

Introduzione

La data analytics è diventata *pop*! Ripulire e analizzare dati, costruire modelli di machine learning, visualizzare evidenze numeriche e utilizzarle per creare storie per influenzare le decisioni: queste abilità non sono più tecnicismi riservati ai *pochi* professionisti del dato, come data scientist e business analyst, ma sono nuove competenze di base per chiunque abbia a che fare con numeri nel proprio lavoro, quindi praticamente per *tutti*. Qualunque sia stato il vostro percorso di studi, la vostra posizione nell'organigramma aziendale o l'area di specializzazione funzionale (vendite, marketing, finanza, logistica, risorse umane e così via) la data analytics può portare benefici concreti e rilevanti nel vostro lavoro. Gli argomenti trattati nel libro che avete tra le mani rispondono al bisogno universale di masticare dati e algoritmi in modo da poterli sfruttare al meglio: la data analytics è davvero per tutti e non c'è dubbio che lo sia anche per voi!

Tutto questo è facile a dirsi ma sicuramente complicato da mettere in pratica. Molte persone interrompono (e talvolta rinunciano a intraprendere) questo percorso educativo scoraggiate da una barriera apparentemente insormontabile: imparare a scrivere codice in un linguaggio di programmazione. Presi dalle mille priorità che ci si trova ad affrontare quotidianamente, si fa fatica a trovare il tempo e l'energia per imparare e programmare. Fortunatamente non siamo obbligati a diventare esperti di programmazione informatica per beneficiare appieno della forza di dati e algoritmi. Oggi esistono strumenti che ci permettono di *fare analytics* senza scrivere una linea di codice: in questo libro imparerete a usare proprio questi strumenti, chiamati *low-code*. Disaccoppiando il bisogno di imparare a programmare da quello di utilizzare gli strumenti della data analytics riusciremo a carpire da subito i notevoli vantaggi offerti dai tanti dati nei quali siamo immersi (o nei quali ci sentiamo addirittura impantanati). Questa è la promessa che vi faccio con questo libro: rendere la data analytics uno strumento utile per semplificare e migliorare sensibilmente il lavoro quotidiano.

Questa guida offre un percorso accessibile e, mi auguro, appassionante, per passare rapidamente dalla teoria alla pratica della data analytics utilizzando strumenti low-code. Malgrado una parte rilevante del libro copra concetti di applicazione generale che incontrereste utilizzando qualsiasi software voi o la vostra azienda decidiate di utilizzare, i tutorial che affronterete sono basati su KNIME e Power BI. Ho selezionato questi due applicativi in quanto rappresentano una combinazione ideale di compagni di viaggio per entrare nel mondo della data analytics.

- KNIME è stato descritto come il “coltellino svizzero” tra le piattaforme analitiche. La sua semplice interfaccia grafica permette a tutti (manager, analisti e studenti) di automatizzare flussi di dati e implementare algoritmi di apprendimento automatico senza scrivere una linea di codice. È importante chiarire da subito che strumenti basati sulla programmazione visuale (*visual programming*) non sono versioni “alleggerite” e “depotenziate” di quegli strumenti più avanzati che, in quanto tali, devono necessariamente utilizzare del codice. Sapendolo usare bene, KNIME vi permette di implementare tutta la data science di cui avete bisogno, incluse la creazione di architetture complesse di data engineering e la predisposizione di modelli di intelligenza artificiale avanzati, come le reti neurali del *deep learning*. Senza entrare ora nel dettaglio di che cosa questo voglia dire, fidatevi: gli strumenti che imparerete permettono di “fare sul serio”.
- Power BI di Microsoft è tra gli strumenti più popolari al momento per la visualizzazione dei dati e la predisposizione di dashboard. Imparando a utilizzarlo potrete facilmente creare interfacce interattive che permettano a voi e ai vostri colleghi di accedere ai dati in modalità *self-service* e di supportare il *decision making* a tutti i livelli e nel migliore dei modi.

Sia Power BI che KNIME possono essere liberamente scaricati da chiunque (voi lo farete tra poco) ed essere utilizzati gratuitamente in gran parte delle loro funzionalità, in modo da iniziare a generare valore da subito senza alcun investimento iniziale (se non di ciò che è più importante e scarso, ovvero il vostro tempo). La combinazione di uno strumento di *back-end* come KNIME e di uno di *front-end* come Power BI (o di Tableau, che verrà introdotto come valida alternativa) è versatile quanto robusta e permette di coprire a 360 gradi i bisogni informativi di un’organizzazione, dalle analitiche descrittive a quelle predittive e, infine, prescrittive. Il libro è corredato di diversi tutorial basati su casi di business e dati reali: questi tutorial vi prenderanno per mano e condurranno nella pratica quotidiana della data analytics, rendendovi utenti autonomi della potente “accoppiata” formata da Power BI e KNIME.

Ho progettato questo percorso di apprendimento concentrandomi sui bisogni di tre tipologie di lettori.

1. Professionisti di qualsiasi background, livello e area di specializzazione funzionale che utilizzano dati nel loro lavoro per prendere decisioni o permettere ad altri di farlo in maniera informata, attraverso la predisposizione di report, presentazioni e analisi di approfondimento *ad hoc*. Se vi ritrovate in questa categoria, sentirete l’esigenza di modernizzare ed espandere la vostra conoscenza in modo da semplificarvi la vita nelle attività regolari di pulizia e trasformazione dei dati. Volete aumentare l’impatto sul business del vostro lavoro e sentite che gli strumenti che utilizzate oggi (probabilmente Excel) vi stanno stretti. Vorreste passare a strumenti più potenti in modo da automatizzare i processi e aumentare la vostra produttività personale ma, non avendo esperienza nella programmazione, non sapete da dove partire. Avreste voglia di costruire dashboard interattive che si aggiornino automaticamente in modo da potervi concentrare su aspetti più interessanti rispetto all’aggiornamento regolare di tabelle (per cui vi sentite davvero sprecati), come quello di interpretare le evidenze numeriche e proporre azioni concrete. Se vi sentite chiamati in causa, questo libro è proprio quello che fa per voi e vi fornirà strumenti e tecniche per raggiungere questi obiettivi.

2. Manager aziendali con qualsiasi background funzionale, inclusi marketing, vendite, finanza, IT e risorse umane, che desiderano acquisire esperienza diretta nell'analisi dei dati per comprendere e sbloccare il vero potenziale dei dati nella propria area di responsabilità. Se appartenete a questa classe, non avrete necessariamente competenze tecniche di data analytics, né l'intenzione di diventarne specialisti. In ogni caso vorreste capire "che cosa è possibile fare" e guidare i vostri team attraverso il processo di trasformazione che è necessario affrontare. Vorreste essere una fonte di aiuto per i vostri collaboratori e non sentirvi, invece, un peso: per questo motivo, siete pronti a rimboccarvi le maniche, studiare e dare l'esempio. Se vi sentite così, non siete soli: ho incontrato numerosi senior leader che intravedono il potenziale valore della data analytics ma che non sanno come realizzarlo nelle loro organizzazioni. Questo libro vi aiuterà a toccare con mano le opportunità di creazione di valore e ad affinare adeguatamente le aspettative nei confronti delle vostre persone. Non è forse vero che tutti noi desideriamo avere capi che conoscano sempre bene quello che chiedono senza pretendere obiettivi irrealizzabili? Questo libro vi renderà manager consapevoli, fiduciosi e in grado di definire aspettative sfidanti e ragionevoli circa l'uso dei dati aziendali per impattare il business.
3. Studenti di economia o data science e professionisti junior che posseggono conoscenze teoriche e accademiche ma che hanno un livello di esperienza aziendale limitata o del tutto assente. Se vi sentite chiamati in causa, il vostro desiderio è quello di scoprire in prima persona come le vostre conoscenze possano essere effettivamente messe in pratica, in modo da impressionare positivamente i vostri responsabili e garantirvi un inizio di carriera accelerato. Dopo aver appreso tanta teoria, suddivisa in compartimenti stagni tra discipline (come informatica, statistica ed economia aziendale) avete un chiaro desiderio: sperimentare la realtà di quel processo end-to-end che porta dai dati grezzi a vere decisioni di business. Se siete alla ricerca di questi aspetti pratici e di un libro ricco di esempi che vengono dal mondo reale, quello che avete tra le mani è esattamente ciò che vi serve.

Fortunatamente, nessuno di questi tre tipi di lettori avrà bisogno di conoscenze preliminari specifiche per affrontare al meglio il percorso che sta per iniziare: la matematica che troverete nei vari capitoli è sempre introdotta in maniera molto delicata e rimane al livello necessario per mettere in pratica la data analytics, lasciando ai lettori che lo vorranno fare la possibilità di approfondire. Ho scritto questo libro per rendere l'analisi dei dati accessibile a tutti coloro che vogliono investire un po' di tempo per la loro crescita: sono contento che anche voi siate tra questi!

Il viaggio di fronte a noi

Il percorso che vi aspetta nelle prossime pagine vi farà alternare nell'attraversamento di due tipi di territori: uno più teorico e concettuale, l'altro più pratico e operativo. Alcuni capitoli vi forniranno gli irrinunciabili concetti necessari per capire le fondamenta della data analytics come l'organizzazione logica dei dati, l'apprendimento automatico, la valutazione dei modelli statistici e i principi cognitivi a supporto della data visualization. Altri capitoli vi porteranno attraverso la manualità del "come" eseguire le operazioni sul vostro computer per ottenere, per esempio, una dashboard in Power BI o l'automatizzazione di un processo ripetitivo di elaborazione dei dati tramite KNIME. I due tipi

di contenuto sono complementari tra loro e si alternano lungo il cammino scandito dei nove capitoli del libro.

1. Il Capitolo 1 apre le danze, portando chiarezza sulla terminologia corretta e fornendo una serie di schemi concettuali (come i tre livelli di analytics, lo stack tecnologico a supporto dei dati, la matrice di ruoli e competenze analitiche in azienda, e i percorsi *data-to-value*) che offrono una prospettiva ampia e strutturata circa il potenziale reale di dati e algoritmi nel business.
2. Nel Capitolo 2 incontreremo KNIME e capiremo come funziona il meccanismo del visual programming, attraverso alcuni esempi pratici. Qui troverete il primo tutorial del libro grazie al quale creeremo una routine in grado di automatizzare il noioso processo di pulizia di alcuni dati generati da consumatori.
3. Il Capitolo 3 introdurrà il concetto di *data model*, mostrando come applicare in maniera sistematica trasformazioni che rendano i nostri dati pronti per essere utilizzati a fare analytics. Impareremo come combinare tabelle, aggregare quantità, applicare cicli e utilizzare variabili per automatizzare le situazioni più complesse. Nel tutorial creeremo da zero un processo automatico per la predisposizione di alcuni semplici report finanziari a partire da diverse fonti di dati grezzi.
4. Il Capitolo 4 separerà mito e realtà al riguardo dell'intelligenza artificiale. Sveleremo come funziona quel processo (vi anticipo, molto affascinante) che permette alle macchine di imparare qualcosa seguendo solo una serie di passi apparentemente banali e ripetitivi. Il capitolo fornisce le basi del machine learning, come la tassonomia degli algoritmi disponibili e la valutazione dell'accuratezza dei modelli, e vi consentirà di iniziare a intravedere quelle situazioni in cui l'intelligenza artificiale può davvero supportare il vostro lavoro quotidiano.
5. Nel Capitolo 5 metteremo in pratica l'apprendimento automatico, creando tre modelli di machine learning partendo da zero e utilizzando dati reali. Proveremo l'emozione di prevedere i prezzi degli affitti immobiliari a Roma, anticipare la reazione dei clienti di una banca di fronte a una campagna di marketing e segmentare gli utenti registrati su un sito di e-commerce. Seguendo i tutorial, otterrete i vostri primi *template*, ovvero dei modelli riutilizzabili che potrete riapplicare in futuro, adattandoli di volta in volta alle necessità specifiche che avrete di fronte.
6. Il Capitolo 6 insegnerà a rispondere a un bisogno immancabile in ogni azienda: creare interfacce chiare ed efficaci per accedere ai dati in maniera sistematica. Nel tutorial costruiremo una dashboard completa utilizzando Power BI. Capiremo come preparare il data model sottostante e come predisporre grafici e tabelle connessi tra loro, in modo da permettere l'esplorazione delle tante informazioni a disposizione in maniera versatile e interattiva. Una dashboard di questo tipo potrà aiutare noi stessi a presentare i dati nella maniera più opportuna e permetterà anche ad altri di accedere facilmente alle informazioni di cui hanno bisogno: il risultato ottenuto sarà la "democratizzazione" dei dati in azienda, con tutti quei diffusi vantaggi di efficienza e di consapevolezza che ne conseguono.
7. Nel Capitolo 7 impareremo una serie di tecniche di *data visualization*: queste ci serviranno a costruire grafici accattivanti con i quali spiegare efficacemente quello che i dati vogliono dirci. Il capitolo ci fornirà una metodologia per selezionare il tipo corretto di grafico in base alle nostre specifiche esigenze e una serie di linee guida di

visual design che assicurino che i nostri messaggi sul business arrivino a destinazione in maniera nitida, forte e chiara.

8. Il Capitolo 8 mostrerà come utilizzare le tecniche di *data storytelling* per preparare e condividere storie avvincenti e persuasive attraverso l'uso dei dati, in modo da guidare opportunamente il decision making. Partendo dai principi di base della persuasione, questo capitolo proporrà diverse tecniche per raccontare le nostre evidenze in maniera diretta ed efficace. Il capitolo ci lascerà con uno schema strutturato (*Data Storytelling Canvas*) da tenere a portata di mano e riutilizzare nel momento del bisogno.
9. Nel Capitolo 9, ultima tappa del nostro viaggio, apriremo i nostri orizzonti e daremo un'occhiata a ciò che ci attende *oltre* gli strumenti e le tecniche incluse in questo libro. Averne contezza ci permetterà di pianificare al meglio i nostri prossimi passi nello sviluppare le competenze di cui abbiamo bisogno. Il capitolo contiene alcune semplici dimostrazioni pratiche di Tableau, Python (inclusa la sua integrazione con KNIME) e H2O.ai per capire come sia possibile automatizzare ulteriormente il processo di machine learning. Questo capitolo dimostrerà il valore generale (ed espandibile) dell'ampio bagaglio di competenze che avremo acquisito completando l'intero percorso.

Consigli di viaggio

A conclusione di questa breve introduzione, permettetemi di condividere alcuni suggerimenti che vi aiuteranno lungo il percorso che state per iniziare.

- Il libro include numerosi tutorial passo-passo i quali vanno considerati come parte integrante del percorso formativo e non come esercizi facoltativi. Alcuni degli aspetti più sottili e affascinanti della data analytics (come la necessità di interagire continuamente con i propri colleghi o il bisogno di procedere “a tentativi” nella costruzione di un modello di machine learning) possono essere compresi appieno solo affrontando in prima persona esempi reali, come quelli offerti dai tutorial. Il mio consiglio è organizzare la propria agenda per concedersi il tempo necessario per ogni tutorial, considerando che ciascuno potrebbe richiedere fino a due ore per il suo completamento. Assicuratevi, inoltre, di procurarvi un computer su cui poter liberamente installare tutto il software necessario allo svolgimento dei tutorial.
- In coda al libro troverete una breve serie di note organizzate per capitoli. Queste sono state selezionate per offrirvi l'opportunità di integrare il percorso del libro con letture di approfondimento. Non dimenticate di dare un'occhiata alle note dopo aver completato un capitolo in modo da scegliere quelle che raccolgono la vostra curiosità e approfondirle opportunamente.
- A seconda del vostro background, potreste trovare alcune parti del libro meno “naturali”: questo è assolutamente normale. Se sentite che qualche passaggio vi dà più filo da torcere, non fatevi prendere dallo sconforto. I capitoli sono tutti strettamente interconnessi tra loro e potreste trovare le risposte ai vostri dubbi qualche pagina più in là, quindi il consiglio è di non scoraggiarsi e continuare a leggere.
- Tutti i dati necessari per completare i tutorial sono disponibili nelle pagine dedicate sul sito dell'editore (<https://bit.ly//apo-dapt>), su quello dell'autore

(www.aboutbigdata.net) e sul portale KNIME Hub (<http://tiny.cc/knimehub>). In queste pagine troverete anche il “risultato finale” di ogni tutorial (come il file contenente il workflow di KNIME che raccoglie tutti i passi previsti o la dashboard di Power BI ottenuta). Se vi sentite persi in qualche passaggio dei tutorial potrete sempre fare riferimento a questi file per ritrovare la via.

- I software migliorano e si aggiornano continuamente. Questo libro si basa sulle ultime versioni di KNIME e Power BI disponibili al momento della sua scrittura (per la precisione: KNIME 4.4 e Power BI 2.93). La maggior parte del contenuto rimarrà valida a lungo, ma è possibile che alcuni passaggi di un tutorial cambino leggermente, rendendo le finestre leggermente diverse rispetto a quelle che trovate nelle figure. Questo è inevitabile ma non comprometterà il vostro percorso di apprendimento. Tenete d’occhio le pagine web relative al libro per eventuali errata corrige o aggiunte che pubblicherò in caso di divergenze significative dovute a una versione più recente del software. Potete anche mettervi in contatto direttamente con me utilizzando i recapiti che troverete nelle conclusioni.
- Alcune delle immagini che incontrerete nel libro fanno uso dei colori per aumentarne la chiarezza didattica e semplificarne l’interpretazione. Per questo motivo la versione cartacea del libro è dotata di un inserto in cui è riprodotta a colori una selezione delle immagini presenti in bianco e nero nel testo. Quando specificato nella didascalia, potrete trovare l’immagine fra le tavole a colori.
- Data la natura degli argomenti trattati, incontreremo un ampio numero di parole in inglese lungo il percorso. Il loro significato, quando non ovvio, viene sempre esplicitato quando compaiono per la prima volta, in modo da rinforzare quella confidenza necessaria ad affrontare letture di approfondimento che inevitabilmente faranno uso dell’inglese. Ho sempre mantenuto la regola di ortografia italiana di non declinare al plurale le parole straniere: è per questo che parlo di “data scientist” o “capability” pur riferendomi talvolta a più di un’unità.
- Questo libro ha una chiara propensione verso l’applicazione pratica della data analytics in contesti operativi di business. Per questo motivo, ogni volta che ho dovuto scegliere tra una rigorosa definizione matematica e una spiegazione pragmatica e intuitiva di un metodo analitico, ho sempre preferito intraprendere quest’ultima via. La logica è che ciascuno di voi avrà sempre il tempo per apprendere metodologie più complesse e migliorare l’accuratezza dei propri modelli, approfondendo la matematica alla base di algoritmi più sofisticati. La priorità sarà sempre quella di consentirvi di utilizzare la data analytics nel lavoro quotidiano più che fornirvi la descrizione formale e completa di ogni singolo concetto matematico che incontreremo.

Spero che tutte queste premesse vi abbiano dato la carica per partire. Non ci resta che iniziare con il primo capitolo che ci offrirà tutte le basi che ci servono per affrontare al meglio il nostro percorso nell’affascinante mondo della data analytics. Buon viaggio!