

TEXT MINING E MACHINE LEARNING PER I DATI NON STRUTTURATI

di MARIA MORO

LE APPLICAZIONI BASATE SULL'INTELLIGENZA ARTIFICIALE PERMETTONO OGGI DI "LEGGERE" E "COMPNDERE" IN MODO AUTOMATICO LE FONTI TESTUALI E LE IMMAGINI LEGATE A UN SINISTRO. UNA MOLE DI INFORMAZIONI CHE POSSONO ESSERE USATE A SUPPORTO DELL'ANALISI PREVISIONALE

Una delle aree di grande sviluppo dell'informazione è quella correlata all'estrazione di valore dai dati non strutturati. Quando si parla di dati aziendali è evidente che si considerano sia quelli *strutturati*, già depositati in data base e accessibili immediatamente alle analisi, sia quelli *non strutturati*, a cui afferiscono normalmente le informazioni che si possono definire testuali o che derivano da immagini: il peso di queste informazioni è notevole, tanto che nel 2018 si è stimato che l'80% del patrimonio informativo aziendale è costituito da dati non strutturati e che la loro crescita avviene con un tasso del 65% annuo; un contributo notevole alla sfera dei big data, che si prevede raggiungerà entro il 2025 il peso di 163 zettabyte.

Nicola Biscaglia, principal di **Milliman** e practice leader P&C in Italia e Cee, è intervenuto all'Innovation Summit 2021 con una testimonianza sull'importanza dei dati non strutturati e su come sia possibile estrarre il valore implicito dalle fonti testuali per utilizzarle a livello statistico e computazionale: "Imprese come **Microsoft**, **Google**, **Amazon**, hanno già avviato questa strada - ha detto Biscaglia - e lo stesso sta avvenendo nel mondo assicurativo, in cui come Milliman abbiamo sviluppato una soluzione *ad hoc* per l'interpretazione

dei dati testuali relativi ai sinistri". Per chi lavora in ambito attuariale la novità non è di poco conto considerando che, ha osservato Biscaglia, "per la stima del costo atteso dei risarcimenti e le proiezioni future si utilizza ancora lo stesso set di dati degli ultimi 20-30 anni". È cresciuta nel tempo la mole di dati disponibili presso le compagnie, ma la maggioranza delle informazioni resta a livello testuale e non viene strutturata in processi di standardizzazione che consentirebbero la loro storicizzazione, riducendo di conseguenza la possibilità di utilizzarli.

PIÙ INFORMAZIONI PER PREVISIONI PIÙ ACCURATE

A questa esigenza vengono incontro due specifiche branche dell'intelligenza artificiale, in grado di "leggere" e ordinare i dati non strutturati: il *machine learning* e il *text mining*. L'uso combinato dei due sistemi consente di analizzare i dati testuali (quali testi liberi, foto, appunti, questionari, email etc.) per individuare informazioni rilevanti e strutturarle a scopo di analisi e previsione, così da ampliare la disponibilità dei modelli di analisi delle compagnie.

Biscaglia ha presentato un esempio di applicazione sui sinistri Rc auto finalizzata a fornire ai dipartimenti sinistri valutazioni automatiche del costo ultimo prevedibile: "I dati sono ristrutturati in categorie che descrivono il sinistro quali la dinamica dell'evento, il meteo, lo stato dell'ambiente, le condizioni della strada, i dati raccolti dagli smartphone. A queste si possono integrare le informazioni che arrivano dalla polizia o dal soccorso medico, quelle relative ai riscontri peritali, i materiali relativi alle azioni legali e alle eventuali cure necessarie per le persone coinvolte". Tali indicazioni contribuiscono ad arricchire i *data set* delle compagnie e permettono di lavorare su sistemi previsionali più accurati.



Nicola Biscaglia, principal di Milliman