UNIVERZA V LJUBLJANI

FILOZOFSKA FAKULTETA

ODDELEK ZA PSIHOLOGIJO

PSIHOLOŠKA DIAGNOSTIKA IN UKREPI V DELOVNEM OKOLJU

'BIG DATA' – MASIVNI PODATKI V KADROVSKEM KONTEKSTU

SEMINARSKA NALOGA

AVTORICA: Maja Vovko, dipl. psih. (UN)

MENTORICA: doc. dr. Eva Boštjančič

Ljubljana, december 2016

UVOD

Psiholog lahko koristi napredke v tehnologiji na mnoge načine, tako v pomoč zaposlenim kot vodstvu oz. organizaciji sami. Razširitev družabnih omrežij, predvsem tistih namenjenih strokovnem povezovanju (npr. *LinkedIn*), psihologu lahko koristijo za rekrutacijo in selekcijo primernih kandidatov ter preverjanje določenih informacij. Glede na tehnološki porast in napredek v družbi lahko psiholog predvideva potrebe, ki jih bo imela organizacija, določena delovna mesta znotraj nje ali celo zaposleni sami. Z razvojem programov simulacije lahko izobražujemo zaposlene kako ravnati v situacijah, ki so preredke ali preveč nevarne, da bi jih njim zares izpostavljali. Precej nova, predvsem na kadrovskem področju, pa je uporaba masivnih podatkov, ki je bila do sedaj v uporabi predvsem v oddelkih za razvoj in marketing. Vendar so lahko te informacije zelo priročne tudi za psihologa, o čemer govori moja seminarska naloga.

KAJ JE BIG DATA?

V dandanašnjem svetu skorajda z vsakim dejanjem, pa če je to v fizičnem svetu ali pa na spletu, puščamo digitalne sledi, ki raziskovalcem lahko veliko povedo o vedenju in lastnostih posameznika, ki jih je pustil. Lahko bi rekli, da sta danes komunikacija in vedenje posameznikov v veliki meri digitalizirana (Chuapetcharasopon, 2015). Ideja masivnih podatkov (ang. *big data)* je torej zbiranje teh digitalnih sledi, ki jih je nato moč analizirati (Marr, 2015). Neposredno s tem je povezana tudi rast podatkov – večajo oziroma širijo se viri podatkov, količine podatkov, celo vrsta podatkov. Analogno s tem razvojem in rastjo pa se razvijajo tudi tehnologije in orodja, s katerimi lahko te vsakodnevne sledi zbiramo, organiziramo in urejamo. Ključna sprememba je torej tudi v tem, da sedaj obstajajo načini zelo sistematične in točne analize teh nestrukturiranih podatkov (ki jih je v primerjavi s strukturiranimi težje analizirati), npr. elektronskih sporočil, slik, videoposnetkov. V teh sledeh lahko iščemo vzorce, kar pred nedavnim tehnološkim razvojem ni bilo mogoče (Marr, 2015).

Masivne podatke je zelo težko povsem natančno opredeliti, saj vsaka organizacija lahko zbira in uporablja zelo raznolike podatke, iz raznolikih virov in raznolikih velikosti, pa jih (zase) smatra kot 'masivne'. V splošnem sicer obstajajo kriteriji za velike podatke. Slednje kriterije velikih podatkovnih baz opredeljuje Chuapetcharasopon (2015).

Prvi kriterij je **volumen** oz. velikost, obširnost podatkov. Masivni podatki naj bi bili veliki več terabitov. V HR kontekstu to pomeni, da kadrovska služba hrani podatke večjega števila zaposlenih; primer organizacije, ki zagotovo operira z neverjetno količino podatkov, je ameriška veriga *Wal-Mart* s kar 2,2 milijona zaposlenih. Tudi če njihova baza drži le najbolj osnovne informacije o zaposlenih, je izjemno velika.

Drug kriterij je **hitrost** pridobivanja podatkov; masivni podatki so tudi 'hitri'. Včasih se je podatke zaposlenih zbiralo enkrat ali parkrat letno (npr. ob vstopu na delovno mesto, ob vsaki spremembi osebnih podatkov ali zamenjavi delovnega mesta, ob oceni delovne uspešnosti itd.). Po drugi strani pa se dandanes za masivne smatrajo podatki, ki v hrambo pritekajo dnevno ali celo sekundno. Ne gre torej le za hitrost ampak tudi *pretok*. Na nivoju poslovanja je to lahko vsak račun, posnetki nadzornih kamer, štetje strank, ki v trgovino vstopijo, itd. Na nivoju kadrovske službe trenutno ni take hitrosti pobiranja podatkov o zaposlenih samih – pogled v prihodnost pa je recimo priročna napravica, ki jo zaposleni tekom dneva nosijo, ona pa beleži nivo stresa, vedenje, druge fiziološke znake. To bi bil primer pobiranja sekundnih podatkov o zaposlenem.

Raznolikost kot naslednji kriterij naj bi s kazala v raznolikosti oblike (formata) in virov podatkov. Podatki torej niso zgolj demografske narave, kot so starost, spol, izobrazba; vključeni so tudi podatki o kompenzaciji, o uspešnosti na delu, podatki raznolikih ocenjevanj dela, vedenjski podatki (npr. število odsotnosti, zamujanja itd.) in drugih notranjih informacij (vprašalniki, podatki o kompetencah, izkušnjah, elektronska sporočila) (McCaffery idr, 2015). Podatki so lahko pridobljeni s strani zaposlenega, nadrejenega, tudi arhivov; v grobem podatke v velikih bazah delimo namreč na 'zunanje' in 'notranje'. Poleg tega nove oblike podatkov predstavljajo tudi CV-ji, objave na različnih družbenih omrežjih (npr. *Twitter, Facebook)* ali podatki in kontakti s podobnih strani (npr. *LinkedIn),* tudi razni komentarji na forumih, blogih, itd. Podatke lahko zbiramo celo preko pametnih telefonov ali senzorjev. Te nove oblike zahtevajo nove oblike analize, saj so podatki nestrukturirani.

Zadnji, a prav tako pomemben, pa je kriterij **verodostojnosti**. Ta pride v ospredje predvsem pri pregledu in oceni zbranih podatkov, ne pa samem zbiranju. Preko zbiranja vseh teh raznolikih in predvsem številnih podatkov (preko več virov in več oblik), se nabere tudi veliko 'šuma' – podatkov, ki so morda nepopolni, največkrat pa neuporabni ali neustrezni za želeno analizo ali problem. Zaradi tega šuma je potrebno pogosto in podrobno vodenje in urejanje podatkovne baze. Kot že rečeno so je v osnovi masivni podatki kljub tem kriterijem zelo izmuzljiv pojem in številne 'prave' masivne podatkovne baze vseeno ne dosegajo vseh kriterijev.

Kje se take informacije torej uporablja? Zbirajo jih namreč zelo raznolika podjetja, v raznolike namene. Najbolj pogosta oz. razširjena je raba na strankah ali kupcih – preko analize masivnih podatkov želijo podjetja torej optimizirati ali povečati prodajo preko spoznavanja kupcev in njihovih nakupovalnih vedenj. S tem, ko sledijo trendom nakupov in nakupovalnih vedenj, lahko prodajna podjetja napovedujejo prodajo nekega svojega izdelka; napovedujejo ali bo prišlo do menjave ponudnika (npr. v mobilnem ali telekomunikacijskem podjetju lahko sklepajo, ali bo nek ponudnik želel menjati naročniški paket). Za izboljšanje poslovnih procesov lahko na podlagi trendov na družbenih

omrežjih, pogostih iskalnih gesel v različnih brskalnikih (torej kaj ljudje '*googlajo*') in celo na podlagi vremenske napovedi predvidevajo, kaj se bo prodajalo in torej s čim je vredno zapolniti skladišče. O široki uporabi priča tudi dejstvo, da z masivnimi podatki razpolaga in operira tudi policija v večjih mestih in sicer v namen lovljenja kriminalcev ali napovedovanja kriminalne aktivnosti. Podobno jih lahko banke uporabijo za odkrivanje nelegalnih transakcij bančnih kartic (Murr, 2015). Masivni podatki pa so vse bolj uporabni tudi v kadrovski službi organizacije, npr. za pridobivanje, trening zaposlenih, ohranjanje delovne sile, ugotavljanje učinkovitosti, razvoj, načrtovanje nasledstva, načrtovanje posameznikove karierne poti ipd. (Roberts, 2013). Rabo znotraj kadrovskega konteksta bom nadalje prikazala na praktičnih primerih in dodatno pojasnila uporabnost ter omejitve.

KAKO NAJ MASIVNE PODATKE UPORABI PSIHOLOG?

Masivni podatki so lahko uporabni za psihologa v kadrovski službi in sicer za več namenov oz. problemov. Na začetku poglavja bom predstavila nekatere resnične uporabe teh podatkov v večjih podjetjih (z velikim številom zaposlenih). Naj omenim, da imajo ta podjetja navadno oddelke namenjene zbiranju ali združevanju takih baz, tako imenovane '*HR analytics departments*' oz. kadrovske analitične oddelke. Če ima podjetje tak oddelek, v njem pa zaposlenega psihologa s potrebnim znanjem, torej je strokovnjak za HR analizo, navadno sam zbira te podatke. Če podatke, potrebne za obdelavo, zbira več različnih oddelkov znotraj podjetja (npr. oddelek za marketing, oddelek za razvoj in inovacijo), je navadno delo analitičnega HR oddelka, da od drugih oddelkov vse potrebne podatke pridobi in združi. Podjetje lahko strokovnjaka za HR analitiko tudi najame, navadno kadar je podjetje premajhno ali premlado, da bi zaposlilo svojega.

Primeri uporabe masivnih podatkov v kadrovskih službah organizacij

Pogosta je raba namene podatkov v namene upravljanja zaposlenih, tudi upravljanja stroškov. Prvi tak primer v kadrovski službi je *Defense Acquisition University*; gre za ameriško izobraževalno ustanovo, ki se ukvarja z vsakoletnim treningom tisočih vojakov in civilnih strokovnjakov v obrambni tehnologiji in logistiki. Z uporabo in analizo raznolikih podatkov, s katerimi so že razpolagali (stroški nastanitve; plača za inštruktorje; stroški potovanja, ki se izplačajo inštruktorjem; pričakovano prisotnostjo študentov; stroški potovanja za študente ipd.), so našli najbolj optimalno lokacijo za izvedbo treningov glede na višino stroškov. Tekom izvajanja treningov so ugotovili, da inštruktor sprva potrebuje povprečno 4 ure, da se pripravi na eno uro pouka, nato pa so s spremljanjem in analizo podatkov odkrili, da se sčasoma (po več opravljenih pripravah in inštrukcijah) čas priprave zniža na pol ure. Te ugotovitve so nato uporabili pri dodeljevanju sredstev, vodenju zaposlenih, določanju pričakovane uspešnosti inštruktorja ipd. Podjetje lahko tudi kvantificira, kako dolgo novi prodajalci v

povprečju potrebujejo, da dosežejo prihodek, ki se od njih pričakuje, ter glede na to oceni učinkovitost novega prodajalca (Roberts, 2013).

Še en primer upravljanja zaposlenih je primer *IBM*-a. Njihov vodja kadrovske službe je s strani ekipe dobil obvestilo, da je med zaposlenimi veliko nezadovoljstva zaradi tega, ker jim podjetje ne povrača stroškov za prevoze, kadar so ti storjeni preko prevozne storitve *Uber*. To odločitev je zaradi potencialne nevarnosti sprejela pravna ekipa, vendar se je *Uber* med zaposlenimi pogosto izkazal kot boljša (cenejša) ali celo edina izbira, da dosežejo svoje stranke. Ko je informacija prišla do vodje kadrovske, je ta takoj odobril nadaljnje kritje uporabe te storitve. Do informacij o nezadovoljstvu je ekipa prišla preko analize internih komunikacijskih sistemov. Problem je bil pravzaprav razrešen, preden so zaposleni sploh vedeli, da je bila kadrovska služba seznanjena z njihovim nezadovoljstvom (Chuapetcharasopon, 2015).

V veliki meri se podatke uporablja za rekrutacijo, ohranjanje in spremljanje karierne poti zaposlenih. Podjetje *Juniper Networks*, ki se ukvarja z razvojem mrežnih infrastruktur, uporablja *LinkedIn* podatke, da spremlja in analizira veščine, znanja, izkušnje in karierne poti trenutno zaposlenih, bivših zaposlenih in kandidatov oz. potencialno zaposlenih. Svojim zaposlenim torej 'sledi' z namenom ugotoviti, kakšne karierne poti imajo tisti, ki odidejo, in kakšen profil zaposlenih zaposlujejo. Take informacije jim lahko koristijo pri nadaljnjem zaposlovanju. Napovedovanje vedenja zaposlenih s pomočjo velikih podatkov izvaja tudi *Deloitte Counseling* – oblikovali so model do 40 dimenzijami, s pomočjo katerega lahko njihova kadrovska služba že mesece vnaprej predvidi, kdo bo zapustil organizacijo (Roberts, 2013). Podatki so lahko koristni tudi pri odločanju v zaposlitvenem procesu – npr. poslovalnica *Wal-Mart* je iskala skladiščnika za nočno delo. Kandidati, ki so bili sicer podobno ustrezni, so se razlikovali le v tem, da je nekaterim bolj ustrezalo delo v nočni izmeni, drugim pa vse izmene. *Wal-Mart* je analiziral že obstoječe podatke zaposlenih na tem delovnem mestu ter odkril, da je manj odpovedi tistih zaposlenih, ki jim ustreza le nočna izmena. (Chuapetcharasopon, 2015).

Uporaba masivnih podatkov se lahko širi preko okvirjev upravljanja z zaposlenimi. Dostavno podjetje *FedEx* v procesu združevanja ali prevzema zunanjega podjetja pregleda podatke njihovih zaposlenih in jih nato primerja s svojimi (npr. v stopnji angažiranosti) ter se preko te primerjave odloči, ali bo prišlo do prevzema. Če podjetje odpira nov oddelek ali poslovalnico lahko združi različne informacije (npr. iz kadrovske in računovodske službe), nato pa preko pregleda podatkov o potrebah in zahtevah delovnega mesta, veščinah in stroških zaposlenih ipd. izbere najbolj optimalno lokacijo (Roberts, 2013). Kadar širjenje še ni vezano na lokacijo, je namreč pri izbiri le-te zelo pomembno analizirati lokacijo z vidika dostopnega kadra (katera lokacija že ima potreben kader, kako ga lahko podjetje izkoristi ipd.). V teh primerih gre za sodelovanje kadrovske službe z drugimi oddelki.

Zakaj uporabljati masivne podatke?

Kot je razvidno iz zgornjih primerov, se je (vsaj v večjih ameriških podjetjih) uporaba masivnih podatkov že dodobra usidrala tudi v kadrovskih službah organizacij. Pregled primerov nam daje tudi vpogled v številna področja in problematike organizacije, kjer je analiza teh podatkov smiselna in uporabna. Poraja pa se morda vprašanje o dodani vrednosti uporabe masivnih podatkov – ali gre res za nadgradnjo tradicionalnih načinov zbiranja in analize podatkov ter odločitev, ki slonijo na takih (tradicionalnih) analizah?

Strokovnjaki s področja se strinjajo, da gre pri masivnih podatkih za uporabo zelo realnih in zelo nazornih podatkov, ki so vezani na ljudi, na zaposlene ('*people-related and employee-related data')*. Z uporabo teh naj bi bolj strateško sprejemali odločitve, ki vplivajo na poslovanje, in izboljšali organizacijsko učinkovitost. Poglavitna 'logika' ali ideja v ozadju analitične kadrovske službe je ta, da v nekatere procese, ki so tradicionalno sloneli predvsem na intuiciji, vnese znanost, natančnost in strogost. Drugi pomemben poudarek zagovornikov pa je '*evidence-based managment';* torej upravljanje zaposlenih, ki prvenstveno sloni na dokazih. Gre za uporabo dokazov, podatkov ter analitično in kritično evalvacijo le-teh pri sprejemanju menedžerskih odločitev. Postavlja se napram uporabi intuicije ali 'najboljše prakse'. Pomembno pa je tudi dejstvo, da so zgoraj omenjeni primeri uporabe velikih podatkov le nekaterih izmed univerzuma možnosti – idejno lahko v kadrovski analitiki ugotavljamo karkoli, če le imamo dostop do potrebnih informacij. Strokovnjaki naštevajo še sledeča relevantna in potencialna področja;

- zaznava blagovne znamke organizacije kot zaposlovalca (*Ali smo mi zaposlovalec izbire? Kako nas kot zaposlovalca vidijo študenti z različnimi nivoji izobrazbe? Na kakšen način se študentje iz različnih univerz spoznavajo/slišijo o naši organizaciji?)*;

- upravljanje s talentom (*Kdaj in zakaj novo zaposleni zapustijo organizacijo? Zakaj se visoko potencialni zaposleni ustavijo pri steklenem stropu? Zakaj določen oddelek izgublja več uspešnega kadra kot drugi?)* in profiliranje talentov (*kakšne značilnosti, izkušnje, osebnost, veščine imajo vodje organizacije ali najbolj uspešni, učinkoviti zaposleni?);*

- segmentacija delovne sile; ocenjevanje učinkovitosti v realnem času; evalvacija raznolikih programov organizacije (npr. ocena učinkovitosti trenutne strategije zaposlovanja).

Ključen je tudi od zgoraj-navzdol pristop kadrovske analitike; najprej se torej pojavi problem (identificira ga navadno vodstvo organizacije), nato pa se za potrebe iskanja rešitev postavi vprašanje in pregleda ali celo zbira podatke. Ostale prednosti HR analitik so torej, da se lahko izboljša sprejemanje odločitev; zmanjšuje tveganja; odkrije pomembne vpoglede, ki bi sicer ostali skriti; nadzoruje

zaposlene ali procese v realnem času; napoveduje dogodke, ki bodo vplivali na poslovanje ali uspešnost organizacije (Chuapetcharasopon, 2015).

Najbolj vabljiv aspekt masivnih podatkov pa je vsekakor napovedljivost. Podatke lahko namreč obdelamo na osnovni ali kompleksni ravni, z nivojem obdelave pa so pogojeni tudi rezultati in odločitve. V organizacijah veliko kadrovske analize ostaja na deskriptivni analizi, katere omejena vrednost je ta, da ne nudi enake moči za iskanje rešitev. Stopnjo nad opisom pa je napoved, ki že omogoči podjetju, da predvidi, kaj se bo zgodilo. Prav v tem leži prava vrednost masivnih podatkov (Roberts, 2013). Le pomislimo na primer v prejšnjem poglavju, kjer je podjetje mesece vnaprej napovedovalo, kdo bo zapustil organizacijo.

Kje naj kadrovska služba išče masivne podatke?

Kot že rečeno se lahko podatke v masivnih bazah loči preko več kriterijev; npr. ali gre za notranje (izvirajo iz organizacije) ali zunanje podatke (zunaj organizacije, npr. iz forumov, družbenih omrežij, statističnih uradov, drugih organizacij). Delimo jih lahko tudi na strukturirane in nestrukturirane, kjer je po sestavi nestrukturiranih veliko več (navadno so to taki podatki, ki se lahko neprestano zbirajo, npr. preko videokamer). Strukturirani podatki lahko izvirajo iz informacijskega sistema kadrovske službe (npr. delovne ure, absentizem) ali iz računovodskih sistemov (plača), v obeh primerih pa gre torej za notranje podatke. Strukturirani podatki so v grobem tisti, ki omogočajo kvantitativno analizo. Nestrukturirani podatki so lahko prav tako lahko notranji (CV, odprti odgovori na interne vprašalnike o zadovoljstvu ali vključenosti) ali zunanji (blog objave, e-maili), so pa navadno proste oblike in neprimerni za kvantitativno analizo.

Kot zunanji vir z mnogimi lastnimi podatkovnimi bazami je lahko *LinkedIn* (ki podjetjem omogoča vpogled v te baze podatkov, iskanje preko filtrov znotraj baz itd.), pa tudi *Facebook*, s pomočjo katerega lahko sledimo, kaj zaposleni o podjetju širijo v svojih zasebnih mrežah – marsikdo namreč vidi svoje zaposlene kot ambasadorje svojega podjetja. Tako si prizadeva zaposliti in ohraniti le tiste, ki o organizaciji širijo dobro besedo (Roberts, 2013).

Dober primer nestrukturiranih podatkov, ki pa jih je zelo smiselno sistematično (računalniško) obdelati, so CV-ji. Pravzaprav avtorji trdijo, da je ravno selekcija oziroma odločanje o zaposlovanju najbolj smiselno in uporabno področje analize masivnih podatkov v kadrovski službi (McCaffery idr., 2015). Psihologi naj bi se pri ocenjevanju CV-jev preveč zanašali na svoje občutke in tradicionalne ocenjevalne metode, npr. metodo mej. Kot primer; v procesu selekcije psiholog izloči vse s povprečjem ocen manj kot 8.0, vendar je meja arbitrarna in ne izhaja iz znanstvene ali utemeljene osnove. Hkrati pa bi drug psiholog za isto delovno mesto lahko izbral neko drugo oceno kot mejno.

KAKO OBDELOVATI MASIVNE PODATKE?

Masivni podatki so najpogosteje nestrukturirani in raznoliki, zato je postopek obdelave lahko nekoliko kompleksen. Marr (2015b) navaja štiri sloje v procesu razvoja nestrukturiranih surovih podatkov v enote, ki jih lahko analiziramo, medtem ko Chuapetcharasopon (2015) govori o fazah analize masivnih podatkov. Spodaj predstavljam postopek kot integracijo idej obeh avtorjev:

1. Sloj virov podatkov

Gre za začetni proces pridobitev podatkov v njihovi osnovni obliki in iz različnih virov. Cilj je nabor podatkov, kjer z nekaterimi že razpolagamo, nekatere pa je morda še potrebno pridobiti. Potrebno je zavedanje, da potrebne podatke lahko pridobimo tudi preko drugih služb znotraj organizacije, vendar je prehodnost teh informacij odvisna od zrelosti podatkov organizacije (nekatere organizacije nimajo večjih baz in morajo podatke zbirati po potrebi; druge imajo ogromne baze, že vnaprej pripravljene za analizo) in centraliziranosti strukture organizacije.

Na tem mestu moramo imeti jasen pregled nad tem, kar imamo dostopno (vred s kakovostjo dostopnih podatkov) in tem, kar nas zanima (kaj je problem in kateri podatki bodo zanj uporabni).

2. *Sloj hranjenja podatkov*

Na tej točki imamo zbrane vse podatke, ki jih lahko nadalje koristimo. Gre za ogromno količino podatkov, zato so že tu pogosto potrebna orodja kot so *Apache Hadoop DFS* ali *Google File System*. Gre za distributivne sisteme, ki lahko hranijo ogromno količino podatkov; navadni računalnik, četudi z večjim trdim diskom, bi prenesel le »manjše« podatkovne baze. Poleg njih potrebujemo ločen sistem, ki bo podatke lahko organiziral in kategoriziral v baze (npr. *HBase*, *DynamoDB* od *Amazon*-a, *MongoDB* in *Cassandra*, ki jo uporablja *Facebook*), saj jih distributivni sistemi le hranijo. Podatke je potrebno ročno očistiti, kar je navadno dolgotrajen proces, gre pa za običajno predpripravo podatkovne baze – popravki manjkajočih podatkov, poenotenje vhodov, re-kodiranje itd.

3. Sloj analiziranja podatkov

S pomočjo orodja *MapReduce* izberemo elemente, ki nas zanimajo, in jih postavimo v format, ki ga lahko nadalje obdelujemo. Obstajajo orodja, ki v podatkih avtomatično prepoznajo trende in vzorce, prav tako pa lahko ekipa analitikov ročno dela na podatkih (torej brez pomoči orodij).

Pri analiziranju je pomembno, da nas vodijo vprašanja oziroma problemi. Velik nabor podatkov nas ne sme prelisičiti v to, da preverjamo povezave levo in desno. Druga pomembna omemba pa je šum v podatkih – le tega je v masivnih podatkovnih bazah dosti več, kot smo ga vajeni v manjših.

4. *Sloj 'output-a' oziroma izhoda*

Na tej točki so analize že opravljene. Sledi sporočanje ugotovitev tistim, ki jim bodo koristile. Navadno gredo najprej do nadrejenih, nato do dotičnih zaposlenih. Vizualizacija podatkov je navadno ključna pri predstavitvi rezultatov laikom. Predstavitev in razlaga naj bosta enostavni, razumljivi in preprosti. Če podatki sami po sebi ne povedo veliko, jih opremimo tudi s slikami (obratno velja za slike; če same niso dovolj predstavne, dodamo konkretne številke).

Torej kako naj organizacija izvaja analizo masivnih podatkovnih baz? Ena možnost je zaposlitev strokovnjaka za HR podatke (ang. *HR data scientist*[1]), ki pa je lahko nekoliko tvegana – taki strokovnjaki so dragi, vnaprej pa se ne da predvideti, kako dobro jim bo šlo. Manj tvegan je najem takega strokovnjaka kot svetovalca organizaciji oziroma njeni kadrovski službi (v ZDA ocenjujejo tako storitev od $120 do preko $300 na uro). Tretja možnost je zaposlitev HR analitika oziroma usposabljanje kadrovske službe v HR analitiki. Ti sicer niso strokovnjaki za masivne podatkovne baze, imajo pa veliko znanja o statistiki in analitiki. Četrta opcija je, da kadrovska služba uporabi orodja, ki jih ponujajo zunanji ponudniki. To so npr. *Oracle, SAP* ali *Workday* (McCaffery idr, 2015).

OMEJITVE

Prva omejitev (sicer bolj analize same kakor masivnih podatkov v ozadju) je to, da čeprav lahko napovedujemo vedenje, ga je še vedno nemogoče spremeniti s pomočjo analitike. Torej, masivni podatki nam lahko povedo, da bo nek zaposlen v sledečem mesecu z veliko gotovostjo zapustil organizacijo, ne ponudijo pa nam 'preračunane' rešitve ali strategije, da to spremenimo – še. Tehnološki napredki se namreč počasi a zagotovo pomikajo v to smer. Druga omejitev je seveda pomanjkanje masivnih podatkov o zaposlenih – mnogo organizacij ne pobira mnoštva podatkov in tako nima 'zrele baze', na kateri bi lahko izvajala analize in napovedovala vedenja. Če organizacija začne z zbiranjem podatkov danes, bodo podatki dovolj 'zreli' šele čez par let (Roberts, 2013).

Omejitve se navezujejo tudi na ugotovljene vzorce ali korelacije, odkrite v podatkih. Četudi se baza in podatki vnaprej sčistijo in sproti ohranjajo, je zaradi samega volumna podatkov možno, da bodo odkrite povsem neveljavne oziroma naključne korelacije.

Pomanjkljivost, ki se navezuje na psihologe v organizaciji, pa je predvsem ta, da ti v osnovi niso izobraženi v smeri razumevanja in zbiranja takih podatkov. Četudi so izučeni v poznejših fazah analize, sporočanja ugotovitev in vizualnega prikazovanja le-teh, morda ne poznajo tehnologij in orodij v ozadju analiziranja velike baze. Morda ne poznajo ali ne znajo načrtovati zbiranja masivnih podatkov od

[1] *HR data scientist* je strokovnjak, ki iz seta podatkov lahko izlušči vrednost ali vpogled. Navadno je vešč v analitiki, računalniških znanostih, matematiki in statistiki, strategiji, vizualizaciji podatkov in komunikaciji (Marr, 2015b).

zaposlenih. Če upoštevamo, da naj bi imel strokovnjak za analizo velikih podatkov na razpolago tudi matematično in računalniško znanje, je zagotovo potrebna identifikacija prednosti pridobivanja teh znanj ter posledično vstop v izobraževanje (ali povezovanje med oddelkom za informacijsko tehnologijo in kadrovskim oddelkom organizacije). Trenutna omejitev je tudi to, da je raba masivnih podatkov v kadrovski službi dokaj nova praksa, posledično je dostop do relevantnih raziskav, primerov praks in etičnih usmeritev zelo omejen. Poleg tega je le malo organizacij v Sloveniji dovolj velikih oziroma zaposluje tako število ljudi, da bi bila izgradnja masivne podatkovne baze sploh možna.

'DATA MINING' IN ETIČNI POMISLEKI

Lanskoletni dogodek 'Statistični dan 2015: Masivni podatki – Velika priložnost', organiziran pod taktirko Statističnega urada RS in Statističnega društva Slovenije, je obravnaval marsikatero področje masivnih podatkov, velik poudarek pa so dali tudi na etične vidike takih praks. Opozarjajo, da je pri delu z masivnimi podatki nujno posvečati povečano pozornost varovanju zasebnosti, skrbeti, da so prakse v skladu z Zakonom o varstvu osebnih podatkov, in da so sami nameni obdelovanja podatkov etični. Upoštevanje zasebnosti je tu velik izziv, saj niti anonimizacija podatkov ne reši več težave. Tuji avtorji analogno poudarjajo, da zakritje imena ne skrije več nujno identitete, saj je le-to sedaj zaradi dostopnosti množičnih in raznolikih podatkov o posamezniku lažje razkriti ali odkriti (Roberts, 2013).

Etična načela moramo upoštevati tudi pri postavitvi problema – čisto vsega namreč ni etično raziskovati. Kot primer, s pomočjo privatnih objav na omrežjih kot sta *Facebook* ali *Twitter* bi lahko potencialno odkrili problematične izjave ali vedenje zaposlenega, vezano na delo ali organizacijo. Znani so primeri, ko so se ljudje 'izživljali' nad šefi v takih objavah, posledično pa bili odpuščeni – resničnost teh primerov je sicer težko preveriti, vendar istočasno obstajajo raziskave, ki kažejo, da je iskanje po brskalnikih oziroma '*googlanje*' zaposlenih ali kandidatov praksa, katere se poslužuje kar 84 odstotkov kadrovikov (SHRM, 2008, v McCaffery, 2015). Vendar ali je sredstvo, ki posega v zasebno sfero zaposlenega, primerno za doseg cilja varovanja podobe organizacije? Pogledamo si lahko tudi predhodno omenjen primer iz podjetja *IBM,* kjer je HR ekipa iz e-poštnih sporočil razbrala nezadovoljstvo in ukrepala k povečanju le-tega. Ali je, primerjalno, ta cilj bolj 'primeren' za uporabo sredstva, ki sicer poseže v zasebnost? Zagotovo gre za neprestano tehtanje pravic in zasebnosti zaposlenega na eni strani in koristi za zaposlene ali organizacijo kot tako na drugi strani, kjer nek postopek ali rešitev, ki bi ustrezala vsaki situaciji, preprosto ne obstaja.

To poseganje pa ni problem, ki bi ostal neopažen. Kalifornija je nedavno postala prva zvezna država v ZDA, ki je pravno zavarovala pravico kandidatov in zaposlenih, da zadržijo zase gesla od družbenih portalov kot sta *Facebook* in *Twitter* (National Conference of State Legislature, 2013, v Shah, Irani in Sharif, 2016). Psihologi v organizacijah se morajo zavedati pomembnosti implikacij dostopanja

do zasebnih podatkov zaposlenih. Možne pravne posledice, s katerimi bodo soočeni, če pri pregledu zasebnih profilov naletijo na potencialno diskriminatorne podatke (o rasi, invalidnostih itd.), ne odtehtajo nujno vseh prednosti pregledovanja in identificiranja »rdečih zastav« pri zaposlenih.

'*Data mining*', oziroma računalniški proces odkrivanja vzorcev v masivnih podatkovnih bazah, je pojem, ki se pojavlja vzporedno z masivnimi podatki. Je način odkrivanja relevantnih informacij, ki se lahko uporabljajo za zniževanje stroškov ali večanje prodaje in je tako zelo popularno orodje raznoraznih prodajalcev in ponudnikov storitev. Primer tega je npr. da trgovina analizira mnoštvo podatkov o svoji prodaji, iz katerih računalniški program razbere sledeče: moški, ki ob četrtkih in sobotah kupijo plenice, z veliko verjetnostjo zraven kupijo tudi pivo. Posledično za povečanje prodaje lahko premaknejo pivo bližje plenicam ali pa zagotovijo, da se ob četrtkih in sobotah pivo ter plenice prodajajo po polni ceni (Palace, 1996). Zaradi razcveta tehnologije pa je to v največji meri prisotno na spletu – brskalniki spremljajo našo aktivnost, te podatke prodajo tretjim deležnikom, ki nam nato tarčno oglašujejo tiste izdelke, katere analiza prikaže kot najbolj verjetne, da jih bomo kupili. Kot tak *'data mining'* ni neposredno prisoten pri delu psihologa v organizaciji, vendar so to po mojem mnenju aspekti, na katere mora biti psiholog vseeno pozoren, v kolikor se v njegovi organizaciji zbirajo masovni podatki o zaposlenih. Opozarjajo namreč na pomembnost varovanja identitete, varovanja dostopnosti podatkov (nadzor in omejitev posredovanja tretjim osebam), tudi na zaznave ljudi o pobiranju in uporabi takih podatkov. Posledično je morda potreben previden pristop in predstavitev praks varovanja etičnih načel zaposlenim, v kolikor organizacija v večji meri zbira in hrani podatke o njihovem vedenju, da preprečimo negativne predstave zaposlenih o teh praksah.

VIRI

Chuapetcharasopon, P. (2015). *HR analytics and »Big« Data through the lens of Industrial/Organizational psychology*. Dostopno na LinkedIn.

Marr, B. (2015a). *Big data explained in less than 2 minutes*. Dostopno na LinkedIn.

Marr, B. (2015b). *Big data: The 4 layers everyone must know*. Dostopno na LinkedIn.

McCaffery, G., Moorstein, G. in Harris, A. (2015). Advances in Big Data. *HRfocus 92*(5), 1 – 4.

Palace, B. (1996). *Data mining.* Zapiski za slušatelje predmeta 'Management' na UCLA. Dostopno na: http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/

Roberts, B. (2013). The benefits of big data. *HR Magazine 58*(10)*,* 21 – 30.

Shah, N., Irani, Z. in Sharif, A. M. (2016). Big data in an HR context: Exploring organizational change readiness, employee attitudes and behaviors. *Journal of Bussines Research 70*(2017)*,* 366 – 378.

SURS (2015). *Zaključki – Statistični dan 2015: Masovni podatki -Velika priložnost.* Dostopno na: http://www.stat.si/dokument/8733/StatDanBrdo2015_Zakljucki.pdf