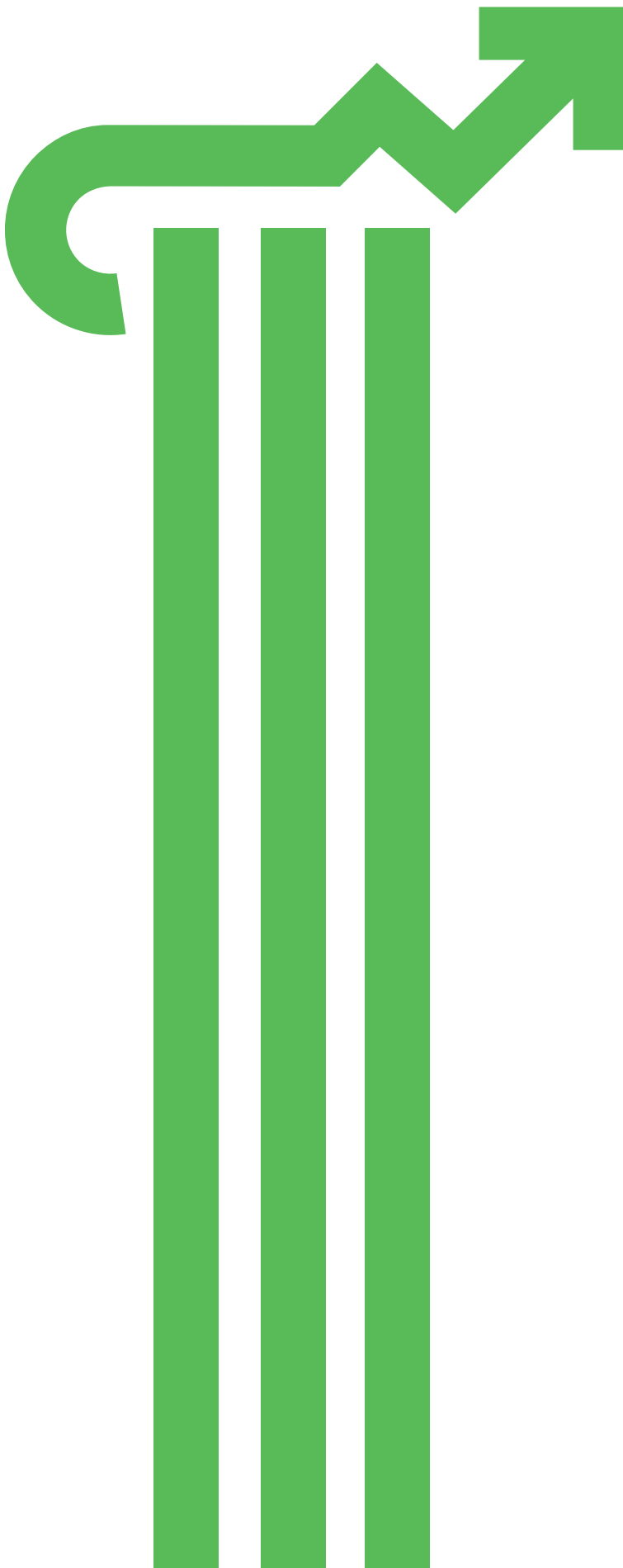




efos

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU
Ekonomski fakultet u Osijeku

JOSIP JURAJ STROSSMAYER UNIVERSITY OF OSIJEK
Faculty of Economics



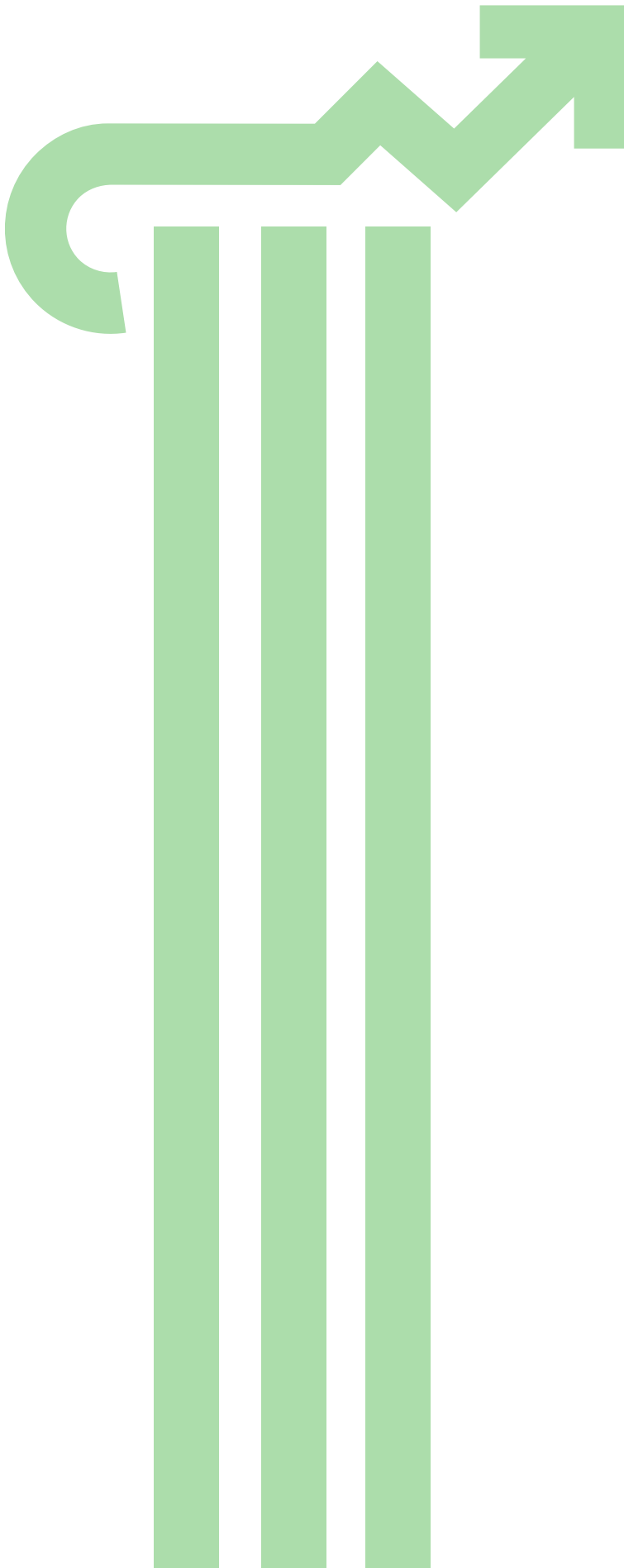
Working Paper Series

Dario Šebalj

Ana Živković

Promjene u upravljanju podacima
koje nosi „BIG DATA“

No. 16-01/ May 2016



Seriya članaka u nastajanju/Working Papers Series

Izdavač/Publisher

Sveučilište Josipa Jurja Strossmayera u Osijeku,
Ekonomski fakultet u Osijeku/*Josip Juraj
Strossmayer University of Osijek, Faculty of
Economics*
Trg Ljudevita Gaja 7
HR-31000 Osijek

Urednici/Editors:

Ivan Kristek, Ph.D.
Hrvoje Serdarušić, Ph.D.

Uređivački odbor/Editorial Board

Antun Biloš, Ph.D.
Boris Crnković, Ph.D.
Davor Dujak, Ph.D.
Ivan Ferenčak, Ph.D.
Ivana Fosić, Ph.D.
Domagoj Karačić, Ph.D.
Dražen Koški, Ph.D.
Branko Matić, Ph.D.
Josip Mesarić, Ph.D.
Sunčica Oberman Peterka, Ph.D.
Mladen Pancić, Ph.D.
Domagoj Sajter, Ph.D.
Antun Šundalić, Ph.D.
Davorin Turkalj, Ph.D.

ISSN 2459-721X

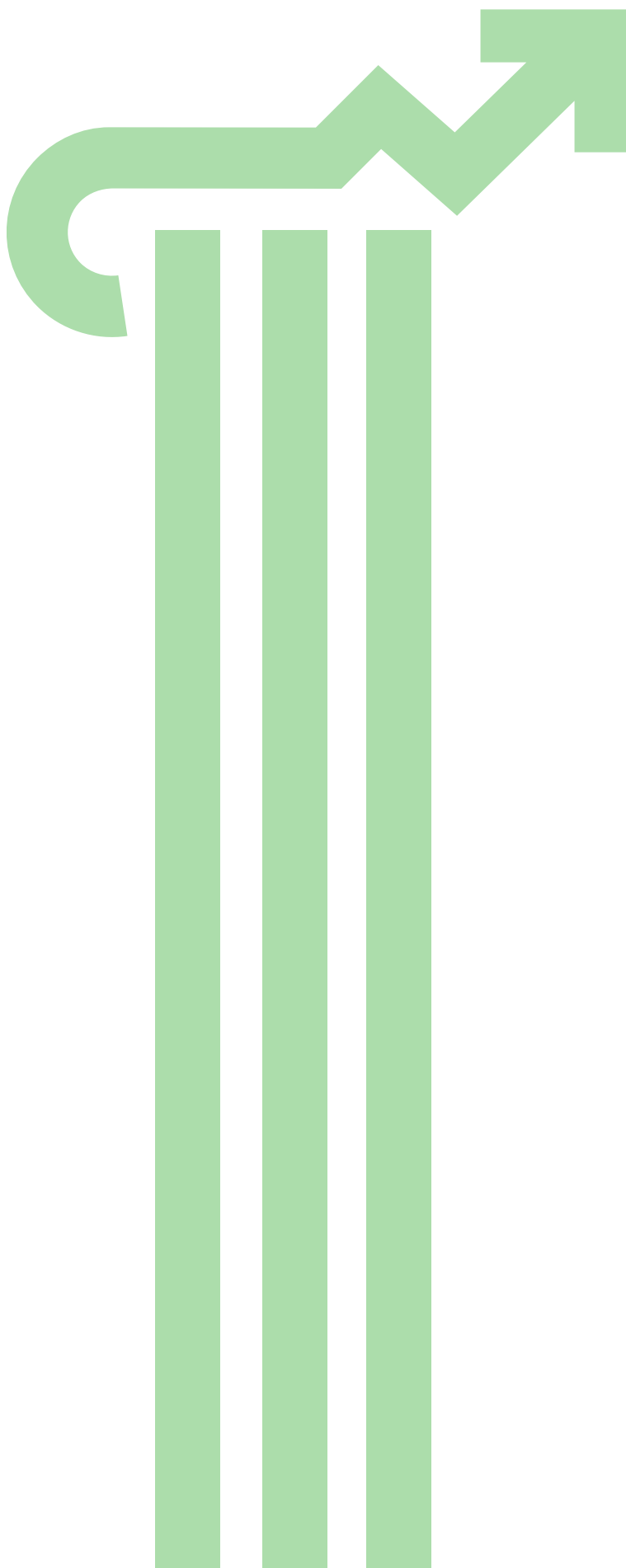
ISBN 978-953-253-143-5

Stajališta i mišljenja autora/koautora iznesena u ovom radu ne odražavaju i nisu nužno istovjetna službenim stajalištima Sveučilišta Josipa Jurja Strossmayera u Osijeku, Ekonomskog fakulteta u Osijeku. Sve primjedbe, komentare i sugestije treba uputiti isključivo autoru/koautorima.

Sva prava pridržana od strane autora/koautora.
Rad je moguće koristiti bez dopuštenja autora isključivo za osobnu upotrebu uz obavezno navođenje izvora. Ukoliko se želi komercijalno ili u ostale svrhe koristiti dio ili cjeloviti tekst rada obavezano je prije toga kontaktirati i dobiti pismeno dopuštenje autora/koautora.

The views and opinions of the author / co-author expressed in this paper do not reflect, and are not necessarily identical to official positions of Josip Juraj Strossmayer University of Osijek, Faculty of Economics. All remarks, comments and suggestions should be referred exclusively to the author / co-authors.

All rights reserved by the author / co-author.
Parts of paper can be used without a permission of the author/ co-author exclusively for personal use only provided the source is stated. If part or full text of a paper is to be used for commercial or other purposes the author/ co-author have to be contacted, and written permissions obtained beforehand.



Cite as: Šebalj, D., Živković, A. (2016) Promjene u upravljanju podacima koje nosi „Big data“, Josip Juraj Strossmayer University of Osijek, Faculty of Economics, Working Paper No. 16-01

Copyright © Dario Šebalj, Ana Živković, May 2016
All rights reserved. Brief excerpts may be reproduced or translated provided the source is stated.

Promjene u upravljanju podacima koje nosi „BIG DATA“

Working Paper Series / No. 16-01

[Dario Šebalj, mag.oec.](#)

Sveučilište J. J. Strossmayera u Osijeku
Ekonomski fakultet u Osijeku
Trg Ljudevita Gaja 7
31000 Osijek, Hrvatska
dsebalj@efos.hr

[Ana Živković, mag.oec.](#)

Sveučilište J. J. Strossmayera u Osijeku
Ekonomski fakultet u Osijeku
Trg Ljudevita Gaja 7
31000 Osijek, Hrvatska
azivkovi@efos.hr

Sažetak

Živimo u digitalnom okruženju u kojem gotovo sve što radimo ostavlja digitalni trag. Podaci se prikupljaju iz raznih izvora - objava na društvenim mrežama, e-mailova, različitih senzora, slikovnih i video sadržaja, pretraga na tražilicama, online kupovina i mnogo drugih. Najveći razlog ovakvog rasta podataka može se naći u tehnološkom napretku budući da se danas podaci mogu lako i jeftino pohranjivati i dijeliti. Ovaj novi trend u generiranju i prikupljanju podataka zahtijeva potpuno novi pristup njihovoj obradi i analizi. Cilj ovoga rada jest da se, na temelju analize aktualnih i relevantnih izvora, prikažu stanje i trendovi u prikupljanju, obradama, analizi i korištenju podataka koji su kompleksni, brzorastući, različiti po tipu i sadržajima. U uvodu je najprije definiran problem istraživanja, zatim je istražen i objašnjen pojam „Big data“, a potom su objašnjene i njegove osnovne 4 dimenzije, odnosno aspekta. U središnjem dijelu rada prikazani su primjeri analize Big data i zanimljivih rezultata koje su te analize polučile, a poseban osvrt dan je na Big Data u menadžmentu.

JEL Classification: M12, M54, D83, L86

Ključne riječi: Big data, menadžment, podaci, SMART model

1. Uvod

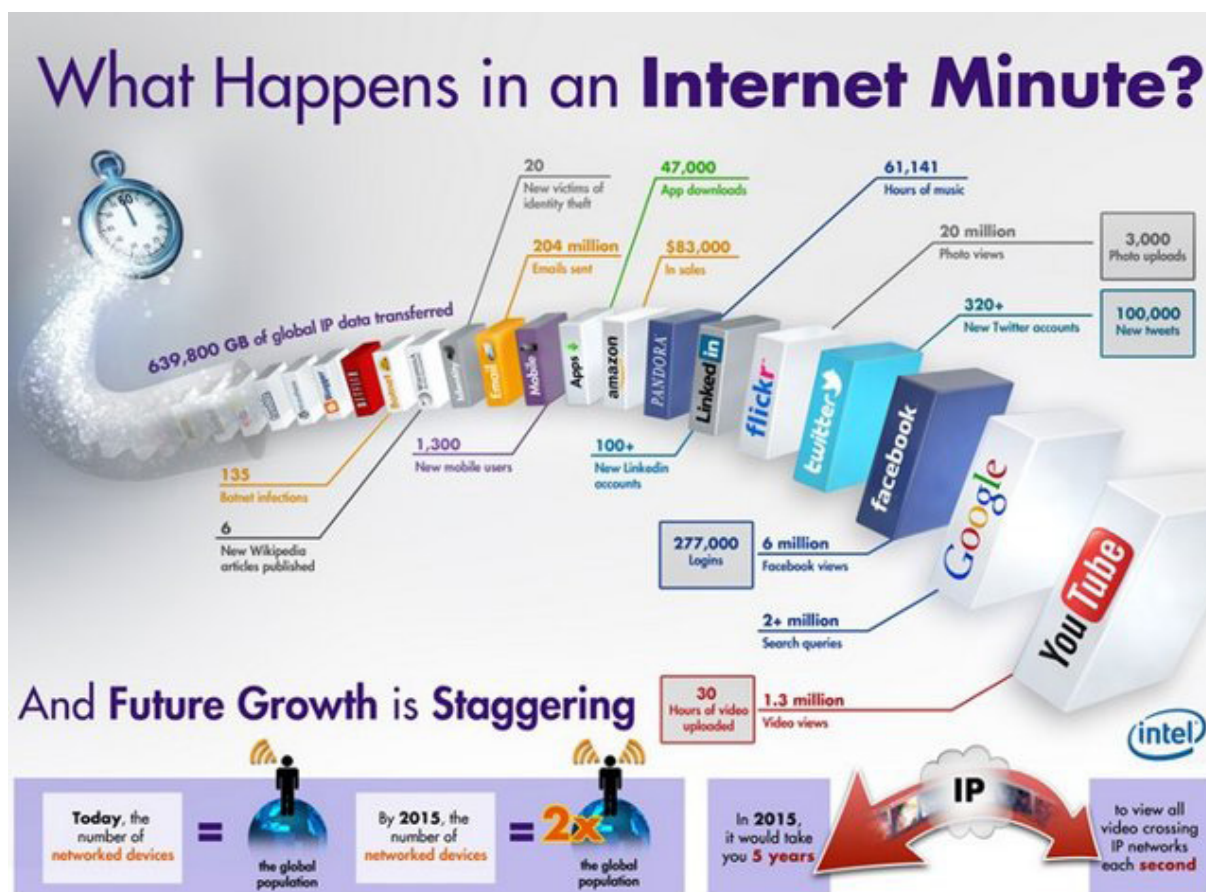
Digitalno okruženje u kojem živimo svakim trenutkom stvara sve veći i veći broj podataka. Najveći razlog ovakvog rasta podataka može se naći u tehnološkom napretku budući da se danas podaci mogu lako i jeftino pohranjivati i dijeliti, što do prije svega par godina nije bio slučaj. Podaci se danas prikupljaju o gotovo svemu što nas okružuje. Gotovo sve što činimo ostavlja digitalni trag koji se kasnije može koristiti za razne analize. Uređaji koje najčešće koristimo (poput mobilnih telefona) sadrže nekoliko senzora koji generiraju hrpu podataka. Sve više i više uređaja je povezano s internetom te takvi uređaji mogu pohranjivati i dijeliti podatke velikom brzinom.

Prema IBM¹-u, ljudi kreiraju 2,5 kvantilijuna bajta podataka svaki dan – toliko da je 90% svih podataka u svijetu danas kreirano u zadnje dvije godine! Podaci se kreiraju na temelju svega – senzora koji prate vremenske i klimatske informacije, postova na društvenim mrežama, digitalnih slika i videa, zapisa o transakcijama o kupnji, GPS signala s mobilnih uređaja, pretraga na tražilicama i dr. Prema Zikopoulosu (2015), Facebook je godinama dodavao nove korisnike svake tri sekunde, danas ti korisnici kolektivno generiraju dvoznamenkasti broj terabajta podataka svaki dan. Točnije, dnevno se na Facebooku napravi 3,5 milijarde objava i oko 155 milijuna „svidanja“ (like-ova).

Uzmimo samo u obzir svoj primjer. Svaki dan primamo i šaljemo e-mailove, intenzivno koristimo tražilice, dopisujemo se s prijateljima i poznanicima putem chat-a i SMS poruka, kupujemo putem web trgovina, plaćamo račune preko mobitela ili interneta, dijelimo slike i video sadržaje na Facebook-u, obavljamo video pozive na Skype-u, koristimo „cloud“ spremišta podataka, pišemo dopise i radimo tablične izračune uz pomoć Google Docs-a... Svi ti podaci, uz mnoštvo dodatnih, spremaju se (najčešće na računalni oblak) te se kasnije koriste za detaljne analize.

¹ Dostupno na: <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html> (pristup: 13.11.2015.)

Slika 1. Što se događa u jednoj minuti na internetu



Izvor: <http://www.sparkyhub.com/what-happens-on-the-internet-in-60-seconds-infographic/> (pristup: 27.11.2015.)

Prema McKinsey izvješću (2011), u 2010. godini, poduzeća i korisnici pohranili su više od 13 eksabajta novih podataka, što je preko 50.000 puta više podataka nego što ih je u kongresnoj knjižnici. McKinsey predviđa i velik utjecaj Big Data na zaposlenost, pri čemu će u SAD-u biti potrebno 140.000-190.000 radnika sa iskustvom u dubinskim analizama. Nadalje, 1,5 milijuna menadžera će morati postati „podatkovno“ pismeni.

Ogroman broj podataka može biti vrlo koristan za poduzeća ukoliko imaju alate i vještine da iz tih podataka „izvuku“ njihovu vrijednost.

Cilj ovoga rada jest da se, na temelju analize aktualnih i relevantnih izvora, prikažu stanje i trendovi u prikupljanju, obradama, analizi i korištenju podataka koji su kompleksni, brzorastući, različiti po tipu i sadržajima.

2. Big Data

O pojmu Big Data u zadnje se vrijeme puno govori, a veoma malo zna. Taj naizgled jednostavan pojam krije mnogo više nego što se na prvi pogled čini. Kako je rečeno u uvodu, danas se prate i pohranjuju podaci o gotovo svemu pa samim time imamo pristup velikom broju podataka. Upravo zbog toga se uvriježio pojam Big Data („veliki podaci“). Marr (2015) u svojoj najnovijoj knjizi ističe kako prava vrijednost podataka nije u njihovom velikom obujmu već u tome što se s tim podacima može napraviti. Također ističe kako razliku ne čini količina podataka već sposobnost analiziranja ogromnih i kompleksnih skupova podataka što se prije nije moglo učiniti. Knapp (2013) definira Big Data kao alate, procese i procedure koji omogućavaju organizaciji kreiranje, manipulaciju i upravljanje jako velikim skupovima podataka i objektima za pohranu. SAS Institute Inc. navodi kako je Big Data pojam koji opisuje veliku količinu podataka (i strukturirane i nestrukturirane, o čemu će kasnije biti riječi u tekstu) koji preplavljaju poduzeća na dnevnoj bazi. I ova kompanija se slaže da nije važna količina podataka već ono što poduzeće s tim podacima može napraviti jer se oni mogu analizirati kako bi se došlo do saznanja koja vode do boljih poslovnih odluka te strateških poslovnih poteza.² Čak je ovaj pojam 2013. godine dodan i u Oxford English Dictionary.³ Magazin Forbes (Arthur, 2013) definira Big Data kao kolekciju podataka dobivenih iz tradicionalnih i digitalnih izvora unutar i izvan poduzeća koja predstavlja izvor za razne analize. IBM na svojoj web stranici navodi kako se „veliki podaci“ generiraju na temelju svega što nas okružuje. Proizvode ih svi digitalni procesi i razmjene putem društvenih mreža. Sustavi, senzori i mobilni uređaji ih odašilju. Kako bi se dobila smisljena vrijednost iz podataka, potrebna je optimalna snaga procesiranja, analitičke sposobnosti i vještine.⁴ O'Reilly na svojim mrežnim stranicama objašnjavaju kako su Big Data podaci koji premašuju kapacitete konvencionalnih baza podataka. Podaci su preveliki, kreću se prebrzo i više ne stanu u strukture klasične baze podataka. Kako bi se dobila vrijednost iz tih podataka, potrebno je izabrati alternativne načine za procesiranje (Dumbill, 2012).

Jennifer Dutcher sa Škole informacijskih znanosti, Berkeley, pokušala je također doći do definicije Big Data tako što je upitala preko 40 vodećih stručnjaka u izdavaštvu, modnom biznisu, prehrambenoj, automobilskoj industriji, medicini i slično da definiraju pojam „Big data“. Tako su neki rekli sljedeće⁵: Jon Bruner (glavni urednik, O'Reilly Media): Big Data je rezultat prikupljanja informacija na najsitnije mogućoj razini; Reid Bryant (ekspert za podatke, Brooks Bell): Big Data će u konačnici opisati bilo koji skup podataka dovoljno velik da zahtijeva visoku razinu programerskih vještina i statističkih metodologija kako bi se podaci transformirali u vrijednost; Rohan Deuskar (direktor i osnivač, Stylitics): Big Data se odnose na pristup podacima „prikupi sada, riješi kasnije“, što znači da prikupljanje i spremanje podataka sa jako velikih količina radnji i transakcija različitih vrsta vrši kontinuirano da bi imali smisla kasnije; AnnaLee Saxenian (dekanica, UC Berkeley Škola informacijskih znanosti): Big Data su podaci koji se ne mogu procesirati koristeći standardne baze podataka jer su preveliki, prebrzo se kreću ili su prekompleksni za tradicionalne alate za procesiranje podataka; Anna Smith (analitički inženjer, Rent the Runway): Big Data označava kada podaci rastu do te točke kada se tehnologija, koja podupire te podatke, mora mijenjati; Mark van Rijmenam (direktor i osnivač, BigData-Startups): Big Data nisu vezane samo za količinu podataka, već i za kombinaciju različitih skupova podataka kako bi se analizirali u realnom vremenu s ciljem dobivanja važnih informacija za poduzeće; Timothy Weaver (izvršni informacijski direktor, Del Monte Foods): Mnoštvo različitih podataka koji dolaze brzo i u različitim strukturama. Gartner, vodeća svjetska tvrtka na području analize, procjene i praćenja trendova u informacijskoj i komunikacijskoj tehnologiji, također je dala svoju definiciju Big Data za koje kažu da su informacijska imovina velikog obujma, velike

² Dostupno na: http://www.sas.com/en_us/insights/big-data/what-is-big-data.html (pristup: 13.11.2015.)

³ Dostupno na: <http://mashable.com/2013/06/13/dictionary-new-words-2013/#lqSgJfPZsPq9> (pristup: 16.11.2015.)

⁴ Dostupno na: <http://www.ibm.com/big-data/us/en/> (pristup: 16.11.2015.)

⁵ Dostupno na: <https://datascience.berkeley.edu/what-is-big-data/#JohnAkred> (pristup: 16.11.2015.)

brzine i/ili velike raznovrsnosti koja zahtjeva isplative, inovativne oblike obrade informacija koje omogućuju poboljšane spoznaje, donošenje odluka i automatizaciju procesa.⁶

Google pretraga za pojam „Big Data“ na dan 13.11.2015. vratila je oko 770 milijuna rezultata. Na prvoj stranici bili su pretežito rezultati velikih IT kompanija SAS-a, IBM-a i Oracle-a. Power (2014) je u svom istraživanju također proveo identičnu pretragu putem Google tražilice. U travnju 2013 pretraga istog pojma vratila je 17,7 milijuna rezultata što znači da je u 2,5 godine broj rezultata narastao za više od 43 puta! To samo govori o popularnosti „velikih podataka“ koja raste doslovno svakim danom. Power je također i dodatno suzio pretragu pa je za pojam „Define Big Data“ (definiraj Big Data) dobio 2,68 milijuna rezultata, a za „What is Big Data“ (Što je Big Data) 24,1 milijun. Sada se za prvi pojam dobiva oko 71 milijun rezultata (26 puta više), dok drugi pojam vraća čak 805 milijuna (za 33 puta više).

Gartner je objavio 2014. godine i svoje godišnje izvješće vezano uz tehnologije koje dolaze, odnosno njihov „*hype*“ ciklus koji predstavlja kretanje nadolazeće tehnologije kroz tipične faze u kojima se odvija njen životni ciklus:⁷ faza „okidanja tehnologije“, „vrhunac prenapuhanih očekivanja“, faza „razočaranja“, „nagib prosvjetljenja“ te „plato produktivnosti“. Prema tom izvješću (slika 2.), Big Data je službeno prošao fazu „vrhunca prenapuhanih očekivanja“ i sada je na putu kroz „razočaranje“. Gartner tvrdi da se to dogodilo prilično brzo jer već postoji dosljednost u načinu na koji se pristupa toj tehnologiji i jer je većina novog napretka samo dodatak, a ne revolucija.⁸

Slika 2. Gartnerov *hype* ciklus za 2014. godinu



Izvor: <http://siliconangle.com/blog/2014/08/19/gartners-hype-cycle-big-datas-on-the-slippery-slope/> (pristup: 17.11.2015.)

⁶ Dostupno na: <http://www.gartner.com/it-glossary/big-data/> (pristup: 16.11.2015.)

⁷ Dostupno na: <http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp> (pristup: 16.11.2015.)

⁸ Dostupno na: <http://siliconangle.com/blog/2014/08/19/gartners-hype-cycle-big-datas-on-the-slippery-slope/> (pristup: 16.11.2015.)

Međutim, u Gartnerovom izvješću za 2015. godinu, Big Data je izbačen iz ciklusa (slika 3.) te se više ne pojavljuje. Razlog tomu leži u činjenici da se Big Data već nalazi kao sastavni dio drugih tehnologija poput autonomnih vozila, interneta stvari (Internet of Things), podatkovne sigurnosti, naprednih analiza i sl.⁹. Još jedan od razloga jest taj što su, prema Gartneru, podaci ključ svih naših odluka, bez obzira na to zovemo li ih „big data“ ili „smart data“ te na njih moramo posebno paziti. Stoga nije potrebno da se ovaj pojam, sam po sebi, nalazi u hype ciklusu.¹⁰

Slika 3. Gartnerov *hype* ciklus za 2015. godinu



Izvor: <http://www.gartner.com/newsroom/id/3114217> (pristup: 17.11.2015.)

Bogatstvo i moć Big Data leži upravo u ogromnoj količini podataka. Danas je moguće tu količinu podataka relativno jeftino prikupiti i pohraniti, što se prije zasigurno nije moglo. Zahvaljujući tome, poduzeća su u mogućnosti vršiti razne poslovne analize na SVIM podacima, a ne kao prije, na određenom uzorku. Taj uzorak, koliko god se mi trudili napraviti ga reprezentativnim, nikada ne može imati kvalitetu cjeline. Tu nastupaju posebni programski alati koji sada mogu analizirati taj veliki broj podataka, bez potrebe izdavanja manjeg dijela. Većina poduzeća nema ni izvora velike količine podataka, financijskih sredstava, podatkovnih eksperata koji znaju analizirati te podatke, kao što je to slučaj sa globalnim i velikim kompanijama poput Amazon-a, Google-a, Facebook-a, eBay-a i sl. Međutim, i na malom broju raznovrsnih tipova podataka se mogu vršiti nezamislive analize. To je još jedan od dokaza da je zapravo pojam Big Data pomalo nespretno stvoren. Prema istraživanju Marr-a (2015), unatoč eksponencijalnom rastu podataka i informacija, manje od 20% podataka koje kompanije trenutno imaju koristi se u svrhu donošenja odluka. I tih 20% su podaci poput financijskih podataka, tradicionalno strukturirani KPI-evi i sl. Međutim, prava moć dolazi prilikom analiziranja tzv. nestrukturiranih podataka.

⁹ Dostupno na: <http://www.contexti.com/blog/big-data-dropped-from-gartners-2015-hype-cycle-for-emerging-technologies/> (pristup: 17.11.2015.)

¹⁰ Dostupno na: <http://www.datasciencecentral.com/profiles/blogs/big-data-falls-off-the-hype-cycle> (pristup: 17.11.2015.)

Mayer-Schönberger i Cukler (2013) objašnjavaju u svojoj knjizi kako nam podaci mogu reći nevjerojatne stvari, otkriti neočekivane korelacije te riješiti, naizgled nerješive, probleme. Pritom se, pri analizi podataka, potrebno držati činjenice ŠTO, a ne **ZAŠTO**. To znači da pri analizi podataka ne bi trebalo tražiti uzroke dobivenim rezultatima. Bitno je ono što rezultati pokazuju, bez obzira što se mogu činiti nelogičnima. Ne treba gubiti vrijeme na traženje uzroka zašto su dobiveni rezultati takvi. Jedan od primjera koji to pokazuje jest poznati američki maloprodajni trgovački lanac Walmart. Istraživanjem kupovnih navika potrošača otkriveno je kako kupci dan ili dva prije nadolazeće oluje masovno kupuju baterijske svjetiljke i točno određenu vrstu čipsa. Za baterijske svjetiljke je logično zašto ih kupuju, međutim to se ne može reći za čips. Sukladno tom saznanju, Walmart je prije svake oluje na ulaz stavio velike količine tog čipsa pri čemu se prodaja dosta povećala.

2.1. Aspekti Big Data

Prije 14 godina, Doug Laney iz META Group-a (danas u Gartner-u) napisao je članak (Laney, 2001) u kojem je odlično predvidio budućnost Big Data. Prvi je spomenuo 3 velika aspekta, odnosno 3 V, i to na primjeru e-trgovine:

- VOLUMEN (Volume)
- BRZINA (Velocity)
- RAZNOVRSNOST (Variety)

Soubra (2012) tvrdi kako ova tri svojstva u potpunosti definiraju Big Data na način da definiraju ekspanziju skupova podataka na raznim mjestima koja se ubrzava kako bi se generiralo još više podataka različitih tipova.

Volumen se odnosi na ogromnu količinu podataka koja se generira svake sekunde (Marr, 2015). Upravo je korist od sposobnosti obrade velike količine informacija glavna značajka Big Data analitike. Spremanje i obrada većeg broja podataka daje bolje poslovne modele (Dumbill, 2012). Laney (2001) je na primjeru e-trgovine ovaj aspekt objasnio na način da niži troškovi e-kanala omogućavaju poduzeću da ponudi svoja dobra i usluge većem broju pojedinaca ili poslovnih partnera te se tako može prikupiti 10 puta veća količina podataka o pojedinačnim transakcijama čime se povećava obujam podataka kojim treba upravljati.

Tablica 1. Prikaz veličina podataka

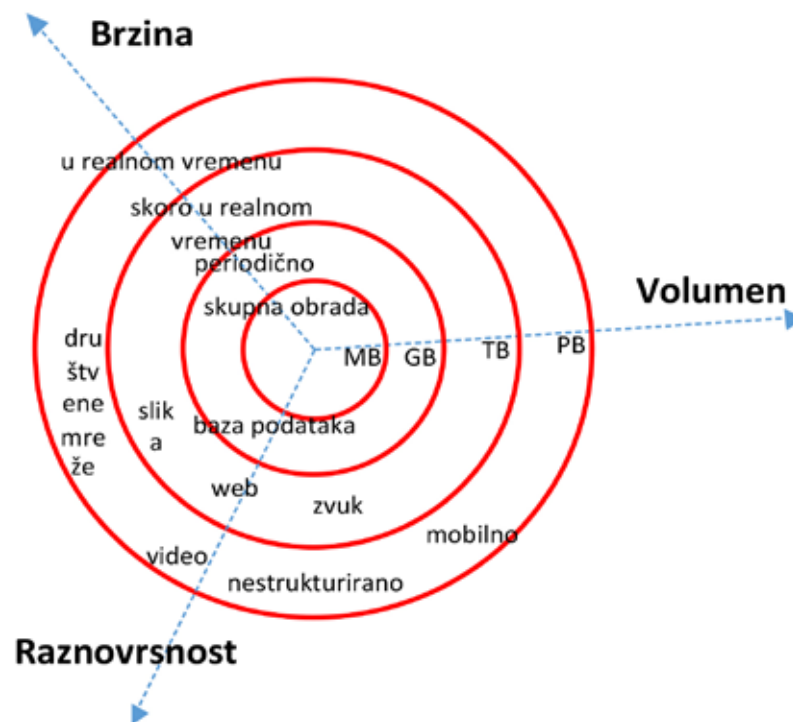
1 bajt	1 znak, slovo, broj
2 kilobajta	1 stranica teksta
1 megabajt	Kratki roman
10 megabajta	Digitalna RTG slika
1 gigabajt	7 minuta HD videa
100 gigabajta	Svi akademski časopisi na jednom katu knjižnice
10 terabajta	Ispisana cjelokupna kolekcija kongresne knjižnice SAD-a
1,5 petabajt	10 milijardi slika na Facebook-u
50 petabajta	Sav napisani materijal čovječanstva, od početka povijesti, u svim jezicima
5 eksabajta	Sve riječi ikad izgovorene
1 zetabajt	250 milijardi DVD-a
1 jotabajt	Veličina cijelog WWW-a; bilo bi potrebno oko 11 trilijuna godina da se s interneta preuzme datoteka veličine 1 jotabajta

Izvor: <https://datascience.berkeley.edu/big-data-infographic/> (pristup: 27.11.2015.)

Brzina se odnosi na brzinu kojom se novi podaci generiraju i kreću oko svijeta. Primjerice, praćenje prijave putem kreditnih kartica prati milijune transakcija u gotovo realnom vremenu kako bi se pronašli neobični obrasci ponašanja (Marr, 2015). Važnost brzine podataka – povećanje stope po kojoj podaci dolaze u organizaciju – prati sličan obrazac kao volumen. Era interneta i mobilnih uređaja omogućava da online trgovci sažmu povijest svakog kupčevog klika ili interakcije (ne samo konačne prodaje). Oni trgovci koji uspiju brzo iskoristiti tu informaciju, primjerice preporučivanjem dodatne kupnje, u ogromnoj su konkurentnoj prednosti (Dumbill, 2012). Laney (2001) objašnjava kako je e-trgovina također povećala brzinu uređaja koji služe kao posrednici u prodaji (POI - point-of-interaction) te time posljedično brzinu i tempo podataka koji se koriste kao podrška interakcijama ili su generirani određenim interakcijama. Soubra (2012) navodi kako su, u početku, tvrtke analizirale podatke koristeći skupnu (*batch*) obradu. Jedna osoba ima dio podataka, dostavlja svoj dio posla na server te čeka isporuku rezultata. Taj način je funkcionirao dok je brzina dolaznih podataka sporija od skupne obrade te kada je rezultat koristan unatoč kašnjenju. S novim izvorima podataka, kao što su društvene i mobilne aplikacije, skupna obrada se raspada. Podaci se sada pohranjuju na serveru u stvarnom vremenu, kontinuirano, a rezultat je koristan samo ako je kašnjenje vrlo kratko.

Raznovrsnost se odnosi na različite tipove podataka koji se generiraju - od financijskih podataka do podataka s društvenih mreža; od slika do podataka sa senzora; od videa do glasovnog snimanja (Marr, 2015). Podaci rijetko kad dolaze u savršeno uređenom obliku i spremnom za obradu. Izvori podataka su različiti, to može biti tekst sa društvenih mreža, slikovni podaci, sirovi podatak direktno iz nekog senzora. Nijedan od ovih podataka nije spreman za integraciju u neku aplikaciju (Dumbill, 2012).

Slika 3. Koncept 3V



Izvor: Prilagođeno prema: Soubra, D. (2012). Dostupno na: <http://www.datasciencecentral.com/forum/topics/the-3vs-that-define-big-data> (Pristup: 17.11.2015.)

U novije vrijeme, s rastom interesa i popularnosti Big Data, pojavljuju se „dodatna slova V“, pa se tako kao četvrto po redu (dodao ga je IBM) navodi **ISTINITOST (Veracity)**. **Istinitost** se odnosi na podatkovni nered koji se generira (Marr, 2015). Odnosi se na nesigurne i neprecizne podatke (Syed, 2013). Autentičnost podataka povećava se automatizacijom prikupljanja podataka. Koristeći više izvora podataka, moguće je povećati istinitost podataka (Moorthy i dr., 2015).

Borne (2014) je u svom članku naveo čak još dodatnih 6V: valjanost (Validity), vrijednost (Value), varijabilnost (Variability), mjesto (Venue), rječnik (Vocabulary) i neodređenost (Vagueness).

2.2. Tipovi podataka

Upravo zbog jednog „V“ tipovi podataka su postali bitna stavka u analizi Big Data. Riječ je o raznovrsnosti (variety). Goes (2014) smatra raznovrsnost najinteresantnijom dimenzijom Big Data. Spajanje podataka iz raznih senzora, „interneta stvari“ (Internet-of-Things), ogromnog repozitorija kojeg nazivamo Web, sadržaja kreiranog od strane korisnika, društvenih mreža, podataka generiranih na mobilnim platformama te podataka iz poslovnog sustava, omogućava znanstvenicima da postavljaju i odgovaraju na pitanja koja objašnjavaju i predviđaju pojedinačna ponašanja i detektiraju popularne trendove. Prema Syed i dr. (2013), Big Data se sastoji od strukturiranih i nestrukturiranih podataka (oko 90% je nestrukturiranih) i uključuju e-mail poruke, objave na društvenim mrežama (Facebook, Tweeter, blogovi), telefonski pozivi, promet na web sjedištima te video prijenose.

Najpoznatiji oblik **strukturiranih** podataka je baza podataka u kojoj su podaci prikazani u stupcima i redovima. Ova vrsta podataka je lako pretraživa, lako čitljiva, „uredno“ složena, sigurna. Međutim, strukturiranim podacima pridaje se manja važnost kada je riječ o Big Data upravo zbog toga što su ti podaci već posloženi, lako ih je analizirati i ono najvažnije – najmanje ih ima. To su, primjerice, podaci o učinjenim POS transakcijama, financijski podaci, podaci o kupcima ili dobavljačima, ulazni i izlazni računi i sl.

Analiza **nestrukturiranih** podataka za sva je poduzeća veliki izazov budući da se ne mogu tako lako svrstati u stupce i redove. Nestrukturirani podaci mogu biti slike i ostali grafički podaci, video sadržaji, web stranice, tekstualne datoteke, e-mailovi, objave na društvenim mrežama, PDF dokumenti, Power-Point prezentacije i sl. Prema Marr-u (2015), podaci se analiziraju iz naših aktivnosti, razgovora, slika i video sadržaja, raznih senzora, interneta stvari. Sve što se pretražuje putem nekog internet preglednika se bilježi. Većina web stranica bilježi koliko je ljudi posjetilo stranicu, odakle ti korisnici dolaze, koliko dugo su se zadržali na web stranici i što su sve „klikali“ na toj stranici. Kada je riječ o glazbenim sadržajima, bilježe se podaci o tome što korisnik sluša, koliko dugo nešto sluša, ali i koje pjesme preskače. Na taj način se ti podaci mogu koristiti kako bi se kreirala lista pjesama na raznim koncertima ili gažama. Upravo to radi Lady Gaga, primjerice. Svaki tjedan se pošalje i pohrani milijarda e-mailova. Zapravo, 20 milijuna e-mailova se napiše u trenutku čitanja ove rečenice. Svakog dana kreira se milijun Twitter računa, svakih 6 sekundi 293.000 promjena statusa se objavi na Facebooku, a svaki dan se učita 350 milijuna slika. U prosječnim mobilnom telefonu nalazi se barem 6 senzora (GPS senzor, senzor koji mjeri akceleraciju, žiroskop, NFC senzor i sl.) i svaki od njih stalno bilježi nove podatke.

Slika 4. Analiza strukturiranih i nestrukturiranih podataka



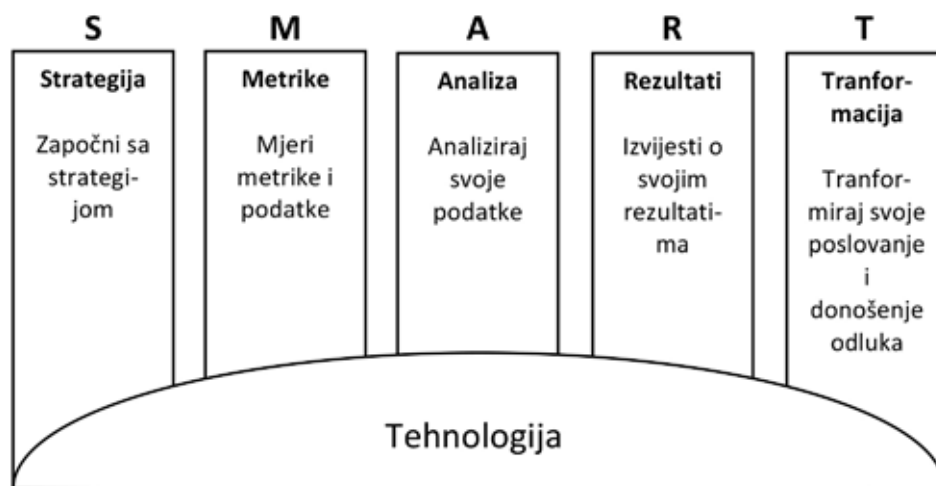
Izvor: izrada autora

Svi ovi gore navedeni primjeri samo pokazuju koliko se danas nestrukturiranih podataka pohranjuje. Za analizu takvih podataka potrebni su posebni programski alati, ali zato mogu dati nevjerojatne rezultate.

3. Primjena SMART modela u analizi podataka

Robert Marr (2015) smatra da je pojam Big Data pomalo zastario te da ne odražava pravo značenje. Njegovo mišljenje je da bi se takvi podaci trebali zvati SMART Data. On tvrdi da Big Data može revolucionirati poslovanje, ali samo ako se fokusira na SMART podatke. Stoga je predstavio SMART model koji bi trebao olakšati kreiranje SMART poslovanja te iskoristiti ogromnu moć Big Data, sukladno veličini poduzeća i budžetu. Marr-ova interpretacija SMART modela vidljiva je na slici 4.

Slika 5. SMART model



Izvor: Prilagođeno prema: Marr, B. (2015). Big Data. Using smart Big Data, analytics and metrics to make better decisions and improve performance. John Wiley & Sons Ltd.

Prema Marr-u, svako bi poduzeće, neovisno o veličini ili pristupu podataka, trebalo početi sa strategijom. Također predlaže kako treba preusmjeriti samo 10% sredstava poduzeća u tehnike otkrivanja podataka, osobito na početku. Razlog tome leži u činjenici da se, bez dobre strukture i procesa, lako može izgubiti dragocjeno vrijeme u otkrivanje podataka. Umjesto kretanja od podataka, treba započeti sa određenim poslovnim ciljevima, a podatke koristiti kako bi se shvatilo koliki je napredak prema ostvarenju tih ciljeva. Time će se automatski prikazati pitanja na koja je potrebno odgovoriti, čime se sužavaju podatkovni zahtjevi na upravljivu razinu. Osim toga, to će osigurati fokusiranost na podatke koji daju stvarnu vrijednost poduzeću. Nakon što se zna što se želi postići, potrebno je istražiti sve moguće tipove i formate podataka koji trenutno postoje, a koji će pomoći u odgovoru na strateška pitanja. Nakon identifikacije tih metrika, dolazi se do analize relevantnih i potencijalno korisnih podataka. Analiza podataka sama po sebi nije dovoljna. Potrebno je napraviti izvješća rezultata kako bi ih ljudi kojima su potrebni (menadžeri) znali čitati s ciljem poboljšanja donošenja odluka i performansi poduzeća.

Središnje tri razine SMART poslovnog modela potpomognute su tehnologijom. Uz pomoć nje moguće je prikupiti podatke koje je potrebno mjeriti, analizirati ih na nezamislive načine te u konačnici kreirati vizualno privlačna izvješća.

4. Alati za Big data analizu u Hrvatskoj – Oracle akademija

Oracle je pokrenuo program Oracle Academy koji omogućava besplatan on-line pristup obrazovnim sadržajima. Dana 12. veljače 2016. godine održan je drugi Oracle Academy dan s temom Big Data. Program je bio podijeljen na 2 velika predavanja:

- Oracle Big Data mijenja poslovanje (predavač: Ljiljana Perica, Oracle Hrvatska)
- Oracle Big Data – Case Studies (predavač: Tomasz Przybyszewski, Oracle Corporation)

Kroz prvo predavanje, Lj. Perica je dala osvrt na aktualne trendove u prikupljanju, obradi i analizi podataka. Navela je kako se svakim danom prikupljaju enormne količine podataka i to su najvećim dijelom, čak 80%, nestrukturirani podaci. Takvi su podaci prikupljeni iz mnoštva različitih senzora (automobili, mobilni uređaji i sl.), društvenih mreža, e-mailova, digitalnih knjiga, slika, video materijala, poziva i dr. Na predavanju je iznesena i informacija da poduzeća koja koriste analize takvih podataka dobiju 10,66 dolara za svaki dolar koji potroše na analizu!

U sljedećoj tablici, prikazanoj na predavanju, vidljivi su novi načini generiranja podataka u raznim industrijama, od maloprodaje pa sve do bankarstva, zdravstva i komunalnih usluga. Jasno su vidljivi novi tipovi podataka (nestrukturirani) na kojima se mogu vršiti razne analize (primjerice, prediktivne).

Tablica 2. Primjeri novih podataka pogodnih za analize

Industrija	Novi podaci	Što se može napraviti?	Zašto?
Maloprodaja Marketing „jedna veličina odgovara svima“	Weblog-ovi; klikovi mišem	Mikro-segmentacija; preporuke	Povećanje neto marže za 60%
Bankarstvo Otkrivanje prijevара, analiza rizika	Weblog-ovi; transakcijski sustavi; izvješća o prijevarama	Otkrivanje uzoraka; Semantička analiza	Milijarde dolara izgubljenih godišnje zbog bankovnih prijevара
Zdravstvo Povećanje kvalitete i efikasnosti	Bilješke liječnika; razne statistike	Najbolja praksa; smanjena hospitalizacija	Povećanje vrijednosti za 300 milijardi dolara godišnje
Komunalne usluge Elastična i prilagodljiva mreža	Pametna očitavanja; Pozivi iz call-centra	Prediktivne analize u realnom vremenu	Očekuje se da će se korištenje energije do 2030. godine povećati za 22%
Usluge bazirane na lokaciji Temeljem poštanskog broja	Privatni lokacijski podaci	Geolokacijski mar- keting; Promet; Lokalne pretrage	Promet povećan za više od 100 milijardi dolara

Izvor: Izrada autora, prilagođeno od Perica, Lj., Oracle akademija

U nastavku predavanja je i objašnjena Oracle-ova platforma za upravljanje velikim podacima i njihovo analiziranje.

Drugi dio Oracle dana bilo je predavanje Oracle Big Data – Case Studies, gdje je gosp. Tomasz Przybyszewski naveo nekoliko primjera analize velikih podataka iz različitih područja. Neki od primjera objašnjeni su u sljedećem poglavlju: Big data – case study.

5. Big data – case study

Brojni su primjeri analiza „velikih podataka“ koje su dovele do nevjerojatnih i vrlo korisnih otkrića. Neke od njih navedene su u nastavku.

Mayer-Schönberger i Cukier (2013) u svojoj knjizi navode nekoliko primjera gdje su analize big data dovele do nevjerojatnih otkrića:

- Google je uzeo 50 milijuna najčešćih izraza koje Amerikanci pretražuju i usporedio ih sa listom podataka američkog centra za kontrolu i prevenciju bolesti (CDC) o širenju gripe između 2003. i 2008. godine. Ideja je bila pronaći područja koja su inficirana gripom na temelju onoga što ljudi pretražuju na internetu. Njihov sustav je tražio korelacije između učestalosti pojedinih pretraživanih izraza i širenja gripe u određenom vremenu i prostoru
- 2003. godine Oren Etzioni letio je iz Seattlea u Los Angeles. Mjesecima prije leta kupio je kartu putem interneta, vjerujući kako će jeftinije proći ako kartu kupi ranije. Međutim, u zrakoplovu mu se pobudila znatiželja te je upitao osobu koja je sjedila pored njega koliko je platila kartu i kada ju je kupila. Ispostavilo se da je ta osoba jeftinije platila kartu iako ju je kupila puno kasnije. Tako je ispitao nekoliko putnika i većina njih je platila kartu manje, a kupili su je kasnije. Par dana nakon leta odlučio je napraviti model (budući da je bio računalni znanstvenik) koji će predvidjeti hoće li se cijena, koja je aktualna u određenom trenutku, povećati ili smanjiti u budućnosti. Pri tome ga nisu zanimali čimbenici koji utječu na formiranje cijene. Koristeći uzorak od 12.000 cijena letova prikupljenih sa web stranice zrakoplovne tvrtke u razdoblju od 41 dana, uspio je kreirati željeni prediktivni model. Njegov maleni projekt evoluirao je u startup nazvan Farecast koji je kasnije kupio Microsoft za 110 milijuna dolara
- Koliko je korisna velika količina podataka pokazuje primjer Google-ovog alata za prevođenje. Kada je projekt pokrenut, 2006. godine, iako pompozno najavljivan, bio je poprilično loš. Gotovo svi prijevodi su bili netočni. Tada je Google počeo proučavati multilingvalne web stranice raznih korporacija, prijevode službenih dokumenata, izvješća Ujedinjenih naroda i Europske unije, pa čak i prijevode raznih knjiga. Google-ov sustav uspio je prikupiti milijarde stranica prijevoda. Njihov algoritam za prevođenje je tada gledao te prikupljene podatke i računao vjerojatnosti da će, primjerice, određena riječ u engleskom slijediti drugu te se tako povećala preciznost njihovog prevoditelja
- Dr. Carolyn McGregor i njezin tim znanstvenika sa sveučilišta Ontario Institute of Technology rade s mnoštvom bolnica na softveru koji bi pomagao liječnicima u donošenju boljih dijagnostičkih odluka prilikom brige za nedonošćad. Softver bilježi i obrađuje pacijentove podatke u realnom vremenu pri čemu prati 16 različitih skupova podataka, poput otkucaja srca, disanje, krvni tlak, razinu kisika u krvi, što zajedno iznosi cca. 1.260 jedinica podataka u sekundi. Na tako velikom broju podataka softver je uspio signalizirati moguću infekciju 24 sata prije nego što se infekcija dogodi. Metoda se ne temelji na uzroku, već na korelacijama. Govori što se događa, ne i zašto se to događa

- Najpoznatiji proizvođač online igara za Facebook - Zynga, otkrio je kroz svoju big data analizu da kupovina virtualnih dobara koje igrači mogu kupiti svojom bankovnom karticom, uvelike ovisi o boji. Primjerice, u igri FishVille igrači kupuju prozirne ribe čak 6 puta više nego ribe običnih boja. Nakon što su to njihove analize pokazale, Zynga je ponudila još više prozirnih stvari te tako višestruko profitirala. Također, u igri Mafia Wars, pokazalo se da igrači kupuju više oružja sa zlatnim okvirom.

Marr (2015) također navodi nekoliko primjera:

- Američki lanac trgovina Walmart je tako svojim kupcima, ako se nalaze 3 milje od Walmarta koji ima, npr. čistač za roštilj u zalihu, a vrijeme je sunčano, davao vaučer za popust na taj artikl
- Jedna telefonska kompanija otkrila je da je moguće, temeljem telefonskih i tekstualnih uzoraka, kao i analize društvenih mreža, klasificirati klijente u razne kategorije. Istraživanja pokazuju da ljudi u određenoj kategoriji imaju veću sklonost raskidanju ugovora i prelasku konkurenciji
- Jedan auto proizvođač, radeći s vanjskim poduzećem koje radi analize, primijetio je da senzor u rezervoaru za gorivo ne radi najbolje. Proizvođač je mogao reći dobavljaču i zamoliti ga da popravi grešku, ali bi tada to poboljšanje imali i ostali proizvođači automobila. Umjesto toga, proizvođač je izumio softversku zakrpu (*patch*) koja popravljala grešku, primio patent za to i prodao patent dobavljaču
- Marr je osobno, kao vlasnik tvrtke koja se bavi analizom big data, radio analizu prodaje za jednu trgovinu odjećom. Ugradio je senzor u izlog trgovine odjećom koji prati signale s mobilnih uređaja kada osobe prođu pored izloga. Sve će njih, kada prođu pored izloga, uhvatiti taj senzor i brojati te dati odgovor na pitanje "*Koliko je osoba prošlo pored trgovine*". Senzor također bilježi koliko je osoba *stalo pred izlogom* i koliko se dugo *zadržalo ispred*, koliko je osoba *ušlo u trgovinu*, a prodajni podaci dat će odgovor na pitanje koliko je njih *kupilo neki proizvod*. Odgovori na sva ova 4 pitanja pomogli su vlasniku trgovine da vidi isplati li mu se uopće ta poslovnica
- Ako se sluša glazba preko smartphona ili digitalnog glazbenog playera, podaci se također prikupljaju o tome što slušamo, koliko dugo i koje smo pjesme preskočili. Lady Gaga koristi ovakve podatke kako bi kreirala playlistu za koncerte
- Američka tinejdžerica je uspjela iskoristiti tehnologiju i ogromnu količinu podataka kako bi kreirao program za dijagnozu raka dojke koji u 99% slučajeva ispravno identificira rak. Brittany Wenger se sa 15 godina zainteresirala za neuronske mreže i računalno programiranje. Nakon što je njezinoj sestričnici otkriven rak dojke, kreirala je neuronsku mrežu koja je uspjela učiti i detektirati uzorke koji ne mogu biti otkriveni golim okom
- Jedna banka je uspjela upola srezati troškove zaposlenika analizirajući uspješnost zaposlenika koji dolaze s različitih sveučilišta. Prije je banka pretpostavljala da će najbolji zaposlenici biti oni sa odličnim ocjenama Ivy League sveučilišta. Analiza podataka je pokazala krivo - najbolji kandidati su bili sa ne baš prestižnih sveučilišta te je omogućeno banci selekciju pravih kandidata za manje novca
- Maloprodajno poduzeće koristi analizu društvenih mreža kombiniranu s drugim analitičkim alatima kako bi pronašla prave kandidate za posao. Analizirajući profile s društvenih mreža mogli su precizno predvidjeti razinu inteligencije te emocionalnu stabilnost potencijalnih kandidata

- Analizirajući razne skupove podataka osoba koje žele zaposliti i one koje žele izbjeći, jedna je tvrtka našla da je vrsta pretraživača (browsera) koji su kandidati koristili za prijavu na posao bila jaki prediktor za pravog kandidata. Kandidati koji su koristili pretraživače koji nisu bili instalirani na računalu (poput Chromea, Firefoxa i sl.) bili su bolji za taj posao
- Jedna tvrtka je otkrila da zaposlenici call-centra sa kriminalnim dosjeom postižu bolje rezultate od onih bez dosjea te da prodavači s manje Facebook kontakata postižu lošije rezultate od onih s manje kontakata

Big Data u menadžmentu

Kao što je spomenuto na predavanju Oracle Academy, Ljiljana Perica, Oracle Business Solution Leader, ističe kako je preko 80% podataka u organizaciji nestrukturirano: od telefonskih poziva, e-mailova, društvenih mreža, senzora pa sve do različitih video, audio i slikovnih zapisa. Financijske transakcije te službeni i neslužbeni organizacijski dokumenti samo su djelomično strukturirani tako da se uistinu u malom postotku može govoriti o strukturiranim podacima. Raznovrsnost i brzina otežavaju upravljanje podacima, a podrazumijeva se, naravno, da što je više prikupljenih podataka, još je teže njima upravljati. Tehnologija je toliko napredovala da je rečenice i fraze telefonskih poziva (recimo u call centrima) moguće pretočiti u matematičke zapise koji će dati informacije o preferencijama potrošača i njihovim stavovima o proizvodima (Katz, 2015). Mobilno okruženje igra važnu ulogu u upravljanju podacima pružajući nevjerojatnom brzinom ogromne količine različitih podataka (Park et al., 2014). U organizaciji se zapravo svi potrebni podatci već nalaze, samo ih je potrebno pronaći i pretvoriti u valjane informacije. Menadžeri uglavnom posjeduju hrpe podataka, ali najčešće ne znaju što bi s njima (Finley et al., 2015). Jedan jedini podatak različite osobe mogu drugačije tumačiti, odnosno iz jednog se podatka može izvući više različitih informacija (De et al., 2016). Zadovoljstvo kao zavisna varijabla u organizacijama je i interno i eksterno jedna od najvažnijih stavki svake organizacije. Interno zadovoljstvo odnosi se na zaposlenike i sve djelatnike neke organizacije, a eksterno se nastoji postići od strane potrošača. Povećanju i jedne i druge skupine zadovoljstva može doprinijeti Big Data. Primjerice, organizacije koje nude on-line usluge kao što je Amazon, mogu povećati zadovoljstvo svojih potrošača na jednostavan način, praćenjem „klikova“, točnog vremena utrošenog na njihove usluge, promatranjem preciznog termina kojim se povećava posjećenost njihovih stranica i slično.

Jedan od idealnih primjera jesu organizacije koje nude on-line gledanje filmova. Oni zahvaljujući Big Data platformi mogu odgonetnuti profile svojih potrošača, pa time i nagovijestiti njihova buduća ponašanja. Korisni podatci javljaju se u okviru vrste filmova koju najčešće gledaju, specifičnom žanru koji gledaju u određeno vrijeme, njihovim eventualnim ocjenama i recenzijama, popisu dosada već odgledanih filmova, popisu filmova koji nisu pogledani do kraja, a započeti su i slično. Zapravo, sve aktivnosti koje potrošač obavlja u, ili preko neke organizacije, bilježe se i pohranjuju. Zahvaljujući praćenju prethodno spomenutih čestica, moguće je potrošaču preporučiti film koji dosada još nije pogledao, a s obzirom na njegove dosadašnje navike i preferencije koje su „iščitane“ iz spomenutih podataka. To se može shvatiti i kao svojevrsno istraživanje tržišta, samo je riječ o puno jednostavnijoj i jeftinijoj varijanti kojom se brže može doći do željenih informacija. Troškovi pohrane podataka su niski, a pohranu podataka ne stvaraju ljudi, nego uređaji. Upravljati se zadovoljstvom stoga može, ne samo na razini propitivanja postojećih potreba i interesa, već i nametanja novih, makar pilot-istraživanjima. Prodiranje u profil potrošača povećava se povezanošću s njegovim društvenim mrežama čime se stvara temeljitiji uvid u njegova dosadašnja iskustva, životne, potrošačke, ali i kupovne navike, emocije i sve oblike interakcija. Tako se stječe uvid ne samo u demografske i transakcijske podatke, već i kontekstualne.

Nije Big Data primjenjiva samo u on-line organizacijama. Maloprodaja neke organizacije kontaktira s potrošačima preko više različitih kanala pa se interakcija i transakcije potrošača pohranjuju u različite baze podataka kako bi se lakše tumačio potrošačev/kupčev izbor. To znači da trgovina na malo može točno povezati eventualne kupovine kupaca s marketinškim kampanjama, a mogu obaviti i analizu košarice. To služi upravo prepoznavanju njihovih kupovnih navika, ali i predviđanju istih. Svi se ti podatci, od strukturiranih kao što su kupčeve recenzije pa sve do nestrukturiranih, pohranjuju u takozvane silose, a Big Data pospješuje njihovu segmentaciju. Zahvaljujući sensorima u prodavaonicama, QR kodovima i njihovoj dubokoj analitici, povećava se prodaja, smanjuju se troškovi skladištenja, i najvažnije – povećava se zadovoljstvo kupaca. S aspekta kupca, moguće je u organizacijama u kojima se uspješno upravlja podatcima, putem on-line zahtjeva, a u suvremenijima i očitavanjem kodova putem mobilnih aplikacija, saznati postoji li traženi broj cipela u nekoj drugoj prodavaonici na svijetu te automatska on-line kupovina istih.

Zaposlenici u organizacijama nisi ni svjesni koliko su danas „praćeni“. Nije samo riječ o klasičnom nadgledanju, već upravljanju produktivnošću. Naime, senzori u prostorijama omogućavaju nadgledanje svih njihovih kretanja. To daje uvid u apsentizam, ali i prezentizam, odnosno u vrijeme koje zaposlenici fizički potroše unutar same organizacije. Ukoliko je riječ o fiksnom tipu radnoga mjesta, bilježe se i interna kretanja u organizaciji. Ovakva praksa ne bi trebala služiti neetičkim nadziranjima, već uistinu prepoznavanju kretanja koje je potrebno smanjiti. Novčana vrijednost vremena itekako je važna jer se gubitkom spomenutoga, troše sredstva. Sensori će omogućiti uviđanje primjerice pretjeranoga kretanja prema kopirnom stroju čime će se ukazati potreba za dislociranošću stroja ili nabavkom dodatnoga. Ljudsko oko teško može izmjeriti vrijeme utrošeno na takva kretanja, ali zato uređaji mogu precizno u sekundama izračunati vremenske gubitke. Big Data može biti i alat za analizu individualnih i timskih ponašanja zaposlenika. Moderne organizacije posjeduju senzore i bedževe kojima se bilježi samostalni i timski rad te vrijeme provedeno na općim i specifičnim radnim zadacima (Gerard et al., 2014). Big Data je u tom okviru kontrolni sustav koji pomaže povezivanju ponašanja s organizacijskim rezultatima, odnosno mjerenju performansi koju predstavljaju produktivnost zaposlenika (Warren et al., 2015). Ranije spomenuti telefonski pozivi kao izvori podataka, ne služe samo za kontrolu potrošača i klijenata, već i kontrolu zaposlenika. Banke i druge organizacije koje nude svoje usluge i savjete putem telefonskih poziva na taj način mogu nadzirati zaposlenike provjeravajući daju li oni potrošačima potrebne i valjane informacije. Matematički zapisi razgovora uputit će nadređene na eventualna odstupanja i pogreške te im ukazati na zaposlenike koje je potrebno upozoriti, dodatno educirati, ili ih se u konačnici riješiti. Dinamiku zaposlenika stoga nikada nije bilo lakše pratiti pa se dosadašnji dnevници o radu smatraju itekako zastarjelim alatom. U organizaciji se mjerenjem utjecaja formalnih i neformalnih komunikacijskih kanala također može odrediti organizacijsko ponašanje te predvidjeti buduće. Ipak, mnogi stručnjaci HRM-a, smatraju kako je Big Data manje primjenjiva platforma u upravljanju ljudskim potencijalima nego što je to u drugi ekonomskim i neekonomskim granama (Chynoweth, 2015). Oni ističu kako kvaliteta podataka varira jer su ljudi osjetljiva čestica koja se ne može lako objektivizirati i generalizirati kao što je to slučaj primjerice s materijalnim resursima. Ljudsko ponašanje je kompleksno i predvidjeti buduća ponašanja zaposlenika nije nimalo lako oslanjajući se samo na objektivne podatke. Zapravo, HRM menadžeri navode kako je učenje upravljanja podatcima u njihovom području neisplativo jer je odluke o ljudima često bolje donijeti na temelju empirije i intuicije. Unatoč tome, ova kritika ne smije osporiti koristi Big Data-e u menadžmentu jer Big Data više pomaže, nego što odmaže.

Big Data ne pomaže samo u upravljanju apsentizmom, produktivnošću i zadovoljstvom, već može pomoći i u procesu selekcije kandidata za pojedina radna mjesta. Istovremeno iz strukturiranih i nestrukturiranih podataka, moguće je saznati dob, spol, obrazovanje, dosadašnje radno iskustvo,

interese, hobije i rekreaciju, mjesto stanovanja, društvene događaje koje osoba posjećuje, članstva i pripadnosti različitim zajednicama, kupovne navike, obiteljske odnose, društvene odnose, postignuća, politička opredjeljenja i tako dalje. Ukoliko se dodatna testiranja kandidata obavljaju računalnim i virtualnim putem, moguće je zapravo eliminirati ogroman broj kandidata sužavanjem izbora na eventualno nekoliko poželjnih. Na taj se način izbjegava utrošak vremena, novca i energije što je posebno važno na suvremenom tržištu radne snage koje obiluje kandidatima sve različitijih profila osobnosti, razina obrazovanja, sposobnosti, kompetencija i vještina. Podrazumijeva se kako računala i roboti nikada neće u nekim segmentima moći zamijeniti čovjeka pa kontakt „licem u lice“ nikada neće izgubiti svoju važnost, međutim istina je i kako se adekvatnim upravljanjem podacima u organizacijama ne moraju intervjuirati stotine kandidata, već samo posljednjih nekoliko koji prema određenim zahtjevima ispunjavaju sve uvjete i kriterije.

Posebno važna korist Big Data-e očituje se na primjerima predstavljenima na Oracle Academy predavanju koji izdvajaju veliku upotrebu Big Data-e u telekomunikacijskim kompanijama, bankama, osiguravajućim društvima i u industriji zabave. I u javnom sektoru je ova platforma itekako primjenjiva: zdravstvo, obrazovanje i promet. Specifičan su primjer i policijski odjeli. Naime, Odjel čikaške policije za kriminalne istrage, ubrzao je detektiranje kriminalnih djela putem podataka prijašnjih kriminalnih incidenata, njihovih bilješki, podataka o osumnjičenima, transkriptima 311 poziva i slično (Oracle Academy predavanje). Pomoć programa očituje se u bržem pronalaženju osumnjičenih, ali i njima sličnih koji bi mogli biti potencijalno osumnjičeni.

Big Data donosi u menadžmentu potpuno nove, dosada neistražene mogućnosti koje mogu doprinijeti stvaranju dodatne vrijednosti. Pitanje konkurentnosti na tržištu jedno je od ključnih problema suvremene organizacije pa su nerijetki slučajevi dvaju ili više konkurentskih organizacija koje na tržištu nastupaju s komplementarnim proizvodima. Dapače, te organizacije djeluju na istom tržištu, i makar ne postoji druga poveznica, tržišta im je zajedničko. U praksi tih je poveznica više: od zajedničkih dobavljača, klijenata, geografske blizine, jednake tehnologije i proizvodnih procesa. Detaljne informacije o takvim odnosima možda formalno nisu dostupne, no nalaze se u nestrukturiranim podacima koji mogu biti različitih oblika. Konkurenciju je potrebno pratiti na sve mogu načine, od praćenja članaka i intervjuja konkurentskih organizacija u medijima, do praćenja njihovih web stranica, blogova i društvenih mreža. Intenzitet praćenja konkurentске organizacije, ne bi trebao biti znatno manji od praćenja vlastite. On-line pregledi, „klikovi“, „lajkovi“ i „tvitovi“ izvor su informacija o vlastitim proizvodima i uslugama, ali i o konkurentskim. Svako primijećeno odstupanje vlastitog proizvoda od željenog stanja uzbuna je za brze akcije na potrošačke reakcije. Takve se situacije događaju i kod konkurenata, pa će aktivno praćenje njihovih potrošačkih reakcija otvoriti prostor vlastitoj organizaciji za „napad“ na slabe točke konkurenata.

Najvažniji rezultati primjene Big Data-e očituju se u brzom donošenju valjanih odluka, brzom rješavanju problema i predviđanju budućih događaja, većoj produktivnosti zaposlenika i cijele organizacije te većoj konkurentnosti na tržištu. U području menadžmenta, odnosno upravljanja, Big Data postaje sve dominantniji koncept u upravljanju opskrbnim lancima (Waller, Fawcett, 2013). Prediktivna analitika u logistici zahvaljujući dosadašnjim pojavama i ponašanjima nastoji predvidjeti buduća, odnosno tijekom skladištenja zaliha i njihove troškove. Strukturirani podatci u logistici nisu novost, ali prikupljanje nestrukturiranih podataka uvelike bi moglo izmijeniti dosadašnju percepciju SCM-a. Senzori mogu detektirati zalihe u skladištu, lokacije na kojima na policama ponestaje određenih proizvoda, pretrpanost u distribucijskom centru, zagubljeni inventar i tako dalje.

Ne postoji obrazovni sustav koji proizvodi isključivo „podatkovne upravitelje“ jer upravljanje podatcima zahtjeva multidisciplinarni pristup. Teško će inženjeri moći samostalno rješavati medicinska pitanja. Preduvjet za uspješnu primjenu Big Data-e primjerice u upravljanju opskrbnim lancima stoga zahtjeva prvenstveno odlično poznavanje materije, dakle teorije i prakse SCM-a, na koju se nadograđuju znanja o upravljanju podatcima i Big Data-i. Zato u organizacijama ponekad nije dovoljno imati osobu koja posjeduje samo iskustvo u analiziranju podataka u informatičkom i tehničkom smislu, već se zaposlenik istovremeno mora razumjeti i u primarni posao kojim se organizacija bavi. Zbog toga možda i nije previše korisno „outsoursati“ ovakvu vrstu posla, jer samo interni zaposlenici, s dugim iskustvom u poslu koji je predmet analize, i dugim iskustvom u samoj organizaciji, mogu se adekvatno koristiti Big Data-om. Göb (2014) ističe kako uspješnoj upotrebi Big Data-e, prethodi visoka razina statističkih znanja i vještina. Analitičke vještine nikako ne smiju isključivati menadžerske vještine. Gerard et al. (2014) ističu kako uspješni „podatkovni“ znalci moraju biti sposobni uvidjeti pravi poslovni problem, ali s podatkovnog stajališta. S obzirom da se od osobe koja bi se procesu odlučivanja uspješno koristila Big Data-om očekuju menadžerske vještine, tehničke i informatičke vještine, te povrh svega poznavanje same organizacije, može se pretpostaviti da taj posao u organizaciji može timski obavljati više osoba. Big Data u upravljanju opskrbnim lancima očekuje poznavanje statistike, metoda predviđanja i vjerojatnosti, optimizacije, analitičkog matematičkog modeliranja, financija, marketinga i računovodstva (Waller, Fawcett, 2013). Daje se zaključiti kako je potrebno posjedovati istovremeno mogućnost kvalitativnog i kvantitativnog razmišljanja. Teško je u jednoj osobi pronaći sve potrebne zahtjeve jer nije dovoljno samo uvidjeti i prepoznati, već i razumjeti neke pojave i predvidjeti buduće. Kombinacija mladosti i iskustva zbog toga je često dobitna kombinacija. Stariji i iskusniji zaposlenici imaju velika znanja o svojoj profesiji, ali nisu fleksibilni i brzi te skloni tehnološkim promjenama kao mlađe generacije kojima, s druge strane, nedostaje iskustvo u samoj profesiji, ali i u organizaciji.

Big Data značajno unapređuje računovodstvene poslove te audit (Griffin, Wright, 2015). Zahvaljujući Big Data-i, eksterni, ali interni auditori imaju uvid u povijest čitave organizacijske kulture. Kontrola kvalitete postaje time realnija jer dostupnost i tumačenje podataka minimizira rizik subjektivne procjene što auditorima uvelike olakšava posao, ali i savjest. Dijeljenje podataka mora biti zaštićeno, a privatnost osigurana. Pristup podatcima ne može imati svatko, a oni koji imaju, moraju njima etički upravljati. Socijalni aspekt Big Data vrlo je važna komponenta upravljanja podatcima jer je neetičko ponašanja glavna kritika digitalne ekonomije. Bojazan od krađe podataka postaje veća i većom upotrebom Big Data-e, otvara se više puteva kradljivcima različitih podataka. Hakeri su sposobni preko noći preuzeti sve financijske informacije o nekoj organizaciji, ali i o njihovim klijentima i potrošačima (Katz, 2015) i time poljuljati lojalnost potrošača nekoj organizaciji. Na razini organizacijskog sustava, na meti su najčešće službeni i neslužbeni e-mailovi kroz koje je moguće dobiti i „meta-podatke“ (primjerice nečiju adresu, broj telefona ili broj računa). Zbog toga Big Data istovremeno treba služiti i kao zaštita od svih kriminalnih aktivnosti. Za korisnike bankarskih usluga to je posebno važno pa mnoge banke upozoravaju korisnike koji promjene uobičajeni bankomat, zemlju ili način pristupa računima kontrolirajući autentičnost. Čak i računi i profili na mrežama i on-line uslugama upozoravaju korisnike pri prijavi s novih IP adresa nastojeći time sačuvati privatnost svojih korisnika.

Big Data nije moda u menadžmentu, već nužna potreba. Pomaže odgovoriti na pitanja kako nešto učiniti bolje, brže i jeftinije. Pitanje je dana kada će se organizacije početi dijeliti na one koje koriste Big Data u poslovanju i na one koje ne koriste (Finley et al., 2015). Organizacije koje otvaraju mogućnost za snimanje, obradu, analizu i distribuciju podataka u svrhu donošenja odluka, vjerojatno će nadmašiti svoje konkurente i odgovarati brže na potrebe svojih potrošača.

6. Zaključak

Big Data zasigurno mijenja način na koji poduzeća i njihovi menadžeri donose odluke. Sve veća pažnja pridaje se nestrukturiranim podacima koji su donedavno bili zanemarivani zbog malih kapaciteta pohrane takvih podataka te zbog nemogućnosti njihove dubinske analize. Razvojem tehnologije i omogućavanjem manipulacije takvim podacima, nestrukturirani podaci postali su glavni interes poslovnih analiza. Ogromnu prednost stječu ona poduzeća koja imaju mogućnosti prikupljanja velikog broja podataka te njihove pohrane. Kao što se moglo i vidjeti iz ovog rada, takvi podaci pokazuju kako postoje nevjerojatne (i na prvi pogled nelogične) korelacije između dvije ili više varijabli. Pritom je važno naglasiti da je većina takvih rezultata posljedica analize svih podataka ($N = \text{sve}$), a ne analize reprezentativnog uzorka te su podaci analizirani bez početnih hipoteza koje su se testirale. Ovdje je riječ o potpuno drugačijem pristupu. Najprije se analiziraju svi podaci te dobivaju jake (ili slabe) korelacije između dvije ili više varijabli. Kao što se moglo vidjeti na primjeru Walmart-a i veze između nadolazeće oluje te veće kupovine određene vrste čipsa, dobivene korelacije ne moraju imati logično objašnjenje. Ono što je bitno jest da se na temelju takvih analiza mogu donijeti poslovne odluke koje pozitivno utječu na poslovanje poduzeća.

Literatura

“Big Data” Dropped From Gartner’s 2015 Hype Cycle for Emerging Technologies. (16.09.2015.). Preuzeto 17.11.2015. iz Contexti: <http://www.contexti.com/blog/big-data-dropped-from-gartners-2015-hype-cycle-for-emerging-technologies/>

Arthur, L. (15.08.2013.). What is Big Data. Preuzeto 16.11.2015. iz <http://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/>

Big Data (n.d.). Preuzeto 16.11.2015. iz Gartner: <http://www.gartner.com/it-glossary/big-data/>

Big data: What it is and why it matters (n.d.). Preuzeto 13.11.2015. iz SAS: http://www.sas.com/en_us/insights/big-data/what-is-big-data.html

Borne, K. (11.04.2014.). Top 10 Big Data Challenges - A Serious Look at 10 Big Data V’s. Preuzeto 17.11.2015. iz Converge blog: <https://www.mapr.com/blog/top-10-big-data-challenges-%E2%80%93-serious-look-10-big-data-v%E2%80%99s#.VLk8ly6mRYo>

Chynoweth, C. (2015). Stop doing dumb things with data. *People Management* (prosinac 2015), 42-44.

De, K., Klimentov, A., Maeno, T., Mashinistov, R., Nilsson, P., Oleynik, D., Panitkin, S., Ryabinkin, E & Wenaus, T. (2016). Accelerating Science Impact through Big Data Workflow Management and Supercomputing. *EPJ Web of Conferences*, 108.

Dumbill, E. (11.01.2012.). What is big data? Preuzeto 17.11.2015. iz O’Reilly: <https://www.oreilly.com/ideas/what-is-big-data>

Dutcher, J. (06.11.2013.). Data Size Matters [Infographic]. Preuzeto 27.11.2015. iz Berkeley School of Information: <https://datascience.berkeley.edu/big-data-infographic/>

Dutcher, J. (03.09.2014.). What is Big Data. Preuzeto 16.11.2015. iz Berkeley School of Information: <https://datascience.berkeley.edu/what-is-big-data/#JohnAkred>

Finley, F., Blaeser, J., & Djavairian, A. (2014). Harnessing Big Data : building & maintaining capabilities that deliver results. *Supply Chain Management Review*, 19(5), 24-31.

Gartner Hype Cycle (n.d.). Preuzeto 16.11.2015. iz Gartner: <http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp>

- George, G., Haas, M. R., & Pentland, A. (2014). Big Data and Management. *Academy of Management*, 57(2), 321-326.
- Göb, R. (2014). Discussion of “Reliability Meets Big Data: Opportunities and Challenges”. *Quality Engineering*, 26(1), 121-126.
- Goes, P. B. (2014). Big Data and IS Research. *MIS Quarterly*, 38(3), 3-8.
- Griffin, P. A., & Wright, A. M. (2015). Commentaries on Big Data's Importance for Accounting and Auditing. *Accounting Horizons*, 29(2), 377-379.
- Katz, D. M. (2015). Big Data, Smaller Risk. *CFO*, 31(8), str. 38-41.
- Knapp, M. (2013). Big Data. *Journal of Electronic Resources in Medical Libraries*, 10(4), 215-222.
- Laney, D. (06.02.2001.). 3D Data Management: Controlling Data Volume, Velocity, and Variety. Preuzeto 17.11.2015. iz META Group: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers Hung, A. (n.d.). Big Data: The next frontier for innovation, competition, and productivity. Preuzeto 18.11.2015. iz McKinsey Global Institute: http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_exec_summary.ashx
- Marr, B. (2015). Big Data. Using smart Big Data, analytics and metrics to make better decisions and improve performance. John Wiley & Sons Ltd.
- Mayer-Schönberger, V., & Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think. John Murray Publishers .
- Moorthy, J., Lahiri, R., Biswas, N., Sanyal, D., Ranjan, J., Nanath, K., & Ghosh, P. (2015). Big Data: Prospects and Challenges. *The Journal for Decision Makers*, 40(1), 74-96.
- Park, J., Kim, H., Jeong, Y.-S., & Lee, E. (2014). Two-phase grouping-based resource management for big data processing in mobile cloud computing. *International Journal of Communication Systems*, 27(6), 839-851.
- Power, D. J. (2014). Using 'Big Data' for analytics and decision support. *Journal of Decision Systems*, 23(2), 222-228.
- Soubra, D. (05.07.2012.). Data Science Central. Preuzeto 17.11.2015. iz The 3Vs that define Big Data: <http://www.datasciencecentral.com/forum/topics/the-3vs-that-define-big-data>
- Syed, A., Gillela, K., & Venugopal, C. (2013). The Future Revolution on Big Data. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(6), 2446-2451.
- Vorhies, W. (17.08.2015.). Big Data Falls Off the Hype Cycle. Preuzeto 17.11.2015. iz Data Science Central: <http://www.datasciencecentral.com/profiles/blogs/big-data-falls-off-the-hype-cycle>
- Waller, M. A., & Fawcett, S. E. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 34(2), 77-84.
- Warren, D. J., Moffitt, K. C., & Byrnes, P. (2015). How accounting records will change with Big Data. *Accounting Horizons*, 29(2), 397-407.

Wasserman, T. (13.01.2013.). Oxford English Dictionary Adds 'Crowdsourcing,' 'Big Data'. Preuzeto 16.11.2015. iz Mashable: <http://mashable.com/2013/06/13/dictionary-new-words-2013/#344s3oQLgPqn>

What is Big Data (n.d.). Preuzeto 16.11.2015. iz IBM: <http://www.ibm.com/big-data/us/en/>

What is big data? (n.d.). Preuzeto 13.11.2015. iz IBM: <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>

Wheatley, M. (19.08.2014.). Gartner's Hype Cycle: Big Data's on the slippery slope. Preuzeto 16.11.2015. iz Silicon Angle: <http://siliconangle.com/blog/2014/08/19/gartners-hype-cycle-big-datas-on-the-slippery-slope/>

Zikopoulos, P., deRoos, D., Bienko, C., Andrews, M., & Buglio, R. (2015). Big Data Beyond the Hype. McGraw-Hill Education.