

Podatkovni analitik v Obamovi kampanji Svet vse bolj številkast, številčna pismenost pa še vedno zelo redka

Lani je ameriškega podatkovnega analitika Rayida Ghanija čakala pomembna karierna odločitev. Uspešno so zaključili ameriško predsedniško tekmo, v kateri je njegova skupina skrbela za informatizacijo volilne kampanje Baracka Obame. Spletne storitve in trgovino je poznal že od prej, zato je hotel poskusiti nekaj novega.

Lenart J. Kučič

Soustanovil je analitsko podjetje Edgeflip, ki nevladnim organizacijam pomaga uporabljati spletna družabna omrežja, in sprejel ponudbo čikaške univerze, kjer združuje družboslovno raziskovanje z uporabo podatkovnih orodij.

Uporaba velikih podatkovnih zbirk je lahko zelo koristno orodje za izboljšanje javnih storitev – od državne uprave do zdravstva in izobraževanja. Vendar je treba znati podatke pravilno zbrati, obdelati in uporabiti, sicer nam prikažejo napačno sliko sveta, je povedal Ghani. Če ima neka bolnišnica visok odstotek smrtnosti med pacienti, to lahko opozori na zdravniške napake ali zgolj pove, da se v tej bolnišnici zdravi več smrtno bolnih bolnikov. Fakulteta, na kateri diplomira največ študentov, pa ni nujno najbolj kakovostna, ampak ima morda nižja merila ali vpisuje samo najboljše kandidate.

Vendar je funkcionalna številčna pismenost med odločevalci in državljani še vedno zelo redka, meni sogovornik. Čeprav postaja naš svet vse bolj številkast.

Če spremljamo razkritja skupnosti wikileaks, nastope nekdanjega obveščevalca Edwarda Snowdena in analize ponudnikov spletnih storitev, se zdi, da velike zbirke podatkov uporabljamo predvsem za nadzor državljanov in izkoriščanje potrošnikov. So velike korporacije in države res največji zaposlovalci podatkovnih specialistov?

Največji zaposlovalec podatkovnih specialistov v ZDA so velika podjetja, ki imajo za uporabo podatkovnih orodij zelo jasen interes: izboljšati učinkovitost poslovanja, povečati dobičke in prehiteti tekmece. Ko je bil finančni balon na vrhuncu, je bil Wall Street največji porabnik doktorjev matematike in fizike, saj je vsaka najmanjša izboljšava trgovalnih algoritmov prinašala velikansko tržno prednost. Ti profili so zelo dragoceni tudi v spletnem poslovanju, kjer je vsak uporabniški klik zelo natančno merljiv in so poslovni rezultati neposredno odvisni od pravilnega razumevanja teh klikov. Velike korporacije so začele prve izrabljati potenciale podatkovnih orodij in so najele najboljše ljudi. Zato imate prav – veli-

ka večina podatkovnih specialistov danes v ZDA dela za velika spletna podjetja, trgovce, finančni sektor in zavarovalnice.

Ne za državo?

Ne vem, ali poznam kakega vrhunškega podatkovnega strokovnjaka, ki je zaposlen v državni službi (*nasmešek*).

Zaradi prenizkih plač?

Država res ne more ponuditi tako dobre plače kot zasebno podjetje, a denar ni najpomembnejši razlog. Podatkovne analitike zanima izboljševanje učinkovitosti, reševanja problemov pa se lotevajo podobno kot znanstveni raziskovalci – z meritvami in preizkušanjem delovnih hipotez. Javne ustanove jim težko ponudijo ustrezno delovno okolje. Na javnem razpisu je skoraj nemogoče kupiti najboljša orodja, ki jih potrebujejo za učinkovito zbiranje in obdelavo podatkov. Vodja skupine zelo težko najame ali odslovi sodelavce. Odločanje je ponavadi počasno in birokratsko, podatkovne analize pa hitro opozorijo na težave, ki jih odločevalci ne marajo slišati ali so politično nesprejemljive: kdo vse ne opravlja dela, katere javne politike so neučinkovite ... Vendar danes tudi države občutijo podobne pritiske kot zasebni sektor. Denarja je premalo, zato potrebujejo orodja, s katerimi bodo lahko učinkoviteje nadzorovali in usmerjali porabo.

Ali države res tako zelo zaostajajo za zasebnim sektorjem? Ali niso prav države in njihove ustanove največje upravljavke velikih podatkovnih zbirk: registra prebivalstva, zdravstvenih in policijskih kartonov, neštetihih registrov in popisov?

Države imajo ogromno podatkov, vendar večinoma še ne vedo, kaj z njimi početi, razen da jih pač zbirajo. V zasebnem sektorju so merila učinkovitosti preprosta: če z analizo procesov zmanjšamo stroške in povečamo prihodke, je uporaba analitike uspešna. Za javne ustanove pa je zelo težko določiti merila, po katerih jih bomo smiselno ocenjevali. Vzemimo lestvice, po katerih vrednotimo fakultete. Tam večinoma merimo tisto, kar je najlažje meriti: koliko študentov se vpiše, koliko jih konča študij in koliko se jih zaposli v enem letu po zaključku študija. Vprašanje pa je, kaj nam v resnici pove podatek, da na eni fakulteti študij dokonča šestdeset odstotkov študentov, na drugi pa osemdeset.

Da je druga fakulteta boljša?

Takšen je naš prvi intuitivni sklep (*nasmešek*). A če malo pomislimo, na končno število diplomantov vpliva zelo veliko dejavnikov. Katere predmete poučujejo na eni in drugi fakulteti? Kakšna so njihova ocenjevalna in vpisna merila? Kakšna je njihova struktura prihodkov – se financirajo predvsem iz šolnin ali iz raziskav? Kakšni so rezultati primerljivih fakultet? Na kateri fakulteti so študenti od začetka do konca študija najbolj napredovali? Če tega ne vemo, nam tudi podatek o diplomantih ne pove prav dosti. Na drugi fakulteti morda dosegajo navidezno boljši rezultat zato, ker



Rayid Ghani Foto Jure Eržen

že pri vpisu izberejo samo najboljše študente, ki bodo bolj verjetno diplomirali. Ali pa, nasprotno, izboljšujejo številke tako, da je pri njih lažje diplomirati.

Ali ni takšno prirejanje številke logična posledica vsakega ocenjevanja in razvrščanja na lestvicah?

Zato je treba znati brati podatke. Pred časom je v ZDA zaživela pobuda, da so bili javno objavljeni podatki o umrljivosti pacientov v bolnišnicah. Razlog za pobudo je bilo prepričanje, da bi takšno objavljanje pripomoglo k večji preglednosti in bi bolnikom omogočalo informirano izbiro boljše zdravstvene oskrbe. A se je tudi tam pokazalo, da podatek o umrljivosti pacientov ni najboljši pokazatelj kakovosti bolnišnice. V nekaterih ustanovah je bila umrljivost večja, ker so zdravile hujše oblike bolezni, sprejemale zahtevnejše paciente ali delovale v revnejših okoljih. Predstavljajte si, da bi začele bolnišnice zmanjševati umrljivost tako, da bi nehale sprejemati bolnike, ki jim kvarijo številke. To se je tudi dejansko dogajalo in lahko si predstavljate, kako negativne so posledice takih odločitev za bolnike in za medicinsko stroko.

Države in podjetja so pod vse večjimi pritiski, da politične in poslovne odločitve utemeljujejo s številkami. Vi pa opozarjate na številčno nepismenost. Ali ni to nevarna kombinacija?

Podatkovna analiza je samo orodje, odločevalci pa se morajo odločiti, kaj bodo



Veliko najbolj nadarjenih programerjev in matematikov poskuša priti na Google ali kako podobno veliko spletno podjetje. [...] A številni kmalu ugotovijo, da je njihovo delo v resnici precej dolgočasno. Eno leto so se po dvanajst ur na dan ukvarjali z optimizacijo oglasnega algoritma, ki bo na koncu leta Googlu prinesla kak dodaten milijon dolarjev ali za desetinko sekunde skrajšala iskanje.

počeli z rezultati. Ko so me povabili k projektu za izboljšanje študijskih rezultatov na neki manjši univerzi, sem naročnika najprej vprašal, kaj želi doseči. Naj mu pomagam izbrati študente, ki bodo najverjetneje uspešno diplomirali, ali ga zanima, zakaj se na njihovo univerzo vpiše tako malo mladih iz revnejših sosesk? Je vloga njegove fakultete izboljšati dostop do izobrazbe za vse mlade v njihovem okolju ali poiskati samo najbolj nadarjene študente in jim pomagati do dobrih služb? To so zelo različni cilji. S podatkovno analizo jih morda lahko poma-

Marisse Mayer, ki ga je novinar sklenil s stavkom: odkar je Mayerjeva prevzela vodenje Yahooja, je vrednost delnic zrasla za 50 odstotkov. Kaj naj sklepam iz takega zaključka? Predstavljam si, da se je Mayerjeva novinarju pohvalila s tem podatkom in ga je uporabil, da bi s konkretno številko pokazal, kako uspešno je njeno vodenje. Nikjer pa ni bilo podatka, da je v tem času povprečna vrednost tehnoloških podjetij zrasla za 80 odstotkov. Se lahko Mayerjeva še vedno pohvali s 50-odstotno rastjo? Enako površno je pokrivanje znanstvenih novic. Saj vem,

nam pri vsakem obroku merijo kalorije in štejejo korake, koliko smo jih porabili. Na koncu dneva podatke delijo z drugimi člani shujševalnega družabnega omrežja in nas pohvalijo, če smo bili pridni. Je tak način učinkovitejši od seznanjanja s podatki?

Na to vprašanje lahko odgovorim iz izkušenj, ker tudi sam uporabljam takšno aplikacijo. Veljam za podatkovnega fanatika in rad izmerim vse svoje dejavnosti, a hkrati rad dobro jem, kar se mi že malce pozna – čeprav se zelo dobro zavedam zdravstvenih tveganj. Zato sem začel uporabljati program, ki meri mojo telesno aktivnost in mi sproti pove, koliko dodatnih korakov moram prehoditi, če si zdajle privoščim še tole sladico.

Upošteвате ta nasvet?

Ne vedno, vsekakor pa prehodim veliko več kot prej. Raziskovalci ugotavljajo, da so mobilne aplikacije razmeroma uspešne pri hujšanju, saj vas takoj nagradijo za pravilno dejanje. Premišljena uporaba malih nagrad za doseganje malih osebnih ciljev je preverjeno dober motivator pri človeški vrsti. Zato predvidevam, da bomo z aplikacijami, igrami, oblikovanjem spletnih storitev in urbanističnimi odločitvami vse bolj izrabljali mehanizem malih nagrad za spodbujanje zdravega življenja.

Aplikacija ve, kaj je dobro zate, zato se je treba podrediti njenim navodilom. Ali ni v tej misli nekaj potencialno avtoritarnega?

Ali ni v vsakem močnem orodju vedno tudi nekaj potencialno avtoritarnega? Zelo dvomim, da se bodo posamezniki množično odpovedovali svobodni volji in odločanje prepustili aplikacijam. Pa tudi človeško obnašanje je veliko preveč zapleteno, da bi ga lahko učinkovito upravljali. Na to nas je opomnila lanska predsedniška kampanja Baracka Obame, pri kateri sem sodeloval. To je bila verjetno najbolj informatizirana kampanja doslej, pri kateri smo uporabili vse razpoložljive podatke o volivcih: volilno zgodovino, družabna omrežja in komercialne podatkovne zbirke, kjer so zbrane njihove potrošniške in druge navade. A smo morali velika pričakovanja razmeroma kmalu omejiti na zelo konkretne cilje.

Kakšne?

Prvi cilj je bil pridobiti dovolj denarja za kampanjo. Pri tem smo bili verjetno še najbolj uspešni, saj smo z analizo podatkov o volivcih zelo natančno določili naše potencialne podpornike in jih dosegli po vseh možnih komunikacijskih kanalih. Zabeležili smo vsako elektronsko sporočilo in objavo na družabnem omrežju ter merili njene učinke: odzive prejemnikov, njihove donacije in mreženje. Drugi cilj je bil mobilizacija volivcev v tistih zveznih državah, za katere smo vedeli, da lahko volilni rezultat odloči že nekaj tisoč glasov. Tam smo poskušali z natančnim filtriranjem volilnih registrov najti potencialne demokratske volivce tudi v tradicionalno republikanskih okrožjih in jih prepričati, naj se prijavijo v volilni imenik. Preizkusili smo dobresedno na stotine načinov, kako oblikovati sporočilo, da bo posameznik odšel na volišče. Veliko smo se ukvarjali z družabnimi omrežji, saj nas je zanimala moč osebnih priporočil: kako se ljudje odzivajo, če jih prijatelj na facebooku prosi, naj prepričajo tri svoje prijatelje, naj se udeležijo volitev in glasujejo za Obama.

Kakšni so bili odzivi?

Sprva smo se bali, da bodo takšni nagovori preveč vsiljivi, in prvi odzivi niso bili najboljši. A smo se med kampanjo naučili bolje uporabljati družabna omrežja in oblikovati sporočila na načine, ki niso bili vsiljivi. Najbolj učinkovita so bila zelo per-

sonalizirana sporočila, kjer je prejemnik dobil občutek, da smo nagovarjali prav njega in ni le eden izmed tisočih prejemnikov enakega elektronskega pisma ali poziva na facebooku.

Obamov štab je po neuradnih ocenah porabil daleč največ denarja za podatkovne analitike. Ste upravičili vložene milijone?

Ne upam trditi, da smo odločili predsedniško tekmo. Zagotovo pa smo pridobili veliko več donatorskega denarja, kot smo ga porabili. Mobilizirali smo številne volivce, ki bi jih tradicionalne kampanje spregledale, in političnim strategom pomagali učinkoviteje porabiti dolarje, da jih niso trošili za napačne volivce ali neučinkovite medijske kampanje.

Kaj čaka vaše naslednike v prihodnjih predsedniških tekmah? Je mogoče še bolj optimizirati lov na nekaj odstotkov neopredeljenih volivcev?

Dvomim, da se bo v prihodnjih nekaj kampanjah drastično spremenilo delovanje politične mašinerije. Postale bodo še bolj informatizirane in bodo uporabljale vse najnovejše prijeme merjenega oglaševanja in prilagojenega nagovarjanja posameznega volivca. Vendar ne smemo pozabiti, da največji del volilnega telesa niso več republikanci, demokrati ali privrženci katere druge politične stranke, ampak tisti, ki se sploh ne udeležujejo volitev. Za morebitna nova politična gibanja bodo verjetno najbolj zanimivi prav nevolivci.

Si bodo lahko tudi alternativna civilnodružbena in politična gibanja kdaj privoščila vaše storitve? Na začetku pogovora ste povedali, da ste podatkovni specialisti predragi celo za državo, kaj šele za nevladnike.

To sicer drži, vendar potrebujemo podatkovni zanesenjaki predvsem izzive in občutek, da ima naše delo konkretne učinke. Veliko najbolj nadarjenih programerjev in matematikov poskuša priti na Google ali kako podobno veliko spletno podjetje. Pritegnejo jih velikanske količine podatkov, mogočna računalniška infrastruktura, veliki razvojni proračuni in skoraj zvezdniški status, ki ga tam priznavajo vrhunskim analitikom. A številni kmalu ugotovijo, da je njihovo delo v resnici precej dolgočasno. Eno leto so se po dvanajst ur na dan ukvarjali z optimizacijo oglasnega algoritma, ki bo na koncu leta Googlu prinesla kak dodaten milijon dolarjev ali za desetinko sekunde skrajšala iskanje. To pa je za pravega informacijskega zanesenjaka premalom zanimivo, zato bo pripravljen poskusiti kaj novega.

Družbeno koristnega?

Vsekakor. Ko smo lani na čikaški univerzi razpisali možnost študija in raziskovanja, kako uporabiti velike podatkovne zbirke za pozitivne družbene spremembe, smo že prvo leto prejeli več kot 600 prijav – samih dobrih študentov, ki so službe zlahka našli v komercialnih podjetjih. Ti študenti zdaj presenečeno ugotavljajo, koliko zanimivih podatkov je v zdravstvu, energetiki in izobraževanju, hkrati pa se nekateri prvič zares srečujejo s problemi ljudi, ki niso del njihove informacijske kaste. Ko smo v nekem projektu ugotovili, da se številni mladi ne odločajo za študij, ker se ne znajo prijaviti na fakulteto ali ne vedo, kako zaprositi za štipendijo, so bili moji mladi sodelavci čisto šokirani. Kako je mogoče, da kdo tega ne ve, so se spraševali. Zato komaj čakam, kakšne družbene inovacije bodo začeli razvijati študenti informatike, ko se bodo namesto z Googlom in Facebookom začeli ukvarjati s problemi okolja, v katerem živijo. ■



gam uresničiti, ne morem pa jih izračunati namesto tistih, ki morajo razmišljati o izobraževalni politiki. Zato menim, da moramo v informacijski dobi nujno povečati funkcionalno številčno pismenost. Ne le med odločevalci, ampak tudi med državljanji.

Kaj pomeni funkcionalna številčna pismenost?

Da razumemo, kaj nam govorijo številke. To ne pomeni, da moramo vsi postati matematiki ali analitiki. Če bi otrokom v osnovni šoli predstavili osnove uporabne matematike in jih naučili, da morajo ob podatku, da ima neka sladkarija za polovico manj sladkorja, takoj pomisliti na najpomembnejša metodološka vprašanja: Manj v primerjavi s čim? Je manj sladkorja zame dobro ali slabo? Je morda 'za polovico sladkorja manj' še vedno preveč za zdrav obrok? Enako kritično moramo brati tudi vse druge podatke, ki jih vidimo v medijih, naj opisujejo politiko, gospodarstvo ali vplive telovadbe na zdravje. Prav v medijih najbolj pogrešam kritično branje števil. Novinarji ste na tem področju ... Boste jezni, če bom iskren?

Obupni?

No, pa ste povedali namesto mene (*smeh*). Od novinarjev res ne pričakujem, da boste zgolj poročali o statistikah in podatkih, ki vam jih posredujejo ministrstva, statistični uradi in podjetja, ampak mi pomagajte dati podatke v ustrezen kontekst. Nedavno me je ujezil portret predsednice uprave internetnega podjetja Yahoo

da novinarji ne morete kritično ovrednotiti rezultatov vsake klinične raziskave, ker tako delo zahteva specialistično znanje. Lahko pa jo pridobite in povabite strokovno javnost, da si sama ustvari mnenje ali vam pomaga razumeti njene ugotovitve.

Lahko boljše obveščeno vpliva na človeško obnašanje? Praksa večinoma ne pritrjuje optimističnim napovedim podatkovnih razsvetljencev in zagovornikom transparence. Naše navade večinoma ostajajo enake, tudi če se seznanimo s podatki, ki nasprotujejo našim prepričanjem.

Kako spremeniti človeško obnašanje? To je verjetno eno najpogostejših vprašanj, s katerim se srečujem. Verjetno veste, da imamo v ZDA velike probleme zaradi prekomerne telesne teže. Ta slabo vpliva na kakovost življenja in obremenjuje zdravstveno blagajno, a jo je izredno težko omejiti, ker gre za kompleksen preplet življenjskega sloga, kulture, socialnih dejavnikov, genov in evolucije. Podatki nam lahko povedo marsikaj: koliko se ljudje gibljejo, kaj pojedjo, kakšno imajo presnovo. O njih lahko pridobimo vse demografske podatke, nakupovalne navade in objave na družabnih omrežjih, s katerimi problem zelo natančno opišemo. A kljub temu še ne poznamo učinkovitega načina, kako jih uporabiti za učinkovito preprečevanje debelosti.

V reviji Atlantic so pred nekaj meseci poročali o razmahu posebnih shujševalnih aplikacij, ki jih namestimo na mobilnik, da