



1. RAZUMIJEVANJE I TUMAČENJE PODATAKA

Autor: Dario Šebalj

U ovom se poglavlju objašnjavaju temeljni koncepti razumijevanja i tumačenja podataka. Podaci su temelj učinkovitog donošenja odluka i imaju ključnu ulogu u ostvarivanju organizacijskog uspjeha. Stjecanjem vještina u analizi i tumačenju podataka, pojedinci (podatkovni analitičari, menadžeri, poslovni stručnjaci i sl.) mogu steći vještine potrebne za izvlačenje vrijednih dijelova iz ogromnog oceana dostupnih informacija. Podaci omogućuju analizu s činjeničnim temeljem. Organizacijama omogućuju donošenje odluka koje su objektivne i temeljene na činjenicama te da idu dalje od samih pretpostavki i intuicije. Poduzeća mogu koristiti podatke za uočavanje korelacija, trendova i obrazaca koje bi inače mogla propustiti.

Pronalaženje neučinkovitih procesa i područja za poboljšanje još je jedna od prednosti analize podataka. Analizom operativnih podataka, organizacije mogu pronaći uska grla, poboljšati radne tokove te opću učinkovitost poslovnih procesa. Dodatno, analiza podataka olakšava procjenu uspješnosti projekata ili provedenih strategija što ubrzava donošenje odluka i poboljšava planiranje.

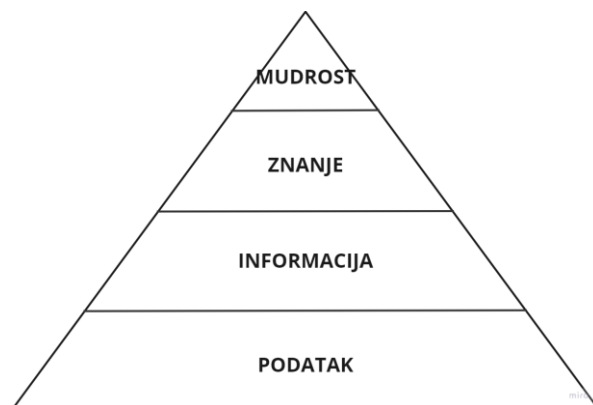
Prije postupka analize podataka, ključno je opisati različite vrste i izvore podataka, objasniti načine njihova modeliranja te naglasiti važnost kvalitete podataka. Ovo poglavlje će postaviti temelje poslovne inteligencije, pružajući korisnicima alate za iskorištavanje potencijala koji leži u podacima te donošenje poslovnih odluka.

1.1. Podatak, informacija, znanje, mudrost

Pri definiranju pojma „podatak“, kao polazište često se koristi poznata piramida podatak-informacija-znanje-mudrost (kratica DIKW, prema početnom slovu engleskim riječi) ili DIKW hijerarhija. Rowley (2007) navodi da ova hijerarhija služi u svrhu identificiranja i opisa procesa uključenih u transformaciju entiteta s niže razine u hijerarhiji (npr. podaci), u entitet na višoj razini u hijerarhiji (npr. informacije), kao i kontekstualiziranja podataka, informacija, znanja i



mudrosti u međusobnom odnosu. Svaka razina piramide nadograđuje se na razinu ispod nje, a da bi donošenje odluka na temelju podataka bilo učinkovito, potrebne su sve četiri razine (Cotton, 2023). Slika 1.1 prikazuje DIKW piramidu.



Slika 1.1 DIKW piramida

Izvor: Rowley (2007).

Podatak predstavlja sirovinu bez ikakvog značenja. To je sadržaj koji je izravno vidljiv ili provjerljiv, neorganizirana činjenica koje nema nikakav kontekst i koju je teško razumjeti (Brackett, 2015; Dalkir, 2023). Podaci mogu biti u obliku brojeva, teksta, slika itd. Bez tumačenja, podaci ostaju besmisleni. Primjer podataka: skup podataka koji sadrži očitavanja temperature prikupljena s meteoroloških stanica.

Skup podataka u kontekstu koji je relevantan za jednu ili više osoba u određenom trenutku ili određeno vrijeme naziva se **informacijom**. To su obrađeni, organizirani, strukturirani i kontekstualizirani podaci. Informacija daje odgovore na pitanja „tko”, „što”, „gdje” i „kada” (Brackett, 2015; Cotton, 2023). Primjer informacije: analizom temperaturnih podataka vidljivo je kako je prosječna temperatura u proteklom mjesecu viša nego u istom razdoblju prošle godine.

Cotton (2023) tvrdi da je **znanje** rezultat analize i tumačenja informacija, a koje otkriva obrasce, trendove i veze. Nudi uvid „kako” i „zašto” se određeni događaji događaju. To podrazumijeva bolje razumijevanje temeljnih ideja. Chaffey i Wood (2005, citirano u Rowley, 2007) definiraju znanje kao „kombinaciju podataka i informacija, kojima je dodano stručno mišljenje, vještine i iskustvo, da bi rezultiralo vrijednom imovinom koja se može koristiti za pomoć u donošenju odluka”. Primjer znanja: pomoću znanja stečenog analizom povijesnih podataka o temperaturi, meteorolog može predvidjeti vremenske uvjete za nadolazeći tjedan.



Sposobnost razumijevanja temeljnih činjenica i donošenje informiranih odluka i učinkovitih radnji poznata je kao **mudrost** (Cotton, 2023). Primjer mudrosti: koristeći vremenske prognoze i razumijevajući lokalne klimatske uvjete, poljoprivrednik može donijeti odluku o sadnji određene vrste usjeva.

Razumijevanje međusobne povezanosti podataka, informacija, znanja i mudrosti čini temeljnu osnovu za iskorištavanje potencijala poslovne inteligencije. Ovo će razumijevanje poslužiti kao osnova za transformaciju podataka u korisne zaključke koji mogu pomoći pri vođenju poduzeća prema uspjehu.

1.2. Izvori i vrste podataka

U modernom dobu, podaci se često nazivaju „novom naftom” – vrijednim resursom koji potiče inovacije i donošenje odluka. U središtu svakog pothvata koji se temelji nad podacima, nalazi se niz tipova podataka, svaki sa svojim jedinstvenim karakteristikama i značajem. Svaki dan se proizvede ogromna količina podataka. Trenutne projekcije govore da u svijetu postoji 97 zetabajta podataka i da se svaki dan stvori više od 2,5 kvintilijuna podataka. 90% podataka u svijetu generirano je u posljednje dvije godine (Marr, n.d.).

Prema Kenettu i Shmueliju (2016), „podaci mogu proizaći iz različitih instrumenata prikupljanja: anketa, laboratorijskih testova, terenskih eksperimenata, računalnih eksperimenata, simulacija, web pretraživanja, mobilnih snimaka itd. Podaci mogu biti primarni, prikupljeni u svrhu studija, ili sekundarni, prikupljeni iz nekog drugog razloga. Podaci mogu biti jednovarijantni ili multivarijantni, diskretni, kontinuirani ili mješoviti. Podaci mogu sadržavati semantičke nestrukturirane informacije u obliku teksta, slika, zvuka i videa. Mogu imati i različite strukture, uključujući vremenske serije, panel podatke, umrežene podatke, geografske podatke i sl. Podaci mogu uključivati informacije iz jednog izvora ili iz više njih. Mogu biti bilo koje veličine i bilo koje dimenzije”.

Podaci se generiraju brzo i kontinuirano. Društvene mreže, pametni uređaji, tehnologije snimanja u medicinskoj dijagnozi samo su neki od primjera. Novi podaci se stoga moraju negdje i spremati. Uređaji i senzori automatski stvaraju podatke koje je potrebno odmah analizirati i pohraniti. Održavanje ove ogromne količine podataka dovoljno je teško, ali njihova analiza s ciljem pronalaska trendova i vrijednih informacija daleko je teža, osobito kada podaci nisu u obliku tradicionalnih struktura podataka (EMC Education Services, 2015).



Prema Blazquezu i Domenechu (2018), tehnologije povezane s internetom, pametnim telefonima i senzorima sve su više uključene u većinu poslovnih i privatnih dnevnih operacija. Primjerice, mnoga poduzeća koriste društvene mreže za promicanje svojih brendova, nude svoje proizvode online, koriste pametne telefone za praćenje ruta prodajnih predstavnika ili koriste specijalizirane senzore za praćenje rada strojeva. S druge strane, ljudi koriste računala, pametne telefone i tablete kako bi na internetu pretraživali proizvode, komunicirali s prijateljima, dijelili mišljenja i snalazili se u prostoru. Osim toga, senzori postavljeni po gradovima, na autocestama i na javnim mjestima bilježe dnevna kretanja i aktivnosti građana. Zbog toga se, uporabom ovih tehnologija, proizvodi ogromna količina novih digitaliziranih i svježih podataka o aktivnostima ljudi i poduzeća. Ako se ispravno analiziraju, ovi podaci mogu pomoći u prepoznavanju trendova i praćenju ekonomskog, industrijskog i društvenog ponašanja.

Sherman (2015) naglašava da može biti problem ako poduzeće ima više podataka nego što to može podnijeti. Kroz svoje svakodnevne interakcije s klijentima, partnerima i dobavljačima, poduzeće prikuplja ogromne količine podataka, kako interno, tako i eksterno. Poduzeća provode i istraživanja tržišta i prate informacije o svojim konkurentima. Web strance im omogućuju praćenje točnog broja posjetitelja i mjesta odakle dolaze. Danas je vrlo aktualan internet stvari (IoT) koji prikuplja podatke sa senzora ugrađenih u predmete iz stvarnog svijeta poput ogrlice za pse, pacemakera i termostata. To je doslovno bujica podataka. Prema EMC Education Services (2015), među izvorima velikih podataka s najbržom stopom rasta su društveni mediji i genetsko sekvencioniranje, koji spadaju u netradicionalne izvore podataka u svrhu analize.

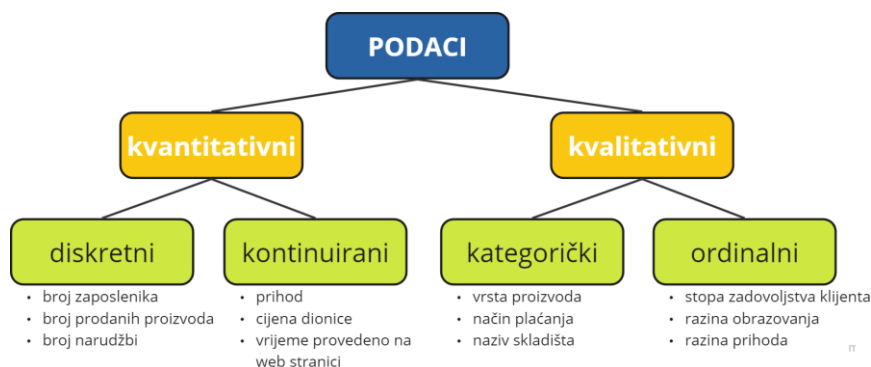
Howson (2014) smatra da uspjeh inicijative usmjerene prema poslovnoj inteligenciji (BI) ovisi o dostupnosti visokokvalitetnih i relevantnih podataka koji obuhvaćaju širok raspon izvora podataka potrebnih za proces poslovnog odlučivanja.

Podaci dolaze u različitim oblicima. Podaci se najčešće dijele na (slika 1.2):

1. **Kvantitativni (numerički) podaci** – podaci izraženi brojevima. Mogu se podijeliti na diskretne i kontinuirane podatke. Diskretni podaci mogu imati samo određenu vrijednost (npr. broj zaposlenih), dok kontinuirani podaci mogu imati beskonačan broj mogućih vrijednosti (npr. cijena proizvoda).
2. **Kvalitativni podaci** – ova vrsta podataka može se podijeliti na kategoričke i ordinalne podatke. Kategorički podaci predstavljaju tekstualne podatke koji se mogu grupirati u



različite kategorije (npr. mjesto rođenja, regija, kategorija proizvoda), dok se ordinalni podaci mogu rangirati ili poredati (npr. zadovoljstvo zaposlenika – vrlo nezadovoljan, nezadovoljan, neutralan, zadovoljan, vrlo zadovoljan; razina obrazovanja – osnovna škola, srednja škola, prvostupnik, magisterij, doktorat).



Slika 1.2 Osnovni tipovi podataka

Izvor: Autor.

Druga klasifikacija podataka je na strukturirane, polustrukturirane i nestrukturirane.

Strukturirani podaci mogu se pohranjivati, obrađivati i njima manipulirati u tradicionalnom relacijskom sustavu upravljanja bazama podataka. Ovi podaci dolaze iz raznih izvora, uključujući web obrasce, POS transakcije, senzore i sl. Mogu ih generirati ljudi ili strojevi. Ova vrsta podataka ima unaprijed određeni format, vrstu i organizaciju (EMC Education Services, 2015; Person i Porway, 2015).

Polustrukturirani podaci organizirani su pomoću oznaka (tagova) koji podacima daju hijerarhiju i poredak, čak i ako se ne uklapaju u sustav strukturirane baze podataka. Baze podataka i datotečni sustavi često sadrže polustrukturirane podatke. Logovi, HTML tekst, XML datoteke i JSON podatkovne datoteke primjeri su polustrukturiranih podataka (McKinsey Global Institute, 2011; Person i Porway, 2015).

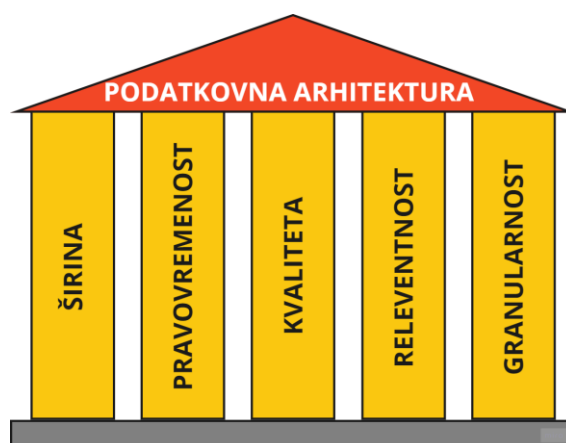
Budući da **nestrukturirane podatke** obično generira ljudska aktivnost i ne uklapaju se u format strukturirane baze podataka, smatraju se potpuno nestrukturiranima. Primjeri ove vrste podataka su tekstualni dokumenti, PDF datoteke, zapisi na blogovima, e-mailovi, slike i video (EMC Education Services, 2015; Person i Porway, 2015). Prema Shermanu (2015), s nestrukturiranim i polustrukturiranim podacima mora se postupati drugačije od tradicionalnih strukturiranih podataka.



Ta golema količina strukturiranih, a posebice nestrukturiranih podataka koji se generiraju, prikupljaju i obrađuju velikom brzom i složenošću, naziva se Big data (veliki podaci). McKinsey Global Institute (2011) definira Big data kao "podatke čija veličina, distribucija, raznolikost i/ili pravodobnost zahtijevaju upotrebu novih tehničkih arhitektura i analitike kako bi se došlo do rezultata koji stvaraju nove izvore poslovne vrijednosti". Prema Kitchenu i McArdleu (2016), Doug Laney je 2001. godine detaljno opisao da Big Data karakteriziraju tri značajke (tri V):

- **Volumen (eng. Volume)** - ogromne količine podataka,
- **Brzina (eng. Velocity)** - stvoreni u realnom vremenu,
- **Raznolikost (eng. Variety)** - strukturirani, polustrukturirani i nestrukturirani.

Prema Howsonu (2014), temelj za uspješnu poslovnu inteligenciju je podatkovna arhitektura (slika 1.3) koja se sastoji od šest važnih aspekata vezanih uz širinu, pravovremenost, kvalitetu, relevantnost i granularnost podataka. Kvaliteta podataka središnji je stup jer se ogroman trud ulaže u osiguranje i poboljšanje kvalitete podataka.



Slika 1.3 Arhitektura podataka kao temelj uspješne poslovne inteligencije

Izvor: Autor, prema Howson (2014).

Kao što je već spomenuto, podaci se mogu prikupljati iz različitih izvora. To je povezano sa širinom podataka kao jednim od stupova podatkovne arhitekture. Ono se odnosi na višestruke izvore podataka što je danas, u eri Big data, uobičajen način prikupljanja podataka. S druge strane, kombiniranje podataka iz više različitih izvora pridonosi problemima vezanim uz kvalitetu podataka (Howson, 2014).

U sljedećem potpoglavlju fokus će biti na iskorištavanju potencijala kojeg podaci pružaju. Istražit će se sustavi za upravljanje relacijskim bazama podataka (RDBMS) za pohranu i



dohvaćanje podataka te vizualni prikaz struktura baza podataka kroz dijagrame entiteta i odnosa (ER). Povezivanje razumijevanja izvora i tipova podataka s njihovim modeliranjem i oblikovanjem, značajan je korak prema praktičnoj primjereni poslovne inteligencije.

1.3. Modeliranje i oblikovanje podataka

Podatkovni model je formalni prikaz podataka koje poslovni sustav koristi i generira (Dennis et al., 2018). Kao što je već spomenuto, strukturirani podaci obično se pohranjuju u sustav za upravljanje relacijskim bazama podataka (eng. relational database management system - RDBMS). Prema Tilleyju (2020), „sustav za upravljanje bazom podataka zbirka je alata, značajki i sučelja koja korisnicima omogućuje dodavanje, ažuriranje, upravljanje, pristup i analizu podataka“. Neki od popularnih RDBMS sustava su Oracle (Oracle), DB2 (IBM) i SQL Server (Microsoft).

U RDBMS-u, podaci su organizirani u tablice koje sadrže zbirku zapisa koji pohranjuju informacije o određenom entitetu. Tablice su predstavljene kao dvodimenzionalne strukture sa stupcima i redovima. Svaki stupac predstavlja polje ili atribut entiteta, a svaki redak predstavlja zapis koji je instanca entiteta (Tilley, 2020). Slika 1.4 prikazuje primjer tablice Proizvod.

PROIZVOD			
ID	Naziv	KategorijaID	Cijena
1032	Laptop	1	800.00
1086	Majica	2	20.00
1099	Mobitel	1	600.00
2033	Kruh	3	2.00
2058	Tenisice	4	80.00
2069	Slušalice	1	40.00

Slika 1.4 Primjer tablice u RDBMS

Izvor: Autor.

Ova tablica predstavlja jednostavni katalog proizvoda sa 6 proizvoda, od čega svaki ima jedinstvenu šifru (ID), naziv, kategoriju i cijenu.

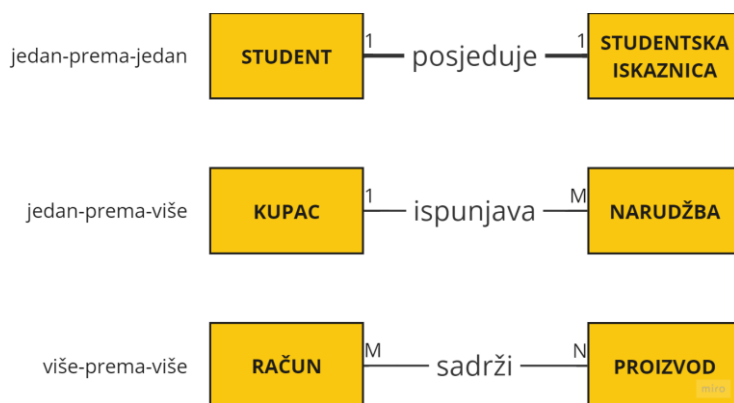


Atribut je posebna vrsta podatka o entitetu. Na primjer, ID kupca, ime, prezime, adresa, poštanski broj, grad, država, e-mail atributi su entiteta Kupac. Red u tablici ili skup povezanih polja koja opisuju jednu instancu entiteta (kao što je Kupac) naziva se **zapis**.

Svaka tablica u bazi podataka mora imati atribut koji služi kao **primarni ključ**. To je polje (ili skup polja) koje svakom proizvodu u tablici daje jedinstvenu vrijednost. To znači da u tablici ne mogu postojati dva proizvoda s istim ID-om.

Tablice u bazi podataka često su povezane s drugim tablicama, tj. između njih postoji veza. Veze (relacije) definiraju kako su podaci u jednoj tablici povezani s podacima u drugoj tablici.

Prema Tilleyju (2020), između entiteta mogu postojati tri vrste odnosa: jedan-prema-jedan, jedan-prema-više i više-prema-više. Primjeri ovih odnosa prikazani su na slici 1.5.



Slika 1.5 Primjeri odnosa među entitetima

Izvor: Autor.

Logička struktura baze podataka i odnosi među tablicama mogu se vizualno predstaviti pomoću **dijagrama entiteta i veza** (eng. Entity Relationship Diagram - ERD).

Postoje različite notacije za stvaranje ERD-ova, a najčešće su Chenova i Martinova (vranina noga). U ovoj će knjizi biti predstavljena Martinova notacija.

Prema Martinovoj (eng. Crow's Foot) notaciji, entitet je predstavljen pravokutnikom. To može biti osoba, mjesto, događaj ili stvar o kojoj se podaci prikupljaju. Atributi su navedeni kao imenice unutar entiteta. Odnosi između entiteta prikazani su linijama koje povezuju entitete. Odnosi imaju kardinalnost veze koja pokazuje koliko je instanci jednog entiteta povezano s instancom drugog entiteta (Dennis et al., 2018). U Crow's Foot notaciji, kardinalnosti su prikazane različitim simbolima. Na primjer, jedna crta označava jedan, dvostruka crta označava



jedan i samo jedan, krug označava nulu, a vrana noga označava više. Tablica 1.1 prikazuje različite simbole kardinalnosti i njihovo značenje.

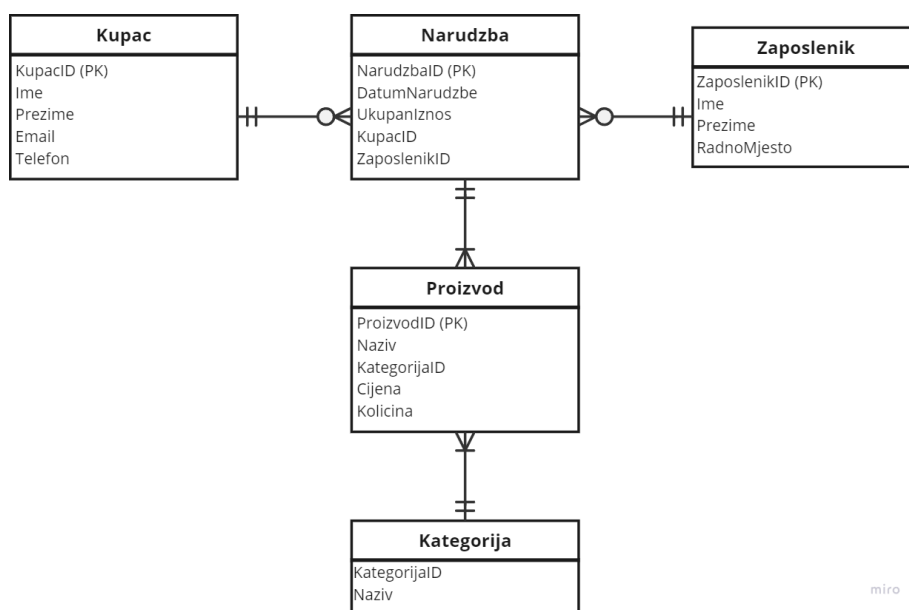
Tablica 1.1 Primjeri kardinalnosti

Simbol	Značenje
	Jedan i samo jedan
	Jedan ili više
	Nula ili više
	Nula ili jedan

Izvor: Tilley (2020).

Prema Dennis et al. (2018), tri su koraka pri izradi ERD-a:

1. Identifikacija atributa,
2. Dodavanje atributa i dodjela primarnog ključa,
3. Identifikacija veza.



Slika 1.6 Primjer ERD-a za sustav prodaje

Izvor: Autor.



Slika 1.6 prikazuje dio ERD-a za sustav prodaje. Svaki entitet u ovom ERD-u prikazan je kao pravokutnik s popisom svojih atributa. Linije koje povezuju entitete koriste se za prikaz njihovih odnosa:

- Kupac može napraviti više narudžbi. Neku narudžbu može napraviti samo jedan kupac.
- Određena narudžba može sadržavati više proizvoda. Svaki proizvod vezan je uz točno određenu narudžbu.
- Narudžba je povezana s kupcem i zaposlenikom koji ju je obradio.
- Proizvod je povezan s kategorijom jer svaki proizvod pripada određenoj kategoriji.

Temeljno razumijevanje modeliranja i dizajna podataka, kao što su ER dijagrami i RDBMS, bitno je za učinkovito korištenje podataka. Ono je temelj za pragmatičnu implementaciju podataka kroz donošenje odluka temeljenih na podacima, pri čemu strukturirani modeli podataka daju vrijedne uvide koji vode informiranim poslovnim strategijama i procesu donošenja odluka.

1.4. Odlučivanje na temelju podataka

Odlučivanje temeljeno na podacima (eng. Data-driven decision-making - DDDM) strateški je pristup koji se oslanja na analizu i tumačenje podataka s ciljem provedbe određenih radnji. Iskorištavanjem rezultata proizašlih iz golemih skupova podataka, tvrtke se mogu snalaziti u neizvjesnim situacijama, čime se smanjuju rizici i povećavaju prilike. Nelson (2022) definira donošenje odluka temeljeno na podacima kao proces donošenja strateških poslovnih odluka koje su u skladu s ciljevima i inicijativama organizacije korištenjem činjenica, metrika i podataka. Donošenje odluka temeljeno na podacima, prema Provostu i Fawcettu (2013), proces je donošenja odluka koje se više oslanjaju na analizu podataka nego na intuiciju. Na primjer, marketinški stručnjak može birati oglase koristeći svoje opsežno poznavanje industrije i istančan osjećaj za ono što će se svidjeti potrošačima. Međutim, svoju bi odluku mogao temeljiti i na analizi podataka koji pokazuju kako korisnici reagiraju na razne oglase.

U ovom pristupu odluke se ne donose na temelju intuicije, već na temelju čvrstih činjenica. Ono uključuje prikupljanje, analiziranje i tumačenje podataka kako bi se identificirali obrasci, trendovi i korelacije. Bilo da se radi o optimizaciji operativne učinkovitosti, poboljšanju korisničkih iskustava ili usavršavanju proizvodnih strategija, odluke temeljene na podacima omogućuju tvrtkama da se prilagode i napreduju na dinamičnim tržištima.



Integriranjem podataka u proces donošenja odluka, organizacije postaju prilagodljivije i otpornije na promjene, čime potiču inovacije i održivi rast. Velik broj istraživačkih radova pokazao je da je donošenje odluka temeljeno na podacima povezano s povećanom produktivnošću (npr. Brynjolfsson i McElheran, 2019; Sala et al., 2022; Colombari et al., 2023). To je razlog zašto većina organizacija, osobito velikih, ulaže u prikupljanje i analizu svojih podataka. Više od 2/3 od 300 rukovoditelja koje je ispitalo Bain & Company (2017) kaže da njihova tvrtka ulaže velika sredstva u analitiku podataka, dok više od polovice predviđa transformacijski povrat svojih ulaganja.

Pet je koraka za donošenje odluka temeljenih na podacima (Asana, 2022):

1. Razumijevanje vizije tvrtke,
2. Pronalaženje izvora podataka,
3. Čišćenje i organizacija podataka,
4. Analiza podataka,
5. Izvođenje zaključaka.

McKinsey Global Institute (2014) izvještava da organizacije koje koriste podatke u svrhu odlučivanja imaju 23 puta veću vjerojatnost da će steći kupce, 6 puta veću vjerojatnost da će zadržati kupce i 19 puta veću vjerojatnost da će biti profitabilne.

Globalno prepoznatljive tvrtke koje svoje odluke donose na temelju podataka su Google, Amazon i Netflix.

Kako bi organizacija ostvarila potpuni potencijal poslovne inteligencije, ključno je uzeti u obzir kvalitetu podataka, što je opisano u sljedećem dijelu. Kvaliteta podataka osigurava preciznost i pouzdanost zaključaka i uvida do kojih se došlo na temelju podataka.

1.5. Kvaliteta podataka

Stupanj točnosti, dosljednosti, pouzdanosti i prikladnosti za određenu svrhu naziva se kvalitetom podataka. Kvaliteta podataka važna je u kontekstu poslovne inteligencije i analize podataka budući da zaključci i prosudbe donesene na temelju podataka uglavnom ovise o njihovoj točnosti i pouzdanosti. Loša kvaliteta podataka može dovesti do netočnih zaključaka, pogrešnih strategija i naposljetku štetnih poslovnih ishoda. Prema Gartneru (2021), svake godine loša kvaliteta podataka košta organizacije u prosjeku 12,9 milijuna dolara.



Kvaliteta podataka može se okarakterizirati pomoću šest najčešće korištenih dimenzija (Foote, 2022):

- **Točnost:** koliko su točne vrijednosti atributa u podacima?
- **Potpunost:** jesu li podaci potpuni, bez nedostajućih informacija?
- **Dosljednost:** koliko su dosljedne vrijednosti u i između baza podataka?
- **Pravodobnost:** koliko su podaci pravovremeni?
- **Valjanost:** kako su podaci u skladu s unaprijed definiranim poslovnim pravilima?
- **Jedinstvenost:** je li svaki zapis jedinstveno identificiran, bez redundancija?

Prema Shermanu (2015), podaci se mogu smatrati visokokvalitetnim ako imaju sljedeće karakteristike (5C podataka):

- **Čisti** (eng. Clean) – odnosi se na stavke koje nedostaju, nevažeće unose i druge slične probleme
- **Dosljedni** (eng. Consistent) - jednoobraznost i koherentnost podataka u različitim izvorima i unutar samog skupa podataka
- **Sukladni** (eng. Conformed) – odnosi se na podatke koji se pridržavaju unaprijed definiranih standarda podataka i pravila
- **Aktualni** (eng. Current) - neophodno je koristiti najnovije podatke za donošenje odluka i analizu
- **Sveobuhvatni** (eng. Comprehensive) - uključuje sve bitne podatkovne elemente potrebne za namjeravani proces donošenja odluka bez izostavljanja kritičnih informacija.

Prema Kenettu i Shmueliju (2016.), gotovo svi podaci moraju se očistiti prije nego što se mogu koristiti za daljnju analizu. Međutim, cilj određuje stupanj čistoće i strategiju čišćenja podataka. Isti podaci mogu sadržavati visokokvalitetne informacije za jednu svrhu i nekvalitetne informacije za drugu.

Gartner (2023) je nedavno identificirao skup od 12 radnji usmjerenih na poboljšanje kvalitete podataka. Ove radnje su klasificirane u četiri različite kategorije, koje treba uzeti u obzir pri procjeni integriteta podataka:

- Usredotočite se na prave stvari u postavljanju jakih temelja,
- Primijenite odgovornost za kvalitetu podataka,
- Uspostavite kvalitetu podataka koja odgovara svrsi,
- Integrirajte kvalitetu podataka u korporativnu kulturu.



Prema Howsonu (2014), dosljedni, sveobuhvatni i točni podaci smatraju se visokim stupnjem kvalitete. Teško je dobiti kvalitetne podatke zbog velikog utjecaja organizacijskih i vlasničkih problemi.

Sposobnost razumijevanja i analize podataka ključna je za donošenje informiranih odluka u području poslovne inteligencije. Bitno je razumjeti da je put od neobrađenih podataka do dobivanja praktičnih rezultata iz njih, a koji obuhvaća različite izvore podataka, vrste, modeliranje i dizajn, sastavni dio procesa izvlačenja vrijednosti iz informacija.

REFERENCE

1. Asana (2022). Data-driven decision making: A step-by-step guide [dostupno na: <https://asana.com/resources/data-driven-decision-making>, pristupljeno November 5, 2023]
2. Bain & Company (2017). Closing the Results Gap in Advanced Analytics: Lessons from the Front Lines [dostupno na: <https://www.bain.com/insights/closing-the-results-gap-in-advanced-analytics-lessons-from-the-front-lines/>, pristupljeno November 5, 2023]
3. Blazquez, D. & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. Technological Forecasting & Social Change, 130, pp. 99-113.
4. Brackett, M. (2015). The Data-Information-Knowledge Cycle. Dataversity [dostupno na: <https://www.dataversity.net/the-data-information-knowledge-cycle/>, pristupljeno November 5, 2023]
5. Brynjolfsson, E. & McElheran, K. (2019). Data in Action: Data-Driven Decision Making and Predictive Analytics in U.S. Manufacturing. Rotman School of Management Working Paper No. 3422397 [dostupno na: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3422397, pristupljeno November 5, 2023]
6. Chaffey, D. & Wood, S. (2005). Business Information Management: Improving Performance using Information Systems. FT Prentice Hall.
7. Colombari, R., Geuna, A., Helper, S., Martins, R., Paolucci, E., Ricci, R. & Seamans, R. (2023). International Journal of Production Economics, 255.



8. Cotton, R. (2023). The Data-Information-Knowledge-Wisdom Pyramid. Datacamp [dostupno na: <https://www.datacamp.com/cheat-sheet/the-data-information-knowledge-wisdom-pyramid>, pristupljeno November 5, 2023]
9. Dalkir, K. (2023). Knowledge Management in Theory and Practice, 4th Edition. MIT Press.
10. Dennis, A., Wixom, B. H. & Roth, R. M. (2018). Systems Analysis and Design, 7th Edition. Wiley.
11. EMC Education Services (2015). Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. Wiley.
12. Foote, K. D. (2022). Data quality dimensions. Dataversity [dostupno na: <https://www.dataversity.net/data-quality-dimensions/>, pristupljeno November 5, 2023]
13. Gartner (2021). How to Improve Your Data Quality [dostupno na: <https://www.gartner.com/smarterwithgartner/how-to-improve-your-data-quality>, pristupljeno November 5, 2023]
14. Gartner (2023). Gartner Identifies 12 Actions to Improve Data Quality [dostupno na: <https://www.gartner.com/en/newsroom/press-releases/2023-05-22-gartner-identifies-12-actions-to-improve-data-quality>, pristupljeno November 5, 2023]
15. Howson, C. (2014). Successful Business Intelligence: Unlock the Value of BI & Big Data, 2nd Edition. McGraw-Hill Education.
16. Kenett, R. S. & Shmueli, G. (2016). Information Quality: The Potential of Data and Analytics to Generate Knowledge. Wiley.
17. Kitchen, R. & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society, 3(1).
18. Marr, B. (n.d.). How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. Bernard Marr & Co. [dostupno na: <https://bernardmarr.com/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/>, pristupljeno November 5, 2023]
19. McKinsey Global Institute (2011). Big data: The next frontier for innovation, competition, and productivity. [dostupno na: https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/big%20data%20the%20next%20frontier%20for%20innovation/mgi_big_data_full_report.pdf, pristupljeno November 5, 2023]



20. McKinsey Global Institute (2014). Five facts: How customer analytics boosts corporate performance [dostupno na: <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/five-facts-how-customer-analytics-boosts-corporate-performance>, pristupljeno November 5, 2023]
21. Nelson, M. (2022). Beyond The Buzzword: What Does Data-Driven Decision-Making Really Mean?. Forbes [dostupno na: <https://www.forbes.com/sites/tableau/2022/09/23/beyond-the-buzzword-what-does-data-driven-decision-making-really-mean/?sh=2d35c4eb25d6>, pristupljeno November 5, 2023]
22. Provost, F. & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1), pp. 51-59.
23. Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), pp. 163–180.
24. Sala, R., Pirola, F., Pezzotta, G. & Cavalieri, S. (2022). Data-Driven Decision Making in Maintenance Service Delivery Process: A Case Study. *Applied Sciences*, 12(15).
25. Sherman, R. (2015). *Business Intelligence Guidebook: From Data Integration to Analytics*. Elsevier Inc.
26. Tilley, S. (2020). *Systems Analysis and Design*, 12th Edition. Cengage.