



1. RAZUMEVANJE IN INTERPRETACIJA PODATKOV.

Avtor: Dario Šebalj

V tem poglavju so pojasnjeni temeljni koncepti razumevanja in razlage podatkov. Podatki so temelj učinkovitega odločanja in imajo ključno vlogo pri doseganju organizacijskega uspeha. S pridobivanjem znanja na področju analize in razlage podatkov lahko posamezniki (podatkovni analitiki, vodje, poslovni strokovnjaki, navdušenci nad podatki itd.) pridobijo veščine, potrebne za pridobivanje dragocenih spoznanj iz velikega oceana razpoložljivih informacij. Podatki analizi zagotavljajo dejansko podlago. Organizacijam omogočajo sprejemanje bolj objektivnih in na dejstvih utemeljenih odločitev, saj jim omogočajo, da presežejo domneve in intuicijo. Organizacije lahko s pomočjo podatkov najdejo korelacije, trende in vzorce, ki bi jih sicer morda spregledale.

Druga prednost analize podatkov je iskanje organizacijskih neučinkovitosti in področij za izboljšave. Organizacije lahko z ocenjevanjem operativnih podatkov najdejo ozka grla, skrajšajo delovne postopke in izboljšajo splošno učinkovitost poslovnih procesov. Poleg tega analiza podatkov olajša ocenjevanje uspešnosti projektov ali strategij, ki so se začeli izvajati, kar pospeši sprejemanje odločitev in izboljša načrtovanje.

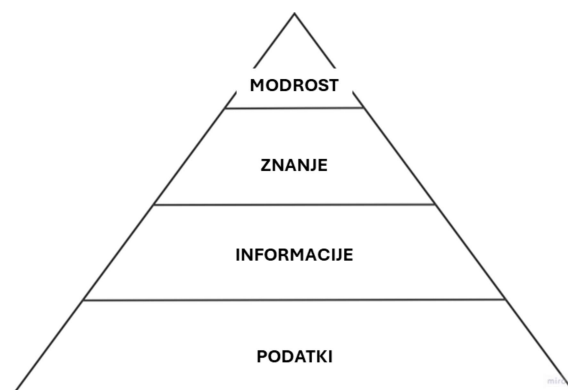
Pred postopkom analize podatkov je treba opisati različne vrste in vire podatkov, razpravljati o modeliranju podatkov in poudariti pomen kakovosti podatkov. To poglavje bo postavilo trdne temelje poslovnega obveščanja in uporabnikom zagotovilo orodja za sprostitve moči podatkov in sprejemanje informiranih odločitev.

1.1. Podatki, informacije, znanje, modrost

Pri opredelitvi pojma podatki se kot izhodišče pogosto uporablja znana piramida podatkov, informacij, znanja in modrosti (DIKW) ali hierarhija DIKW. Rowley (2007) navaja, da hierarhija služi opredelitvi in opisu procesov, ki so vključeni v preoblikovanje entitete na nižji ravni v hierarhiji (kot so podatki) v entiteto na višji ravni v hierarhiji (kot so informacije), ter kontekstualizaciji podatkov, informacij, znanja in občasno modrosti v medsebojnem razmerju.



Vsaka raven piramide temelji na ravni pod njo, za učinkovito odločanje na podlagi podatkov pa so potrebne vse štiri ravni (Cotton, 2023). Slika 1.1 prikazuje piramido DIKW.



Slika 1.1 Piramida DIKW (Data, Information, Knowledge, Wisdom)

Vir: Rowley (2007).

Podatki so surovina, ki nima pomena. Gre za vsebino, ki jo je mogoče neposredno opazovati ali preverjati, za neorganizirano dejstvo, ki je izvzeto iz konteksta in ga je težko razumeti (Brackett, 2015; Dalkir, 2023). Podatki so lahko v obliki števil, besedila, slik itd. Brez razlage bodo podatki ostali brez pomena. Primer podatkov: nabor podatkov, ki vsebuje odčitke temperature, zbrane na vremenskih postajah.

Zbirka podatkov v kontekstu, ki so pomembni za eno ali več oseb ob določenem času ali za določen čas, se imenuje **informacija**. Gre za obdelane, organizirane, strukturirane in kontekstualizirane podatke. Odgovore na pogosta vprašanja, kot so "kdo", "kaj", "kje" in "kdaj", je mogoče najti s pomočjo informacij. (Brackett, 2015; Cotton, 2023). Primer informacij: Z analizo podatkov o temperaturi lahko ugotovimo, da je povprečna temperatura v zadnjem mesecu višja kot v enakem obdobju lani.

Cotton (2023) trdi, da je **znanje** rezultat analize in interpretacije informacij, ki razkriva vzorce, trende in povezave. Ponuja vpogled v to, "kako" in "zakaj" se določeni dogodki dogajajo. Vključuje boljše razumevanje temeljnih idej. Chaffey in Wood (2005, povzeto po Rowley, 2007) opredeljujeta znanje kot "kombinacijo podatkov in informacij, ki jim je dodano strokovno mnenje, veščine in izkušnje, da nastane dragoceno sredstvo, ki se lahko uporabi za pomoč pri odločanju". Primer znanja: Meteorolog lahko z znanjem, pridobljenim z analizo preteklih podatkov o temperaturi, predvidi vremenske razmere za prihodnji teden.

Sposobnost razumevanja dejstev, na katerih temeljijo, ter sprejemanja dobro informiranih odločitev in učinkovitih ukrepov je znana kot **modrost** (Cotton, 2023). Primer modrosti: Na



podlagi vremenskih napovedi in razumevanja lokalnih podnebnih razmer lahko kmet sprejme odločitve o sajenju določene vrste pridelka.

Razumevanje medsebojnega vpliva med podatki, informacijami, znanjem in modrostjo je temeljna podlaga za izkoriščanje potenciala poslovne inteligence. V nadaljevanju te knjige bo to razumevanje služilo kot osnova, na kateri je mogoče izkoristiti moč podatkov in jih preoblikovati v uporabna spoznanja, kar bo organizacije vodilo k uspehu v nenehno razvijajočem se okolju informacijske dobe.

1.2. Viri podatkov in vrste podatkov

V sodobnem času se podatki pogosto imenujejo "nova nafta" - dragocen vir, ki spodbuja inovacije in odločanje. V središču vsakega prizadevanja, ki temelji na podatkih, je spekter vrst podatkov, od katerih ima vsak svoje edinstvene značilnosti in pomen. Vsak dan nastane ogromna količina podatkov. Po trenutnih napovedih je na svetu 97 zettabajtov podatkov, vsak dan pa se ustvari več kot 2,5 kvintilijonov bajtov podatkov. 90 % podatkov na svetu je bilo ustvarjenih v zadnjih dveh letih (Marr, n.d.).

Kenett in Shmueli (2016) pravita, da "podatki lahko izhajajo iz različnih instrumentov zbiranja: anket, laboratorijskih testov, terenskih poskusov, računalniških poskusov, simulacij, spletnih iskanj, mobilnih posnetkov itd. Podatki so lahko primarni, zbrani za namen študije, ali sekundarni, zbrani iz drugega razloga. Podatki so lahko univariatni ali multivariatni, diskretni, zvezni ali mešani. Podatki lahko vsebujejo semantične nestrukturirane informacije v obliki besedila, slik, zvoka in videa. Podatki imajo lahko različne strukture, vključno s presečnimi podatki, časovnimi vrstami, panelnimi podatki, omrežnimi podatki, geografskimi podatki itd. Podatki lahko vključujejo informacije iz enega samega vira ali iz več virov. Podatki so lahko poljubne velikosti in razsežnosti.

Podatki se ustvarjajo hitreje in neprekinjeno. Vsi ti podatki in še več ustvarjajo nove podatke, ki jih je iz nekega razloga treba nekje hraniti. Družabna omrežja, pametni telefoni in slikovna tehnologija, ki se uporablja v medicinski diagnostiki, so le nekateri primeri. Naprave in senzorji samodejno ustvarjajo diagnostične podatke, ki jih je treba takoj analizirati in shraniti. Ohranjanje te ogromne količine podatkov je dovolj težavno, vendar je njihovo analiziranje za prepoznavanje trendov in pridobivanje dragocenih informacij še veliko težje, zlasti če podatki ne ustrezajo tradicionalnim predstavam o strukturi podatkov (EMC Education Services, 2015).



Blazquez in Domenech (2018) pravita, da se tehnologije, povezane z internetom, pametnimi telefoni in pametnimi senzorji, vse bolj vključujejo v večino poslovnih in osebnih vsakodnevnih dejavnosti. Veliko podjetij na primer uporablja družbene medije za promocijo svojih blagovnih znamk, ponuja izdelke prek spleta, s pametnimi telefoni spremlja poti, ki jih opravijo prodajalci, ali uporablja specializirane senzorje za beleženje delovanja strojev. Po drugi strani pa ljudje uporabljajo računalnike, pametne telefone in tablične računalnike za brskanje po spletu za izdelke, komuniciranje s prijatelji, izmenjavo mnenj in iskanje poti. Poleg tega senzorji, nameščeni po mestih, na avtocestah in na javnih površinah, kot so supermarketi, beležijo vsakodnevno gibanje in dejavnosti državljanov. Zaradi tega vse te tehnologije ustvarjajo ogromno količino na novo digitaliziranih in svežih podatkov o dejavnostih ljudi in podjetij. Če so ti podatki ustrezno analizirani, lahko pomagajo prepoznati trende in slediti gospodarskemu, industrijskemu in družbenemu obnašanju.

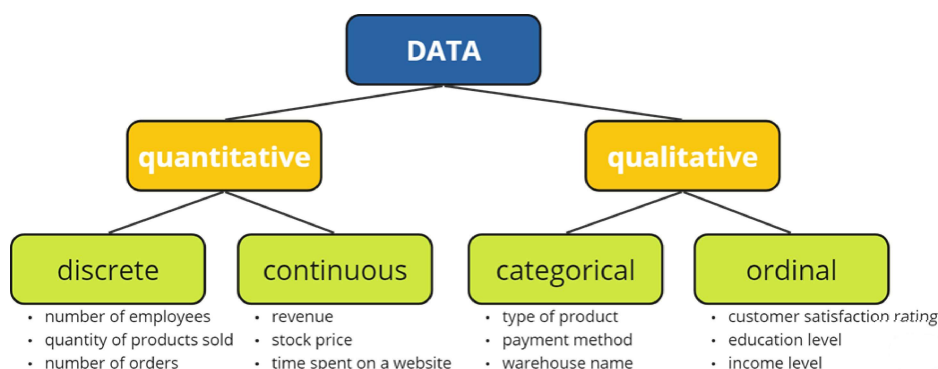
Sherman (2015) poudarja, da je lahko težava, če ima podjetje več podatkov, kot jih lahko upravlja. Z vsakodnevnimi stiki s strankami, partnerji in dobavitelji zbirajo ogromne količine notranjih in zunanjih podatkov. Izvajajo tržne raziskave in spremljajo informacije o svojih tekmecih. Spletna mesta s kodami za sledenje jim omogočajo, da sledijo natančnemu številu obiskovalcev in njihovem izvoru. Hranijo in upravljajo podatke, potrebne za industrijske pobude in vladne zahteve. Danes obstaja internet stvari (IoT), ki zbira podatke iz senzorjev, vgrajenih v predmete iz resničnega sveta, kot so pasje ovratnice, srčni spodbujevalniki in termostati. Gre za poplavo podatkov. Po podatkih podjetja EMC Education Services (2015) sta med viri velikih podatkov z najhitrejšo stopnjo rasti družbeni mediji in genetsko sekvenciranje, ki sta primera netradicionalnih virov podatkov, ki se uporabljajo za analizo.

Howson (2014) meni, da je uspeh pobude za poslovno obveščanje (BI) odvisen od razpoložljivosti visokokakovostnih in ustreznih podatkov, ki vključujejo širok nabor podatkovnih virov, potrebnih za informiranje procesov odločanja.

Podatki so na voljo v različnih oblikah. Glavne vrste podatkov so (slika 1.2):

1. **Kvantitativni (numerični) podatki** - podatki, ki so izraženi v številkah. Nadalje jih lahko delimo na diskretne in zvezne podatke. Diskretni podatki so podatki, ki imajo lahko samo določeno vrednost (npr. število zaposlenih). Neprekinjeni podatki imajo lahko neskončno število možnih vrednosti (npr. cena izdelka).
2. **Kvalitativni podatki** - to vrsto podatkov lahko razdelimo na kategorične in ordinalne podatke. Kategorični podatki predstavljajo besedilne podatke, ki jih je mogoče razvrstiti

v posamezne kategorije (npr. kraj rojstva, regija, kategorija izdelka), ordinalne podatke pa je mogoče razvrstiti ali urediti (npr. zadovoljstvo strank - zelo nezadovoljen, nezadovoljen, nevtralen, zadovoljen, zelo zadovoljen; stopnja izobrazbe - osnovna, srednja šola, dodiplomski študij, magisterij, doktorat).



Slika 1.2 Glavne vrste podatkov

Vir: Avtor: Avtor.

Podatki se delijo na strukturirane, polstrukturirane in nestrukturirane.

Strukturirane podatke je mogoče shranjevati, obdelovati in obdelovati v tradicionalnem sistemu za upravljanje relacijskih podatkovnih zbirk. Ti podatki prihajajo iz različnih virov, vključno s tokovi klikov, spletnimi obrazci, transakcijami na prodajnih mestih, senzorji in stroji. Ustvarjajo jih lahko ljudje ali stroji. Ti podatki imajo vnaprej določeno obliko, vrsto in organizacijo (EMC Education Services, 2015; Person & Porway, 2015).

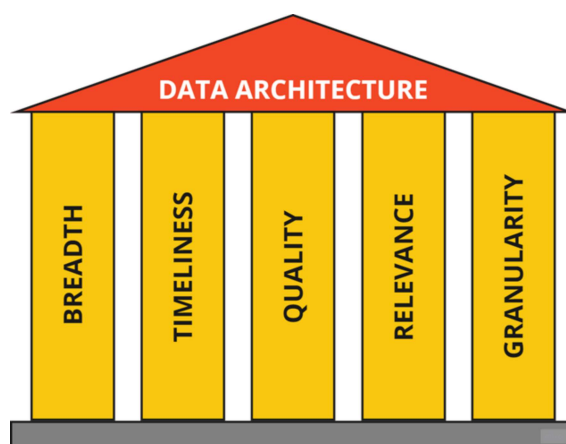
Polstrukturirani podatki so organizirani z oznakami, ki pomagajo podatkom dati hierarhijo in red, čeprav ne ustrezajo strukturiranemu sistemu podatkovne zbirke. Podatkovne zbirke in datotečni sistemi pogosto vsebujejo polstrukturirane podatke. Datoteke dnevnikov, besedilo z oznakami HTML, datoteke XML in podatkovne datoteke JSON so nekateri od možnih formatov za shranjevanje. (McKinsey Global Institute, 2011; Person & Porway, 2015).

Ker **nestrukturirani podatki** običajno nastanejo zaradi človeške dejavnosti in ne ustrezajo strukturirani obliki podatkovne zbirke, so v celoti nestrukturirani. Primeri te vrste podatkov so besedilni dokumenti, datoteke PDF, zapisi na spletnih dnevnikih, elektronska sporočila, slike in videoposnetki. (EMC Education Services, 2015; Person & Porway, 2015). Po mnenju Shermana (2015) je treba nestrukturirane in polstrukturirane podatke obravnavati drugače kot tradicionalne strukturirane podatke.

Tej veliki količini strukturiranih in zlasti nestrukturiranih podatkov, ki se ustvarjajo, zbirajo in obdelujejo z veliko hitrostjo in kompleksnostjo, pravimo veliki podatki. McKinsey Global Institute (2011) opredeljuje velike podatke kot "podatke, katerih obseg, porazdelitev, raznolikost in/ali pravočasnost zahtevajo uporabo novih tehničnih arhitektur in analitike za omogočanje vpogledov, ki odpirajo nove vire poslovne vrednosti". Kitchen in McArdle (2016) navajata, da je Doug Laney leta 2001 podrobno opisal, da so za velike podatke značilne tri lastnosti (Three V's):

- **obseg** (sestavljen iz ogromnih količin podatkov),
- **hitrost** (ustvarjena v realnem času),
- **tipi** (strukturirani, polstrukturirani in nestrukturirani).

Po Howsonu (2014) je temelj uspešnega poslovnega obveščanja podatkovna arhitektura (glej sliko 1.3), ki jo sestavlja šest pomembnih vidikov v zvezi s podatki: širina, pravočasnost, kakovost, ustreznost in zrnatost. Kakovost podatkov je osrednji steber, saj je veliko truda vloženega v zagotavljanje in izboljševanje kakovosti podatkov.



Slika 1.3 Podatkovna arhitektura kot temelj za uspešno BI

Vir: Avtor, prirejeno po Howson (2014).

Kot smo že omenili, se lahko podatki zbirajo iz različnih virov. To je povezano s širino podatkov kot enim od stebrov podatkovne arhitekture. Nanaša se na možnost pridobivanja več virov podatkov, kar je danes, v dobi velikih podatkov, običajen način zbiranja podatkov. Po drugi strani pa združevanje podatkov iz več različnih izvornih sistemov prav tako prispeva k težavam s kakovostjo podatkov (Howson, 2014).

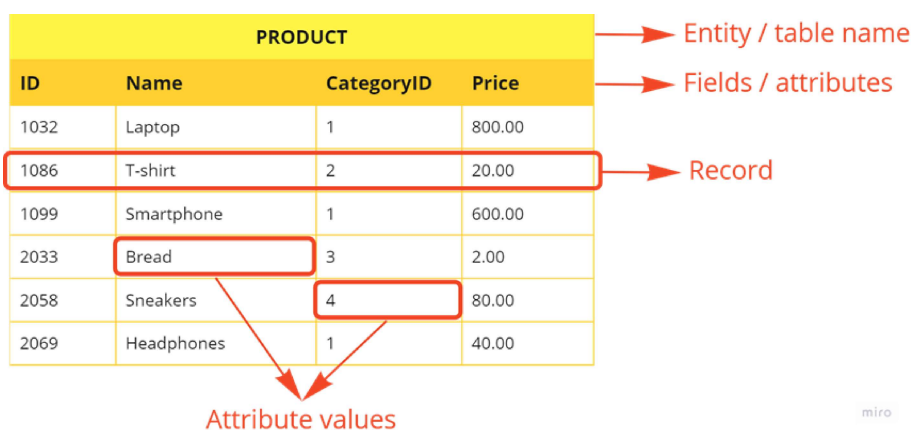
V naslednjem podpoglavju se bomo osredotočili na razumevanje podatkov in izkoriščanje njihovega potenciala. Raziskani bodo sistemi za upravljanje relacijskih podatkovnih zbirk

(RDBMS) za shranjevanje in iskanje podatkov ter vizualna predstavitev struktur podatkovnih zbirk z diagrami Entity- Relationship (ER). S povezovanjem razumevanja podatkovnih virov in podatkovnih vrst z modeliranjem in oblikovanjem podatkov bo narejen pomemben korak k praktični uporabi poslovne inteligence.

1.3. Modeliranje in oblikovanje podatkov

Podatkovni model je formalna predstavitev podatkov, ki jih poslovni sistem uporablja in ustvarja. (Dennis et al., 2018). Kot že omenjeno, so strukturirani podatki običajno shranjeni v sistemu za upravljanje relacijskih podatkovnih zbirk (RDBMS). Po Tilleyju (2020) je "sistem za upravljanje podatkovnih zbirk zbirka orodij, funkcij in vmesnikov, ki uporabnikom omogoča dodajanje, posodabljanje, upravljanje, dostop in analizo podatkov". Nekateri priljubljeni RDBMS so Oracle (Oracle), DB2 (IBM) in SQL Server (Microsoft).

V RDBMS so podatki organizirani v tabele, ki vsebujejo zbirko zapisov, v katerih so shranjene informacije o določeni entiteti. Tabele so predstavljene kot dvodimenzionalne strukture z navpičnimi stolpci in vodoravnimi vrsticami. Vsak stolpec predstavlja polje ali atribut entitete, medtem ko vsaka vrstica predstavlja zapis, ki je primerek entitete (Tilley, 2020). Na sliki 1.4 je prikazan primer tabele Izdelek.



PRODUCT			
ID	Name	CategoryID	Price
1032	Laptop	1	800.00
1086	T-shirt	2	20.00
1099	Smartphone	1	600.00
2033	Bread	3	2.00
2058	Sneakers	4	80.00
2069	Headphones	1	40.00

Slika 1.4 Primer tabele v RDBMS

Vir: Avtor: Avtor.

Ta tabela predstavlja preprost katalog izdelkov s 6 izdelki, od katerih ima vsak edinstven ID izdelka, ime, kategorijo in ceno.

Atribut je posebna vrsta podatkov o entiteti. Na primer, ID stranke, ime, priimek, naslov, poštna številka, mesto, država, e-poštni naslov so atributi entitete Customer. Vrstica v tabeli

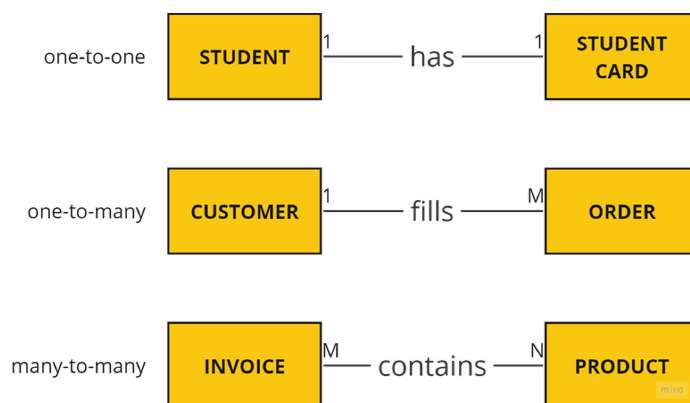


ali zbirka povezanih polj, ki opisujejo en primerek entitete (na primer stranke), se imenuje **zapis**.

Vsaka tabela v zbirki podatkov mora imeti atribut, ki služi kot **primarni ključ**. To je polje (ali niz polj), ki daje vsakemu izdelku v tabeli posebno identiteto. To pomeni, da v preglednici ne moreta biti dva izdelka z enakim ID.

Tabele v zbirki podatkov so pogosto povezane z drugimi tabelami v zbirki podatkov, kar pomeni, da med njimi obstaja povezava. Razmerja določajo, kako so podatki v eni tabeli povezani s podatki v drugi tabeli.

Po Tilleyju (2020) lahko med entitetami obstajajo tri vrste razmerij: ena-za-eno, ena-za-več in mnogo-za-več. Primeri teh razmerij so prikazani na sliki 1.5.



Slika 1.5 Primeri odnosov med entitetami

Vir: Avtor: Avtor.

Logično strukturo podatkovne zbirke in razmerja med tabelami je mogoče vizualno prikazati s **diagramom razmerij entitet (Entity Relationship Diagram - ERD)**.

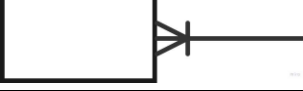
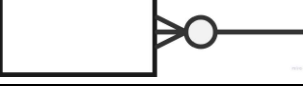
Za izdelavo ERD obstajajo različni zapisi, najpogostejša sta Chenov zapis in Martinov zapis (Crow's Foot). V tej knjigi bo predstavljen Martinov zapis.

V skladu z Martinovim zapisom (Crow's Foot) je entiteta predstavljena s pravokotnikom. To je lahko oseba, kraj, dogodek ali stvar, o kateri se zbirajo podatki. Atributi so navedeni kot samostalniki znotraj entitete. Odnosi med entitetami so prikazani s črtami, ki povezujejo entitete med seboj. Razmerja imajo kardinalnost, ki kaže, koliko primerkov ene entitete je povezanih s primerkom druge entitete (Dennis et al., 2018). V zapisu Crow's Foot so kardinalnosti prikazane z različnimi simboli. Na primer, enojna črtica označuje ena, dvojna



črtica označuje ena in samo ena, krog označuje nič, vranja noga pa veliko. V preglednici 1.1 so prikazani različni simboli za kardinalnost in njihov pomen.

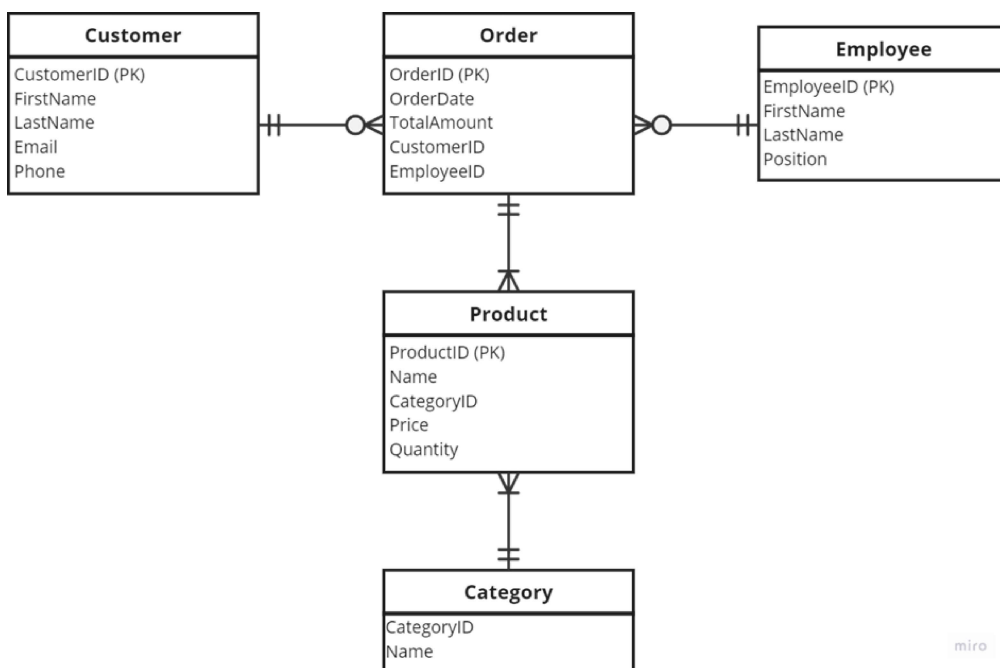
Preglednica 1.1 Primeri kardinalnosti

Simbol	Pomen
	Ena in edina
	En ali več
	Nič ali veliko
	Nič ali ena

Vir: Tilley (2020).

Dennis in drugi (2018) navajajo, da obstajajo trije koraki pri oblikovanju ERD:

1. Določite subjekte,
2. Dodajte attribute in dodelite primarne ključne,
3. Prepoznavajte odnose.



Slika 1.6 Primer ERD prodajnega sistema

Vir: Avtor: Avtor.



Slika 1.6 prikazuje del ERD prodajnega sistema. Vsaka entiteta v tem ERD je prikazana kot pravokotnik s seznamom njenih atributov znotraj. Črte, ki povezujejo entitete, so uporabljene za prikaz njihovih razmerij:

- Stranka lahko odda več naročil. Vsakemu naročilu je pripisana edinstvena stranka.
- Eno naročilo lahko vključuje več kot en izdelek. Vsak izdelek je povezan z določenim nakupom.
- Naročilo je povezano s stranko in zaposlenim, ki je naročilo obdelal.
- Izdelek je povezan s kategorijo, saj vsak izdelek pripada eni sami kategoriji.

Osnovno razumevanje modeliranja in oblikovanja podatkov, kot je razvidno iz diagramov ER in RDBMS, je bistvenega pomena za učinkovito uporabo podatkov. To razumevanje vzpostavlja temelje za prehod na pragmatično izvajanje podatkov prek podatkovno usmerjenega odločanja, pri katerem sistematično strukturirani podatkovni modeli zagotavljajo dragocene vpogleda, ki spodbujajo utemeljene poslovne strategije in postopke odločanja.

1.4. Sprejemanje odločitev na podlagi podatkov

Na podatkih temelječe odločanje (DDDM) je strateški pristop, ki temelji na analizi in razlagi podatkov za usmerjanje odločitev in ukrepov. Z uporabo vpogledov, pridobljenih iz obsežnih zbirk podatkov, lahko podjetja natančno krmarijo po negotovosti ter tako zmanjšajo tveganja in povečajo priložnosti. Nelson (2022) opredeljuje odločanje na podlagi podatkov kot proces sprejemanja strateških poslovnih odločitev, ki so v skladu z nameni, cilji in pobudami organizacije, z uporabo dejstev, metrik in podatkov. Odločanje, ki temelji na podatkih, je po Provost & Fawcett (2013) proces sprejemanja odločitev, ki se bolj kot na intuicijo opira na analizo podatkov. Na primer, tržnik lahko izbere oglase samo na podlagi svojega obsežnega znanja o panogi in dobrega občutka za to, kaj bo pritegnilo potrošnike. Druga možnost je, da svojo odločitev utemelji na analizi podatkov, ki kažejo, kako se potrošniki odzivajo na različne oglase.

Pri tem pristopu se odločitve ne sprejemajo na podlagi intuicije, temveč na podlagi trdnih dejstev. Vključuje zbiranje, analiziranje in razlaganje podatkov, da bi ugotovili vzorce, trende in korelacije. Ne glede na to, ali gre za optimizacijo operativne učinkovitosti, izboljšanje izkušenj strank ali izpopolnjevanje strategij izdelkov, odločitve, ki temeljijo na podatkih, podjetjem omogočajo, da se prilagodijo dinamičnim trgom in na njih uspevajo.



Z vključevanjem podatkov v proces odločanja postanejo organizacije bolj prilagodljive, odzivne in odporne na spremembe, s čimer spodbujajo inovacije in trajnostno rast. Veliko število raziskovalnih člankov je pokazalo, da je odločanje na podlagi podatkov povezano z večjo produktivnostjo (npr. Brynjolfsson in McElheran, 2019; Sala et al., 2022; Colombari et al., 2023). Zato večina organizacij, zlasti velikih, vlaga v zbiranje in analiziranje svojih podatkov. Več kot dve tretjini od več kot 300 vodstvenih delavcev, ki so sodelovali v raziskavi družbe Bain & Company (2017), pravi, da njihovo podjetje veliko vlaga v podatkovno analitiko, več kot polovica pa pričakuje transformacijske donose svojih naložb.

Obstaja pet korakov za sprejemanje odločitev, ki temeljijo na podatkih (Asana, 2022):

1. Razumevanje vizije podjetja,
2. Iskanje virov podatkov,
3. Čiščenje in urejanje podatkov,
4. Izvajanje analize podatkov,
5. Sklepanje.

McKinsey Global Institute (2014) poroča, da imajo organizacije, ki temeljijo na podatkih, 23-krat večjo verjetnost, da bodo pridobile stranke, 6-krat večjo verjetnost, da jih bodo obdržale, in 19-krat večjo verjetnost, da bodo dobičkonosne.

Svetovno priznana podjetja, ki svoje odločitve sprejemajo na podlagi podatkov, so Google, Amazon in Netflix.

Da bi organizacija v celoti izkoristila potencial poslovne inteligence, je treba upoštevati kakovost podatkov, kar je podrobneje opisano v naslednjem razdelku. Ta kakovost zagotavlja natančnost in zanesljivost zaključkov in vpogledov, pridobljenih iz podatkov.

1.5. Kakovost podatkov

Stopnja natančnosti, doslednosti, zanesljivosti in primernosti za določen namen se imenuje kakovost podatkov. Kakovost podatkov je pomembna v okviru poslovnega obveščanja in analize podatkov, saj so zaključki in sodbe, pridobljeni na podlagi podatkov, večinoma odvisni od njihove natančnosti in zanesljivosti. Slaba kakovost podatkov lahko privede do napačnih sklepov, zgrešenih strategij in na koncu do škodljivih poslovnih rezultatov. Po podatkih družbe Gartner (2021) slaba kakovost podatkov vsako leto organizacije v povprečju stane 12,9 milijona dolarjev.



Kakovost podatkov lahko opišemo s šestimi najpogosteje uporabljenimi dimenzijami (Foote, 2022):

- **Natančnost:** kako natančne so vrednosti atributov v podatkih?
- **Popolnost:** ali so podatki popolni, brez manjkajočih informacij?
- **Doslednost:** kako dosledne so vrednosti v podatkovnih zbirkah in med njimi?
- **Pravočasnost:** kako pravočasni so podatki?
- **Veljavnost:** kako so podatki skladni z vnaprej določenimi poslovnimi pravili?
- **Edinstvenost:** ali je vsak zapis enolično identificiran, brez odvečnega shranjevanja?

Po Shermanu (2015) se lahko podatki štejejo za visokokakovostne, če imajo naslednje značilnosti (pet C podatkov):

- **Čistost (Clean)** - nanaša se na manjkajoče elemente, neveljavne vnose in druge podobne težave.
- **Doslednost (Consistent)** - enotnost in skladnost podatkov v različnih virih in znotraj samega nabora podatkov.
- **Skladni (Conformed)** - nanašajo se na podatke, ki so skladni z vnaprej določenimi podatkovnimi standardi in pravili.
- **Aktualni (Current)** - za sprejemanje odločitev in analizo je treba uporabljati najnovejše podatke.
- **Celovit (Comprehensive)** - vključuje vse bistvene podatkovne elemente, ki so potrebni za predvideni postopek odločanja, ne da bi izpustili ključne informacije.

Kenett in Shmueli (2016) pravita, da je treba skoraj vse podatke očistiti, preden jih je mogoče uporabiti za nadaljnjo analizo. Vendar pa cilj določa stopnjo čiščenja in strategijo čiščenja podatkov. V istih podatkih so lahko visokokakovostne informacije za en namen in nizkokakovostne informacije za drug namen.

Gartner (2023) je nedavno opredelil niz 12 ukrepov za izboljšanje kakovosti podatkov. Ti ukrepi so bili razvrščeni v štiri različne kategorije, ki jih je treba upoštevati pri ocenjevanju celovitosti podatkov:

- Pri postavljanju trdnih temeljev se osredotočite na prave stvari,
- Uporaba odgovornosti za kakovost podatkov,
- Vzpostavitev kakovosti podatkov, ki ustreza namenu,
- Vključite kakovost podatkov v kulturo podjetja.



Po mnenju Howsona (2014) se dosledni, celoviti in natančni podatki štejejo za zelo kakovostne. Dobro kakovost podatkov je težko doseči, saj imajo velik vpliv organizacijske in lastniške težave.

Sposobnost razumevanja in analiziranja podatkov je ključnega pomena za sprejemanje premišljenih odločitev na področju poslovne inteligence. Razumevanje procesa od surovih podatkov do uporabnih vpogledov, ki vključuje različne vire podatkov, vrste, modeliranje in oblikovanje, je sestavni del procesa pridobivanja vrednosti iz informacij.

LITERATURA

1. Asana (2022). Data-driven decision making: A step-by-step guide [available at: <https://asana.com/resources/data-driven-decision-making>, access November 5, 2023]
2. Bain & Company (2017). Closing the Results Gap in Advanced Analytics: Lessons from the Front Lines [available at: <https://www.bain.com/insights/closing-the-results-gap-in-advanced-analytics-lessons-from-the-front-lines/>, access 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 [available at: <https://www.dataversity.net/the-data-information-knowledge-cycle/>, access 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 [available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3422397, access 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 [available at: <https://www.datacamp.com/cheat-sheet/the-data-information-knowledge-wisdom-pyramid>, access 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 [available at: <https://www.dataversity.net/data-quality-dimensions/>, access November 5, 2023]
13. Gartner (2021). How to Improve Your Data Quality [available at: <https://www.gartner.com/smarterwithgartner/how-to-improve-your-data-quality>, access November 5, 2023]
14. Gartner (2023). Gartner Identifies 12 Actions to Improve Data Quality [available at: <https://www.gartner.com/en/newsroom/press-releases/2023-05-22-gartner-identifies-12-actions-to-improve-data-quality>, access 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. [available at: <https://bernardmarr.com/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/>, access November 5, 2023]
19. McKinsey Global Institute (2011). Big data: The next frontier for innovation, competition, and productivity. [available at: 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, access November 5, 2023]



20. McKinsey Global Institute (2014). Five facts: How customer analytics boosts corporate performance [available at: <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/five-facts-how-customer-analytics-boosts-corporate-performance>, access November 5, 2023]
21. Nelson, M. (2022). Beyond The Buzzword: What Does Data-Driven Decision-Making Really Mean?. Forbes [available at: <https://www.forbes.com/sites/tableau/2022/09/23/beyond-the-buzzword-what-does-data-driven-decision-making-really-mean/?sh=2d35c4eb25d6>, access 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.