

[as] illuminazioni

Sfide.

Che cosa hanno in comune le scelte politiche di Hillary Clinton, il suggerimento di Facebook di “taggare” un amico su una foto appena postata, la possibilità di osservare il leptone tau decadere in tre muoni e la scelta del ristorante per una cena romantica? Apparentemente nulla. Eppure sono tutti argomenti di sfide lanciate sul sito Kaggle (www.kaggle.com). Kaggle è una piattaforma web per l'analisi di dati e la produzione di modelli che descrivano dei dati.

Vediamo di che cosa si tratta. Pensate ai suggerimenti delle “partenze intelligenti”, che vengono dati in genere da tutti i telegiornali a ridosso delle vacanze. Su che cosa si basano? Su una semplice – e spesso errata – “analisi dei dati”, che parte dall'assunto che molti sono pigri e non riescono a organizzarsi per partire presto e molti altri preferiscono pranzare seduti comodi: quindi il suggerimento è di partire molto presto o durante l'ora di pranzo, per non trovare traffico. La realtà è spesso diversa e ci troviamo incolonnati insieme a migliaia di altre persone, anche se siamo partiti prestissimo. Perché? Perché l'analisi dei dati è una scienza complessa e l'intuizione può non bastare o darci addirittura indicazioni sbagliate. Negli ultimi decenni si è sviluppata la scienza del *data mining* e del *machine learning*, che potrebbe essere tradotto come “imparare dai dati attraverso l'uso dei computer”. Molte sono le tecniche sviluppate e diversi gli

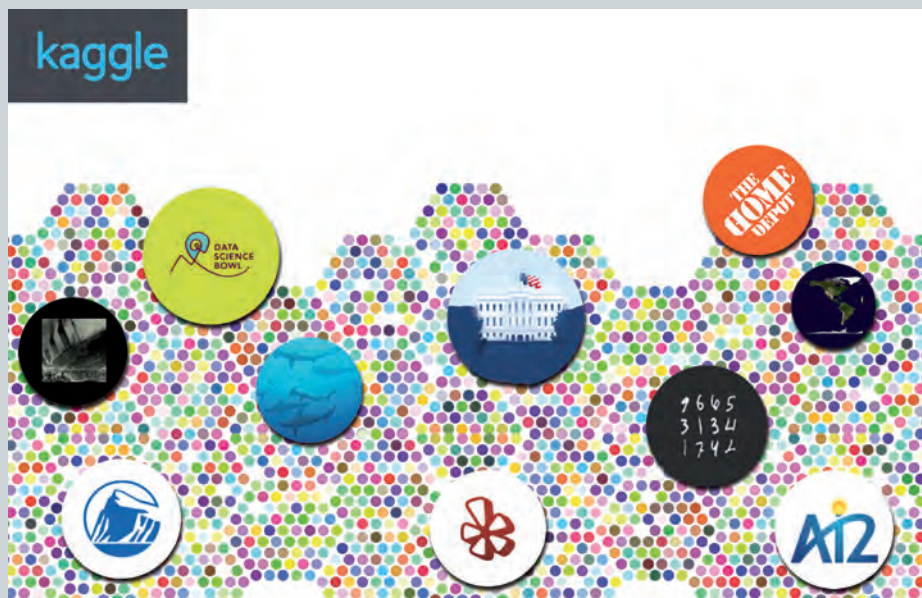
ambiti di applicazione, dallo sviluppo demografico alla fisica delle particelle, dalla medicina al traffico aereo.

L'idea di Kaggle si basa sul fatto che ci sono innumerevoli strategie che possono essere applicate a qualsiasi problema di analisi dati ed è impossibile conoscere a priori quale tecnica sarà più efficace. Le sfide proposte sul sito Kaggle funzionano così: chi è interessato a lanciare la sfida mette a disposizione i dati e una descrizione del problema da risolvere e decide per quanto tempo sarà aperta la sfida. I partecipanti (chiunque può farlo registrandosi sul sito) provano a risolvere il problema utilizzando la propria tecnica preferita e chi, alla fine, avrà trovato la soluzione migliore vince il premio (in denaro), messo a disposizione da chi ha lanciato la sfida. Il vantaggio per il promotore della sfida è il poter usare liberamente la soluzione vincitrice.

Il Cern ha già lanciato due sfide su Kaggle, una proposta dall'esperimento Atlas per migliorare l'analisi dei dati del bosone di Higgs (www.kaggle.com/c/higgs-boson, finita nel settembre 2014) e una proposta da Lhcb per la ricerca di un rarissimo decadimento del leptone tau (www.kaggle.com/c/flavours-of-physics, chiusa nell'ottobre 2015).

In una sezione apposita, <https://inclass.kaggle.com>, Kaggle ospita anche i progetti di centinaia di università di tutto il mondo, progetti in cui gli studenti possono cimentarsi liberamente nell'applicare le tecniche del *machine learning* a problemi reali.

Il Cern ha in programma di lanciare altre sfide, ma nel frattempo, la prossima volta che userete le indicazioni di un sito web per decidere in quale ristorante andare, sappiate che potrebbe essere che stiate utilizzando proprio il risultato di una sfida lanciata su Kaggle! [Barbara Sciascia]



Vedi anche la sfida sul sito:

- di Atlas: <https://higgsml.lal.in2p3.fr/>

- di Lhcb: <http://lhcb-public.web.cern.ch/lhcb-public/Welcome.html#Kaggle>