

# Misurare i dati per capire le nostre vite



**BIG DATA**

**EVA RICCOMAGNO\* ED DANIELE VENTUROLI\*\***

**G**IORNO dopo giorno, gli strumenti più diversi generano una mole di dati, i famosi "big data". Le centraline per misurare l'inquinamento, le carte di credito, le attività svolte da banche ed assicurazioni, e anche i social network sono generatori di dati. Per mettere ordine in questo continuo flusso d'informazioni, occorre saperle gestire e interpretare in modo competente e soprattutto corretto e occorre costruire modelli matematici e statistici per quantificare il rischio, fare previsioni e simulare scenari alternativi.

Il "data analyst", ovvero lo statistico, si occupa di questo. Un esempio è dato dal progetto Google Flu Trends (Gft), un'applicazione ideata per prevedere i picchi influenzali basandosi

sul numero di ricerche online sui sintomi dell'influenza. Il progetto sembrava fornire previsioni più precise di quelle stimate dai Cdc (Centers for Disease Control and Prevention di Atlanta, l'analogo statunitense della rete dei medici-sentinella in Italia), ma poi ha cominciato a sovrastimare il numero di persone che si sarebbero ammalate.

In parte perché le ricerche sui sintomi venivano fatte anche da chi ammalato non era, ma soprattutto perché la qualità del dato non era sicura, a differenza dei dati forniti dai Cdc. Le lezioni derivate dall'insuccesso del progetto Gft (utilissimo invece come strumento complementare per le previsioni dei picchi influenzali) può riassumersi nella frase "big data, old problems" al centro di fervide discussioni in corso in diversi ambienti scientifici.

Un approccio consapevole all'analisi

dei dati è integrante delle attività di didattica e di ricerca del Dipartimento di Matematica di Genova.

Si è rivelato vincente per creare un algoritmo per la prevenzione di frodi nell'online banking, per valutare e combinare le diverse opinioni di esperti per lo sviluppo di un'unità navale. I temi qui accennati sono approfonditi nel corso di laurea triennale in Smid (Statistica Matematica e Trattamento Informatico dei Dati).

I nostri laureati si inseriscono negli ambiti lavorativi in cui analisi dati e previsioni in presenza di incertezza sono necessarie. La presenza di Smid a Genova, unica triennale in Italia, è giustificata dalla convinzione che statistica, matematica e informatica sono tre strumenti necessari al Data Analyst.

*\*professore di Data Mining  
 \*\* studioso e ricercatore*

