli elementi critici del machine learning a supporto delle decisioni**

Alcune riflessioni critiche sulla rivoluzione in atto sono d'obbligo. La prima è positiva: il fatto che i decisori utilizzino dati e previsioni sempre più affidabili per decidere non può che portare a un miglioramento del processo decisionale. Come sottolineava Laplace già due secoli fa, il supporto della teoria o di una formulazione quantitativa è un sostegno indispensabile alla debolezza della mente umana nel momento in cui deve prendere decisioni difficili. D'altra parte però, il pericolo è di lasciare che le decisioni vengano prese in automatico dalle macchine senza una supervisione o un intervento umano. Infatti un algoritmo di advanced analytics segue ciecamente regole predefinite e dettate da linee di codice che rappresentano comunque un'astrazione del problema e non possono tenere in conto tutte le combinazioni possibili e di tutte le peculiarità di situazioni specifiche. Si pensi a un algoritmo che è in grado di assegnare o meno un prestito. Se utilizzassimo come regola semplificata dell'algoritmo la seguente: "qualora il soggetto richiedente non sia stato in grado di ripagare un prestito negli anni precedenti, il richiedente non può avere diritto a un nuovo prestito", sicuramente tutti i soggetti che, anche se per poco, non hanno ripagato una rata nei precedenti anni, non potranno avere accesso al credito per un periodo indefinito, anche se la loro situazione finanziaria è nettamente migliorata. E quindi avremmo una negazione ingiusta, ma automatica, del diritto di accesso al credito. In tal senso, quindi, il legislatore ha posto dei limiti temporali. Quindi, riproviamo in un modo alternativo: "qualora il soggetto richiedente non sia stato in grado di ripagare un prestito nei tre anni precedenti, il richiedente non può ricevere il prestito". In questo caso, tutti i soggetti che, anche se per poco, non hanno ripagato una rata nei precedenti tre anni, non potranno avere accesso al credito per i tre anni successivi dal loro ritardo di pagamento. Tuttavia, un esame separato della situazione patrimoniale del richiedente o un miglioramento netto della sua situazione finanziaria potrebbe indurre un risk manager a proporre un override della regola, rilasciando il prestito, mentre tale miglioramento delle condizioni di contorno non viene catturato dalla regola di decisione dell'algoritmo. Ma, a fronte di un algoritmo che analizza migliaia di dati e produce come segnale un "no, non concediamo il prestito" come può un essere umano, in base a sue sensazioni specifiche, ribaltare tale analisi? È l'umano in grado di prendersi questa responsabilità?

---

## Le potenzialità del machine learning

---

Alcuni algoritmi avanzati di machine learning, come per esempio le evoluzioni più recenti dei neural networks (deep learning), sono diventati il metodo più diffuso per una vasta gamma di compiti. In applicazioni come la classificazione delle immagini, il riconoscimento del parlato e l'elaborazione del linguaggio naturale, le reti di deep learning mostrano prestazioni sorprendenti, vale a dire una precisione predittiva che spesso supera persino l'abilità degli esseri umani. Nel 2016, AlphaGo di Google Deep Mind ha battuto Lee Sedol, uno dei migliori giocatori professionisti del gioco Go di tutti i tempi. Go è un gioco da tavolo strategico amato soprattutto in paesi quali Cina e Corea. Rispetto agli scacchi è molto basato sull'intuizione visiva e fino al 2016 non si pensava il computer avrebbe potuto battere l'uomo. Per gli scacchi invece, sappiamo che l'uomo viene battuto regolarmente dal computer già dalla famosa sfida tra Deep-Blue e Kasparov nel 1997. Nel duello tra AlphaGo di Google e Lee Sedol, la ventiseiesima mossa è stata la chiave della successiva vittoria, anche se durante la partita non era chiaro perché questa mossa potesse essere vantaggiosa. Un esperto dopo quella specifica partita ha dichiarato: "Non è una mossa umana: non ho mai visto un umano fare questa mossa" (Fan Hui, 2016). È da sottolineare pertanto che gli algoritmi di machine learning più avanzati scansionano enormi masse di dati e potrebbero rilevare relazioni e percorsi sconosciuti agli esseri umani. Questo apprendimento del sistema diventa di interesse in svariati ambiti, per esempio per fisici e chimici, rivelando leggi nascoste della natura, oppure, come accennato in precedenza, mostrando nuove possibili strategie di azione: i giocatori di Go hanno incorporato nel loro repertorio la mossa di AlphaGo, i giocatori professionisti di scacchi si allenano regolarmente contro il computer.

---

## L'interpretabilità dei risultati

---

L'aumento delle prestazioni delle macchine è dovuto al miglioramento dei modelli e delle metodologie di analisi, alla presenza di dati sempre migliori e a una maggiore potenza di calcolo, soprattutto grazie all'uso delle GPU e del computing parallelo. Ma questo aumento delle prestazioni non è a costo zero, di solito questi modelli ad alte performance predittive sono caratterizzati da una mancanza di interpretabilità e trasparenza nelle relazioni tra input e output che portano a specifiche previsioni.

Tra gli altri, Breiman (2001) ha enfatizzato in passato il trade-off tra prestazioni e interpretabilità. Da lungo tempo è pertanto opinione comune che i modelli più semplici siano più facili da interpretare ma al costo di prestazioni predittive inferiori (si veda anche Rudin, 2019). Negli anni più recenti, uno degli sforzi dei ricercatori che lavorano sullo sviluppo di modelli avanzati, è stato rivolto a indebolire questa affermazione, sviluppando metodi che mantengano elevate performance predittive ma che forniscano maggiore interpretabilità.

Occorre sottolineare che la rilevanza degli aspetti di interpretabilità dei modelli non è omogenea nelle diverse applicazioni. Se si pensa a un algoritmo di riconosci-

mento delle immagini o del parlato o anche in alcuni compiti come giocare a Go, il punto chiave è la performance e l'interpretabilità è un obiettivo minore. Ma ci sono altre applicazioni in cui è inevitabile cercare un equilibrio tra interpretabilità e capacità predittiva e sicuramente la maggioranza delle applicazioni manageriali ricade tra queste: si pensi per esempio alle problematiche di studio della propensione ad acquistare un prodotto specifico da parte di un cliente, al tema della loyalty della clientela (churn/attrition) o della segmentazione comportamentale della clientela. Tutte queste applicazioni richiedono che l'analista sia in grado di fornire un'interpretazione dei risultati sia per comunicare al management i risultati, sia per essere in grado di implementare le azioni più opportune per raggiungere gli obiettivi di miglioramento delle condizioni del business che sono la premessa all'idea stessa di applicare un modello di machine learning.

Inoltre l'interpretabilità di un modello può anche rivelare le sue debolezze, fornendo per esempio una spiegazione del perché il modello fallisca o addirittura mostrando che la decisione suggerita dal modello è matematicamente corretta, ma basata su dati imprecisi o misleading. Conoscere a fondo queste debolezze dell'algoritmo è il primo passo per cercare di migliorare il modello.

L'interpretabilità va anche di pari passo con la *fiducia*. Un decisore non può fidarsi ciecamente delle previsioni quando non ha assolutamente idea di come sono stati ottenuti i risultati. Questo è particolarmente vero in campi delicati come la medicina, dove è fondamentale sapere perché una certa diagnosi è stata fatta dalla macchina. Un medico deve verificare l'output della black-box, in modo che capisca a fondo le ragioni di un certo output e possa con fiducia seguire il suggerimento della macchina. Le condizioni non sono molto diverse in ambito manageriale, in quanto la condivisione delle decisioni e delle loro motivazioni è quasi sempre un elemento fondamentale nel processo decisionale aziendale.

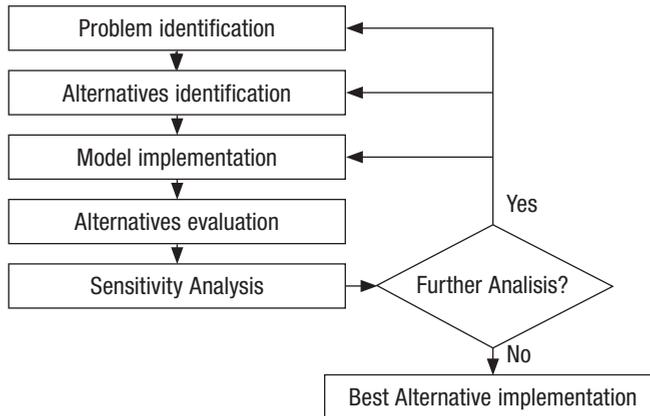
Ancora, e questo si collega al tema della fiducia, si deve cercare di disporre (e a tutt'oggi non siamo vicini) di un'adeguata quantificazione dell'*incertezza*. Anche questo aspetto è particolarmente importante in settori come la medicina, come sottolineato da Begoli, Bhattacharya e Kusnezov (2019). Gli autori evidenziano che queste moderne scatole nere basate sull'analisi dei dati, e in particolare l'IA medica, hanno bisogno di sviluppare un principio di quantificazione e disciplina dell'incertezza (Uncertainty Quantification) rigorosa.

In tutto questo, non sono poi trascurabili i temi legali. Con il ruolo crescente degli algoritmi di IA arriva la richiesta di regolazione. I decisori sono spesso obbligati a giustificare le decisioni prese, soprattutto in settori delicati, quali, come indicavamo, quello bancario o penale. È noto che gli algoritmi possano apprendere regole o usino dati che nascondano una qualche forma di distorsione, per il modo in cui sono stati raccolti, quindi persone con caratteristiche specifiche potrebbero avere una probabilità molto minore di ottenere un prestito o *più alta di essere condannate* (Rudin, 2019). L'Unione Europea ha in questo senso adottato il "diritto alla spiegazione", un nuovo regolamento che consente agli utenti di chiedere spiegazioni sulle decisioni prese in con il supporto dell'intelligenza artificiale e l'interpretabilità dei modelli gioca un ruolo estremamente rilevante.

## Il processo decisionale moderno

Arriviamo quindi al moderno processo di decisione. Il processo è stato lungamente studiato in letteratura (si veda, tra gli altri, Clemen, 1997). In **Figura 1** vengono illustrati i passi principali.

**Figura 1** Il moderno processo di decisione



La prima fase è quella dell'individuazione del problema. Per esempio, un'azienda che voglia capire meglio il posizionamento percettivo del proprio brand nel mercato potrebbe essere interessata a effettuare un'analisi di customer churn.

Tra le alternative disponibili possiamo considerare quella di effettuare l'analisi in-house tramite personale dell'azienda stessa o di affidare l'analisi a un fornitore esterno. A parità di risorse temporali ed economiche, la prima ha come condizione che l'azienda possieda internamente le competenze necessarie all'analisi. Il passo successivo è quello della formulazione di un modello. Il "modello" è l'insieme dell'algoritmo di predizione e dei dati disponibili su cui tali predizioni vengono effettuate. Qui entriamo in un mondo complesso, ovvero quello della preparazione dei dati e della selezione dell'algoritmo predittivo. In un certo senso, tra le alternative disponibili abbiamo, in linea di principio, infiniti modelli matematici. Consideriamo, per esempio un'analisi di churn. In questo caso l'analista potrebbe scegliere se optare per un modello di regressione logistica, per un modello di classificazione ad albero, per un Random Forest o un Gradient Boosting. E abbiamo anche infinite variabili su cui applicare i modelli. Ma è chiaro che non tutte le variabili hanno senso per il problema in questione. Una volta individuate le variabili che hanno senso concettuale per il problema, occorre poi raccogliere informazioni quantitative su di loro orientando la raccolta dati.

Una volta implementato il modello, si esamineranno i risultati ottenuti. Dall'analisi di sensibilità e robustezza del modello e da un attento esame del processo

---

avvenuto sino a quel momento, si potrà capire se i risultati sono soddisfacenti e si può passare alla fase di implementazione (per ulteriori dettagli, si veda Borgonovo, 2017). In caso contrario, si torna a ripetere e raffinare alcune delle attività precedenti in maniera tale da arrivare a una deliberazione il più possibile robusta.

Lo sviluppo della disciplina è tale che molte aziende stanno creando al proprio interno team di data science. Questa scelta ha il vantaggio di permettere all'azienda di internalizzare le competenze necessarie, arricchendo le risorse umane aziendali in una funzione critica per il supporto al management. Ha anche il vantaggio di una gestione tendenzialmente internalizzata dei dati che da una parte contribuisce a minimizzare i rischi legati al tema della riservatezza e dall'altra ne permette anche un monitoraggio costante e diretto.

Ovviamente, la creazione di un team di data science ha anche dei costi associati in termini di risorse umane e informatiche (computazionali e di storage) che vanno a quel punto interamente coperti dall'azienda e sono continui nel tempo. Al tempo stesso, se la dimensione dell'azienda o le possibilità di investimento sono relativamente limitate, la soluzione più flessibile e comunque praticabile è quella di un'esternalizzazione del processo di gestione e/o di analisi dei dati, attraverso l'impiego di provider esterni qualificati.

Nel nostro volume ci prefiggiamo di introdurre il lettore a questa nuova e complessa disciplina. La prima parte prevede un'introduzione alle principali metodologie di data ingestion, data management e data analysis, con una particolare attenzione anche alle tematiche della data preparation e della valutazione dei modelli predittivi. Nella seconda parte vengono presentate alcune applicazioni selezionate in ambito business, basate su Small e Big Data.