

Focus: una “nuova” ondata di informazione statistica

La disponibilità di informazione statistica procede di pari passo con la capacità di elaborazione e di calcolo degli strumenti informatici. A questa fa seguito l’aumento nella capacità di utilizzo delle informazioni da parte degli analisti economico-statistici.

..la nuova filiera delle decisioni:

- 1. Produzione dei dati**
- 2. “Data Management”**
- 3. Data Analysis”**
- 4. “Advanced Data Analysis”**
- 5. Utilizzo: decision making**



“Ondate” di informazione statistica

La disponibilità di informazione statistica procede di pari passo con la capacità di elaborazione e di calcolo degli strumenti informatici. A questa fa seguito l'aumento nella capacità di utilizzo delle informazioni da parte degli analisti economico-statistici.

La fine del secolo scorso ha visto la crescente diffusione di banche dati contenenti informazioni puntuali (es. bilanci depositati) su grandi campioni di imprese, e a volte sull'universo delle imprese, di una determinata area (regioni, paesi, etc.). L'utilizzo di una tale mole di informazioni era impensabile fino a qualche decina di anni fa, sia per capacità di calcolo sia per strumenti statistici. Oggi le istituzioni finanziarie e creditizie, ma anche le istituzioni pubbliche, ad esempio l'Agenzia delle Entrate, fanno largo utilizzo di basi dati contenenti informazioni sulle singole imprese, quali ad esempio le voci di bilancio. Dal punto di vista dell'analista, ciò richiede competenze specifiche, diverse da quelle richieste per l'utilizzo di dati aggregati.

In anni più recenti si è assistito ad un aumento esponenziale nella quantità e nella varietà dell'informazione potenzialmente disponibile, nell'ambito dei dati c.d. *non-strutturati*. Tale aumento è legato alla diffusione della rete Internet e da una serie di altre novità legate al mondo digitale:

➤ *Open Data*

La diffusione degli Open Data ha messo a disposizione un inestimabile patrimonio di informazioni in una miriade di ambiti: amministrativo, elettorale, biomedico, ambientale, sportivo, etc...

➤ Dati provenienti dai *social networks*

Colossi quali Twitter, Facebook, Instagram hanno intrapreso un percorso di liberalizzazione dei dati di traffico tramite l'accesso alle proprie API (Application Programming Interface)

➤ Dati sulle *transazioni commerciali* in formato digitale

Anche le imprese di piccole e piccolissime dimensioni hanno oramai la possibilità di fare ricorso alle vendite online (e-commerce) o di dotarsi di strumenti per la raccolta di informazioni sui clienti e le transazioni commerciali.

➤ Dati provenienti da *devices* di uso comune

Ad es.: strumenti dotati di sensori, applicazioni per smartphone etc...

➤ *Informazioni personali* catalogate in formato digitale

Chiunque oggi è in grado di salvare le proprie informazioni personali sotto forma di dati digitali, tramite programmi specifici o programmi generici di storage (es: informazioni sui partecipanti a un evento, informazioni su sessioni di allenamento etc...)



Le nuove tecnologie stanno determinando cambiamenti importanti

...aumenta la dimensione dei dataset (Big Data)

Le aziende di grandi dimensioni sono in grado di accumulare diversi terabyte di dati sulle transazioni dei propri clienti in poche ore.

Una tale mole di dati pone almeno due categorie di problemi:

- Problemi di management: gli strumenti informatici utilizzati tradizionalmente per organizzare e interrogare le banche dati non sono più sufficienti => Gli attuali software per l'analisi economico-statistica saranno rimpiazzati da altri più performanti, in grado di sintetizzare grandi quantità di informazioni e renderle utilizzabili per l'analisi economico-statistica;
- Nuovi strumenti di analisi economica e statistica. Gli strumenti statistico-econometrici per l'inferenza e la previsione, e gli apparati teorici di riferimento nei diversi ambiti disciplinari devono essere adeguati a questo nuovo genere di informazione disponibile => Nuove metodologie stanno nascendo in ambito accademico e si diffonderanno in ambito professionale.

...aumenta il numero di potenziali utilizzatori dei dati

Se il fenomeno dell'aumentata mole dei dataset riguarda principalmente le imprese di grandi e grandissime dimensioni, la loro crescente disponibilità ha coinvolto una quantità di soggetti senza precedenti.

Micro imprese, enti pubblici, organizzazioni non governative, società sportive: praticamente chiunque oggi ha la possibilità di accedere con facilità a una collezione di dati inimmaginabile sino a pochi anni fa.

Contestualmente, è sempre più diffusa l'idea che nei dati si nascondano informazioni che possono restituire una conoscenza profonda del proprio mercato di riferimento.

Cosa cambia dal punto di vista delle figure che lavorano con i dati?

È necessario un adeguamento nella struttura decisionale. Le figure professionali che pongono i dati al centro della propria attività si collocano lungo una ipotetica "filiera" che porta dai dati alle decisioni.

...una nuova filiera decisionale:

1. Produzione dei dati

Ad esempio le società di rilevazioni statistiche, le società di raccolta dati in settori specifici (alberghiero, commercio al dettaglio), società dedicate all'organizzazione di focus group, etc...



2. “Data Management”

Le attività di programmazione e gestione dei dati (c.d. Database Management System, o DBMS) sono affidate a professionalità, operanti nel campo dell'informatica, che utilizzano linguaggi di programmazione specifici (SQL) per organizzare diversi tipi di informazioni “non strutturate” in database “strutturati”. Il compito primario dell'esperto di DB management è, appunto, quello di organizzare in maniera strutturata flussi di informazioni destrutturate, in modo che queste siano più facilmente reperibili ed analizzabili. Tradizionalmente al DB manager non vengono richieste competenze avanzate di analisi in campo statistico, tantomeno in campo economico. Ad esempio, le azioni di interrogazione e query possono fornire una serie di utili statistiche di sintesi di un database che, però, non possono in alcun modo essere ritenute idonee a costituire una base informativa adeguata per le decisioni strategiche.

3. “Data Analysis”

L'analista estrae informazioni dai dati. È questo forse il campo in cui più si avverte il cambiamento determinato dalla “rivoluzione dei dati”. Per comprendere e sfruttare il potenziale informativo contenuto nei dati non strutturati si rende infatti necessario possedere competenze multidisciplinari che permettano di interagire con i soggetti implicati in ogni fase della filiera “dei dati”: dalla produzione, al management, all'utilizzo. Per riferirsi a tali figure professionali si utilizza il termine *data scientist*, o anche *data analyst*¹. A chi opera in questo campo si richiede la capacità non solo di sintetizzare e descrivere la struttura di un database ma anche, e soprattutto, di comprendere in maniera approfondita i legami che sottendono le relazioni fra gli elementi che compongono il database. Questi soggetti devono conoscere gli strumenti classici di dell'inferenza statistica ma anche strumenti nuovi, ad esempio nel campo della computer science. Il data analyst deve essere in grado di applicare una serie di metodologie standard al fine di estrarre conoscenze utili a chi dovrà prendere delle decisioni. Statistical Learning, Data Mining, Machine Learning sono definizioni (nate in ambiti scientifici differenti) utilizzate per descrivere, a grandi linee, gli strumenti di lavoro del data scientist.

4. “Advanced Data Analysis”

Il data-scientist non svolge, normalmente, analisi avanzate sui dati. Le analisi avanzate necessitano di conoscenze specialistiche in determinati ambiti e di una formazione economico-statistica di livello post-universitario. L'Advanced Data Analyst individua/seleziona le informazioni utili e “finalizza” l'analisi. Di solito non dialoga direttamente con l'informatico (che si occupa della fase di produzione/gestione del database) ma si serve del data scientist quale la figura di contatto e raccordo con l'informatico. L'Advanced Data Analyst è l'anello finale della nuova filiera decisionale; è il punto di contatto con l'operatore economico che intende (o deve, per essere competitivo) fondare le proprie decisioni sull'analisi dei dati.

5. Utilizzo finale: decisioni strategiche.

Gli utilizzatori finali delle analisi quantitative sono i consulenti di impresa e le imprese stesse, ad esempio nella persona del CFO o del responsabile marketing.



NOTE

ⁱ Con il termine "data science" ci si riferisce ad un nuovo ambito disciplinare che accoglie contributi provenienti da diverse discipline - statistica, scienze computazionali, matematica, etc.- Come spesso avviene, nelle fasi primordiali di una serie di concetti e definizioni, l'utilizzo di tale definizione è suggerita più dal background scientifico dei singoli attori che non da una segmentazione rigorosa degli argomenti che la compongono.



M|RESEARCH

M|Research s.r.l. Sede (legale) Pescara: V.le G. d'Annunzio, 69 - 65127 Pescara
Sede Milano: Via Festa del Perdono, 10 - 20122 Milano
Tel +39 02 87167506 Fax +39 02 86882806
www.mresearch.it info@mresearch.it Skype: mresearchsr



M|Research è uno spin-off dell'Università degli Studi G. d'Annunzio di Chieti-Pescara