

PODATKOVNA ZNANOST I NJEZIN UTJECAJ NA POSLOVANJE

Glavinić, Ante

Undergraduate thesis / Završni rad

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Polytechnic of Šibenik / Veleučilište u Šibeniku**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:143:822842>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-09-22**

Repository / Repozitorij:

[VUS REPOSITORY - Repozitorij završnih radova Veleučilišta u Šibeniku](#)



VELEUČILIŠTE U ŠIBENIKU
ODJEL MENADŽMENTA
PREDDIPLOMSKI STRUČNI STUDIJ INFORMATIČKI
MENADŽMENTA

ZAVRŠNI RAD

PODATKOVNA ZNANOST I NJEZIN
UTJECAJ NA POSLOVANJE

Ante Glavinić

Šibenik, rujan, 2021.

VELEUČILIŠTE U ŠIBENIKU

ODJEL MENADŽMENTA

PREDDIPLOMSKI STRUČNI STUDIJ INFORMATIČKI MENADŽMENT

**PODATKOVNA ZNANOST I NJEZIN UTJECAJ NA
POSLOVANJE**

Kolegij: Poslovni informacijski sustavi

Mentor: Jerko Acalin

Student: Ante Glavinić

Broj indeksa: 1219054724

TEMELJNA DOKUMENTACIJSKA KARTICA

Veleučilište u Šibeniku

Završni rad

Odjel Informatički Menadžmenta

Preddiplomski stručni studij Menadžment

PODATKOVNA ZNANOST I NJEZIN UTJECAJ NA POSLOVANJE

ANTE GLAVINIĆ

Adresa, aglavini@vus.hr

(stranice / slika / tablice / literaturni navod / jezik izvornika: hrvatski)

Rad je pohranjen u:

Ključne riječi: podatkovna znanost, podatak, poslovanje, veliki podatci

Mentor: Jerko Acalin

Rad je prihvaćen za obranu:

BASIC DOCUMENTATION CARD

Polytechnic of Šibenik

Final paper

Department of Management

Professional Undergraduate Studies of
Management

Data science and its impact on business

Address, aglavini@vus.hr

(pages /figures / tables / references / original in Croatian language)

Paper deposited in:

Keywords: data science, data, business, big data

Supervisor: Jerko Acalin

Paper accepted:

Sadržaj

1.	UVOD	1
2.	PODATKOVNA ZNANOST	2
2.1	Podatak i informacija.....	2
2.2	Općenito o podatkovnoj znanosti	3
2.3	Osnovne strukture podataka	4
2.4	Razlika poslovne inteligencije i podatkovne znanosti.....	6
2.5	Koraci podatkovne znanosti	8
3.	VELIKI PODATCI (<i>BIG DATA</i>).....	11
3.1	Općenito o Big Data	11
3.2	Karakteristike Big Data	12
3.3	Primjena Big Data	15
3.4	Distribuirana obrada podataka.....	17
3.4.1	Apache Hadoop.....	18
4.	UTJECAJ PODATKOVNE ZNANOSTI NA POSLOVANJE	20
4.1	Podatkovni znanstvenik.....	20
4.2	Učinak podatkovne znanosti.....	22
4.3	Pozitivni i negativni učinci podatkovne znanosti.....	23
4.4	Primjena podatkovne znanosti u tvrtki Coca Cola	26
4.4.1	Razvoj proizvoda uz pomoć podatkovne znanosti.....	27
4.4.2	Društvene mreže i oglašavanje.....	28
5.	ZAKLJUČAK	30
6.	LITERATURA.....	31

1. UVOD

Podatkovna znanost je interdisciplinarno područje koje koristi znanstvene metode, procese, algoritme i sustave za izvlačenje znanja i uvida iz strukturiranih i nestrukturiranih podataka te primjenu znanja i djelotvoran uvid iz podataka u širokom rasponu primjene.

U prvom poglavlju rada objašnjen je sam pojam podatka, kojeg smo razlučili od značenja riječi informacija. Zatim je objašnjeno značenje i sam koncept podatkovne znanosti te strukture podataka s kojima se podatkovni znanstvenici susreću. Definirana je i razlika između podatkovne znanosti i umjetne inteligencije.

U drugom poglavlju pojašnjeno je što su veliki podatci (eng. Big Data). Specificirane su karakteristike velikih podataka, kao i primjena tih znanja u različitim sektorima. Naveden je i proces obrade podataka te na koji način se podaci pretvaraju u korisne informacije.

U trećem poglavlju je elaborirana je primjena podatkovne znanosti, fokusirano na primjenu u poslovnom području. Definiran je pojam podatkovnog znanstvenika te njegova uloga u samom analiziranju velikih podataka. Da bi što bolje razradili i shvatili podatkovnu znanosti i njezin utjecaja na poslovanje, kao primjer navedena je tvrtka Coca Cola. Coca cola je tvrtka za bezalkoholna pića koja se bavi proizvodnjom, prodajom i prodajom bezalkoholnih pića te je jedna od prvih velikih globalno priznatih marki, van tehnološkog sektora, koja je ukomponirala korištenje velikih podataka u donošenju strateških odluka.

2. PODATKOVNA ZNANOST

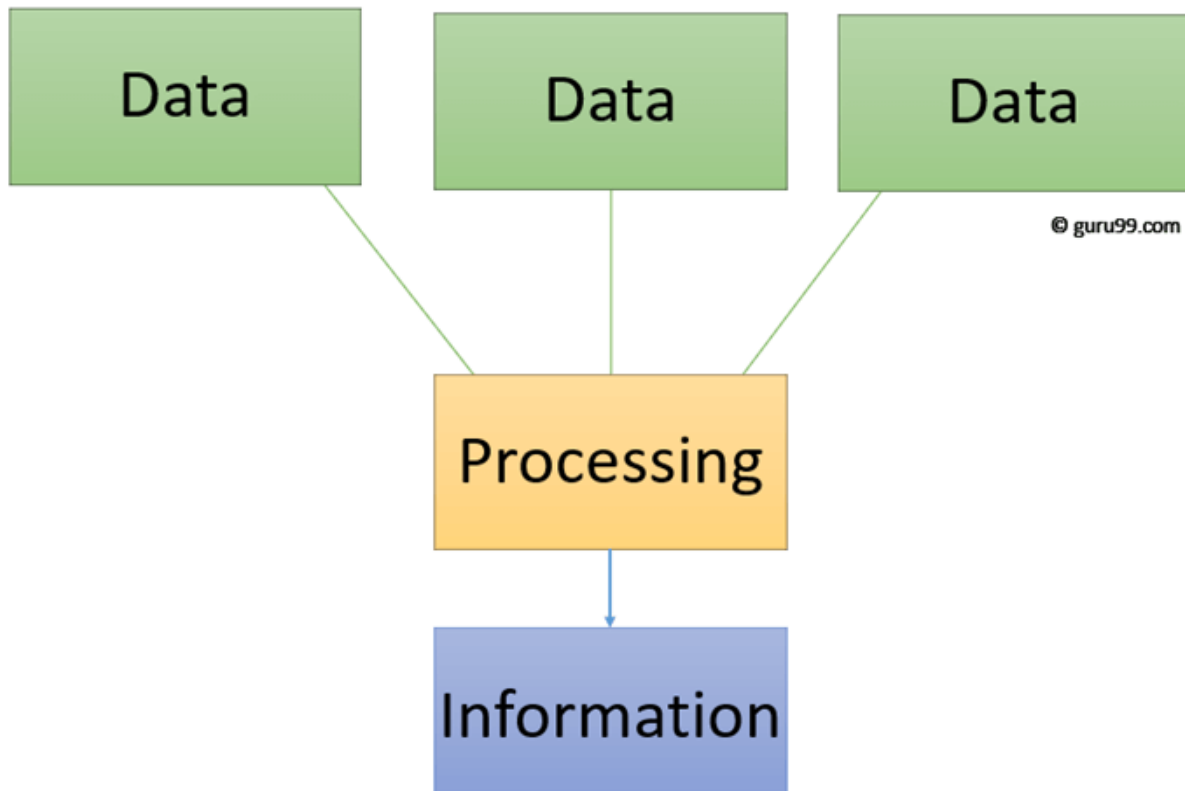
2.1 Podatak i informacija

Iako se izraz "podatak" i "informacija" koriste naizmjenično, njihova značenja su različita. Riječ podatak vuče korijene iz latinske riječi mn. lat. *datum*, što znači dio informacije, što nam govori da je podatak zapravo samo djelić informacije. Podatak nam predstavlja jednostavnu, neobrađenu činjenicu koja ima svoje značenje. Koriste se u znanstvenim istraživanjima, upravljanju poduzećima (primjerice prodaja, prihod, dobit,..), financijama te upravljanju u gotovo svim oblicima organizacijskih aktivnosti ljudi. Podatci se mjere, prikupljaju, izvještavaju te analiziraju, nakon čega ih je moguće vizualizirati, kako bi iz njih dobili konkretnu informaciju. Podaci referiraju na činjenicu da su neke postojeće informacije i znanja predstavljena ili kodirana u nekom obliku pogodnom za bolju upotrebu ili obradu. Neobrađeni podatci su zbirka brojeva ili znakova koje istraživač "očisti" i ispravi. Neobrađene podatke je potrebno ispraviti zbog očitih greški u instrumentu te kako bi uklonili izvanredne podatke.

Oblici podataka pojavljuju se u raznim oblicima, primjerice: zvučni, slikovni, brojčani te tekstualni, a njihova struktura je apstraktna i čine ju: značenje, vrijednost i vrijeme. Možemo reći da se podatci pretvaraju u informaciju kada se gledaju u kontekstu ili analizi.

Obrada podataka je proces pretvorbe podatka u informaciju. Da bi se podatak shvatio kao informacija, on mora imati novu vijest za primatelja, tj. njegovo znanje mora biti uvećano nakon primitka podataka. Obrada podataka obično se odvija u fazama, a obrađeni podaci iz jedne faze se mogu smatrati neobrađenim podacima u drugoj fazi.

Informacija dolazi od latinske riječi lat. *informare*, što znači informirati ili obavještavati. Informacija, kao rezultat analize i organizacije podataka, daje novu znanje primatelju. Interpretirana je kao takva kada je stavljena u kontekst ili kada joj je dodano značenje.



Slika 1. Razlika između podatka i informacije [preuzeto s <https://www.guru99.com/difference-information-data.html>]

2.2 Općenito o podatkovnoj znanosti

Prije podatkovne znanosti (eng. *Data Science*), postojao je pojam eng. *Data Mining*, koji se smatra kao korijen današnje podatkovne znanosti. Nakon toga, znanstvenik pod imenom William S. Cleveland, podiže *Data Mining* na veći nivo, spajajući računalnu znanost i *Data Mining*. Pritom smo statistiku učinili više tehničkom. Sve ovo se događalo u vrijeme kada je i sam web postao naprednije i počeo služiti kao platforma za dijeljenje slika i videozapisa.

Znanost o podacima (*Data Science*) je interdisciplinarno područje koje koristi različite znanstvene metode, procese, algoritme i sustave sve u svrhu izvlačenja znanja i uvida u strukturirane i nestrukturirane podatke. Spoj je više disciplina, koje uključuju: statistiku i računalnu znanost,

informacijsku tehnologiju i analizu podataka te ostalih srodnih područja kako bi se "razumjeli i analizirali stvarni fenomeni".

Područje znanosti o podacima usmjereno je na izdvajanje znanja iz skupova podataka, koji su obično jako veliki. Također obuhvaća analizu, pripremu podataka za analizu te predstavljanje nalaza za donošenje odluka na visokoj razini u organizaciji. Kao takav predstavlja široko područje koje upotpunjuje spektar znanja iz matematike, statistike, vizualizacije informacije, grafičkog dizajna, složenih sustava, komunikacije i poslovanja.

Statistika, kao disciplina, se uglavnom počela razvijati u prošlom stoljeću. Teorija vjerojatnosti, koja predstavlja matematički temelj za statistiku, razvijena je od 17. do 18. stoljeća na temelju djela Thomasa Bayesa, Pierre-Simona Laplacea i Carl Gausa. Za razliku od čisto teoretske prirode vjerojatnosti, statistika je primijenjena znanost koja se bavi analizom i modeliranjem podataka.

Moderna statistika, kao rigorozna znanstvena disciplina, vuče korijene iz kasnih 1800-ih. Vodeći pionir moderne statistike bio je R. A. Fisher uvodeći ključne ideje eksperimentalnog dizajna i procjene najveće vjerojatnosti.

Jim Gray zamišljao je znanost o podacima kao "četvrtu paradigmu" znanosti (empirijska, teorijska, računska i u budućnosti podatkovna). Ustvrdio je da se sve u znanosti mijenja zbog utjecaja informacijske tehnologije i zbog velike količine podataka.

Ovi i mnogi drugi statistički koncepti temelj su za razvoj podatkovne znanosti.

2.3 Osnovne strukture podataka

Dvije su osnovne strukture podataka:

1. Numeričke strukture podataka

Obilježje prema kojemu vrijednost dodjeljujemo prema omjernoj skali zovemo numeričko obilježje. Numeričke strukture podataka pojavljuju se kroz dva oblika: kontinuirano i diskretno. Kontinuirani su oni podatci koji se neprestano pojavljuju, tj. nema prekida. U sustavima su

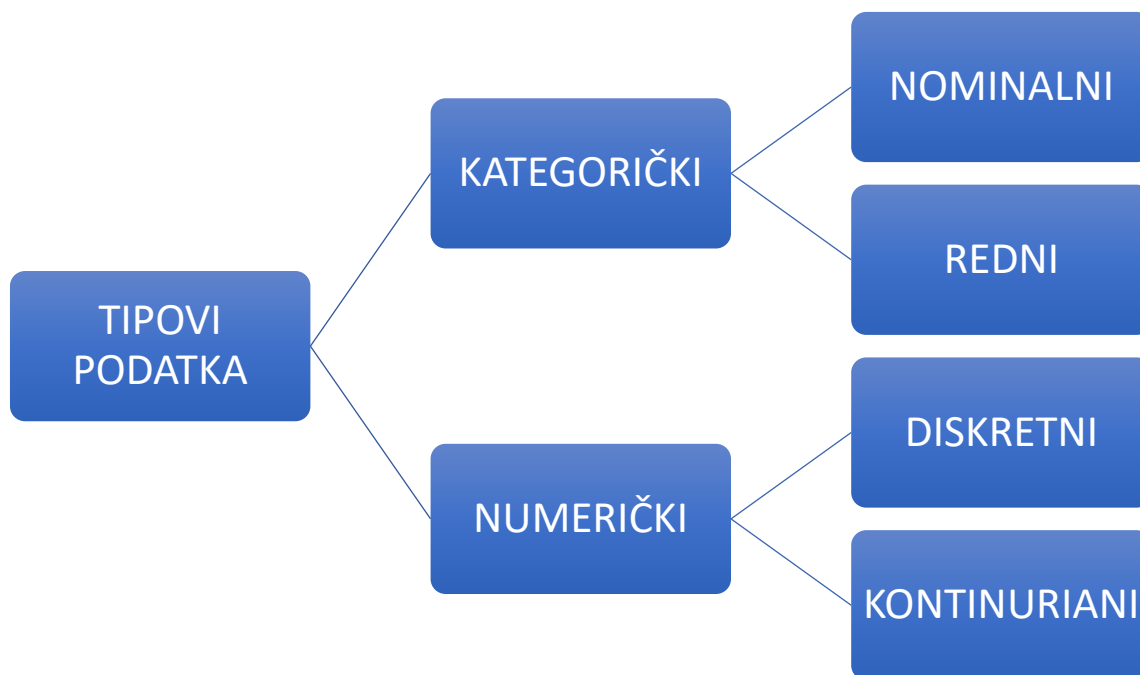
predstavljani kao neprekinuta funkcija koja ovisi o vremenu. Diskretni podatci su oni koji mogu poprimiti konačne vrijednosti, a njihovo pojavljivanje u skupu je naznačeno indeksom.

2. Kategoričke strukture podataka

Statička obilježja koja pripadaju nominalnoj ili ordinarnoj skali se nazivaju kategorička (kvalitativna) obilježja. Kod kategoričkih podataka, uzimaju se samo fiksni skupovi vrijednosti, primjerice vrste TV zaslona (plazma, LCD,..), ili naziv države (Albanija, Aljaska,..).

Binarni podatci su specifičan slučaj kategoričkih podataka jer poprimaju samo dvije vrijednosti kao što su 0 ili 1, da ili ne, istina ili laž.

Također u ovu skupinu spadaju i redni brojevi, vrlo korisna kategorijska vrsta koja nam obilježava raspored podataka.



Slika 2. Hijerarhijski prikaz strukture podataka.

Koncepti obuhvaćaju proces od zamišljanja problem, primjena tehnika znanosti podataka do postavljanje rezultata radi poboljšanja odlučivanje. Koncepti također podupiru širok spektar metoda poslovne analitike i tehnike.

Temeljni koncepti znanosti o podacima su:

1. Koncepti o tome kako podatkovna znanost odgovara organizaciji i konkurentnom okruženju, uključujući načine za privlačenje, strukturiranje i njegovanje timova za znanost o podacima, načini za razmišljanje o tome kako podatkovna znanost dovodi do konkurentske prednosti i taktički koncepti za dobar rad s projektima podatkovne znanosti.
2. Općeniti način analitičkog razmišljanja koji pomaže u identificiranju prikladnih podataka i razmatranje mogućih metoda. Koncept uključuje postupak rudarenja podataka kao i prikupljanje različitih zadataka na visokoj razini.
3. Opći pojmovi za stvarno odvajanje znanja iz podataka, koji podrazumijevaju ogroman niz zadataka iz znanosti o podacima i njihovih algoritama.

Temeljni koncept je utvrđivanje sličnosti dvaju entiteti opisanih podacima. Ova sposobnost čini osnovu za razne specifične zadatke. Može koristiti izravno za pronalaženje kupaca sličnih određenom kupcu. Čini jezgru nekolicine algoritama predviđanja koji procjenjuju ciljanu vrijednost, poput očekivane upotrebe resursa klijent ili vjerojatnost kupca da odgovori na ponudu. To je ujedno i osnova za tehnike klasteriranja koje grupiraju entitete prema zajedničkim značajkama bez fokusiranosti cilj. Sličnost je osnova za pronalaženje informacija, u kojoj se dokumenti ili podatci dohvaćaju se web stranice relevantne za upit za pretraživanje.

2.4 Razlika poslovne inteligencije i podatkovne znanosti

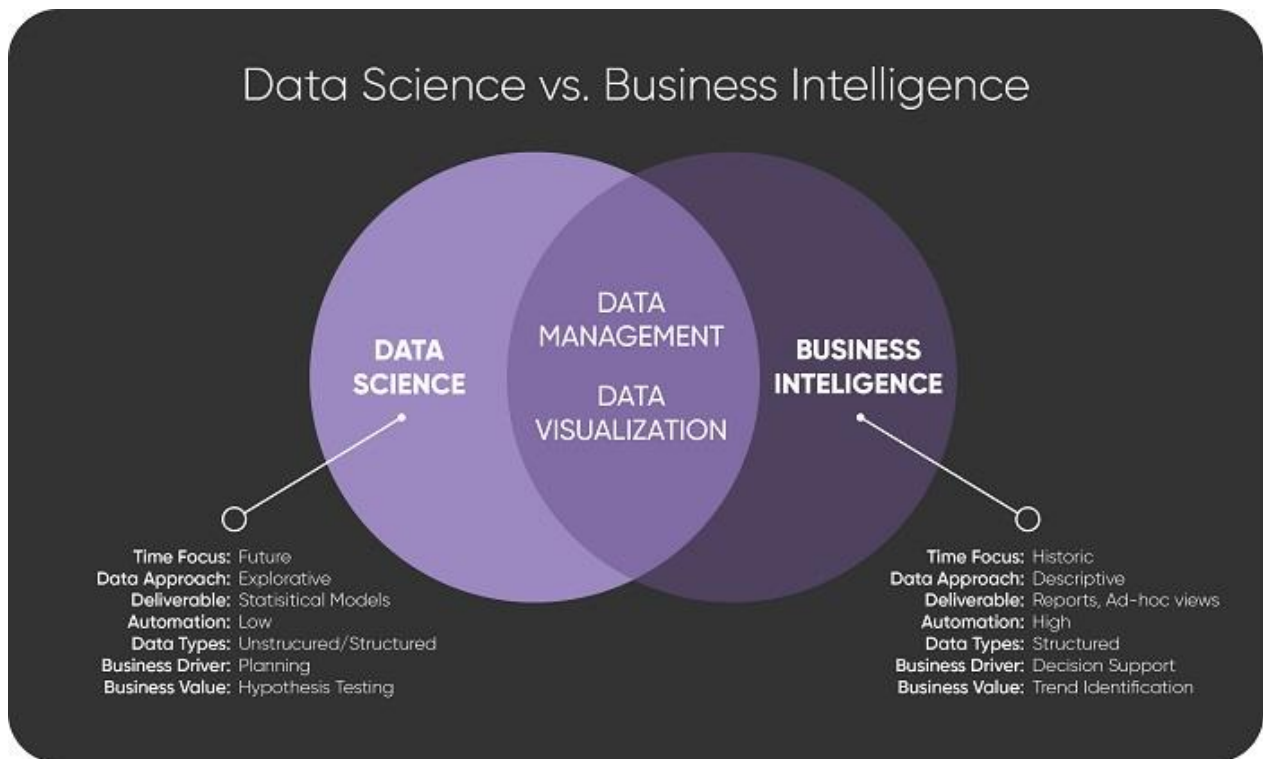
Poslovna inteligencija (*Bussines Intelligence*) obuhvaća strategije i tehnologije koje poduzeća koriste za analizu podataka poslovnih podataka. Tehnologije poslovne inteligencije pružaju povijesne, trenutne i prediktivne poglede na poslovanje. Uobičajene funkcije tehnologija poslovne

inteligencije uključuju izvješćivanje, mrežnu analitičku obradu, analitiku, rudarenje podacima, rudarstvo procesa, obradu složenih događaja, upravljanje poslovnim performansama, benčmarking, rudarenje tekstom, prediktivnu analitiku i analitiku na recept. Tehnologije mogu rukovati velikim količinama strukturiranih i ponekad nestrukturiranih podataka kako bi pomogle u identificiranju, razvoju ili stvaranju novih strateških poslovnih prilika. Cilj im je omogućiti lako tumačenje ovih velikih podataka. Prepoznavanje novih prilika i provedba učinkovite strategije zasnovane na uvidima mogu poduzećima pružiti konkurentsku prednost na tržištu i dugoročnu stabilnost.

Poslovna inteligencija se upotrebljava u poduzećima za potporu u donošenju širokog spektra poslovnih odluka, od operativnih do strateških. Osnovne operativne odluke uključuju pozicioniranje proizvoda ili cijene. Strateške poslovne odluke uključuju prioritete, ciljeve i upute na najširoj razini. U svim slučajevima, poslovna inteligencija je najefikasnija kada kombinira podatke izvedene s tržišta na kojem tvrtka posluje (vanjski podaci) s podacima iz unutrašnjih (internih) izvora poduzeća, kao što su financijski i operativni podaci (interni podaci). Kada se kombiniraju, vanjski i unutarnji podaci mogu pružiti cjelovitu sliku koja zapravo stvara „inteligenciju“ koja se ne može izvesti iz bilo kojeg pojedinačnog skupa podataka. Među bezbroj primjena, alati za poslovnu inteligenciju osnažuju organizacije da steknu uvid na nova tržišta, procijene potražnju i prikladnost proizvoda i usluga za različite segmente tržišta i procijene učinak marketinških napora.

Nasuprot tome, podatkovna znanost se koristi rascjepkanim podacima za predviđanje i istraživanje, a pri tome se fokusira na analiziranje sadašnjosti kako bi nam olakšala donošenje informiranih odluka oko budućnosti. Umjesto sakupljanja informacija koliko se određenog proizvoda prodalo u prethodnom kvartalu, koriste se tehnike poput vremenskih nizova koje omogućavaju predviđanje prodaje i prihoda.

I znanost podataka i poslovna inteligencija fokusiraju se na „podatke“, s ciljem pružanja povoljnih rezultata, što u slučaju poslovanja može biti profitna marža, zadržavanje kupaca, zauzimanje novog tržišta itd. Oba ova polja imaju sposobnost „tumačenja podataka“ i obično angažiraju tehničke stručnjake koji prevode ili transformiraju podatke obogaćene podacima u prijateljski uvid ili konkurentnu inteligenciju.



Slika 3. Usporedba podatkovne znanosti i poslovne inteligencije [preuzeto s <https://tdwi.org/-/media/TDWI/Upside/newsletters/2017/12/120517Pugsley.jpg>]

2.5 Koraci podatkovne znanosti

Ovisno o situaciji i podacima koje razmatramo imamo drugačiji pristup, ali generalno uvijek prolazimo kroz iste korake. Sami koraci nisu jasno određeni niti svaki korak ima svoj naziv ali u suštini možemo izdvojiti sljedeće korake:

- Istraživanje

Svaki proces podatkovne znanosti započinje s istraživanjem. Prije rješavanja svakog problema, prvo je potrebno ustanoviti šta je točno naš problem. Proučava se svaka domena, obuhvaćamo i povijesne činjenice te procjenjujemo koliki su dostupni resursi za projekt. Glavni cilj ove faze je proučavanje poslovnog problema kao analitičkog te formuliranje početne hipoteze za testiranje i početak usvajanja znanja o navedenom procesu.

- Priprema podataka

Nakon šta smo ustanovili koji je naš problem, potrebni su podaci koji će nam pružiti uvide potrebne za rješavanje problema. Podaci se istražuju, obrađuju te postavljaju u stanja prikladna za modeliranje i analizu. Priprema također obuhvaća pretvorbu podataka u tablične forme, zaključivanje podataka koji nedostaju itd. Kako bi što bolje mogli razumjeti podatke, oni se vizualiziraju. Vizualizacija pomaže u uočavanju trendova, devijacija i odnosa između varijabli. Ovaj korak je najzahtjevniji te prosječno oduzima 50% utrošenog vremena u procesu. Priprema podataka također je i najvažniji korak u cijelom procesu, jer bez kvalitetno obrađenih podataka ni sljedeći koraci ne mogu obaviti svoje zadaće.

- Izgradnja modela

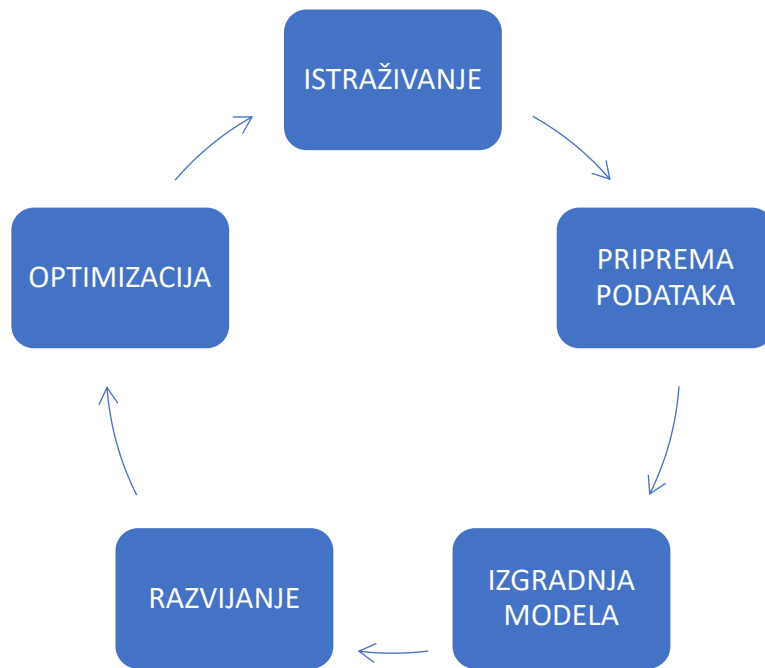
Važno je odabrati adekvatne metode i tehnike koje će se koristiti za razvoj modela. Istražuju se podaci kojima bi mogli saznati odnose između varijabli kako ni oabrali ključne varijable i najbolji model.

- Razvijanje

Dobivene rezultate uspoređujemo s kriterijima koji su identificirani kao uspjeh ili neuspjeh. U slučaju da je nas model kategoriziran kao neuspjeh, pokušavaju se identificirati razlozi i eventualne greške te se poboljšava model. U slučaju da nam je projekt kategoriziran kao uspješan, pohranjujemo rezultate koji se vizualiziraju klijentima.

- Optimizacija

Donose se konačni zaključci našeg projekta, informiranja i kodiranja te se pokušava optimizirati stvoreni model.



Slika 4. Prikaz koraka podatkovne znanosti

3. VELIKI PODATCI (*BIG DATA*)

Big data nam označava velike količine podataka s kojima je gotovo nemoguće raditi uz pomoć standardnih alata ili relacijskih baza podataka. U hrvatskom rječniku najčešće se upotrebljava engleski naziv, a u optičaju je i nekoliko hrvatskih prijevoda, primjerice: veliki podatci, golemi podatci, velika količina podataka te široki podatci. U nastavku ovog rada koristit ću engleski naziv Big Data.

3.1 Općenito o Big Data

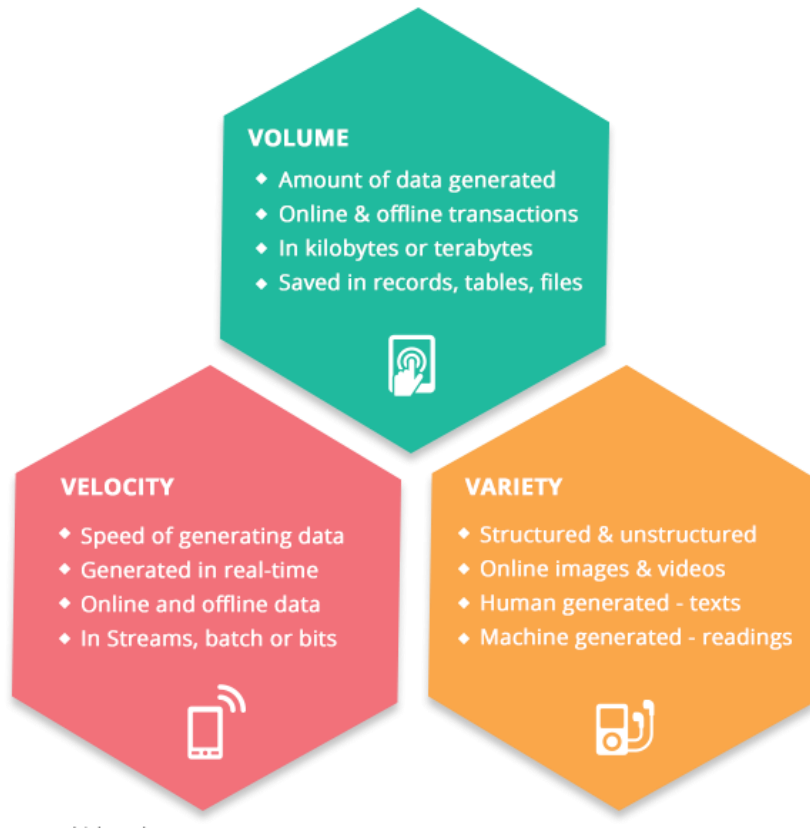
Eksponencijalnim rastom tehnologije tokom godina raste i količina samih podataka koje te tehnologije pohranjuju. Osim količine podataka, raste i njihovo vrijeme pristizanja, koje se veoma približava realnom vremenu. Navedene karakteristike podataka uzrok su otežanog spremanja, kao i analize i obrade podataka u stvarnom vremenu.

Big Data, koji je donedavno bio isključivo akademska tema, danas se koristi u konkretnim projektima. Mogućnost primjene proteže se na mnoga područja, od znanstvenih istraživanja s gomilom podataka, pa sve do politike, koja analizira ponašanja i javno dostupna mišljenja birača.

Pojam veliki podatak odnosi se na ogromne i složene setove podataka kod kojih uobičajne aplikacije za obradu podataka nisu primjenjive. Predstavlja tehnologiju za prikupljanje, obradu te analizu velikih količina podataka. Podaci u Big Data su raznoliki, razlikujemo više vrsta: strukturirani te nestrukturirani. Generiraju se i stižu velikom brzinom i u različitim intervalima te su upravo zbog ovih svojstava vrlo složeni za analizu i razmatranje. Također, podaci pristižu iz različitih uređaja i sustava koje koristimo (pametni telefoni, pametni satovi, pametni hladnjaci,..) koji prikupljaju podatke, šalju ih na analizu i obradu te ih vraćaju u oblik koji nam je razumljiviji i kojeg možemo upotrijebiti.

3.2 Karakteristike Big Data

Karakteristike Big Data najbolje se opisuju uz pomoć izraza 3V koji uključuje volumen, velicitet i varijantnost.

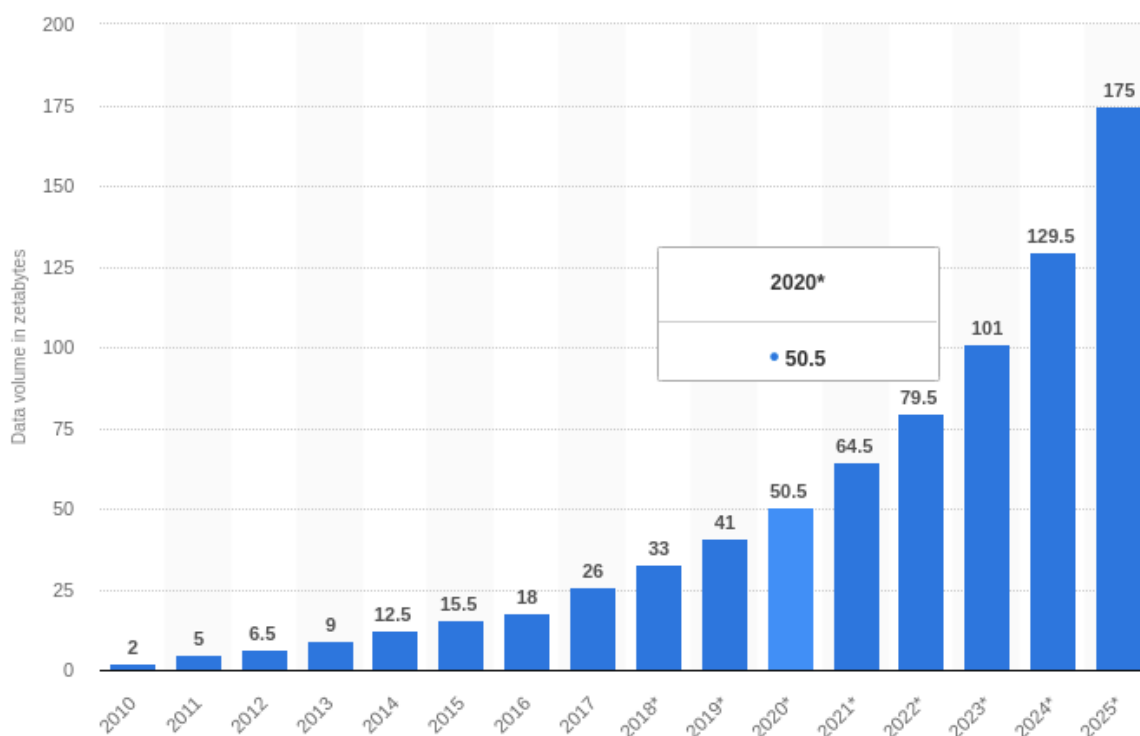


Slika 5. Karakteristike Big Data opisane pomoću 3V(volumen,varijantnost, velicitet) [preuzeto s <https://www.whishworks.com/blog/big-data/understanding-the-3-vs-of-big-data-volume-velocity-and-variety/>]

- Volumen (eng. *volume*)

Veliki podatci, u usporedbi s normalnim podacima, su znatno veći, tj sežu od par terabajta (TB) sve do par zeta bajta (ZB). Na primjer, unutar prostora društvenih medija, količina se odnosi na količinu podataka generiranih putem web stranica, portala te aplikacija. Ako uzmemo u obzir da Facebook ima 2 milijarde korisnika, YouTube 1 milijardu korisnika, Twitter 350 milijuna korisnika i Instagram 700 milijuna korisnika, njihovi korisnici

svakodnevno doprinose prikupljenim količinama podataka putem slika, postova, tweetova itd. Procjenjuje se da se svaki dan stvori 2,5 kvinti lionskih bajt podataka, što će rezultirati s 50 zeta bajt podataka stvorenih do 2020.godine (slika 5.). Ovakve podatke nije moguće pratiti i analizirati s jednim računalom te dolazi do razvoja paralelne računске arhitekture.



Slika 6. Prikaz volumena podataka u zeta bajtima kroz godine[preuzeto s <https://www.experfy.com/blog/big-data-and-blockchain-an-unbeatable-match/>]

- Velocitet (eng. *velocity*)

Velocitet se odnosi na brzinu kojom se generiraju podatci. Tradicionalni podatci su spori te se razvijaju u relativno statičnim uvjetima. Usporedno s tim, veliki podatci se uvijek "gibaju" i to velikom brzinom. Brzo nastaju te se brzo i obrađuju. Da bi se njihova učinkovitost optimizirala potrebno ih je obraditi u stvarnom vremenu. Uzmimo u obzir društvene mreže, na kojima se svakodnevno pojavi prosječno 900 milijuna fotografija na Facebook-u, 500 milijuna tweetova na Twitteru, 0,4 milijuna sati videozapisa prenesenih

na YouTube-u te 3,5 milijarde pretraživanja na Google-u. Big Data pomaže tvrtkama da održi ovu eksploziju podataka, prihvati podatke te ih istovremeno i obradi.

- Varijantnost (eng. *variety*)

Varijantnost se odnosi na raznolikost podataka. U prošlosti, podatci su se prikupljali s jednog mjesta i dostavljali u jednom formatu. Danas podatke nailazimo u raznolikim oblicima (videozapis, tekst, PDF,..) te stižu iz različitih mjesta. Strukturiranost podataka možemo podijeliti na strukturirani podatci (jasno definiran tip podatka i konzistentan format kao što je baza podataka) i nestrukturirani podatci (podatci koji se ne mogu lako pretraživati, uključuju formate kao što su audio, video i objave na društvenim mrežama). Strukturirani podatci se obično nalaze u relacijskim bazama podataka. Polja pohranjuju telefonske brojeve s brojem podatka, socijalnog osiguranja ili poštanske brojeve s definiranom duljinom. Podatci su definirani od strane čovjeka ili stroja. Ovaj format podatka je izuzetno dostupan za pretraživanje, ali uz prethodno znanje o vrijednostima koje se mogu pojaviti u tablicama.

Nestrukturirani podatci imaju unutarnju strukturu ali nisu strukturirani putem unaprijed definiranih modela podataka. Mogu biti generirani od strane čovjeka (tekstualni dokumenti, email, društvene mreže, internet stranice...) ili od strane stroja (satelitske slike za analiziranje vremena, znanstveni podatci, podatci sa senzora..).

Najveća razlika između dvije navedene kategorije je u jednostavnosti analize strukturiranih podataka u odnosu na nestrukturirane podatke.

Tablica 1. Usporedba strukturiranih i nestrukturiranih podataka

STRUKTURIRANI PODATCI	NESTRUKTURIRANI PODATCI
Moguće ih je prikazati u redovima, stupcima i relacijskim bazama podataka	Nije ih moguće prikazati u redovima, stupcima i relacijskim bazama podataka
Brojevi, datumi i nizovi	Slike, audio, video, tekstualni dokumenti, e-mailovi, proračunske tablice
Zauzima oko 20% ukupnih podataka	Zauzima oko 80% ukupnih podataka
Zahtjeva malo prostora za pohranu	Zahtjeva mnogo prostora za pohranu
Jednostavniji za obradu i zaštitu	Zahtjevniji za obradu i zaštitu

3.3 Primjena Big Data

Upisivanjem željenog pojma u Google tražilicu, automatski se dopuni pojam koji unosimo na temelju naših dosadašnjih preferencija ili pretraživanja. Pretraživanje knjiga putem online knjižare nailazimo i na preporuke knjige na osnovu već pročitanih. Kupovinom preko web stranica dobit ćemo prijedloge sličnih predmeta na bazi kupljenog predmeta. Skrolanjem Netflix-om nailazimo na filmove koji bi nam se dopali s obzirom na već pogledane filmove. Na sve gore navedene primjere i još mnoge druge, utjecaj je imao Big Data.

Sve naše aktivnosti na mreži - web stranice koje posjećujemo, postovi koji nam se sviđaju, stvari koje dijelimo, videozapise koje gledamo - sve se snima, prati te analizira. Ogroman volumen podataka dovodi do složenosti, a isto tako i do mnoštva prednosti. Sve industrije nastoje iskoristiti mogućnosti koje ovi podatci nude.

Praktične primjene velikih podataka kategorizirao sam u 5 skupina:

1. Praćenje lokacije

Primjer tvrtki koje koriste praćenje lokacije su logističke tvrtke. One koriste analitiku lokacije za praćenje i izvještavanje o narudžbama. Uporabom Big Data, moguće je pratiti stanje robe u prijenosu te prikupiti podatke o prometu i vremenskim uvjetima u stvarnom vremenu za definiranje rute prijevoza. Time ublažavaju rizike u prijenosu te poboljšavaju brzinu i pouzdanost u isporuci.

2. Medicina

Uz pomoć Big Data, bolnice mogu poboljšati nadzor nad pacijentima. 24-satni nadzor moguć je bez potrebe za izravnim nadzorom. Također učinkovitost lijekova se može poboljšati analizom prošlih zapisa o pacijentima i lijekovima koji su im pruženi. U slučaju biofarmaceutike, postoje mnoge varijable koje utječu na konačan proizvod. Primjerice, tokom proizvodnje inzulina, potrebno je intenzivno paziti kako bi se osigurao proizvod željene kvalitete.

3. Otkrivanje prevara

Bankarski i financijski sektori koriste Big Data za predviđanje i sprječavanje cyber kriminala, otkrivanje prijevara s karticom itd. Analiziranjem prošlih podataka svojih klijenata i podatke o prethodnim napadima, banke mogu predvidjeti buduće pokušaje. Big Data također pomaže i u rješavanju problema povezanih s propuštenim transakcijama i neuspjesima u mrežnom bankarstvu. Komisija za razmjenu vrijednosnih papira (SEC) koristi velike podatke za praćenje financijskih tržišta radi mogućih ilegalnih trgovina i sumnjivih aktivnosti.

4. Oglašavanje

Oglašavanje je jedna od najvećih korisnika Big Data. Bilo da se koristi Facebook, Google, Twitter, Instagram ili neki drugi mrežni div, svi prate ponašanje i transakcije korisnika. Internetski divovi oglašivačima pružaju mnogo podataka o korisnicima koji bi mogli voditi do određene kompanije. Primjerice na Facebooku možemo ciljati na ljude na temelju namjere kupnje, posjete web mjesta, interesa, uloge posla, demografskih čimbenika itd. Sve ove podatke Facebook algoritmi prikupljaju i analiziraju. Na temelju rezultata dobivenih analizom, preporuča vam web stranice, mjesta ili neke objekte, sve na temelju naših pretraživanja putem internetskih divova.

5. Zabava i medij

Na području zabave i medija, Big Data je usredotočen na ciljane ljude s pravim sadržajem u pravo vrijeme. Na temelju vaših prošlih pregleda i ponašanja na mreži prikazat će vam

različite preporuke. Ovu tehniku koriste Netflix i YouTube za povećanje angažmana i prihoda. U novije vrijeme, čak i televizijski programi žele segmentirati bazu podataka i u skladu s tim prikazivati različite oglase i emisije.



Slika 7. Big Dana u Netflix-u [preuzeto s: <https://www.muvi.com/blogs/deciphering-the-unstoppable-netflix-and-the-role-of-big-data.html>]

3.4 Distribuirana obrada podataka

Veliki podatci nadmašuju kapacitet obrade podataka koji su u mogućnosti konvencionalnih računala. Razlog toga su podatci koji dolaze velikom brzinom te im je volumen prevelik. Kako bi nam podatci pružali informaciju, potrebno ih je obraditi na alternativne načine. Područje prostiranja velikih podataka prikazano je na slici 6. koja je zadnji put dorađena 2019.godine, iz koje možemo dobiti dojam koliko široko zapravo i seže to područje.

Nutch. Sadržavao je i verziju MapReduce, koja predstavlja pojednostavljenju obradu podataka na velikim klasterima.

Osnovni Apache Hadoop okvir se sastoji od sljedećih modula:

- Hadoop Common – sadrži knjižice i uslužne programe koje su potrebne ostalim Hadoop modulima
- Hadoop distribuirani sustavi datoteka (HDFS) – pohranjuju podatke na strojevima, pružajući vrlo veliku ukupnu propusnost kroz klaster
- Hadoop YARN – uveden je 2012.godine, a ta platforma je odgovorna za upravljanje računalnim resursima u klasterima i njihovim korištenjem za raspoređivanje korisničkih aplikacija
- Hadoop Mapreduce – implementacija programskog modela MapReduce za veliku obradu podataka

4. UTJECAJ PODATKOVNE ZNANOSTI NA POSLOVANJE

Aplikacija velikih podataka postao je jedan od glavnih izazova u svim poslovnim i istraživačkim sferama. Pojavom Hadoopa, kao open-source tehnologije, promijenio se način razmišljanja o podacima, kao i njihovoj količini, načinu obrade te analizi. Podatci se obrađuju kako bi se otkrili skriveni obrasci korelacije te dobio uvid kako bi se donijele pravilne poslovne odluke. Tvrtke žele biti objektivnije i usmjerenije prema podacima zbog čega i prihvaćaju snagu podataka i tehnologije.

Koncept velikih podataka postoji već dugi niz godina. Desetljećima prije prvog spominjanja velikih podataka, postojala su poduzeća koja su primjenjivala analitiku nad prikupljenim podacima kako bi stekla uvid i otkrila trendove.

Danas, analitika velikih podataka se vrši pomoću naprednih softverskih sustava. Time je tvrtkama omogućeno smanjeno vrijeme analize za brže donošenje odluka.

4.1 Podatkovni znanstvenik

Pojam podatkovni znanstvenik, eng. *data scientist*, nije postojao prije 10-ak godina. Prvi put se pojavljuje 2018.godine od strane dvojice analitičara LinkedIn-a i Facebook-a. Danas *data scientist* obuhvaća veliki spektar poslova te je jedno od najtraženijih poslova današnjice. Područje rada pokriva raznovrsna područja primjerice obavještajni rad, rad u znanosti, rad u osiguravajućim kućama pa sve do rada u bankama. Slika 7. nam prikazuje porast zanimanja za ovaj tip posla.



Slika 9. Porast potražnje podatkovnih znanstvenika [preuzeto s:

<https://www.certace.com/blog/what-can-you-expect-from-a-data-expert/>]

Sam naziv "podatkovni znanstvenik" upućuje nas na to da je osoba znanstvenik. Znanstvenici vole istraživati, otkrivati i stvarati nove stvari. Riječ "podatkovni" nam implicira da se radi o zanimanju orijentiranom prema informatičkoj strani, ali ustvari ima poslovni fokus. Podatkovni znanstvenici služe da olakšaju razumijevanju tržišta, razmatraju potencijale i postojeće kupce usluga i proizvoda. Istraživajući ponašanje kupaca, u okvirima vlastitih usluga, proširujući istraživanje na širu populaciju, olakšavajući modele prediktivne i statističke analize. Glavni cilj podatkovnih znanstvenika je povećanje dobiti i vrijednosti poduzeća.

Jednostavnije rečeno, podatkovni znanstvenici bave se rješavanje problema unutar tvrtke koristeći podatke.

4.2 Učinak podatkovne znanosti

1. Strateški pristup upravljanja rizikom

Svaka tvrtka mora imati strateški pristup upravljanja rizikom. Veliki podatci omogućit će tvrtkama da kvantificiraju i modeliraju rizike koji se mogu suprotstavljati njihovom poslovanju. Korištenjem velikih podataka poslodavci se mogu osloniti na prediktivnu analitiku kako bi imali inteligentna predviđanja rizika. Uvidom u predviđanja lakše će se odrediti optimalna strategija za ublažavanje rizika.

2. Bolje poznavanje kupaca

Za napredovanje tvrtke, nužno je razumijevanje svojih kupaca, a veliki podatci pružaju nam pristup tim informacijama. Veliki podaci izvlače informacije o korisnicima na temelju posjeta web stranica te postova na društvenim mrežama. Također, omogućuju tvrtki izradu profila kupca kao osobe, kao i poznavanje njenih stavova i prohtjeva. Bolje poznavanje kupaca pridonosi lakšem poboljšanju svojih usluga i proizvoda prilagođavajući ih upravo njima. Oni također pomažu pri stvaranju lojalnosti prema kupcu.

3. Upoznavanje konkurencije

Pomoću velikih podataka dolazi do lakšeg upoznavanja konkurencije. Oni pružaju informacije o modelima određivanja cijena konkurenata te pomažu pri stvaranju percepcije kupaca. Veliki podaci prate ponašanje konkurencije na web-u, kao što je angažman na društvenim mrežama.

4. Personalizacija marketinga

Marketing je jedna od ključnih stavki koja pridonosi rastu ili pak padu poslovanja. Za njegov uspjeh najvažnija je personalizacija kojoj između ostalog pomažu veliki podaci. Isti se koriste i pri stvaranju marketinških kampanja koje zahvaljujući njima mogu biti usmjerena točno prema određenim ciljevima. Marketinški stručnjaci imaju uvid u kupce te stvaraju visoko konvertibilne materijale.

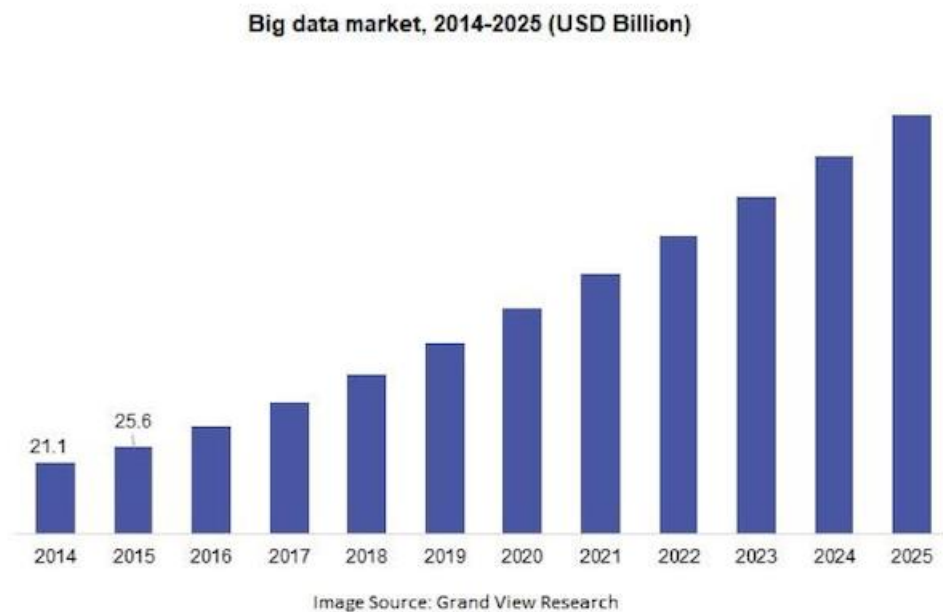
5. Prepoznavanje trendova

Pomoću velikih podataka, moguće je prepoznati vodeće trendove koji mogu biti korisni u istraživanju i razvoju proizvoda. *Big Data* pružaju potrebne informacije o ponašanju kupca pa sve do uzroka njihove kupnje. Tim podacima omogućena je analiza kako će se trendovi

mijenjati tijekom vremena, a zauzvrat će dati tvrtkama više vremena da se pripreme za nadolazeće promjene.

6. Upravljanje ljudskim resursima

Korištenjem velikih podataka možemo razviti robustan pristup upravljanju ljudskim resursima. Stvaraju pristup informacijama koje su potrebne za razumijevanje i poboljšanje radne snage. Primjerice posjeduju informacije o zadovoljstvu osoblja te podatak o njihovoj produktivnosti, a sve u svrhu stvaranja što boljeg poslovanja.



Slika 10. Učinak podatkovne znanosti na poslovanje [preuzeto s: <https://indatalabs.com/blog/impact-of-big-data-on-business/>]

4.3 Pozitivni i negativni učinci podatkovne znanosti

Podatkovna znanost imala je ogroman utjecaj na način na koji tehnologija utječe na naše živote. Neki utjecaji su bili dobri, a neki malo lošiji. Količina tehnologije koja utječe na naše živote nedvojbeno će se povećavati s vremenom i ne bismo trebali ograničavati svoje razumijevanje bez

svijesti o pozitivnim i negativnim posljedicama koje to može imati. U daljnjem radu predstaviti ćemo te dobre i loše strane podatkovne znanosti.

Pozitivni učinci:

- Astrologija

Nedavno smo imali napredak u astrofizici s prvom slikom crne rupe. To pomaže fizičarima da potvrde više od stoljeća čisto teorijskog rada oko crnih rupa i teorije relativnosti. Za snimanje ove slike, znanstvenici su koristili teleskop velik poput zemlje (Event Horizon Telescope ili EHT) kombinirajući podatke iz niza od osam zemaljskih radioteleskopa i sve to smislivši za izradu slike. Analiza podataka, a zatim vizualizacija tih podataka - zvuči kao neka znanost o podacima upravo ovdje.

- Biologija

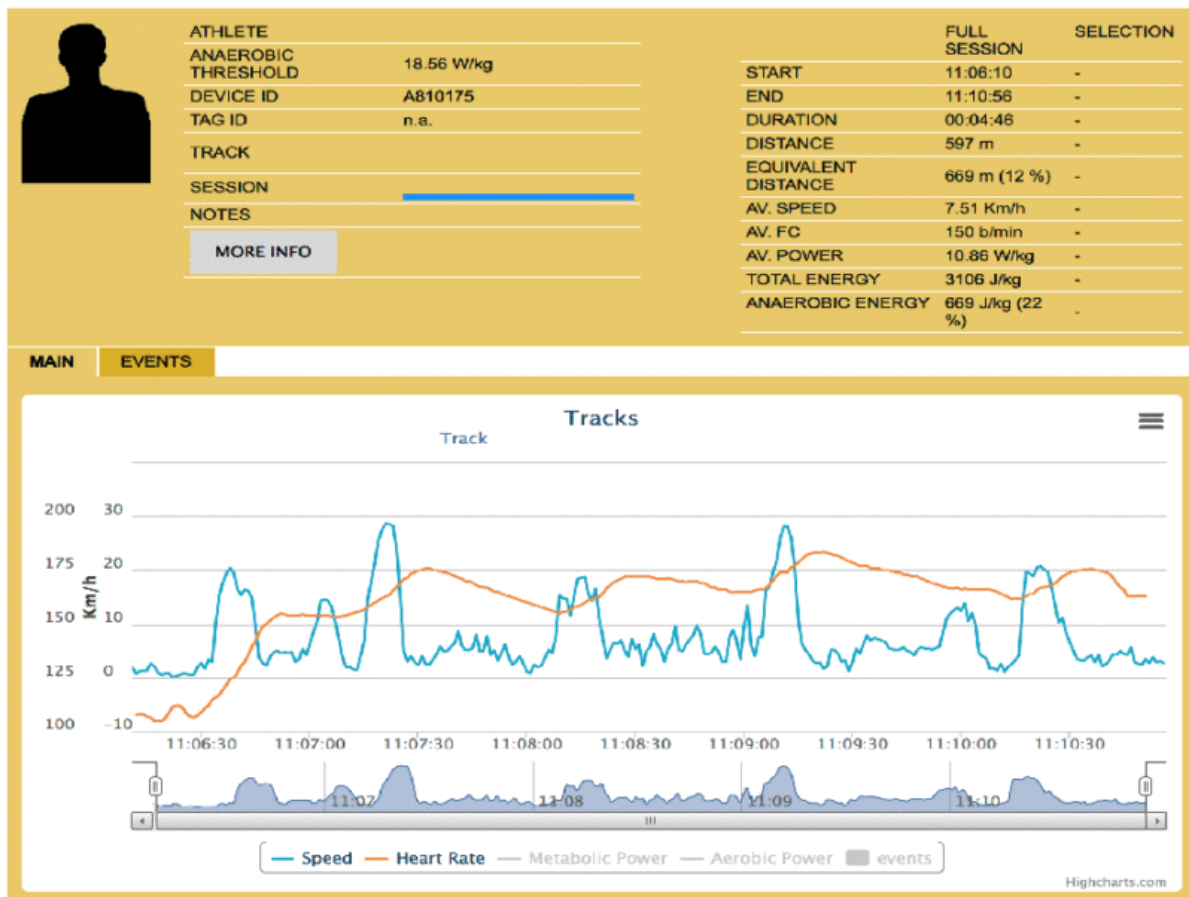
Olivier Elemento sa Sveučilišta Cornell koristi Big Data Analytics za identifikaciju mutacija u genomima koje rezultiraju širenjem tumorskih stanica kako bi se mogle ranije ubiti - ovo je veliki pozitivan utjecaj znanosti na ljudski život.

- Sport

Leicester City naslov u sezoni 2015-2016 osvojio je upravo uz pomoć podatkovne znanosti. Na početku sezone kladionice su imale vjerojatnost da će Leicester City osvojiti EPL s 10 puta manjom šansom od pronalaska čudovišta iz Loch Nessa. Svi su htjeli znati kako je Leicester to uspio, a pokazalo se da je znanost o podacima imala veliku ulogu! Zahvaljujući njihovom ulaganju u analitiku i tehnologiju, klub je mogao mjeriti razinu kondicije i tjelesno stanje igrača dok su trenirali kako bi spriječili ozljede, procjenjujući najbolje taktike koje se koriste u igri na temelju razine energije igrača. Svi treninzi imali su planove potkrijepljene stvarnim podacima o igračima, pa je kao rezultat toga Leicester City pretrpio najmanje ozljeda igrača od svih klubova te sezone.

Mnogi vrhunski timovi koriste analizu podataka kako bi pomogli u izvedbi igrača, izviđačkom talentu i razumijevanju načina planiranja za određene protivnike.

Upotreba analize podataka u nogometnim utakmicama nakon navedenog događaja naglo je porasla. Primjer programa i upotrebe analize podataka prikazan je na slici ispod (slika 10).



Slika 11. Primjer upotrebe analize podataka u nogometu [preuzeto s:

<https://www.gpexe.com/2014/03/25/football-analysis-software/>]

Negativni učinci:

- Ciljanje na temelju ranjivosti

Primjerice pristranosti u predviđanjima i klasifikacijama te utjecaj personalizacije i oglašavanja na društvo.

S velikim podacima dolaze i novi načini društvenog sortiranja sa sve većom preciznošću. Kombiniranjem više oblika skupova podataka može se puno naučiti. To se naziva "algoritamsko profiliranje" i izaziva zabrinutost zbog toga što malo ljudi zna o tome kako se njihovi podaci prikupljaju dok pretražuju, komuniciraju, kupuju, posjećuju web stranice, putuju itd. Veći dio ovog razvrstavanja prolazi ispod radara, iako su prakse posrednika u podacima privlačile pozornost. U svom svjedočenju Kongresu SAD -a, Pam Dixon sa Svjetskog foruma za privatnost izvijestila je da je pronašla posrednike u podacima koji

prodaju popise žrtava silovanja, adrese skloništa za obiteljsko nasilje, oboljele od genetskih bolesti, oboljele od ovisnosti i drugo.

- Zloupotreba osobnih podataka

Pojavila se zabrinutost zbog toga kako tvrtke za kreditne kartice koriste osobne podatke, primjerice gdje netko kupuje, te jesu li ili nisu platili savjetovanje o braku kako bi postavili stope i ograničenja. Jedna studija detaljno opisuje slučaj čovjeka kojem je kreditni rejting smanjen jer je American Express utvrdio da su drugi koji su kupovali tamo gdje je on kupovao imali lošu povijest otplate.

- Diskriminacija

Budući da korporacije, državna tijela i drugi koriste velike podatke, ključno je znati da se diskriminacija može i događati - i nenamjerno i namjerno. To se može dogoditi jer algoritamski vođeni sustavi ljudima različito nude, uskraćuju ili posreduju pristup uslugama ili mogućnostima.

Neki izazivaju zabrinutost zbog toga kako nove uporabe velikih podataka mogu negativno utjecati na sposobnosti ljudi da dobiju smještaj ili osiguranje - ili da pristupe obrazovanju ili zaposlenju. Istraživanje ProPublice i Consumer Reports iz 2017. pokazalo je da manjinska naselja plaćaju više za osiguranje automobila nego bijela naselja s istim razinama rizika. ProPublica također pokazuje kako su novi alati za predviđanje koji se koriste u sudnicama za izricanje kazni i obveznice "pristrani prema crncima". Drugi izražavaju zabrinutost zbog toga kako procesi velikih podataka olakšavaju ciljanje određenih skupina i diskriminiraju ih.

4.4 Primjena podatkovne znanosti u tvrtki Coca Cola

Coca Cola je najveća svjetska tvrtka za proizvodnju pića, s više od 500 marki bezalkoholnih pića koja se prodaju u više od 200 zemalja. Svakoga dana svijet popije više od 1,9 milijardi porcija njihovih napitaka, uključujući Colu, Diet Colu, Fanta, Sprite, Dasani, Powerade, Schweppes, Minute Maid i više.

Prema članku Forbesa, Coca Cola je bila jedna od prvih globalno priznatih marki, izvan tehnološkog sektora, koja je prihvatila *Big Data*. U 2012. godini, njegov glavni službenik za velike

podatke, Esat Sezer, rekao je: „Društveni mediji, mobilne aplikacije, računalstvo u oblaku i e-trgovina kombiniraju se kako bi tvrtkama poput Coca-Cole dali skup alata bez presedana koji će promijeniti način na koji pristupaju IT-u. Iza svega ovoga veliki podaci daju vam inteligenciju da sve to prekinete. ” S obzirom na veličinu svog poslovanja, Coca Cola generira znatnu količinu podataka u svom lancu vrijednosti - uključujući nabavu, proizvodnju, distribuciju, prodaju i povratne informacije kupaca. Tijekom godina, tvrtka je prihvatila Big Data kako bi donijela svoje poslovne strateške odluke.

Coca Cola ulaže velika sredstva u istraživanje i razvoj, osobito u umjetnu inteligenciju, kako bi bolje iskoristila brdo podataka koje prikuplja od kupaca diljem svijeta. Taj je napor pomogao tvrtki da bolje razumije trendove potrošača u smislu okusa i sklonost kupaca prema zdravijim opcijama u određenim regijama. Uz to, s obzirom na prisutnost Coca Cole u više od 200 zemalja s različitim trendovima kupaca i porastom inovativne robne marke, postalo je sve važnije za tvrtku za piće da razumije i prati razvoj ukusa svojih kupaca i uvede društvenu svijest u svoju ponudu proizvoda.

4.4.1 Razvoj proizvoda uz pomoć podatkovne znanosti

Coca Cola je 2008. godine predstavila novi aparat za piće u fontani, koji je korisnicima omogućio da sa svog pametnog telefona pripremaju pića, miješajući različite okuse. Pomoću pametnih telefona ljudi mogu naručiti točne postotke različitih smjesa i dodataka okusima te ih spremite za sljedeći put. Na temelju podataka praćenja prikupljenih iz fontana ovog samposlužnog bezalkoholnog pića, Coca Cola je lansirala Cherry Spirite kao novi okus.

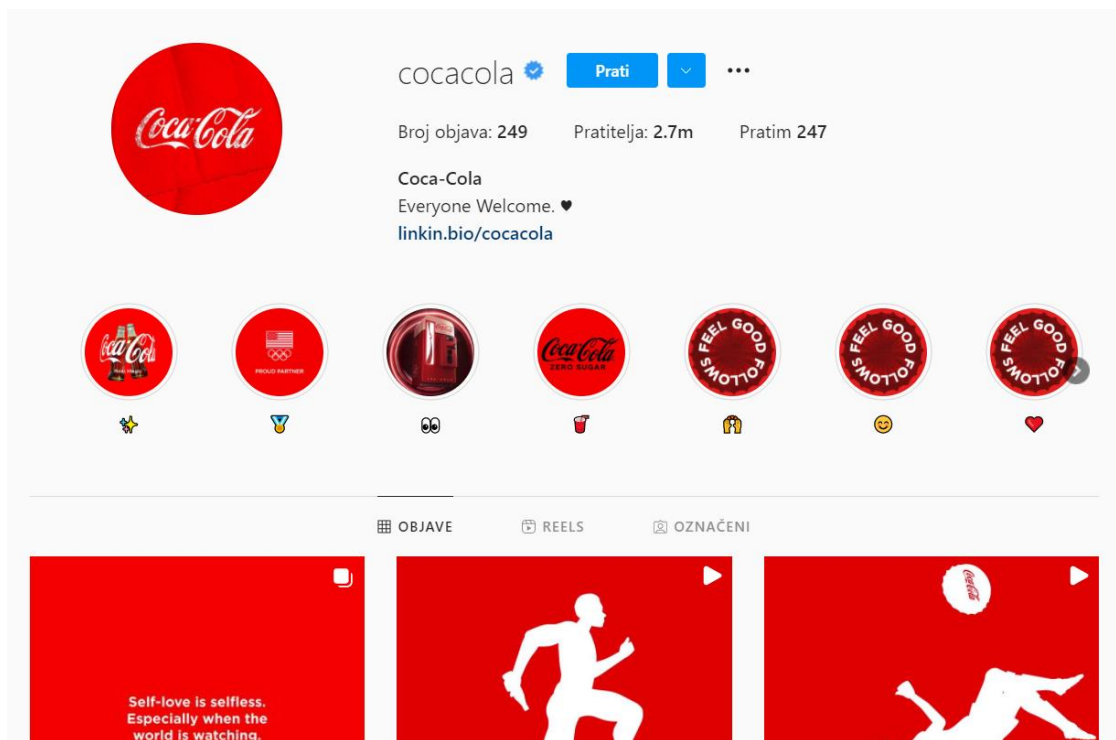


Slika 12. CocaCola freestyle uređaj za pripremu vlastitog pića miješajući okuse

4.4.2 Društvene mreže i oglašavanje

Coca Cola ima preko 100 milijuna sljedbenika na Facebooku i 35 milijuna na Twitteru. 2015. godine uspjeli su utvrditi da su se proizvodi Coca Cola na internetu spominjali jednom u dvije sekunde. Pristup tim podacima pomaže im da razumiju tko su im kupci, gdje žive i što ih potiče na raspravu o robnoj marki. Također bi mogli upotrijebiti umjetnu inteligenciju na slikama za identifikaciju kada su fotografije njihovih proizvoda učitane na društvene mreže, a zatim posluživati oglase ljudima na temelju slika koje su prenijeli. Prema riječima predstavnika tvrtke, četiri je puta veća vjerojatnost da će se kliknuti na takve ciljane oglase od ostalih metoda ciljanog oglašavanja.

Društveni mediji važni su jer vam omogućuju dosezanje, njegovanje i interakciju sa vašom ciljnom publikom - bez obzira na njihovu lokaciju. Kada poduzeće može koristiti društvene medije za povezivanje sa svojom publikom, može koristiti društvene medije za stvaranje svijesti o robnoj marki, potencijalnih kupaca, prodaje i prihoda. Marketingom korisnicima koji su već ljubitelji vaše tvrtke znate da dosežete kvalificiranu demografsku kategoriju. Činjenica da vas prate na Facebooku, Twitteru ili nekoj drugoj društvenoj platformi znači da već žele znati više.



Slika 13. Stranica Coca Cole na Instagramu

5. ZAKLJUČAK

Podatkovna znanost (eng. Data Science) je područje koje kombinira stručnost u domeni, vještine programiranja i znanje matematike i statistike kako bi izvukli smislene uvide iz mnoštva podataka. Sve veći broj tvrtki shvaća važnost podatkovne znanosti, kooperativno s umjetnom inteligencijom i strojnim učenjem. Bez obzira na industriju ili veličinu, ukoliko organizacije žele ostati konkurente na tržištu, moraju učinkovito razvijati i implementirati podatkovnu znanost. Samu važnost podatka ukazuje i to da je podatkovni znanstvenik jedan od najtraženijih poslova današnjice.

Uz pomoć podataka te analize istih, pri donošenju novih strateških odluka rizik od neuspjeha je znatno smanjen. Pruža nam uvid u kupce, njihove navike, želje i zadovoljstvo s proizvodima koje nudimo te nam u velikoj mjeri olakšava sljedeće poslovne poteze.

6. LITERATURA

- [1] S interneta: <https://www.ibm.com/cloud/learn/data-science-introduction>
- [2] S interneta: <https://cacm.acm.org/magazines/2013/12/169933-data-science-and-prediction/fulltext>
- [3] S interneta: <https://towardsdatascience.com/how-data-science-will-impact-future-of-businesses-7f11f5699c4d>
- [4] S interneta: <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>
- [5] S interneta: <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/?sh=72c8c07c55cf>
- [6] S interneta: <https://www.forbes.com/sites/greatspeculations/2019/08/02/understanding-coca-colas-business-model-and-performance-across-key-operating-markets/?sh=ba93cc618c89>
- [7] Tom Fawcett - Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking
- [8] S interneta: <https://www.forbes.com/sites/forbestechcouncil/2021/08/23/how-big-data-empowers-organizations-to-work-smarter-not-harder/?sh=21fc175a532f>
- [9] S interneta: <https://www.ibm.com/analytics/hadoop/big-data-analytics>