

3

Analiziranje naših podataka

Godine 2012, *New York Times* je objavio priču o tome kako korporacije analiziraju naše podatke za potrebe oglašavanja. U članku je otkriveno da korporacija Target može da utvrdi način na koji žene kupuju za vreme svoje trudnoće i da koristi te informacije da bi trudnicama slale oglase i kupone za proizvode u vezi beba. U članku je bila i anegdota o čoveku iz Minneapolisa koji se požalio korporaciji zato što je slala kupone u vezi beba njegovoj kćerki tinejdžerki, da bi kasnije čovek otkrio da je korporacija bila u pravu.

Opšta praksa prikupljanja i čuvanja svih vrsta podataka naziva se „veliki podaci“ (engl. *big data*), a naučni postupak i inženjering izdvajanja korisnih informacija iz nje nazivaju se „rudarenje podataka“ (engl. *data mining*). Kompanije rudare po svojim podacima da bi forukisrale svoje reklamne kampanje. Barak Obama se iz istog razloga bavio rudarenjem podataka za potrebe svojih predsedničkih kampanja za 2008. i 2012. godinu. Auto kompanije rudare podatke iz vašeg automobila kako bi dizajnirale bolje automobile; opštine prikupljaju i rudare po podacima od senzora na putevima kako bi razumeli uslove vožnje. Naši genetski podaci su izrudareni

za razne vrste medicinskih istraživanja. Kompanije kao što su Facebook i Twitter rudare po našim podacima u reklamne svrhe i omogućavaju akademskim ustanovama da ih koriste za istraživanje društva.

Većina rudarenja je zapravo sekundarna upotreba podataka. To jest, to nije razlog zbog kog su podaci prikupljeni. U stvari, to je osnovna vrednost velikih podataka: sačuvajte sve što možete i jednog dana ćete imate neku korist od svega toga.

Veliki skupovi podataka imaju vrednost, delom iz zaključaka koji mogu da se izvedu od njih. Neki od njih su očigledni. Ako imate podatke o nečijem detaljnom kretanju tokom godine, možete da zaključite koji su njegovi omiljeni restorani. Ako imate spisak ljudi koje poziva i kojima šalje e-poštu, možete da zaključite ko su njegovi prijatelji. Ako imate spisak internet stranica koje posećuje, ili možda spisak knjiga koje je kupio, možete da zaključite kakvi su njegovi interesi.

Neki zaključci su suptilniji. Spisak kupljenih namirnica neke osobe može ukazati na njenu etničku pripadnost. Ili starost i pol, i verovatno religijsko opredeljenje. Ili medicinsku istoriju i sklonost ka alkoholu. Ljudi u marketingu su stalno u potrezi za obrascima koji ukazuju da će neko napraviti veći trošak, kao što je venčanje, odlazak na odmor, kupovina kuće, rođenje deteta i tako dalje. Policija u raznim zemljama koristi ove obrasce kao dokaze, bilo na sudu ili u tajnosti. Facebook može da proceni nečiju rasu, ličnost, seksualnu orijentaciju, političku ideologiju, stanje veze sa partnerom i upotrebu droga samo na osnovu pritska na Like dugme. Kompanija zna da ćete da se verite pre nego što ste to objavili, ili da je neko homeskusualac i pre nego što okolina sazna, a njegove poruke mogu to da otkriju drugim ljudima bez njegovog znanja ili dozvole. U zavisnosti od zemlje u kojoj živi, to bi mogla da bude sramota ili gubitak glave.

Postoji mnogo grešaka u zaključivanjima, a to svi mi koji smo videli internet oglase koji su nam slabo zanimljivi, možemo da potvrdimo. Ali kada su oglasi pogodili šta nas

intresuje, mogu da budu jezivi i često nam se ne sviđaju. Jedna je stvar kada vidite na televiziji oglase za čepiće za hemoroida ili oglas za agenciju za bračno posredovanje, gde znamo da svi vide te oglase. Ali kada znamo da su oglasi posebno ciljani na nas, na osnovu onoga što smo objavili ili označavali sa Like na internetu, tada oglašavanje može da bude mnogo invazivnije. Ovo nam stvara protivrečnu situaciju: podaci koje smo spremni da delimo mogu da nas dovedu do zaključka da ne želimo da ih delimo. Mnoge žene sa zadovoljstvom kažu šta ih interesuje da kupe na popustu i žele da budu obavешtavane o novim proizvodima koje bi možda želele da kupe, ali većina njih ne želi da kompanija shvati da su trudne. Takođe ne želimo velike krađe podataka i prevare koje neizbežno prate ove velike baze podataka.

Kada pomislimo na računare koji koriste sve naše podatke da bi napravili zaključke, mi imamo veoma humani način razmišljanja o tome. Zamišljamo kako bismo podacima dali smisao i preneli takav proces na kompjutere. Ali to nije tako. Računari i ljudi imaju različite snage, slabosti i ograničenja. Računari ne mogu apstraktno da razmišljaju kao ljudi, ali mogu da obrade ogromne količine podataka sve brže i brže. (Ako razmišljate o tome, to znači da su računari bolji u radu sa metapodacima nego što su u obradi snimljenih razgovora) Računari se stalno poboljšavaju. Računarska snaga se još uvek udvostručuje svakih osamnaest meseci, dok je veličina mozga naše vrste i dalje ista. Računari su već daleko bolji od ljudi pri obradi kvantitativnih podataka i nastaviće da se poboljšavaju.

Upravo sada, rudarenje podataka je akutelna tehnologija, a oko nje je mnogo oduševljenja i oportunitizma. Još nije sasvim jasno koje će vrste istraživanja da budu moguće, ili koji je pravi potencijal svega toga. Ono što je jasno je da tehnologija rudarenja podataka postaje sve moćnija i omogućava posmatračima da izvuku sve više iznenađujućih zaključaka iz velikih skupova podataka.