

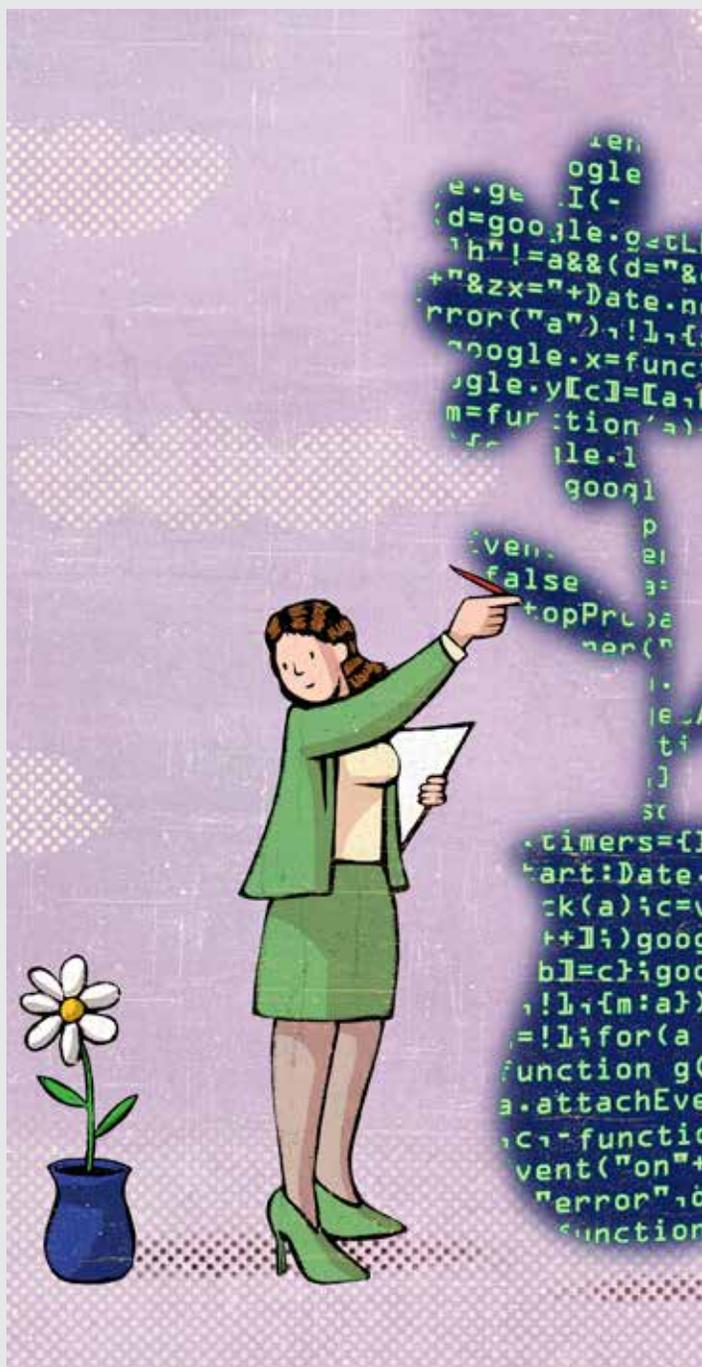
Gli advanced analytics per le decisioni manageriali

di Emanuele Borgonovo, Luca Molteni*

GESTIONE & APPLICAZIONE

La data-driven economy è iniziata da qualche anno. La rivoluzione digitale, l'aumento della connettività, l'istantaneità della diffusione delle informazioni stanno rivoluzionando il modo in cui operiamo e lavoriamo, a livello individuale, aziendale e istituzionale. Ci si rende sempre più conto dell'importanza dei dati. Il momento di emergenza stesso in cui stiamo vivendo ci spinge sempre più a utilizzare informazioni quantitative: i gradi di libertà della vita di milioni di cittadini vengono decisi in base a dati sulla diffusione della pandemia che, opportunamente studiati tramite modelli matematico-statistici, determinano il colore della regione in cui viviamo.

L'umanità sta creando quantità di dati sempre maggiori, e lascia enormi tracce digitali dei propri gusti, bisogni, consumi, attività, interessi. Macchine opportunamente addestrate sono in grado di svolgere compiti sempre più sofisticati che in molti casi aiutano l'uomo e ne amplificano capacità e risultati. Si parla dunque sempre più spesso di intelligenza artificiale (artificial intelligence, AI). Da una parte abbiamo quella parte di AI che colpisce la nostra immaginazione: robot autonomi che sostituiscono gli uomini in città futuristiche, completamente automatizzate (ma è una realtà da cui fortunatamente siamo ancora lontani). Dall'altra abbiamo l'AI che, basata sul machine learning, è molto vicina a noi e sta già avendo un impatto non trascurabile sulla nostra 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.



* Emanuele Borgonovo è Professore Ordinario presso il Dipartimento di Scienze delle Decisioni presso l'Università Bocconi, Professore di Decision Sciences & Business Analytics e Direttore del Management Science Laboratory presso SDA Bocconi School of Management.

Luca Molteni è Ricercatore presso il Dipartimento di Scienze delle Decisioni dell'Università Bocconi e Department of Decision Sciences Liaison Officer presso SDA Bocconi School of Management.

Sempre più imprese, piccole e grandi, si rendono conto della potenzialità che risiedono nei dati a propria disposizione. Scoprire trend, pattern, comportamenti in anticipo è fattore di vantaggio competitivo.

LA «DIGITAL ECONOMY»

Anche dal punto di vista istituzionale da qualche anno ci si è resi conto del potenziale impatto delle nuove metodologie di analisi. Nel 2012 è nata infatti l'Agenzia per l'Italia Digitale (Agid); 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 sta ancora oggi) per essere travolta e stravolta dall'enorme massa di dati generati dall'umanità e dall'altrettanta veloce digitalizzazione dei processi di decisione. Il documento 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 per decisori, istituzioni, imprese. Il report della Commissione evidenzia inoltre come sia necessario uno sforzo formativo congiunto dell'università e dell'industria per riuscire a colmare il gap professionale, già rilevante 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.

IL MACHINE LEARNING A SUPPORTO DELLE DECISIONI

Alcune riflessioni critiche sulla rivoluzione in atto sono d'obbligo. La prima è positiva: il fatto che si 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, lo stesso non potrà 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 a un nuovo credito per un periodo indefinito, anche se nel frattempo 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. Riproviamo ora in un modo alternativo: «Qualora il soggetto richiedente non sia stato in grado di ripagare un prestito nei tre anni precedenti, lo stesso non potrà ricevere un nuovo prestito per i tre anni successivi». 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 comunque il prestito; un miglioramento delle condizioni di contorno che non sarebbe stata catturata 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?

L'IMPATTO DEL COVID-19 SULLA DATA SCIENCE

Uno degli aspetti più critici che stiamo affrontando in questi mesi è sicuramente l'impatto del Covid-19 sulle nostre vite, attività, istituzioni. Siamo più isolati di prima, con contatti molto più digitali, costretti a rivedere i nostri consumi. È chiaro che la pandemia avrà un impatto non trascurabile sul mondo della data science. I modelli di previsione – e, in particolare, di quelli basati sulle serie storiche – è stato e sarà nei prossimi mesi fortemente impattato dalla discontinuità delle serie storiche stesse, suggerendo di integrare strumenti di ricerca di natura più qualitativa e di scenario sul comportamento della domanda¹. Inoltre, se da un lato l'importanza dell'analisi dei dati in generale e dei big data in particolare emerge ancora più nettamente in un periodo critico come questo, dall'altro gli investimenti delle aziende in difficoltà potrebbero incontrare una contrazione per effetto della riduzione dei loro margini; tali investimenti potrebbero contrarsi sia in relazione all'assunzione di personale dedicato, sia per l'acquisizione di servizi esterni di analytic non giudicati essenziali, sia, infine, per l'acquisizione di strumenti di gestione e analisi dei dati. L'aumento del lavoro in smart working sta inoltre creando maggiori minacce alla sicurezza delle aziende, in quando si è assistito a una forzata accelerazione verso il digitale e lo scambio di documenti e informazioni all'esterno dell'azienda; questa condizione di lavoro potrebbe pertanto favorire investimenti ulteriori nell'ambito della cybersecurity.

Più in generale, in questi mesi di grande difficoltà si chiede alle aziende maggiore efficienza e capacità di scaricare a terra il potenziale delle applicazioni di machine learning, con soluzioni efficaci e capaci di aiutare concretamente il processo decisionale di business.

LE POTENZIALITÀ DEL MACHINE LEARNING

Prescindere dalle difficoltà del momento legate alla pandemia, alcuni algoritmi avanzati di machine learning, come per esempio le evoluzioni più recenti delle reti neurali artificiali (deep learning), sono ormai già 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, con una precisione predittiva che supera spesso l'abilità degli esseri umani. Nel 2016, AlphaGo di Google DeepMind ha battuto Lee Sedol, uno dei miglior giocatori di go di tutti i tempi. Go è un gioco da tavolo strategico amato soprattutto in Cina e Corea. Rispetto agli scacchi è molto basato sull'intuizione visiva e fino al 2016 non si pensava che un computer avrebbe potuto battere un uomo per il tipo di compiti che il gioco richiede. 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 Alpha Go 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. Dopo quella specifica partita, Fan Hui, esperto goista, ha dichiarato: «Non è una mossa umana: non ho mai visto un umano fare questa mossa». È da sottolineare pertanto che gli algoritmi di machine learning più avanzati scansionano enormi masse di dati e potrebbero rilevare relazioni e percorsi che gli esseri umani non riescono a percepire. 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 i 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 Graphics Processing Unit-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 ha enfatizzato il trade-off tra prestazioni e interpretabilità². Da lungo tempo è pertanto opinione comune che i modelli più semplici siano più facili da interpretare ma al costo di prestazioni predittive inferiori³. Negli anni più recenti, uno degli sforzi degli 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 riconoscimento 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 legate alla propensione ad acquistare un prodotto specifico da parte di un cliente, al tema della loyalty della clientela (churn/attrition) o a quello 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 questo fallisca o addirittura mostrando che la decisione suggerita sia matematicamente corretta, ma basata su dati imprecisi o misleading. Conoscere a fondo le debolezze dell'algoritmo è il primo passo per cercare di migliorarlo.

L'interpretabilità va anche di pari passo con il tema della fiducia. Un decisore non può fidarsi ciecamente delle previsioni quando non ha assolutamente idea di come siano 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 ne capisca a fondo le ragioni e possa con fiducia seguire il suggerimento della macchina. Le condizioni non sono molto diverse in ambito manageriale, in quanto la condivisione delle scelte fatte 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

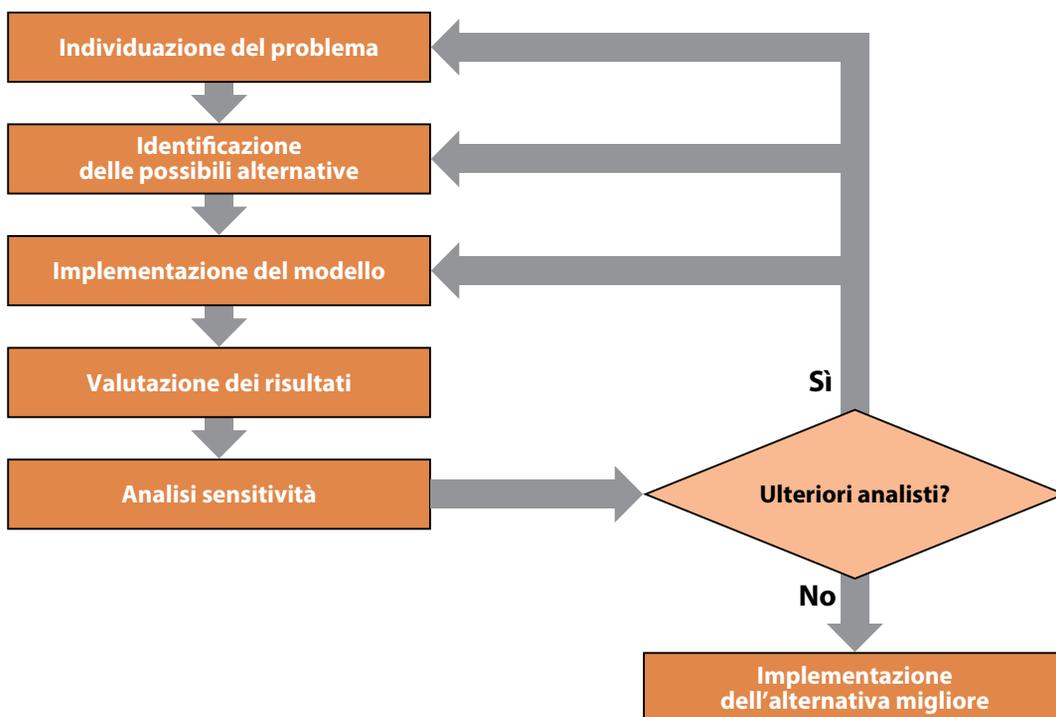
ad averla) di un'adeguata quantificazione dell'incertezza. Anche questo aspetto è particolarmente importante in settori come la medicina⁴. Le moderne scatole nere basate sull'analisi dei dati, e in particolare l'IA medica, hanno infatti bisogno di sviluppare un principio di quantificazione e disciplina 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 quello bancario o penale. È noto che gli algoritmi possano apprendere regole o usare dati che, per il modo in cui sono stati raccolti, nascondono una qualche forma di distorsione, quindi, tornando all'esempio iniziale, persone con caratteristiche specifiche potrebbero avere una probabilità molto minore di ottenere un prestito⁵. 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 con il supporto dell'intelligenza artificiale. Un ambito in cui l'interpretabilità dei modelli gioca un ruolo estremamente rilevante.

IL MODERNO PROCESSO DECISIONALE⁶

Arriviamo quindi al moderno processo di decisione. Il processo è stato lungamente studiato in letteratura⁷. Nella Figura 1 vengono illustrati i passi principali.

Figura 1 - Il moderno processo decisionale



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

Tra le alternative disponibili possiamo considerare quella di svolgerla 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 opzione ha come condizione dirimente quella di possedere internamente le competenze necessarie per farlo.

Il passo successivo è quello della formulazione di un modello. Questo 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 quelle pertinenti 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. In questo caso si può passare alla fase di implementazione. In caso contrario, si torna a ripetere e raffinare alcune delle attività precedenti in maniera tale da arrivare a una deliberazione il più robusta possibile.

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 interna 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. Di contro, se la dimensione dell'azienda o le possibilità di investimento sono relativamente limitate, la soluzione più flessibile e comunque praticabile è quella di esternalizzare il processo di gestione e/o di analisi dei dati attraverso l'impiego di provider qualificati⁸.

¹ A. Dearmer, «How Has Covid-19 Impacted Data Science?», *Xplenty*, 16 dicembre 2020.

² L. Breiman, «Statistical Modeling: The Two Cultures», *Statistical Science*, 16(3), 2001, pp. 199-231.

³ C. Rudin, «Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead», *Nature Machine Intelligence*, 1(5), 2019, pp. 206-215.

⁴ Si veda in proposito E. Begoli, T. Bhattacharya, D. Kusnezov, «The Need for Uncertainty Quantification in Machine-Assisted Medical Decision Making», *Nature Machine Intelligence*, 1(1), 2019, pp. 20-23.

⁵ Rudin, op. cit.

⁶ Questa sottosezione è in parte ripresa da E. Borgonovo, L. Molteni, Quando ai manager danno i numeri, Milano, Egea, 2020.

⁷ Si veda, tra gli altri, R.T. Clemen, Making Hard Decisions: An Introduction to Decision Analysis, Pacific Grove, Duxbury Press, 1997.

⁸ Nel volume Quando ai manager danno i numeri, op. cit. ci siamo prefissi di introdurre il lettore a questa nuova e complessa disciplina, presentando una panoramica delle 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; all'interno del medesimo volume è possibile ritrovare alcune applicazioni selezionate in ambito business, basate su small e big data.

IN SINTESI

- L'equilibrio tra capacità predittiva della macchina e l'interpretabilità del dato da parte del management è oggi un requisito fondamentale per raggiungere gli obiettivi aziendali e migliorare il proprio business.
- In molti ambiti manageriali, l'interpretabilità del dato va di pari passo con il tema della fiducia: un decisore ha chiaramente qualche remora a fidarsi ciecamente delle previsioni elaborate da sistemi di intelligenza artificiale se non ha una minima idea di come siano state ottenute.
- Sempre più spesso, la definizione di un adeguato modello decisionale basato sulle informazioni prodotte dagli algoritmi di advanced analytics è affidato dalle aziende a un proprio team interno di data science. Questa scelta ha il vantaggio di permettere all'azienda di internalizzare le competenze necessarie e di gestire autonomamente i propri dati, minimizzando i rischi di cybersecurity.